International Journal of Forecasting 1 (1985) 179–190
North-Holland

FORECASTING MONTHLY RIVERFLOW TIME SERIES *

Donald J. NOAKES

Pacific Biological Station, Nanaimo, B.C., Canada

A. Ian McLEOD

University of Western Ontario, London, Ont., Canada N6A 3K7

Keith W. HIPEL

University of Waterloo, Waterloo, Ont., Canada N2L 3G1

Mean monthly flows from thirty rivers in North and South America are used to test the short-term forecasting ability of seasonal ARIMA, deseasonalized ARMA, and periodic autoregressive models. The series were split into two sections and models were calibrated to the first portion of the data. The models were then used to generate one-step-ahead forecasts for the second portion of the data. The forecast performance is compared using various measures of accuracy. The results suggest that a periodic autoregressive model, identified by using the partial autocorrelation function, provided the most accurate forecasts.

Keywords: ARMA, Mean monthly riverflow, Periodic autoregression, Seasonality, Stochastic hydrology, Subset autoregression.

1. Introduction

Seasonality of geophysical data adds a degree of complexity to the selection and development of an appropriate stochastic model to fit to a given series. By nature, geophysical data are often seasonally stationary. That is, riverflows or temperatures for a particular season of the year are statistically similar from year to year but may vary drastically across seasons. Hydrologists have also found that riverflow time series exhibit an autocorrelation structure that depends not only on the time lag between observations but also on the season of the year [Moss and Bryson (1974)]. For example, if snowmelt is an important factor in runoff that might occur in March or April, the correlation between observed riverflows for these months may be negative whereas at other times of the year it is usually positive. These characteristics offer a challenge to the practitioner while at the same time provide a convenient structure for the modeller to exploit. A family of Periodic Autoregressive (PAR) models specifically designed to account for seasonal variations is described in this paper.

To examine the efficacy of PAR models, a comprehensive forecasting study is carried out by comparing their performance with that of several models used to model seasonal data. The PAR models are compared to seasonal autoregressive integrated moving average (SARIMA) and desea-

* The authors greatly appreciate the constructive suggestions of three reviewers as well as associate editor Everette S. Gardner, Jr. and editor J. Scott Armstrong. By incorporating their pertinent comments into the paper, the quality of the paper was significantly improved. The forecasting study was partially funded by the Natural Science and Research Council of Canada. The data used in this study are available on request from the authors.

sonalized autoregressive moving average (ARMA) models using thirty monthly riverflow series. Methods of model order selection for the PAR models are also compared. The experiments described in this paper are the most comprehensive yet reported in the hydrological literature. Other published comparisons have used only a few series and usually only two models [see for example Delleur et al. (1976)]. Also, the majority of the hydrological forecasting research to date has concentrated on shorter time intervals in the order of a few hours or days [see for example the Proceedings of the Oxford Hydrological Forecasting Symposium, April 15–18 (1980) and Thompstone et al. (1983)]. However monthly riverflow forecasts are often used for operational planning of reservoir systems. Even modest improvements in the operation of large reservoir systems can result in multi-million dollar savings per year [Brochu (1978)]. Thus, the results of the forecasting study given in this paper should be important of those concerned with the optimal medium and long-term operation of reservoir systems.

The performance of our forecasts is assessed using the mean absolute percentage error (*MAPE*), median absolute percentage error (*MEDIAN APE*), mean absolute deviation (*MAD*) and root mean square error (*RMSE*) criteria. Although these criteria give an indication as to which models seem to perform better, no statement concerning statistically significant differences can be made from such a comparison. To address this question, a nonparametric Wilcoxon test is used to determine if a particular model produces significantly better forecasts when compared to another model.

The next section of the paper describes the types of models used in the forecasting study. The results of the forecasting study of the thirty monthly riverflow series are then described and interpreted.

2. Time series models

Three main approaches to modelling seasonal time series are considered in this paper. In the first approach, a separate autoregressive (AR) model is fitted to each month in order to form a PAR model. This is done in an attempt to account for seasonal changes in the autocorrelation of the series. Different ways are considered for identifying the order of the AR model for each season.

In the second approach, the series is deseasonalized by subtracting the seasonal means and perhaps dividing the seasonal adjustment by the seasonal standard deviation. A non-seasonal ARMA model is then fitted to the deseasonalized series.

In the last approach, the series is seasonally differenced. A linear stochastic model containing both seasonal and non-seasonal parameters is then fitted to the differenced series to determine the estimated SARIMA model. This type of seasonal model is discussed by Box and Jenkins (1970, ch. 9). It has been used by various other researchers for modelling seasonal riverflow time series. The SARIMA model is often considered the most parsimonious model.

