CRES Report Number TR/176

Advances in Real-Time Flood Forecasting

By Peter C. Young

Centre for Research on Environmental Systems & Statistics (CRES), Lancaster University, Lancaster LA1 4YQ, UK†

A shorter version of this report has been published (Young, 2002a)

The report discusses the modelling of rainfall-flow (rainfall-runoff) and flow routing processes in river systems within the context of real-time flood forecasting. It is argued that deterministic, reductionist (or 'bottom-up') models are inappropriate for real-time forecasting because of the inherent uncertainty that characterizes river catchment dynamics and the problems of model over-parametrization. The advantages of alternative, efficiently parameterized *Data-Based Mechanistic* (DBM) models, identified and estimated using statistical methods, are discussed. It is shown that such models are in an ideal form for incorporation in a real-time, adaptive data assimilation and forecasting system based on recursive state space estimation (an adaptive version of the stochastic *Kalman Filter* algorithm). An illustrative example, based on the analysis of a limited set of hourly rainfall-flow data from the River Hodder in NW England, demonstrates the utility of this methodology and illustrates the advantages of incorporating real-time state and parameter adaption.

Keywords: rainfall-flow processes, data-based mechanistic models, recursive estimation, Kalman filter, real-time forecasting, parameter adaptation, heteroscedasticity, variance adaptation

† Also, Centre for Resource & Environmental Studies, Institute of Advanced Studies, Australian National University, Canberra ACT 2020, Australia.

TEX Paper

1. Introduction

The primary objective of this paper is to describe recent research on the design of flood forecasting procedures; procedures that can be applied to the problem of predicting future flow levels and, therefore, future flood events in river systems. The aim of this research is to produce an on-line, real-time approach to flood forecasting that is inherently stochastic and so able to predict not only the likely level of future flow, but also the uncertainty associated with this prediction. In this manner, the probability of a flood occurring in the near future is quantified and this additional information can then be used as a basis for decision-making and operational management in flood-prone locations.

The paper briefly outlines the problem of modelling the rainfall-flow processes and the use of such models in flow forecasting system design. It then goes on to describe, in some detail, one particular approach to the problem that includes two main components. First, stochastic modelling based on the statistical identification and estimation of physically-meaningful, nonlinear, transfer function models. Second, an adaptive *Kalman Filter* forecasting algorithm based on a stochastic state space formulation of these models.

This approach has the virtue of being inherently stochastic and, because it is formulated in Bayesian, recursive estimation terms (e.g. Kalman, 1960; Bryson and Ho, 1969; Young, 1984), it provides an ideal basis for real-time implementation and the introduction of *adaptive* procedures. Such adaption is motivated by a view that the rainfall-flow and riverine flow processes are inherently 'nonstationary': i.e. no completely fixed model with constant parameters will be able to characterize the catchment behaviour for all times into the future. As a result, the forecasting system should be based on models that are able to adjust to any, normally small, changes in the catchment behaviour not predicted accurately enough by the initially estimated model.

The paper has another, underlying objective that is of deeper philosophical and methodological significance and is, in part, a response to the recent increased interest in the so-called 'top-down' (or 'holistic') approach to modelling hydrological systems (e.g. Jothityangkoon *et al.*, 2001, which follows from the earlier contributions of Klemes, 1983). Interest in top-down modelling has been revived largely because the alternative 'bottom-up' or 'reductionist' philosophy that dominated much research during the last century, has failed to solve the many problems of modelling natural environmental systems. Top-down modelling in hydrology has its parallels in the environmental (e.g. Young, 1978, 1983; Beck, 1983) and ecosystems (e.g. Silvert,1993 & the prior references therein) literature of the 1970s and early 1980s, where the present author's contributions were presented within the context of 'badly defined' environmental systems.

To quote from Young (1978):

> "Many of the modeling problems that have arisen in connection with large and complex natural systems, such as those met in environmental research, can be attributed in part to the 'badly defined' nature of such systems. This poor definition arises for two major reasons. First, the size and complexity of the systems are such that the mechanisms which govern the change in the observed system variables and their relationships one to another are rarely fully understood *a priori*. There can, in other words, be a basic ambiguity, a situ-

ation in which a number of possible explanations for the observed behaviour seem feasible but where there exists little *a priori* evidence as to which of these explanations seems most plausible. ...

... This problem is exacerbated by a tendency for most simulation modelling methodology to be based on a "reductionist" philosophy. Here the system is repetitively sub- divided into elemental components which are assumed to have physical significance to the modeller and can be analyzed as relatively separate entities. Having separately evaluated the 'physical' parameters associated with each of the elemental models (such as dispersion or advection coefficients in hydrological systems analysis), usually by experimentation either *in situ* or in a laboratory, the modeler then re-assembles the model components in a manner which he and his advisors perceive to be appropriate.

This reductionist philosophy can be contrasted with the alternative "holistic" approach (see e.g. Rigler, 1976), in which the model is obtained from a study of the intact system. It can and, indeed, it normally does, include an appraisal of the components of the system (even small components). But as long as the measurements on which the model building is based are made *in situ* and the model is statistically assessed against these measurements as a single entity, then the approach is holistic. Having established such an holistic, empirical model (or theory) of behaviour, however, the holistic modeler usually attempts to falsify his model and, if necessary to search for a more satisfactory explanatory model, usually by recourse to sophisticated statistical methodology.

The limitations of the reductionist approach have been emphasized by Herbert Simon (1967) who says that a complex system is "one made up of a large number of parts that interact in a non-simple way. In such systems, the whole is more than the sum of the parts, not in an ultimate metaphysical sense, but in the important pragmatic sense that, given the properties of the parts and their laws of interactions, it is not a trivial matter to infer the properties of the whole. In the face of complexity an in-principle reductionist may be at the same time a pragmatic holist". And the situation is even worse in the case of badly defined systems since we are usually far from certain about the nature of the interactions.

The main purpose of the present paper is to describe a new holistic approach to the problem of modeling badly defined dynamic systems; ... ... It is suggested here that good modeling will be strongly objective orientated and that, in the area of environmental (and indeed socio-economic) systems analysis, this objective is normally linked with problems of control or management. The new approach is, therefore, specifically designed for such control and management applications and entails the formulation of a linear but possibly time-variable parameter 'estimation model'. This model, which is obtained directly from a simple 'control model' of the system, is aimed specifically at describing the 'dominant modes of behaviour' of the system in as simple and parametrically efficient manner as possible."

And again (Young, 1983):

"Such a reductionist approach is rarely, however, accompanied by sufficient evaluation of the resulting model as a complete entity. "Holistic" validation is normally restricted to exercises in deterministic 'model fitting', in which overall 'calibration' of the model is achieved using manual or automatic methods of

> parameter 'tuning' or optimization; an approach that is sometimes enhanced by deterministic sensitivity analysis, in which the sensitivity of the model outputs to variations in the parameters is examined using various analytic procedures (see, for example, Miller et al., 1976).
>
> Although such analysis is perfectly respectable, it must be used very carefully; the dangers inherent in its application are manifold, but they are not, unfortunately, always acknowledged by its proponents. It is well known that a large and complex simulation model, of the kind that abounds in current ecological and environmental system analysis, has enormous explanatory potential and can usually be fitted easily to the meager time-series data often used as the basis for such analysis. Yet even deterministic sensitivity analysis will reveal the limitation of the resulting model: many of the 'estimated' parameters are found to be ill-defined and only a comparatively small subset is important in explaining the observed system behavior.
>
> Of course, over-parameterization is quite often acknowledged, albeit implicitly, by the reductionist simulation model-builder. Realizing the excessive degrees of freedom available for fitting the model to the data, he will often fix the values of certain 'better known' parameters and then seek to fit the model by optimizing the chosen cost function (usually the sum of the squares of the difference between the model outputs and the observations) in relation to the remaining parameters, which are normally few in number. In this manner, the analyst ensures that the cost function-parameter hypersurface is dominated by a clearly defined optimum (a minimum in the least-squares case), so that estimation of the parameters which define the optimum becomes more straightforward.
>
> But what is the value of this optimization exercise in relation to the specification of the overall model? Clearly a lower-dimensional parameter space has been located which allows for the estimation of a unique set of parameter values. However, this has been obtained only at the cost of constraining the other model parameters to fixed values that are assumed to be known perfectly and are defined in relation to the analyst's prior knowledge of the system. As a result, the model has a degree of 'surplus content' not estimated from the available data, but based on a somewhat *ad hoc* evaluation of all available prior knowledge of the system and coloured by the analyst's preconceived notions of its behavioral mechanisms.
>
> On the surface, this conventional simulation-modeling approach seems quite sensible: for example, the statistician with a Bayesian turn of mind might welcome its tendency to make use of all *a priori* information available about the system in order to derive the *a posteriori* model structure and parameters. On the other hand, he would probably be concerned that the chosen procedures could so easily be misused: whereas the constrained parameter optimization represents a quantitative and relatively objective approach, it is submerged rather arbitrarily within a more qualitative and subjective framework based on a mixture of academic judgment and intuition. Such a statistician would enquire, therefore, whether it is not possible to modify this framework so that the analyst cannot, unwittingly, put too much confidence in *a priori* perceptions of the system and so generate overconfidence in the resulting model."

These early papers rejected the idea of 'deterministic reductionism'; i.e the widely held view that a model can be constructed on the basis of deterministic

equations based on the modeller's perception of the physical system. And they presented initial thoughts on a more objective, statistical approach to modelling poorly defined systems of all kinds. This approach is much in sympathy with the tenets of top-down modelling but it is rather different in its methodological basis. The papers also adumbrated very similar anti-reductionist arguments that have appeared recently in the hydrological literature and express some of these same views within a hydrological context (Jakeman & Hornberger, 1993, Beven, 2000). Quite similar anti-reductionist views are also appearing in other areas of science: for instance, in a recent lecture (Lawton, 2001), the current chief executive of the *Natural Environment Research Council* (NERC) recounted the virtues of the top-down approach to modelling ecological systems (although, for some reason, he did not appear to accept that such reasoning could also be applied to other natural systems, such as the physical environment).

In the twenty years since the earlier papers on top-down modelling were published, the author has sought to develop his approach within a more rigorous statistico-systems setting that he has called *Data-Based Mechanistic* (DBM) modelling (this term is first used in Young & Lees, 1993 although it follows directly from Young & Minchin, 1991). Underlying the main topic of the present paper, therefore, is a desire to promulgate the idea of 'inductive' DBM modelling as an alternative to the 'hypothetico-deductive' (and often reductionist) approach that has dominated much scientific modelling research over the last century. Other recent publications that have concentrated more centrally on this topic and can be considered as adjuncts to the present paper, in this more general regard, are: Young (1998a,b; 1999a); Young & Pedregal (1998,1999a); Young & Parkinson (2002); Young *et al.*(1996); Shackley *et al.*(1998); Parkinson & Young, (1999).

2. Rainfall-Flow Modelling

It is possible to design flood forecasting systems for river catchments without the *explicit* identification and estimation of physically meaningful rainfall-flow (rainfall-runoff) and flow routing (flow-flow) models. In practical engineering terms, however, it is often an advantage if the end-user understands the nature of the forecasting algorithm, so that the overt presence of such models helps to engender confidence in the nature of the resulting design. Also, such model construction is an essential component in the DBM modelling procedures considered later in this paper.

Characterization of the nonlinear dynamic relationship between rainfall and river flow is one of the most interesting modelling problems in hydrology. It has received considerable attention over the past thirty years, with mathematical and computer-based models ranging from simple black-box representations to complex, physically-based catchment models. It would be impossible to review this enormous literature here. Fortunately, however, there are many books available that deal in whole, or in part, with this challenging area of science and engineering. Useful texts of this type are: Anderson & Burt (1985); Shaw (1994); Singh (1995); and Beven (2001). The latter book, in particular, provides a clearly written review of the whole topic that not only deals critically with many recent developments but also provides an excellent introduction to the subject at the start of the twenty-first Century. In addition, two recent reports by the UK Environment Agency (Moore & Bell, 2000; Carrington *et al.*, 2000: see also Moore *et al.*, 2000) are of considerable

importance in both reviewing and comparing rainfall-flow models within the real-time forecasting context. Unfortunately, as the authors point out, only a limited sub-set of *Transfer Function* (TF) models, of the type discussed in the present report, were considered (isolated event-mode, linear TF models), so the comparative results are not particularly relevant to the present paper.

Wheater *et al.* (1993) have categorized rainfall-flow models into the following four, broad types.

- *Metric Models*, which are based primarily on observational data and seek to characterize the flow response largely on the basis of these data, using some form of statistical estimation or optimization (e.g. Wood & O'Connell, 1985; Young, 1986). These include purely black-box, time-series models, such as discrete and continuous-time transfer functions, neural network and neuro-fuzzy representations (e.g Jang *et al.*, 1997). Often, such models derive from, or can be related to, the earlier unit hydrograph theory but this is not always recognized overtly.

- *Conceptual Models*, which vary considerably in complexity but are normally based on the representation of internal storages, as in the original Stanford Watershed Model of the nineteen sixties (Crawford & Linsley, 1966). However, assumptions about catchment-scale response are not often included explicitly, notable exceptions being TOPMODEL (Beven & Kirkby, 1979) and the ARNO model (Todini, 1996). The essential feature of all these models, however, is that the model structure is specified *a priori*, based on the hydrologist/modeller's perception of the relative importance of the component processes at work in the catchment; and then an attempt is made to optimize the model parameters in some manner by calibration against the available rainfall and flow data.

- *Physics-Based Models*, in which the component processes within the models are represented in a more classical, mathematical-physics form, based on continuum mechanics solved in an approximate manner via finite difference or finite element spatio-temporal discretization methods. A well known example is the *Systéme Hydrologique Européen* (SHE) model (e.g. Abbot *et al.*, 1986). The main problems with such models, which they share to some degree with the larger conceptual models, are two-fold: first, the inability to measure soil physical properties at the scale of the discretization unit, particularly in relation to sub-surface processes; and second, their complexity and consequent high dimensional parametrization. This latter problem makes objective optimization and calibration virtually impossible, since the model is normally so over-parameterized that the parameter values cannot be uniquely identified and estimated against the available data (see below).

- *Hybrid Metric-Conceptual* (HMC) Models, in which (normally quite simple) conceptual models are identified and estimated against the available data to test hypotheses about the structure of catchment-scale hydrological storages and processes. In a very real sense, these models are an attempt to combine the ability of metric models to efficiently characterize the observational data in statistical terms (the 'principle of parsimony' (Box & Jenkins, 1970); or

the 'Occam's Razor' of antiquity), with the advantages of conceptual models that have a prescribed physical interpretation within the current scientific paradigm.

The models in the two middle categories, above, are often characterized by a large number of unknown parameters that need to be estimated ('optimized' or 'calibrated') in some manner against the observational rainfall-flow time series. Because the number of parameters is normally very large in relation to the information content of the data, however, such models are often over-parameterized and not normally identifiable, in the sense that it is impossible to estimate their parameters uniquely without imposing prior restrictions on a large subset of the parameter values prior to estimation (see e.g. Young *et al.*, 1996). The author and his co-workers have addressed these problems of over-parameterization and poor identifiability associated with large environmental models many times over the past quarter century (see previous references in §1). And recently, Beven and his co-workers (e.g. Franks *et al*, 1997) have revisited this idea within the hydrological context, using the term 'equifinality' rather than 'non-identifiability' to describe the consequences of such over-parametrization: namely the existence of many different parametrizations and model structures that are all able to explain the observed data equally well, so that no unique representation of the data can be obtained within the prescribed model set.

