CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Real-Time Forecasting
Tools and Applications

Peter C. Young
Centre for Research on Environmental Systems and Statistics
Lancaster University; and ICAM, CRES, Australian National University

Prepared for

G-WADI Modelling Workshop
28th February-3rd March, 2005

National Institute of Hydrology
Roorkee 247 667, INDIA

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Contents

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Real-Time Updating

There are two major types of real time model updating:

- parameter updating, where the estimates of some (possibly all) of the *model parameters* are updated regularly on the basis of incoming data (e.g. rainfall, flow, etc.)

- state updating, where estimates of the *state variables* in the model are updated regularly on the basis of incoming data.

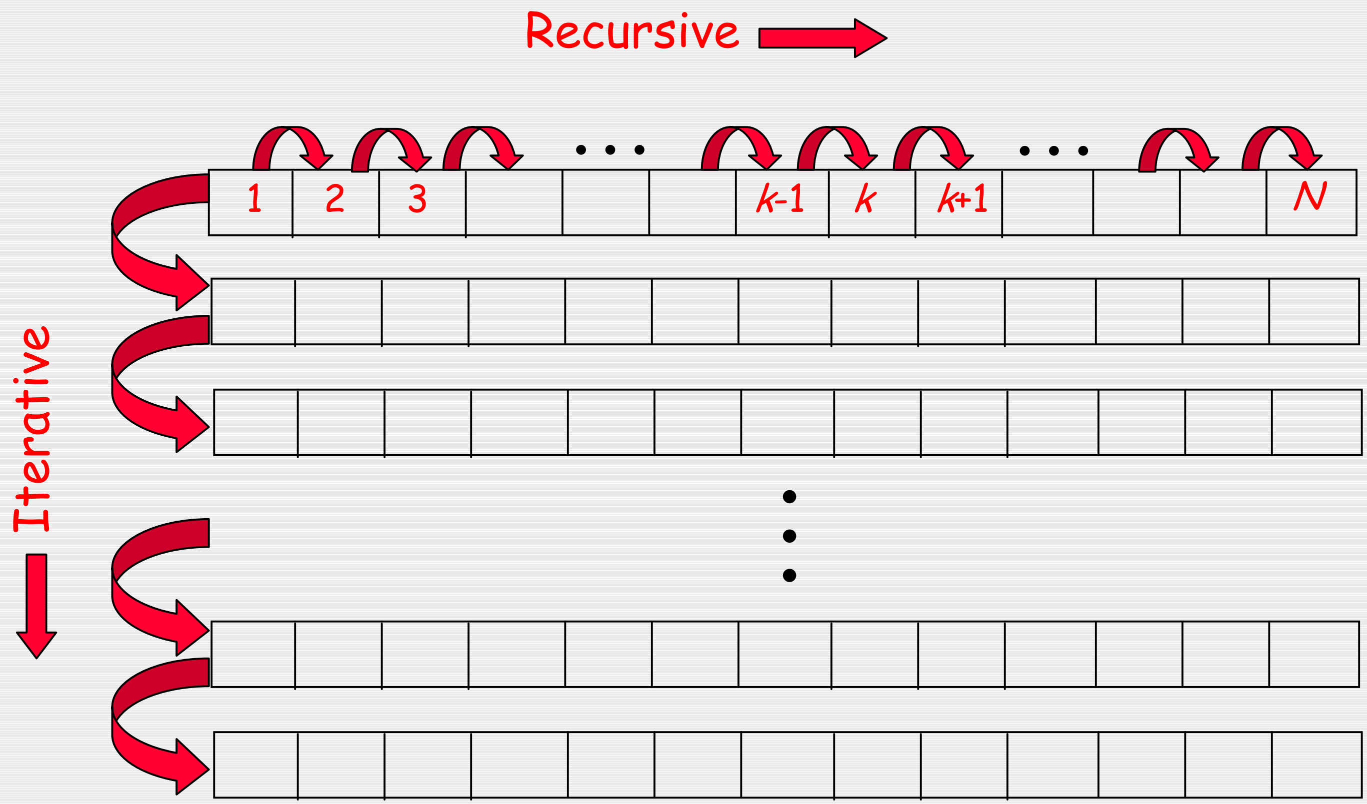
Sometimes these updating operations are carried out in a fully integrated manner (e.g. in some form of parameter-state estimation algorithm: e.g the *Extended Kalman Filter*). Alternatively, they are carried concurrently but in separate algorithms. These algorithms are normally termed *Recursive Estimation Algorithms.*

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Recursive Processing of Data

Model Equation: $\mathbf{x}_t = f\{\mathbf{x}_{t-1}, \mathbf{a}_t, \xi_t\}$

Observation Equation: $\mathbf{y}_t = g\{\mathbf{x}_t, \eta_t\}$



CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Recursive Estimation Algorithms

(originated by Gauss, sometime before 1826: see Young, 1984)

The generic form of the recursive parameter estimation algorithm is as follows:

Innovations process (1 step ahead prediction)

$$\hat{\mathbf{a}}_t = \hat{\mathbf{a}}_{t-1} + \mathbf{G}_t\{\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\} \quad ; \quad \hat{\mathbf{y}}_{t|t-1} = f\{\hat{\mathbf{a}}_{t-1}, \mathbf{y}_{t-1}\}$$

While the generic form of the state estimation algorithm is:

c.f Model Equation

Innovations process

$$\text{Prediction:} \quad \hat{\mathbf{x}}_{t|t-1} = f\{\hat{\mathbf{x}}_{t-1}\hat{\mathbf{a}}_{t-1}\}$$

$$\text{Correction:} \quad \hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t-1} + \mathbf{G}_t\{\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\}$$

where $\mathbf{y}_t = g\{\mathbf{x}_t\}$ is the observed data that is related to the state variables of the model in some defined manner; and $\mathbf{G}_t$ is a time variable matrix that is also computed recursively and is a function of the uncertainty in the parameter or state estimates.

An algorithm that combines these two recursive estimation operations is often called a *Data Assimilation Algorithm*.

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

The Problem of Identifiability

- An *identifiable* model is one whose parameters can be estimated unambiguously from the available data.
- The number of parameters that can be identified unambiguously is related to the information content of the data and a model that has more parameters than this is *over-parameterized* and unidentifiable.
- Normally identifiable models are characterized by relatively few parameters (but they can often explain and forecast the data well).
- An over-parameterized model does not provide a unique description of the data and this can result in *ambiguity or equifinality*: i.e. the existence of more that one model that can characterize the data equally well.
- The recursive estimates of parameters in identifiable models are well-defined statistically; while those in poorly identified models are either poorly defined (very high variance) or unobtainable without the imposition of constraints.

The Danger of Imposing Subjective Constraints

- In order to estimate the parameters in an over-parameterized model, some of the parameters need to be constrained, either deterministically or stochastically.
- In the deterministic case, this means defining a subset of 'well-known' parameters whose parameters are known exactly, with no uncertainty.
- In the stochastic case, the parameters are defined as probability distributions and the recursive (Bayesian) estimation proceeds accordingly, either using a recursive algorithm or some form of numerical Bayesian estimation. However, for this to work satisfactorily in the over-parameterized situation, the priors on a subset of the parameters need to be to be tightly defined: i.e. it is the stochastic equivalent of the deterministic approach.
- In either case, the dangers are obvious: the model will depend upon the efficacy of the subjectively defined constraints on the parameter estimates.
- Adaptive estimation can be difficult in such models
- In these circumstances, the modeller is placing great faith in prior judgement. An alternative is to exploit Data-Based Mechanistic (DBM) Modelling

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Data-Based Mechanistic (DBM) Modelling

(see e.g. Young and Lees, 1993; Young, 2003 and the prior references therein)

Data-Based Mechanistic (DBM) modelling has been developed at Lancaster over the past twenty years but is derived from research dating back to the early 1970's.

The general approach of DBM modelling is as follows:

1. Define the objectives of the modelling exercise.
2. Prior to the acquisition of data (or in situations of insufficient data), develop a simulation model that satisfies the defined objectives; evaluate the sensitivity of the model to uncertainty and develop a reduced order, dominant mode model that captures the most important aspects of its dynamic behaviour.
3. When sufficient data become available: identify and estimate from these data a parsimonious, dominant mode, stochastic model that again satisfies the defined objectives; is identifiable (reflects the information content in the data); and can be interpreted in physically meaningful terms.
4. Attempt to reconcile the models obtained in 1. and 2.
5. Apply the model in data assimilation,forecasting and control system design.

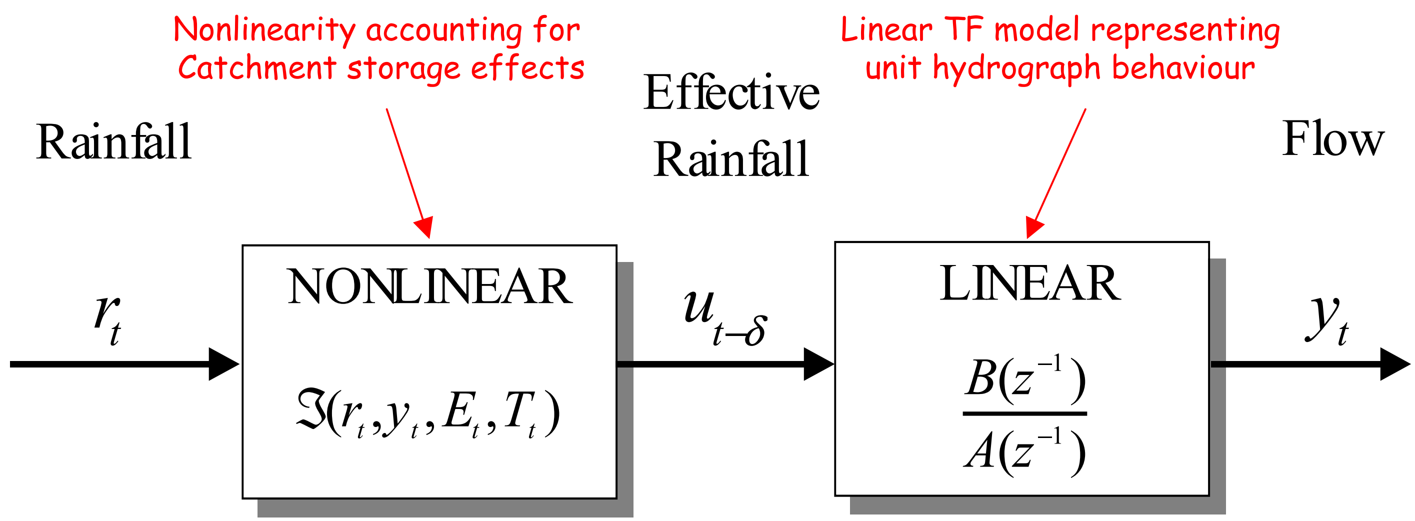
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

DBM Models of Rainfall-Flow (or Stage) and Real Time Updating

(see e.g. Young, 1993; Young and Beven,1994; Young, 2003 and the prior references therein)

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

A Generic Form for DBM Rainfall-Flow Models



z^{-1} is the backward shift operator (sometimes denoted by L or B):
i.e. $z^{-r} y_t = y_{t-r}$

Note that in DBM modelling, the model structure and order, such as this, *is inferred from the data and not assumed prior to estimation*: i.e. a wide variety of different model structures are evaluated statistically and the identified model is the one which best satisfies the statistical identification criteria ***and*** has an acceptable mechanistic interpretation (see e.g. Young, 2003 and the prior references therein).

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

General Multi-Order Dicrete and Continuous Time TF Model Forms

Transfer functions can represent discrete-time linear systems:

$$y_k = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}} u_{k-\delta}$$

i.e. $y_k = -a_1 y_{k-1} - \ldots - a_n y_{k-n} - b_0 u_{k-\delta} - \ldots - b_m u_{k-\delta-m}$

But they can also represent continuous-time differential equations:

$$y(t) = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1} s + b_m}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s^{n-2} + a_n} u(t-\tau)$$

i.e. $$\frac{d^n y(t)}{dt^n} + a_1 \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots \; a_n y(t) = b_0 \frac{d^m u(t-\tau)}{dt^m} + b_1 \frac{d^{m-1} u(t-\tau)}{dt^{m-1}} + \cdots + b_m u(t-\tau)$$

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Stochastic Transfer Function Models
1. Rainfall-Flow Example (daily data)

The TF model for the Canning River considered later takes the form:

$$y_t = \frac{0.06 + 0.10z^{-1} - 0.141z^{-2}}{1 - 1.603z^{-1} + 0.623z^{-2}} u_t + \xi_t$$

or

$$y_t = 1.603y_{t-1} - 0.623y_{t-2} + 0.06u_t + 0.10u_{t-1} - 0.141u_{t-2} + \eta_t$$

which can be decomposed into the following form:

$$y_t = 0.06u_t + \frac{0.026z^{-1}}{1 - 0.942z^{-1}} u_t + \frac{0.171z^{-1}}{1 - 0.661z^{-1}} u_t + \xi_t$$

Instantaneous (within one day) | Slow Pathway | Quick Pathway

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Stochastic Transfer Function Models
2. Rainfall-Stage and Stage Routing Example (hourly data)

The complete DBM model for the River Severn consists of several TF modules some for rainfall-stage modelling where the TF model is typically of the form:

$$y_t = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_{t-\delta} + \xi_t$$

These are, once again, decomposed into the physically meaningful parallel pathway form.