2.1. Periodic Autoregressive (PAR) models

Models similar to the family of PAR models presented in this paper have previously been employed by other researchers in hydrology and economics [see for example Clarke (1973) and Parzen and Pagano (1979)]. However, additional results to be used at the identification, estimation, and diagnostic check stages of model development are presented here and in more detail in Noakes (1984).

Let Z_t, $t = 1, 2, \ldots$ be a seasonal time series with period s. The time index t may be regarded as a function of the year T ($T = 1, 2, \ldots, N$), and the season, $m (m = 1, 2, \ldots, s)$. Thus, the time index may be written as $t = (T-1)s + m$. The PAR $(p_1, p_2, \ldots, p_s)$ model is conveniently described by

$$\phi^{(m)}(B)(Z_t^{(\lambda)} - \mu_t) = a_t, \tag{1}$$

where $\phi^{(m)}(B) = 1 - \phi_1^{(m)}B - \ldots - \phi_{p_m}^{(m)}B^{p_m}$; B is the backward shift operator on t; $\mu_t = \mu^{(m)}$ is the mean for period m; and $a_t \sim \text{NID}(0, \sigma^{2(m)})$. The superscript m obeys modulo arithmetic (i.e., $\mu^{(1)} \equiv \mu^{(s+1)} \equiv \mu^{(-s+1)}$). The superscript λ is the exponent of an appropriate Box–Cox transformation. The Box–Cox transformation [Box and Cox (1964)] is given by

$$Z_t^{(\lambda)} = \lambda^{-1}\left[(Z_t + \text{constant})^{\lambda} - 1\right], \qquad \lambda \neq 0,$$

$$= \ln(Z_t + \text{constant}), \qquad \lambda = 0. \tag{2}$$

Two convenient and efficient estimation techniques for PAR models are available. In the first approach, approximate maximum likelihood estimates of the AR parameters are obtained directly from the linear regression of $Z_t^{(\lambda)}$ on $Z_{t-1}^{(\lambda)}, Z_{t-2}^{(\lambda)}, \ldots, Z_{t-p_m}^{(\lambda)}$. The maximum order of autoregression for each season, p_m, does not need to be equal. Subset regression [Morgan and Tatar (1972)] can be used to constrain insignificant autoregression parameters to zero.

In the second scheme, the Yule–Walker equations are formulated and solved to obtain estimates of the model parameters. In the Yule–Walker approach p_m may vary from season to season but all of the AR parameters are estimated. The order of the AR model fit to each season (p_m) may be determined by using some automatic selection criterion such as the *AIC* [Akaike (1974)] or the *BIC* [Schwarz (1978)] or by examining plots of the partial autocorrelation function (*PACF*) for each season.

In general, the *AIC* is defined by

$$AIC = -2 \log L + 2k, \tag{3}$$

where $\log L$ is the maximized log likelihood and k is the number of free parameters estimated. The difference between the *AIC* and the *BIC* is the penalty assigned to the number of model parameters. The second term in the *BIC* is $k \ln n$, where n is the number of observations. Previously, the *AIC* has been used in stochastic hydrology for determining the orders of various kinds of seasonal and non-seasonal models [see for example Hipel et al. (1977)]. In the case of the PAR model, the *AIC* for the mth season is given by Noakes (1984) as

$$AIC_m = 2N \log \hat{\sigma}^{2(m)} - 2(\lambda - 1)\sum_{i=1}^{N} \log Z_{(i-1)s+m} + 2n_m + 4, \tag{4}$$

where $\hat{\sigma}^{2(m)}$ is the residual variance for season m; λ is the exponent of an appropriate Box–Cox transformation; n_m is the total number of parameters for season m; N is the number of years of data and the constant 4 allows for the mean of season m and the residual variance parameter. Once the best AR models are found for each season, the overall *AIC* value for the PAR model can be calculated as

$$AIC = \sum_{i=1}^{s} AIC_i + 2\delta_1, \tag{5}$$

where $\delta_1 = 0$ when $\lambda = 1$ and $\delta_1 = 1$ when $\lambda \neq 1$. The *BIC* may be calculated in a similar way.

In both estimation procedures described in this section, the seasonal mean parameter $\mu^{(m)}$ is estimated by

$$\hat{\mu}^{(m)} = \frac{1}{N} \sum_{i=1}^{N} Z_{(i-1)s+m}^{(\lambda)}, \qquad m = 1, 2, \ldots, s. \tag{6}$$

The residuals $\hat{a}_t$ are calculated from (1) by setting initial values to zero. Then the residual variance is

$$\hat{\sigma}^{2(m)} = \frac{1}{N} \sum_{i=1}^{N} \hat{a}_{(i-1)s+m}^{2}, \qquad m = 1, 2, \ldots, s. \tag{7}$$

It should be noted that the model parameters for the mth season (i.e., $\mu^{(m)}$, $\sigma^{2(m)}$, $\phi_1^{(m)}$, $\phi_2^{(m)}, \ldots, \phi_{p_m}^{(m)}$) can be estimated entirely independently of the model of any other season. Also, since the Fisher information matrix is block diagonal [Pagano (1978)], the estimates of the parameters in different seasons are statistically independent.

