A Streamflow Forecasting Algorithm And Results For The Upper Murray Basin

P.C.Young[1], S.Yu.Schreider[2] and A.J. Jakeman[2,3]
[1]Control and Systems Group, CRES/IEBS, Lancaster University, Lancaster, LA1 4YQ, UK
[2]Centre for Resource and Environmental Studies, The Australian National University, Canberra, ACT 0200, Australia
[3]Adjunct Professor, Department of Economics, University of Western Australia, Nedlands, WA 6009

Abstract In Schreider, *et al.* (1997a), the IHACRES rainfall-runoff model is calibrated for the purpose of predicting streamflow discharge in ten catchments of the Upper Murray Basin, using a four hourly time step. A map of the basin and list of the catchments considered are presented in Figure 1 and Table 1 in Schreider, *et al.* (1997a). The major aim of the present paper is to describe the subsequent development and testing of a four hourly time step, flow forecasting model which exploits the Kalman Filter (KF) algorithm to upgrade the IHACRES models from a simple predictive to a real-time forecasting capability. In Schreider *et al.* (1997b), the IHACRES model and a self adaptive filtering approach, based on the Auto-Regressive Integrated Moving Average (ARIMA) representation of the model residuals, were combined and utilised for forecasting daily streamflow in nine catchments of the Upper Murray Basin. Such linear filtering of the model residuals provided a considerable improvement in forecasting both low and high values of streamflow. A KF forecasting algorithm, incorporating the sub-daily Upper Murray Basin IHACRES model, has been used in this second stage of the project as a tool for operational streamflow forecasting because it provides a more flexible approach and yields even better results (in terms of Nash-Sutcliffe efficiency statistics and relative errors) than the ARIMA linear filtering approach.

1. INTRODUCTION

In the almost forty years since Kalman (1960) first derived his famous filtering algorithm, it has been applied in many different areas of science and social science and there are myriads of publications devoted to its application in the field of hydrology. Thus, a comprehensive review of such work is beyond the scope of this paper. The Kalman Filter (KF) algorithm used here is formulated in the predictive-corrective form (e.g. Young 1984 and the earlier references therein) which underlies the general *Unobserved Component* approach to state space estimation and forecasting (Young, 1988, 1989, 1991; Young *et al.* 1998). The well known monograph by Bras and Rodriguez-Iturbe (1985) contains an overview of the KF and other statistical techniques in hydrology (see also the papers of Young and Wallis, 1985; and Young, 1986); while an inventory of recent hydrological applications of the KF, especially for streamflow forecasting, can be found in Sen (1991), where the KF method is also applied for the prediction of monthly flow for two catchments in Turkey and the USA; and for monthly rainfall prediction in Saudi Arabia.

1.1 Prediction Versus Forecasting

The KF methodology applied in the current work is based on the separation of two steps: simulation, or simple prediction; and multi-step ahead forecasting. Here, the term 'prediction' is used for modelling exercises which follow the establishment of relationships between the climatic input and the hydrological/ geomorphological parameters, measured during some period, and the streamflow output for the same period of time. This then allows either for the generation of streamflow during the same period (fitting), or over another period using the relevant climatic inputs and model parameters (simulation). The term 'forecasting' is employed to describe algorithms which allow for the real-time, multi-step-ahead forecasting of streamflow values into the future, with the estimates of streamflow and other associated 'state-variables' updated at regular intervals. This methodology, in various forms, is widely used in hydrology for operational streamflow forecasting. In 'adaptive forecasting', the model parameters are also updated recursively to ensure that the underlying model is calibrated on the latest data (see e.g. Lees *et al.*, 1994). Such statistical forecasting algorithms are normally based, as in this paper, on different forms of the KF; or on modifications of the AutoRegressive Integrated Moving Average (ARIMA) approach of Box and Jenkins (1970; see also Bras and Rodriguez-Iturbe, 1985). Some publications on previous uses of the KF are discussed in the next sub-section.

