Nordic Hydrology, 7, 1976, 43-56

Stochastic Modelling of Monthly Streamflow in the Federal Republic of Germany

S. A. Robinson and F. G. Rohde
Technical University, Aachen, Germany

Monthly streamflow series from five catchment areas, representative of some of the typical types encountered in Germany, are analysed. The time series structure of the streamflow can be described by a set of statistical parameters from which the catchment areas can be parametrically classified and regionalised according to the physical characteristics, local climate and geographical location. Simple stochastic models, with emphasis on the seasonal multiplicative model, are considered. These models describe the streamflow generating process and may be used for synthetic data generation for water resources planning and operational purposes. Their performance and limitations are discussed.

Introduction

Research in hydrology in the Federal Republic of Germany has accelerated rapidly in the last ten years as modern mathematical methods have been applied to theoretical and real problems. A representative committee was founded in 1965 for the I.H.D. (I.H.P.) and many coordinated research programmes have been carried out through universities and federal institutes. Some of the German contributions to the I.H.D. are outlined in a publication (DFG/IHD, 1972) and since 1965 a yearbook containing specific data of hydrologic conditions has been published (DFG/IHD, 1968). Since most of the research is documented via institute reports and German periodicals, relatively little has been published internationally via international journals and conferences. This initial paper outlines some current lines of research concerned with streamflow data analysis and modelling.

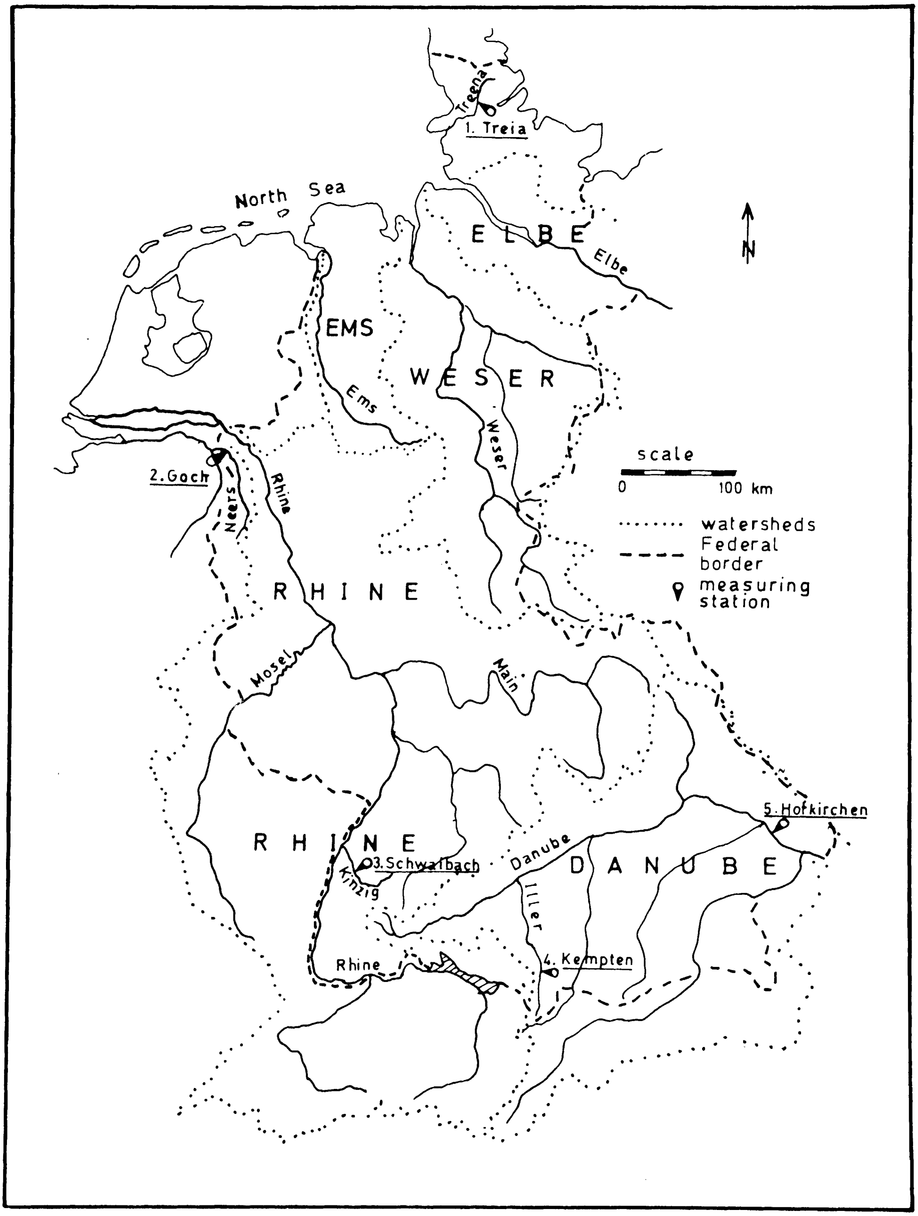


Fig. 1. Map of the Federal Republic of Germany showing the principal catchment areas and location of the measuring stations.

