INSTITUTE OF HYDROLOGY

REPORT No 29

October 1975

SUBCATCHMENT MODELLING FOR DEE RIVER FORECASTING

by

A I McKerchar B.E., Ph.D.

ABSTRACT

A simple model to forecast outflows from five River Dee (North Wales) subcatchments has been calibrated and tested. In the model a catchment is considered as a storage and the outflow from the catchment has a logarithmic relationship with the contents of the storage. Parameters of this relationship and empirical lags were estimated both graphically and with a numerical optimizing routine. A two season year with different parameters for each season was used. Good agreement is demonstrated between observed and calculated hydrographs. Examples of simulated on-line forecasts of outflows assuming deterministic rainfalls are given. For one example a lag two autoregressive scheme appeared suitable for forecasting the errors. Because the catchments respond quickly to rainfall it is suggested that in practical situations most of the forecast error will result from errors in rainfall forecasts.

CONTENTS

<u>FIGURES</u> Page

<u>TABLES</u>

1. INTRODUCTION

The Dee River in North Wales is the centre of a number of British studies for developing methods for analysing and controlling complex water resource systems. These studies include: 1) the measurement and forecasting of precipitation using radar; 2) the forecasting of downstream flows for a large river basin; and 3) the development of control strategies for the efficient utilization of reservoirs. The river forecasting model has been developed for the Water Research Centre (WRC) by the Institute of Hydrology (IH) and is being implemented on a PDP 11/40 computer at the Bala office of the Welsh National Water Development Authority (WNWDA). The model uses automatically supplied observations of rainfalls and upstream flows to forecast downstream hydrographs for 24 hours ahead. A network of telemetering river and reservoir level gauges supply current flows and storage contents. Rainfalls are measured either by telemetering raingauges, or remotely, by radar. This gauging network enables downstream forecasts to be recentred onto the most recent upstream observations.

The river forecasting model, which represents one sort of distributed catchment model, has three major components:

rainfall forecasting over subcatchments;
rainfall/runoff conversion for subcatchments;
hydraulic routing of channel flows.

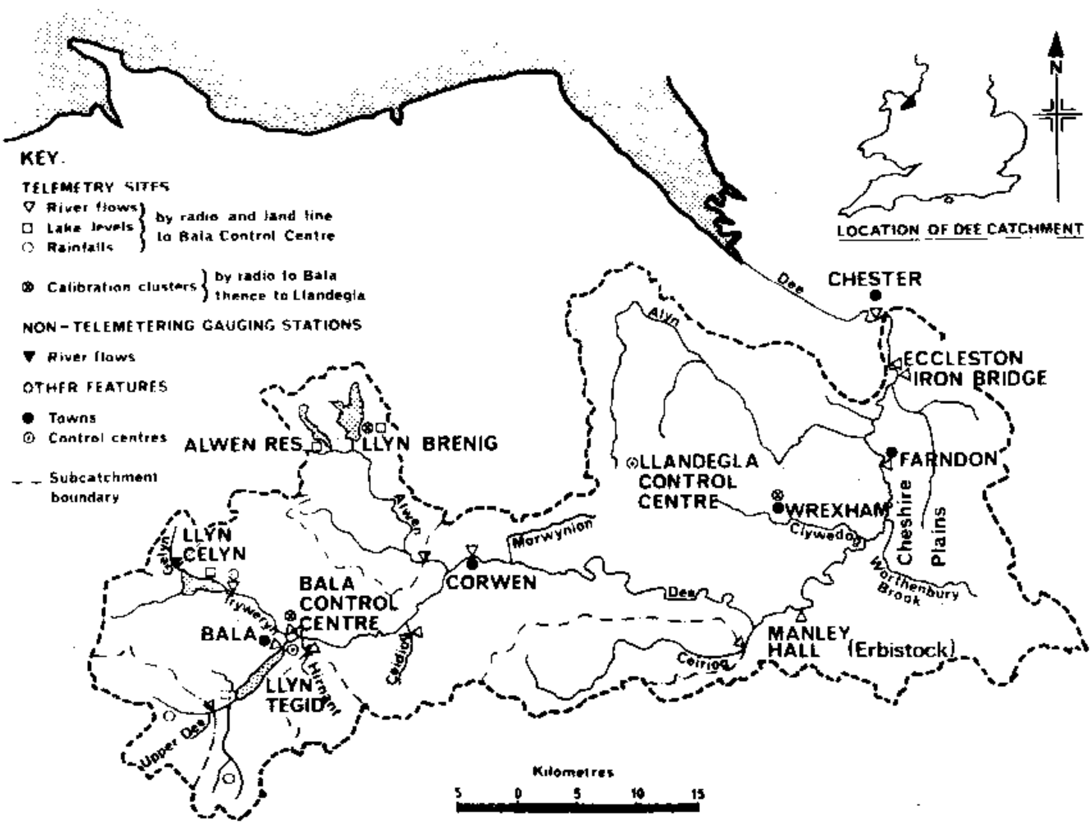
Hydraulic routing is well understood (Price, 1973) so that an upstream hydrograph can, in general, be accurately transformed into a downstream hydrograph. In this particular study a variable parameter diffusion technique has been used.

The second component, and the subject of this report, the rainfall-runoff conversion, is in general of a lesser accuracy than the channel routing. The third component, the rainfall forecasting, is the least well understood and has the lowest accuracy.

The ordering and the degree of understanding of these three components is a reflection of the number of spatial dimensions involved. Routing a hydrograph down a river channel involves an understanding of the movement of water along essentially one major spatial dimension, the distance along the river channel; a subcatchment model requires an understanding of the processes occuring at all points on and below a catchment surface; rainfall forecasting requires an understanding of the physical processes within a three-dimensional atmosphere. In the total flow forecasting model therefore, the accuracy for the first few intervals, being based largely on flows already observed in upstream channels, is expected to be high, but to become fair and then low as the forecast period becomes longer and proportionately more of the forecasted downstream discharge depends upon observations and then forecasts of subcatchment precipitation (Lowing *et al* 1975). The purpose of this report is to describe how the rainfall to runoff conversion was achieved by using a relatively simple subcatchment model; details are given of its application to five subcatchments within the Dee basin with an objective procedure for determining the model parameters.

These five subcatchments cover only 39% of the catchment area above the gauging station for which forecasts are required. Runoff from ungauged subcatchment areas was estimated by multiplying the nearest gauging subcatchment runoff by the ratio of estimated annual runoff volumes. Details of this and the linking of these subcatchment flows to produce downstream hydrographs is described by Lowing *et al*, 1975.

Besides the simplicity of the model, the other major advantage is the ease with which the most recent observations of rainfall and upstream discharge can be incorporated into the model to give new improved forecasts for the future. Many hydrological models are developed to estimate a sequence of flows that resulted from historical sequences of rainfall, and the 'real-time' use of this model needs to be emphasised.



<u>Figure 1:</u> Map of River Dee Catchment

2. THE SUBCATCHMENTS

The location of the River Dee catchment in North Wales is shown on the inset map in Figure 1. The river rises in the hill country of North Wales and flows east and north to its estuary below Chester. The locations of an extensive telemetry network are shown on the main map in Figure 1. River and reservoir levels, and observed rainfalls at six gauges, are transmitted to the Bala Control Centre where they provide a comprehensive picture of the hydrological state of the catchment at any time. The weather radar provides an alternative to the telemetering raingauges for estimating rainfalls over subcatchment areas.

Manley Hall is a flow gauging station sited on the Dee River where it emerges from the hill country and meanders northwards over the Cheshire Plains. It is the first of the downstream gauging sites for which forecasts of flows 24 hours ahead are required. It is also the lowest point on the river for which satisfactory full range river gauging has been achieved at a single site. Five subcatchments above this station are identified in Figure 1 and listed in Table 1 with the hydrometric network number and the catchment area. The rainfall-runoff modelling is required for these five subcatchments.

The selection of these five catchments was dictated by firstly, the availability of a historical flow record and secondly, that the record be unaffected by upstream controls. Note that Figure 1 shows two reservoirs on the Alwen subcatchment; one of these, the Brenig Reservoir, is currently under construction (1975) and has not affected the historic data prior to the commencement of construction; the other, the Alwen Reservoir, is a direct supply reservoir from which water is diverted through a pipeline out of the Dee catchment. Normally small steady compensation flows are released downstream from this reservoir but occasionally when it overflows an unknown water volume enters the lower catchment. The data from such periods is not used in the modelling exercise.

These are small hill country catchments with thin soil cover and steep channel slopes and, not surprisingly, fast responses to rainfall. Soils on the upper parts of catchments are peaty; these drain slowly and sustain low but continuous flows through dry periods. The annual total rainfall is well distributed throughout the year but there is a tendency for greater falls in the winter. Nevertheless, severe storms leading to downstream flooding can occur at any time. Snowfall is erratic; snow-pack can develop in exceptionally severe winters and can lead to snow-melt flooding. This is recognized in this work although the modelling will not include snowmelt explicitly.

Rainfall for the period July 1972 to June 1974 was obtained from the Dee Weather Radar Project. The data were in the form of 15 minute totals from 67 Plessey recording raingauges set out over the Dee catchment above Manley Hall. Half-hourly subcatchment totals for the five subcatchments in Table 1 were calculated as the arithmetic mean of half-hourly catches for those gauges lying within the particular subcatchment. Half-hourly flow data for several years were obtained by processing Fischer and Porter punched tapes held by the Water Data Unit (WDU). Daily pan evaporation data for the period August 1969 to March 1975 and potential evaporation data for the period January 1972 to December 1973 were supplied by the Dee and Clwyd Rivers Division (DCRD) of the WNWDA. The concurrent rainfall and runoff data for each subcatchment for the period July 1972 to June 1974 were plotted and after visual checking for gross errors were stored on magnetic tape. Appendix 1 describes the archiving of the rainfall and flow on magnetic tapes at IH.

Table 1: Dee Subcatchments

Hydrometric Network No.	River Name	Station Name	Catchment Area (sq kms)	Months not included in the data From July 72 to June 74
1. 67/05	Ceiriog	Brynkinalt Weir	113.7	July ... Sept, Dec 72 Jan, May ... Aug 73 Apr 74
2. 67/06	Alwen	Druid	160.0*	
3. 67/10	Gelyn	Cynefail	13.1	July, Aug 72 Aug ... Dec 73
4. 67/13	Hirnant	Plas Rhiwaedog	33.9	
5. 67/18	Upper Dee (Dyfrdwy)	New Inn	53.9	July, Aug 72

* excluding Alwen Reservoir Catchment

3. THE SUBCATCHMENT MODEL

Jamieson and Wilkinson (1972) have postulated a *semi-distributed* conceptual model for representing subcatchments of the Dee River. This model has one, two or three hydrological response zones, each having surface and subsurface storages but a common groundwater storage. In its distributed form up to 23 parameters need to be determined. Although this model is hydrologically more realistic than many alternatives, its level of complexity and the associated problems of calibration suggested that simpler alternatives should be considered. Only limited time was available in which to undertake this work and with only two years of subcatchment data for calibration, and the requirement of a model for short-term forecasting rather than for record reconstruction, a very much simpler alternative was considered. This model has already been applied to several Dee River subcatchments (Lambert 1969, 1972) and was known to perform satisfactorily.

It assumes that the flow q_t at time t is related uniquely to the quantity of water stored within the catchment area as groundwater, soil moisture and surface water, collectively termed catchment storage. Together with continuity for the storage, this unique relationship leads to a closed form of solution which relates q_t to q_{t-1} and the rainfall total for the interval t-1 through t. This recursive form is ideal for the real-time operational forecasting.

The relation between storage and outflow is derived as follows: consider the catchment storage to be at an arbitrarily defined state S_o at time t = 0 when the flow is q_o. Define Q_{01} as the total outflow flow time t = 0 to t = 1: that is

$$Q_{01} = \int_0^1 q_t \, dt,$$

and E_{01} as the loss through evaporation and transpiration losses from time t = 0 through t = 1. Then, in recession, the storage is depleted to a new state

$S_1 = S_o - Q_{01} - E_{01}$ corresponding to the outflow q_1

Further,

$S_2 = S_1 - Q_{12} - E_{12}$ corresponds to the outflow q_2

and in general,

$S_i = S_{i-1} - Q_{i-1,i} - E_{i-1,i}$ corresponds to the outflow q_i.

Lambert suggested that the relation between the points S_i and q_i could be approximated by one or several curves of the form

$$S = k \log_e q + c \qquad \dots (1)$$

or $$S = K q + d \qquad \dots (2)$$

The form of the function and the values of k or K appropriate for different ranges of q could be determined by analysing either a series of individual recessions, or a master recession curve for winter months during which the interception and evaporation losses were expected to be low. A crude estimation of the losses through the period of analysis could be made by estimating the total loss for the period and assuming that a mean value E was lost from the storage at each interval.

Log-linear, rather than linear, S vs q relationships generally provided better fits to the observed data and the development proceeds assuming this to be the case. Differentiating

Eqn (1); $$\frac{dS}{dq} = \frac{k}{q} \qquad \dots (3)$$

the continuity equation in continuous form is

$$\frac{dS}{dt} = -q + r - e \qquad \dots (4)$$

where r and e are the rainfall and loss rates at time t.

Combining (3) and (4) to eliminate dS,

$$\frac{dq}{dt} = -\frac{q}{k}(q - r + e) \qquad \dots (5)$$

and integrating (5) over the time period t = 1 through t = 2, we arrive at

$$\int_{q_1}^{q_2} \frac{dq}{q(q-r+e)} = \int_1^2 - \frac{dt}{k} \qquad \dots (6)$$

Although demonstrably untrue, we assume r-e to be constant over the period of integration. The effect of this approximation may be minimal however, because of the dampening effect of the catchment on the short period fluctuations in the inputs. Hence, for r - e = 0, Eqn (6) becomes

$$\int_{q_1}^{q_2} \frac{dq}{q^2} = \int_1^2 \frac{dt}{k}$$

or

$$q_2 = \frac{q_1}{1 + \frac{q_1 . \Delta t}{k}} \qquad \dots (7)$$

4. FITTING THE MODEL

4.1 Graphical Estimation of Parameters:

To gain familiarity with the model, estimates of k and L were made for the Ceiriog and the Upper Dee, these being the two catchments reported upon by Lambert (1969, 1972). Master recession curves were derived from these catchments for a "winter" season by taking events from the months November through to April. Through this period the mean daily open pan evaporation for Bala, a central point in the catchment, was 0.86 mms, from which the daily potential evapotranspiration is estimated as 0.69 mms using a pan factor of 0.8. This potential estimate is used as a crude estimator of the loss $E_{i,i+1}$. Figures 2 and 3 show the plot of the S^* and S vs $\log_e q$ for the Ceiriog and Upper Dee respectively. The plotted points are the S_i^* values where $S_i^* = S_{i-1,i}^* - Q_{i-1,i}$ and the subtraction of $E_{i-1,i}$ is done on the graphs to give the curve of S vs q. Both of these figures show support for the empirical relationship of $S = k \log_e q + c$.

For the Ceiriog one straight line with k = 22.7 and L = 4 hours appears appropriate. This compares with a range of k values between 10 and 25 and a lag of 2 hours that Lambert reported for a number of individual recessions. The value obtained for the master recession, drawn as an envelope about the flattest recessions, is close to Lambert's maximum of 25, which presumably was the flattest individual recession in his analysis. In Lambert's work a value of k = 14 was adopted as this gave a better estimate of the peak flows; this was considered more important on the Ceiriog than achieving a good fit of the whole computed hydrograph.