In principle, the PAR model could be extended to the case of the periodic autoregressive moving-average (PARMA) models as suggested by Tao and Delleur (1976) and Salas et al. (1982). However, when there are moving average parameters present in the model, the resulting non-linearities cause difficulties in obtaining maximum likelihood estimates of the parameters. If the method of moments is used, the estimates for the PARMA model parameters are inefficient. Consequently, although basic statistical research should continue for PARMA models, this paper is concerned solely with the more practical PAR case.

Six types of PAR models were considered in this study. In the first model, a separate AR(1) model was fitted to each month (PAR/1) using linear regression. This model was originally suggested by Thomas and Fiering (1962) and has been used extensively by hydrologists.

The second and third PAR models were fitted to the data using the algorithm of Morgan and Tatar (1972). This algorithm calculates the residual sum of squares of all possible regressions for each season. The *AIC* and *BIC* can thus be calculated for all possible models. The PAR model which gave the minimum value of the *AIC* or *BIC* (with $p_m \leq 12$) was selected as the most appropriate. This type of procedure has been called subset autoregression by McClave (1975), and thus will be referred to as SUBSET/*AIC* or SUBSET/*BIC* modelling.

The next PAR models were obtained by solving the appropriate Yule–Walker equations. In the first case p_m was selected on the basis of the minimum value of the *AIC* or *BIC*. Unlike the previous case, however, intermediate parameters were not allowed to be constrained to zero. Thus, all of the parameters from $\phi_1^{(m)}$ to $\phi_{p_m}^{(m)}$ were estimated in this model for a given season to produce the PAR/*AIC* and PAR/*BIC* models.

The last PAR models were identified by examining plots of the monthly partial autocorrelation function (*PACF*). In general, an AR(p_m) model was fitted to month m, where p_m was the last lag for which the *PACF* was significantly different from zero. The adequacy of the selected model was checked by testing for significant residual correlation or non-normality [Noakes (1984)]. Thus the PAR/*PACF* is the natural extension to PAR models of the modelling philosophy recommended by Box and Jenkins (1970). Once again no intermediate parameters were constrained to zero.

2.2. Deseasonalized models

A common approach to modelling seasonal geophysical data is first to deseasonalize the series and then fit an appropriate non-seasonal stochastic model to the deseasonalized data. This procedure has been followed by a number of researchers [see for example Delleur et al. (1976)] and it has proven to be particularly useful in modelling seasonality in geophysical data for the purpose of simulation.

Two standard deseasonalization techniques that have been employed are

$$w_{i,j} = Z_{i,j}^{(\lambda)} - \tilde{\mu}_j, \tag{8}$$

and

$$w_{i,j} = \left(Z_{i,j}^{(\lambda)} - \tilde{\mu}_j \right) / \tilde{\sigma}_j, \tag{9}$$

where $Z_{i,j}^{(\lambda)}$ is the transformed observation for the ith year; jth month; $\tilde{\mu}_j$ is the fitted mean for season j; $\tilde{\sigma}_j$ is the fitted standard deviation for season j; and λ is the exponent of an appropriate Box–Cox transformation. In this study the series were deseasonalized using both (8) and (9). The monthly means and standard deviations were estimated by

$$\tilde{\mu}_j = \frac{1}{N} \sum_{i=1}^{N} Z_{i,j}^{(\lambda)}, \tag{10}$$

and

$$\tilde{\sigma}_j = \left[\frac{1}{N} \sum_{i=1}^{N} \left(Z_{i,j}^{(\lambda)} - \tilde{\mu} \right)^2 \right]^{0.5}. \tag{11}$$

After the series were deseasonalized, non-seasonal ARMA models were fitted to the data. Models fitted to data deseasonalized by (8) and (9) will be referred to as ARMA/DSM and ARMA/DES respectively. In both cases, the parameters in the ARMA models were estimated using the maximum likelihood estimation algorithm of McLeod and Sales (1983).

The deseasonalization procedures given in (8) and (9) probably reflect the inherent properties of many geophysical time series. For example, when considering average monthly riverflow data, the observations for any particular month tend to fluctuate about some fixed mean level. Consequently, the deseasonalization method in (8) may be appropriate if the monthly standard deviations of the $Z_{i,j}^{(\lambda)}$ series are more or less constant throughout the year. When both the means and the standard deviations of the $Z_{i,j}^{(\lambda)}$ sequence are different from month to month, then the transformation in (9) should be used.