Regional Hydrologic Conditions

Fig. 1 shows the five principal catchment and subcatchment areas in Germany. The hydrologic conditions and types of catchments have a great variation depending on their geographical location. The Rhine, as one of the largest and most commercially important rivers in Europe has much of its catchment area in Germany. The other large rivers - the Danube, Weser, Elbe and the Rhine tributaries - the Main and Mosel are utilised extensively. The upper reaches of the Rhine and the southern tributaries of the Danube are strongly influenced by the high Alpine region and its extreme climate of high rain and snowfall. In the north the river regimes and catchments are characterised by the flat landscape and low rainfall area of the German plain, whereas the hilly areas of moderate slopes in the central and southwest regions produce intermediate catchments influenced by the local orographic rainfall conditions.

Qualitative hydrologic descriptions and basic statistics of the principle rivers and flow regimes have been well documented in the past, but in order to make a complete classification of catchment types and their regional characteristics a comprehensive data analysis is necessary. In the last 10-15 years hydrologic events have been recognised as stochastic processes - having a time varying random probabilistic structure - and studied by advanced statistical and time series analysis techniques. In this study mean monthly streamflows from various representative catchment areas are analysed. Monthly time intervals reflect medium term properties such as seasonal-annual variations and smooth out short term effects and complex interactions of the hydrologic cycle. A monthly streamflow record therefore gives a measure of the net output and behaviour of a catchment system in a single measured series and a time series analysis results in a set of statistical parameters which give a quantitative description of the flow regime and structure related to the physical characteristics and properties of the catchment.

Methods of Analysis

Statistical time series analysis techniques are employed to analyse the series in the amplitude, time and frequency domains.

The time domain is of principal interest when investigating the stochastic structure of a hydrologic time series. By autocorrelation analysis the autocorrelation function (a.c.f.)

$$r_k \equiv \frac{C_k}{C_0} \quad \text{where} \quad C_k \equiv \frac{1}{n}\sum_{t=1}^{n-k} (X_t-\bar{X})(X_{t+k}-\bar{X}) \tag{1}$$

is the autocovariance at lag k and x_t, $t=1$, n is the time series with mean $\bar{X}$ measures the serial correlation or time dependence of variables separated by k units of time and can be illustrated in the form of a correlogram r_k plotted against k.

A transformation of the a.c.f. in the frequency domain gives the spectral density function (s.d.f.). Roesner and Yevjevich (1966) outlined the procedure for monthly hydrologic time series.

Some of the typical characteristics of hydrologic series resulting from a time series analysis are directly dependent on the physical characteristics and location of the catchment. Cyclicity or periodicity produces expected high-low flows, as an annual cycle, caused by the local seasonal climatic conditions controlled by the 1 year terrestrial cycle.

Cyclicity can be identified by distinct cyclic patterns in the a.c.f. and, correspondingly, by prominent peaks in the s.d.f.

Another typical property is persistence. The lag 1 serial correlation (r_1 in Eq. (1)) is an indication of the average, month to month persistence reflecting carry over, storage characteristics of a catchment. This parameter is not constant but can have a seasonal variation. The month by month lag 1 autocorrelation illustrates this.

Mathematical Models

For describing and fitting streamflows in the F.R. Germany three models were considered:

a) Thomas-Fiering model (Maass et al., 1962; Fiering, 1967)
b) Deterministic-random series model (Roesner and Yevjevich, 1966)
c) Seasonal-multiplicative model (Bacon, 1965; Box and Jenkins, 1970)

a) The Thomas-Fiering model is given by:

$$Q_{t+1} = \bar{Q}_{j+1} + r_j \frac{s_{j+1}}{s_j} (Q_t - \bar{Q}_j) + \varepsilon_t s_{j+1} (1-r_j^2)^{\frac{1}{2}} \quad (2)$$

where Q_{t+1}, Q_t are flows at time $t+1$, t; Q_{j+1}, s_{j+1}; Q_j, s_j are means and standard deviation of flows in month $j+1$, j; r_j is the regression coefficient between flows in month $j+1$ and j and ϵ_t is a random independent variable.

For non-normally distributed (skewed) streamflows the random component may be transformed from normal to an approximately gamma distributed random variable (Matalas, 1967).

b) Deterministic-random series model

Here streamflow is considered to be the sum of a deterministic and random component: $X_t = D_t + Z_t$ in which a deseasonalising transformation, involving monthly means and standard deviations $\bar{X}_j$, s_j, is applied to remove the deterministic annual cycle: $Z_t = (X_t - \bar{X}_j)/s_j$.