For the Upper Dee at New Inn, Lambert gave four segments for the storage-outflow function, three log-linear segments followed by a linear segment for higher q values. The individual k and the threshold q values are given in Table 2, and a lag of 2 hours was recommended. The slope of the plot in Figure 3 agrees well with the k values for the lower ranges of flows in Table 2; Figure 3 does not cover the upper flow range as the historic flows are known to be unreliable due to overbank flooding around the gauging site at flows greater than about 40 cumecs (2.67 mm/hr).

Table 2: Lambert's storage-outflow function for the Upper Dee

Flow range (mms/hr)	Function Type	k(K)
< 0.15	log-linear	13.1
0.15 to 0.30	log-linear	5.8
0.30 to 0.584	log-linear	3.7
> 0.584	linear	4.62

The plot in Figure 3 suggests that the storage outflow function should be log-linear with at least 2 k values above and a threshold flow of about 0.25 mm/hr.

The questions raised by this section are: 1) how well does this model perform and 2) whether the storage outflow curve derived from analysis of the master recession curve gives the best value for k. Evaluation of the model is required for all seasons of the year, for three other Dee subcatchments, and for a range of flow conditions. The master recession curve approach is questioned as this traditionally is estimated as a lower envelope curve to a number of individual recessions, and possibly k values estimated for an "average" recession curve might provide a better fit of the model. A numerical search routine giving objective estimates of the parameters provides a convenient approach for resolving this question.

4.2 Objective Estimation of Parameters:

An alternative approach is possible for estimating the parameters k and L. This is to use the model to simulate a sequence of flow data corresponding to an observed sequence of precipitation and then to compare the computed flows with those observed, and to adjust the k and L values to improve the fit between the observed and computed hydrographs. This fitting can be done not over selected isolated recessions but over the whole historic hydrograph which can include both rising and falling limbs of individual hydrographs, and also indeterminate showery conditions.

One measure of the fit between the observed flow q_t and the computed q_t is the sum of squares function $F = \sum_{t=1}^{N} (q_t - \hat{q}_t)^2$. and a best fit may be defined when F is minimized and the parameter values are said to be optimized. Automatic computer routines are available to carry out the numerical work. Features of numerical fitting of hydrographs are discussed by Douglas (1974); a particular routine which has been used successfully in rainfall-runoff modelling is adopted for use in this present study.

The sum of squares criterion is not beyond criticism, but is felt to be appropriate for this work because the forecasting model is required not only for the forecasting of flood peaks, but also to operate in showery conditions, at low and average flow conditions, when subcatchment forecasts are required to assist in deciding on regulation releases that may be required from reservoirs to meet prescribed values of downstream flows.

The proposed approach places the parameter estimation into a form suitable for computer use; it also has the advantage of allowing implicitly for the losses e. Unlike the rainfall data, these losses are not known over the short time intervals required; if they were, estimates of k and L, say $\hat{k}$ and $\hat{L}$ would be obtained. If all e are set to zero, a different biased value for $k = \hat{k}^1$ will be obtained.

The use of this crude scheme is justified since the model will be used to forecast flows for up to 24 hours ahead and the losses during this time interval will be of secondary importance.

The estimate of k is expected to vary with seasons, as the losses are seasonal phenomena. As a first approximation, a two-season year of summer (May to October) and winter (November to April) is used and two sets of the model parameters are estimated for each catchment. Such a simplistic seasonal representation is tentative and ideally needs a more extensive analysis, perhaps setting k as a function of an antecedent precipitation index. But the argument against this is (1) the model immediately becomes more complex and (2) the recorded summer events were relatively sparse and more data would be needed. However, the winter data were more complete and contained many more runoff events and the modelling results reported in the following sections certainly justified the lumping of six or seven months into one "winter" season.

4.3 Numerical Estimation of Parameters, Winter Data:

The optimizing routine was applied to sequences of data from the five stations listed in Table 1 to obtain parameter values that minimized the sum of squares of differences between observed and computed hydrographs. In these runs the computed hydrographs for several months are calculated from the historic rainfall records, and then comparisons are made between the computed and observed hydrographs. The calculation of the computed hydrograph is independent of the historic flows, except for the first interval of the first month in the sequence, when the historic flow is used to initialize the recursive calculations. Because some of the sequences of data did not have the months arranged in chronological order, the initialization with observed flows was undertaken (for consistency) at the start of every month. Another way to view the calculated hydrograph is as a set of forecasts, made at the beginning of each month, of the flows to occur during the month, given perfect rainfall forecasts.

The principle results of these applications are summarised in Table 3. Columns 1, 2 and 3 respectively give the station name, a run number, and the months of data included in the run. An assessment of the model performance is given in columns 4, 5, 6 and 7. Column 4 contains the initial, no-model, sum-of-squares $F_o = \sum_{t=1}^{N} (q_t - \bar{q})^2$, which is a reference value against which the sum-of-squares given by various models and various sets of parameter values may be compared. Column 5 gives the final sum-of-squares $F = \Sigma (q_t - \hat{q}_t)^2$ obtained in the optimizing routine; a dimensionless measure of the model performance is given in Column 6 as the model efficiency $E = \frac{F_o - F}{F_o}$. Column 7 gives a volumetric comparison $\Sigma \hat{q}_t / \Sigma q_t$. Columns 8, 9, 10 and 11 list the parameter values finally obtained by the optimizing routine. If the S vs $\log_e q$ curve is divided into two linear segments, with slopes k_1 and k_2, the threshold flow value in column 10 gives the position of the division. This value

Table 3: Summary of optimising results for winter months

Catchment	Run No	Months included in data	Initial (no model) sum of squares Fo	Final sum of squares F	Efficiency $(\frac{Fo-F}{Fo})$ E	Total computed runoff / Total observed runoff $\frac{\Sigma q_t}{\Sigma q_t}$	Parameter Estimates: Lag L (hrs)	Slope k (or k_1)	Threshold q (mm/hr)	k_2	Notes
1. Ceiriog	1.1	Nov 72, Feb 73, Mar 73 Nov 73 - Mar 74	85.37	6.43	0.925	1.03	1.55	20.44	----	----	See Figs 4 and 5 for plots of Nov 72 and Jan 74
2. Alwen	2.1	Nov 72 - Apr 73	158.49	22.29	0.859	1.05	0.75	9.31	----	----	
	2.2	Nov 72 - Apr 73	158.49	19.13	0.879	1.02	1.15	20.43	0.15 *	6.57	Insufficient improvement over 2.1 to justify 2 k values
	2.3	Nov 73 - Apr 74	70.83	8.82	0.876	0.93	1.25	7.93	----	----	
	2.4	Nov 73 - Apr 74	70.83	7.84	0.889	0.93	1.5 *	13.46	0.15 *	6.15	Insufficient improvement over 2.3 to justify 2 k values
3. Gelyn	3.1	Nov 72 - Apr 73	1154.19	115.41	0.900	1.10	0.00	7.67	----	----	
	3.2	Jan 74 - Apr 74	870.94	88.25	0.899	1.00	0.00	6.06	----	----	
4. Hirnant	4.1	Nov 72 - Apr 73	372.12	25.99	0.930	1.04	0.05	11.84	----	----	
	4.2	Nov 73 - Apr 74	452.73	29.90	0.934	1.01	0.67	11.55	----	----	
5. Upper Dee	5.1	Oct 72 - Apr 73	1380.15	104.18	0.925	0.95	0.86	5.31	----	----	)Negligible improvement with)
	5.2	Oct 73 - Apr 74	1580.17	155.36	0.902	0.90	1.17	4.59	----	----	)k values

* Not included in optimisation

was not included within the optimization, but was estimated by inspection of historic hydrographs.

The efficiency E is a convenient single valued index of a model performance; it is analogous to the coefficient of determination in multiple regression and thus can be thought of as the proportion of the variance of the observed outflows explained by the model. If E = 1.0 a perfect fit is obtained, but if E < 0.0, the flows predicted by the model are less accurate than assuming the average outflow at a point in time. Although the E (or F) measures do provide single indices for comparisons between observed and computed hydrographs, excessive reliance on them can be misleading. Visual assessment of the results is useful in detecting any unusual features and for this purpose plots of the computed and observed hydrographs were made.

In Table 3 the high levels achieved for the efficiency E, and the closeness of the volumetric ratio $\Sigma\hat{q}_t/\Sigma q_t$ to unity, gave some support to the choice of this simple model and the crude allowance for losses. Plots of observed and computed flows for two months of the Ceiriog winter data (Figures 4, 5) illustrate the close fit. The results give little support to the further subdivision of the winter season. The poorest, but by no means unsatisfactory, results were those for the Alwen subcatchment (runs 2.1 and 2.3). It was thought that these would be improved by using two k values for lower and upper segments of the storage-outflow function, but this gave only negligible improvement over just one k (runs 2.2 and 2.4). Similarly, two k values with the Upper Dee had negligible effect.

The other important results of Table 3 are the parameter values. The lags ranged from 0.0 hours for the Gelyn to 1.55 hours for the Ceiriog and were less than the values estimated in the two manual examples. Estimated k values ranged from 20.44 for the Ceiriog to 4.59 for the Upper Dee. This k value defines the slope of the storage outflow function and hence nature of the recession; the lower the k value, the steeper the recession. Examples of relatively slow Ceiriog recessions are given in Figures 4 and 5; fast Upper Dee recessions will be illustrated later. For particular catchments estimates of the k and L for different winter seasons were reassuringly close and certainly the values do differ significantly between catchments.

As a simple check on the optimizing routine, the sum-of-squares F for the Ceiriog winter data is plotted for a range of k and L values in Figure 6. A sum-of-squares surface interpolated from these grid points appears "well-behaved"; it shows no interdependency between k and L and is insensitive to changes in L compared with changes in k. It has just one minimum at k ≈ 20.5 and L = 1.5 where F = 6.5 approximately. This compares well with the minimum given in the numerical optimizing in Run 1.1 of Table 3 (k = 20.44, L = 1.55, F = 6.43). It also compares with that estimated graphically (22.7) but the lag is less than the four hours estimated as being an average time between peak rainfall and peak runoff. From Figure 6 these graphical values give F = 8.8, and thus an efficiency E = 0.897, only a little less than the optimized result

E = 0.924. This is only a small difference which does not, in itself, justify the extra computational effort needed in the numerical optimization: it does however give some basis for confidence in the numerical estimation technique, use of which gave considerable savings in the man-hours needed to estimate parameters for each of several periods for the five subcatchments.

The results for the Ceiriog show that the recessions fit well, but that some peak flows tend to be underestimated. The Ceiriog is used as a representative station for estimating ungauged lateral inflows along the reach of the main river between Corwen and Manley Hall, and the estimation of peak flows is considered to be particularly important. Better estimation of peaks can be obtained at the expense of poorer recessions by using a lower k value; this was the reason for Lambert's selection of the low value of k = 14. *The sum-of-squares surface in* Figure 6 shows that for k = 14.0, L = 1.5, the F = 13.7 which is a large increase from the minimum of F = 6.43. However, for initial steps in the direction of this point away from the minimum, the change in F is relatively small. Thus for example k = 17.0, L = 1.5 gives F = 8.1 and it is suggested that this k value should be adopted for the Ceiriog.

4.4 Verification of the Model, Winter Data:

A rigorous test of a hydrological model is to check its performance with data from outside the calibration period: this may variously be described as model verification, split-record testing, and model prediction. With the exception of the Ceiriog, the winter data are conveniently divided into two sequences for the two successive winters, and the results from modelling with these sequences are summarised in Table 3.

Allied to the need for this test, the two sets of model parameter estimates for each catchment did have some variation between them, and another question needing investigation was the significance of this variation. Figure 6 suggested that the fit of the model was insensitive to values over a wide range chosen for the lag and that in the immediate vicinity of the minimum point the fit was fairly insensitive to changes in k. A more complete answer is given by using the parameters estimated for one winter to estimate the flows resulting from the rainfalls recorded for the other winter. Examples of comparisons of the estimated flows are given in Figures 7-13 and a full summary of the results is given in Table 4.

The results in Table 4 shows that the model is robust and does perform well with data from outside the calibration range. The efficiencies are within 3% of the values obtained for the optimizing runs on the corresponding sequences in Table 3, and the computed volumes are again close to the observed volumes, indicating that the order of variation for parameter estimates for a particular catchment is acceptable. The lowest efficiencies for all the winter data were for the Alwen, and Figure 7 gives the poorest fitting month, November 1972, for the Alwen data. In this month the volume of runoff and the peak flows were overestimated, probably because insufficient allowance was made for the depleted state of the soil moisture following a dry period of more than 10 days. Note however that the peaks are correctly positioned in time and that the recession

Table 4 Verification of winter model

Catchment	Months included	Initial (No model) sum of squares F_o	Final sum of squares F	Efficiency $\left(\frac{F_o - F}{EF_o}\right)$	$\frac{\text{Total computed}}{\text{Total observed}} = \Sigma\hat{q}_t/\Sigma q_t$	Parameters: Lag L (hours)	Parameters: Slope k	Run No. in Table 3 giving k & L	Months plotted	Fig. Nos.	Notes
2. Alwen	Nov 72 - Apr 73	158.49	24.16	0.848	1.05	1.0	7.93	2.3	Nov 72 Dec 72	Fig 7 Fig 8	Worst winter result
	Nov 73 - Apr 74	70.83	9.39	0.867	1.08	1.0	9.31	2.1	- - - - - -		
3. Gelyn	Nov 72 - Apr 73	1154.19	151.31	0.869	1.10	0.00	6.06	3.2	Dec 72 Apr 73	Fig 9 Fig 10	April plot gives example of unresolved timing error
	Nov 73 - Apr 74	870.94	107.34	0.877	1.00	0.00	7.67	3.1			
4. Hirnant	Nov 72 - Apr 73	372.12	27.28	0.927	1.04	0.5	11.55	4.2	Dec 72	Fig 11	
	Nov 73 - Apr 74	452.73	30.06	0.934	1.01	0.5	11.84	4.1	- - -		
5. Upper Dee	Oct 72 - Apr 73	1380.15	112.64	0.918	0.95	1.0	4.59	5.2	Feb 24	Fig 12	
	Oct 73 - Apr 74	1580.17	169.08	0.893	0.90	1.0	5.31	5.1	Mar 74	Fig 13	Example of snowmelt
	Oct 72 - Apr 73	1380.15	92.12	0.933	0.92	Lambert's values (Table 2)			- - -		
	Oct 73 - Apr 74	1580.17	188.35	0.881	0.89	Lambert's values (Table 2)			- - -		

calculated by the model does occur at about the same rate as that observed. Thus it is to be expected that the errors will be far less when the model is used in its intended form for real-time forecasting and the forecasts for 24 hours ahead from time 't' will always be calculated from the observed flow at time 't'.

Figures 9-13 show observed and computed hydrographs for a selection of months for the Gelyn, the Hirnant, and the Upper Dee. Figure 10 for the Gelyn for April 1973 shows a timing error of about two hours between observed and computed peak flows which would contribute towards the F value. This was the only such error observed and in general both rainfall and the flow data appeared to be of a very high quality.

