Journal of Hydrology (N.Z.) Vol. 23, No. 1, 1984

TIME-SERIES MODELS FOR THE PREDICTION OF STREAM FLOW IN A KARST DRAINAGE SYSTEM

A. J. Jakeman[1], M. A. Greenaway[1] and J. N. Jennings[2]

[1] *Centre for Resource and Environmental Studies, Australian National University.*
[2] *Department of Biogeography and Geomorphology, Research School of Pacific Studies, Australian National University.*

ABSTRACT

Time series methods are used to predict values for numerous missing daily flow records from rainfall in the Cave Creek drainage system, Cooleman Plain, New South Wales. Separate models are developed which are statistically valid for fixed portions of the data. Missing flows within a consistently-behaved hydrological period can be predicted from the statistical model, which is identified from other data within the period. The method circumvents difficulties associated with validating a detailed physical model of a complex hydrological/geomorphological system.

By regression with sample chemical analyses the discharge estimates obtained are used to ascertain the temporal distribution of limestone solution, the dominant geomorphic process here.

INTRODUCTION

The high solubility of limestone and other carbonate rocks gives rise to the distinctive geomorphology and hydrology of karst terrain. To understand both landforms and drainage, it is necessary to determine when, where, and at what rates limestone is removed in solution. An important component of such studies is the chemical monitoring of karst drainage. This paper is concerned with the practical difficulties of utilising incomplete records collected during such a study.

Cave Creek and its tributaries drain Cooleman Plain, southeastern New South Wales, Australia, an upland limestone plain in natural grassland at 1100-1500 m, and its surrounding forested ranges of igneous rocks rising up to 500 m higher (Fig. 1). Its discharge and chemistry have been monitored during monthly visits between 1965 and 1969 (Jennings 1972) and subsequently between 1969 and 1977 using automatic instruments as well (Jennings 1983). Discharge is measured at Cave Creek where it leaves Cooleman Plain; at this point Cave Creek drains a catchment of 55.5 km^2. A Sumner-Rimco recorder was mounted over float and well to record river stage, and was also linked to a pluviograph. A second recorder operated another pluviograph and automatic pan evaporimeter at Coolamine Homestead which is 2 km away in the middle of the Plain.

The instrumental records were incomplete. Instrument failure was caused by the severe climate, and human and animal interference; relative

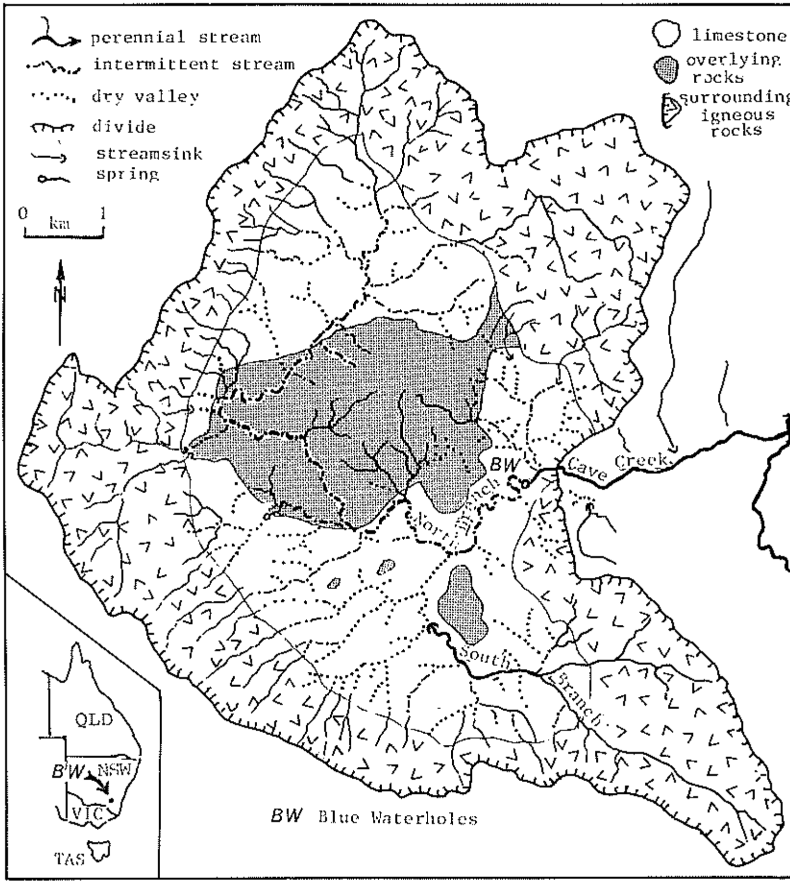


FIG. 1—Drainage and simplified geology of Cooleman Plain, New South Wales.