The resulting random series Z_t now has, in general, a simple markov structure with autoregressive parameter φ:

$$Z_t = \phi Z_{t-1} + (1-\phi^2)^{\frac{1}{2}} \varepsilon_t \qquad (3)$$

which is the particular case, first order autoregressive AR (1), of the class of linear random models or ARIMA models as proposed by Box and Jenkins (1970). For non-normally distributed streamflows a logarithmic transformation of the original series X_t can be applied.

As an alternative to the transformation involving sample monthly means and standard deviations an analytic function such as a polynomial or harmonic series may be fitted to the annual cycle. A harmonic series for the mean is fitted by

$$X(\tau) = \bar{X} + \sum_i (a_i \cos \frac{2\pi i \tau}{12} + b_i \sin \frac{2\pi i \tau}{12}) \; ; \quad \tau = 1,12 \qquad (4)$$

where $\overline{X}$ is mean of series X_t, and a_i and b_i are harmonic coefficients:

$$a_i = \frac{2}{n} \sum_{i=1}^{n} (X_t - \bar{X}) \cos \frac{2\pi i t}{12} \qquad b_i = \frac{2}{n} \sum_{i=1}^{n} (X_t - \bar{X}) \sin \frac{2\pi i t}{12}$$

$$i = 1,6 \text{ (for 1 and 6, } a_i = a_i/2, \; b_i = 0).$$

Usually only the first 2 harmonics - annual and 6-monthly cycle, i=1,2 are necessary to describe the annual cyclicity.

A useful statistical parameter is the quantity »proportion of variance of series explained by the harmonics« given by

$$\sum_i (a_i^2 + b_i^2)/2s^2$$

where i is the number of significant harmonics and s^2 is the variance of series X_t.

This measures the proportional contribution of the deterministic cycle to the total variance and is therefore a further structural property of the series.

c) Seasonal-multiplicative model

In both the previous models an underlying deterministic cycle, estimated from monthly means or a fitted harmonic function, is assumed. In some streamflow series distinct deterministic patterns may be doubtful or difficult to confidently estimate even from a long historic data record. A different method of deseasonalising and modelling is possible by applying the differencing technique as proposed by Box and Jenkins (1970). Here an annual-cyclic series can be made stationary by differencing annually:

$$w_t = \nabla^{12} X_t = X_t - X_{t-12} \qquad (5)$$

and a multiplicative ARIMA $(p,d,q) \times (P,D,Q)_{12}$ model fitted to the w_t series.

An intensive study was carried out to discover the model order $(p,d,q)\,(P,D,Q)$ for a typical streamflow series (Rohde and Robinson, 1975). The most appropriate model was identified as (1,0,1) x $(0,1,1)_{12}$ given by

$$(1-\phi B)w_t = (1-\theta B)(1-\Theta B^{12})a_t \tag{6}$$

where B is a backward shift operator and a_t random noise. In generative form

$$X_t \equiv X_{t-12} + \phi X_{t-1} - \phi X_{t-13} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13} \tag{7}$$

Here only 3 parameters φ, θ, Θ and residual variance S^2a define the model which describes the streamflow series in which the autoregressive parameter φ accounts for both cyclicity and persistence. The parameters can be tentatively estimated from the a.c.f. of $w_t = \nabla^{12}X_t$ series or, more efficiently, from a grid search over the surface of the sum of square of residuals

$$\sum_{t=1}^{n} a_t^2 (\phi, \theta, \Theta) \; .$$

Diagnostic checking by a residual analysis tests the suitability of the model to fit the series.

Catchment Areas and Data

A total of 23 typical river basins of the F.R. of Germany are described in the I.H.D. yearbook (D.F.G./I.H.D., 1968). From these 5 catchment areas were selected for this study representative of the different types of drainage areas encountered in respect of size, geomorphology and climatic conditions. Table 1 lists the stations and their geographical location is shown in Fig. 1.

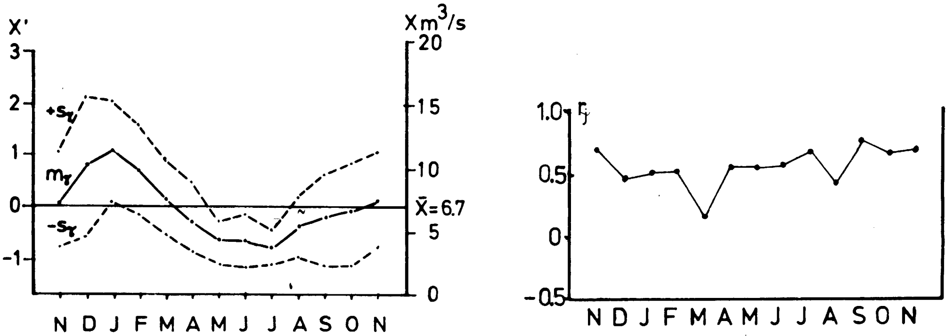
Table 1 – Basic statistics of the rivers and catchment areas