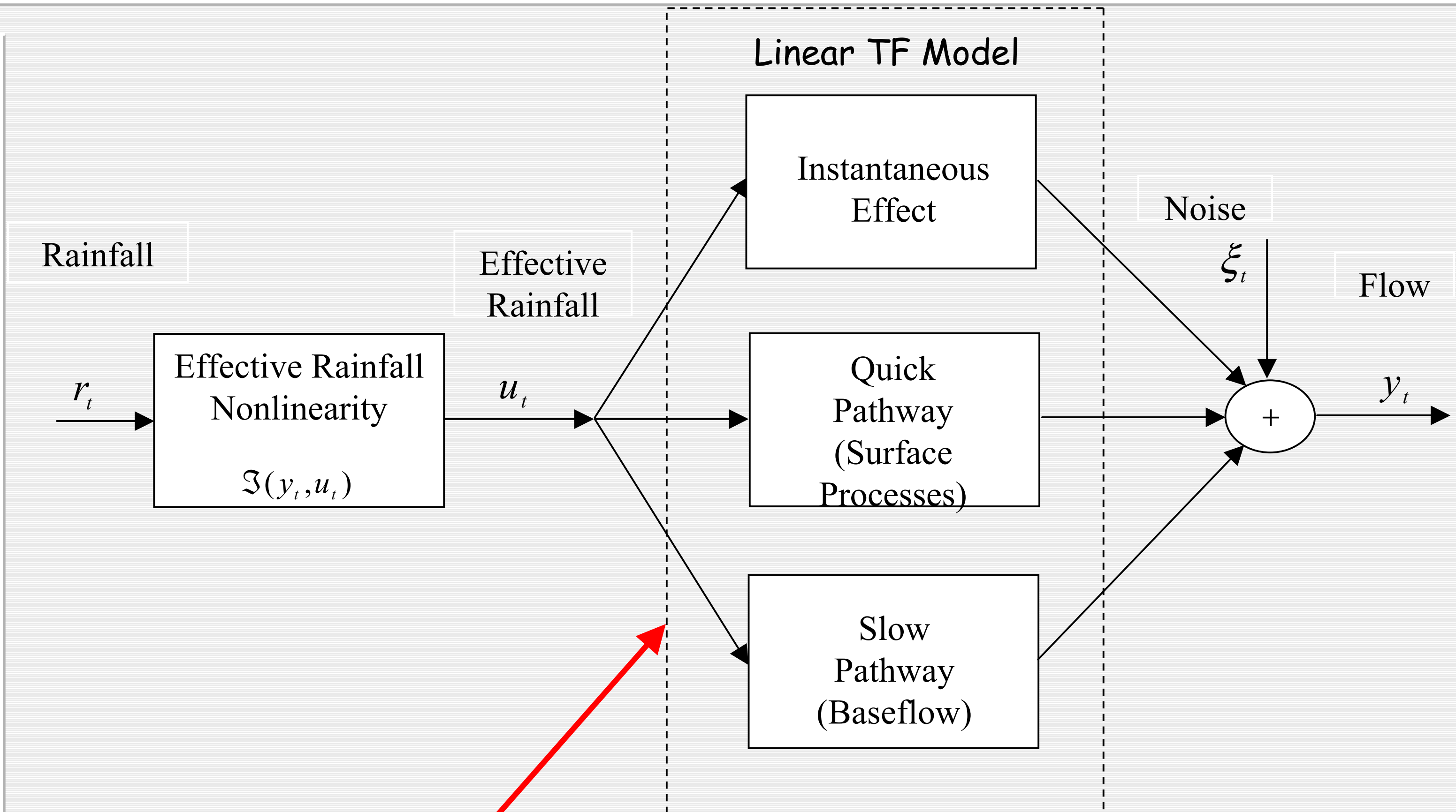
$$y_t = \frac{\alpha_1 z^{-1}}{1 - \beta_1 z^{-1}} u_t + \frac{\alpha_2 z^{-1}}{1 - \beta_2 z^{-1}} u_t + \xi_t$$

In the case of the stage-stage routing modules, the model is identified in a 'cascade' of 1st order TFs (similar to the Nash cascade but with normally non-identical elements), each of the form:

$$y_{i,t} = \frac{b_0}{1 + a_1 z^{-1}} y_{i,t-\delta i} + \xi_{i,t} \quad i = 1, 2, \ldots, n$$

where $y_{i,t-\delta i}$ is the upstream stage and $y_{i,t}$ is the downstream stage.

Parallel-Flow Decomposition and Physical Interpretation



Decomposed TF model: this physically meaningful parallel structure is not imposed: it is a natural decomposition of the TF model

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

The DBM State Space Model

The DBM State Space(SS) model is derived from the TF models. For example, in the case of the rainfall-stage model it is obtained from the TF model *decomposed into its physically meaningful parallel pathway form*:

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}u_{t-\delta} + \zeta_t$$

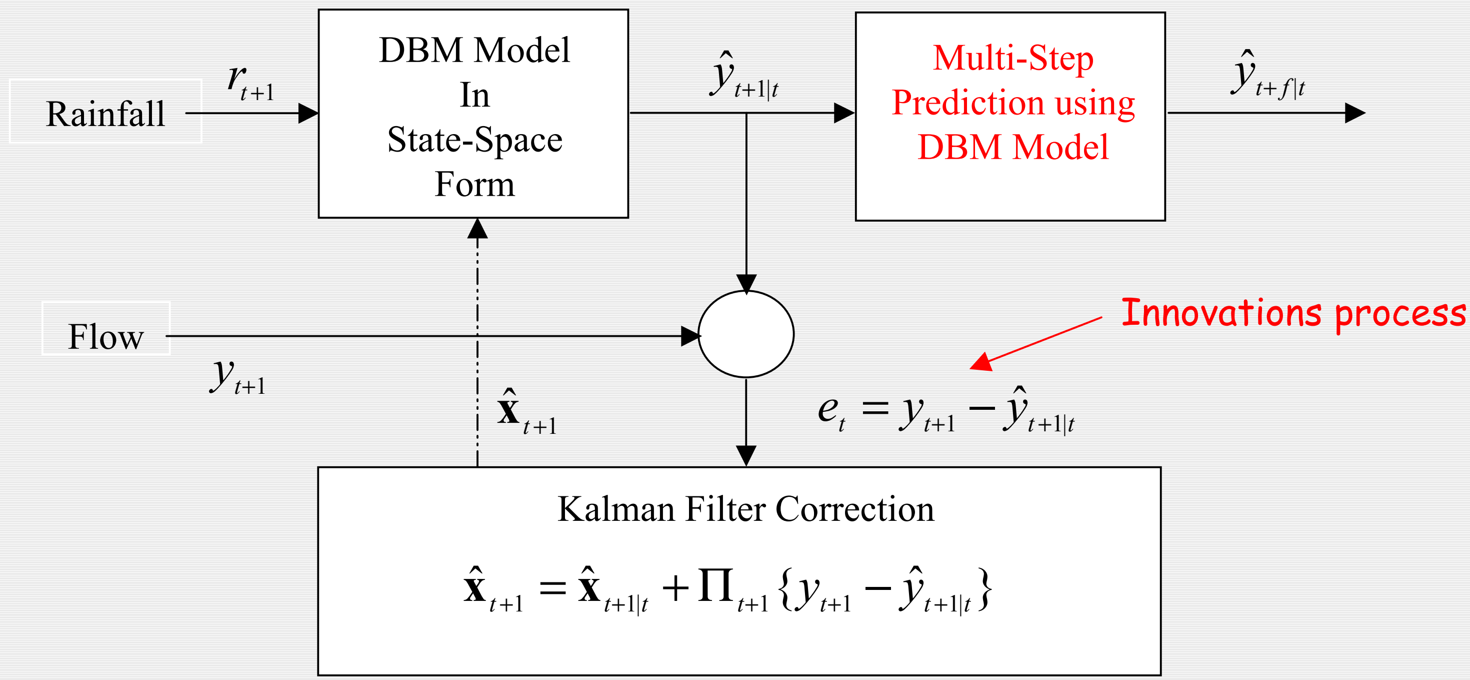
$$y_t = \mathbf{h}^T\mathbf{x}_t + e_t \qquad e_t = N(0,\sigma_t^2)$$

where,

$$\mathbf{x}_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} -\alpha_1 & 0 \\ 0 & -\alpha_2 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \zeta_k = \begin{bmatrix} \zeta_{1,k} \\ \zeta_{2,k} \end{bmatrix} \quad \mathbf{h}^T = [1 \quad 1]$$

and the parameters α_i and β_i, $i=1,2$, are the parameters of the decomposed TFs in each of the two (quick and slow) parallel pathways. *Note that not only does this model have better physical significance than the originally estimated TF form and its normal 'canonical' SS form, but the forecasting performance is **much** better.*

Adaptive Kalman Filter-Based Forecasting Using a DBM Model



Adaptive parameter updating can be carried out in parallel for model both DBM model parameters or other 'hyper-parameters

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

DBM Models of Rainfall-Flow (or Stage) and Real Time Updating

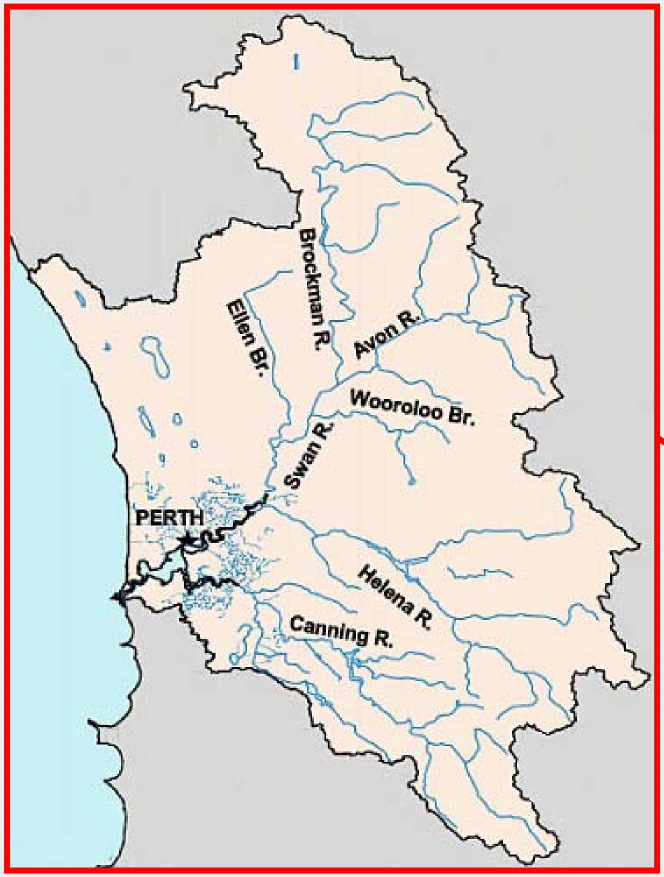
(see e.g. Young, 1993; Young and Beven,1994; Young, 2003 and the prior references therein)

Example 1: DBM Modelling of the Ephemeral Canning River, Western Australia

The Canning River at Glen Eagle in SW Australia is a 544 km^2 benchmark catchment. The discharge is dominated by zero flows, with zero flow periods occupying more than half of the recorded period. The rainfall is Winter dominated, with the Winter 4 months receiving approx. 70% of the total annual rainfall.