inaccessibility also made maintenance difficult. Suspect values were removed from the record so that accepted values are regarded as reliable. Totalisers provided check values for pluviograph totals between visits. One daily rainfall record was derived from the Cave Creek pluviograph by filling its gaps using records from the Cooleman Plain pluviograph and from the nearest meteorological station at Tantangara Dam, 18 km to the south in an area of similar relief. The discharge record was not complete for any single hydrological year, the best year being 94% complete. Two years were much worse than the others, only 31 and 17% complete.

Cave Creek was sampled for chemical analysis on 104 occasions which were well distributed seasonally. Total carbonate hardness was calculated from concentrations of Ca^{++} and Mg^{++} ions. Hardness (T) was regressed on discharge at the time of sampling (Q). Subsets of samples from rising and falling limbs of hydrograph events did not yield significantly different regressions, but there were only 6 samples in the rising limb subset. 'Summer' and 'winter' subsets did give significantly different regressions:

December-May	$\log T = 2.79 - \log Q^{0.31}$	$R^2 = 0.67$ $P = 0.001$
June-November	$\log T = 2.45 - \log Q^{0.19}$	$R^2 = 0.49$ $P = 0.001$

These regressions were used to calculate daily hardnesses and from them daily carbonate loads as a basis for analysing monthly and annual variations in the rate of limestone erosion. The chemical data available were inadequate to analyse event hydrochemistry.

It was desirable to make best estimates for missing sections of the discharge record by predicting the likely response of the catchment to rainfall. For accuracy and convenience, estimates of daily discharge were needed, even though the carbonate loads calculated from these estimates would be used only on a coarser time scale.

THE HYDROLOGICAL SYSTEM

Limestone catchments behave differently from normal surface-stream catchments because of underground drainage. At Cooleman Plain, limestone outcrops over 29% of the area but the topography and structure are such that it affects all drainage (Fig. 1). A large number of small streams from the surrounding igneous ranges, and from a central area of overlying impervious rocks, flow for varying distances onto the limestone where they lose all or part of their water underground. No drainage from the southern half of the Plain ever reaches the gauging station entirely as surface flow; subsurface flow resurges at the Blue Waterholes, a group of springs 300 m upstream of the discharge recorder. From the northern part of the Plain, there is winter and spring flow all the way down a surface channel to the Blue Waterholes for part of all but drought years. Underground flow from the northern Plain feeds the Blue Waterholes at all times. Between 1965 and 1969, the surface flow down North Branch of Cave Creek occurred for 35% of the time and was 20% of the total output. This North Branch flow is only partly surface runoff; permanent and temporary limestone springs provide a significant part of it. Substantially different routing patterns may come into play as a result of varying antecedent conditions.

The hydrograph of Cave Creek is more complex than that of normal rivers. Cave streamflow is significantly slower than surface streamflow. In water-filled cave sections, however, a hydraulic ram effect comes into play; increased input into such a section is transmitted almost immediately to the bottom end. During rising stage, caves fill with water and feed adjoining fissures of substantial volume. This water is returned to the cave slowly during falling stage. Percolation water, passing more or less vertically through the limestone to be gathered into cave streamflow or to join streams fed from surrounding impervious rocks, varies greatly in its rates of movement but generally moves several magnitudes more slowly than cave streams.

Thus, varying proportions of surface and underground flow reach the gauging station, making prediction of runoff from basin rainfall very difficult. Index functions of antecedent precipitation did not appear adequate to model the rainfall-runoff relationships.

MODELLING PHILOSOPHY

Rainfall and discharge data for the years 1969 to 1977 (see Figures 2-6) confirm that the Cave Creek drainage system has a variable rainfall-runoff relationship. Perhaps the simplest example of this is given by the steady gain of the system. This property is the relative amount of flow produced per unit of rainfall input, and was found to vary from around .12 for the December to May period of 1976 to as much as .92 for the December to May period of 1971 (see Table 1). Clearly such variations are the result of more complex storage within the system than just soil moisture retention.