catchment (river)	station	catchment size km²	type	record length (years)	mean flow m³/s	standard deviation of flow	skew-ness	explained variance %	r_1
1 Treena	Treia	480	lowland	24	6.73	4.10	1.18	36	0.60
2 Neers	Goch	1 220	flatland	14	7.96	3.75	1.10	34	0.66
3 Kinzig	Schwaibach	955	hilly	14	22.20	15.20	1.40	24	0.39
4 Iller	Kempten	953	alpine	30	46.38	26.00	0.66	45	0.55
5 Danube	Hofkirchen	49 496	mixed	30	644.10	270.90	1.29	13	0.52

1. Treena (Treia)

The Treena catchment is relatively small and homogeneous, situated in a low undulating area in the far north of the country. The river meanders in a valley plain approximately 200 m wide and many small lakes of surface areas 12-70 hectares exist which help to regulate the flow. The average discharge of 6.26 m³/s is low due to the relatively low but regular annual rainfall of appr. 800 mm.

Winter high flows are approximately three times summer low flows and a skewness of 1.18 indicates non-normally distributed flow. Fig. 2.1 (a) shows the monthly means bounded by the ± monthly standard deviations computed from standardised variables $\overline{X}_t = (X_t - \overline{X})/s_X$ thereby illustrating the expected annual cycle and its variability. In Fig. 2.1(b) the variability in the lag 1 monthly autocorrelations indicates non-stationarity in the autocovariance.



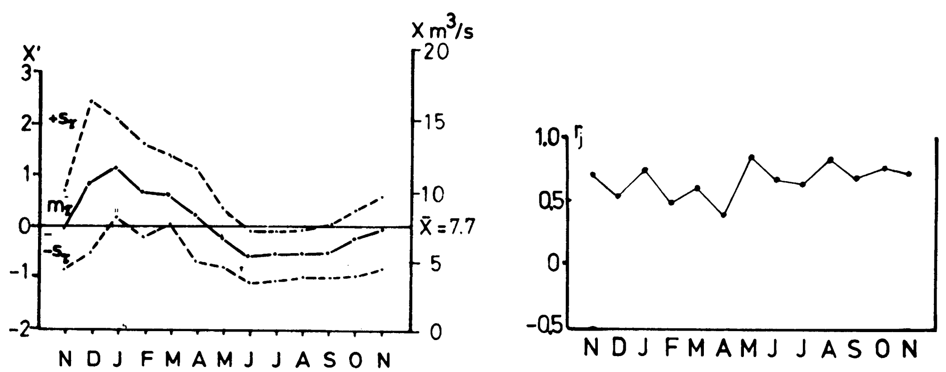
(a) monthly mean flow m_τ and standard deviation s_τ

(b) monthly autocorrelations

Fig. 2.1. River Treena (Station Treia)

2. Niers (Goch)

The source of the river Niers lies in the flat ridges which form the basin divide between the Rhine and the Maas. The mean slope of the catchment is only o.55 $^0/_{00}$ signifying a flatland river. The runoff is relatively low with winter flows approximately twice summer flows. The lag 1 autocorrelation (r_1) of 0.66 indicates a high degree of persistence due to the retention properties and slow response of the flat land form. Correspondingly Fig. 2.2(b) shows that this persistence is relatively constant throughout the year.



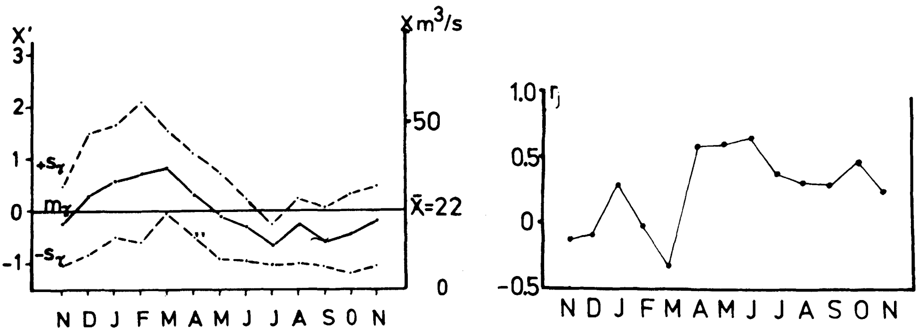
(a) monthly mean flow m_τ and standard deviation s_τ

(b) monthly autocorrelations

Fig. 2.2. River Niers (Station Goch)

3. Kinzig (Schwaibach)

The Kinzig is the largest tributary flowing from the Black Forest to the upper Rhine. Approximately 85% of the drainage area lies in the Black Forest - a heavily forested and hilly area with hard impervious underlying rock formations. The moderate slopes and high rainfall gives a relatively high average discharge of 22 m³/s with very high flood flows of more than 750 m³/s. The orographic and very variable nature of the rainfall in this area contributes to the very variable monthly autocorrelations producing flows of low persistence. This is particularly noticeable in winter where snowfall causes an unusual negative autocorrelation between March and February flows.