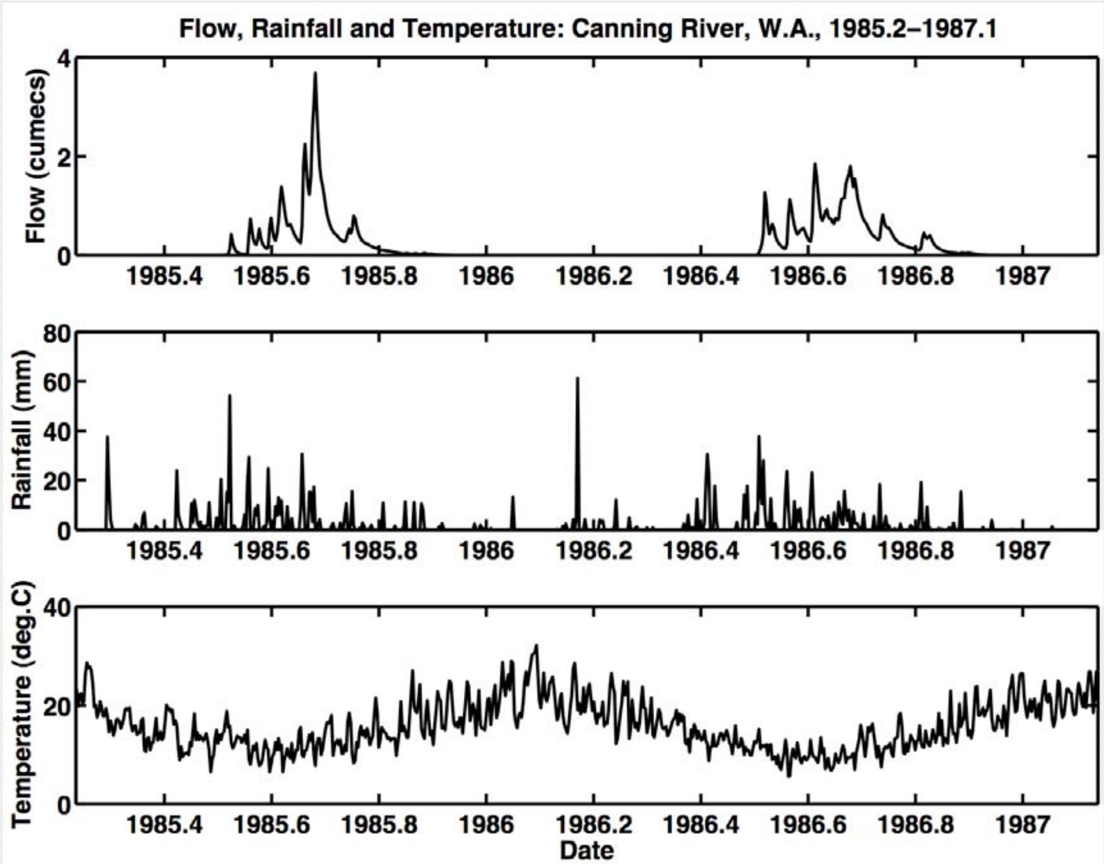
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

The Swan-Canning River System, W.A.



CRESearch
The Key to Success in Environmental and Biological Systems Analysis

DBM Model Identification and Estimation: Rainfall, Flow and Temperature Data (1985-1987)

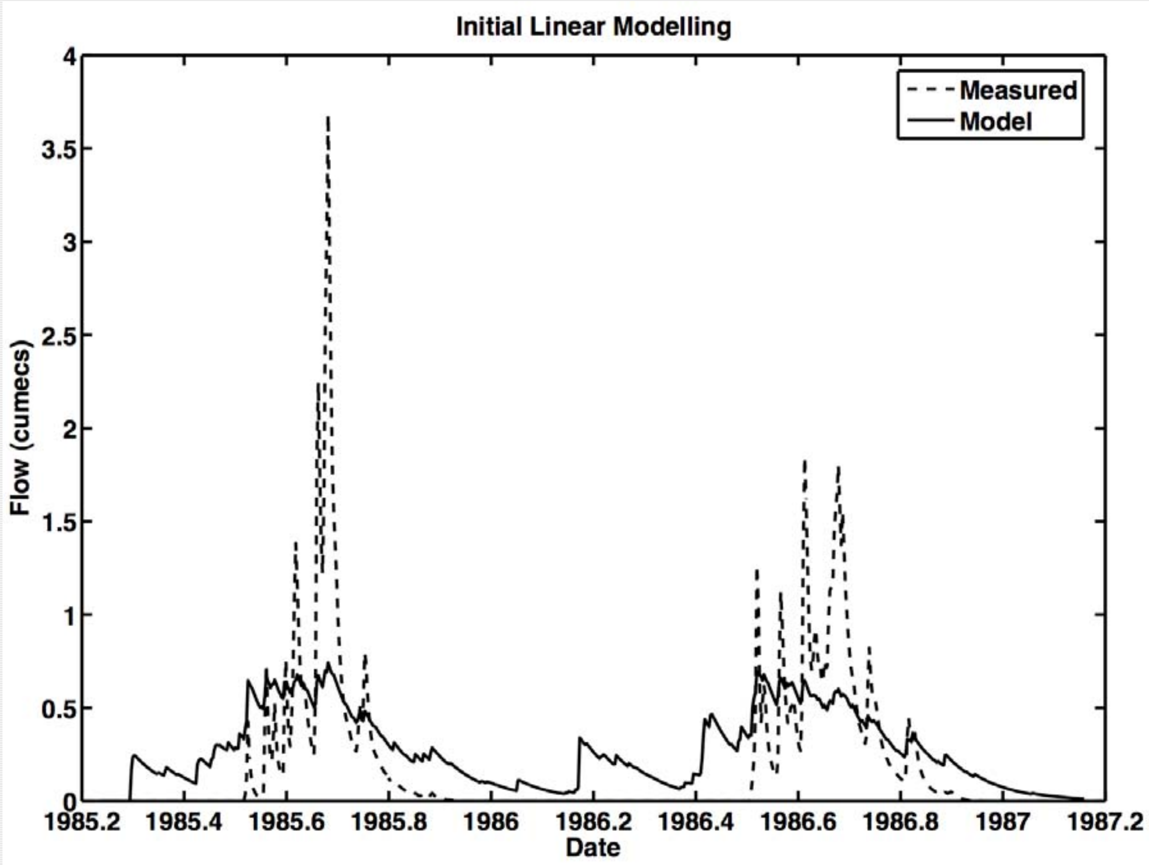


Stage 1: DBM Model Identification and Estimation

Since there is adequate data in this example, real-time forecasting system design starts with the data-based identification and estimation of the DBM Model.

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Initial Linear Modelling



Time Variable Parameter (TVP) Estimation

Young (1984)

- Although it is not possible to explain the rainfall-flow dynamics by a linear model, it is possible to do so with a *Time Variable Parameter* (TVP) model in which the parameters in the linear TF model are allowed to change and are estimated using a recursive TVP estimation algorithm.
- The denominator parameters do not change appreciably but the numerator parameters, which define the gain of the transfer function, are estimated to change considerably. Moreover, these changes are highly correlated with the temporal changes in the flow series.
- This suggests that the numerator parameters, which multiply the rainfall input, are dependent on the flow and so an 'effective' rainfall should be estimated from the measured rainfall through a *State-Dependent Parameter* (SDP) function of the flow.
- Of course, it is not possible that this SDP is a direct function of the flow, so it seems reasonable, on physical grounds, to assume that the flow is acting as a surrogate measure of the catchment storage.
- The multiplication of this SDP by the rainfall creates the *Effective Rainfall Nonlinearity* and this accounts for the effects of prior rainfall on the catchment storage.

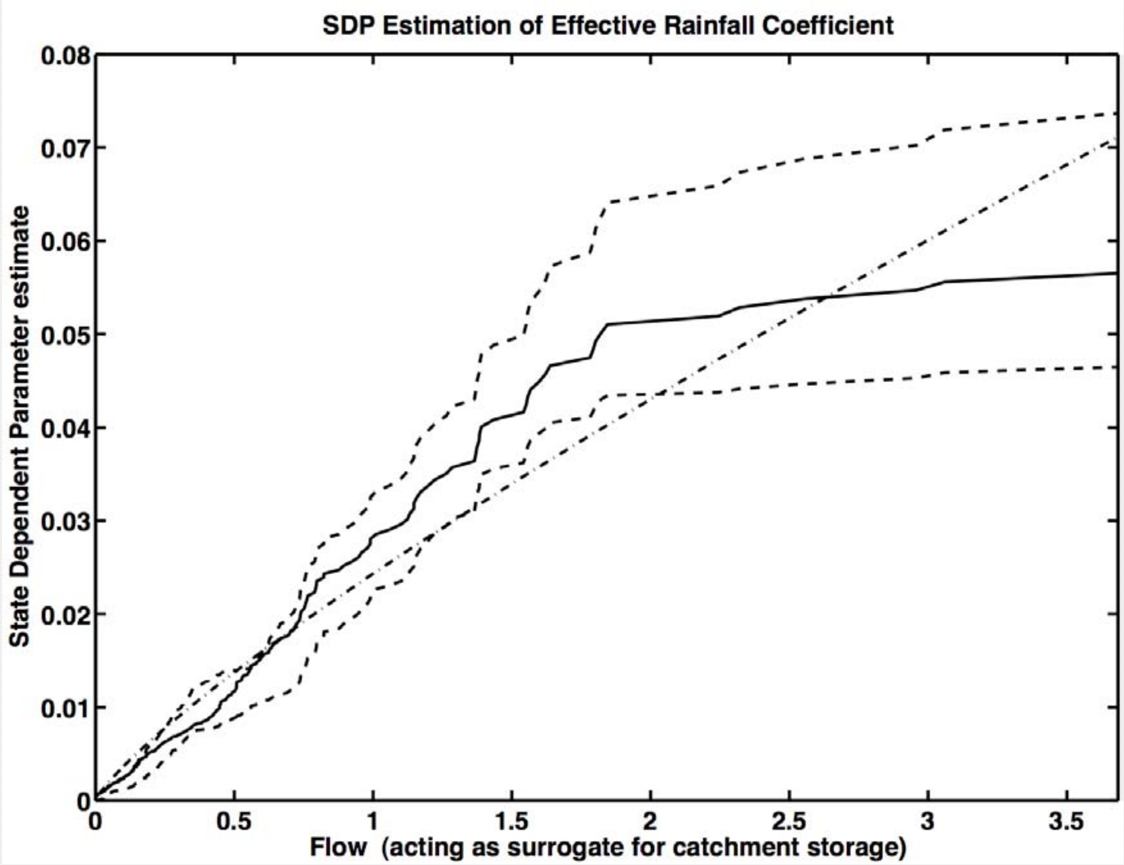
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

State Dependent Parameter (SDP) Estimation

Young (1993, 2000,2001)

- Although algorithms for the recursive estimation of *slowly* time variable model parameters are well known, the *optimal* versions of these algorithms are not well known.
- The backward-recursive, *Fixed Interval Smoothing* (FIS) part of the algorithms are particularly important.
- At first sight, such 'filtering/smoothing' algorithms cannot be applied to SDP estimation unless the SDPs vary slowly in time, which is unlikely in general applications.
- However, *time series do not have to be processed in temporal order*: if the variables associated with the time dependency are sorted, such that they vary slowly (e.g. ascending magnitude order: the Matlab 'sort'), then the SDPs will also vary slowly in this same ordering and can be estimated using the optimised TVP algorithms.

SDP Estimation of the Effective Rainfall Nonlinearity



Non-parametric SDP estimate (full line); SE bounds (dashed lines); parametric estimate based on power law parameterization (dash-dot line).

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Parameterized DBM Model: Identification and Estimation

The complete DBM model for the River Canning is identified and estimated as follows:

$$y_t = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_t + \xi_t \qquad u_t = \{c.y_t{}^{\gamma}\} r_t$$

Estimated SDP *Effective Rainfall Nonlinearity* parameterized as a power law

Where ξ_t is an AR(25) noise process. This decomposes to:

$$y_t = k u_t + \frac{\beta_1 z^{-1}}{1 - \alpha_1 z^{-1}} u_t + \frac{\beta_2 z^{-1}}{1 - \alpha_2 z^{-1}} u_t + \xi_t$$

fast pathway   slow pathway

The associated parameter estimates are as follows:

$$\hat{a}_1 = -1.6031(0.012); \quad \hat{a}_2 = 0.6228(0.011); \quad \hat{\gamma} = 0.823(0.09);$$

$$\hat{b}_1 = 0.0601(0.005); \quad \hat{b}_2 = 0.100(0.010); \quad \hat{b}_3 = -0.1405(0.006);$$

$$k = 0.06; \quad \hat{\beta}_1 = 0.171; \quad \hat{\alpha}_1 = 0.661; \quad \hat{\beta}_2 = 0.026; \quad \hat{\alpha}_2 = 0.942;$$

Note: the parameter c is a normalization coefficient selected so that $\sum_{t=1}^{t=N} u_t = \sum_{t=1}^{t=N} y_t$

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Parameterized DBM Model: Parallel Decomposition

The TF parallel decomposition details area are as follows:

Instantaneous Pathway

Root	SSG	TC	%flow
0	0.06	0	5.9

Fast Pathway

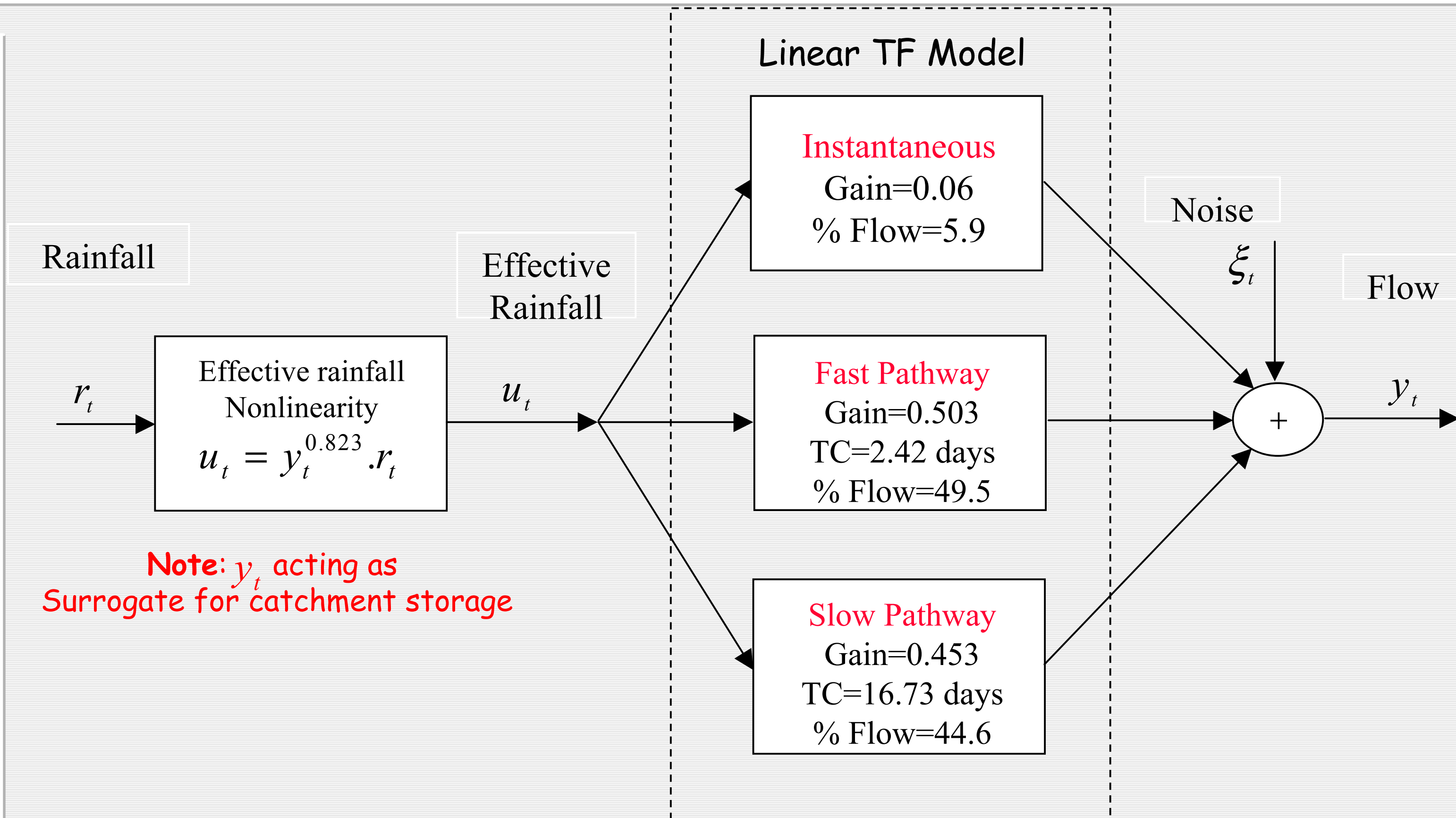
Root	SSG	TC	%flow
0.661	0.503	2.42d	49.5

Slow Pathway

Root	SSG	TC	%flow
0.942	0.453	16.73d	44.6

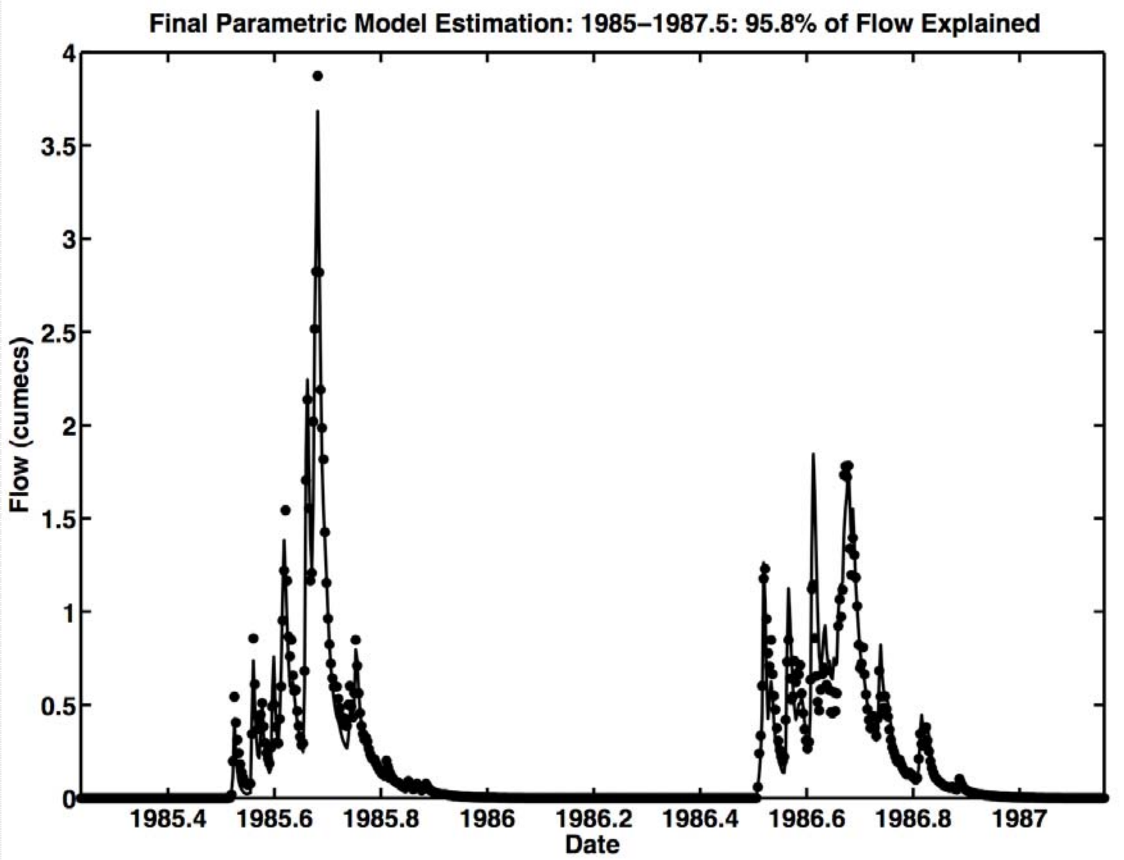
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Parallel-Flow Decomposition and Physical Interpretation



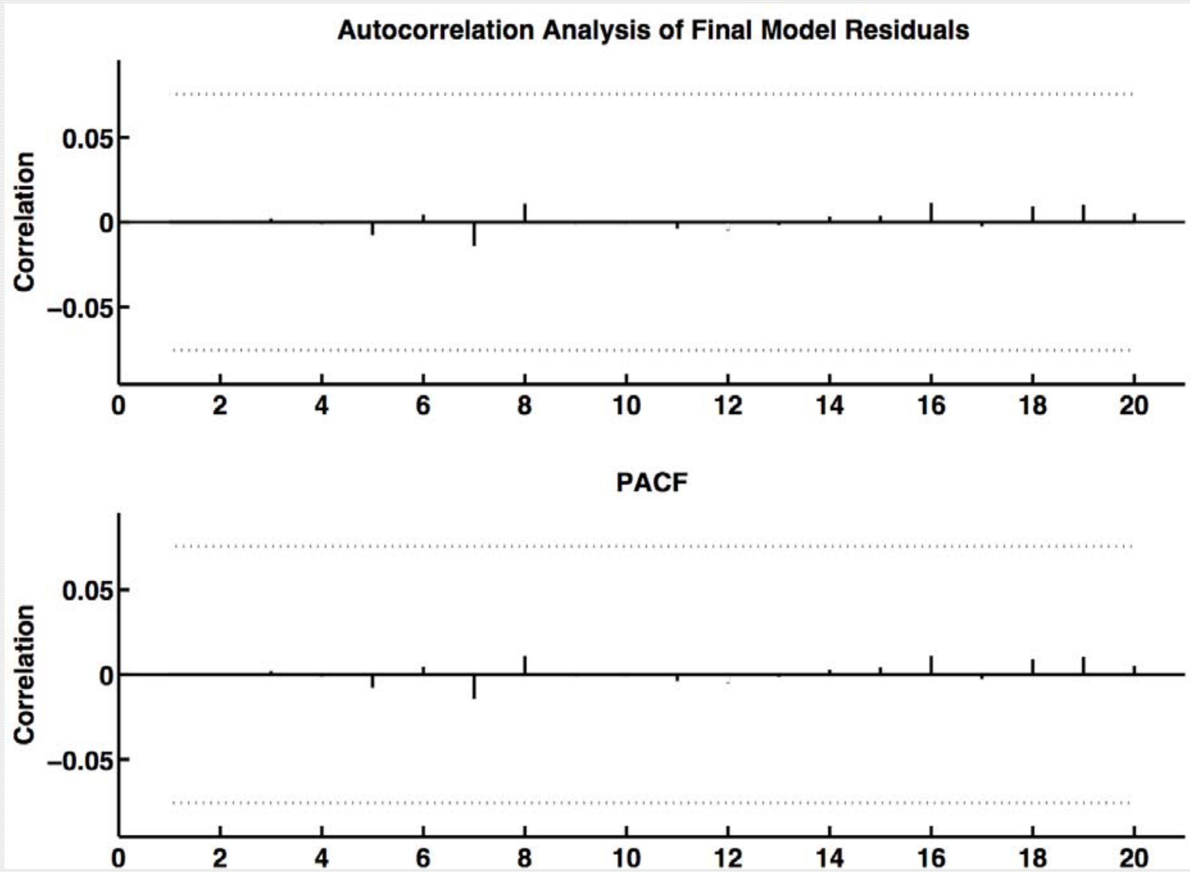
Note: y_t acting as
Surrogate for catchment storage

Model Output of Final DBM Parametric Model



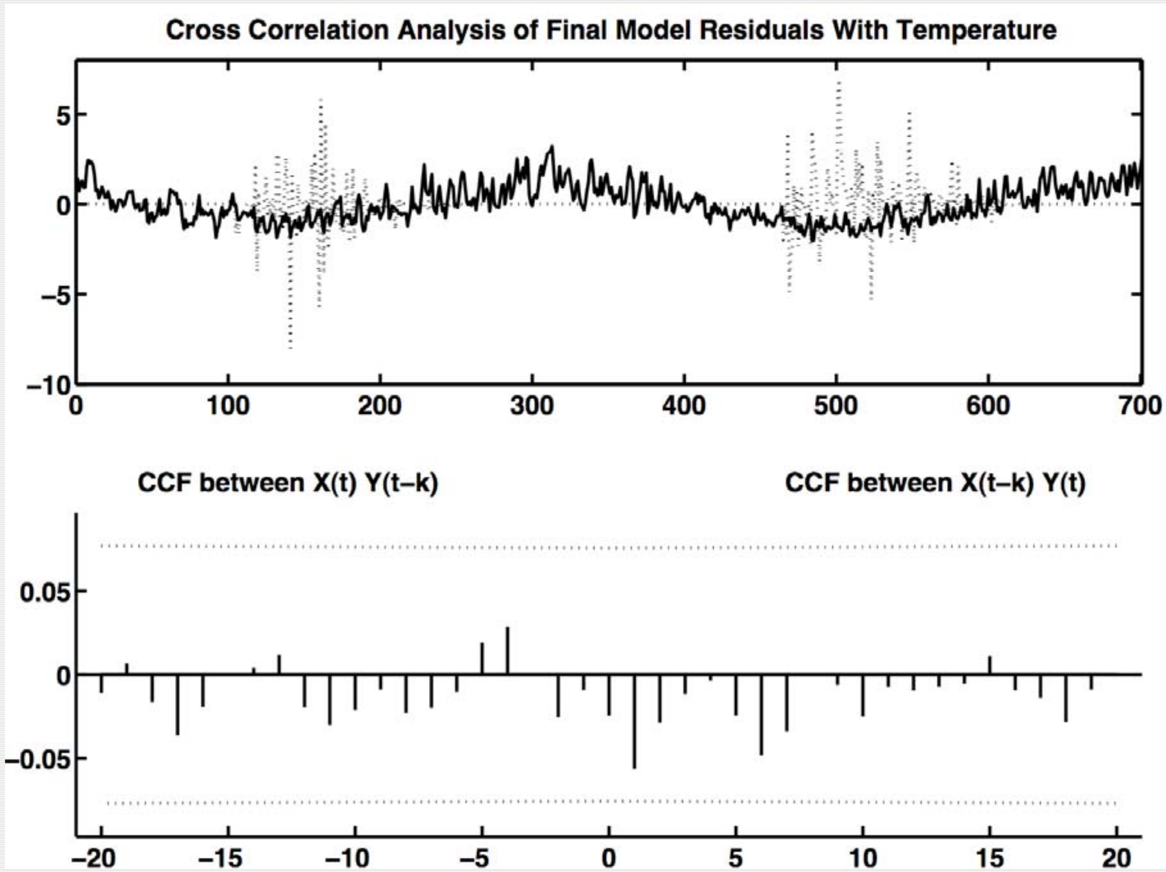
DBM model output (full line); Daily measured flow (dots).