A plot for the Hirnant is given in Figure 11; here the fit is good except for underestimates of the peak flows. Other examples for the Upper Dee follow in Figures 12 and 13. One of these plots, Figure 13, illustrates another source of lack of fit: snowmelt, not modelled, occurred on 6th and 7th of March 1974 following snow which fell during the first few days of the month, and shows as a substantial pulse on the observed hydrograph following a small amount of rain, whereas the model, having over-predicted earlier immediate runoff from precipitation which occurred as snow, remains in recession. It was observed that the model has tended to overestimate the historic Upper Dee flows that were greater than about 40 cumecs. However as mentioned earlier, the recorded flows above this value at this station are underestimates of the true flows and therefore the errors in the estimated peaks were less than was apparent.

These winter modelling results were considered to be acceptable and the final point to be resolved is the parameter values to be adopted for further model use. The results summarized in Table 4 show that the level of variation found in the parameter estimates has relatively little effect on flow predictions, and the best course appears to be to pool the two sets of estimates and adopt mean values for each station. The lags are rounded to the nearest half hour and the final values are given in Table 5.

Table 5: <u>Suggested parameter values for winter months</u>

Catchment	L (hours)	k
1. Ceiriog	1.5	17.0
2. Alwen	1.0	8.6
3. Gelyn	0.0	6.9
4. Hirnant	0.5	11.7
5. Upper Dee	1.0	4.9

Table 6: Summary of optimising results for summer months

Catchment	Run No.	Months included in data	Initial sum of squares F_o	Final sum of squares F	Efficiency E	Total computed / Total observed $\Sigma\hat{q}_t/\Sigma q_t$	Parameters Estimated: Lag (hrs)	Slope k (or k1)	Threshold q	Slope k2	Notes
1. Ceiriog	1.1	Oct 72, Apr 73, Sept 73 Oct 73, May 74, June 74	27.65	6.88	0.751	1.22	0.02	35.7	-	-	
	1.2	ditto	27.65	3.50	0.874	1.09	1.25	60.38	0.08	17.94	Plots of Sept, Oct 73 in Figs 14, 15
2. Alwen	2.1	May 73 ... Oct 73	94.93	34.84	0.633	1.48	0.00	16.16	-		
	2.2	May 73 ... Oct 73	94.93	23.04	0.757	1.33	1.10	23.59	0.10	9.09	
	2.3	May 74, Jun 74 July 72 ... Oct 72	98.74	26.14	0.735	1.60	1.30	10.72			
	2.4	ditto	98.74	23.07	0.766	1.46	1.38	15.80	0.10	7.99	
3. Gelyn	3.1	Sept 72, Oct 72, May 73 Jun 73, Jul 73, May 73 Jun 74	262.53	82.53	0.686	2.07	0.78	11.08			
	3.2	ditto	262.53	53.70	0.795	1.72	1.09	24.18	0.10	6.84	Plots of May 73 July 73 in Figs 16, 17.
4. Hirnant	4.1	May 73 ... Oct 73	251.46	40.16	0.840	1.37	0.00	15.90			
	4.2	May 74, June 74, July 72 ... Oct 72	46.70	23.23	0.503	1.48	0.00	16.37			
	4.3	May 73 ... Oct 73	251.46	29.04	0.885	1.22	0.00	32.84	0.10	13.06	
	4.4	May 74, June 74, Jul 72 ... Oct 72	46.70	22.45	0.519	1.42	0.02	18.73	0.10	11.89	

4.5 Numerical Estimation of Parameters: Summer Data:

Modelling the "summer" data was more difficult than modelling the "winter" data for the following reasons: 1) losses from interception, evaporation and transpiration are greater; 2) less rain gives fewer runoff events; and 3) much of the missing data was for summer months. However the success in modelling the winter data suggested that the approach should be attempted for the summer months. This was done and the optimized results are presented in Table 6.

These "summers" were made up from the data not used in the winter modelling. For the Alwen and the Hirnant the months May 73 - October 73 formed one "summer" and the months May 74 - June 74 and July 72 - October 72 formed the other. For the Upper Dee, sequences finishing one month earlier, in September, were used. For the Ceiriog and the Gelyn just one summer of all recorded data between May and October was used. Because the months were not necessarily in chronological order, the calculated flows for each month were centred on to the historic flow observed at the start of the month.

The results achieved in terms of efficiency were not so good as for the winter months, but nevertheless were encouraging. The accounting for losses by means of a biased k value is not as satisfactory as the computed flow volumes all exceeded the observed volumes, generally by between 20% and 107%. Introducing an arbitrary threshold runoff value and using two k values, one of which applied above the threshold and one below, gave a considerable improvement to the fit of the model for the Ceiriog, the Alwen, and the Gelyn. The improvement for the Hirnant was less marked and for the Upper Dee appeared negligible. With two k values, the observed volumes were again overestimated, but by lower margins.

Examples of plots of computed and observed hydrographs for the optimized parameters (with two k values and with the lag rounded to the nearest half

Table 7: Comparison of peak flow for two sets of summer data

Catchment	"Summer" I May-June 74 July-Sept/Oct 72		"Summer" II May-Sept/Oct 73	
	Peak Flow (Cumecs)	Date	Peak Flow (Cumecs)	Date
Alwen	110.9	1.8.72	67.6	19.10.73
	17.2	7.8.72	57.0	5. 8.73
	16.4	4.7.72	43.3	16. 7.73
Hirnant	14.8	1.8.72	33.8	5. 8.73
	6.8	7.8.72	11.8	3. 9.73
	3.2	4.7.72	11.0	18.10.73
Upper Dee	10.2	8.9.72	53.9	5. 8.73
	8.8	16.6.74	52.0	3. 7.73
	8.3	6.6.74	44.9	16. 7.73

hour) are shown for the Ceiriog for September 1973 (Figure 14) and October 1973 (Figure 15), and for the Gelyn for May 1973 (Figure 16) and July 1973 (Figure 17). The Ceiriog results show the contrast between two months: those for September 1973 were disappointingly poor, probably the poorest of all the months for all the stations, yet for the following month, October 1973, the result is far better. This illustrates the need for caution in interpreting a single dimensionless efficiency index; despite the poor prediction for September 1973, the efficiency achieved over all the six months included in the sequence was at a respectable level of 0.874 and the volume was overestimated by only 9%. In contrast, for the Gelyn, the efficiency was lower at 0.795 and the volume was overestimated by 72%, but visually the fit of the predicted hydrograph illustrated in Figures 16 and 17 was considered to be very good. Possibly this good fit is misleading as these two illustrations show the two major storms of the seven summer months and fitting the upper range of the model will be based on these two events. In the remaining five months (not shown) the flows were very low in which case a small absolute error in estimating the flow volume gives a large percentage error.

Results for the "summer" made up from May-June 1974 and July-Sept/Oct 1972 need to be considered with caution as these were particularly dry periods containing only one significant storm event on 1st August. The point is made by comparing the magnitude of the three peak flows that occurred in this period with the three peak flows on record for May-Sept/Oct 1973; these are set out for the Alwen, the Hirnant and the Upper Dee in Table 7. High flows for the summer of May-June 1974, July-Sept/Oct 1972 are dominated by the event of 1st August and other peaks are much less. Data for the 1st August were absent from the Upper Dee record, and peaks apart from this date were all exceeded by the peaks for the 1973 summer. Parameters estimated for the summer data for 1973 are therefore expected to be more robust and it is suggested those estimated for the 1973 summer should be used for these three stations. Thus the suggested summer parameters are given in Table 8.

Table 8: Suggested parameter values for summer months

Catchment	L (hours)	k or k_1	Threshold q (mm/hrs)	k_2
1. Ceiriog	1.0	60.4	0.08	17.9
2. Alwen	1.0	23.6	0.10	9.1
3. Gelyn	1.0	24.2	0.10	0.8
4. Hirnant	0.0	15.9	-	-
5. Upper Dee	1.0	7.6	-	-

4.6 Verification of the Model: Summer Data:

The limitations for model calibration of the data from the combined summer of 1974 and 1972 were described in the previous section and the parameter values adopted are given in Table 8. The parameters for the Alwen, the Hirnant and the Upper Dee were estimated solely from the 1973 summer. As examples of out of range testing these were used to generate flows corresponding to the sequences of rainfalls for the summer of 1974 and 1972. The results from this exercise, which are summarised in Table 9, provide a partial test of the model. As in the calibration runs for these data the flow volumes are overestimated. The efficiencies are equal to, or less than, the efficiencies for the corresponding calibration runs in Table 6.

Examples of the reconstructed hydrographs are given in Figures 18-20. Figures 18 and 19 show the major event of 1st August, 1972, for the Alwen and the Hirnant. Results for the Alwen are considered good, despite the overestimate of the peak flows for the first two days. Historical data is excluded for August 3-5 and 9-11 and minus unity substituted for the historic flow because the Alwen Reservoir overflowed during these times. In a forecasting mode, given re-centering of the forecasted flows on to the current true value, the projections for future flows should be acceptable. Figure 19 for the Hirnant shows a drift in the calculated hydrograph away from the observed flows and results in a gross overestimation of the flow volume. Again, with re-centering, the real-time forecasts would be much closer to the historic values since the calculated hydrograph does have the correct shape with the peaks correctly located in time. These comments also apply to Figure 20 which shows the Upper Dee hydrographs for June 1974: again the flow volume is overestimated; the efficiency is low (in fact it was negative for this sequence) but the shape is correct and real-time re-centering is expected to give acceptable flow forecasts. Certainly the accuracy of the forecasted flows, given perfect rainfall forecasts, is likely to be much greater than the accuracy that will in fact be achieved for precipitation forecasts. On the catchments studied here the empirical catchment lags were estimated as varying between 0.5 and 1.5 hours and forecasts of flows to occur more than these times ahead of the present will be entirely dependent on the forecasts of rainfall.

Table 9: Verification of summer model

Catchment	Months included in data	Initial sum of squares	Final sum of squares	Efficiency	Total computed / Total observed	Parameters				Months Plotted
		F_o	F	E	$\Sigma q/\Sigma q$	Lag (hrs)	(k_1)	Threshold flow	k_2	
Alwen	May-June 74 July-Oct 72	98.74	35.04	0.645	1.30	1.0	23.59	0.10	9.09	Aug 72 (Fig 18)
Hirnant	May-June 74 July-Oct 72	46.70	23.25	0.502	1.50	0.0	15.9	-	-	Aug 72 (Fig 19)
Upper Dee	May-June 74 Aug-Sept 72	12.33	22.13	- 0.79	2.10	1.5	7.58	-	-	June 74 (Fig 20)

5. SIMULATION OF REAL-TIME OPERATION

In real-time flow forecasting a wealth of data is produced and a real problem exists in displaying it. The previous plots of calculated hydrographs showed the model operating in a record-reconstruction, or data generation form, but these can also be considered as forecasts made at the beginning of a month of the flows to occur during the month given perfect rainfall forecasts. In its intended form, 24 hours of forecasts will be produced at every half-hour interval and only part of this information can be illustrated as part of a simulation exercise.

Figures 21, 22 and 23 show typical results from this exercise for two successive winter months and one summer month. At every half-hour interval, flows and rainfall up to present time "t" are assumed known from the telemetry system, perfect rainfall forecasts are assumed, and the flows for the next 48 intervals (24 hours) are calculated as $\hat{q}_t(1)\ldots\hat{q}_t(48)$. The information shown in the figures is the sequence of one-step-ahead forecasts $\hat{q}_t(1)$. These two plots are very close, a not unexpected result since $\hat{q}_t(1)$ is calculated from q_{t-1}, and the model should not be greatly in error over half an hour. Note that Figure 21 shows the Alwen for November 1972 and Figure 23 the Ceiriog for September 1973 and that these months were amongst the poorest results in the calibration study (Figures 7 and 14). Many other plots of $q_t(J)$ and q_{t-J} are possible for a range of lead times J; particular interest centres on the one-step-ahead forecasts, however, since if these are in error, forecasts for larger lead times will also be in error.

6. ANALYSIS OF FORECASTING ERRORS

A sequence of one-step-ahead forecasting errors, w_t, can be defined by the differences between q_t and $\hat{q}_{t-1}(1)$: thus

$$w_t = q_t - \hat{q}_{t-1}(1). \qquad t = 2, \ldots N.$$

Persistence, measured by the autocorrelation function, may exist in these errors. If the persistence is significant, it would suggest that the model errors contain information about future errors that is not included in the model and that either the model should be modified to include this extra information, or the errors w_t should be modelled with a stochastic model. Given a suitable stochastic model, at time "t" the known sequence $\ldots w_{t-2}, w_{t-1}, w_t$ may be used to estimate values for $w_{t+1}, w_{t+2}, \ldots$. These forecasts of errors can then be used to improve the accuracy of the discharge forecasts.

As an example, consider the one-step-ahead forecasts for the Alwen between 1 November and 11 December which are plotted in Figures 21 and 22. This series covers 41 days and contains N = 41 x 48 = 1968 data points. The mean flow during the period was $\bar{q} = 0.2154$ mm/hr (9.56 cumecs) with variance var(q) = 54.23×10^{-3}. The mean of the one-step-ahead forecast errors was $\bar{w} = 0.754 \times 10^{-3}$ mm/hr with variance var(w) = 0.5753×10^{-3}. Thus the proportion of the variance of the flows explained by the one-step-ahead forecasts was 0.989. This is a deceptively good result which is brought about in part by using the very short time interval of half an hour. The object of a stochastic model for the errors is to account for as much of the remaining variance as is possible.

Assuming the w_t to be stationary, and following the identification, estimation and checking procedures of Box and Jenkins (McKerchar and Delleur, 1972) the autocorrelation function (acf) estimated for w_t was:

Lag (i)	1	2	3	4	5	6	7	8	9
Acf (r_i)	0.862	0.674	0.505	0.343	0.230	-0.147	-0.075	0.026	-0.002

Lag (i)	10	11	12
Acf (r_i)	-0.019	-0.045	-0.069

This autocorrelation function and the associated partial autocorrelation function suggested that either a second-order autoregressive (AR(2)) model or a combined first-order autoregressive and first-order moving average model (ARMA(1,1)) would fit the series. The AR(2) model has the form:

$$w_t - \bar{w} = \phi_1(w_{t-1} - \bar{w}) + \phi_2(w_{t-2} - \bar{w}) + a_t, \qquad \ldots (9)$$

where ϕ_1 and ϕ_2 are the AR coefficients to be estimated and a_t is residual pure-random series. Approximate maximum likelihood estimates of the coefficients using a non-linear iterative algorithm were $\phi_1 = 1.09$, (std error = 0.02), $\phi_2 = -0.26$ (std error = 0.02) and var $(a_t) = 0.1367 \times 10^{-3}$. The acf estimated for the residual a_t series was:

Lag (i)	1	2	3	4	5	6	7	8
Acf (r_i)	-0.001	-0.028	0.113	-0.087	-0.057	0.038	-0.033	-0.033

Lag (i)	9	10	11	12
Acf (r_i)	-0.028	-0.043	-0.023	0.048

The approximate standard error for this acf is ± 0.023, indicating that r_3, r_4 and r_5 in particular may differ significantly from zero and that the a_t may not be pure random.