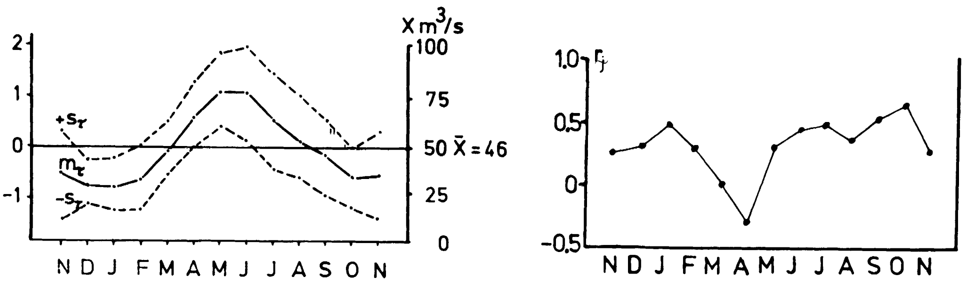
(a) monthly mean flow m_τ and standard deviation s_τ

(b) monthly autocorrelations

Fig. 2.3. River Kinzig (Station Schwaibach)

4. Iller (Kempten)

The Iller rises in the northern limestone Alps and almost all the drainage area is mountainous thus giving the river alpine characteristics with summer high flows three times winter low flows. The average discharge is very high caused by the large amount of precipitation in the form of both rain and snow (appr. 2000 and 500 mm annually). The steep slopes and local climatic conditions give the catchment almost instantaneous response to rainfall with very little storage and losses. The dependability of the large snowmelt runoff gives the river a very distinct and regular annual cycle, the monthly means being closely bounded by the monthly standard deviations (Fig. 2.4. (a)). This is confirmed by the relatively high value of explained variance (highest of all the series tested) of the fitted annual cycle of 45%. In Fig. 2.4 (b) can be seen a highly negative autocorrelation between April and March flows due to the large accumulation of snow in the winter low flow period whereas in the snowmelt period in late spring and summer persistently high flows result.



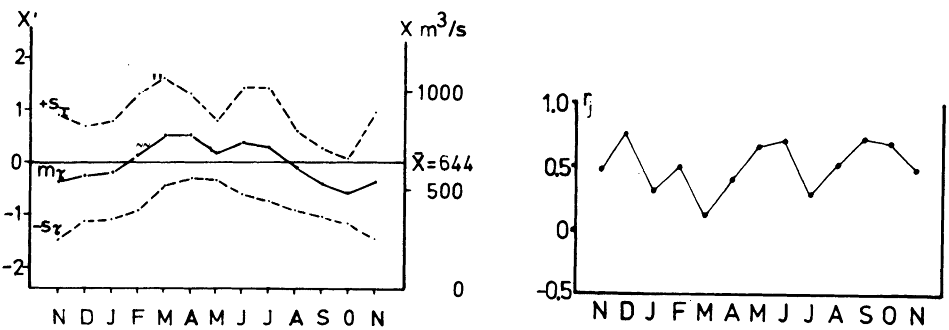
(a) monthly mean flow m_τ and standard deviation s_τ

(b) monthly autocorrelations

Fig. 2.4. River Iller (Station Kempten)

5. The Danube (Hofkirchen)

The Danube river rises in the Black Forest and flows eastwards towards the measuring station at Hofkirchen, approximately 500 km from the source and covering an area of 47 496 km². The southern tributaries rise in the Alps and contribute much of the volume of flow together with upland tributaries from the north. The individual effects of the many contributing subcatchments of differing characteristics are smoothed out thus giving the Danube a flow of low seasonality with a corresponding high low/high flow ration of only 1:1.5. Fig.2.5(a) shows that the individual monthly flows could have a large variation about their expected mean values. Correspondingly the explained variance of the fitted annual cycle was the lowest of all the series at 12.5%.



(a) monthly mean flow m_τ and standard deviation s_τ

(b) monthly autocorrelation

Fig. 2.5. River Danube (Station Hofkirchen)

Results of model fitting

From the foregoing analysis there was no reason to suspect that the Thomas-Fiering model (a) and the deterministic-random model (b) were inappropriate for the streamflow series considered. Each series could be satisfactorily fitted by either model, the models describing the flow regimes and having the ability to generate long synthetic records in which the low order statistics and autocorrelation structure is preserved. The Thomas-Fiering model would be preferred if the monthly autocorrelations varied greatly, as they do, (Figs. 2.1(b) - 2.5(b)) since then seasonal persistence and non-stationarities in autocovariance would be accounted for. Fig. 3(a) shows a 10 year generated flow trace by the Thomas-Fiering model for the Treena; superimposed is the annual cycle of monthly means. A comparison of monthly autocorrelations between the historic and synthetic series is shown in Fig. 3(b).