Statistical Diagnostic Testing: Autocorrelation of DBM Model Stochastic Residuals



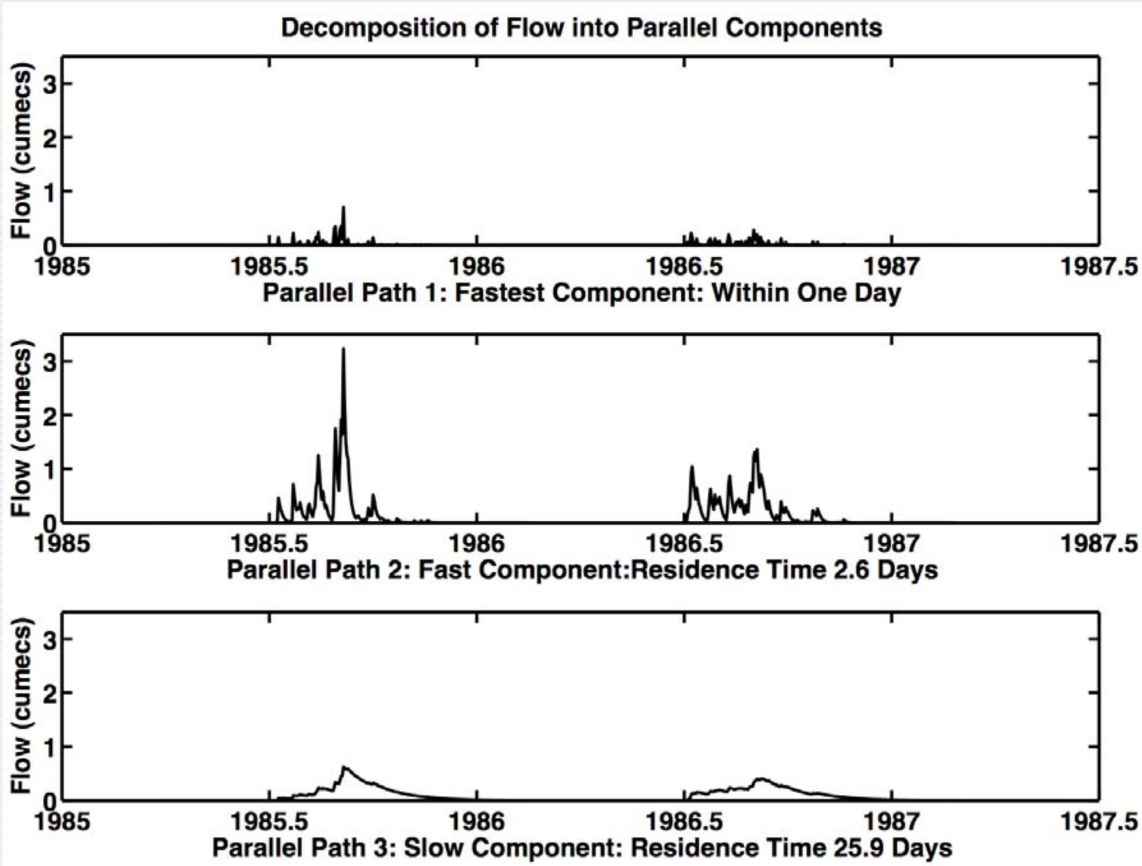
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Statistical Diagnostic Testing: Cross-correlation of Stochastic Residuals and Temperature Series



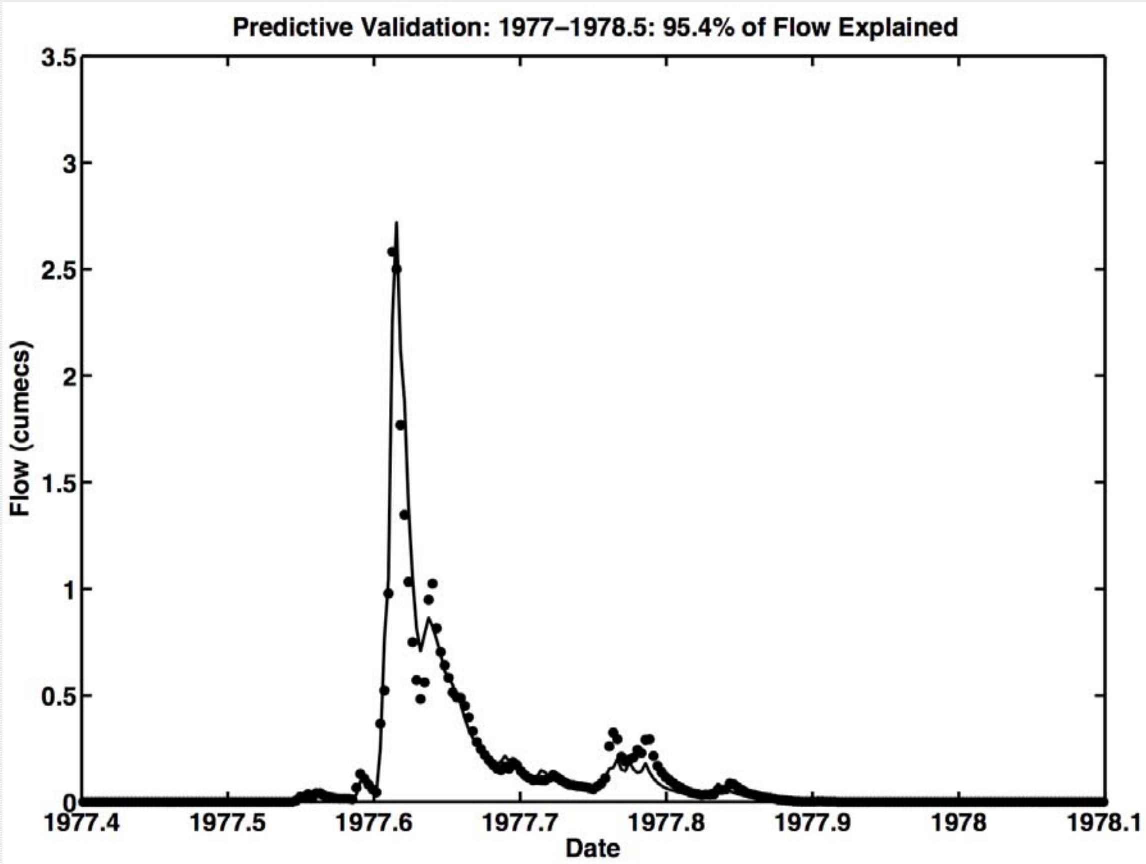
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Parallel Flow Decomposition

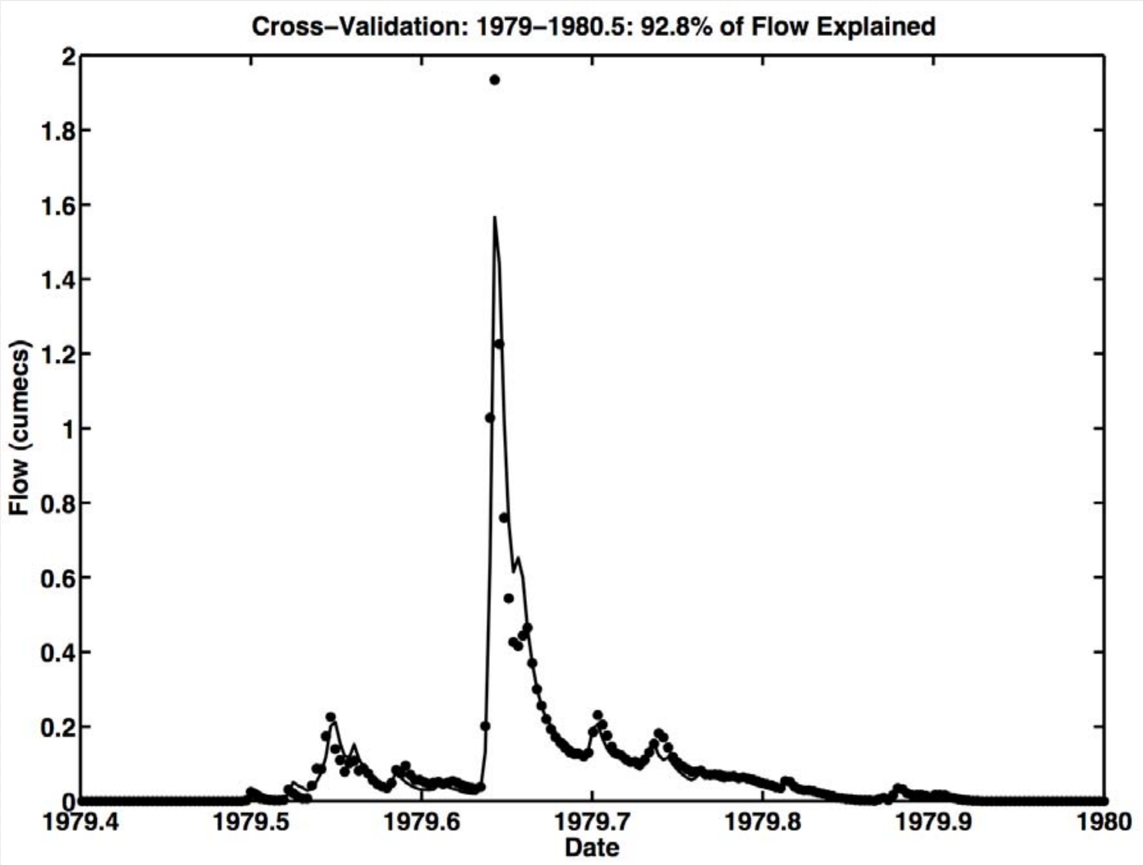


Note: all components plotted to same scale for comparative purposes

Predictive Validation on Data Not Used in Model Estimation (1977-1978)

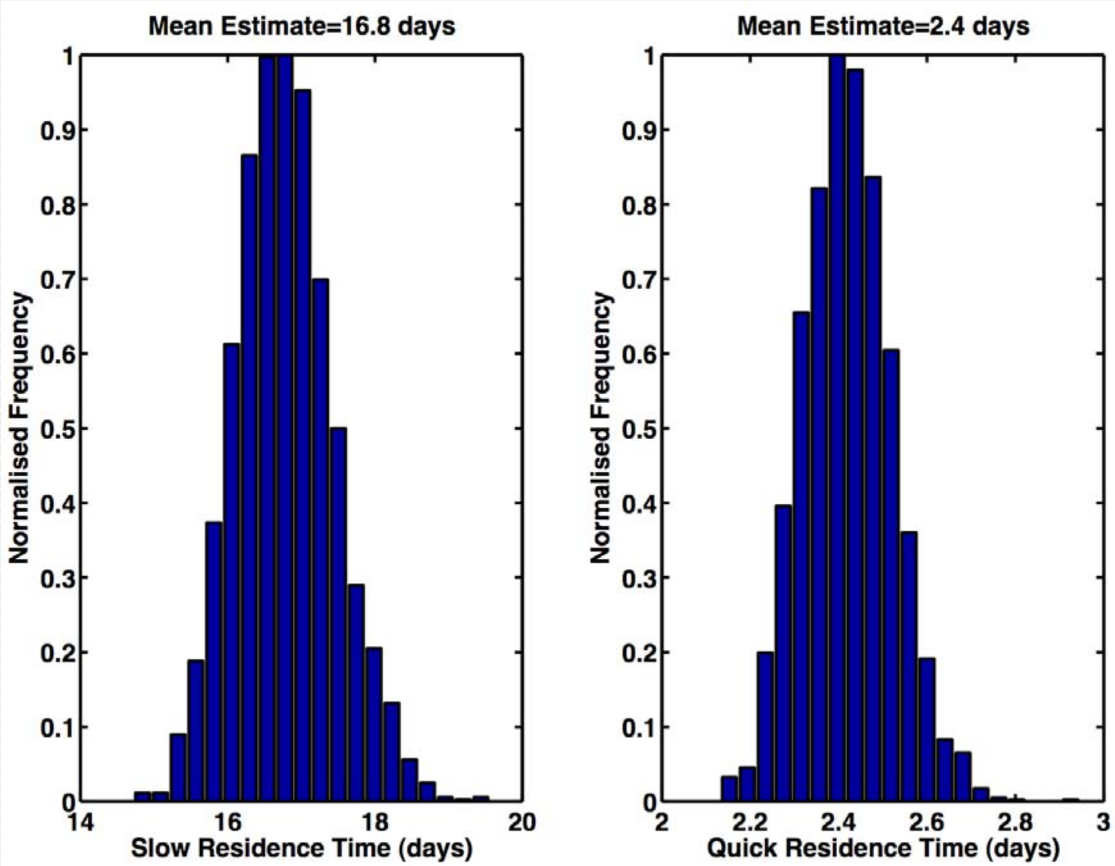


Predictive Validation on Data Not Used in Model Estimation (1979-1980)