There appear to be two main reasons for these identifiability problems. First, any limitations of the observational data can be important, since the available time series may not be sufficiently informative to allow for the estimation of a uniquely identifiable model form. In particular, the inputs to a system may not be 'sufficiently exciting' (see e.g. Young, 1984), in the sense that they do not perturb the system sufficiently to allow for unambiguous estimation of all the model parameters within an otherwise identifiable model structure. Secondly, even if the input does sufficiently excite the system, there are usually only a limited number of dynamic modes - the *dominant modes* of the system - that are excited to any significant extent; and the observed output of the system is dominated by their cumulative effect.

The importance of this dominant mode concept in model identification and estimation is illustrated by appendix 1 of Young (2001b), which shows how the response of a 26th order hydrological simulation model can be duplicated with exceptional accuracy (0.001% error by variance) by a much simpler 7th order dominant mode model. This is typical of most high order linear systems and appears to carry over to nonlinear systems. For example, Young *et al.*(1996), Young (1998b) and Young and Parkinson (2002) have used similar analysis to show how the response of high dimensional, nonlinear global carbon cycle simulation models are accurately reproduced by differential equation models of much reduced order. This is also reflected in other recent work on the simplification of global climate models (Hasselmann *et al.*, 1997; Hasselmann, 1998).

In the above references, the author has stressed that dominant modal behaviour is a generic property of dynamic systems and that it is probably the main reason for the limitation on the number of clearly identifiable parameters that can be estimated from observational data. The largest order identifiable system that the author has encountered from the analysis of real time series data, over the past

forty years, is a 10th order differential equation model for the vibrations in a man-made and specially designed cantilever beam (Young, 1998a), where the design and associated very low damping of the system results in four dominant, complex modes and the resulting model explains 99.74% of the experimental data.

However, the identifiable order is normally much lower than this for natural systems, and many previous rainfall-runoff modelling studies (e.g. Kirkby, 1976; Hornberger *et al.*, 1985; Jakeman & Hornberger, 1993; Young, 1993, 1998b; Young & Beven, 1994; Young *et al.*, 1997a,b; Ye *et al.*, 1998) suggest that a typical set of rainfall-runoff observations contain only sufficient information to estimate up to a maximum of six parameters within simple, nonlinear dynamic models of dynamic order three or less. In the rainfall-runoff example discussed later, for instance, there is clear evidence in the data of only two dominant modes between the effective rainfall input and the flow response (as described by a second order transfer function model with only four parameters): a 'quick' mode with a residence time (time constant) of a few hours; and a 'slow' mode, with a residence time of many hours.

By their very nature, both the metric and HMC approaches avoid many of these 'large model' problems. As a result, they provide a potentially attractive vehicle for real-time flood forecasting: they can be justified well in statistical terms and they are inherently simple in both structure and application. Such simplicity means that the forecasting system can be more easily optimized on a regular basis in order to ensure near-optimal performance. And, as we see later, it facilitates the incorporation of advanced features such as on-line state and parameter adaption. Of the two approaches, however, the attractiveness and practical utility of the basic metric model as a vehicle for flood forecasting is marred by its lack of any clearly defined internal physical interpretation. For instance, neural network (e.g. Tokar & Johnson,1999) and neuro-fuzzy models have attracted a great deal of attention in recent years but they are the epitome of black box modelling, revealing very little of their internal structure that has any physical meaning (see the discussion in Young (2001c) on the paper by Hu *et al.*(2001) where a neuro-fuzzy model with 102 parameters can be replaced by a nonlinear TF model with only 15 parameters if the internal structure of the model is identified and taken into consideration). For this reason, many hydrologists tend to mistrust such a black box model as a basis for something as important as flood forecasting. Moreover, their lack of any obvious internal physical meaning means that metric models are difficult to interrogate and diagnose when errors are encountered. HMC models, on the other hand, do not suffer from these problems and, indeed, are often simpler in dynamic terms than the metric model.

Within the category of HMC models two main approaches to 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 (1620), to Karl Popper (1959) and Thomas Kuhn (1962). In the first hypothetico-deductive approach, the *a priori* conceptual model structure is effectively a theory of hydrological behaviour based on the perception of the hydrologist/modeller and is strongly conditioned by assumptions that derive from current hydrological paradigms (e.g. the IHACRES model of Jakeman *et al.*, 1990). The alternative *Data-Based Mechanistic* (DBM) approach is basically inductive, in the sense that it tries to avoid theoretical preconceptions 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 hydrological paradigm: e.g. the models of rainfall-flow data in Young (1993, 1998b), Young & Beven (1994) and Young *et al.*(1997a).

This physical interpretation is an essential element in all DBM modelling: no matter how well the DBM model explains the data, it is only considered truly credible if it can be interpreted in physically meaningful terms. In this, the DBM approach harks back to the father of modern statistical inference, Karl Friedrich Gauss, who held that no hypothesis was satisfactory which rested on a formula and was not also a consequence of physical conjecture. For this reason, Gauss abandoned his work on the attraction between charged particles because he was unable to find a plausible physical interpretation of the formula he had obtained for the relationship between the relative motion and position of two particles.

Since the DBM approach is inductive, it is not wedded as strongly to the current paradigms as the hypothetico-deductive approach: indeed, its intention is always to respect these paradigms but not allow them to dictate the structure of models (here rainfall-flow) if the data suggest otherwise. In other words and to use a Kuhnian interpretation of science, the DBM approach encourages the continual questioning of current paradigms and rejoices in its ability to promote paradigm change if this is supported by observational data. Examples of this ability to promote paradigm change in an evolutionary manner, based primarily on the statistical analysis of data, are the development of the *Aggregated Dead Zone* (ADZ) model for solute transport in rivers (e.g. Beer & Young, 1983; Wallis *et al.*, 1989; Young & Wallis, 1994); and recent research on modelling the relationship between government spending, private capital investment and unemployment in the USA during the last half century (Young & Pedregal, 1998, 1999a).

Another important aspect of the DBM approach to rainfall-flow modelling relates to the objectives of the modelling exercise in each case. In the author's opinion, the search for a single, all encompassing model of any dynamic system is futile. Rather, the model builder should be seeking a model that suits the nature of the study objectives. Of course, this objective orientation does not have to be precisely defined, since a model can simultaneously serve more than one purpose. But even more loosely defined objectives need to be considered carefully before the modelling exercise begins. In the present context, the primary objective is to obtain DBM models that perform well in a flood forecasting and warning context.

3. Statistical Identification, Estimation and Validation

The statistical approach to modelling assumes that the model is stochastic: in other words, no matter how good the model and how low the noise on the observational data happens to be, a certain level of uncertainty will remain after modelling has been completed. Consequently, full stochastic modelling requires that this uncertainty, which is associated with both the model parameters and the stochastic inputs, should be quantified in some manner as an inherent part of the modelling analysis. This statistical approach involves three main stages: *identification* of an appropriate, identifiable model structure; *estimation* (optimization, calibration) of

the parameters that characterize this structure, using some form of estimation or optimization; and *conditional predictive validation* of the model on data sets different to those used in the model identification and estimation. In this section, we consider these three stages in order to set the scene for the later analysis. This discussion is intentionally brief, however, since the topic is so large that a comprehensive review is not possible in the present context.

(*a*) *Structure and Order Identification*

In the hypothetico-deductive approach to model building, the model constitutes a hypothesis or theory of behaviour and it is normally selected beforehand, based on the current scientific paradigm. However, the subsequent processes of model estimation and validation are often considered as exercises in Popperian falsification (Popper, 1959) and so the initial perceived model structure may well be modified in the light of these. In the DBM approach, this questioning of the hypothetical model is more overt and the identification stage is considered as a most important and essential prelude to the later stages of model building. Nevertheless, in the case of HMC models, both approaches make use of statistical identification procedures to some extent. These usually involve the identification of the most appropriate model order, as defined in dynamic system terms, although the model structure itself can be the subject of the analysis if this is also considered to be ill-defined. In the DBM approach, for instance, the nature of linearity and nonlinearity in the model is not assumed *a priori* (unless there are good reasons for such assumptions based on previous data-based modelling studies) but is identified from the data using non-parametric and parametric statistical estimation methods.

This important 'identification' stage means the application of objective statistical methods to determine the *dynamic* model order and structure. Within the hydrological 'top-down' context, for example, it is related directly to problems such as the definition of how many storage zones (conceptual 'buckets') are required to characterize the data *at the scale of interest*; and how these sub-models are interconnected (i.e. in series, parallel or feedback arrangements). It must be stressed, however, that such problems arise mainly from the specification of the *dynamic* model order (i.e. the order of the differential equations that are used to describe the major rainfall-flow dynamics; or equivalently, here, the number of storage zones). So a parsimonious model, in this important dynamic sense, is one that has a lowest dynamic order that is consistent with the information content in the data and whose parameters are statistically significant.

Of course, the DBM model may well have *other* parameters that are not associated primarily with the dynamic order of the model and so are not so important in identifiability terms: for instance, coefficients that parameterize any nonlinearity in the system (see later). Here again, however, the presence of such parameters in the model should be justified by whether or not they are statistically significant. The statistical significance of parameter estimates can be evaluated by conventional statistical tests or, in these days of the fast digital computer, by Monte Carlo simulation and sensitivity analysis (see e.g. chapter 6 in Saltelli *et al.*, 2000 and chapter 7 in Beven, 2001). As we shall see, for example, such analysis is very useful for assessing the uncertainty associated with 'derived', physically interpretable parameters that are computed from the estimated model parameters, rather than being esti-

mated directly. In the present hydrological context, these include parameters such as residence times and flow partitioning percentages associated with the inferred catchment storage dynamics.

Once a suitable model structure has been defined, there are a variety of statistical methods for identifying model order. Fitting criteria, such as the coefficient of determination† (R_T^2: see later) based on the model errors, can be very misleading if used on their own, since over-parameterized models can 'over-fit' the data. In general, therefore, it is necessary to exploit some specific order identification statistics, such as: the correlation-based statistics popularized by Box & Jenkins (1970); the well known Akaike Information Criterion (AIC: Akaike, 1974); and the YIC criterion proposed by the present author (Young, 1989). In all cases, the objective is to avoid over-parametrization by identifying a model structure and order that explains the data well within a minimal parametrization: i.e. 'parsimonious models' (Box & Jenkins, 1970). The time series methods used for model order identification in the present report are outlined in Young and Lees (1993), Young and Beven (1994), Young et al. (1996) and Young and Parkinson (2002).

(*b*) *Estimation (Optimization)*

Once the model structure and order have been identified, the parameters that characterize this structure need to be estimated in some manner. There are many automatic methods of estimation or optimization available in this age of the digital computer, from the simplest, deterministic procedures, usually based on the minimization of least squares cost functions; to more complex numerical optimization methods based on statistical concepts, such as Maximum Likelihood (ML). In general, the latter are more restricted, because of their underlying statistical assumptions, but they provide a more thoughtful and reliable approach to statistical inference; an approach which, when used correctly, includes the associated statistical diagnostic tests that are considered so important in statistical inference. Moreover, the power of the modern computer is such that some of these restrictions are gradually being lifted, with the advent of stochastic approaches, such as numerical Bayesian methods that exploit *Monte Carlo Simulation* (MCS) methods (see later). In the present context, however, the estimation methods are based on special *Instrumental Variable* (IV) methods that are formulated within a ML context but do not require such strong assumptions about the nature of the noise processes (e.g. Young, 1984 and the references therein).

(*c*) *Conditional Predictive Validation*

Validation is a complex process and even its definition is controversial. Some academics (e.g. Konikow & Brederhoeft, 1992, within a ground-water context; Oreskes *et al.*, 1994, in relation to the whole of the earth sciences) question even the possibility of validating models. To some degree, however, these latter arguments are rather philosophical and linked, in part, to questions of semantics: what is the 'truth'? What is meant by terms such as validation, verification and confirmation? etc. Nevertheless, one specific, quantitative aspect of validation is widely accepted;

† often termed the 'Nash-Sutcliffe efficiency' in the hydrological literature (Nash & Sutcliffe, 1970)

namely predictive validation, in which the predictive potential of the model is evaluated on data other than that used in the identification and estimation stages of the analysis.

It appears normal these days to follow the Popperian view of validation (Popper, 1959) and consider it as a continuing process of falsification. Here, it is assumed that scientific theories (models in the present context) can never be proven universally true; rather, they are not yet proven to be false. It seems reasonable to consider that this yields a model that is considered conditionally valid, in the sense that it can be assumed to represent the best theory of behaviour currently available that has not yet been falsified. Thus, conditional predictive validation means that the model has proven valid in this more narrow predictive sense. In the rainfall-flow context, for example, it implies that, on the basis of the new measurements of the model inputs (e.g. rainfall, temperature or evaporation) from the validation data set, the model produces flow predictions that are acceptable within the uncertainty bounds associated with the model. Note this stress on the question of the inherent uncertainty in the estimated model: one advantage of statistical estimation, of the kind considered in this chapter, is that the level of uncertainty associated with the model parameters and the stochastic inputs is quantified in the time series analysis. Consequently, the modeller should not be looking for perfect predictability (which no-one expects anyway) but predictability which is consistent with the quantified uncertainty associated with the model.

4. Data-Based Mechanistic (DBM) Modelling

Previous publications (Beck & Young, 1975; Whitehead & Young, 1975; Young, 1978, 1983, 1992, 1993, 1998a,b; Young & Minchin, 1991; Young & Lees, 1993; Young & Beven, 1994; Young *et al.*, 1996; Young & Pedregal, 1998, 1999a; Young and Parkinson, 2002) map the evolution of the DBM philosophy and its methodological underpinning in considerable detail, and so it will suffice here to merely outline the main aspects of the approach.

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 of linear TF 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 characterize the identified model structure in step 2. are estimated using advanced methods of statistical estimation for dynamic systems. The methods used in the present report are the *Refined Instrumental Variable* (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 & Jakeman (1979, 1980); Jakeman & Young (1979); Young, (1984, 1985). They are also outlined in Young & Beven (1994), Young *et al.*(1996) and Young and Parkinson (2002).

4. If significant parameter variation is detected then the model parameters are estimated by the application of an approach to time (or state) dependent parameter estimation based on recursive Fixed Interval Smoothing (FIS): e.g. Bryson & Ho (1969); Young (1984, 2000). Such parameter variation will tend to reflect nonstationary and nonlinear aspects of the observed system behaviour. In effect, the FIS algorithm provides a method of non-parametric estimation, with the *Time Variable Parameter* (TVP) estimates defining the non-parametric relationship, which then can often be interpreted in nonlinear *State-Dependent Parameter* (SDP) terms (see Young, 1993; Young & Beven, 1994; Young, 1998a, 2000, 2001a; Young *et al.*, 2001).