The physical behaviour of the system was deemed too complex to use a mechanistic model. A statistical approach was therefore taken, incorporating in the model the most important mechanism known to be operating — an exponential-type decay in flow following rainfall events. It was hypothesised that the remaining mechanisms contributing to the behaviour of the system could be described adequately by black-box parameters. The rate of decay of the hydrograph and the values of these additional parameters could then be estimated statistically from the data. The advantage of this approach is that it can identify time intervals of consistent hydrologic behaviour. Different periods in the data set produced different statistical models with parameter values relevant to the period of interest. The basic structure of these models is similar; it is the value of the parameters in the structure which vary. The modelling exercise yields criteria for judging the validity of each model.

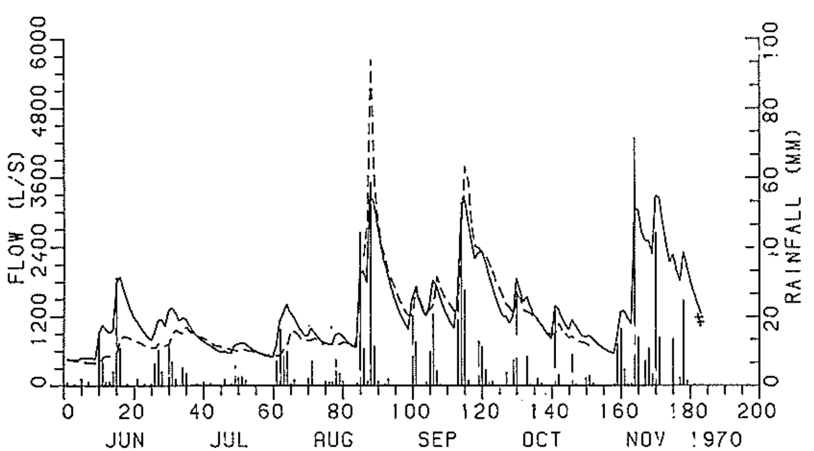


FIG. 2—Daily rainfall (spikes), predicted flow (continuous line) and measured flow (dashed or crosses) for June to November 1970.

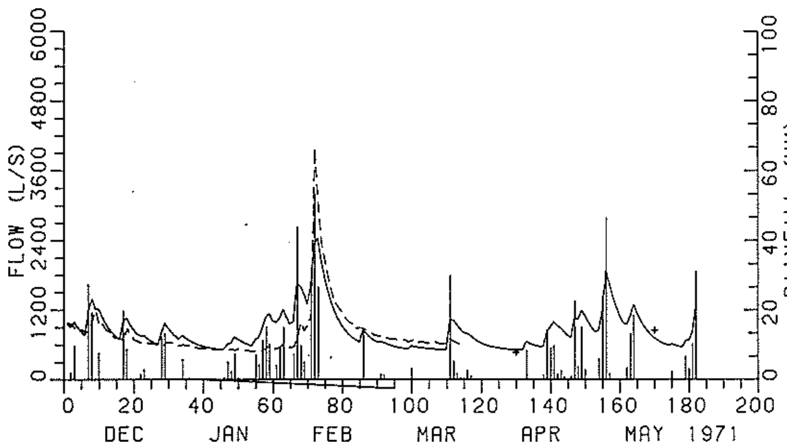


FIG. 3—Daily rainfall (spikes), predicted flow (continuous line) and measured flow (dashed or crosses) for December to May 1971.

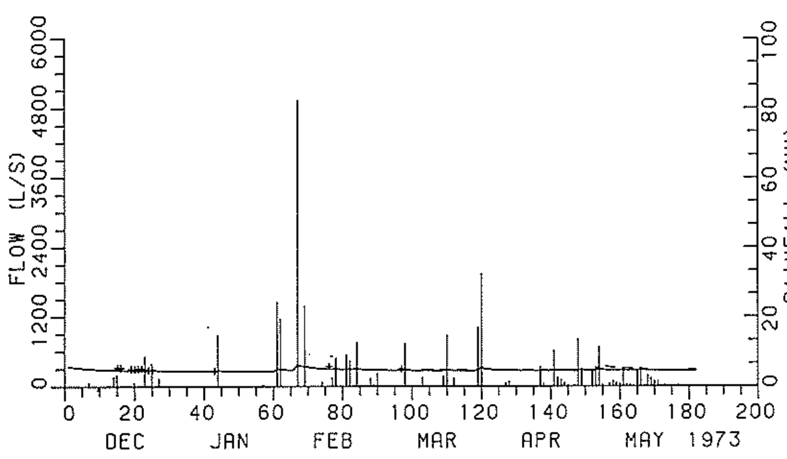


FIG. 4—Daily rainfall (spikes), predicted flow (continuous line) and measured flow (dashed or crosses) for December to May 1973.