The ARMA (1,1) model has the form:

$$w_t - \bar{w} = \phi_1(w_{t-1} - \bar{w}) + a_t - \theta_1 a_{t-1}, \qquad \ldots (10)$$

where ϕ_1 and θ_1 are the coefficients to be estimated and a_t is a pure random series. Again, fitting this model, the estimation algorithm gave $\phi_1 = 0.79$, (std error = 0.02), $\theta_1 = -0.31$ (std error = 0.02) and var $(a_t) = 0.1368 \times 10^{-3}$. The acf estimated for the a_t series was:

Lag (i)	1	2	3	4	5	6	7	8
Acf (r_i)	-0.000	0.034	0.085	-0.102	-0.019	0.015	-0.045	-0.033
Lag (i)	9	10	11	12				
Acf (r_i)	-0.036	0.040	-0.026	-0.007				

with standard error = ± 0.023, again suggesting that r_3 and r_4 may differ significantly from zero and that the a_t may not be pure random.

Thus both model fitting attempts suggest that a higher order model, probably AR(4), should be used. However, in doing so the reduction in variance achieved was minimal (var $(a_t) = 0.1351 \times 10^{-3}$) and r_4 and r_6 were still highly significant. For illustrative purposes we consider the use of the AR(2) model for forecasting. Using the notation $\hat{w}_t(J)$ for the forecast of w_{t+J} made at time t, it can be shown that by taking conditional expectations of Eqn (9) (McKerchar and Delleur, 1972) that the following recursive scheme applies:

$$\hat{w}_t(1) = \bar{w} + \phi_1(w_t - \bar{w}) + \phi_2(w_{t-1} - \bar{w}) \quad \ldots (11)$$

$$\hat{w}_t(2) = \bar{w} + \phi_1(\hat{w}_t(1) - \bar{w}) + \phi_2 (w_t - \bar{w}) \quad \ldots (12)$$

$$\hat{w}_t(J) = \bar{w} + \phi_1(\hat{w}_t(J-1) - \bar{w}) + \phi_2 (\hat{w}_t(J-2) - \bar{w}),\ J \geq 3 \quad \ldots (13)$$

Further, the standard errors for the forecasts are given by:

$$S_W(J) = (\psi_o{}^2 + \psi_1{}^2 + \ldots + \psi_{J-1}{}^2)^{\frac{1}{2}} \sigma_a \quad \ldots (14)$$

where $\psi_o = 1$

$\psi_1 = \phi_1$

$\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2},\ j > 1$

Noting that $\sigma_w = 0.02399 = 2.05\ \sigma_a$, and substituting our values for ϕ_1 and ϕ_2, the asymptotic increase in the forecast error $S_W(J)$ from σ_a towards σ_w is demonstrated in the following table.

J	1	2	3	4	5	6	7
ψ_{J-1}	1.00	1.09	0.83	0.62	0.45	0.33	0.24

$\left\| \sum_{j=0}^{J-1} \psi_j^2 \right\|^{\frac{1}{2}}$	1.00	1.48	1.70	1.86	1.91	1.94	1.96
$s_W(J)$	0.0117	0.0173	0.0199	0.0218	0.0223	0.0227	0.0229

Thus for example, for the four-step-ahead forecast ($J = 4$); the standard error $s_w(4) = 0.0218$ and the proportion of the forecast error variance that would be explained by the stochastic model would be

$$\{1.0 - (s_w(4)/\sigma_w)^2\} = 0.17,$$

and this figure decreases further as the lead time J increases. The point of this illustration is that for all the subcatchments for $J = 4 = 2$ hours, forecasts of flows will be dependent on forecasts of precipitation; and errors in the precipitation forecasts will be large and will completely outweigh any improvement in the forecast accuracy that will be achieved by utilizing a stochastic scheme for forecasting errors in this situation. It follows that improved forecasting accuracy might be better achieved by improving the accuracy of rainfall forecasts, rather than by refining hydrological rainfall-runoff models.

One theoretical consideration which should be mentioned relates to the estimation of the stochastic model parameters: as described herein the parameter values were determined for the AR(2) model and the ARMA (1,1) model as those which minimized the sums-of-squares Σa_t^2. By implication, this minimization is conditional upon the values of the parameters (k and the lag) previously determined for the conceptual part of the model and therefore this is a two-stage fitting process; the true unconditional minimum of Σa_t^2 and the corresponding parameters may not be achieved.

Clarke (1972) suggested that a more efficient estimation method would be to estimate the conceptual and stochastic parameters jointly. With the assumption that the forms of the conceptual and stochastic parts of a combined model remain the same, there would appear to be no particular difficulties, other than the usual difficulties of multi-dimensional optimization, in extending the use of the optimizing routine to estimate all the parameters by minimizing Σa_t^2. However it is necessary first to carry out the two-stage estimation procedure for the two parts of the combined model to determine the order of the ARMA model that is required.

Although the use of a combined conceptual-stochastic model is not suggested at this stage for real-time flow forecasting on the Dee River subcatchments, this section has been included to illustrate what should be a worthwhile avenue of investigation to pursue in other studies. Any refinement of the subcatchment model for real-time forecasting would need to be balanced against the relatively gross approximations in the precipitation forecasts to ensure that efforts to improve the accuracy of forecasted downstream hydrographs are not misdirected. At present it appears that the rainfall forecasts will be the major source of errors in real-time real-world flow forecasts from subcatchments.

In the complete Dee forecasting model, the subcatchment discharge forecasts are combined with reservoir and channel routing procedures to produce forecasts of discharge at Manley Hall. Although this report has been concerned only with the subcatchment forecasting, the stochastic modelling concepts are also applicable to the Manley Hall forecasts, despite the fact that in this case the problem of jointly estimating the parameters is not straightforward.

7. CONCLUSIONS

A simple conceptual model has been calibrated and tested for five subcatchments of the Dee River in North Wales.

A graphical method for estimating the model parameters gave results comparable with those from a numerical search routine. By requiring the drawing of the storage-outflow curve, the graphical method provided a check on the assumptions made in formulating the model as a simple storage representing the catchment. The numerical estimation procedure was adopted because (1) it could be applied to a long sequence of data covering a range of flow conditions, (2) it automated a large amount of tedious numerical work and (3) it did not require the subjective selection of particular flow recessions and the drawing of a master recession curve. The two approaches should be considered as complementary rather than competitive.

To account for seasonal variations in losses through evaporation and transpiration, the model was fitted separately to "summer" and "winter" where "summer" and "winter" were arbitrarily defined as including the months between May and October, and November and April, respectively.

Although the model will be used in an operational "real-time" forecasting context, it was calibrated for individual subcatchments in a data generation mode. It was inferred from the closeness of the computed and observed discharge hydrographs that the differences in estimates between the parameters between these two modes should be negligible.

Examples of simulated real-time operation with perfect rainfall are given. A sequence of one-step-ahead forecast errors is shown to be serially correlated: it is suggested that this fact could be used by a stochastic model for forecasting future errors. Aspects of one stochastic model are outlined, and it is suggested that for an efficient fitting of the combined conceptual and stochastic model, the parameters need to be estimated jointly.

Even without a stochastic model to forecast errors, the simulation of the real-time operation suggested that a high proportion of the variance of flows is accounted for by the one-step-ahead forecasts. Much of this high figure must be due to the short time step (half an hour) and the fact the flows change by a relatively small amount over this time. It must also be remembered that this simulation used excellent rainfall data from a network of raingauges and that in practice the accuracy of the rainfall data, either from telemetering gauges or from the radar, will not be as high. Secondly, beyond the catchment lag times of 0.0 to 1.5 hours, the forecasts of runoff will be dependent on precipitation forecasts. These are likely to be issued at about 24 hour intervals and the accuracy that will be achieved, particularly in timing, is likely to be low in comparison to the accuracy of the rainfall to runoff conversion. This presents a practical limit to the required accuracy of the subcatchment modelling and suggests that in the Dee River forecasting the first priority should be to improve the accuracy of the precipitation forecasting scheme, rather than to pursue the subcatchment model joint estimation problem. This is not to denigrate the need for better hydrological models; it is intended only to specify priorities for improving Dee River forecasts.

8. ACKNOWLEDGEMENT

The author wishes to thank Mr A O Lambert of the Dee and Clwyd Rivers Division, Welsh National Water Development Authority and IH colleagues for helpful discussions, the Water Data Unit and the Meteorological Office for providing data.

9. REFERENCES

Clarke R. T. 1972. General Report on stochastic and statistical models. *Proc. Int. Symp.* on Uncertainties in Hydrology and Water Resource Systems *3*, 1325-1356, Tucson, Ariz.

Cole J. A., McKerchar A. I. and Moore R. J. 1975. An on-line flow forecasting system incorporating radar measurements of rainfall as used to assist the short-term regulation of the River Dee in North Wales. (*Pro.* Bratislava *Symp.*). Application of Mathematical Models in Hydrolgoy and Water Resource Systems. *IAHS Pub. No.* 115, 57-66.

Douglas J. R. 1974. Conceptual Modelling in Hydrology. *Inst. Hydrol, Wallingford, Rep. 24.*

Jamieson D. G. and J. C. Wilkinson, 1972. River Dee Research Program 3. A short-term control strategy for multipurpose reservoir systems. *Wat. Resour. Res., 8* (4), 911-920.

Lambert A. O. 1969. A Comprehensive Rainfall/Runoff Model for an Upland Catchment Area. *Instn. Wat. Engrs. 23*, 231-238.

Lambert A. O. 1972. Catchment Models Based on ISO-Function. *J. Instn. Wat. Engrs. 26*, 413-422.

Lowing M. J., Price R. K. and Harvey R. A. (in press). Real time conversion of rainfall to runoff for flow forecasting in the River Dee. Symp. on Weather Radar and Water Management. Water Research Centre and Royal Radar Establishment.

McKerchar A. I. and Delleur J. W. 1972. Stochastic Analysis of Monthly Flow Data: Application to Lower Ohio Tributaries. Purdue University Water Resources Research Centre, *Tech. Report No.* 26.

Price R. K. 1973. Flood Routing Methods for British Rivers. *Proc. Instn. Civil Engrs.* part 2, *55*, 913-930.