1.2 Previous KF Based Algorithms

One form of the KF technique is used in the European Flood Forecasting Operational Real-Time System (EFFORTS), widely applied for water resource management in Europe and elsewhere (Todini, 1996). Here, the predictive part of the algorithm is based on the conceptual rainfall-runoff model ARNO. The forecasting part is based on two linear, interactive KF's, one in state vector space and another in the parameter space. Dimopoulos *et al.* (1996) have developed a streamflow forecasting method based on the combination of a neural network model and KF techniques. The neural network algorithm is used as a simple predictive (or simulation) part of the algorithm, taking into account the non-linearities of the relationship between input rainfall and the output runoff; whereas the KF is applied for real-time correction of the predictive model residuals. The algorithm has been applied to two catchments in France using weekly and daily time steps, yielding Nash-Sutcliffe efficiency of about 0.800 in both cases. Bidwell and Griffiths (1994) describe an algorithm somewhat similar to the one described in the present paper: for the predictive step, they use a first order transfer function (TF) model with a time step of one hour and adaptive coefficients, in order to establish a forecasting relationship between modelled streamflow output and antecedent values of the modelled streamflow and measured input. The precipitation and measured upstream flow time series are used as model inputs and the TF model parameters are estimated initially using the recursive IV algorithm (see e.g. Young, 1984). The KF algorithm (effectively time variable parameter, recursive least squares) is then used for updating the TF parameters in order to provide real-time, 4 hour ahead forecasting on the Waimakiriri River in New Zealand. The algorithm is tested for four flood events and yields relatively low errors.

In the present paper, the predictive-corrective version of the KF algorithm is used for real time forecasting, based on the IHACRES model. Here, the predictive step utilises the linear module of IHACRES, with the effective rainfall input obtained from the nonlinear module; and the corrective step adjusts the streamflow estimate by reference to the 'innovation' error between the prediction and the streamflow measurements. An important advantage of the KF algorithm used in this manner is that any missing measurements of streamflow do not interrupt the algorithmic implementation, since inherent model-based interpolations within the KF allow for continuation of the forecasting process without interruption.

2. DESCRIPTION OF THE KALMAN FILTER FORECASTING ALGORITHM

Since the two-reservoir structure of the IHACRES linear module (Jakeman *et al.*, 1990; Jakeman and Hornberger, 1993) involves parallel processes (see also Young, 1992), it can be formulated in the following state-space form:

$$x(k) = A\, x(k-1) + b\, u(k-d) + \xi(k) \tag{1}$$

$$y(k) = C\, x(k) + e(k) \tag{2}$$

where $\mathbf{u}(k)$ is effective rainfall input; $\mathbf{x}(k) = (x^q, x^s)^T$ is the state vector representing the outputs of the 'quick' and 'slow' reservoirs; y(k) is the measured streamflow (the output measurement); $\mathbf{C}$ is the 2x1 output or observation row vector $\mathbf{C}=[1\ \ 1]$; $\xi(k)$ is a white noise vector with covariance matrix $\mathbf{Q}$; and e(k) is white measurement noise, with variance σ^2. Finally, the IHACRES model parameters a_q, a_s, b_q, b_s and d, define the 2x2 diagonal state transition matrix $\mathbf{A}$, with the parameters a_q and a_s on its diagonal; the vector $\mathbf{b}=[b_q, b_s]^T$;and $\mathbf{d}$ is the integer time delay.

The particular structure of the KF algorithm used in the present work is a recursive procedure consisting of two steps per recursion: *prediction* and *correction*. The prediction step utilises the above state-space representation (1) of the IHACRES model to predict the state variables for one sample ahead; and the correction step compares this prediction with the measured streamflow at the next sample and utilises this 'innovation error' to update the previous state estimate.