5. If nonlinear phenomena have been detected and identified in stage 4, the non-parametric state dependent relationships are normally parameterized in a finite form and the resulting nonlinear model is estimated using some form of numerical optimization, such as nonlinear least squares or ML based on prediction error decomposition (Schweppe, 1965). In the present report, this approach to nonlinear identification and estimation is required only to define the nature of the effective rainfall nonlinearity, which appears only at the input to the model, as described in subsequent sections.

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 in the sense discussed above. This involves standard statistical diagnostic tests for stochastic, dynamic models, including analysis which ensures that the nonlinear effects have been modelled adequately (e.g. Billings & Voon, 1986), as well as exercises in predictive validation and stochastic sensitivity analysis.

One aspect of the above DBM approach which differentiates it from alternative deterministic top-down approaches is its inherently stochastic nature. This means that the uncertainty in the estimated model is always quantified and this information can then be utilized in various ways. For instance, it allows for the application of Monte Carlo-based uncertainty and sensitivity analysis, as well as the use of the model in statistical forecasting and data assimilation algorithms, such as the Kalman Filter. The uncertainty analysis is particularly useful because it is able to evaluate how the covariance properties of the parameter estimates affect the probability distributions of physically meaningful, derived parameters, such as residence

times and partition percentages in parallel hydrological pathways (see e.g. Young, 1992, 1999a and the example below).

5. Transfer Function (TF) Modelling: Historical Background

TF models are one aspect of time series analysis and there have been many publications on the application of time series methods to rainfall-flow modelling and forecasting, much too numerous to review in this paper. The papers in Wood (1985) are a good reference to the state of the art at that time (e.g. Young and Wallis, 1985) and the contemporaneous chapter by Wood & O'Connell (1985) on real-time forecasting provides a good introduction to the Kalman Filter formulated within a real-time forecasting context. Another important over-view of stochastic hydrology is Yevjevich (1987). A later review, concerned mainly with linear TF models, is given by Cluckie (1993).

TF modelling originally derives from the Systems and Control literature, where it has been used for over half a century as a major tool in modelling and control system design for linear dynamic systems. TF models also have an obvious appeal in hydrological terms, since the unit impulse response of the TF is an amplitude-scaled equivalent of the hydrological *Instantaneous Unit Hydrograph* (IUH). As a result, TF models were quickly assimilated into hydrological research and have figured prominently in the hydrological literature for many years. Early examples are Dooge (1959) and Nash (1959), the latter introducing the now well known 'Nash Cascade', which is a chain of first order transfer functions used for flow routing. Since then, there have been many references to TF models in the hydrological literature, again too numerous to review here. The present author (Young, 1986) interpreted existing flow routing models in transfer function terms, showing how they could be recursively estimated and used for flow forecasting purposes.

While useful for modelling flow processes in river channels, an early application of TF modelling to rainfall-flow data (Young, 1974) demonstrated that linear TF models could only characterize rainfall-flow dynamics in the short term, as a description of the dynamics associated with *individual* storm events. However, if the input (numerator) parameters of the TF were allowed to vary, the model could then capture the effects of temporal changes in the catchment soil-water storage and modify the rainfall-runoff behaviour accordingly. When combined with methods of recursive estimation (e.g. Young, 1974, 1984), such TVP models could then form the basis for *Parameter-Adaptive* flood forecasting procedures (see Cluckie, 1993; Lees *et al.*, 1994).

The TVP model in Young (1974) led quickly to the formulation of the nonlinear 'Bedford-Ouse' model (BM model e.g. Whitehead & Young, 1975; Young, 2001b). This consists of two components connected in series: an *effective rainfall* (sometimes erroneously referred to as 'rainfall excess') nonlinearity, which accounts for the catchment storage effects and helps to remove the requirement for the time variable parameters; and a constant parameter, linear TF, which models the underlying IUH dynamics. This special type of model (known as a 'Hammerstein' model in the Systems literature) is an HMC model, as discussed previously, since the nonlinearity is one particular conceptualization of the catchment storage dynamics and its effect on the rainfall-runoff process.

Of course, such conceptualizations are not unique and it was not surprising that later research led to a small modification of the nonlinear BM model to yield the IHACRES model (Jakeman *et al.*, 1990) that has received considerable attention in recent years. However, using the more general DBM approach to modelling, Young (1993) then showed that the variations in the input parameters of the earlier TVP transfer function model could be considered as being dependent upon the changes in flow, with the flow effectively acting as a *surrogate* measure of the catchment storage (see below, as well as the discussion in Young & Beven, 1994, and the papers of Lees, 2000a,b). In the resulting SDP model, the effective rainfall nonlinearity is identified directly from the rainfall-flow data, so avoiding the intuitive conceptualization of the BM and IHACRES models. As we shall see, this model is also in a useful, minimally parameterized, form that is well suited for flood forecasting. At this point in time, therefore, it constitutes one of the most advanced TF models being used in flood forecasting and can be seen as a logical successor to previous TF models.

6. The Generic Catchment Model Based on TF Concepts

Within the catchment modelling context, TF models are of two types: the *Nonlinear Rainfall-Flow Model*; and the *Linear Flow Routing Model*. The complete model used in flood forecasting and warning applications is comprised of both types linked in a manner that reflects the physical nature of the catchment under study. In this paper, however, we concentrate almost completely on the rainfall-flow component, with only a brief reference to flow routing. It must be emphasized, however, that this is not because flow routing is unimportant in real-time flood forecasting. It is simply that the advances reported in this paper relate almost entirely to rainfall-flow modelling.

(*a*) *The Rainfall-Flow Component*

The first step in DBM modelling is the consideration of the objectives. In this case, it will be assumed that this is limited to obtaining a model which explains the rainfall-flow data well on an hourly basis at the *whole catchment scale* and, at the same time, is capable of reasonable mechanistic interpretation combined with an ability to perform well in a flood forecasting/warning context. Note that this emphasis on the 'catchment scale' is very important because the hydrological significance and interpretation of the rainfall-flow models developed below all relate to catchment scale characteristics, such as storage and flow partitioning. These models do not relate directly to more detailed characteristics such as flow paths in the field, analysis of soil depths etc. Note also the allusion to the 'rainfall-flow relationship', rather than the use of the more conventional 'rainfall-runoff' terminology. This is to emphasize that, as discussed below, the models considered here predict both storm runoff and base-flow, which are interpreted as the major components of the *total* gauged flow.

Based on these objectives, the most obvious and physically meaningful model form in this hydrological context is a continuous-time, differential equation (or set of equations). Such a model is consistent, for example, with many conventional hydrological models: e.g. conceptual models of serial and parallel connected nonlinear

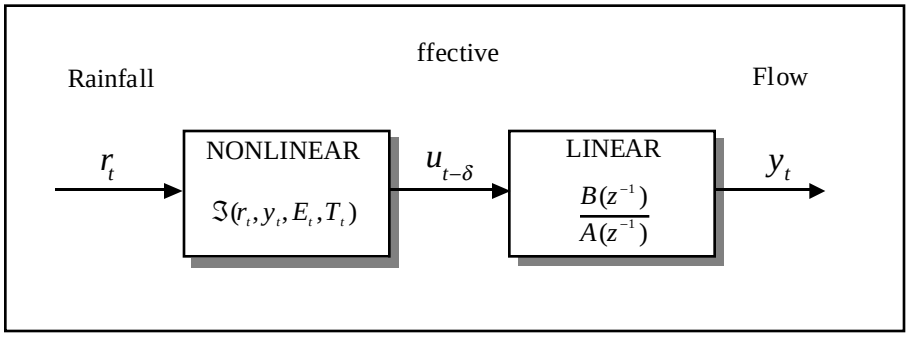

Figure 1. Block diagram of the generic TF rainfall-flow model.

'buckets', as discussed, for instance, in the top-down modelling of Jothityangkoon *et al.*† (see the description of such models in Young, 2002b). However, when dealing with discrete-time, sampled data, it is often convenient to consider modelling in terms of the discrete-time equivalent of the differential equation, the discrete-time TF. Using this generic TF model form, previous DBM modelling of rainfall-flow data based on SDP estimation (see the references in the previous section) has confirmed many aspects of earlier hydrological research and identified the nonlinear DBM model structure shown in figure 1‡. Here, the two components of the TF model are the linear component, which models the basic, underlying, hydrograph behaviour; and the nonlinear component, which models the relationship between the *measured rainfall* r_t and the *effective rainfall* u_t, so controlling the magnitude of the hydrograph contribution through time.

If a constant, uniform sampling interval of Δt time units (e.g. one hour) is utilized, the flow y_t measured at sample time t is related to past, sampled values of itself and present and past sampled values of the u_t by the linear, discrete-time equation

$$y_t = -a_1 y_{t-1} - \cdots - a_n y_{t-n} + b_0 u_{t-\delta} + b_1 u_{t-\delta-1} + \cdots + b_m u_{t-\delta-m} + \eta_t$$

or, in transfer function terms,

$$y_t = \frac{B(z^{-1})}{A(z^{-1})} u_{t-\delta} + \xi_t \tag{1a}$$

In these equations, z^{-1} is the backward shift operator, i.e., $z^{-r} y_t = y_{t-r}$, while $A(z^{-1})$ and $B(z^{-1})$ are *constant coefficient polynomials* in z^{-1} of the following form:

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n} \\ B(z^{-1}) &= b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m}. \end{aligned} \tag{1b}$$

The term δ is a pure time delay, measured in sampling intervals, which is introduced to allow for any temporal (advective) delay that may occur between the incidence

† although these authors discuss modelling at annual, monthly and daily scales, the conceptual arguments are similar.

‡ This model is similar in concept to the variable gain factor model suggested by Ahsan & O'Connor (1993), although its identification, estimation and implementation is quite different.

of a change in u_t and its first effect on y_t. The noise term $\xi_t = \{1/A(z^{-1})\}\eta_t$ represents uncertainty in the relationship arising from a combination of measurement noise, the effects of other unmeasured inputs and modelling error. Sometimes, this noise variable is modelled explicitly as a coloured noise process: e.g. by an Auto-Regressive (AR) model,

$$\xi_t = \frac{1}{A(z^{-1})} e_t \qquad e_t = N(0, \sigma^2) \tag{1c}$$

where e_t is a zero mean, white noise input, sometimes with assumed Gaussian normal amplitude distribution and variance σ^2.

The structure (order) of the TF model (1a) is defined by the triad $[n\ m\ \delta]$ and this is normally identified from the data during the identification and estimation of the model, based on historical rainfall-flow data. This order is normally low, with $n \leq 2$, $m \leq 3$; while the value of δ is defined by the nature of the catchment and the location of the measurement devices, so its range is more difficult to define *a priori*. The general TF model form $B(z^{-1})/A(z^{-1})$ defines the input-output relationship between u_t and y_t and its unit impulse response is a scaled version of the underlying IUH. But, as we see later, it can also be decomposed into a parallel connection of lower order processes. This decomposition not only makes the physical interpretation of the TF more transparent, it can also improve its performance in forecasting terms when implemented within a flood forecasting system (see later §8).

The nonlinear component in figure 1 takes the general form:

$$u_t = \mathcal{F}(r_t, y_t, E_t, T_t) \tag{2}$$

where $\mathcal{F}(r_t, y_t, E_t, T_t)$ denotes an unknown nonlinear functional relationship defining the unobserved catchment storage state s_t (or, as we shall see later, some surrogate for this state) considered as a function of potentially important variables that may affect or be related to catchment storage. In addition to the rainfall r_t, this function may involve other relevant measured variables, such as the temperature T_t (or some function of this, such as the mean monthly temperature T_m), the potential evaporation E_t and the flow y_t; all of which could help to define the changes in soil moisture and storage if they are available. The inclusion here of y_t may seem strange at first sight but this is explained further below.

Three examples will help to take some of the mystery out of this nonlinear function $\mathcal{F}(.)$. First, in the case of the Bedford-Ouse (BM) model (Young, 1974; Young and Whitehead, 1975), $\mathcal{F}(.) = \mathcal{F}(r_t, T_m)$ is defined as a function of r_t and T_m by the following conceptual equations:

$$r_t^* = r_t \frac{T_r - T_m}{c} \tag{3a}$$

$$s_t = s_{t-1} + \frac{1}{\tau_s}(r_t^* - s_{t-1}) \tag{3b}$$

$$u_t = s_t . r_t^*; \tag{3c}$$

Here, in order to allow for seasonal effects, the gauged rainfall r_t is first modified in relation to the changes in the difference between the monthly mean temperature T_m and a reference temperature T_r, where both T_r and c are unknown parameters

estimated from the data. The rainfall modified in this manner, r_t^*, is then assumed to affect the catchment storage through the storage equation (3b), which is simply the discrete-time equivalent of a first order differential equation, with an unknown residence time (time constant) τ_s, again estimated from the data. This equation converts the changes in the modified rainfall r_t^* to changes in the storage, with a lag effect defined by τ_s and is assumed to account for the aggregative effect, at the catchment scale, of all the processes involved in the storage dynamics. Finally, the effective rainfall is defined in (3c) as the product of the gauged rainfall r_t and s_t. Note that, although superficially different, this model is closely related to the conventional *Antecedent Precipitation Index* (API: see e.g Weyman, 1975; Shaw, 1994), since the effect of equation (3b) is to weight the modified rainfall exponentially into the past with a time constant τ_s.

In the case of the standard IHACRES model (Jakeman *et al.*, 1990), which is a direct development of the BM, $\mathcal{F}(.) = \mathcal{F}(r_t, T_t)$ is defined as a function of r_t and T_t by the following conceptual equations:

$$\tau_s(T_t) = \tau_s e^{20-\frac{T_t}{g}} \tag{3d}$$

$$s_t = s_{t-1} + \frac{1}{\tau_s(T_t)}(r_t - s_{t-1}) \tag{3e}$$

$$u_t = c.s_t.r_t \tag{3f}$$

where $\tau_s(T_t)$ in (3d) is a temperature dependent time constant that applies differential exponential weighting to the antecedent rainfall r_t through the first order equation (3e) that, as in the BM, models the changes in the storage state. This time constant is assumed to be inversely related to the rate at which catchment wetness (or potential evapo-transpiration) declines, which is arbitrarily defined as a constant τ_s at 20^oC. The parameter g is a temperature modulation factor which accounts for the fluctuations in potential evapo-transpiration and determines how $\tau_s(T_t)$ changes with temperature. Both parameters τ_s and g are unknown and are estimated from the data. Finally, as in the BM, the effective rainfall u_t is then generated by the product of r_t and s_t, with the constant scaling coefficient c introduced so that the volume of the effective rainfall (rainfall excess) is equal to the total stream flow volume over the estimation period.