THE RAINFALL — FLOW MODEL

Transfer-function models are described in Young (1978), Jakeman (1981), Mahendrarajah et al. (1982) and the references therein. Useful discussions can also be found in Caprihan and Neto (1977), while O'Connell and Clarke (1981) review the application of transfer function models, and a range of

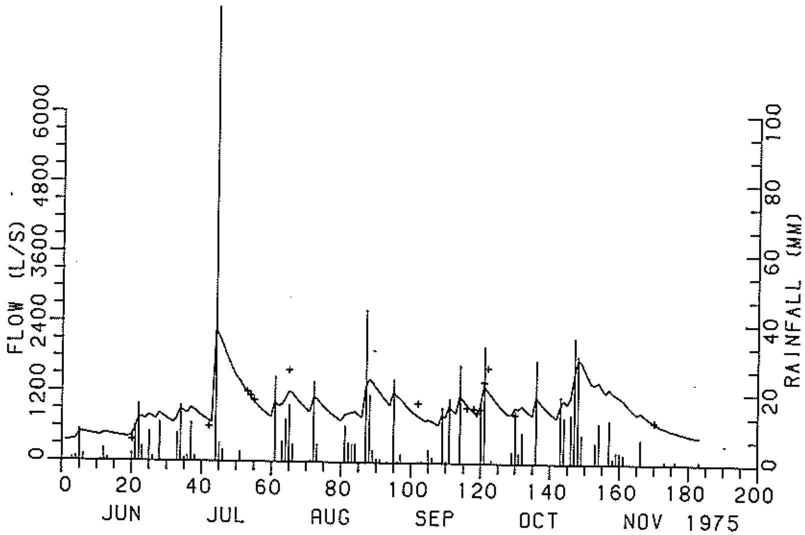


FIG. 5—Daily rainfall (spikes), predicted flow (continuous line) and measured flow (dashed or crosses) for December to May 1975.

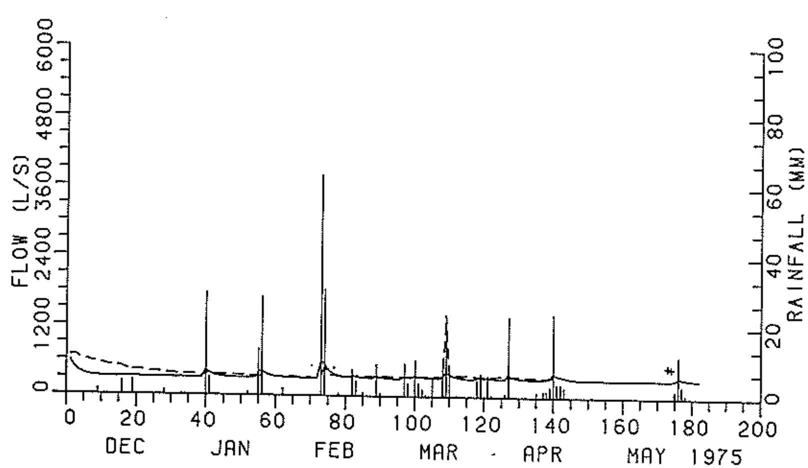


FIG. 6—Daily rainfall (spikes), predicted flow (continuous line) and measured flow (dashed or crosses) for June to November 1975.

TABLE 1—Time series model parameters used in equation (1) to predict flow and the data period over which each model was identified. The time constant and steady state gain model properties and standard errors on model parameters are also shown.