The *prediction* step is as follows:

$$\mathbf{xh}(k|k-1) = \mathbf{A}\, \mathbf{xh}(k) + \mathbf{b}\, u(k-d) \tag{3}$$

$$\mathbf{P}(k|k-1) = \mathbf{A}^T\, \mathbf{P}(k-1)\, \mathbf{A} + \mathbf{Q}/\sigma^2 \tag{4}$$

$$yh(k) = \mathbf{C}\, \mathbf{xh}(k|k-1) \tag{5}$$

The *correction* step is then:

$$G(k) = (1 + \mathbf{C}\, \mathbf{P}(k|k-1)\, \mathbf{C}^T)^{-1} \tag{6}$$

$$\mathbf{xh}(k) = \mathbf{xh}(k|k-1) + [\, \mathbf{P}(k|k-1)\, \mathbf{C}^T\, G(k)]\, \{y(k) - yh(k)\} \tag{7}$$

$$\mathbf{P}(k) = \mathbf{P}(k|k-1) - \mathbf{P}(k|k-1)\, \mathbf{C}^T\, G(k)\, \mathbf{C}\, \mathbf{P}(k|k-1) \tag{8}$$

In the above, $\mathbf{xh}(k)$ is the state vector; $\{y(k) - yh(k)\}$ is the innovation error; $\mathbf{P}(k)$ is the normalised covariance matrix, $cov(\mathbf{xh}(k)-\mathbf{x}(k))/\sigma^2$; and the two diagonal elements of $\mathbf{Q}/\sigma^2$, the *Noise Variance Ratio* (NVR) matrix, are the unknown 'hyper-parameters' to be optimised by manual tuning or numerical optimisation (see e.g. Young, 1988, 1994, Young *et al.*, 1989, 1991, and the references therein). If an *n* time step ahead forecast is required, then the correction step is simply omitted *n* times.

3. COMPARATIVE RESULTS OF THE ARIMA AND KF FORECASTING FOR A DAILY TIME STEP

Comparative analysis of the streamflow forecasting using the KF and ARIMA (see earlier) methods was carried out for nine catchments of the Upper Murray Basin modelled using a daily time step. (These nine catchments are the same as those used for the 4-hourly modelling except Corryong Creek is missing). The efficiency statistics estimated for the whole period when data are available are summarised in Table 1, which illustrates that the results produced by the KF approach are consistently better than the ones provided by the ARIMA method.

Table 1: Efficiency statistics (E) for 9 catchments of the Upper Murray Basin. Results are for IHACRES alone compared with IHACRES combined with an ARIMA algorithm, and IHACRES with KF.

Station number	River and station location	IHACRES model applied solely E	IHACRES model combined with ARIMA (1,0,0) E	Kalman filter estimate incorporating IHACRES E
401203	Mitta-Mitta River at Hinnomunjie	0.669	0.835	0.850
401220	Tallangatta Creek at McCallums	0.611	0.807	0.896
401229	Cudgewa Creek at Berringama	0.638	0.744	0.782
401012	Murray River at Biggara	0.649	0.835	0.911
401217	Gibbo River at Gibbo	0.692	0.805	0.917
401210	Snowy Creek at Granite Flat	0.729	0.809	0.941
401013	Jingellic Creek at Jingellic	0.533	0.602	0.657
401014	Tooma River at Pine Grove	0.664	0.841	0.866
401216	Big River U/S of Joker Ck	0.709	0.845	0.891

4. RESULTS OF KF FORECASTING FOR THE 4-HOURLY TIME STEP

A simulation test (sometimes termed validation) for the IHACRES model was carried for all ten catchments considered. Here, the values of all six parameters of the IHACRES model, optimised during the calibration runs, were used for modelling the streamflow, based on inputs of the associated rainfall and temperature series. The results of this simulation test for the IHACRES model using a 4-hour time step, as shown in Table 2, reveal that the performance of the model, taken alone, is much worse than that of the model calibrated on the daily time step (Schreider *et al.*, 1997), where the efficiency statistic is consistently higher than 0.600. On the other hand, when this 4-hour time step model is integrated with the KF forecasting algorithm, as described above, it yields higher

Table 2: Efficiency E and absolute relative errors ARE for IHACRES alone compared with the KF estimate and 3-day ahead forecast for 10 catchments of the Upper Murray Basin.