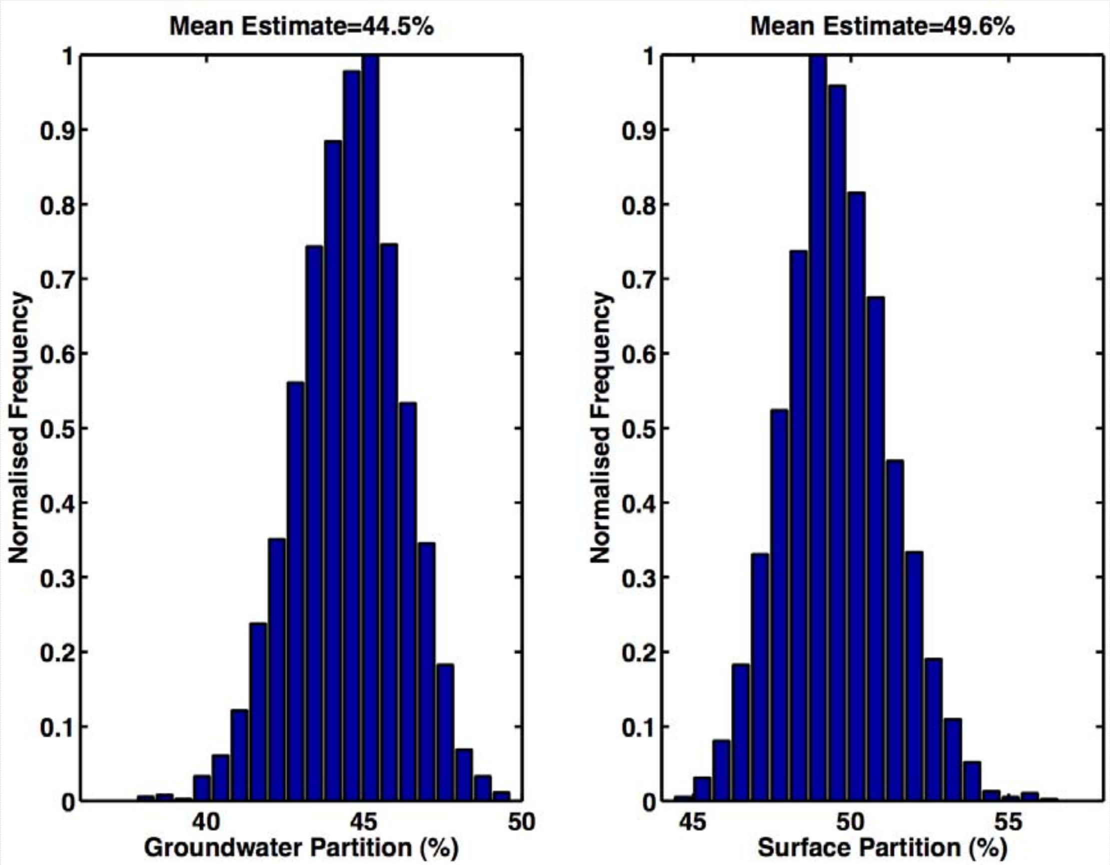
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Estimates of Uncertainty associated with Estimated Residence Time Parameters (MCS analysis)

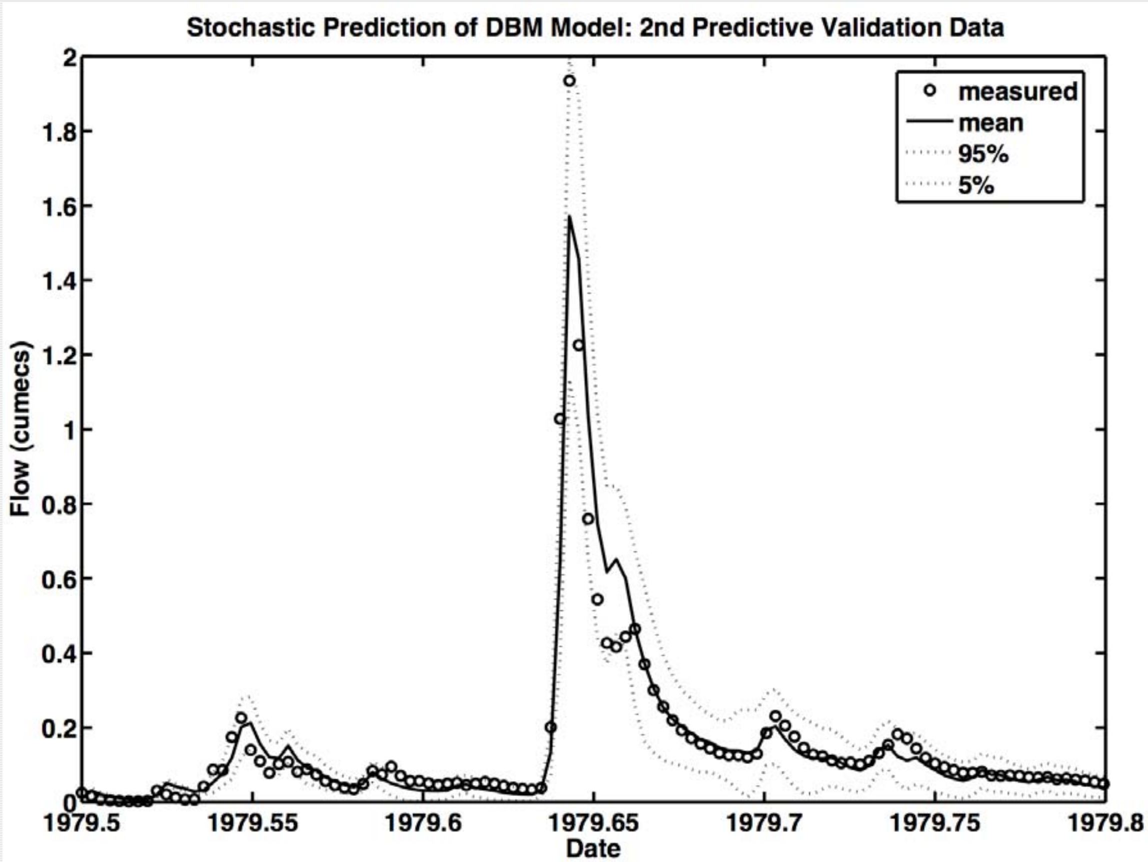


CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Estimates of Uncertainty associated with Estimated Partition Percentage Parameters (MCS analysis)



Stochastic Predictive Validation on 2nd Validation Data Set (1979-1980)



Note: the residual noise is modelled and propagated as a heteroscedastic process based On SDP estimation of the variance as a function of the flow

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Stage 2: Real-Time Forecasting

The best identified DBM model has zero time delay, so forecasting with this model would require one-day-ahead forecasting of the rainfall. To avoid this, the DBM model is re-estimated with a 'virtual' one day time delay added: i.e. a [2 2 1] TF model.

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

DBM Forecasting Model: Identification and Estimation of Model with Added 1 Day Time Delay

The DBM *forecasting* model for the River Canning is estimated as follows:

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

Note the added 'virtual' one day time delay

where ξ_t is an AR(25) noise process. This decomposes to:

$$y_t = \frac{\beta_1}{1 - \alpha_1 z^{-1}} u_{t-1} + \frac{\beta_2}{1 - \alpha_2 z^{-1}} u_{t-1} + \xi_t$$

fast pathway slow pathway

The associated parameter estimates are as follows:

$$\hat{a}_1 = -1.6243(0.018); \quad \hat{a}_2 = 0.6417(0.016); \quad \hat{\gamma} = 0.777(0.096);$$

$$\hat{b}_1 = 0.209(0.004); \quad \hat{b}_2 = -0.1912(0.004);$$

$$\hat{\beta}_1 = 0.171; \quad \hat{\alpha}_1 = 0.661; \quad \hat{\beta}_2 = 0.026; \quad \hat{\alpha}_2 = 0.942;$$

Note: the parameter c is a normalization coefficient selected so that $\sum_{t=1}^{t=N} u_t = \sum_{t=1}^{t=N} y_t$

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

The DBM State Space Model

The DBM forecasting model is again *decomposed into its physically meaningful parallel pathway form and used as the basis for the following State Space (SS) model with heteroscedastic measurement noise*:

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}u_{t-\delta} + \zeta_t$$

$$y_t = \mathbf{h}^T\mathbf{x}_t + e_t \qquad e_t = N(0, \sigma_t^2)$$

where,

Note the time variable error variance

$$\mathbf{F} = \begin{bmatrix} 0.661 & 0 \\ 0 & 0.942 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} 0.171 \\ 0.026 \end{bmatrix} \quad \zeta_k = \begin{bmatrix} \zeta_{1,k} \\ \zeta_{2,k} \end{bmatrix} \quad \mathbf{h}^T = [1 \quad 1]$$

so that it can be embedded within an adaptive Kalman Filter (KF) forecasting engine

Note that, here, for simplicity, the AR(25) noise model is not included in the SS model, so that the forecasting results are 'sub-optimal' in this regard.

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Adaptive Gain Estimation

Whatever models are used for the rainfall-stage and routing processes, they will be uncertain and it is likely that some of the characteristics will change over time. This suggests that an adaptive implementation is required.

We have found that full parameter adaptation is often not necessary: simpler gain adaptation is sufficient in many situations (e.g. Dumfries flood warning system). Here, the estimation algorithm assumes that the gain parameter can be modelled as a random walk process:

$$y_t = g_t \hat{y}_t + \varepsilon_t \qquad g_t = g_{t-1} + \eta_t$$

so that the adaptive estimation algorithm takes the following RLS form (e.g. Young 1984):

$$p_{t|t-1} = p_{t-1} + q_g \quad \text{where } q_g \text{ is the ML optimized NVR hyper-parameter}$$

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

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

Heteroscedasticity and Adaptive Variance Estimation

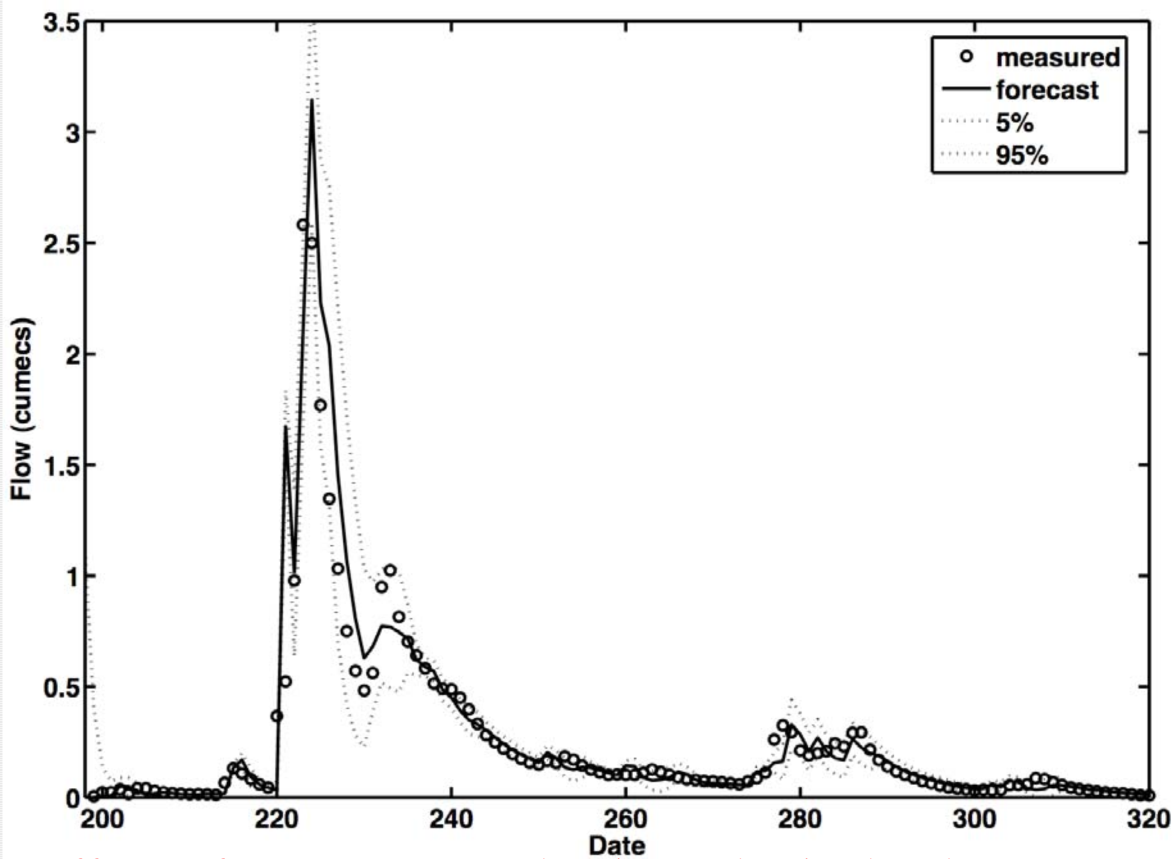
- The noise on the stage (or flow) measurement is normally heteroscedastic , requiring either an SDP implementation (e.g. variance proportional to stage level); an adaptive estimation and update of the variance; or both of these. Here, I have simply updated an estimate of the innovations variance based on a normalization transformation using a similar RLS algorithm to that used for the gain parameter (details in the paper).

- The two simple gain and variance adaptive algorithms are embedded in the normal KF algorithm. All the 'hyper-parameters' in this modified KF algorithm (here simply noise variance ratios) are optimized by ML using prediction error decomposition (Schweppe, 1964).