6 MONTH PERIOD	FLOW GAPS (DAY NUMBERS)	PARAMETER (STANDARD ERROR) a_1		b_0		b_1		b_2		BASE FLOW B	TIME CONSTANT (days)	STEADY STATE GAIN	MODEL IDENTIFICATION PERIOD (DAY NUMBERS: PERIOD)	ASSOCIATED R^2 VALUE
DM 1969*	1-112	-0.80	(.06)	17.56	(3.11)					300	4.5	.60	1-114: DM 1971	0.59
JN 1969	none	-0.90	(.03)	12.54	(1.56)					350	9.5	.74	1-151: JN 1969	0.68
DM 1970	135-145	-0.99	(.07)	0.08	(.32)					344	99.5	.55	1-134: DM 1970	0.24
JN 1970	152-181	-0.85	(.03)	29.27	(3.12)					390	6.2	.91	1-151: JN 1970	0.79
DM 1971	115-184	-0.80	(.06)	17.56	(3.11)					500	4.5.	92	1-114: DM 1971	0.59
JN 1971	65-98	-0.94	(.01)	1.22	(.72)	5.58	(.74)	-1.59	(.74)	400	16.2	.76	3- 64: JN 1971	0.96
DM 1972	98-122	-0.96	(.02)	0.06	(.23)	2.05	(.26)	-1.76	(.25)	349	24.5	.56	123-183: DM 1972	0.85
JN 1972	98-111	-0.95	(.03)	-0.19	(.69)	1.80	(.62)	-0.78	(.70)	350	19.5	.58	112-183: JN 1972	0.91
DM 1973	2-102) 149-151) 169-179)	-0.91	(.06)	1.20	(.19)	-0.18	(.22)	-0.25	(.22)	255	10.6	.41	103-148: DM 1973	0.58
JN 1973*	1-120	-0.85	(.03)	29.27	(3.12)					277	6.2	.77	1-151: JN 1970	0.79
	121-174	-0.80	(.06)	17.56	(3.11)					277	4.5	.57	1-114: DM 1971	0.59
DM 1974	2-80) 159-173)	-0.92	(.07)	4.63	(2.14)					348	12.0	.63	81-158: DM 1974	0.70
JN 1974	44-82	-0.94	(.04)	35.04	(3.77)	1.54	(4.59)	-19.68	(4.60)	0	16.2	.44	83-183: JN 1974	0.85
DM 1975	55-63) 141-182)	-0.67	(.00)	4.32	(1.50)	-0.74	(1.56)			299	2.5	.49	64-140: DM 1975	0.05
JN 1975*	1-183	-0.90	(.03)	12.54	(1.56)					350	9.5	.74	1-151: JN 1969	0.68
DM 1976	1-164	-0.99	(.05)	0.67	(.22)	0.10	(.22)			0	99.5	.12	165-183: DM 1976	0.11
JN 1976*	12-24	-0.99	(.05)	0.67	(.22)	0.10	(.22)			0	99.5	.12	165-183: DM 1976	0.11
	111-150	-0.94	(.04)	35.04	(3.77)	1.54	(4.59)	-19.68	(4.60)	0	16.2	.44	83-183: JN 1974	0.85
DM 1977	none	-0.98	(.01)	1.08	(.23)	0.30	(.23)	-0.37	(.23)	0	49.5	.08	1-145: DM 1977	0.68

DM — December-May
JN — June-November
* — Period for which no satisfactory model could be built from available data.

other models, to rainfall-runoff prediction.

The form is a linear systems input (rainfall) - output (flow) model

$$x_k = -a_1x_{k-1} - a_2x_{k-2} - \ldots - a_nx_{k-n} + b_0u_k + b_1u_{k-1} + \ldots + b_mu_{k-m} \qquad (1)$$

which states that flow x_k on day k, say, is a linear combination of flows on the n preceding days, and of rainfall u_k on the current day and the past m days. This is an alternative to the unit hydrograph form derived from the convolution integral (see, for example, Broome and Spigel, 1982, and the references therein) since the above equation is a rational approximation of

$$x_k = g_0u_k + g_1\,u_{k-1} + g_2\,u_{k-2} + \ldots \qquad (2)$$

where the coefficients g_i ($i = 0, 1, 2, \ldots$) are ordinates of the unit hydrograph.

The linear systems model (1) has distinct advantages. The number of parameters ($m + n + 1$) to be estimated is always much smaller than the number of non-zero g coefficients in (2). In modelling the discharge of Cave Creek it was found that n=1 always, and m varied from 0 to 2. The lower number of parameters obtained using (1) tends to yield lower variance on the statistical parameters and hence on the model predictions.

Error in the data is assumed to be additive at the output of the hydrological system, which is the flow (see Jakeman and Young, 1983). In this case, x_k in (1) becomes an unobserved or hypothetical noise-free output and equation (1) is complemented with the so-called observation equation

$$y_k = x_k + n_k \qquad (3)$$

where, at time k, y_k is the observed output or flow, and the error n_k represents the contribution of all outside disturbances and lack of fit of the model. For the rainfall-runoff models described in this paper, where a complex system is described by a simple linear model, it is necessary that n_k can possess a non-trivial structure. In this case, one of the most robust methods for identifying the orders n and m in (1) and estimating the parameters a_i ($i = 1 \ldots, n$) and b_j ($j = 0, 1, \ldots m$) is the instrumental variable technique (Young, 1974), which allows the error to be of very general character as long as it possesses no systematic bias. The method has been programmed for computer and is described in Jakeman et al. (1982).