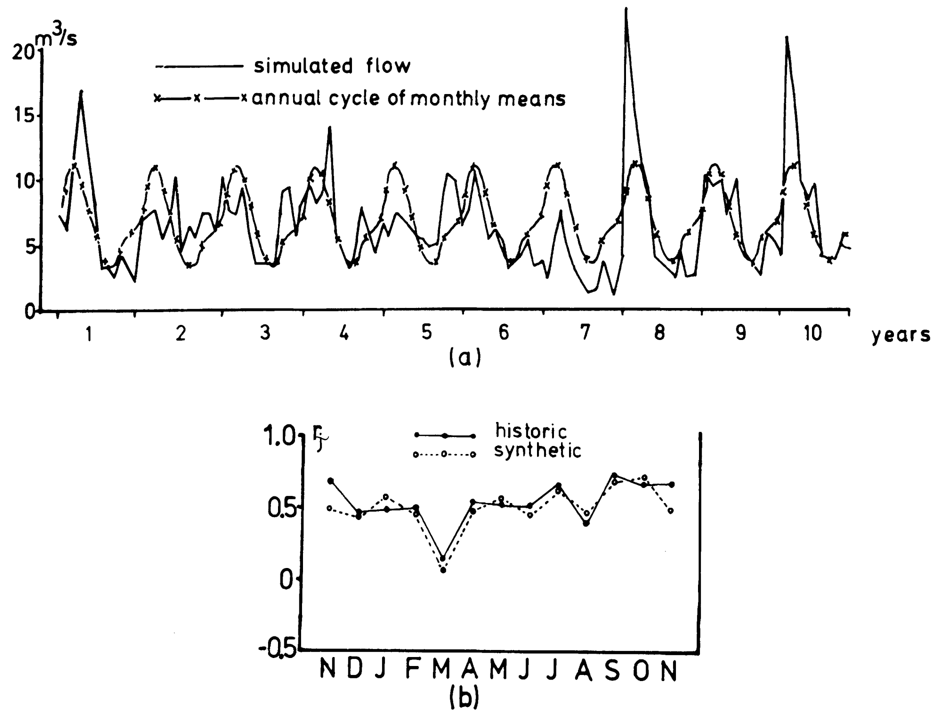


Fig. 3. (a) A ten year synthetic trace generated by the Thomas-Fiering model and (b) a comparison of the historic-synthetic monthly autocorrelations for the river Treena.

For the seasonal multiplicative model all the test data series were tried. It was found that by differencing the historic series with respect to years i.e. $w_t = \nabla^{12} x_t$, the resulting series w_t displayed approximately stationary characteristics as illustrated in the a.c.f. Fig.4.1 shows the a.c.f. of the w_t series for the Treena. This form was identified as the $(1,0,1) \times (0,1,1)_{12}$ model in which the autoregressive parameter φ explains the initial decay and φ and Θ together account for the symmetrical depression around lag 12.

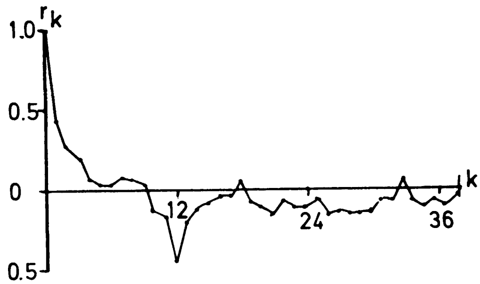


Fig. 4.1. Autocorrelation function of the $\nabla^{12}X_t$ series for the river Treena

Fig. 4.2 shows the residual sum of squares surface for $\varphi = 0.7$. The residual variance is plotted over the range of parameter values θ and Θ and the shaded area represents 95% confidence region bounded by the 0.411 contour around the minimum residual variance of 0.398 at the point $\theta = 0.3, \Theta = 0.9$. Furthermore the surface was found to be sensitive to the parameter Θ but insensitive to φ and θ when considered together meaning that the Treena flows could be fitted with approximate parameters $\varphi \approx 0.5$; $\theta \approx 0.0$; $\Theta \approx 0.9$ (2 parameters) or only slightly better with $\varphi \approx 0.7$; $\theta \approx 0.3$; $\Theta \approx 0.9$. Diagnostic checking, which involved a number of tests of independence of residuals, tested the suitability of the model to represent the series.