***Note**: the overall adaptive forecasting algorithm is extremely simple in this case because the SS model is only second order (computation time for the complete results on the next slide only 0.19 secs on Mac 12" PowerBook). It would not be much more than this even if the colour in the noise was taken into account, but we have found that this is often not essential, as here, because of the adaptive modifications.*

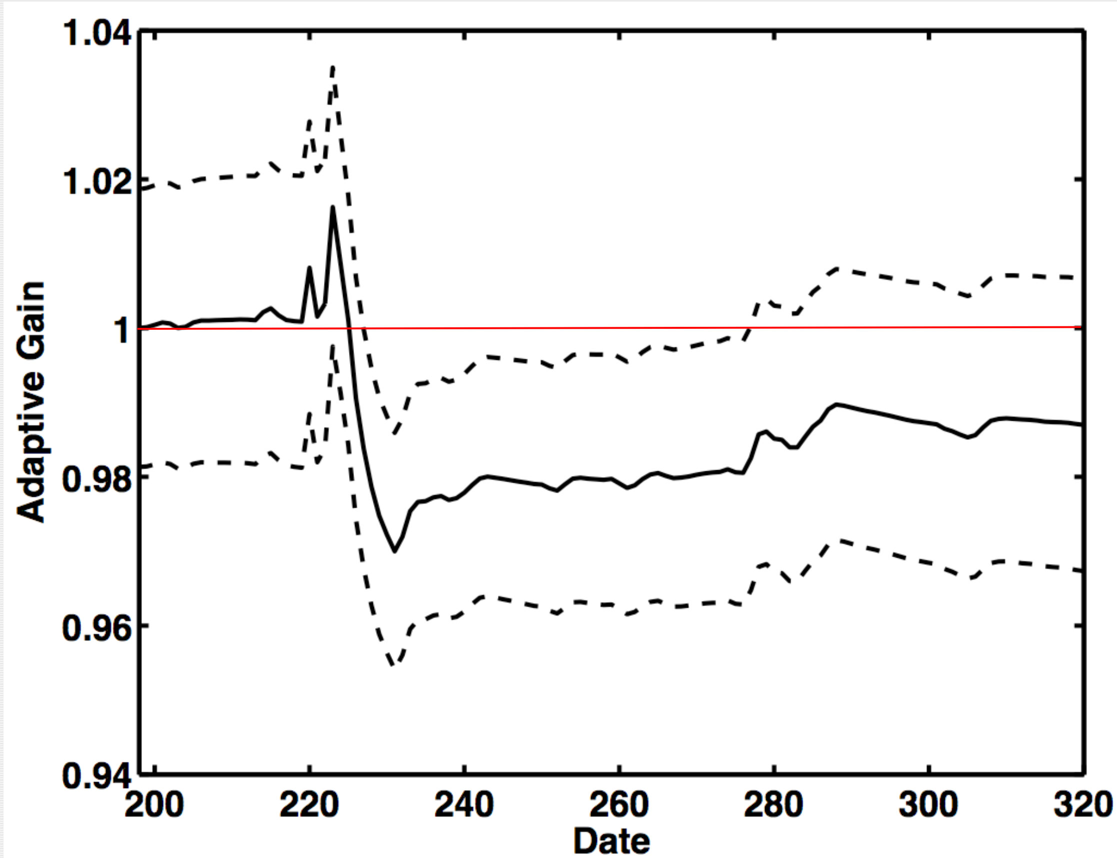
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Adaptive Forecasting Results on 1st Validation Data Set (1977-1978)



Coefficient of Determination based on the one-day ahead prediction errors is $R^2 = 0.92$ compared with $R^2 = 0.84$ for the naïve forecast: i.e
1/2 the forecasting error variance and 1/3 of the peak error compared with naïve.

Adaptive Gain Variation on 1st Validation Data Set (1977-1978)



Note: The non-adaptive gain if 1.0 (in red), so the adaption is not having much effect in this case because the estimated model is so good.

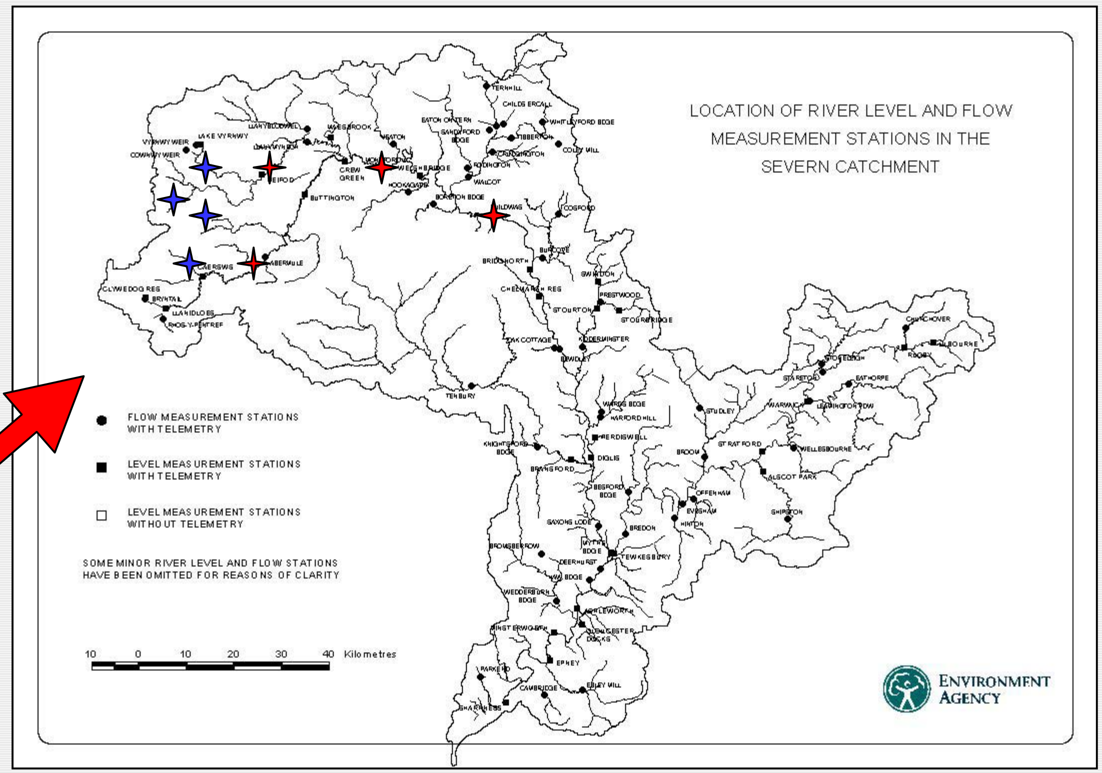
Example 2: DBM Modelling and Adaptive Forecasting, River Severn, UK

These are initial results from Lancaster's contribution to the Flood Risk Management Research Consortium (FRMRC) Project in the UK and the research has been carried out in conjunction with Renata Romanowicz, who generated most of the results shown here, and Keith Beven.

Although the climate of the UK is certainly not 'arid', the methodology is based on the modelling approach used in the Canning example and so are almost certainly applicable to arid catchments.

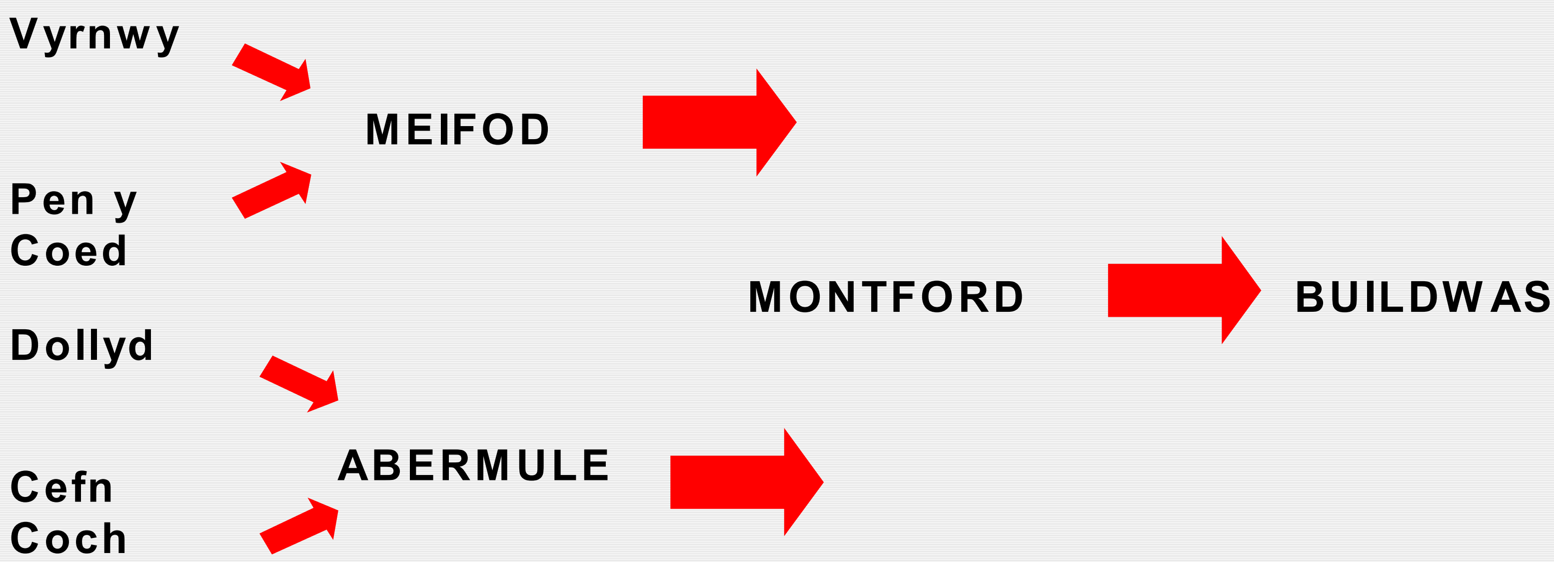
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

The River Severn System, UK



CRESearch
The Key to Success in Environmental and Biological Systems Analysis

River Severn System Diagram



Schematic presentation of the on-line forecasting system for the River Severn upstream of Buildwas

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Stochastic Transfer Function Models
2. Rainfall-Stage and Stage Routing Example

The complete DBM model for the River Severn consists of several TF modules some for rainfall-stage modelling where the TF model is typically of the form:

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

where y_t is the stage level; u_t is the effective rainfall; and r_t is the gauged rainfall. These are, once again, decomposed into the physically meaningful parallel pathway form.

$$y_t = \frac{\alpha_1 z^{-1}}{1 - \beta_1 z^{-1}} u_t + \frac{\alpha_2 z^{-1}}{1 - \beta_2 z^{-1}} u_t + \xi_t$$

In the case of the stage-stage routing modules, the model is identified in a 'cascade' (similar to the Nash cascade) of 1st order TFs, each of the form:

$$y_{i,t} = \frac{b_0}{1 + a_1 z^{-1}} y_{i,t-\delta i} + \xi_{i,t} \qquad i = 1, 2, \ldots, m$$

where $y_{i,t-\delta i}$ is the upstream stage and $y_{i,t}$ is the downstream stage.

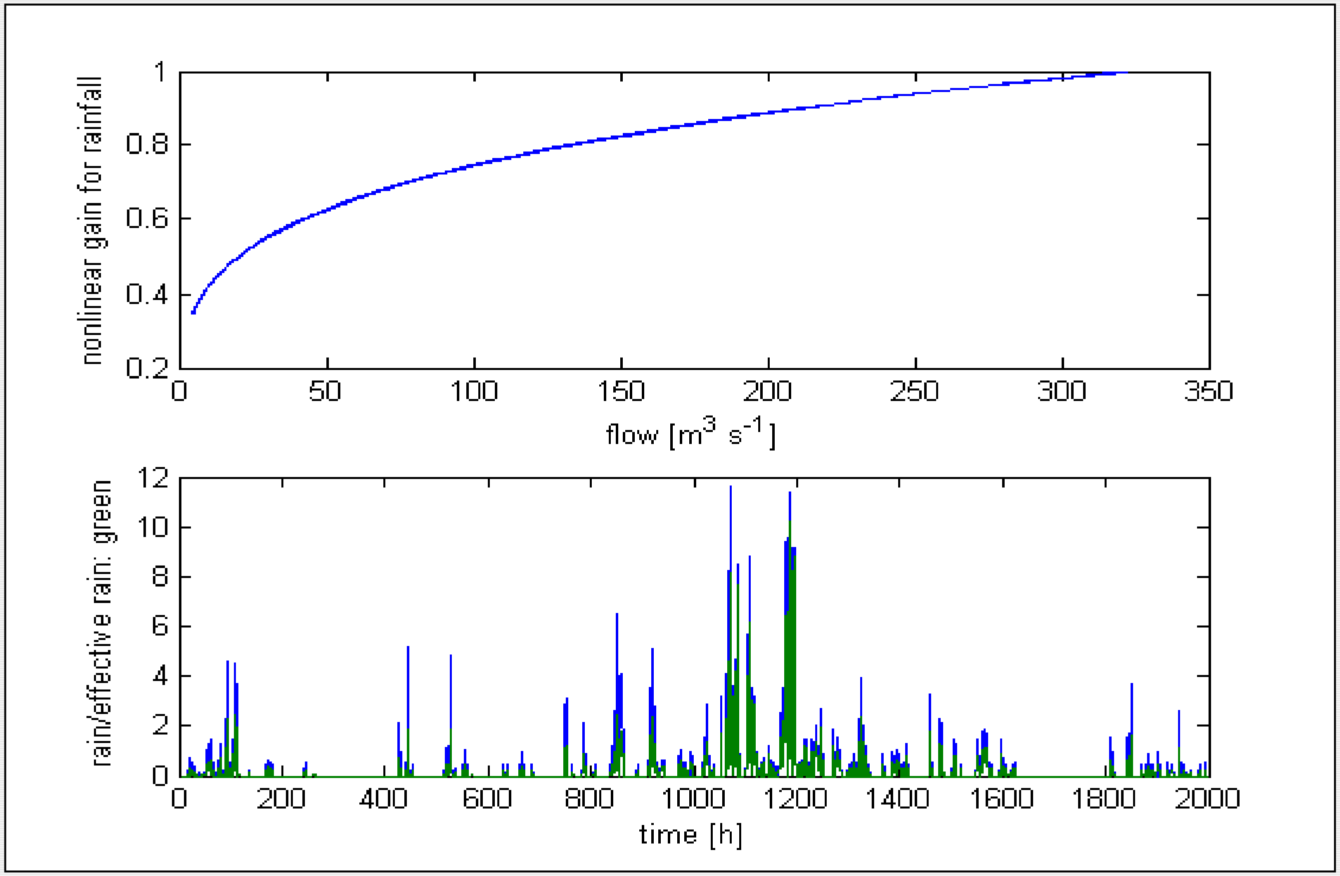
CRESearch
The Key to Success in Environmental and Biological Systems Analysis

Real-Time Forecasting: Initial Results

The FRMRC Consortium Project started last year and has only been proceeding for several moths so far. These are the initial results we have obtained recently after project definition and data acquisition.