Two properties of the linear systems model (1) are *steady state gain* and *system time constants*. The steady state gain is given by

$$\sum_{i=0}^{m} b_i / (1 + \sum_{i=1}^{n} a_i)$$

whereas the time constants are, to a good approximation, given by $-1/\log_e(-a_i)$, ($i = 1, \ldots, n$) (Caprihan and Neto, 1977).

With n = 1 in all models identified in this paper, there is only one time constant per model — the time taken for the peak flow response to an impulse of rain to decay to about one-third of its peak value. It is a measure of the period of influence of previous rainfall. The time constant was commonly around 5 to 10 days, and varied from indicating a quick and precise hydrological response to a slow and heavily damped one. The time constant was as low as 2.5 days for the December to May 1975 period and as high as 50 to 100 days for several periods.

The instrumental variable technique yields an estimate of the covariance matrix of the parameters which provides a measure of the reliability of a model for prediction. If asymptotically efficient estimates are required, a computationally more complex technique known as refined instrumental variables can be used. This yields an estimate of the full probability distribution of the parameters a_i ($i=1,\ldots,n$) and b_j ($j=0,1,\ldots,m$) (Young and Jakeman, 1979). The simpler basic instrumental variable method was used here. Consequently, we report only the estimated mean and standard errors of the model parameters (Table 1). The coefficient of determination, R^2 (Table 1) is a measure of the degree of explanation of the output (flow here) data used to estimate an original model. It is normalised to have a maximum value of unity when, given the input, a model output perfectly reproduces the measured output data. It is defined here as

$$R^2 = 1 - \Sigma\,(y_k - \hat{x}_k)^2 / \Sigma\,(y_k - \bar{y})^2$$

where $\hat{x}_k$ is the instrumental variable method's estimate of x_k, obtained by inserting the parameter estimates in (1), and $\bar{y}$ is the mean of the observed runoff.

RESULTS

Satisfactory and consistent statistical models of the form (1) could be estimated for seasonal divisions of individual years. A useful seasonal division is December to May (Dec-May) and June to November (Jun-Nov). June to November has higher precipitation and runoff; December to May has reduced precipitation and runoff. Pan evaporimeter records, though extremely broken, did show a seasonal pattern. Evaporation is significant in the latter part of the wetter season, but its effects predominate only in the drier season. Thornthwaite water-balance calculations also supported this partition of the year.

Enough seasonal data were available for all periods to build separate models for each (except for Jun-Nov 1973 and Jun-Nov 1975). However, two of those identified models were unsatisfactory. Therefore, four periods required flow prediction by other means.

Table 1 displays details of the models used to predict flow gaps in each seasonal data set. Figures 2 to 6 present plots of the raw data and the model predictions for a representative selection of rainfall and flow regimes. There are some days of recorded flows which occurred within flow periods requiring prediction. Since these small amounts of data were not used in the modelling, they provide a check on the performance of the models.

The rainfall and flow data sets used for prediction are listed in Table 1. There were four periods for which no satisfactory model could be built from data within those periods. Predictions for those periods were obtained by selecting an appropriate model from the suite of 13 models identified for the other periods. The model chosen was that which best reproduced flow characteristics indicated by available flow data.

The coefficients of determination (Table 1) associated with the identified models are generally quite adequate. The degree of explanation of the flow data ranges from 59 to 96 per cent for all but three of the models. Where the explanation is below 25 per cent, the models indicated were accepted because they predict useful trends. Much of the flow data not explained by these models is difficult to interpret. For example, the discrepancy between predicted flow and the measured flow of 1418 *l*/s for 19 March 75 (Figure 5), almost totally accounts for the low R^2 value for the Dec-May 1975 model.

A seasonal model for predicting unrecorded flows within the same season is useful if, in addition to having high correlation coefficients, it also has tight standard errors on its parameters. The standard error on the a_1 parameter, which describes the response time of the system to rainfall, is always less than 10 per cent of the parameter value.

In the model development, it was usually necessary to impose a base flow term on (1) and (3) so that (3) becomes

$$y_k = x_k + n_k + B \qquad (k = 1, 2, \ldots)$$

In the absence of any identifiable trend in base flow, B was assigned a constant value within a seasonal period. Its value for a particular data set was determined by estimating models as B varied downwards from the minimum flow. B was chosen as the value in that model which provided a compromise between good explanation of the flow (R^2 value) and low standard errors on the parameters.