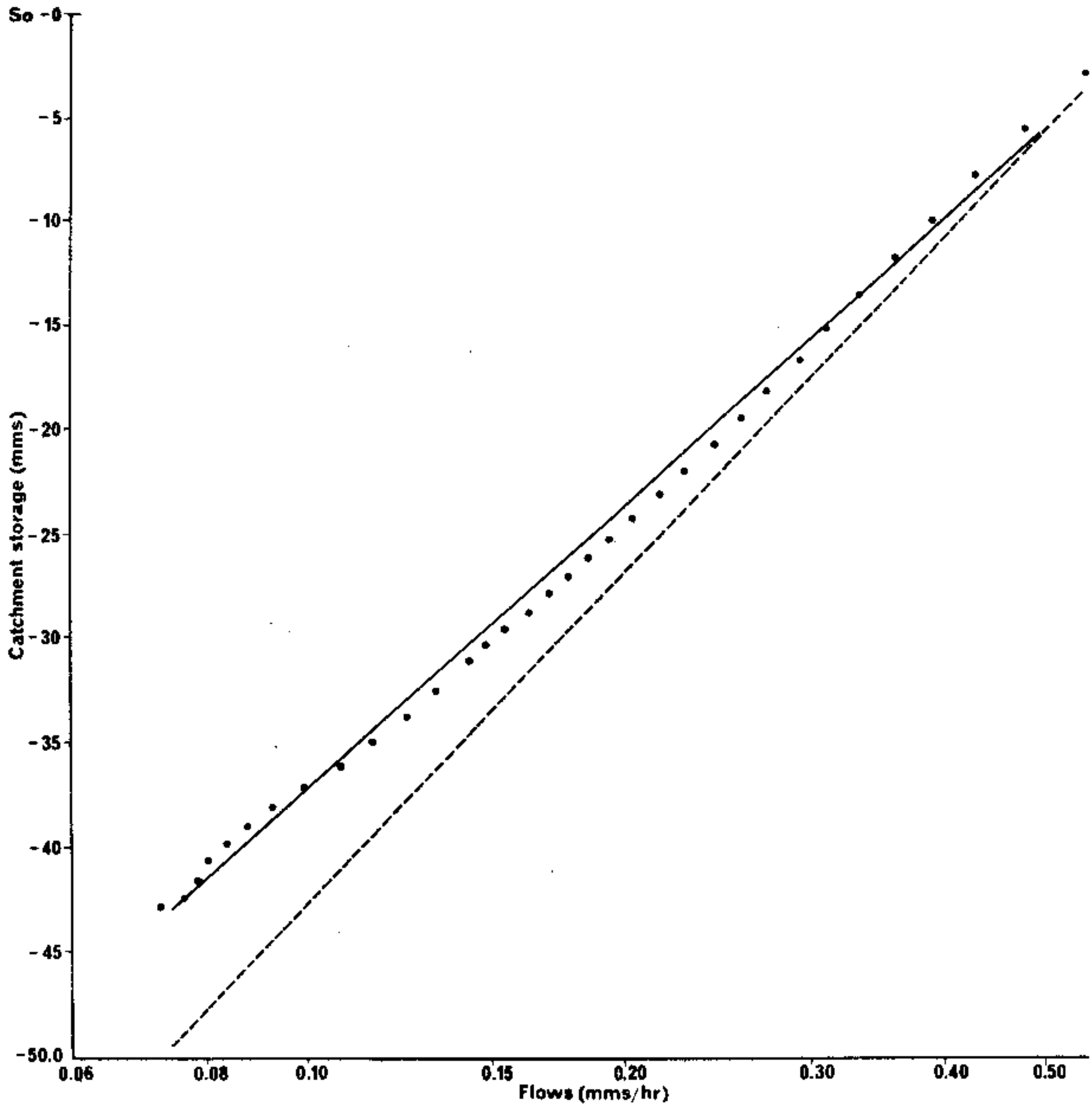


Figure 2: Ceiriog: Storage-outflow function

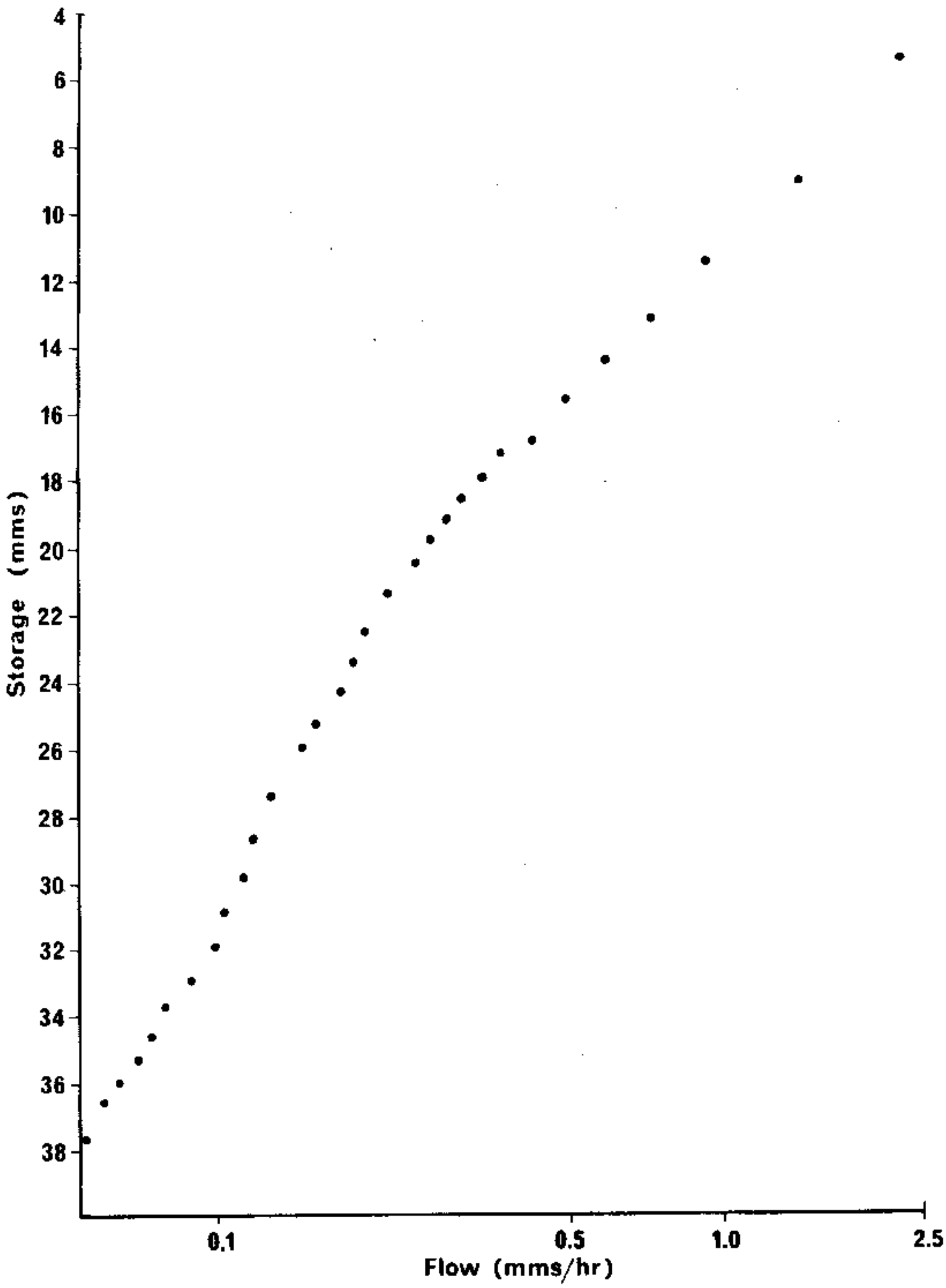


Figure 3: Upper Dee: Storage-outflow function

Figure 4

STØRAGE-ØUTFLØW RELATIØNSHIP LØG-LINEAR S VS Q

67/05 CEIRIØG AT BRYNKINALT WEIR NØVEMBER 72

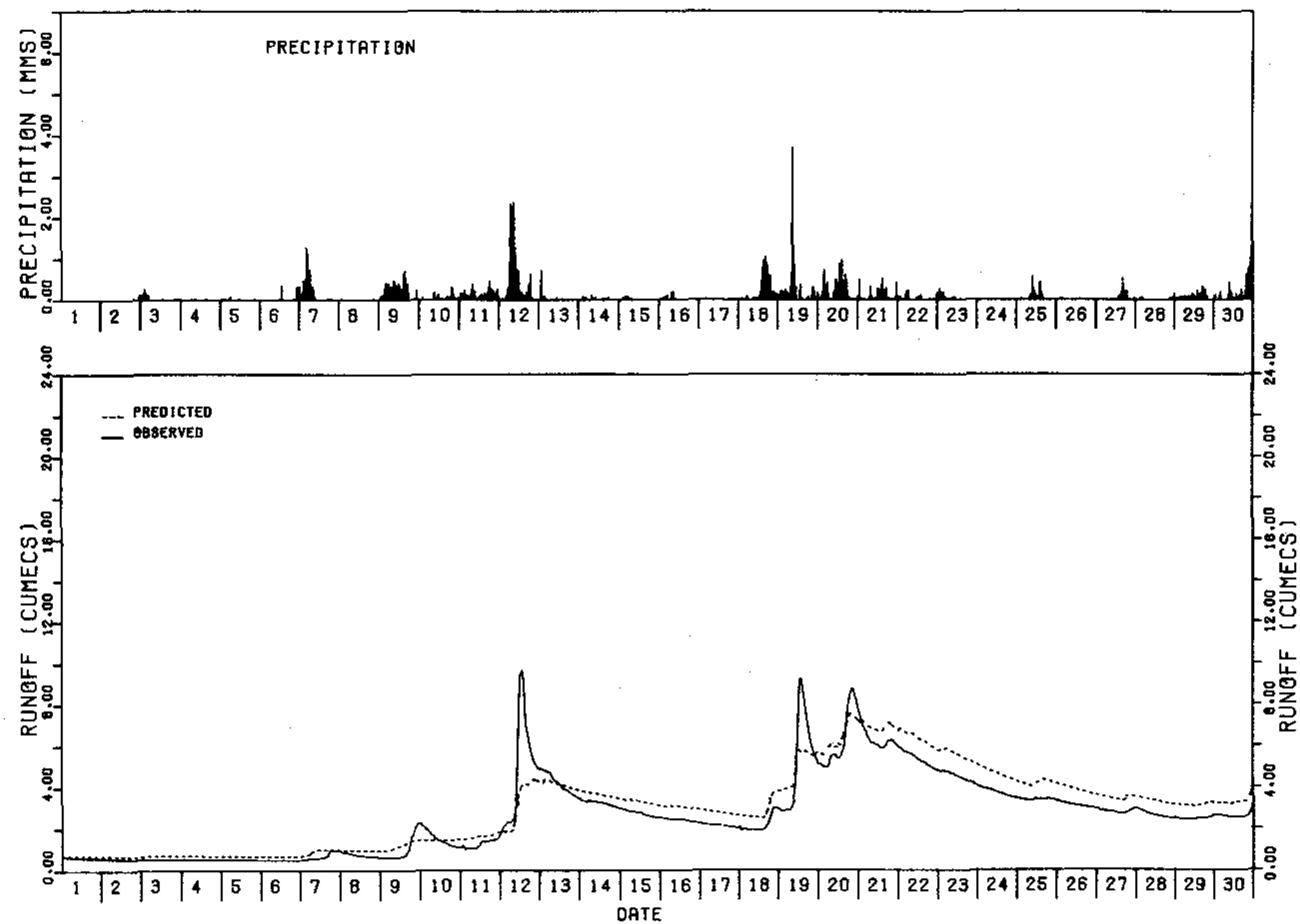
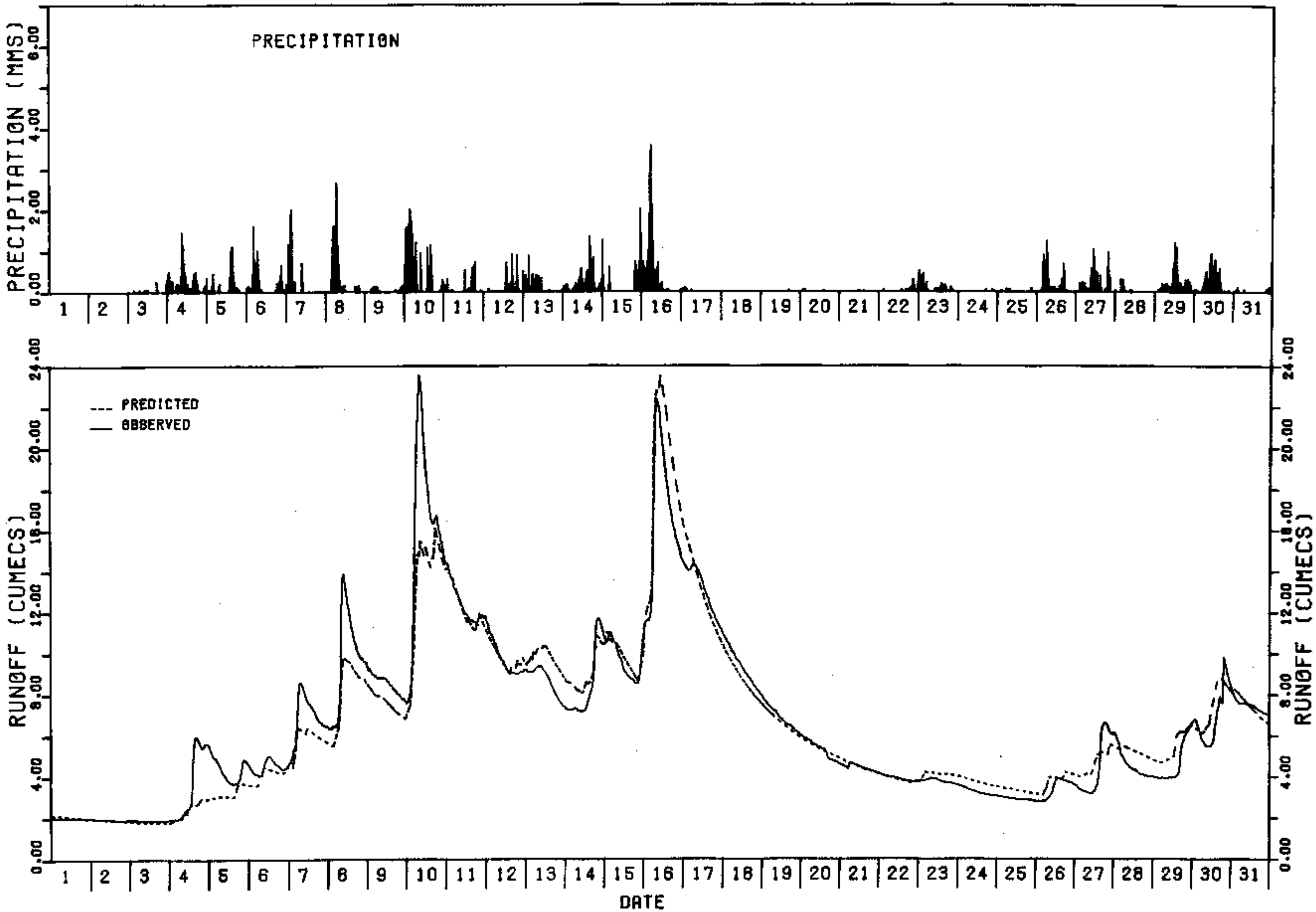


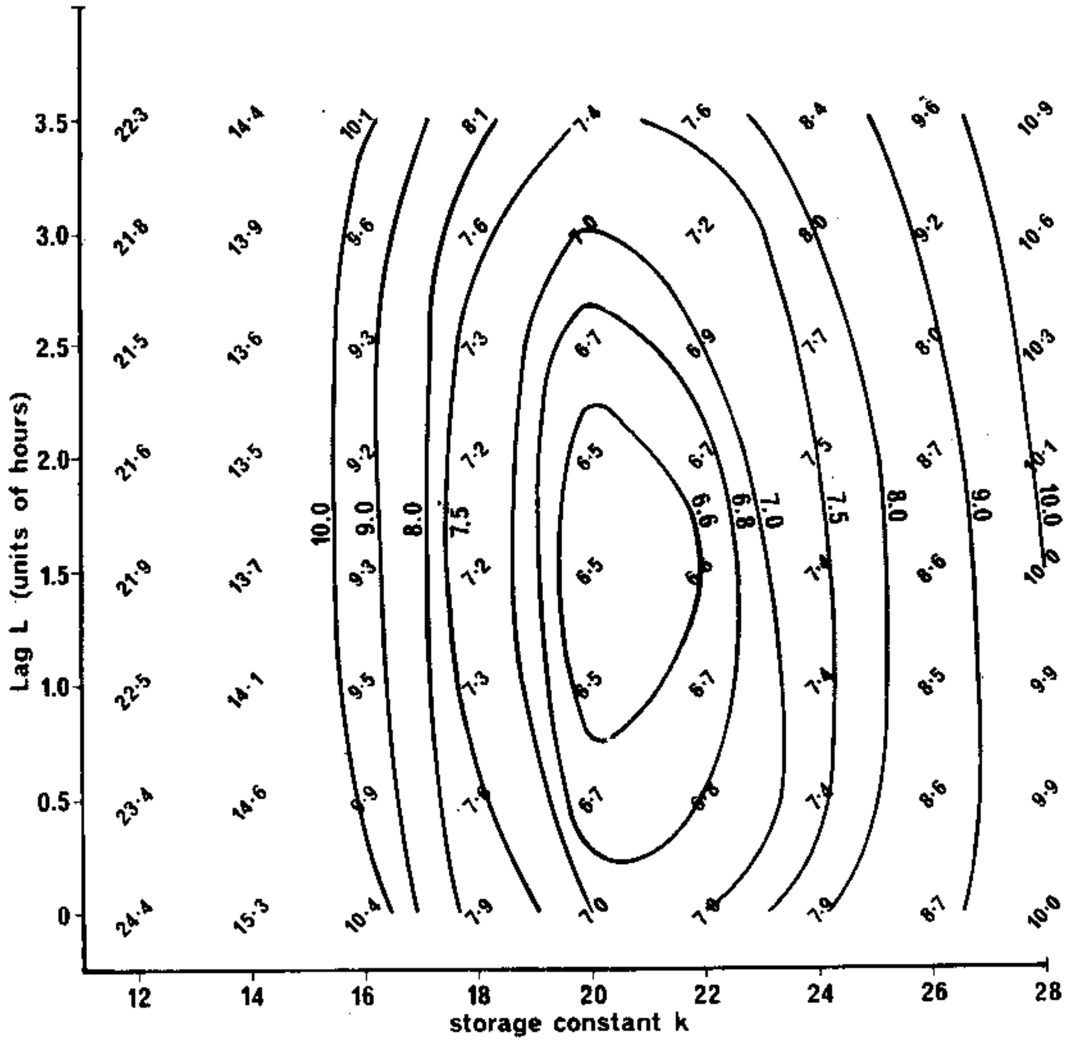
Figure 5

STORAGE-OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/05 CEIRIOG AT BRYNKINALT WEIR

JANUARY 74





'Optimum' reached numerically was, L = 1.55 hrs, k=20.4, F=6.433

Figure 6: Ceiriog winter data: sum of squares surface

Figure 7

STORAGE OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/06 ALWEN AT DRUID NOV72-APR73 NOVEMBER