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

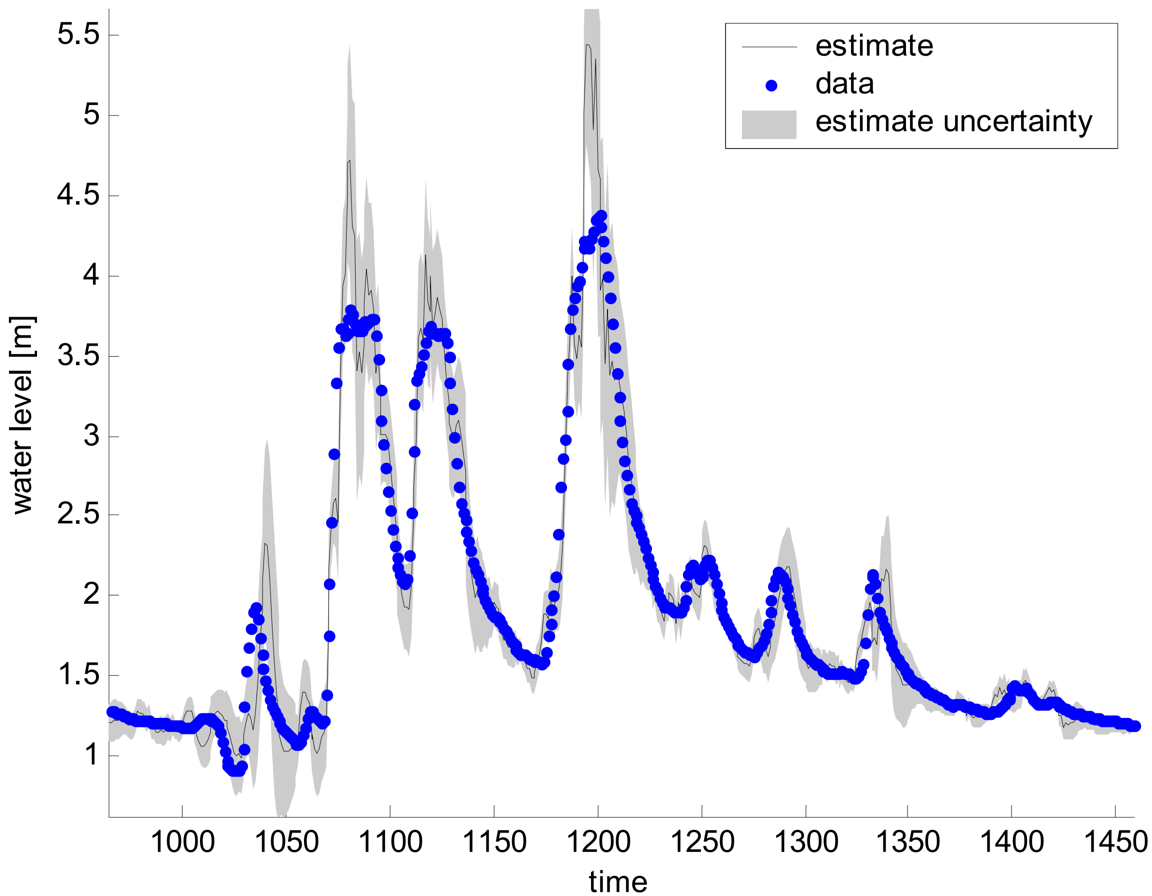
Nonlinear DBM rainfall-stage model, Abermule



Nonlinear gain for the rainfall (upper panel); lower panel shows rainfall (blue line) plotted against effective rainfall (green line).

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

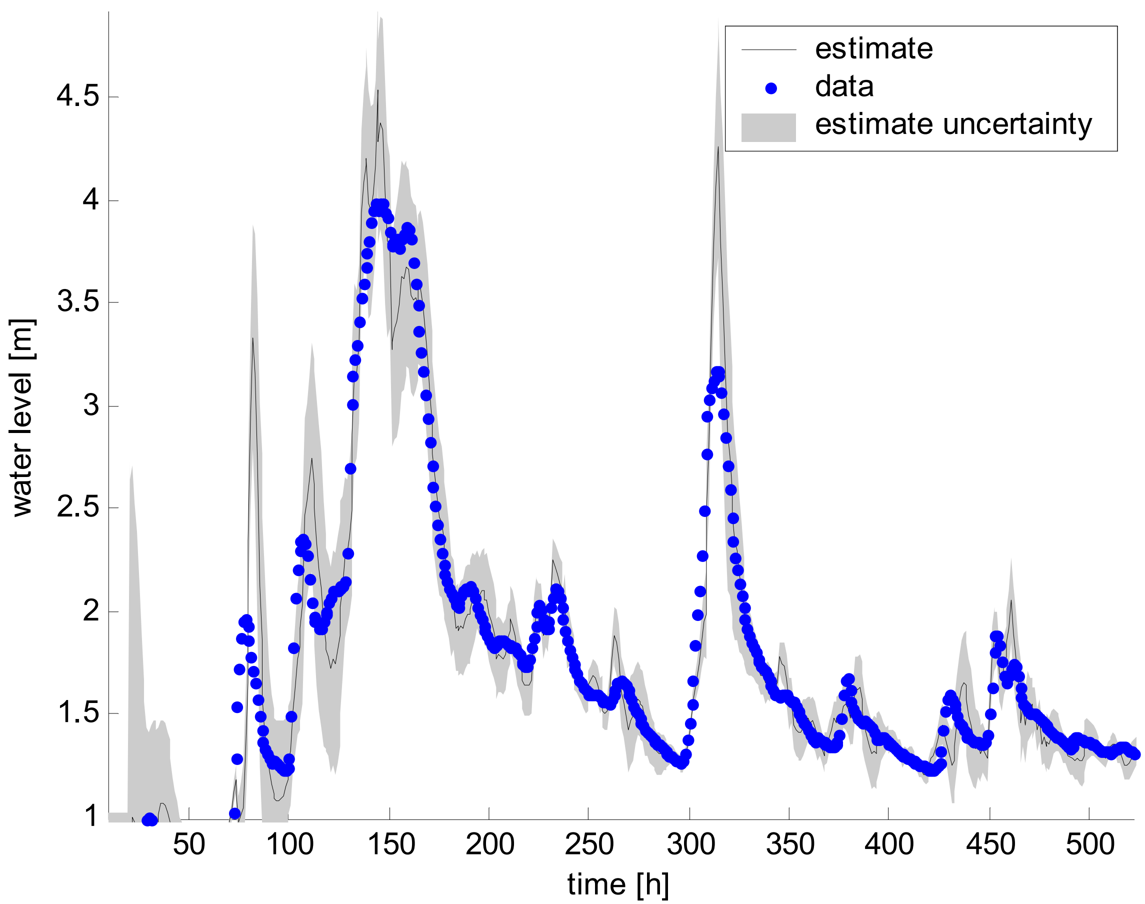
Rainfall-stage adaptive model for Abermule: calibration; 28th Oct.1998



5 hour ahead forecasts:

2 rain gauge inputs

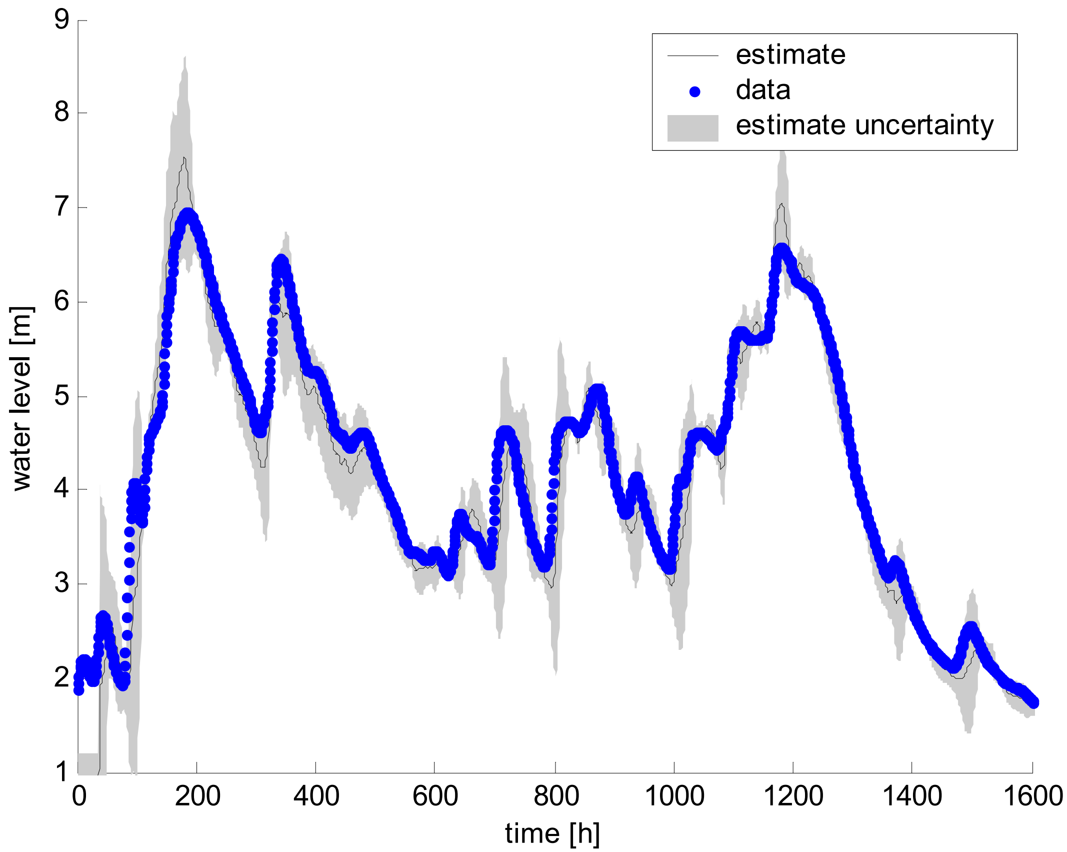
Rainfall-Stage Model for Abermule
Validation: 30th October 2000; R_T^2 = 92%



5 hour ahead forecasts:

2 rain gauge inputs

Rainfall-Stage Model for Abermule-Meifod-Montford Validation: 1st November; R_t^2 = 92%



21 hour ahead forecasts:

2 rain gauge inputs + routing

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

CONCLUSIONS

1. Recursive statistical estimation provides the most obvious algorithm for real-time parameter and/or state updating, data assimilation and adaptive forecasting.
2. Model identifiability is particularly important in real-time DBM modelling and forecasting; and model complexity has an important bearing on identifiability.
3. The dangers of model over-parameterization in complex models are manifold and need to be acknowledged in real-time forecasting.
4. DBM modelling and real-time forecasting in the case of the Canning River demonstrates the applicability of this approach in arid zone hydrology.
5. The application of DBM modelling and recursive parameter/state adaptive methods to River Severn stage forecasting has proven that these methods are capable of providing acceptable long-term forecasts, with reasonable accuracy, in a quite complicated river catchment.
6. DBM modelling and real-time forecasting exploits the CAPTAIN Toolbox for Matlab: see http://www.es.lancs.ac.uk/cres/captain/

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

A Few References

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

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

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

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

Young, P.C. (2002) Advances in real-time forecasting, *Phil. Trans. Royal Soc., Physical and Engineering Sciences,* **360**,1433-1450.

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