Both the BM and IHACRES models are HCM models inferred from the rainfall-flow-temperature data in an hypothetico-deductive manner. The BM has, perhaps, less overtly conceptual reasoning in the definition of its equations than the IHACRES model, which might be considered as hydrologically more acceptable. But both, nonetheless, represent hypotheses about the storage dynamics and effective rainfall generation that are based on the prior perceptions of the modellers (including the present author in the case of the BM model). The DBM model, on the other had, relies much less on such prior perceptions and is based directly on the inductive analysis of the rainfall-flow data. As a result, it is only as complex as required to explain these data from the available measurements. In the case where only rainfall and flow data are available, for instance, $\mathcal{F}(.) = \mathcal{F}(y_t)$ is defined much more simply by the equation,

$$u_t = c.f(y_t).r_t \tag{4}$$

In other words, the catchment storage terms s_t in the BM and IHACRES model equations (3c) and (3f) are replaced by the simpler nonlinear function $f(y_t)$. The physical significance of this nonlinear function is discussed in §7, below: for the moment, it is sufficient to note that the inductive analysis has discovered that the changes in y_t can reflect the changes in the catchment storage sufficient to define the nonlinearity in the rainfall-flow dynamics.

Typically, the form of the nonlinearity $f(y_t)$ is initially identified from the rainfall-flow data through SDP estimation in non-parametric (graphical or 'look-up' table) form, without any prior assumptions about their nonlinear nature. This is then parameterized in some simple manner: for example, in Young (1993), Young & Beven (1994) and Young & Tomlin (2000), $f(y_t)$ is defined as a power law $f(y_t) = y_t^\gamma$ with the power law exponent γ estimated from the data. However, later research has shown that other parametric functions may be more effective and this is a suitable topic for future research (see Conclusions §10) The attraction of this SDP estimation approach is that the nonlinear function $f(y_t)$ is inferred from the rainfall-flow data and not assumed *a priori*, as in HCMs such as the BM and IHACRES models, so leaving less room for unjustified over-confidence in the hypothetical definition of the nonlinear model form.

The DBM model, even in the simple form of equations (1a) and (4), appears to have wide application potential. In addition to rivers in Australia (e.g. Young *et al.*, 1997a,b) and the USA (Young, 2001b), it has been combined with an adaptive gain updating scheme in the parameter-adaptive Dumfries flood warning system (Lees *et al.*, 1994), which has been operating successfully without major modification since 1991; and it is embedded within the Kalman Filter to provide a *State-Adaptive* forecasting system for the River Hodder in NW England (see the example in §9 and Young & Tomlin, 2000).

(*b*) *The Flow Routing Component*

The generic flow (channel) routing model is much simpler than the rainfall-flow model since it is now widely accepted that linear TF models are adequate for the representation of flow dynamics in river systems. The discrete-time routing model for a single stretch of river consists of a serial connection of channel storage elements, each of which has the form:

$$y_t^i = \frac{B(z^{-1})}{A(z^{-1})} y_{t-\delta_i}^{i-1} + \eta_t \qquad i = 1, 2, \ldots, nr \tag{5}$$

where nr is the number of reaches and the i superscript denotes the reach number. This can be considered as the discrete-time equivalent of continuous-time differential equation storage equations (see e.g. Young, 1986). Normally, each of these elements is only first or second order (as defined by statistical identification and estimation based on the up-stream and down-stream flow data). The complete catchment routing model will consist of models such as this for the main river channels and all their tributaries within the catchment, connected accordingly; and it can involve any other measured in-flows as additions, inserted at appropriate nodal locations. The model will receive inputs from the rainfall-flow models discussed above and, in examples such as the Dumfries flood warning model, from flow gauges far upstream which provide advance warning of impending flow changes. A typical

early example of such a model is that used for studies of the Bedford-Ouse river system (Whitehead *et al.*, 1975); more recent examples are the simple River Wyre model (Young, 1986); the much more spatially complex Dumfries model (Lees *et al.*, 1992) and other models discussed in Cluckie (1993).

Note that we are restricting attention here to TF-based flow routing: this is not, of course, the only form of flow routing and other approaches are often utilized, although most of these can be considered in TF terms if this is desired (e.g. the 'kinematic wave' model used in the *Thames Catchment Model*: see Greenfield, 1984; Moore & Jones, 1978). As in the case of the rainfall-flow models, flow routing TF model parameters are normally obtained by the analysis of historical flow records using similar statistical identification and estimation methods to those used in the rainfall-flow example above.

7. Physical Interpretation of the TF Models

As we have stressed, an important aspect of DBM modelling is that the model can be interpreted in physically meaningful terms. In this regard, let us consider first the nonlinear effective rainfall equations (4). The relationship $u_t = c.f(y_t).r_t$ should not be taken literally and interpreted as saying that the effective rainfall is physically a function of flow. Rather, the measured flow y_t is effectively acting here as an objectively identified *surrogate* for the catchment storage s_t. This seems sensible from a hydrological standpoint, since flow is clearly a function of the catchment storage and its pattern of temporal change is likely to be similar. So, the nonlinear function as a whole is similar in its motivation to that used in the BM and IHACRES conceptual models and is justified similarly in physical terms.

The effective rainfall from equation (4) provides the input to the linear TF model component (1a). If this TF is greater than first order and characterized by real eigenvalues (the roots of the $A(z^{-1})$ polynomial), as it normally will be, then the TF can be decomposed into a parallel pathway form, with first order storage equations in each pathway (see e.g. the discussion on the physical interpretation of parallel TF models in Wallis *et al.*, 1989; Jakeman *et al.*, 1990; Young, 1992, 1993; Young & Beven, 1994; and Lees, 2000a,b). From this decomposition, it is possible to compute the residence times (time constants); the advective time delays; the percentage partition of the flow down each of the storage pathways; and even the changing volumes associated with these pathways, all with obvious physical significance. When dealing with hourly data, there are usually two such pathways with very different dynamic characteristics. For example, in the case of the River Hodder example considered later in §9, these take the form of:

- A quick-flow pathway described by a first order TF,

$$x_{1,t} = \frac{\beta_1}{1 + \alpha_1 z^{-1}} u_{t-4} \tag{6a}$$

  which has a partition percentage of 56%, a residence time of 5.5 h and an advective time delay of 4 hours, so producing a total travel time of 9.5 hours;

- A slow-flow pathway described by a first order TF,

$$x_{2,t} = \frac{\beta_2}{1 + \alpha_2 z^{-1}} u_{t-4} \tag{6b}$$

with a partition percentage of 44%, a residence time of 84 h hours and a total travel time of 88 h.

Given these derived model parameters, the most obvious physical interpretation of the DBM is that the effective rainfall affects the river flow via two main pathways. First, the initial rapid rise in the hydrograph derives from the quick-flow pathway, probably as the aggregate result of the many surface processes active in the catchment. And the long, elevated tail in the recession of the hydrograph arises from the slow-flow component, most likely the result of water displacement (probably of old water) from the storage within the groundwater system. Note that the estimate of the flow contribution of this slow-flow component is also practically useful in other ways: it provides a relatively objective estimate of the total base-flow in the river and, as such, can be utilized for base flow quantification and removal, if this is required, as suggested by Jakeman *et al.* (1990). This contrasts with the classical IUH methods, where the base-flow has to be removed rather subjectively.

The linear TF is not always identified from the rainfall-flow data in the [2 2 δ] form of the hourly Hodder model. For instance, in the case of 'flashy' catchments with little storage, the model can be first order, normally [1 1 δ]. And sometimes it can be more complex, implying the presence of more than two parallel pathways. The commonest instance of this is in the case of daily data, where it is quite normal to identify a [2 3 δ] model. Here, in addition to the quick and slow flow pathways, there is an 'instantaneous' effect, with the rainfall causing run-off within one sampling interval. This also occurs in the continuous-time hourly DBM model mentioned later and described in Appendix A.

Finally, it must be emphasized that the estimated TF and its decomposition are stochastic objects and so the uncertainty that is inherent in their derivation needs to be taken into consideration when interpreting the model in physically meaningful terms. As we shall see in the later example of §9, for instance, the data on which the model is based are quite limited, so the flow decomposition described above is uncertain. In particular, while the quick flow pathway dynamics are quite well defined, the slow-flow pathway dynamics are highly uncertain

8. Data Assimilation: the Recursive Kalman Filter, State and Parameter-Adaptive Forecasting

Most conventional methods of flow forecasting utilize the estimated ('calibrated') model for generating forecasts. But if we are concerned with forecasting flow several hours ahead, rather than with simply modelling the rainfall-flow data, then it cannot be assumed that the estimated model of equations (1) and (2), or indeed any model estimated in a similar manner, provides the optimum vehicle for generating such forecasts. The reason for this is obvious. The parameters of the model are normally estimated by minimizing some form of cost function that involves the error between the model generated flow and the measured flow, or the

one-step-ahead prediction errors (as in ML estimation). However, the error that is relevant for multi-step-ahead forecasting purposes is not the 'fitting' or simple prediction error: it is the multi-step-ahead forecasting error based on the required forecast lead time. In this section, therefore, we consider first how the multi-step ahead forecasts can be generated by embedding the model within the Kalman Filter (from hereon denoted by KF) prediction-correction equations. We then discuss the optimization of the complete forecasting algorithm based on the mean square value of multi-step-ahead prediction errors or some function of this measure.

Within the flood forecasting and warning context, a catchment model based on rainfall-flow and flow routing TF models should not be considered as an end in itself: rather, it is a major component of a *data assimilation* system that collects data from remote sensors within the catchment and 'blends' these data with the model in a statistical manner to produce forecasts for multiple time-steps into the future. In the case of stochastic TF models such as those discussed above, an obvious statistical framework for data assimilation is the KF, based on a *Stochastic State-Space* (SS) formulation of the catchment model. It is straightforward to synthesize such a stochastic SS model but it is normally complicated to present its complete constituent equations. For illustrative purposes here, therefore, let us consider the SS formulation in terms of a single, second order rainfall-flow model of the form:

$$y_t = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_{t-\delta} + \xi_t \qquad u_t = \{c y_t^{\gamma}\}.r_t \tag{7}$$

with the parallel flow decomposition shown in (6a) and (6b).

In the simplest situation, where the ξ_t is a white noise process e_t, with variance σ^2, the SS form of the Hodder model can be written most conveniently in the following form:

$$\boldsymbol{x}_t = \mathbf{F}\boldsymbol{x}_{t-1} + \mathbf{G}u_{t-\delta} + \boldsymbol{\zeta}_t \tag{8a}$$

$$y_t = \mathbf{h}^T \boldsymbol{x}_t + e_t \tag{8b}$$

If, as discussed above, the model is decomposed into its parallel form, which has a distinct advantage in forecasting terms (see the example below and Young & Tomlin, 2000), then the matrices $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{h}$ in this SS formulation are defined simply as:

$$\mathbf{F} = \begin{bmatrix} -\alpha_1 & 0 \\ 0 & -\alpha_2 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \boldsymbol{\zeta}_t = \begin{bmatrix} \zeta_{1,t} \\ \zeta_{2,t} \end{bmatrix} \quad \mathbf{h^T} = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

In this manner, the state variables are defined as the unobserved (hidden or latent) quick and slow components of the flow, as defined by the decomposed, first order TFs; and the $\mathbf{h}^T$ vector combines these to form the complete flow output. The white noise variables $\zeta_{1,t}$ and $\zeta_{2,t}$ are introduced to allow for the inevitable uncertainty in the definition of the parallel pathway dynamics and are an important aspect of this 'state-adaptive' approach to forecasting (see later).

For flow forecasting purposes, this state space model is used as the basis for the implementation of the following, recursive, KF state estimation and forecasting algorithm:

***A priori* prediction:**

$$\begin{aligned}\hat{\boldsymbol{x}}_{t|t-1} &= \mathbf{F}\hat{\boldsymbol{x}}_{t-1} + \mathbf{G}u_{t-\delta} \\ \mathbf{P}_{t|t-1} &= \mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^T + \sigma^2\mathbf{Q}_r \\ \hat{y}_{t|t-1} &= \mathbf{h}^T\hat{\boldsymbol{x}}_{t|t-1}\end{aligned}$$

***A posteriori* correction:**

$$\begin{aligned}\hat{\boldsymbol{x}}_t &= \hat{\boldsymbol{x}}_{t|t-1} + \Pi_t \cdot \left\{y_t - \hat{y}_{t|t-1}\right\} \\ \Pi_t &= \mathbf{P}_{t|t-1}\mathbf{h}[\sigma^2 + \mathbf{h}^T\mathbf{P}_{t|t-1}\mathbf{h}]^{-1} \\ \mathbf{P}_t &= \mathbf{P}_{t|t-1} - \Pi_t\mathbf{h}^T\mathbf{P}_{t|t-1} \\ \hat{y}_t &= \mathbf{h}^T\hat{\boldsymbol{x}}_t\end{aligned}$$

In these equations, $\mathbf{P}_t$ is the error covariance matrix associated with the state estimates; and $\mathbf{Q}_r$ is the 2×2 *Noise Variance Ratio* (NVR) matrix defined below.

(*a*) *State Adaption*

In the above KF equations, the model parameters $\alpha_i, i = 1, 2$ and $\beta_j, j = 1, 2$ are known initially from the model identification and estimation analysis based on the estimation data set. However, by embedding the model equations within the KF algorithm, we have introduced additional, unknown parameters, normally termed 'hyper-parameters' to differentiate them from the model parameters†. In this example, these hyper-parameters are the elements of the NVR matrix $\mathbf{Q}_r$ and, in practical terms, it is normally sufficient to assume that this is purely diagonal in form. These two diagonal elements are defined as $NVR_i = \sigma^2_{\zeta_i}/\sigma^2, i = 1, 2$. These specify the nature of the stochastic inputs to the state equations and so define the level of uncertainty in the evolution of each state (the quick and slow flow states respectively) relative to the measurement uncertainty. The inherent state adaption of the KF arises from the presence of the NVR parameters since these allow the estimates of the state variables to be adjusted to allow for presence and effect of the unmeasured stochastic disturbances.

Clearly, the NVR hyper-parameters have to be estimated in some manner on the basis of the data. One well known approach is to exploit *Maximum Likelihood* (ML) estimation based on *Prediction Error Decomposition* (see Schweppe, 1964; Young, 1999b). Another, used later in the example of §9, is to assume that *all* the parameters of the state space model (8a,b) are unknown and re-estimate them by minimizing the variance of the multi-step-ahead forecasting errors. In effect, this optimizes the memory of the recursive estimation and forecasting algorithm (Young & Pedregal, 1999b) in relation to the rainfall-flow data. In this numerical optimization, the multi-step-ahead forecasts $\hat{y}_{t+f|t}$, where f is the forecasting horizon, are obtained by simply repeating the prediction step in the algorithm f times, without

† Of course this differentiation is rather arbitrary since the model is inherently stochastic and so these parameters are simply additional parameters introduced to define the stochastic inputs to the model when it is formulated in this state space form

intermediate correction. The main advantage of this latter approach is, of course, that the integrated model-forecasting algorithm is optimized directly in relation to the main objective of the forecasting system design; namely the minimization of the multi-step prediction errors.

(*b*) *Parameter Adaption*

Although the parameters and hyperparameters of the KF-based forecasting system can be optimized in the above manner, we cannot be sure that the system behaviour may not change sufficiently over time to require their adjustment. In addition, it is well known that the measurement noise e_t is quite highly heteroscedastic: i.e. its variance can change quite radically over time, with much higher variance occurring during storm events. For these reasons, it is wise to build some form of parameter adaption into the forecasting algorithm.