2.3. Seasonal Autoregressive Integrated Moving Average (SARIMA) models

The SARIMA model for a time series $Z_1, Z_2, \ldots$ can be written as

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D\left(Z_t^{(\lambda)} - \mu_t\right) = \theta(B)\Theta(B^s)a_t, \tag{12}$$

where t is the observation number, s is the number of seasons per year, and a_t is the noise component of the stochastic model assumed to be NID(0, σ_a^2). The terms $\phi(B)$ and $\Phi(B^s)$ are the non-seasonal AR operator of order p and the seasonal AR operator of order P, resectively. Similarly, $\theta(B)$ and $\Theta(B^s)$ are the non-seasonal and seasonal MA operators of orders q and Q, respectively. The non-seasonal differencing operator $(1-B)^d$ is of order d and the series is seasonally differenced D times using $(1-B^s)^D$. The Box–Cox parameter λ is as in eq. (2). The SARIMA models are usually represented by the notation $(p, d, q)(P, D, Q)_s$ where the first set of brackets contains the order of the non-seasonal operators and the second set of brackets the orders of the seasonal operators. The SARIMA models used for the monthly riverflow series were all determined to be of the form $(p, 0, q)(0, 1, Q)_{12}$ with $\lambda = 0$ and with typical values of p, q and Q being 1, 0, and 1.

3. The forecasting study

The data used in this study comprised thirty monthly unregulated riverflow time series ranging in length from thirty-seven to sixty-eight years. The rivers are from a number of different physiographic regions and vary in size from rivers with mean annual flows less than two cubic meters per second (cms) to rivers with mean annual flows exceeding 100 cms. The data for the Canadian rivers were obtained from Water Survey of Canada records, the American riverflow series are from the United States Geological Survey. The Brazilian data were obtained from Electrobras (the national electrical company of Brazil). The rivers and their mean annual flows (in cms) for the water year (October to September) are displayed in exhibit 1.

The general procedure was first to truncate the transformed data sets by omitting the last thrity-six

Exhibit 1
Monthly riverflow data sets.

	River	Location	Period	Observations	Mean flow (cms)
1	American	Fair Oaks, California	1906–1960	660	106
2	Boise	Twin Springs, Idaho	1912–1960	588	33
3	Clearwater	Kamish, Idaho	1911–1960	600	231
4	Columbia	Nicholson, British Columbia	1933–1969	444	109
5	Current	Van Buren, Missouri	1922–1960	468	54
6	W.B. Delaware	Hale Eddy, New York	1916–1960	540	30
7	English	Sioux Lookout, Ontario	1922–1977	660	123
8	Feather	Oroville, California	1902–1977	708	167
9	James	Buchanan, Virginia	1911–1960	600	69
10	Judith	Utica, Montana	1920–1960	492	1
11	Mad	Springfield, Ohio	1915–1960	552	14
12	Madison	West Yellowstone, Montana	1923–1960	456	13
13	McKenzie	McKenzie Bridge, Oregon	1911–1960	600	47
14	Middle Boulder	Nederland, Colorado	1912–1960	588	2
15	Missinaibi	Mattice, Ontario	1921–1976	672	103
16	Namakan	Lac La Croix, Ontario	1923–1977	648	108
17	Neches	Rockland, Texas	1914–1960	564	69
18	N. Magnetawan	Burke Falls, Ontario	1916–1977	732	6
19	Oostanaula	Resaca, Georgia	1893–1960	816	78
20	Pigeon	Middle Falls, Ontario	1924–1977	636	14
21	Rappahannock	Fredericksburg, Virginia	1908–1971	768	45
22	Richelieu	Fryers Rapids, Quebec	1932–1977	468	331
23	Rio Grande	Furnas, Minas Gerais, Brazil	1931–1978	576	896
24	Saint Johns	Fort Kent, New Brunswick	1927–1977	600	30
25	Saugeen	Walkerton, Ontario	1915–1976	744	68
26	S.F. Skykomish	Index, Washington	1923–1960	456	278
27	S. Saskatchewan	Saskatoon, Saskatchewan	1911–1963	624	272
28	Trinity	Lewiston, California	1912–1960	588	47
29	Turtle	Mine Centre, Ontario	1921–1977	672	37
30	Wolf	New London, Wisconsin	1914–1960	564	49