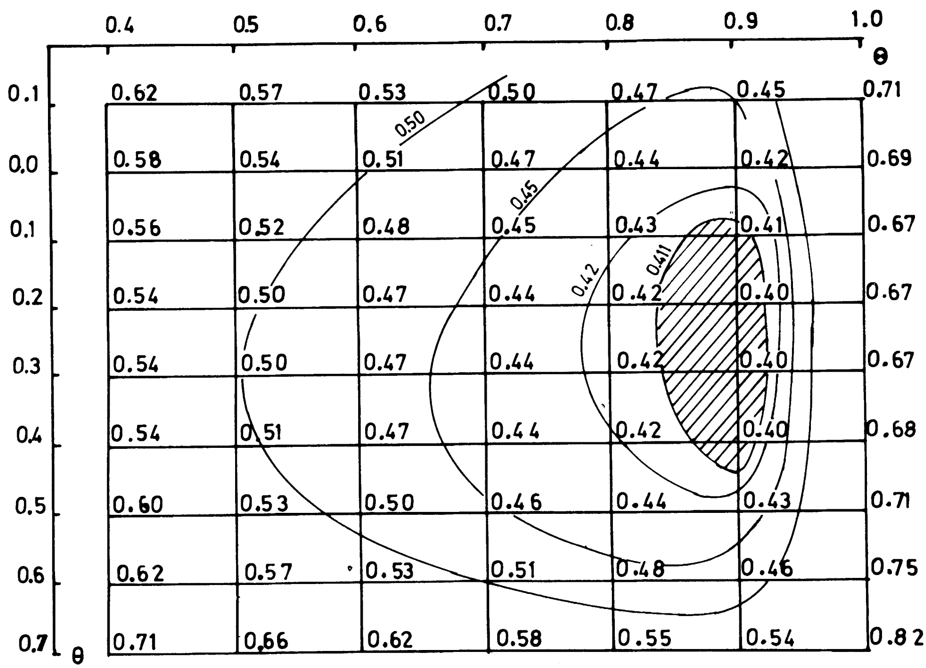


Fig. 4.2. Sum of squares surface of residual variance of the (1,0,1)×(0,1,1) model with φ =0.7 for the river Treena (Treia). The 95% confidence region is bounded by the 0.411 contour.

Table 2 summarises the results in which only the series for the Danube at Hofkirchen failed. In the analysis stage the Danube was shown to be an exceptional case of a large river and catchment area with many different types of contributing subcatchments giving very variable streamflow conditions which could not be explanined by the model.

Table 2 - Results and parameters of the seasonal multiplicative model

catchment (river)	station	indification (model)	parameter estimation	residual variance	diagnostic checking
1 Treena	Treia	$(1,0,1) \times (0,1,1)_{12}$	$\varphi = 0.712$ $\theta = 0.307$ $\Theta = 0.915$	0.396	*P*
2 Neers	Goch	$(1,0,1) \times (0,1,1)_{12}$	$\varphi = 0.810$ $\theta = 0.312$ $\Theta = 0.854$	0.355	*P*
3 Kinzig	Schwaibach	$(1,0,0) \times (0,1,1)_{12}$	$\varphi = 0.216$ $\theta = 0.000$ $\Theta = 0.907$	0.531	*P*
4 Iller	Kempten	$(1,0,1) \times (0,1,1)_{12}$	$\varphi = 0.412$ $\theta = 0.108$ $\Theta = 0.910$	0.511	*P*
5 Danube	Hofkirchen	$(1,0,1) \times (0,1,1)_{12}$	$\varphi = 0.809$ $\theta = 0.414$ $\Theta = 0.913$	0.451	*F*

P ≡ pass; F ≡ fail

Eq. (7) was applied to generate synthetic flow sequences in which the residual term *at* was drawn from a $N(O,S_a)$ random number generator. The monthly means were substituted for the starting values $X_{t-1}, \ldots X_{t-13}$ so as to test the reproduction of historic statistics. The resulting pattern of generated traces was found to be very sensitive in such a way that an abnormally high or low value for a particular month was passed through the generation from year to year giving a distorted trace and clearly not preserving the historic statistics. Recently Kavvas and Delleur (1975) have convincingly shown that the differencing method of removing cyclicity results in distorted spectral densities and divergence in subsequently generated series thereby destroying the applicability of the model for streamflow generation.