(i) *Gain Adaption*

It is straightforward to update *all* of the parameters in the rainfall-flow model since the RIV/SRIV estimation algorithms can be implemented in a recursive form that allows for sequential updating and the estimation of time-variable parameters (Young, 1984). However, this adds complexity to the final forecasting system and previous experience suggests that a simpler solution, involving a simpler scalar gain adaption is often sufficient. This is the approach that has been used successfully for some years in the Dumfries flood warning system (Lees *et al.*, 1994) and it involves the recursive estimation of the gain $g(k)$ in the following relationship:

$$y_t = g_t.\hat{y}_t + \epsilon_t \tag{9a}$$

where ϵ_t is a noise term representing the lack of fit and, in the case of the second order model (7),

$$\hat{y}_t = \frac{\hat{b}_0 + \hat{b}_1 z^{-1}}{1 + \hat{a}_1 z^{-1} + \hat{a}_2 z^{-2}} u_{t-\delta} \qquad u_t = \{cy_t^{\gamma}\}.r_t \tag{9b}$$

In other words, the time variable scalar gain parameter g_t is introduced so that the model gain can be continually adjusted to reflect any changes in the steady state (equilibrium) response of the catchment to the effective rainfall inputs.

The associated recursive estimation algorithm for g_t takes the usual *Recursive Least Squares* (RLS) form in the case where g_t is assumed to vary stochastically as a *Random Walk* (RW) process (e.g. Young, 1984)†:

$$p_{t|t-1} = p_{t-1} + q_g \tag{9c}$$

$$p_t = p_{t|t-1} - \frac{p_{t|t-1}^2 \hat{y}_t^2}{1 + p_{t|t-1}\hat{y}_t^2} \tag{9d}$$

$$\hat{g}_t = \hat{g}_{t-1} + p_t \hat{y}_t \{y_t - \hat{g}_{t-1}\hat{y}_t\} \tag{9e}$$

where $\hat{g}_t$ is the estimate of g_t; while q_g is the NVR defining the stochastic input to the RW process, the magnitude of which needs to be specified (see later). The

† It is also a scalar example of *Dynamic Linear Regression* (DLR) algorithm (see Young, 1999b).

adapted forecast is obtained by simply multiplying the initially computed forecast by $\hat{g}_t$. Note that gain adaption of this kind is quite generic and can be applied to *any* model, not just those discussed here.

(ii) *Variance Adaption*

To allow for the heteroscedasticity in e_t, it is necessary to recursively estimate† its changing variance σ_t^2. Although a logarithmic transform might suffice, a superior approach is to use the transformation is $c_t = \log(\chi_t^2) + \lambda$, where the stochastic process χ^2 defined by,

$$\chi_m^2 = (e_{2m-1}^2 + e_{2m}^2)/2 \qquad m = 1, ...N/2 \tag{10a}$$

in which $\lambda = 0.57722$ is the Euler constant. This is motivated by Davis & Jones (1968), who showed that c_t has a theoretical distribution which is almost normal. As a result, an estimate $\hat{h}_t$ of the transformed variance can be obtained from the following recursive least squares (*cf* the above RLS algorithm for $\hat{g}_t$), where this time it is c_t that is assumed to vary stochastically as a RW process:

$$p_{t|t-1} = p_{t-1} + q_h \tag{10b}$$

$$p_t = p_{t|t-1} - \frac{p_{t|t-1}^2}{1 + p_{t|t-1}} \tag{10c}$$

$$\hat{h}_t = \hat{h}_{t-1} + p_t \left\{ c_t - \hat{h}_{t-1} \right\} \tag{10d}$$

An estimate $\hat{\sigma}_t^2$ of σ_t^2 can then be obtained as $\hat{\sigma}_t^2 = exp(\hat{h}_t - \lambda)$.

(iii) *Hyper-parameter estimation*

The RLS estimation algorithms (9) and (10) are very simple examples of the KF and so it is necessary to estimate the hyper-parameters (here q_g and q_h) in some manner. Their joint estimation with the KF hyper-parameters using some form of combined optimization is difficult, however, and it is simpler in practice to consider a more heuristic approach. For instance, it is well known that q_g and q_h control the memory of their respective RLS estimation algorithms and the associated smoothing of the estimate (e.g. Young, 1984). Consequently, since q_g and q_h are scalar values, it is quite straightforward to manually optimize them to yield the best multi-step-ahead forecasts.

9. An Illustrative Example: Adaptive Flow Forecasting for the River Hodder in North West England

This example is concerned with the analysis and forecasting of hourly flow, measured during 1993, at Hodder Place gauging station on the River Hodder in North West England. The River Hodder has a catchment area of 261 km^2 and it forms part of the larger River Ribble catchment area of 456 km^2. The average annual precipitation is 1600 mm and the mean flow is 8.42 $m^3 s^{-1}$ (95% exceedance 0.969 $m^3 s^{-1}$;

† a non-recursive ML formulation of this heteroscedasticity problem is given by Sorooshian (1985)

10% exceedance 21.46 $m^3 s^{-1}$). The level-over-datum of the gauging station and the maximum altitude in the catchment are 42 m and 544 m, respectively. The catchment is very lightly populated: it has mixed farming at lower levels but is mostly peat moorland, with millstone grit and carboniferous limestone. The flows in the river can be affected by spillway operation and storage release from the upstream Stocks reservoir but this effect is fairly small (mean flow $0.48 m^3 s^{-1}$), occurring almost entirely in Summer. Certainly, as we shall see, there appears to be no discernible effect on the Winter flows considered in the present example. And, of course, if there happens to be a very small random effect, then the stochastic formulation of the DBM model is able to take account of this, provided most of the downstream flow can be explained well by the rainfall.

DBM model identification and estimation in the present example is based on 720 h of hourly rainfall-flow data measured during January 1993, as shown in figure 2. The rainfall series r_t, measured in $mm\ h^{-1}$, is based on a Thiessen Polygon average of three tipping-bucket rain gauges; while the flow series y_t, measured in the same units as the rainfall (computed by dividing the gauged volumetric flow rate by the catchment area), is obtained from an Environment Agency flow gauge located at Hodder Place. The subsequent validation and forecasting analysis is based on a further 480 h of rainfall-flow data measured later, during December 1993, as shown in figure 3. Young & Tomlin (2000) have previously used this second data set to illustrate how a second order, nonlinear DBM model can provide the basis for DBM modelling and KF-based flow forecasting, so the present analysis can be seen as an extension of these earlier studies. It should be emphasized that these data sets were not chosen to produce good results: indeed, the modelling and forecasting problem they pose is quite difficult since the estimation sample size $N = 720$ covers a very short period (just over a month), the measured data (particularly the rainfall) are not particularly good quality, and other associated data that might assist in the analysis, such as temperature or soil moisture measures, were not available for the present analysis. Note also that the validation data set exhibits quite significantly larger maximum flow rates than those in the estimation data set, so that the predictive and extrapolative ability of the nonlinear model is evaluated in the face of this larger envelope of rainfall-flow conditions.

The modelling in this example is carried out in terms of discrete-time transfer functions and associated discrete-time KF-based state space representations. It is interesting to note, however, that it is possible to approach the problem within a hybrid 'continuous-discrete' (CD) framework (e.g. Young, 1981); i.e. continuous-time models estimated from the sampled data and an associated CD Kalman Filter forecasting algorithm, with continuous-time prediction and discrete-time correction. Although this approach has not been used very often in the past, it has the advantage of presenting the models in a more transparent, differential equation form that has immediate physical relevance. Moreover, this CD algorithm can handle non-uniformly sampled data. Appendix A demonstrates the feasibility of this approach by showing how a continuous-time model can be estimated from the same sampled data used in the discrete-time analysis described below.

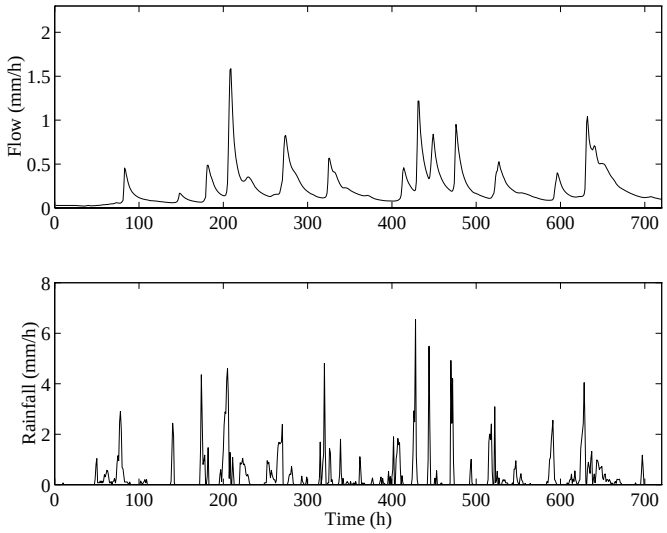

Figure 2. Hourly rainfall-flow data for the River Hodder during January 1993.

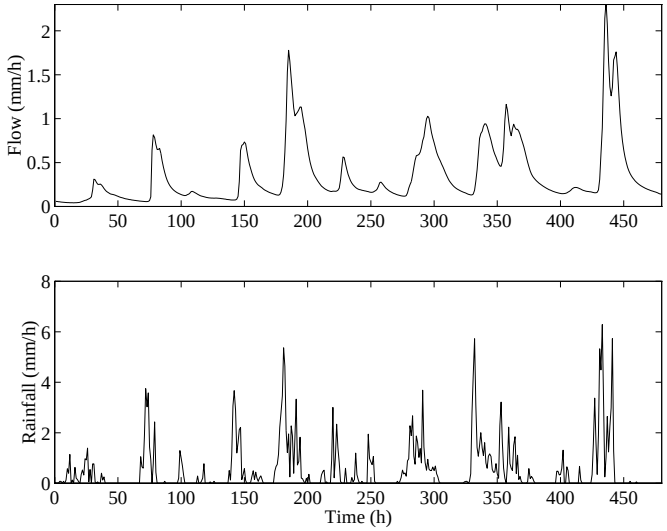

Figure 3. Hourly rainfall-flow data for the River Hodder during December 1993.

(*a*) *Identification and Estimation*

The first step in DBM modelling is the identification of an appropriate model structure. Initial SRIV estimation of a linear, constant parameter TF model suggests that both [1 1 3] and [2 2 4] models are well identified (they have the most negative YIC values), with the latter yielding a respectable $R_T^2 = 0.82$ compared with $R_T^2 = 0.79$ for the former. However, standard statistical tests for the parameter constancy (e.g. Johnston & DiNardo, 1997) suggest strongly that the parameters are varying in some manner. Moreover, initial SDP estimation based on these mod-

Table 1. *Model Identification*
(Linear component of nonlinear model)

Linear Model Component	R_T^2	YIC	AIC	σ^2
[1 1 3]	0.82	-9.1	-4.8	0.0081
[2 2 4]	0.84	-8.2	-4.9	0.0070
[3 2 3]	0.86	-5.7	-5.0	0.0065

els reveals significant state dependent parameter variation similar to that obtained previously in the analysis of rainfall-flow data (e.g. Young, 1993, 2001b; Young & Beven, 1994; Young *et al*, 1997; Young & Tomlin, 2000; Lees, 2000a,b). On this basis, the following model structure is initially identified:

$$y_t = \frac{B(z^{-1})}{A(z^{-1})} u_{t-\delta} + \xi_t \tag{11a}$$

where the effective rainfall u_t is defined in the form of a power law relationship in the flow y_t, which is acting as a surrogate measure of soil water storage, as discussed earlier in §7, i.e.,

$$u_t = f(y_t).r_t \qquad f(y_t) = c.y_t^{\gamma} \tag{11b}$$

Here, the normalization coefficient c is is chosen so that, over the observation interval of N samples, $\sum^N(u_t) = \sum^N(y_t)$. Based on the linear model identification, therefore, it is likely that the linear component model (11a) structure is either [1 1 3] or [2 2 4], although this needs to be verified during identification and estimation of the nonlinear model.

The parameters in the identified nonlinear model equations (11a) and (11b) are estimated by nonlinear least squares estimation using the *leastsq* optimization procedure in *MATLAB*, in which the SRIV estimation algorithm is incorporated to estimate the linear TF parameters†, with the effective rainfall function parameter γ optimized concurrently within the optimization function. The most significant results obtained in this identification stage of the analysis are shown in table 1. Although, at first sight, the [1 1 3] model structure is identified well, simulation of the model shows that it does not capture the recession part of the flow curve, which is so important in hydrological terms and determines the nature of the base-flow characteristics. Similarly, the third order [3 2 3] model has inherent limitations: although it has the highest R_T^2 value, it also has a significantly higher YIC value, suggesting over-parametrization. In addition, the residues of the associated linear TF partial expansion include one negative value, implying a negative flow in the parallel partitioning that is difficult to explain in physically meaningful terms. In contrast to these other two identified models the [2 2 4] model can be interpreted very well in physical terms, as shown below.

† A more sophisticated stochastic estimation procedure based on maximum likelihood optimization of the associated state-space model (e.g. Young, 2000) would be preferable in statistical terms but is more complex and not justified in the present illustrative context.

Based on the above considerations, the finally estimated model takes the form:

$$y_t = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_t + \xi_t \qquad u_t = \{cy_t^{\gamma}\}.r_t \tag{12}$$

The optimized parameter estimates are as follows:

$$\hat{a}_1 = -1.821(0.012); \hat{a}_2 = 0.823(0.011); \hat{b}_0 = 0.102(0.0030); \hat{b}_1 = -0.1002(0.003);$$
$$\hat{\gamma} = 0.281(0.009); \hat{c} = 1.17$$

where the figures in parentheses adjacent to the estimates are the standard error bounds. The noise model is identified as a third order *AutoRegressive* (AR(3)) model (e.g Box and Jenkins, 1970) by the AIC and the associated parameter estimates are:

$$\hat{c}_1 = -1.114(0.037); \hat{c}_2 = 0.507(0.052); \hat{c}_3 = -0.153(0.037)$$

The model (12) has $R_T^2 = 0.844$ based on its response (simulation) error, while the standard coefficient of determination based on the final stochastic residuals from the AR(3) noise model (i.e. one-hour-ahead prediction errors) is $R_1^2 = 0.95$. The auto (acf) and partial (pacf) autocorrelation functions of these stochastic residuals show no significant temporal correlation although, as normal in rainfall-flow models, they are highly heteroscedastic (see later) and correlated to a minor extent with the rainfall input†.

As required by the DBM modelling strategy, the model (12) can be interpreted well in physically meaningful terms. Based on a partial fraction expansion of the linear TF, as discussed in §7, it can be interpreted as a parallel connection of two first order processes. The quick-flow process has a residence time of 5.5 h and accounts for 56% of the total flow; while the slow-flow process has a residence time of 84 h and and accounts for 44% of the flow. The estimated hydrographs (impulse responses) associated with this parallel decomposition are compared in figure 4 with the complete hydrograph of the TF model (12). It is clear that much of the initial response is associated with the quick-flow pathway, while the main effect of the slow flow pathway is to raise the longer-term tail of the hydrograph recession.