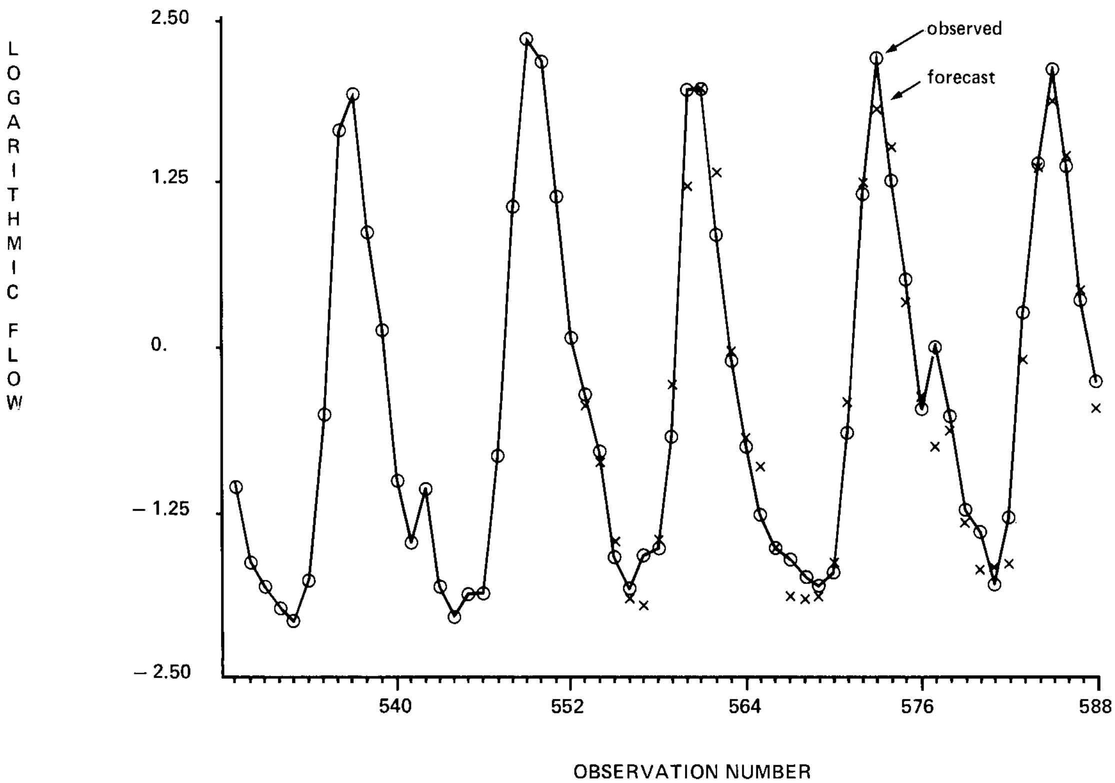

Exhibit 2. Logged monthly riverflow of the Middle Boulder Creek and one-step PAR/*PACF* forecasts.

observations. All data in this study were transformed by taking natural logarithms. This procedure was found to be the most appropriate Box–Cox transformation. The log transformation was needed to ensure that the model residuals were approximately normally distributed and homoscedastic. We used appropriate identification, estimation, and diagnostic checking techniques [Hipel et al. (1977), Noakes (1984)]. SARIMA, ARMA/DSM, ARMA/DES, SUBSET/*AIC*, SUBSET/*BIC*, PAR/*AIC*, PAR/*BIC*, PAR/1, and PAR/*PACF* models that best fit the first portion of the data were then calibrated. The nine models were then used to generate thirty-six one-step-ahead forecasts for the logarithmic flows. Exhibit 2 shows a time series plot of the logarithmic flows and their forecasts using the PAR/*PACF* method for river number 14.

The forecasts, when simply the monthly means of the logarithmic flows (MEANS) were used, were also computed. The logarithmic forecast errors associated with each of the ten forecasting models were then compared using the forecast performance measures *RMSE, MAD, MAPE* and *MEDIAN APE*.

RMSE results are given in exhibit 3 for each river. The results for each performance measure are summarized in exhibits 4 through 7 where rank and rank-sum comparisons appear.

The rank-sums for the models are the sums of the product of the rank and the associated table entry. Thus, models with lower rank-sums performed better than those with larger rank-sums. The models PAR/*PACF*, PAR/1, PAR/*AIC*, PAR/*BIC*, and SUBSET/*BIC* performed very well on the basis of all performance criteria. As expected, using the MEANS proved unsatisfactory in most cases. The MEANS had the worst overall performance and produced the largest *RMSE* for

Exhibit 3
RMSE × 1000 of the logarithmic forecast errors.