Station number	Time delay	E and ARE for IHACRES	E and ARE for KF	E and ARE for 3 day ahead forecast
401203	2	0.112 75%	0.936 4.5%	0.684 11.5%
401230	2	-0.060 80%	0.964 5.3%	0.846 15%
401220	1	0.776 62%	0.934 9%	0.870 20%
401229	2	-1.350 87%	0.857 9.5%	0.563 21%
401012	2	0.260 55%	0.967 4%	0.856 11%
401217	1	0.256 62%	0.981 3.6%	0.908 9%
401210	2	0.435 61%	0.932 6%	0.813 12%
401013	1	0.333 100%	0.852 10%	0.633 25%
401014	2	0.843 130%	0.916 11%	0.761 19%
401216	0	0.410 33%	0.839 5.5%	0.690 12%

values of the efficiency E combined with lower relative errors.

It should be noted here that the Snowy Creek catchment was modelled with and without the snow melt/accumulation module (see Schreider, *et al.*, 1997a). The KF forecasting algorithm was applied for both cases and it yields an efficiency slightly higher for the case when the snow melt/accumulation module was applied (E = 0.941) than without this module (0.932). However, since the relative errors were 14% and 6%, respectively, for these two approaches, the application of the snow melt/accumulation module provides little advantage for forecasting high flow events (the Nash-Sutcliffe efficiency is more sensitive to high absolute errors) and it is a disadvantage for forecasting the medium and low events. The results for the 3 time step (12 hours) ahead forecast yield efficiencies are 0.774 and 0.813, and relative errors are 38% and 12% for the modelling with and without snow melt/accumulation module, respectively. Although all these results suggests that the snow melt/accumulation module should not be included for the purposes of operational streamflow forecasting, it is too soon to reach any firm conclusions in this regard.

Figures 1 and 2 present the results of the KF algorithm applied to the Tooma River and Corryong Creek catchments. They were selected as examples of catchments with very different properties: the Tooma River has a large (1,819 km^2) snow-affected catchment; and Corryong Creek has relatively small (387 km^2) snow-free catchment. Figures 3 and 4 show the results from the KF algorithm for the Upper Murray and Gibbo catchments with higher temporal resolution (zoom display over a shorter period), for the cases of low and high flow events. Note that figure 4a demonstrates how the KF algorithm works during a period when no precipitation is recorded.

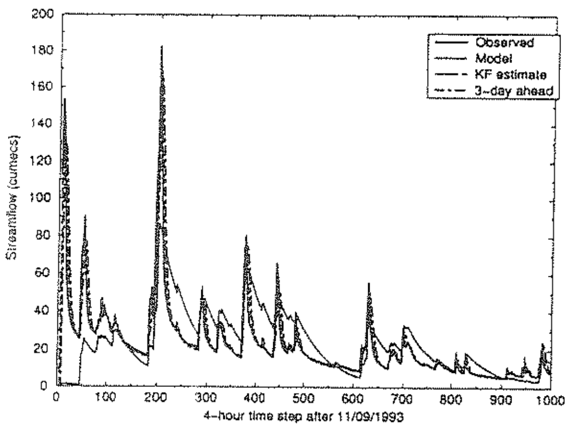

Figure 1: Results of Kalman filter forecasting for the Tooma River catchment.

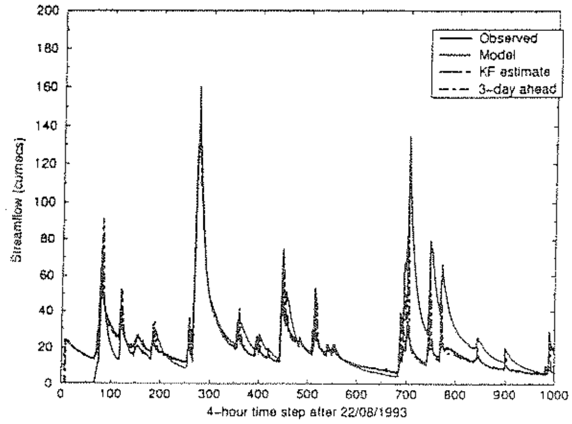

Figure 2: Results of Kalman filtering forecasting for the Corryong Creek catchment.