Note, however, that the derived parameter estimates that define the parallel decomposition are not necessarily all that well defined statistically because of the uncertainty in the estimated TF parameters. In particular, figure 5 shows some of the results obtained from Monte Carlo Simulation (MCS: see Young, 1999a) analysis based on 5000 random realizations and the SRIV estimated covariance matrix of the TF model parameters. Despite high correlation between the basic TF model parameter estimates, the empirical probability distribution function (pdf) associated with the derived quick-flow residence time (right hand panel) is quite well defined, with only small dispersion around the mean value, which coincides with the estimated value of 5.5 h, as expected. In contrast, the empirical pdf of the slow-flow pathway in the left hand panel is very poorly defined: the distribution is highly dispersed and skewed markedly towards higher values with a very long tail.

† This small level of cross correlation between the model residuals and the input rainfall is quite normal in rainfall-flow models of all forms and is probably linked in part with the heteroscedasticity. It is not desirable, however, and suggests that further research on rainfall-flow models is still required to remove this anomaly.

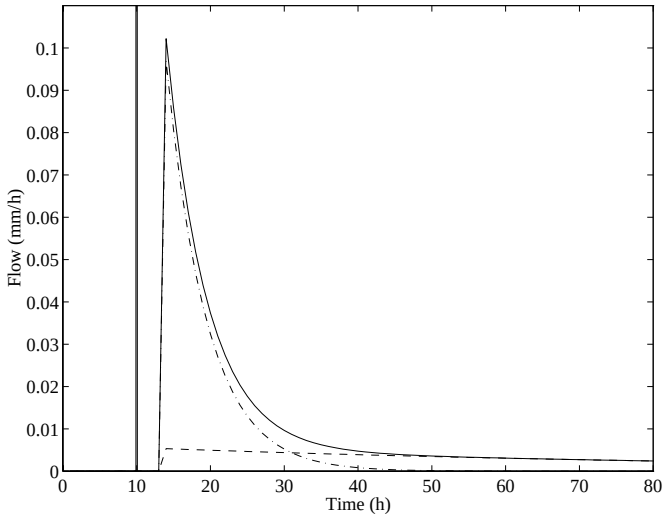

Figure 4. Estimated unit hydrograph (unit impulse response) based on the DBM model (full line); quick-flow component (dash-dot line); slow flow component (dashed line).

This is not surprising since the short term observational data used in the modelling analysis is not providing nearly as much information on this long term mode as it does on the quick-flow dynamics.

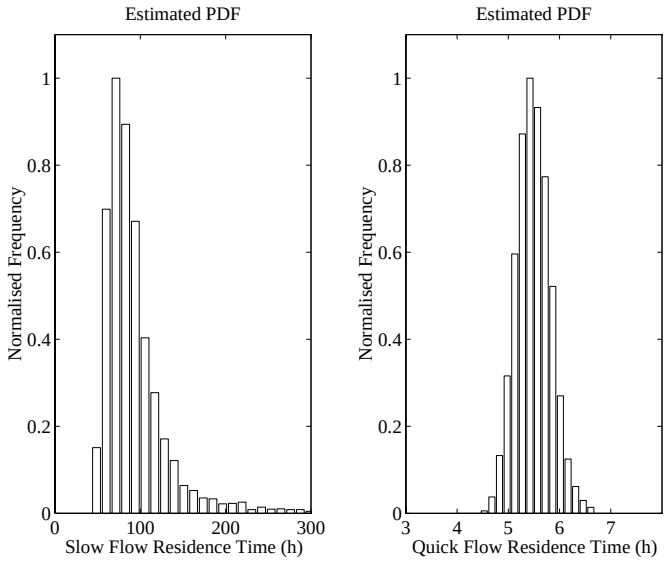

Figure 5. Normalized histograms (empirical pdfs) of the slow (left panel) and quick (right panel) residence times obtained by *Monte Carlo Simulation* (MCS) analysis.

Finally, the full line in figure 6 shows the effective rainfall coefficient plotted against the flow y_t, as defined by the estimated power law nonlinearity $y_t^{0.281}$ (recall

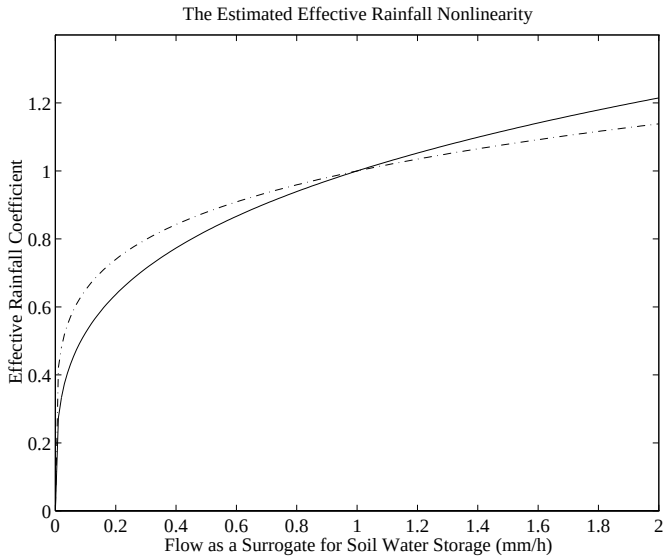

Figure 6. Estimated effective rainfall coefficient plotted as a function of flow, which is acting as a surrogate measure of soil-water storage. The dash-dot line shows the modified nonlinearity arising from the optimized flow forecasting analysis.

that y_t is acting here as a surrogate for a soil-water storage variable). This makes good physical sense, since it shows that, as the implied soil-water storage variable increases, so the proportion of rainfall that is effective in causing run-off increases, but with the decreasing slope as the catchment becomes wetter. It can be shown (Beven, 2001, p. 94; Lees, 2000a) that there are parallels between the form of this nonlinearity and the hydrologic concept of a 'dynamic contributing area', as in TOPMODEL (Beven & Kirkby, 1979) and the PDM (Moore, 1985).

(*b*) *Adaptive Forecasting*

As pointed out in §7, the estimated or 'fitted' model (12) does not necessarily provide the best basis for multi-hour-ahead forecasting. In order to design the flow forecasting system, therefore, it is necessary to re-optimize the model parameters, and any other associated hyper-parameters of the KF based, for example, on a least squares cost function in the error between the specified multi-hour-ahead forecast and the measured flow over the estimation data set. Given the four hour advective time delay in the model (12), it makes sense to assume here that the major objective of forecasting in the present example is to optimize the four hour-ahead forecasts. Of course, other cost functions could be used, such as a likelihood function based on the four step-ahead forecasting errors, but this simple least squares cost function will suffice for the present example and makes immediate physical sense, given the nature of the forecasting problem defined here. In this illustrative example, the length of the validation data set is not really sufficient to consider the updating of *all* the model parameters but we are able to evaluate the effectiveness of the simpler scalar gain and variance adaption procedures.

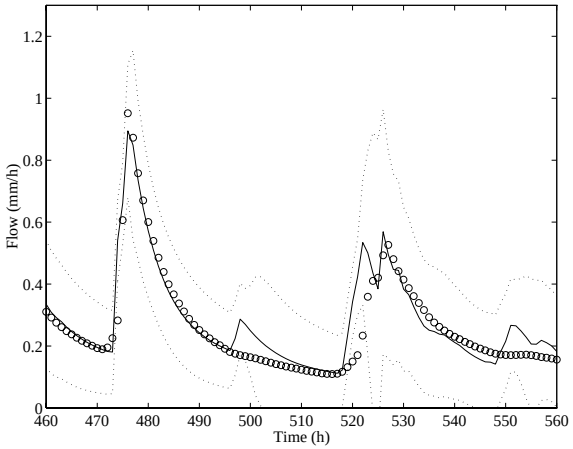

Figure 7. Comparisons of four-hour-ahead flow forecasts (full line) and measured flow (dash-dot line) for the period between 460 h and 560 h during January 1993. The dotted lines show the standard error bounds.

Figures 7 to 9 show the forecasting results obtained over various sections of the data using the KF forecasting system described in §8, optimized in the above manner and incorporating both gain and variance adaption. Figures 7 and 8 show forecasting within the estimation data set, while figure 9 shows how the system performs on the validation data set during December 1993. Figure 9 also shows, plotted from above, the adjustments to the gauged rainfall made by the nonlinear effective rainfall coefficient. Note how, in the first two figures, the model forecasts some false flow increases due to measured increases in the rainfall r_t. This is not unusual in rainfall-flow modelling and illustrates well the deficiency of this rainfall measure as a reliable quantification of the rainfall causing flow at the Hodder Place flow gauging station. It also demonstrates how the ultimate accuracy of flow forecasting and warning is critically dependent on good rainfall measurement and estimation at the catchment scale.

It should be emphasized that the nonlinear rainfall-flow model parameters and hyper-parameters used over the validation data set are those optimized on the basis of the estimation data set alone; and *they are maintained at these values over the whole of the validation data set*. Given the limited size of the estimation data set, however, it is not surprising that this model is not entirely appropriate for the later December 1993 period and the gain and variance adaption mechanisms are active in improving the forecasts, as we see later.

The model used to generate the results in figures 7 to 9 is the same form as the model (12) but the optimized parameters, within this four-hour-ahead forecasting setting, are as follows:

$$\hat{a}_1 = 1.814(0.016); \hat{a}_2 = 0.817(0.015); \hat{b}_0 = 0.1006(0.0031); \hat{b}_1 = -0.0973(0.003);$$
$$\hat{\gamma} = 0.187(0.012); \hat{NVR}_1 = 0.57(0.11); \hat{NVR}_2 = 0.821(0.46); \hat{c} = 1.17$$
$$\hat{q}_g = 0.000001; \hat{q}_h = 2.5$$

The modified effective rainfall nonlinearity is shown as the dash-dot line in figure 6. The estimated noise model in this case is identified by the AIC as a much higher

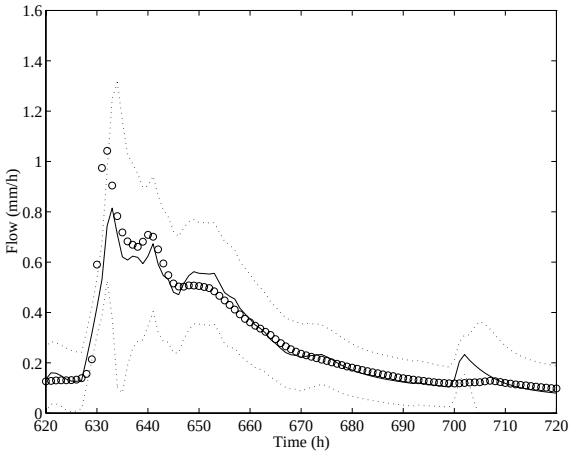

Figure 8. Comparisons of four-hour-ahead flow forecasts (full line) and measured flow (dash-dot line) for the period between 620 h and 720 h during January 1993. The dotted lines show the standard error bounds.

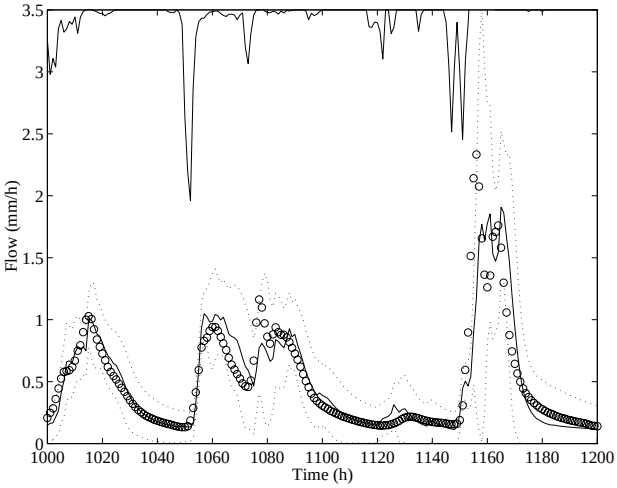

Figure 9. Comparisons of four-hour-ahead flow forecasts (full line) and measured flow (dash-dot line) for the period between 280 h and 480 h during December 1993. The dotted lines show the standard error bounds. The plot above illustrates is d_t, the difference between the gauged rainfall r_t and the effective rainfall u_t subtracted from 3.5, for clarity: i.e. $d_t = 3.5 - (r_t - u_t)$.

order AR(9) process but, for simplicity, forecasting here is based only on the state space model for the decomposed second order TF model (see §7). This means that the forecasting performance discussed below could be improved by the addition of the AR(9) noise model and the associated augmentation (from 2nd to 11th order) of the state space model used for KF forecasting system design.

The gain and variance adaption operative over the whole estimation-validation period (concatenated for clarity) are shown in figures 10 and 11. In figure 10, we see

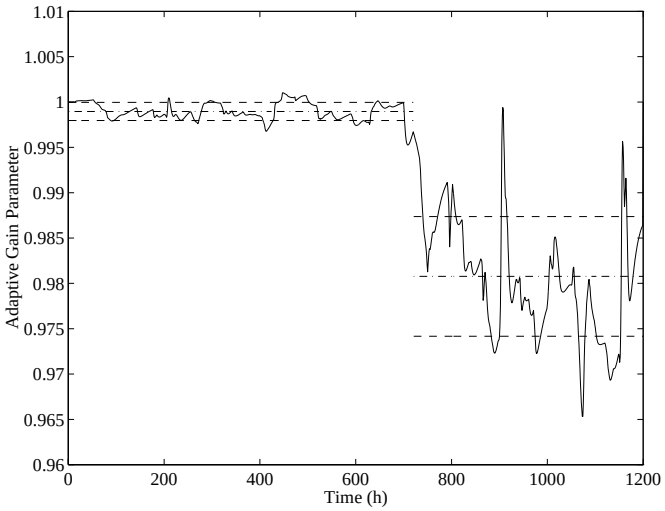

Figure 10. Recursive estimate of the adaptive gain parameter over the full concatenated data set, showing the significant reduction in the estimated value over the December 1993 portion of the data. The mean(std) estimates are 0.996(0.006) for January and 0.982(0.005) for the final 300 h in December: these are plotted as dashed lines with dotted std bounds. The nominal unity value is shown as dash-dot line.

that the adaptive gain $\hat{g}_t$ reduces significantly after sample 720 h when forecasting begins over the December 1993 validation period. This indicates the value of such adaption in correcting for any deficiency in the estimated model. The importance of the heteroscedastic noise is illustrated in figure 11, where we see that the estimated residual series variance changes markedly over the whole validation period, with particularly large changes estimated over the two major rainfall episodes around samples 900 h and 1150 h, where the heteroscedasticity is particularly significant.

The effect of introducing the variance adaption is particularly noticeable in the standard error (*se*) bounds on the four-hour-ahead forecasts plotted in figures 8 and 9, where the bounds widen considerably over the peak flow periods. This would not happen in the standard KF algorithm, as pointed out by Lees (2000a). To illustrate this, Lees carries out off-line analysis of his forecasting results and applies a Box-Cox transformation to the forecasting residuals. He then computes the empirical *se* bounds, showing how they reveal clearly the increased uncertainty over the peak flow periods. In this regard, it should be emphasized that the *se* bounds in figures 7 to 9 are estimated *on-line* and in real time as an inherent part of the variance adaptive KF algorithm and are *not* computed empirically off-line. In other words, the user is informed of this increased uncertainty in real-time and can judge the potential for flooding within the next few hours accordingly.