River	PAR/ *PACF*	PAR/1	PAR/ *AIC*	PAR/ *BIC*	SUBSET/ *AIC*	SUBSET/ *BIC*	ARMA/ DSM	ARMA/ DES	SARIMA	MEANS
1	857	896	813	864	796	796	801	907	690	1240
2	280	279	273	280	307	289	264	289	273	248
3	323	330	334	331	346	330	359	339	367	544
4	183	190	180	198	204	211	184	181	182	209
5	426	418	445	410	464	423	389	408	390	357
6	658	642	628	666	681	664	689	690	698	775
7	191	218	187	203	218	201	209	205	440	633
8	337	338	394	338	415	335	354	347	358	481
9	516	495	536	544	562	548	489	489	488	579
10	470	469	463	469	500	471	582	427	576	746
11	435	428	416	431	481	440	426	431	424	539
12	98	91	120	90	125	98	98	118	107	127
13	175	175	208	176	254	221	167	171	169	186
14	273	273	272	274	281	296	290	290	302	365
15	619	614	604	618	634	626	707	639	752	961
16	242	244	238	243	248	238	253	259	261	515
17	909	909	930	910	1078	906	916	907	969	1147
18	407	407	416	407	419	407	408	411	419	440
19	424	418	425	420	427	425	448	447	446	487
20	600	591	592	604	627	618	673	707	694	1118
21	530	546	536	547	570	535	553	552	564	569
22	250	266	264	270	326	274	277	270	260	600
23	230	226	265	229	294	241	241	236	242	335
24	411	412	398	420	414	428	389	385	398	379
25	425	402	430	421	479	422	433	423	432	532
26	380	391	407	422	434	401	411	416	411	476
27	436	438	420	379	500	391	464	445	461	587
28	626	624	624	633	603	632	628	639	627	822
29	282	283	282	283	318	283	283	301	297	410
30	355	358	408	367	368	372	352	361	352	465

Exhibit 4
RMSE of one-step *MMSE* forecasts of logged series (number of times each method had indciated rank).

Rank	PAR/ *PACF*	PAR/1	PAR/ *AIC*	PAR/ *BIC*	SUBSET/ *AIC*	SUBSET/ *BIC*	ARMA/ DSM	ARMA/ DES	SARIMA	MEANS
1	4	3	7	3	1	4	1	1	3	3
2	3	5	5	2	0	5	4	3	3	0
3	10	2	3	4	1	4	2	2	2	0
4	3	11	0	7	0	2	2	3	2	0
5	5	3	5	6	1	3	3	3	1	0
6	3	3	2	2	5	1	7	6	1	0
7	1	1	2	4	4	3	4	3	7	1
8	1	2	4	1	3	5	5	5	4	0
9	0	0	2	1	11	1	2	4	7	2
10	0	0	0	0	4	2	0	0	0	24
Rank-sum	110	119	127	134	230	145	166	173	178	268

Exhibit 5
MAD of one-step *MMSE* forecasts of logged series (number of times each method had indicated rank).

Rank	PAR/ *PACF*	PAR/1	PAR/ *AIC*	PAR/ *BIC*	SUBSET/ *AIC*	SUBSET/ *BIC*	ARMA/ DSM	ARMA/ DES	SARIMA	MEANS
1	4	4	4	1	1	5	1	2	5	3
2	6	4	4	4	1	4	3	3	1	0
3	5	8	5	3	1	4	2	2	0	0
4	6	6	2	8	1	4	1	1	1	0
5	6	2	4	6	0	2	5	1	4	0
6	2	3	3	2	6	3	4	4	2	1
7	0	1	2	5	3	3	6	6	4	0
8	1	2	4	1	3	2	6	3	8	0
9	0	0	2	0	10	3	1	8	5	1
10	0	0	0	0	4	0	1	0	0	25
Rank-sum	105	111	137	135	221	133	175	185	180	268

twenty-four of the series. Exhibit 3 shows that in the three cases (rivers 2, 5, and 24) where the MEANS had the smallest *RMSE* there is very little difference between any of the forecast methods. Moreover, in these three cases all methods had low *MAPE*s and *RMSE*s. At the other extreme, the best alternative to MEANS for rivers 7, 16, and 22 had an RMSE less than half that of MEANS. Next to the PAR models mentioned above, the ARMA/DSM, ARMA/DES, and SARIMA performed about equally as well. The SUBSET/*AIC* model performance was disappointing although not totally surprising. The large number of parameters associated with the SUBSET/*AIC* model does not provide a sufficiently parsimonious model for accurate forecasts. The importance of parsimony in forecasting models is discussed by Ledolter and Abraham (1981).

We noticed an extreme disparity between the *MAPE* and *MEDIAN APE* for several rivers. This was found to be due to a defect in the absolute percentage error when the observed value is small. For example, the observed logged flow for river 14 for November 1959 was 0.0024 and the PAR/*PACF* forecast was -0.746. This resulted in an absolute percentage error of over 31,000!

Exhibit 6
MAPE of one-step *MMSE* forecasts of logged series (number of times each method had indicated rank).

Rank	PAR/ *PACF*	PAR/1	PAR/ *AIC*	PAR/ *BIC*	SUBSET/ *AIC*	SUBSET/ *BIC*	ARMA/ DSM	ARMA/ DES	SARIMA	MEANS
1	3	5	3	1	3	5	1	1	5	3
2	5	4	3	5	2	3	2	5	1	0
3	4	7	4	4	1	3	5	1	0	1
4	7	2	5	7	0	3	3	1	2	0
5	7	6	4	2	1	3	1	2	4	0
6	2	2	1	5	1	4	6	5	4	0
7	1	1	2	5	2	3	5	7	3	1
8	1	3	6	1	3	2	6	2	4	2
9	0	0	2	0	11	4	0	6	7	0
10	0	0	0	0	6	0	1	0	0	23
Rank-sum	115	115	147	134	218	144	166	177	175	259

Exhibit 7
MEDIAN APE of one-step *MMSE* forecasts of logged series (number of times each method had indicated rank).