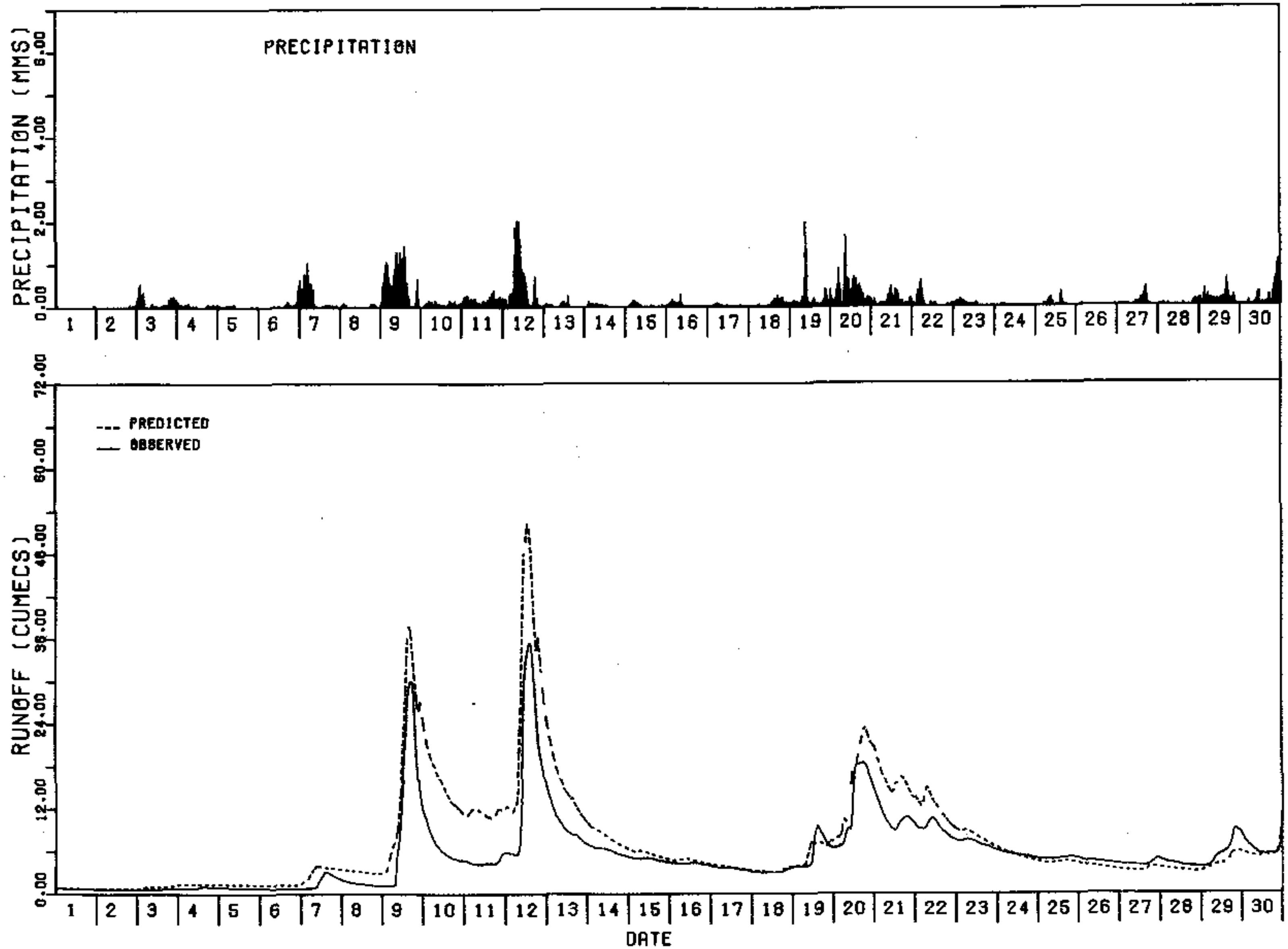


Figure 8

STØRAGE ØUTFLØW RELATIØNSHIP LØG-LINEAR S VS Q

67/06 ALWEN AT DRUID NØV72-APR73 DECEMBER

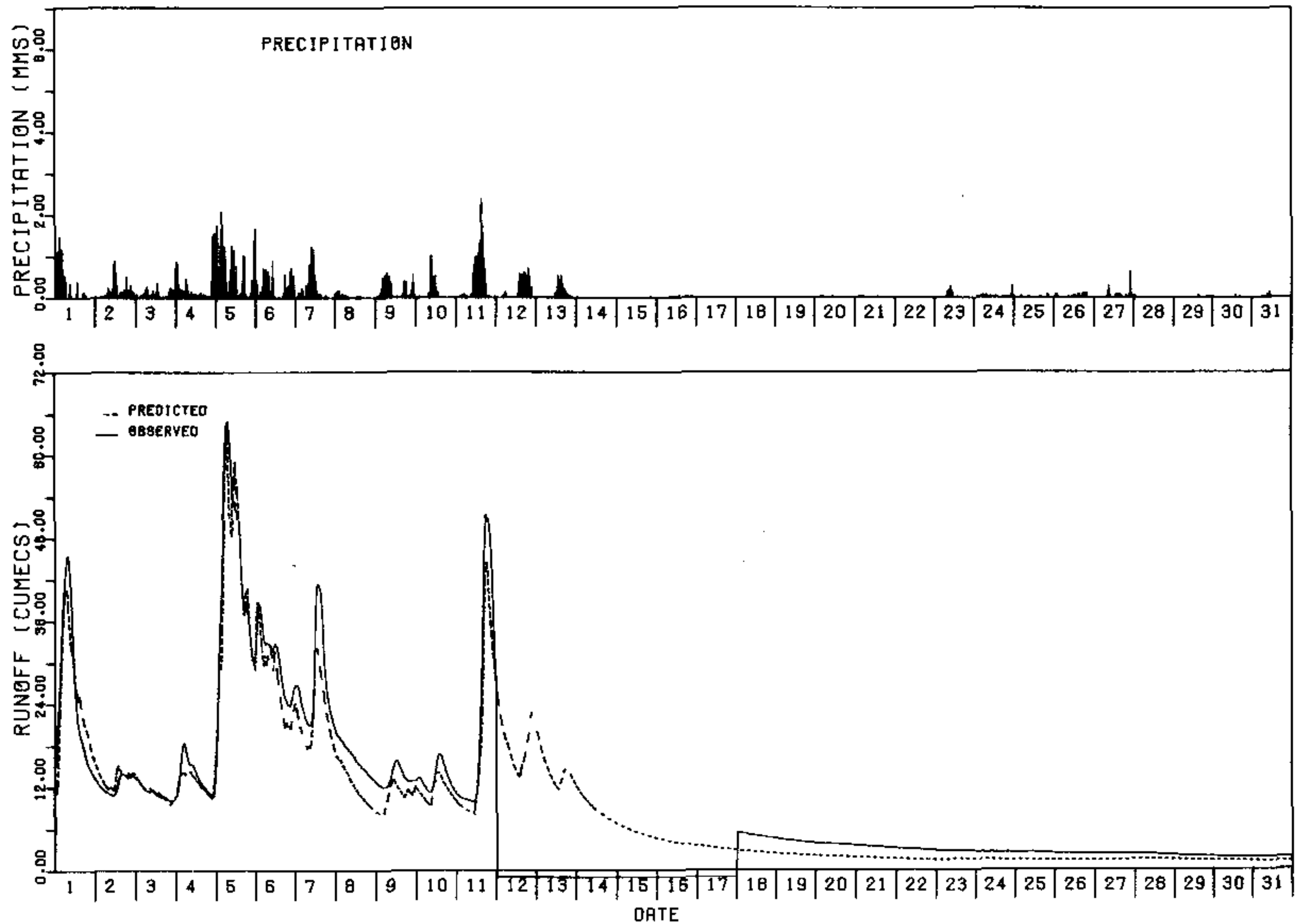


Figure 11

STORAGE-OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/13 HIRNANT NOV72-APR73