Table 2 gives some indication of the forecasting performance achieved here (as measured by appropriately defined coefficients of determination, R_i^2) when compared with other forecasting procedures under various settings of the forecasting system. The two other forecasting options are: (a) the 'standard' TF forecasting system in which the TF model (12) is used directly in its full TF form, without parallel decomposition or incorporation in the KF; and (b) the näive forecasting

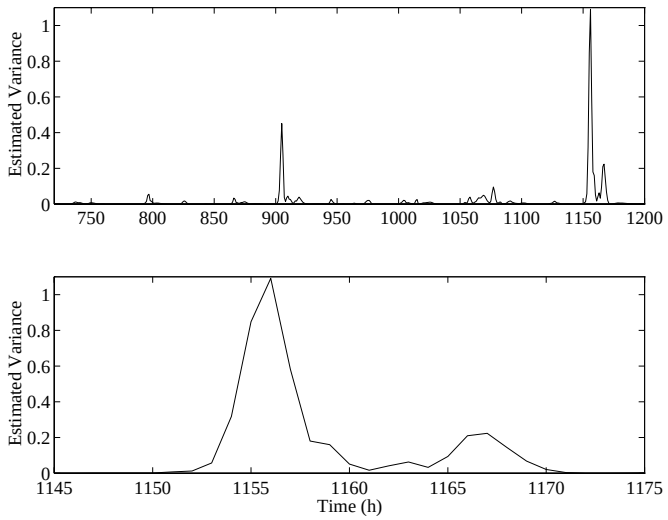

Figure 11. Recursive estimate of the adaptive noise variance parameter. The upper panel shows the estimate over the full concatenated data set. The lower panel is an enlarged view of the significantly heteroscedastic period between samples 1145 h and 1175 h, during December, 1993.

system, in which the four-hour-ahead forecast $\hat{y}_{t+4}$ at any sampling instant t is simply set to the flow measurement y_t. At first, these comparative results are surprising, since it is clear that the näive forecaster performs better than the standard TF forecaster based directly on the model (12). The main reason for this is that, as it stands, the TF model (12) is not good for forecasting because the numerator parameters in the TF model are approximately the same value and different in sign. This induces a near-differencing operation and causes 'spikes' in the forecasts that considerably degrade the forecasting performance. This problem is completely avoided by the physically meaningful decomposition of the TF and its incorporation, in this decomposed form, within the KF forecasting engine.

Finally, although the forecasting system here has been designed for 4-hour-ahead forecasts, it produces forecasts for any requested forecasting interval. For instance, the coefficients of determination for the forecasts over all lead times from 1 to 6, $R_i^2, i = 1, 2, ..., 6$, are given below:

$$R_1^2 = 0.874; R_2^2 = 0.856; R_3^2 = 0.847; R_4^2 = 0.842;$$
$$R_5^2 = 0.764; R_6^2 = 0.658$$

Of course, the forecasts for periods other than four hours will not necessarily be optimal and may be improved by explicit optimization for the specified forecasting interval. For instance, the comparative figures obtained when the optimization is based on separate optimization at each sampling interval is as follows:

$$R_1^2 = 0.939; R_2^2 = 0.877; R_3^2 = 0.845; R_4^2 = 0.842;$$
$$R_5^2 = 0.767; R_6^2 = 0.658$$

So we see that worthwhile advantage is obtained in the case of forecasting intervals from one to three-hours-ahead. This would require 3 additional KF algorithms

Table 2. *Comparative Forecast Evaluation*
(Comparison of R_i^2 values obtained from various forecasting situations)

Level of Adaption	4-step ahead	1-step ahead	4-step: no decomposition, no KF	Näive forecast
Only state adaption (sa)	0.820	0.876	0.313	0.461
sa + gain adaption	0.834	0.882	0.328	0.461
sa + variance adaption	0.841	0.876	0.350	0.461
sa + both	0.842	0.874	0.358	0.461

acting in parallel, the algorithms are so simple that the increase in the computational burden is quite acceptable. Also, note how the forecasts are degraded more for forecasting intervals greater than four hours. This is because, after this interval, it is necessary to forecast the rainfall into the future and here, these forecasts are simply set to zero. Improved performance would be expected, therefore, if rainfall forecasts were available for forecasting intervals greater than four hours.

Table 2 shows that both gain and variance adaption are effective when applied independently, with the variance adaption providing the larger improvement. When used together, the improvement in the four-hour-ahead forecasting performance is only marginally better than that achieved by the variance adaption alone. The reason why the simultaneous implementation of the two adaption algorithms does not lead to further improvement is probably because of interaction between them. And it probably suggests that there is room for improvement in the design of such adaption techniques. Nevertheless, the improvement over the situation with no adaption is obviously worthwhile and would probably be greater than this if a longer and more realistic time period was considered. In the case of the Dumfries flood warning system, for instance, the gain adaption has proven very beneficial over many years and has considerably reduced the frequency of full re-calibration. Moreover, as pointed out above, the variance adaption improves the real-time estimates of the *se* bounds and so is beneficial for this reason alone.

10. Conclusions

This report describes some recent advances in stochastic modelling and forecasting that provide the basis for the implementation of real-time flow and flood forecasting systems. It argues that deterministic reductionist (or 'bottom-up') models are inappropriate for real-time forecasting because of the inherent uncertainty that characterizes river catchment dynamics and the problems of model over-parametrization that are a natural consequence of the reductionist philosophy. The advantages of alternative *Data-Based Mechanistic* (DBM) models, statistically identified and estimated in an inductive manner directly from rainfall-flow data, are discussed. In particular, the report shows how nonlinear, stochastic, transfer function models can be developed using powerful methods of recursive time series analysis. Not only are these models able to characterize well the rainfall-flow dynamics of the catchment in a parametrically efficient manner but, by virtue of the DBM modelling strategy, they can also be interpreted in hydrologically meaningful terms. Most importantly in the forecasting context, the models are also in an ideal form for incorporation

into a data assimilation and forecasting engine based on a special, adaptive version of the Kalman Filter algorithm.

The practical example described in the paper demonstrates how, with the minimum of rainfall-flow data and no available rainfall forecasts, the approach proposed here can generate useful flow forecasts for several hours ahead; forecasts that could form the basis for flood warning system design. Such a system would be a natural development of the Dumfries flood warning system (Lees *et al.*, 1994), which was designed from a similar DBM modelling standpoint and has been operating successfully without major modification since 1991. The methodological advances described in the present paper would ensure much improved performance from such a system but the basic minimalist design and low economic cost of development would be retained. Both of these recursive approaches to real-time forecasting can be contrasted with more conventional, non-recursive, real-time forecasting procedures proposed previously. A typical example is the adaptive scheme suggested by Brath & Rosso (1993) which addresses some of the same statistical issues raised in the present paper. However, it operates on an event basis rather than continuously; it uses repeated *en-bloc* optimization rather than recursive estimation; it is based on a simple conceptual model with *a priori* assumed structure and parameterization; and it is computationally much more demanding.

Of course, there remain a number of methodological problems still to be solved. The DBM models discussed in the paper perform well but they cannot be considered completely satisfactory while the model residuals retain their current unsatisfactory statistical characteristics. In particular, the correlation remaining between the residuals and the rainfall input shows that the model is still not fully explaining the complete rainfall-flow process (although the remaining unexplained variance represents only a small proportion of the total variance). This limitation of the current DBM models (shared, the author believes, by all current rainfall-flow models, whatever their type) is almost certainly due to deficiencies in the effective rainfall nonlinearity and possibly the presence of other, smaller nonlinearities in the system as yet unquantified. There is clear need for more research on this fascinating subject and, although such research would require the analysis of a large and comprehensive rainfall-flow data base covering a wide array of different catchment behaviour, it would provide useful information for *all* existing rainfall-flow modelling studies, not just those discussed in this paper.

This future research could be based on an extension of the DBM models discussed here within the existing Kalman Filter forecasting system. Or it could involve the use of more sophisticated but computationally intensive Bayesian updating procedures which exploit on-line *Monte Carlo Simulation* (MCS). Early use of MCS in hydrology (e.g. Whitehead and Young, 1979) was inhibited by computational limitations but, in recent years, the advances in computers have led to an explosion of research in this area. As a result, there are a wide spectrum of MCS methods are available ranging from *Markov Chain Monte Carlo* (MCMC: see e.g. Gammerman, 1997), through Monte Carlo filtering algorithms (e.g. Kitagawa, 1996, 1998; Thiemann *et al.*, 2001), to simpler non-recursive approaches such as the GLUE procedure (Beven & Binley, 1992). Research is continuing on the best approach in the present context but simple Monte Carlo extensions to the adaptive KF described above are yielding promising results. The gain and variance adaption procedures presented in the present paper also require further development: although they enhance the

forecasting performance, they are still not entirely satisfactory in theoretical and practical terms.

Finally, the models considered in the paper are all of a 'lumped parameter' variety (i.e. they consist of the linear and nonlinear transfer functions that are the discrete-time equivalents of differential equations and describe the temporal behaviour only at selected spatial nodes within the catchment system). The alternative 'distributed parameter' models, which involve spatio-temporal aspects of the catchment and are described by models such as partial differential equations in time and space (or some equivalent of these), have not been considered at all. Such models are clearly attractive in these days of *Geographical Information Systems* (GIS) and weather radar, since they are, potentially at least, able to exploit spatial information of this type.

Within a flood forecasting system, such distributed models are of particular relevance because they can hope to predict the spatio-temporal progress of flood inundation, as in Romaowicz and Beven (1998) and Beven *et al.*(2000). However, the parenthetical comment in the title of the latter paper '*Mapping the probability of flood inundation (even in real time)*' hints at the difficulties of using such computationally intensive models in real-time applications, even if the other theoretical and practical problems associated with such models could be solved. But they are, nonetheless, an important topic of continuing research and the continuing evolution of the digital computer will undoubtedly resolve the computational problems in the not too distant future. In the mean time, there is room for research on the amalgamation of distributed and lumped parameter concepts, with the distributed models of rainfall and its distribution throughout the catchment providing improved estimates and forecasts of the rainfall inputs. For, as we see in the example of §9, it is the inadequacy and inconsistencies of the rainfall inputs that appears to most limit the accuracy of the flow and flood forecasts.

Acknowledgements

The author is very grateful to his colleagues Professor, Keith Beven, Dr. Paul McKenna and Dr. Renata Romanowicz for reading and commenting on a draft of this report. Naturally, the author is responsible for any errors or omissions.

Appendix A. Continuous-Time Modelling

A major advantage of instrumental variable estimation, in general, and the SRIV algorithm, in particular, is that it can be used to estimate models in continuous or discrete time from discrete-time sampled data (Young & Jakeman, 1980; Young, 1996). To illustrate this facility, the following continuous time TF was identified and estimated from the January 1993 estimation data set:

$$y(t) = \frac{\beta_0 s^2 + \beta_1 s + \beta_2}{s^2 + \alpha_1 s + \alpha_2} u(t-4) + \xi(t)$$

where s is used here as the differential operator, i.e. $s^n y(t) = d^n y(t)/dt^n$ and $\xi(t)$ is the continuous-time equivalent of the noise. The estimated parameters in this case are as follows:

$$\hat{\alpha}_1 = 0.167(0.009); \hat{\alpha}_2 = 0.0012(0.0003); \hat{\beta}_0 = 0.066(0.007); \hat{\beta}_1 = 0.106(0.004);$$
$$\hat{\beta}_2 = 0.0013(0.0003); \hat{\gamma} = 0.2807(0.012); \hat{c} = 1.17$$

This model explains the data marginally better than the discrete-time model (12) with $R_T^2 = 0.845$ but, as can be seen, an additional numerator parameter is required.

The parallel flow decomposition of this TF can be considered in the following differential equation form:

- An instantaneous pathway with gain of 0.066 and a partition percentage of 6%
- A quick-flow pathway described by a first order TF,

$$6.3\frac{dy(t)}{dt} = -y(t) + 0.574u(t-4)$$

  which has a partition percentage of 54%, a residence time of 6.3 hours and an advective time delay of 4 hours, so producing a total travel time of 10.3 hours;

- A slow-flow pathway described by a first order TF,

$$135\frac{dy(t)}{dt} = -y(t) + 0.427u(t-4)$$

  with a partition percentage of 40%, a residence time of 135 hours and a total travel time of 139 hours.

Clearly, these estimated dynamic characteristics are not the same as those obtained for the discrete-time model in §8(a) but they are consistent if the uncertainty in the parameter estimates is taken into account (see e.g figure 5: the sizes of the uncertainty bounds obtained via MCS are similar for the continuous-time model: a continuous-time example of such MCS analysis is given in Young, 1999a). These differences would make a small difference to the forecasting performance if a CD implementation was preferred, but it is unlikely that this difference would be very significant. The main advantage of the continuous-time model is clear from the above decomposition: the model for each flow pathway is defined directly by the estimated parameters of first order differential equation model for the pathway, so that the model is transparent and immediately interpretable in a physically meaningful manner.

References

Ahsan, M. and O'Connor, K. M. 1993 A simple nonlinear rainfall-runoff model with a variable gain factor. *Journal of Hydrology*, **155**, 151-183.

Akaike, H. 1974 A new look at statistical model identification. *IEEE Trans. Auto. Control*, **AC 19**, 716-723.

Anderson, M. G. & Burt, T. P. (eds.) 1985 *Hydrological forecasting*. Chichester: J. Wiley.

Bacon, F. 1620 *Novum organum*: see e.g. Basil Montague (ed. and trans.) (1854) The Works, 3 vols., Philadelphia: Parry & MacMillan, 3, 343-71.

Carrington, D. S., Bell, V. A. & Moore, R. J. 2000 Comparison of rainfall-runoff models for flood forecasting, part 2: Calibration and evaluation of models. Environment Agency R&D Technical Rep. W242, 239 pp.

Beck, M. B. 1983 Uncertainty, system identification and the prediction of water quality. In *Uncertainty and forecasting of water quality* (eds. M.B. Beck and G. Van Straten), pp 3-68. Berlin: Springer-Verlag.

Beck, M. B. & Young, P. C. 1975 A dynamic model for BOD-DO relationships in a non-tidal stream. *Water Research*, **9**, 769-776.

Beer, T. & Young, P. C. 1983 Longitudinal dispersion in natural streams. *Jnl. Env. Eng. Div., American Soc. Civ. Eng.*, **102**, 1049-1067.

Beven, K. J. 2000 Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences*, **4**, 203-213.

Beven, K. J. 2001 *Rainfall-runoff modelling: the primer*. Chichester: J. Wiley.

Beven, K. J. & Binley, A. M. 1992 The future of distributed models : model calibration and uncertainty prediction. *Hydrological Processes*, **6**, 279-298.