Rank	PAR/ *PACF*	PAR/1	PAR/ *AIC*	PAR/ *BIC*	SUBSET/ *AIC*	SUBSET/ *BIC*	ARMA/ DSEM	ARMA/ DES	SARIMA	MEANS
1	5	1	3	1	6	4	2	1	3	4
2	3	3	5	4	4	2	1	3	4	1
3	4	5	6	2	0	3	3	4	2	1
4	6	4	3	6	2	3	2	3	0	1
5	4	5	5	6	2	1	3	2	2	0
6	3	3	1	3	4	6	3	2	4	1
7	3	6	2	2	2	6	2	5	2	0
8	1	1	2	3	3	2	7	4	5	2
9	1	1	3	1	4	2	4	5	8	1
10	0	1	0	2	3	1	3	1	0	19
Rank-sum	123	150	131	154	160	156	190	175	177	234

Although the above results of the logarithmic forecast errors are certainly important, the real concern of hydrologists is how well the models forecast actual riverflows. When obtaining forecasts of the flows in the untransformed domain, the following method produces minimum mean squared error (*MMSE*) forecasts [Granger and Newbold (1976)]:

$$\hat{Z} = \exp\left[\hat{Z}^{(\lambda)} + \tfrac{1}{2}\hat{\sigma}_a^2\right], \tag{13}$$

where $\hat{Z}^{(\lambda)}$ is the one-step-ahead forecast produced by the model for the logged data and $\hat{\sigma}_a^2$ is the appropriate residual variance. Then $\hat{Z}$ is the *MMSE* forecast for the flow. The *MSE* of the flow forecast errors are presented in Noakes (1984). A summary of those results is shown in exhibit 8. Once again the same PAR models performed very well. As expected, there are some variations due to the transformation given in (13). Note particularly the improvement of the MEANS model and the dismal performance of the ARMA/DES models. The ARMA/DSM and SARIMA models still performed reasonably well and the SUBSET/*AIC* improved slightly.

Exhibit 8
RMSE of one-step *MMSE* forecasts of the flows (number of times each method had indicated rank).

Rank	PAR/ *PACF*	PAR/1	PAR/ *AIC*	PAR/ *BIC*	SUBSET/ *AIC*	SUBSET/ *BIC*	ARMA/ DSM	ARMA/ DES	SARIMA	MEANS
1	2	5	7	0	4	4	3	0	2	3
2	5	4	5	6	0	3	3	0	4	0
3	11	3	3	5	0	3	2	0	1	2
4	6	6	4	5	1	5	1	1	0	1
5	1	8	2	8	2	6	0	0	3	0
6	4	3	3	3	5	1	6	0	1	4
7	1	0	4	1	2	5	3	5	4	5
8	0	1	1	1	5	2	7	5	4	4
9	0	0	1	1	10	1	4	1	8	4
10	0	0	0	0	1	0	1	18	3	7
Rank-sum	105	112	115	129	202	135	178	268	196	210

Exhibit 9
Results of Fisher's test.

Model	PAR/1	PAR/ YW1	PAR/ *AIC*	PAR/ *BIC*	SUBSET/ *AIC*	SUBSET/ *BIC*	ARMA/ DSM	ARMA/ DES	SARIMA	MEANS
χ^2_{cal}	65.2	60.5	64.8	57.8	116.0	87.8	99.4	101.2	113.3	276
DF	60	40	58	40	60	60	60	60	60	60
sl(%)	30	2	25	3	10^{-4}	1	0.1	0.05	10^{-5}	10^{-13}

A Wilcoxon rank-sum test for paired data was used to test for statistically significant differences in the forecasting ability of the various procedures. In this test the differences in the squares of the logarithmic forecast errors were computed. These differences were ranked in ascending order, without regard to sign, and assigned ranks from one to thirty-six. The sum of the ranks of all positive differences [$T(+)$] were then computed and compared to tabulated values in order to ascertain if the forecasts from one model were significantly better than the forecasts from a competing model. These results were then used to examine the performance of the models across all thirty series. In this test, the $P-$ value associated with each $T(+)$ value was calculated by estimating the area in the tail of the distribution. Then the Fisher (1970, p. 99) method for combining significance levels is

$$-2\sum_{i=1}^{k} \ln p_i = \chi^2_{2k}, \tag{14}$$

where p is the calculated P-value associated with $T(+)$ and k is the number of series considered in the test. This combination technique generally has greater power than alternative methods such as simply summing the $T(+)$s.