DISCUSSION

The steady-state gains (Table 1) are those obtained after base flow has been added to the model output, with flow recalibrated from litres per second to millimetres per day, taking into account catchment area. Both the steady-state gain and time constant values vary remarkably over the nine year study period. Since steady-state gain is the relative amount of flow produced per unit input of rainfall, it is a measure of the current retention properties of the surface and underground system. A low gain implies that rainfall is being stored within the system, whereas a high value indicates that much of the rainfall is generating flow directly.

Although this suggests that low gains should follow seasons of predominantly low rainfall, this is not always the case (Table 1). In the period Dec-May, 1975, a medium level gain of .49 is obtained for the identified model when the rainfall for the preceding Jun-Nov 1974 period is one of the heaviest recorded. Clearly, the retention processes are complex.

For a linear transfer function model of the form (1), the time constant is a measure of the duration of response of the system to rainfall. Often

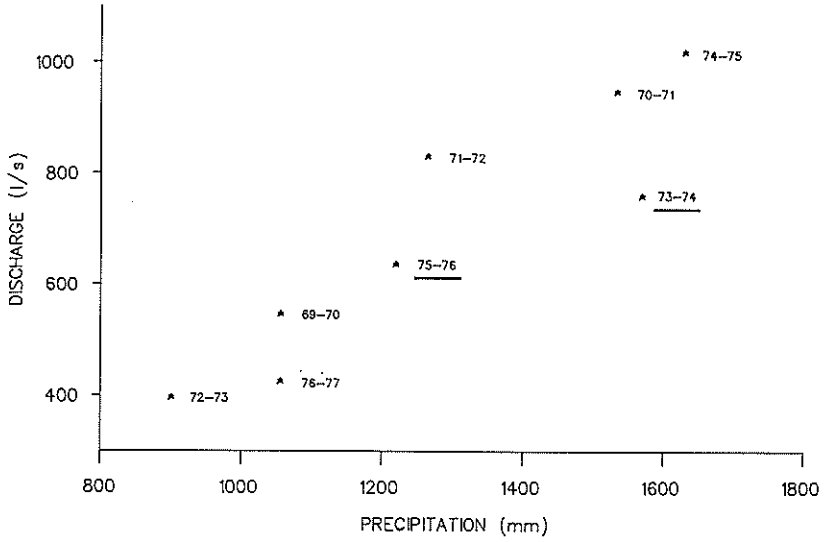
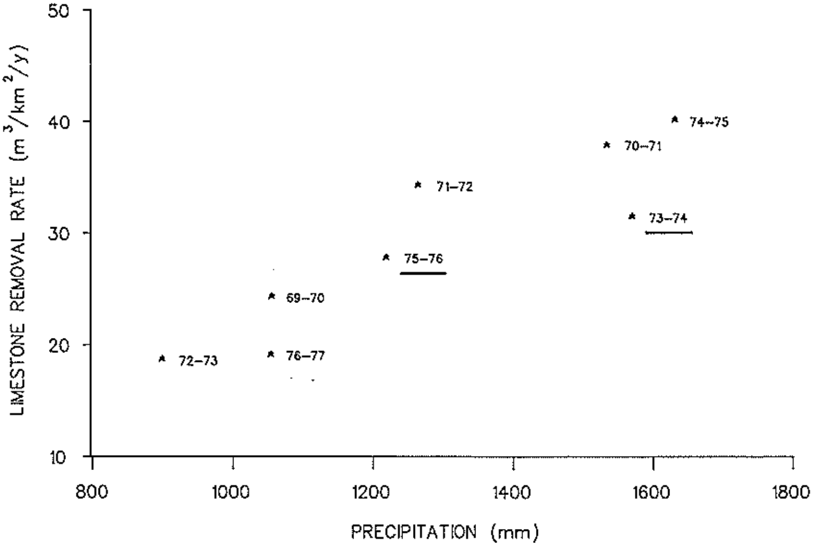


FIG. 7—(Above). Mean annual discharge (including estimates where necessary) plotted against annual precipitation; (below) annual limestone removal rate against annual precipitation.



rainfall affected flow for about ten days (Table 1) but for six of the periods the response was about twice as fast. On the other hand, flow rate can be influenced by previous rainfall for periods of the order of 100 days; here it is likely that the low frequency characteristics of the system are being observed. The steady-state gain values associated with long time constants indicate that the retention capacity of the system is high, thereby yielding a slow and steady response of flow to rainfall. The gains are particularly low for Dec-May 1976, Jun-Nov 1976 and Dec-May 1977, which are periods with long time constants.