5. DISCUSSION

Previously, the conceptual rainfall-runoff model IHACRES has been successfully calibrated for ten catchments in the Upper Murray Basin using a 4-hourly time step for input and output time series (Schreider *et al.*, 1997a). In the present study, a KF based forecasting algorithm has been developed for streamflow forecasting using this calibrated IHACRES model as the underlying prediction technique. This algorithm yields high values of the efficiency statistics (from 0.839 to 0.967) and low relative errors (from 4% to 11%) over the whole period of observation for each of the ten catchments considered. The algorithm has also been implemented for the case of 12 hour ahead forecasts and it produces consistently high efficiency values. Finally, when compared with the ARIMA-based forecasting algorithm, using a daily sampling interval, the KF technique is consistently superior for nine of the catchments in the Upper Murray Basin.

An important advantage of the KF algorithm proposed here is that missing inputs of streamflow do not interrupt the data processing. Another advantage is that information about the climatic data (temperature and precipitation) is required at only 8-12 hours prior to the forecast time. This is achieved by calibrating the IHACRES model for almost all catchments with a convenient 4-8 hours time delay (Table 2), which follows the approach used by Lees *et al* (1994). The Big River catchment is an exception, but this is not crucially important because it is a tributary of the Mitta-Mitta River and is not used, therefore, for practical forecasting of flow into the Dartmouth Lake inlet.

Another, related advantage of the KF algorithm applied on the 4-hourly time step is that, if the model is calibrated with a time delay d of 1 time step, then the KF provides the one-step-ahead prediction directly, without the need for multi-step forecasting (See equations 3-5). All of the Upper Murray Basin catchments are calibrated with such a delay (except the Big River catchment, which is a tributary of the Mitta-Mitta River upstream of its gauging station).
Finally, the well known robustness of the KF algorithm allows one to use the IHACRES predictive model parameters estimated using 'regionalisation' principles when streamflow time series are not available for model calibration. Regionalisation assumes that the model parameters are a function of the landscape and vegetation characteristics of a catchment. For the Mitta-Mitta and Tooma Rivers, the IHACRES parameters were taken from the calibration results obtained for the Big River catchments. Although the KF estimate for the Tooma River is the worst in the Basin, in terms of relative error, the estimate obtained for the Mitta-Mitta catchment is comparable with other catchments. This is explained by the higher geomorphologic similarity between the Big River and Mitta-Mitta catchments (the Big River is a tributary of Mitta-Mitta), than that between the Big and Tooma catchments.

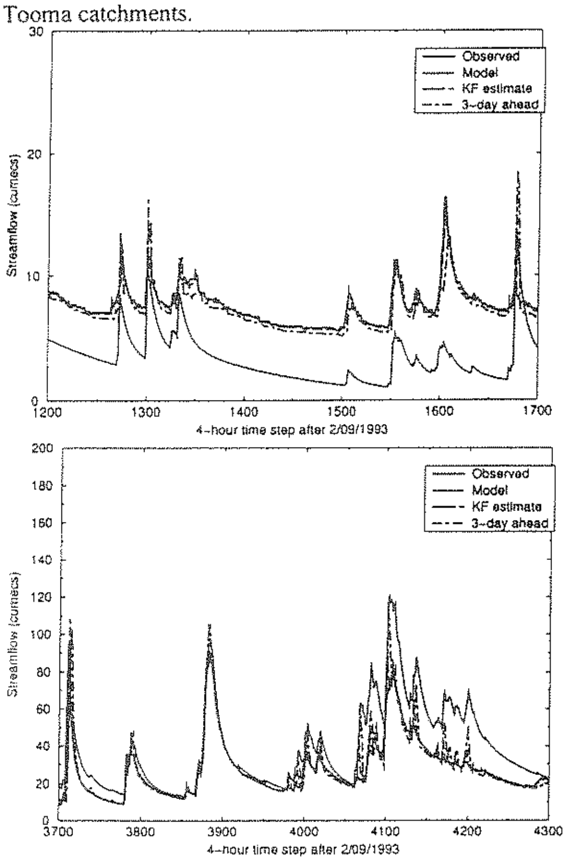

Figure 3: Results of Kalman filter forecasting for the Upper Murray River catchment: (a) low and (b) high flow periods.