DECEMBER 72

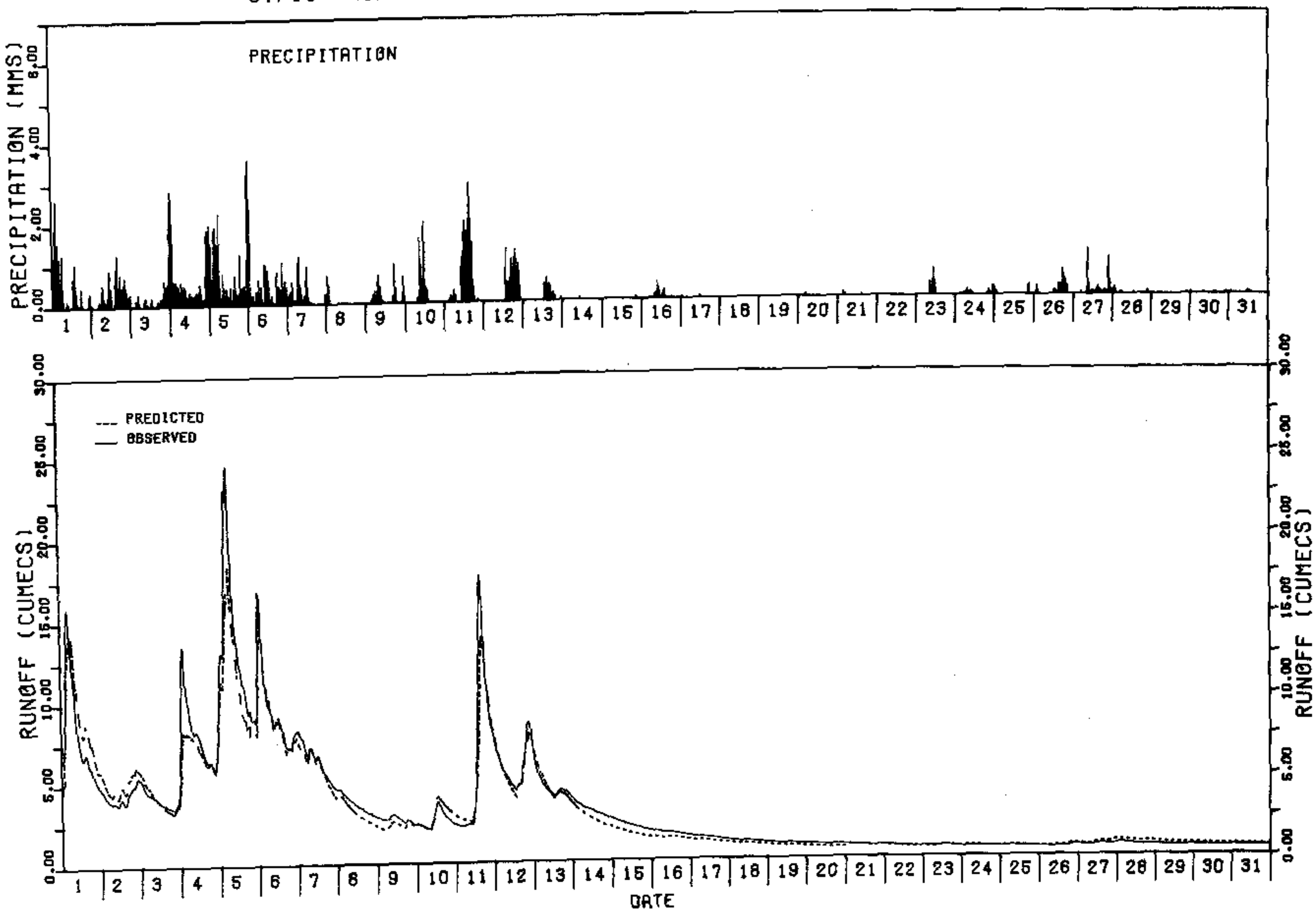


Figure 12

STORAGE-OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/10 DEE AT NEW INN

FEBRUARY 74

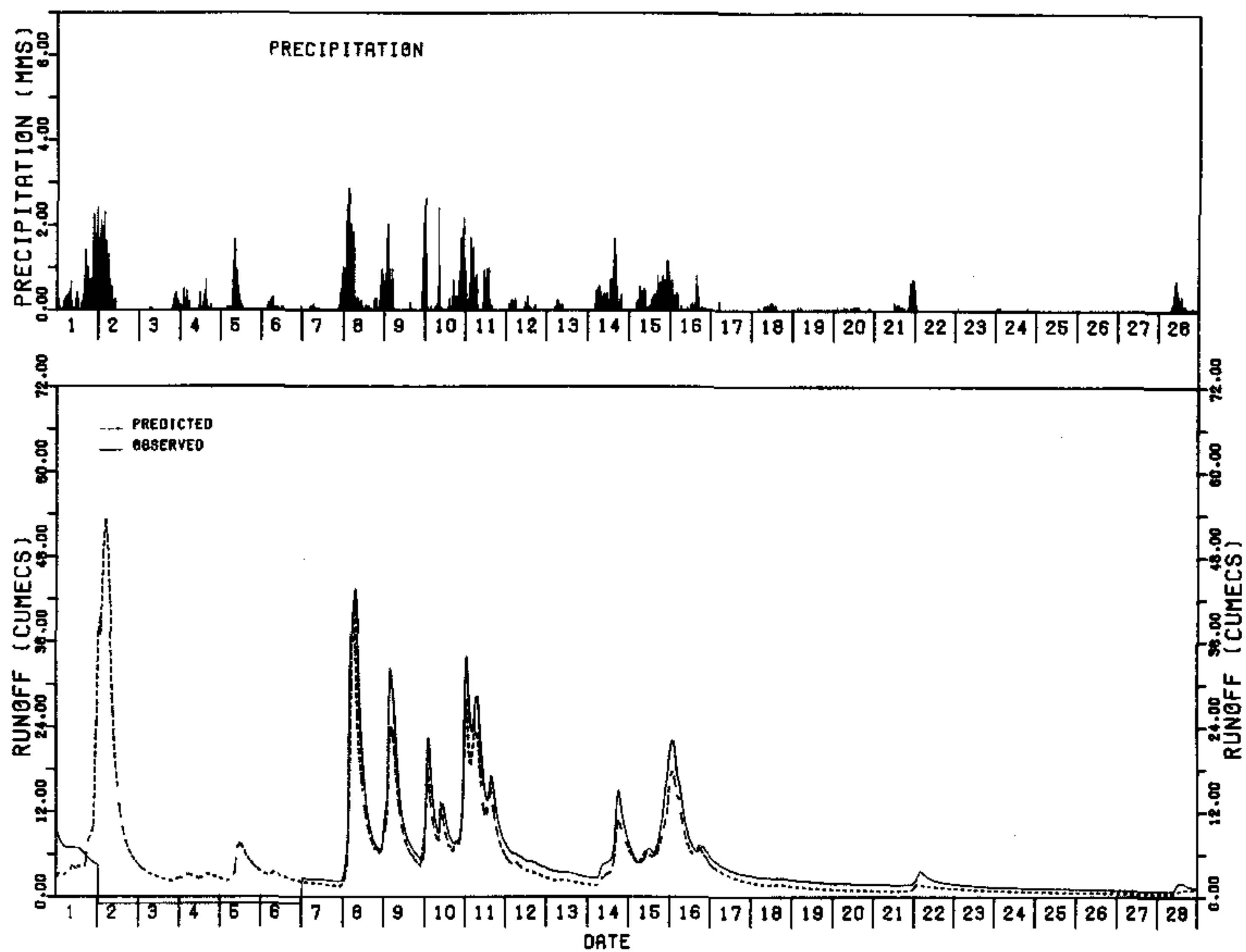


Figure 13

STORAGE-OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/10 DEE AT NEW INN

MARCH 74

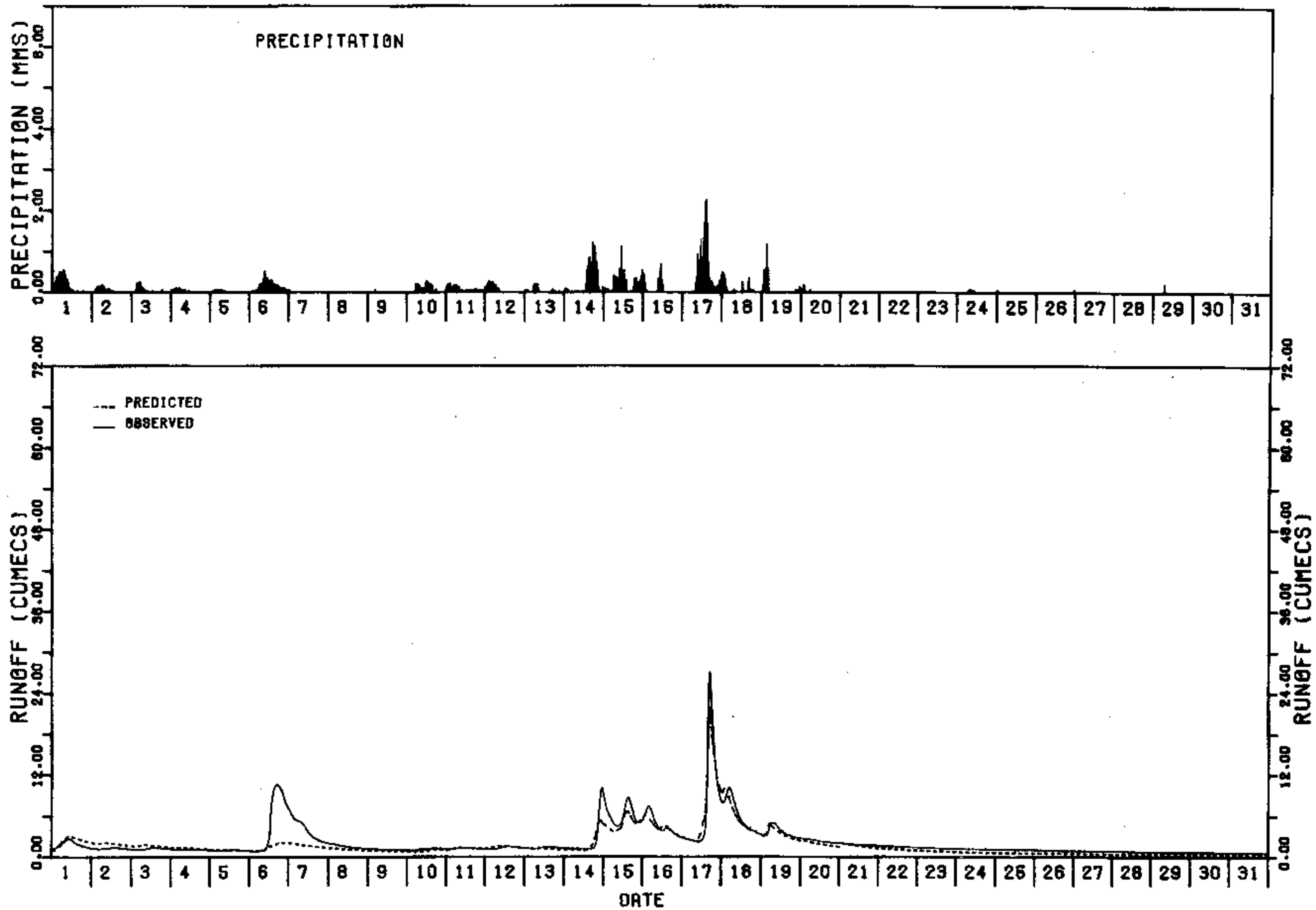


Figure 14

STORAGE OUTFLOW FN LOG-LINEAR FN

67/05 CEIRIOG AT BRYNKINALT WEIR

SEPTEMBER 73

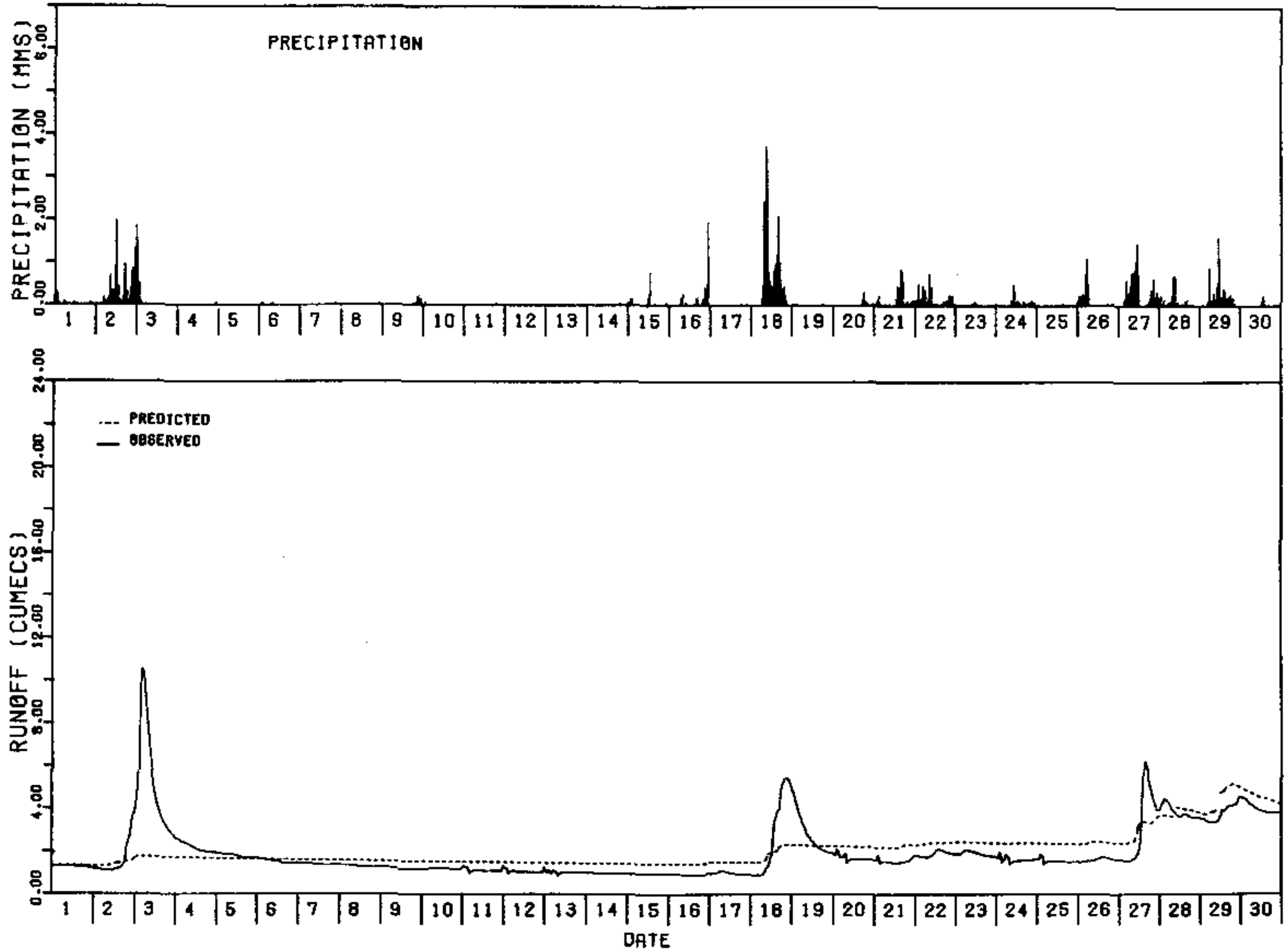


Figure 15

STORAGE OUTFLOW FN LOG-LINEAR FN

67/05 CEIRIOG AT BRYNKINALT WEIR

OCTOBER 73

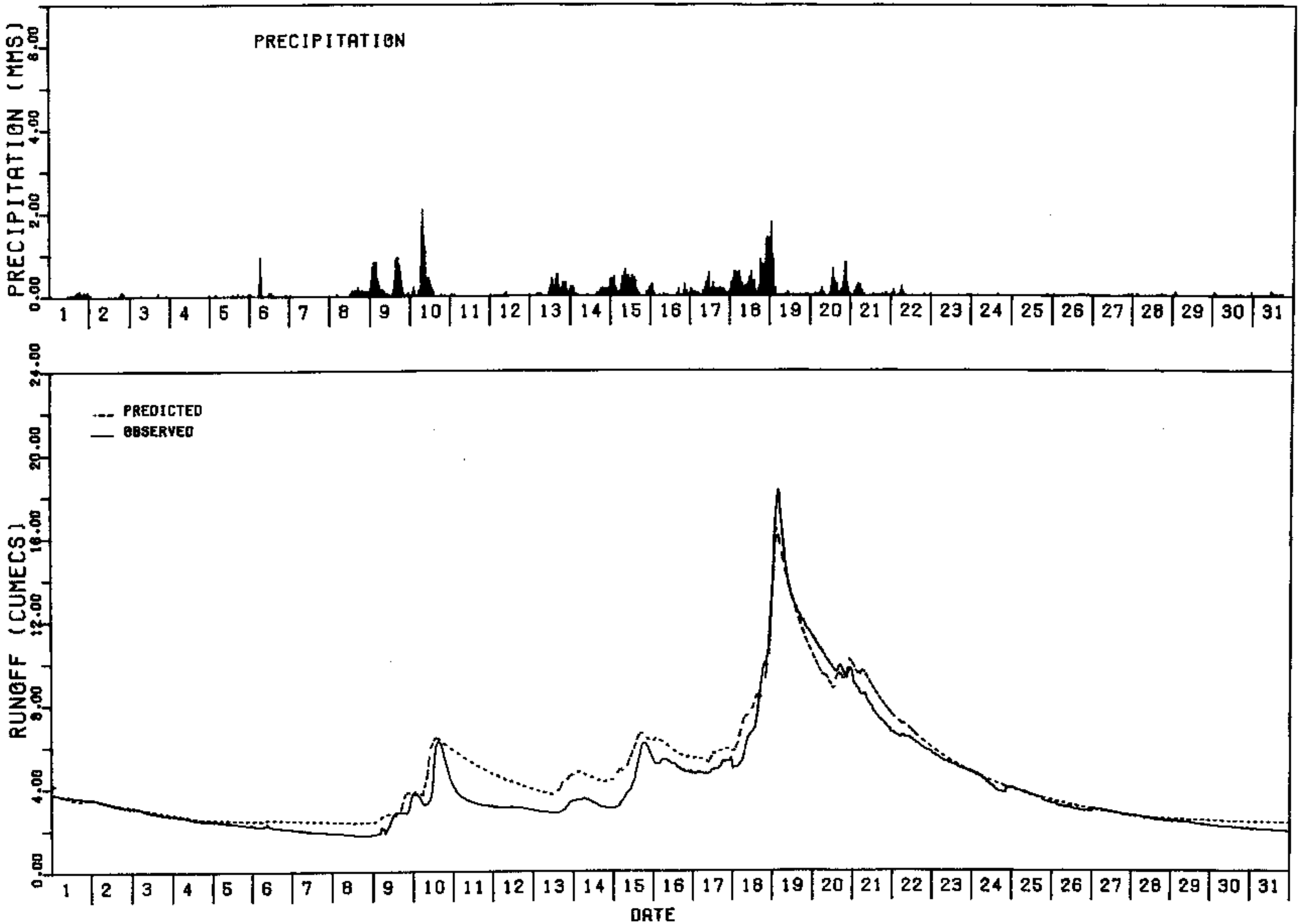


Figure 16

STØRAGE ØUTFLØW RELATIØNSHIP LØG-LINEAR S VS Q

67/10 GELYN AT CYNEFAIL SEPT72 ØCT72 MAY-JUL73 MAY-JUN74 MAY 73

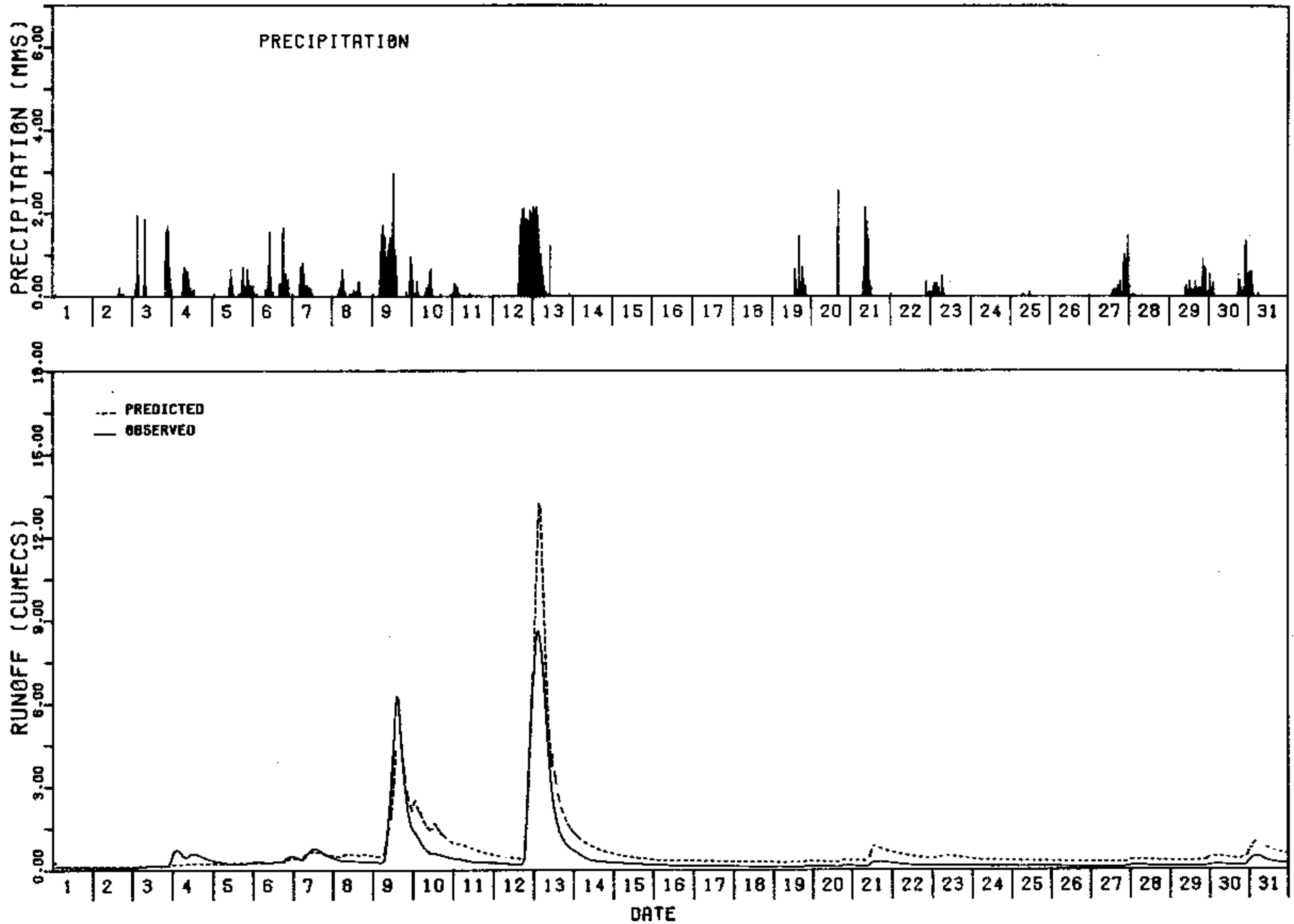


Figure 17

STORAGE OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/10 GELYN AT CYNEFAIL SEPT72 OCT72 MAY-JUL73 MAY-JUN74 JULY 73