The seasonal-multiplicative model, as originally proposed and developed for short term forecasting of economic time series by Bacon (1965), has been applied to forecasting monthly streamflows by Clarke (1973) and McKerchar and Delleur (1974). Here only the magnitude of the forecasted event, not long term generation, is of interest. The model was briefly tried for a one step ahead forecast (forecast lead time of one month). Fig.4.3 shows the monthly forecast to the flows of the Iller at Kempten in which the seasonal high-low flow periods are followed and the actual forecasted flow is reasonably accurate. When applied to other rivers the results were not so good due to the high degree of stochasticity and variability in the series. In the

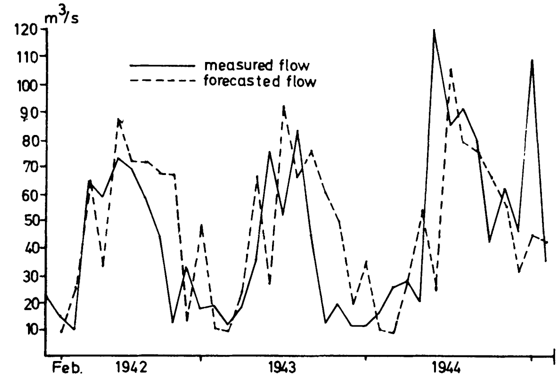


Fig. 4.3. Monthly forecast by the seasonal multiplicative model for the river Iller (Kempten).

analysis stage the alpine catchment of the Iller was found to have the most regular, least stochastic streamflow of all whereas the Danube had the most variable and stochastic flow and failed to be fitted by the model. This would seem to be the limiting factor in the suitability of this model and it is further thought that stochastic models involving autocorrelative relationships between previously occurring streamflow events would in general be unreliable for monthly forecasting and the cause-effect nature of the rainfall-runoff process must be considered in a model reflecting the input-output mechanism.

Conclusions

For the five streamflow series and catchments considered, representative of some of the typical types encountered in the F.R. of Germany, a quantitative classification by means of statistical parameters resulting from the analysis could be made. The small catchments situated in the low-lying areas of the north German plain have moderately cyclic and persistent streamflows. The low rainfall, high evaporative losses and flat landform help to regulate the flows giving slow response, large storage characteristics. In the hilly areas of the central uplands streamflows are less persistent from month to month and are responsive to the high, orographic rainfall conditions and steeper gradients. In the south the rivers flow northwards from the Alps and have distinct characteristics of a regular cyclic nature with persistently high flows in spring and summer due to snowmelt. The steep gradients give very rapid response to the high rainfall with little storage and loss. Large river basins produce very variable streamflows due to the superposition of the characteristics of contributing subcatchments.

Resulting from the time series analysis the generating process can be identified and described by simple stochastic models which are very tractable for streamflow generation, water resources planning and operational purposes. The Thomas-Fiering and the deterministic-stochastic models, both involving parameters associated with the annual cycle, were found to be suitable for generating sequences in which short run properties were reproduced. The seasonal multiplicative model did not assume any deterministic structural properties of the series and was found to fit in all but one case. The minimum of parameters involved offered a concise description of the streamflow process for a given catchment but the model's practical application is limited since generated sequences did not reproduce the historic statistics and the model could accurately forecast monthly streamflows only in the exceptional case of a river with very regular flow characteristics.

References

Bacon, D. W. (1965) Seasonal time series. Ph. D. Thesis, University of Wisconsin, Madison.

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

Clarke, R. T. (1973) Mathematical models in hydrology. Irrigation and Drainage Paper 19, F.A.O. Rome.

D.F.G./I.H.D. (1968) *I.H.D. yearbook of the Federal Republic of Germany* (English/German). Bundesanstalt für Gewässerkunde, Koblenz.

D.F.G./I.H.D. (1972) *I.H.D., German contribution in the years 1965-1969.* English edition, Deutsche Forschungsgemeinschaft, Bonn.

Fiering, M. B. (1967) *Streamflow synthesis.* Harvard University Press, Cambridge/Mass.

Kavvas, M.L., and Delleur, J.W. (1975) Removal of periodicities by differencing and monthly mean subtraction *Journal of Hydrology, 26,* pp. 335-353.

Maass, A., et al (1962) *Design of water resources systems.* Harvard University Press, Cambridge/Mass.

Matalas, N.C. (1967) Mathematical assessment of synthetic hydrology. *Water Resources Research 3,* pp. 937-945.

Mc Kerchar, A. I., and Delleur, J.W. (1974) Application of seasonal parametric linear stochastic models to monthly flow data. *Water Resources Research 10,* pp. 246-254.

Roesener, L. A., and Yevjevich, V. M. (1972) Mathematical models for time series analysis of monthly precipitation and monthly runoff. Hydrology Paper No. 15, Colorado State University, Fort Collins, Colorado.

Rohde, F. G., and Robinson, S. A. (1975) Zwischenbericht 2, DFG Ro 391/2, R.W.T.H. Aachen, Institut für Wasserbau und Wasserwirtschaft.

First received 29 September, 1975

Revised version received 18 November, 1975.

Address:

Stephen A. Robinson, Research asst.,

Fritz G. Rohde, Professor of Water and Energy Resources,

Institut für Wasserbau und Wasserwirtschaft,

Rheinisch-Westfälische Technische Hochschule Aachen

D 51 Aachen, Mies-van-der-Rohe-Strasse,

Germany.