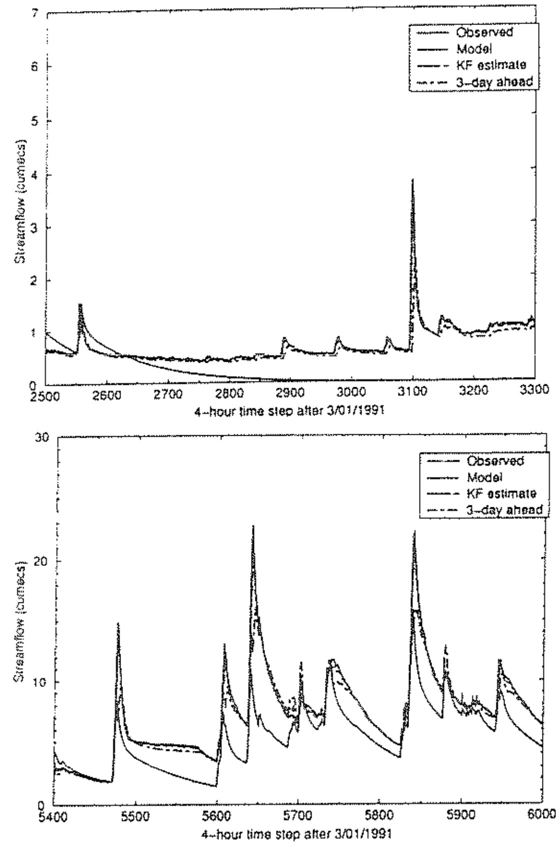

Figure 4: Results of Kalman filter forecasting for the Gibbo River catchment: (a) low and (b) high flow periods.

6. ALGORITHMIC IMPROVEMENTS

The proposed KF-based forecasting algorithm for the Upper Murray Basin can be improved in various ways. Firstly, the rainfall data used at present are recorded in the lowest parts for the catchments, at the site of stream gauging stations. At the same time, the climatic conditions in the region under study are very heterogeneous: mean annual precipitation can double from the lowest to the highest elevations in some catchments, especially in large ones like those of the Tooma, Mitta-Mitta and Upper Murray rivers. Therefore, for many peak flow events, the corresponding rainfall is not recorded at the gauging site. Another similar problem is that maximum rainfall intensity may be located in different parts of the catchments, so that the time delay between a rainfall event and the streamflow response can vary in large catchments from 0 to 20 hours, so that the time delay d presented in Table 2 reflects just an average value for this characteristic. Use of a relevant spatial interpolation procedure and data for precipitation over the whole catchment (or weather radar) is a possible solution to this problem, but this has not been considered during the present stage of this work because of the lack of suitable precipitation measurements.

Secondly, the basic assumption of the KF algorithm is that the streamflow data used in the correction step are free of systematic error; or, in other words, the stochastic inputs e(k) and $\xi(k)$ are composed of zero mean, serially uncorrelated sequences of random variables (white noise). In actuality, however, these inputs are significantly coloured: for instance, e(k) can have a periodic component induced by the sensitivity of the gauging instruments to the daily fluctuations of temperature. More efficient (lower variance) estimation and forecasting will clearly result, therefore, if additional stochastic states are introduced into the model to account for these coloured noise input effects.

Thirdly, the IHACRES model itself has a systematic bias in these catchments, especially at the 4-hourly temporal resolution. The poor convergence obtained in the simulation tests, as reflected in the low values of the efficiency statistics (Table 2), illustrates the difficulties of applying IHACRES alone on independent data sets for forecasting purposes. It means that the measurement noise e(k) does not have a zero mean value when the model is applied to such independent data outside the calibration period.

Further improvement in the forecasting performance of the proposed KF algorithm, including resolution of the bias problem discussed above, will be possible if the algorithm is made self adaptive, either by recursively updating some or all the parameters of the IHACRES model, or by introducing a single adaptive gain parameter, as suggested in Lees *et al.* (1994). If sufficient, this latter approach is likely to provide a more robust and practical real-time solution: for example, it has been used successfully over the last four years as part of the Solway River Purification Board's flood warning system for the town of Dumfries, on the River Nith in Scotland.

7. REFERENCES

Bidwell, V.J. and Griffiths, Adaptive Flood Forecasting: An Application to the Waimakariri River, *Journal of Hydrology (N.Z.)*, 32(2), pp. 1-15, 1994.