Mangin (1974-5) measured drainage from karst springs in the Pyrenees which are as complex as Cave Creek. Despite overall agreement between discharge and precipitation, the relationship within any hydrological year is poor, due to varying storages. He states that a drought year can affect the next five years of spring behaviour.

The number of models that had to be employed in this study reflects the complexity of the system and our present lack of a satisfactory deterministic theory. That the approach has been useful is shown by comparing the results obtained over the eight hydrological years, which have varying degrees of dependence on estimates of discharge.

Fig. 7 displays a plot of mean annual discharge (with gaps filled by estimates) against precipitation for these years. It also shows limestone carbonate load against precipitation. That estimates for years with poor records of discharge do not depart from a straight line plot more than those for years with good or tolerable records is evidence that the modelling used to fill in the record has been satisfactory. Departure from the straight line plots is probably due to poorly representative precipitation measures. Using regressions of total hardness on discharge to calculate limestone erosion rate brings additional error into play but this has not incurred greater departure from a straight line than in the case of discharge on its own.

CONCLUSION

There are many sources of error in the determination of solution regimes and removal rates for limestone areas. Nevertheless, data in this field are so few that provided their level of reliability is recognised it is desirable to amplify them. The present paper has shown that much information can be derived from incomplete field data by the use of transfer function models.

The complex hydrological behaviour of karst drainage is paralleled by the complexity of time series modelling of the rainfall-runoff relationships. Filling the gaps in the discharge record provided monthly load figures for a fuller picture of the seasonal regime of limestone erosion. The previous preliminary study of 4 years (Jennings 1972) had revealed great variation from year to year in limestone removal. By furnishing a longer record the modelling improved the long term mean rate, although it did not extend the range of the annual values.

ACKNOWLEDGEMENT

The authors would like to thank the referees for their very helpful comments on an earlier version of this paper.

REFERENCES

Broome, P.; Spigel, R. H. 1982: A linear model of storm runoff from some urban catchments in New Zealand. *Journal of Hydrology (N.Z.) 21*: 13-33.

Caprihan, A.; Neto, A. G. 1977: Optimum modeling of vascular beds using indicator — dilution measurements. *IEEE Transactions on Biomedical Engineering*. BME-*24*: 219-226.

Jakeman, A. J. 1981: The role of recursive time series analysis in resource management. *Proceedings, Resource Management Conference*, Civil and Systems Engineering Dept., James Cook University, July 13-17 (Appears as *Resource Management: Issues, Techniques, Applications*, 1983, James Cook University). See also CRES Report AS/R47, Australian National University.

Jakeman, A. J.; Young, P. C. 1983: Advanced methods of recursive time series analysis. *International Journal of Control 37*: 1291-1310.

Jakeman, A. J.; Young, P. C.; Bayes, A. J. 1982: A general program for recursive time series analysis. *Proceedings 6th IFAC Symposium on Identification and System Parameter Estimation*, Washington DC, June 7-11.

Jennings, J. N. 1972: The Blue Waterholes, Cooleman Plain, N.S.W., and the problem of karst denudation rate determination. *Transactions Cave Research Group G.B. 14*: 109-117.

Jennings, J. N. 1983: Further studies at the Blue Waterholes, Cooleman Plain, N.S.W., 1969-77. *Helictite 21*: 3-20, 35-54.

Mahendrarajah, S.; Jakeman, A. J.; Young, P. C. 1982: Stochastic time series analysis and the village dam management problem in Sri Lanka. *Proceedings First International Assembly, International Association of Hydrological Sciences*, Exeter, July 19-30, pp. 251-264.

Mangin, A. 1974-5: Contribution a l'e´tude hydrodynamique des aquiferes. *Annals of Speleology 29*:283-332, 495-61; 30: 21-124.

O'Connell, P. E.; Clarke, R. T. 1981: Adaptive hydrological forecasting — a review. *Hydrological Sciences Bulletin 26*: 179-205.

Young, P. C. 1974; Recursive approaches to time series analysis. *Bulletin of the Institute of Mathematics and its Applications 10*: 209-224.

Young, P. C. 1978: A general theory of modelling for badly defined systems. In: *The Modelling of Environmental Systems*, G. C. Vansteenkiste (ed.), North Holland, Amsterdam, pp 103-135.

Young, P. C.: Jakeman, A. J. 1979: Refined instrumental variable methods of recursive time series analysis. Part I: Single input-single output systems. *International Journal of Control 29*: 1-30.