Fisher's test was employed to compare the overall performance of the PAR/*PACF* model to that of the other competing models. In addition, the PAR/1 parameters were also estimated using the Yule–Walker equations to provide an additional model for comparison (PAR/YW1). In this way identical forecasts produced by the PAR/*PACF* and PAR/YW1 models could be ignored, ensuring that only the differences in the forecasting procedures were compared. The results of Fisher's test are presented in exhibit 9. The PAR/*PACF* was significantly better than all of the models except the PAR/1 and the PAR/*AIC* at the five-percent level. Since different estimation procedures were employed for the PAR/*PACF* and PAR/1 models, there were several forecasts that were almost but not quite identical. These were all included in the analysis, thus masking the differences in the performance of the two models. The PAR/YW1, however, emloyed the same estimation procedure, thus resulting in identical forecasts when an AR(1) model was identified for a particular month for the PAR/*PACF* model. This allowed ties to be dropped from consideration, and resulted in the testing of only the differences between the two models. All series with fewer than five untied forecasts were dropped from consideration in this test. The results of this comparison indciated that when ties were ignored, the PAR/*PACF* model is better than the PAR/YW1 model at the two-percent level of significance. Although the PAR/*AIC* compares quite favourably with the PAR/*PACF* when the significance levels are combined, detailed examination of the results revealed that for three rivers the PAR/*PACF* forecasted significantly better at the five-percent level than the PAR/*AIC*. However, in no case were the PAR/*AIC* forecasts significantly superior to those of the PAR/*PACF*. Additional details are given in the thesis of Noakes (1984).

4. Conclusions

Based upon the results of this study and also upon a physical understanding of how seasonal hydrologic time series behave, certain types of PAR models are recommended for use in forecasting. In particular, the PAR/*PACF* model forecasted most accurately for the data examined in this study. Other models may be more parsimonious, but the PAR/*PACF* model gives the most parsimonious *adequate* model due to the seasonal correlation effects.

References

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

Box, G.E.P. and D.R. Cox, 1964, An analysis of transformations, Journal of the Royal Statistical Society A127, 211–252.

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

Brochu, M., 1978, Computerized rivers, Science Dimension 10, 18–20.

Clarke, R.T., 1973, Mathematical models in hydrology, Irrigation and drainage paper, Food and Agriculture Organization of the United Nations, Rome.

Delleur, J.W., P.C. Tao and M.L. Kavvas, 1976, An evaluation of the practicality and the complexity of some rainfall and runoff time series models, Water Resources Research 12, 953–970.

Fisher, R.A., 1970, Statistical methods for research workers (Oliver and Boyd, Edinburgh).

Granger, C.W.J. and P. Newbold, 1976, Forecasting transformed series, Journal of the Royal Statistical Society B38, 189–203.

Hipel, K.W., A.I. McLeod and W.C. Lennox, 1977, Advances in Box–Jenkins modelling, Water Resources Research 13, 567–586.

Hydrological forecasting symposium, 1980, International Association of Hydrological Sciences, Proceeding of the Oxford Symposium, April, IAHS pub. no. 129.

Ledolter, J. and B. Abraham, 1981, Parsimony and its importance in time series forecasting, Technometrics 23, 411–414.

McClave, J.T., 1975, Subset autoregression, Technometrics 17, 213–220.

McLeod, A.I. and P.R.H. Sales, 1983, An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models, Applied Statistics 32, 211–223.

Morgan, J.A. and J.F. Tatar, 1972, Calculations of the residual sum of squares for all possible regressions, Technometrics 14, 317–325.

Moss, M.E. and M.C. Bryson, 1974, Autocorrelation structure of monthly streamflows, Water Resources Research 10, 737–744.

Noakes, D.J., 1984, Applied time series modelling, seasonality and forecasting, Ph.D. dissertation (University of Waterloo, Waterloo, Ont.).

Pagano, M., 1978, On periodic and multiple autoregression, The Annals of Statistics 6, 1310–1317.

Parzen, E. and M. Pagano, 1979, An approach to modelling seasonally stationary time series, Journal of Econometrics 9, 137–153.

Salas, J.D., D.C. Boes and R.A. Smith, 1982, Estimation of ARMA models with seasonal parameters, Water Resources Research 18, 1006–1010.

Schwarz, G., 1978, Estimating the dimension of a model, The Annals of Statistics 6, 461–464.

Tao, P.C. and J.W. Delleur, 1976, Seasonal and nonseasonal ARMA models in hydrology, ASCE, Journal of the Hydraulics Division 102, HY10, 1541–1559.

Thomas, H.A. and M.B. Fiering, 1962, Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation, in: Maass et al., eds., Design of water resources (Harvard University Press, Cambridge, MA) 459–493.

Thompstone, R.M., K.W. Hipel and A.I. McLeod, 1983, Transfer function noise modelling for powerhouse inflow forecasting, INFOR 21, 258–269.