Box, G.E.P. and Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, San Francisco, Holden Day, 553 pp., 1976.

Bras, R.L. and Rodriguez-Iturbe, I., *Random Functions and Hydrology*, Addison-Wesley: New York, 559 pp., 1985.

Dimopoulos, I., Lek, S. and Lauga, J. Modelisation de la Relation Pluie-Debit par le Reseaux Connexionnistes et le Filtre de Kalman, *Hydrological Sciences*, 41(2), pp. 179-193, 1996.

Jakeman, A.J., Littlewood, I.G and Whitehead, P.G., Computation of the Instantaneous Unit Hydrograph and Identifiable Component Flows with Application to Two Small Upland Catchments, *Journal of Hydrology*, 117, pp. 275-300, 1990.

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

Kalman, R.E. A New Approach to Linear Filtering and Prediction Problem, *Journal of Basic Engineering*, 82D, pp. 35-45, 1960.

Lees, M.J. , Young, P.C., Ferguson, S., Beven, K.J. and Burns, J. An Adaptive Flood Warning System for the River Nith at Dumfries, in W.R. White and J. Watts (eds.) *River Flood Hydraulics*, Chichester, Wiley, pp. 65-76, 1994.

Schreider, S.Yu., Jakeman, A.J., Dyer, B.G. and Francis, R.I. A Combined Deterministic and Self-Adaptive Stochastic Algorithm for Streamflow Forecasting with Application to Catchments of the Upper Murray Basin, Australia, 1997, *Environmental Modelling Software*, Vol. 12, pp. 93-104, 1997b.

Schreider, S.Yu., Jakeman, A.J. and Young, P.C., Precipitation - Runoff Model Calibration on a Sub-daily Time Step in the Upper Murray Basin, *Proceeding International Congress on Modelling and Simulation*, MODSIM97, A.D. McDonald *et al.* (eds), University of Tasmania, December 8-11, 1997.

Sen, Z., Adaptive Prediction of Hydrologic Series by Walsh-Kalman Model, *Journal of Hydrology*, 122, pp. 221-234, 1991a.

Todini, E., The ARNO Rainfall-Runoff Model, *Journal of Hydrology*, 175, pp. 339-382, 1991.

Young. P.C., *Recursive Estimation and Time-Series Analysis*, Berlin, Springer-Verlag, 300 pp., 1984.

Young, P.C., The Instrumental Variable method: A Practical Approach to Identification and Parameter Estimation, in H.A. Barker and P.C. Young (eds.), *Identification and System Parameter Estimation*, Oxford, Pergamon, pp. 1-16, 1985.

Young, P.C., Time-Series Methods and Recursive Estimation in Hydrological Systems Analysis, in D.A. Kraijenhoff and J.R. Moll (eds.), *River Flow Modelling and Forecasting*, Dordrecht, D. Reidel, pp. 129-180, 1986

Young, P.C., Recursive Extrapolation, Interpolation and Smoothing of Nonstationary Time-Series, in H.F. Chen (ed.), *Identification and System Parameter Estimation 1988 Vol. 1*, Oxford, Pergamon Press, pp. 33-44, 1988.

Young, P.C., Parallel Processes in Hydrology and Water Quality, *Jnl. Inst. Water Eng. And Management*, 5, pp. 598-612, 1992

Young, P.C. and Wallis, S.G., Recursive Estimation: a Unified Approach to the Identification, Estimation and Forecasting of Hydrological Systems, *Applied Math. and Computation*, 17, pp. 299-334, 1985.

Young, P.C., Ng, C. and Armitage, P., A Systems Approach to Recursive Economic Forecasting and Seasonal Adjustment, *Computers Math. Applic.*, 18, pp. 481-501, 1989.

Young, P.C., Jakeman, A.J. and Post, D.A., Recent Advances in Data-Based Modelling and Analysis of Hydrological Systems, Proc. WaterMatex Symposium, Quebec: to appear, *Water Science and Technology*, 1998.

Young, P.C., Ng, C. N, Lane, K., and Parker, D. Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data, *Journal of Forecasting*, 10, pp. 57-89, 1991.