Beven, K. J., Romanowicz, R & Hankin, B. 2000 Mapping the probability of flood inundation (even in real time). In *Flood forecasting: what does current research offer the practitioner?* (M. J. Lees & P. Walsh eds.), pp 56-63. BHS Occasional paper No 12, produced by the Centre for Ecology and Hydrology on behalf of the British Hydrological Society.

Beven, K.J. & Kirkby, M.J. 1979 A physically-based variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, **24**, 43-69.

Billings, S.A. & Voon, W.S.F. 1986 Correlation based model validity tests for nonlinear models, *Int. Journal of Control*, **44**, 235-244.

Box, G. E. P. Jenkins, G. M 1970 *Time-series analysis, forecasting and control*. San Francisco: Holden-Day.

Brath, A, & Rosso, R. 1993 Adaptive calibration of a conceptual model for flash flood forecasting. *Water Resources Research*, **29**, 2561-2572.

Bryson, A. E. & Ho, Y-C. (1969) *Applied optimal control*. Massachusetts: Blaisdell Publishing.

Cluckie, I. D. 1993 Real-time flood forecasting using weather radar. In *Concise Encyclopedia of Environmental Systems* (ed. P. C. Young), pp 291-298. Oxford: Pergamon Press.

Crawford, N.H. & Linsley, R.K. 1966 Digital simulation in hydrology: the Stanford Watershed Model IV, Tech. Report no. 39, Stanford University, California.

Davis, H. T. & Jones, R. H. 1968 Estimation of the innovations variance of a stationary time series. *J. Amer. Statist.Assoc.*, **63**, 141-149.

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

Franks, S., Beven, K.J., Quinn, P.F. & Wright, I. 1997 On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration *Agric. Forest Meteorol.*, **86**, 63-75.

Gamerman, D. 1997 *Markov chain Monte Carlo*. London: Chapman & Hall.

Gordon, N., Marrs, A. & Salmond, D. 2000 Sequential analysis of nonlinear dynamic systems using particles and mixtures. In *Nonstationary and nonlinear signal processing* (W.J. Fitzgerald, A. Walden, R. Smith & P.C. Young eds.), pp 42-73. Cambridge: Cambridge University Press.

Greenfield, B. J. (1984) *The Thames Water Catchment Model*, Internal Report, Technology Development Division, Thames water, UK.

Hasselmann, K. 1998 Climate-change research after Kyoto. *Nature*, **390**, 225-226.

Hasselmann, K., Hasselmann, S., Giering, R., Ocana, V. & Storch, H. von 1997 Sensitivity study of optimal emission paths using a simplified structural integrated assessment (SIAM). *Climate Change*, **37**, 345-386.

Hornberger, G.M., Beven, K.J., Cosby, B.J. & Sappington, D.E. 1985 Shenandoah watershed study: calibration of a topography-based variable contributing area hydrological model for a small forested catchment. *Water Resources Research*, **21**, 1841-1850.

Hu, J., Kumamaru, K. and Hirasawa, K. (2001) A quasi-ARMAX approach to modelling nonlinear systems. *Int. Journal of Control*, **74**, 1754-1766.

Jakeman, A.J. & Hornberger, G.M. 1993 How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, **29**, 2637-2649.

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

Jakeman, A. J. & Young, P.C. 1979 Refined instrumental variable methods of recursive time-series analysis: part II, multivariable systems. *Int. Jnl. of Control*, **29**, 621-644.

Jang, J.-S.R., Sun, C.-T., & Mizutani, E. (1997) *Neuro-Fuzzy and Soft Computing*, Prentice Hall: Upper Daddle River, NJ.

Johnston, J. & DiNardo, J. 1997 *Econometric methods*. New York: McGraw Hill.

Jothityangkoon, C., Sivapalan, M. and Farmer, D. L. 2001 Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. *Journal of Hydrology*, **254**, 174-198.

Kalman, R. E. 1960 A new approach to linear filtering and prediction problems. *ASME Trans., Journal Basic Eng.*, **82-D**, 35-45.

Kirkby, M.J. 1976 Hydrograph modelling strategies. In *Processes in Physical and Human Geography* (R.J. Peel, M. Chisholm & P. Haggett eds.), pp, 69-90. London: Academic Press.

Kitagawa, G. 1996 Monte carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, **5**, 1-25.

Kitagawa, G. 1998 A self-organizing state-space model. *Journal of American Statistical Association*, **93**, 1203-1215.

Klemes, V. 1983 Conceptualisation and scale in hydrology, *Journal of Hydrology*, **65**, 1-23.

Konikow, L.F. & Bredehoeft, J.D. 1992 Ground water models cannot be validated. *Advances in Water Resources*, **15**, 75-83.

Kuhn, T. 1962 *The structure of scientific revolutions*. Chicago: University of Chicago.

Lawton, J. 2001 Understanding and prediction in ecology. Institute of Environmental and Natural Sciences. Lancaster University, Distinguished Scientist Lecture.

Lees, M. J. 2000a Advances in transfer function based flood forecasting. In *Flood forecasting: what does current research offer the practitioner?* (M. J. Lees & P. Walsh eds.), pp 41-55. BHS Occasional paper No 12, produced by the Centre for Ecology and Hydrology on behalf of the British Hydrological Society.

Lees, M. J. 2000b Data-based mechanistic modelling and forecasting of hydrological systems. *Journal of Hydroinformatics*, **2**, 15-34

Lees, M., Young, P.C., Beven, K. J., Ferguson, S. & Burns, J. 1994 An adaptive flood warning system for the River Nith at Dumfries. In *River flood hydraulics* (W. R. White & J. Watts eds.), pp 65-75. Wallingford: Institute of Hydrology.

Miller, D.R., Butler, G. and Bramall, C. 1976 Validation of ecological systems models, *Journal of Environmental Management*, **4**, 383-401.

Moore, R. J. 1985 The probability-distributed principle and run-off production at point and basin scales. *Hydrol. Sci. Jnl.*, **30**, 273-297.

Moore, R. J. & Jones, D. A. 1978 An adaptive finite difference approach to real-time channel flow routing. In *Modeling, identification and control in environmental systems* (G.C. Vansteenkiste ed.), 153-170. Amsterdam: North Holland.

Moore, R. J. and Bell, V. A. 2000 Comparison of rainfall-runoff models for flood forecasting, part 1: Literature review of models. Environment Agency R&D Technical Rep. W241, 94 pp.

Moore, R. J., Bell V. A. & Carrington D. S. 2000 Intercomparison of rainfall-runoff models for flood forecasting. In *Flood forecasting: what does current research offer the practitioner?* (M. J. Lees & P. Walsh eds.), pp 69-76. BHS Occasional paper No 12, produced by the Centre for Ecology and Hydrology on behalf of the British Hydrological Society.

Nash, J. E. 1959 Systematic determination of unit hydrograph parameters. *Journal of Geophysical Research*, **64**, 111-115.

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

Oreskes, N., Shrader-Frechette, K. & Belitz, K. 1994 Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, **263**, 641-646.

Parkinson, S.D. & Young, P.C. 1998 Uncertainty and sensitivity in global carbon cycle modelling. *Climate Research*, **9**, 157-174.

Popper, K. 1959 *The logic of scientific discovery*. London: Hutchinson.

Rigler, F. H. 1976 Review of "Systems Analysis and and Simulation in Ecology", (B. C. Patten ed.), *Limnology and Oceanography*, **21**, 481-483.

Romanowicz, R and Beven, K J 1998 Dynamic real-time prediction of flood inundation probabilities. *Hydrol. Sci. J.*, **43**, 181-196.

Saltelli, A., Chan, K and Scott, E. M. 2000 *Sensitivity Analysis*, J. Wiley: Chichester.

Sorooshian, S 1985 Development of rainfall-runoff models. *Applied Mathematics & Computation*, **17**, 279-298.

Schweppe, F. 1965 Evaluation of likelihood functions for Gaussian signals. *IEEE Trans. on Information Theory*, **11**, 61-70.

Shackley, S., Young, P.C., Parkinson, S.D. & Wynne, B. 1998 Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools? *Climate Change*, **38**, 159-205.

Shaw, E. M. 1994 *Hydrology in practice: 3rd edition*. London: Chapman & Hall.

Silvert, W. 1993 Top-down modelling in ecology. In *Concise Encyclopedia of Environmental Systems* (P. C. Young ed.), 60. Oxford; Pergamon Press.

Simon, H. A. 1967 The architecture of complexity, in *General Systems: the Year Book of the Society of General Systems Research*, Volume 10, Brazilier: New York, 63-76.

Singh, V. P. (ed.) 1995 *Computer models of watershed hydrology*. Colorado: Water Resources Publications.

Thiemann, M., Trosset, M., Gupta, H. & Sorooshian, S. 2001 Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research*, **37**, 2521-2535.

Todini, E. 1996 The ARNO rainfall-runoff model. *Journal of Hydrology*, **175**, 339-382

Tokar, A. S. & Johnson, P. A. 1999. Rainfall-runoff modeling using artificial neural networks. *ASCE Journal of Hydrological Engineering*, **4**, 232-239

Wallis, S.G., P.C. Young & Beven, K. J. 1989 Experimental investigation of the Aggregated Dead Zone (ADZ) model for longitudinal solute transport in stream channels. *Proc. Inst. of Civil Engrs: Part 2*, **87**, 1-22.

Weyman, D. R. 1974 *Runoff processes and streamflow modelling*. Oxford: Oxford University Press.

Wheater, H.S., Jakeman, A.J. & Beven. K.J. 1993 Progress and directions in rainfall-runoff modelling. In *Modelling change in environmental systems* (A.J. Jakeman, M.B. Beck & M.J. McAleer eds.), pp 101-132. Chichester: Wiley.

Whitehead, P.G. & Young, P.C. 1975 A dynamic-stochastic model for water quality in part of the Bedford-Ouse River system. In *Computer simulation of water resources systems* (G.C. Vansteenkiste ed.), pp 417-438. Amsterdam: North Holland.

Whitehead, P.G. & Young, P.C. 1979 Water quality in river systems: Monte Carlo analysis. *Water Resources Research*, **15**, 451-459.

Whitehead, P. G., Young, P. C. & Hornberger G. H. (1976) A systems model of stream flow and water quality in the Bedford-Ouse river - Part l: Stream Flow Modeling, *Water Research*, **13**, 1155-1169.

Wood, E. F. (ed.) 1985 Special issue: system identification and parameter estimation of environmental systems. *Applied Mathematics & Computation*, **17**, 277-482.

Wood, E. F. & O'Connell, P. E. 1985 Real-time forecasting. In *Hydrological forecasting* (M. G. Anderson and T. P. Burt eds.), pp 505-558. Chichester: J. Wiley.

Ye, W., Jakeman, A. J. & Young, P. C. 1998 Identification of improved rainfall-runoff models for an ephemeral low yielding Australian catchment. *Environmental Modelling & Software*, **13**, 59-74.

Yevjevich, V. 1987 Stochastic models in hydrology. *Stochastic Hydrology & Hydraulics*, **1**, 17-36.

Young, P.C. 1974 Recursive approaches to time-series analysis. *Bull. of Inst. Maths and its Applications*, **10**, 209-224.

Young, P.C. 1978 A general theory of modeling for badly defined dynamic systems. In *Modeling, identification and control in environmental systems* (G.C. Vansteenkiste ed.), pp 103-135. Amsterdam: North Holland.

Young, P.C. 1981 Parameter estimation for continuous-time models - a survey. *Automatica*, **17**, 23-29.

Young, P.C. 1983 The validity and credibility of models for badly defined systems. In *Uncertainty and forecasting of water quality* (M.B. Beck & G. Van Straten eds.), pp 69-100. Berlin: Springer-Verlag.

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

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

Young, P.C. 1986 Time-series methods and recursive estimation in hydrological systems analysis. In *River flow modelling & forecasting* (D.A. Kraijenhoff & J.R. Moll eds.), pp 129-180. Dordrecht: D. Reidel.

Young, P. C. 1989 Recursive estimation, forecasting and adaptive control. In *Control and dynamic systems: advances in theory and applications, Vol. 30* (C.T. Leondes ed.), pp 119-166. San Diego: Academic Press.

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

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. 1996 Identification, estimation and control of continuous-time and delta operator systems. In *Identification in Engineering systems* (M. I. Friswell & J. E. Mottershead eds.), pp 1-17. Swansea: University of Wales.

Young, P.C. 1998a Data-based mechanistic modelling of engineering systems. *Journal of Vibration & Control*, **4**, 5-28.

Young, P.C. 1998b 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.), pp 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 & P. D. Bates eds.), pp 117-161. Chichester: J. Wiley.

Young, P. C. 2001c Comment on 'A quasi-ARMAX approach to the modelling of nonlinear systems' by J. Hu *et al.*, *Int. Journal of Control*, **74**, 1767-1771.

Young, P. C. 2002a Advances in real-time flood forecasting. *Phil. Trans. R. Soc. Lond.*, **360**, 1433-1450.

Young, P. C. 2002b Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment Scale. *Hydrological Processes*, in press.

Young, P.C. & Beven, K.J. 1994 Data-based mechanistic modelling and the rainfall-flow nonlinearity, *Environmetrics*, **5**, 335-363.

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

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

Young, P.C. & Lees, M. J. 1993 The active mixing volume: a new concept in modelling environmental systems. In *Statistics for the Environment* (V. Barnett & K. F. Turkman eds.), pp 3-43. Chichester: J. Wiley.

Young, P.C. & Minchin, P. 1991 Environmetric time-series analysis: modelling natural systems from experimental time-series data. *Int. Jnl. Biol. Macromol.*, **13**, 190-201.

Young, P. C. & Parkinson, S. D. 2002 Simplicity out of complexity. In *Environmental foresight and models* (M. B. Beck ed.), Amsterdam: Elsevier. (in press)

Young, P.C & Pedregal, D. 1998 Data-based Mechanistic Modelling. In *System dynamics in economic and financial models* (C. Heij & H. Schumacher eds.), pp 169-213. J. Wiley: Chichester, .

Young, P.C & Pedregal, D. 1999a Macro-economic relativity: Government spending, private investment and unemployment in the USA 1948-1998. *Structural Change & Economic Dynamics*, **10**, 359-380.

Young, P.C & Pedregal, D. 1999b Recursive and en-bloc approaches to signal extraction. *Journal of App. Stats.*, **26**, 103-128.

Young, P. C. & 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. & 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. & Wallis S, G. 1994 Solute transport and dispersion in channels. In *Channel networks* (K.J. Beven & M.J. Kirkby eds.), pp 129-173. Chichester: J. Wiley.

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

Young, P.C., McKenna, P. & Bruun, J. 2001 Identification of non-linear stochastic systems by state dependent parameter estimation. *Int. Journal of Control*, **74**, 1837-1857.

Young, P. C., Parkinson, S.D., & Lees, M. 1996 Simplicity out of complexity in environmental systems: Occam's Razor revisited. *J. Appl. Stat.*, **23**, 165-210.

Young, P.C., Schreider, S. Yu., & Jakeman, A.J. 1997b A streamflow forecasting algorithm and results for the Upper Murray Basin. In *Proc. MODSIM 97 Congress on Modelling and Simulation, Hobart, Tasmania.*