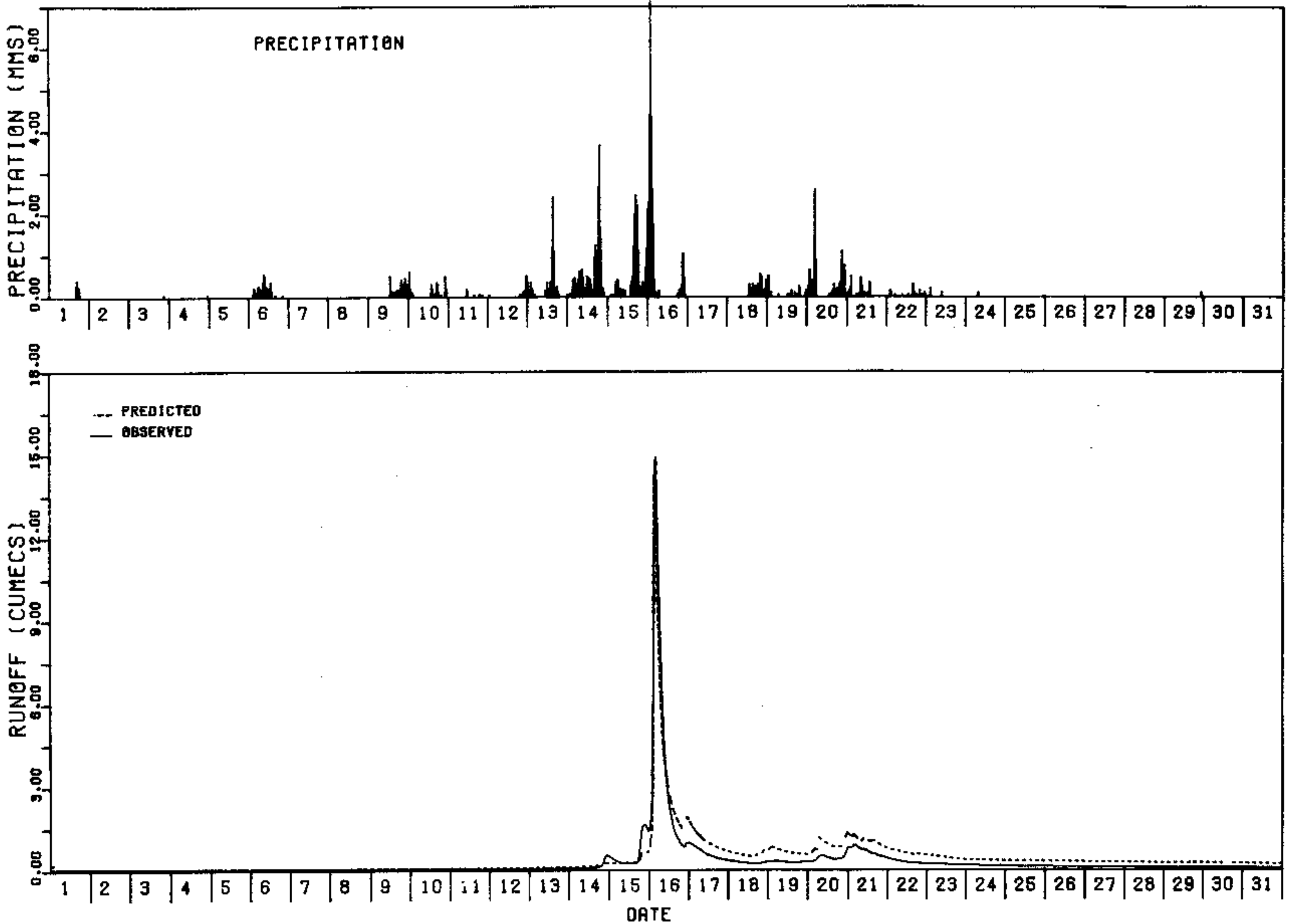


Figure 18

STORAGE OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/06 ALWEN AT DRUID MAY-JUN74 JUL-OCT72 AUGUST 72

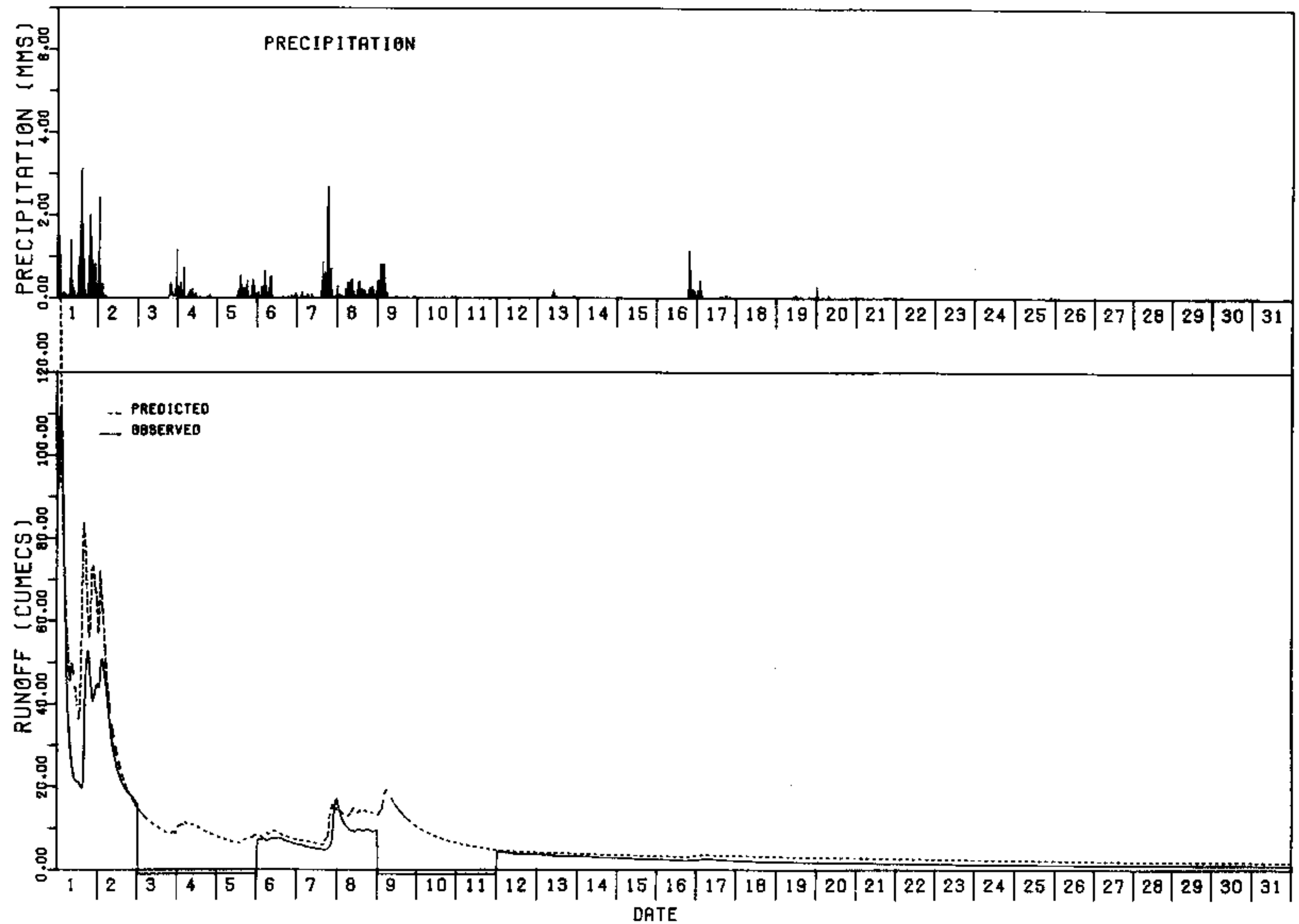


Figure 19

STORAGE-OUTFLOW RELATIONSHIP LOG-LINEAR S VS Q

67/13 HIRNANT MAY-JUN74 JUL-OCT72

AUGUST 72

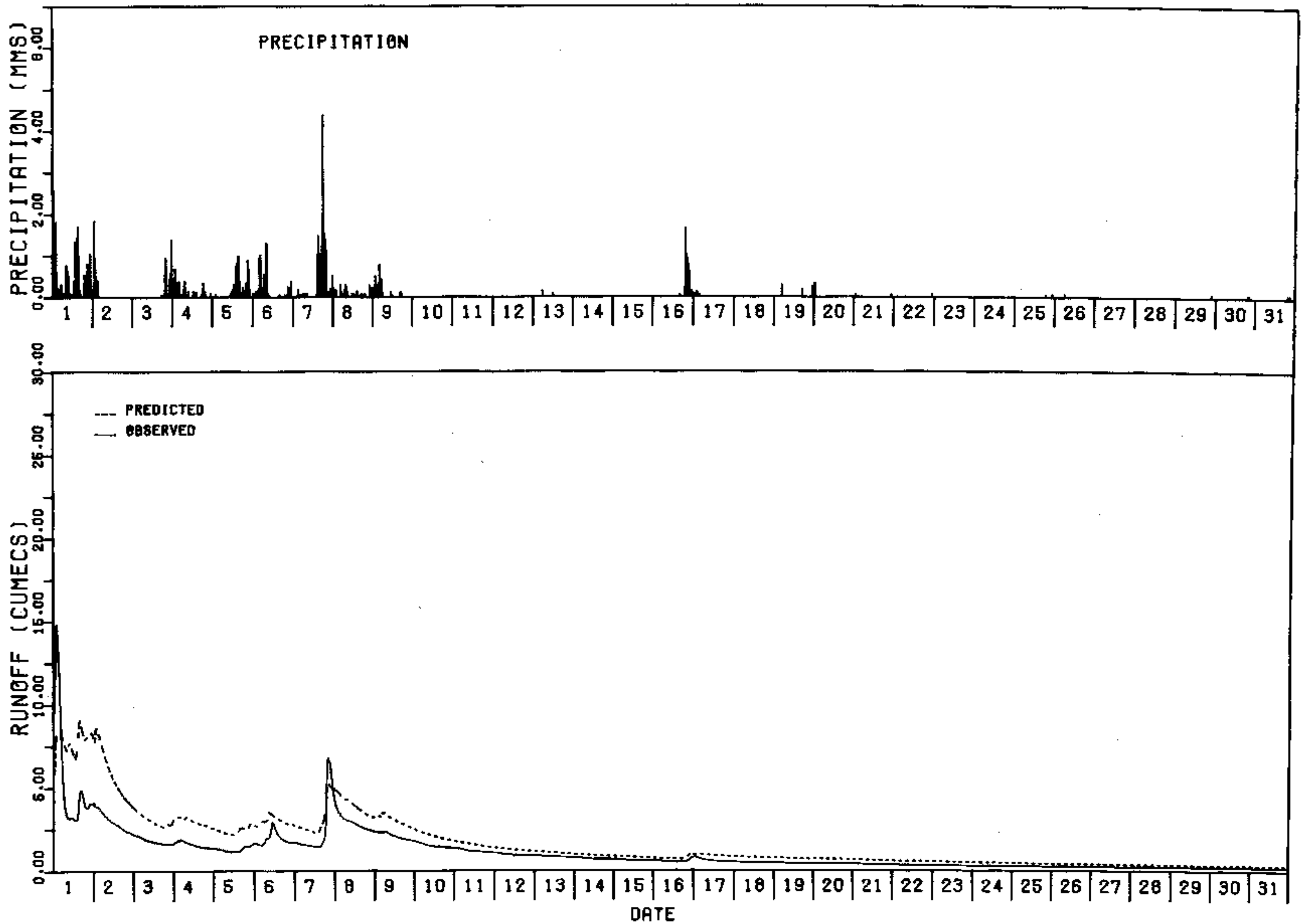


Figure 20

STØRAGE-ØUTFLØW RELATIØNSHIP LØG-LINEAR S VS Q

67/18 DEE AT NEW INN

JUNE 74

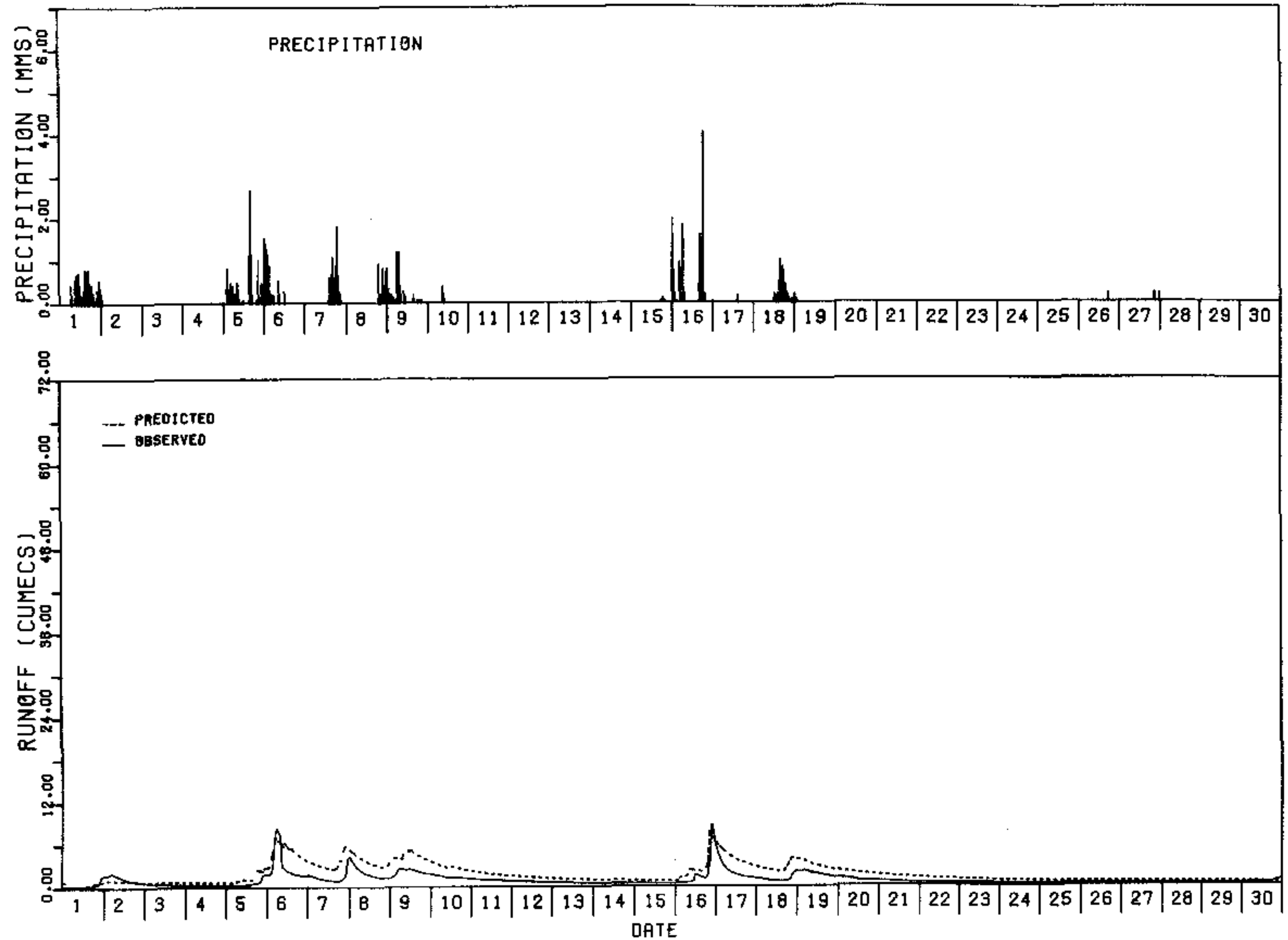


Figure 21

SIMULATED REAL-TIME OPERATION

67/06 ALWEN AT DRUID NOV72-APR73

NOVEMBER

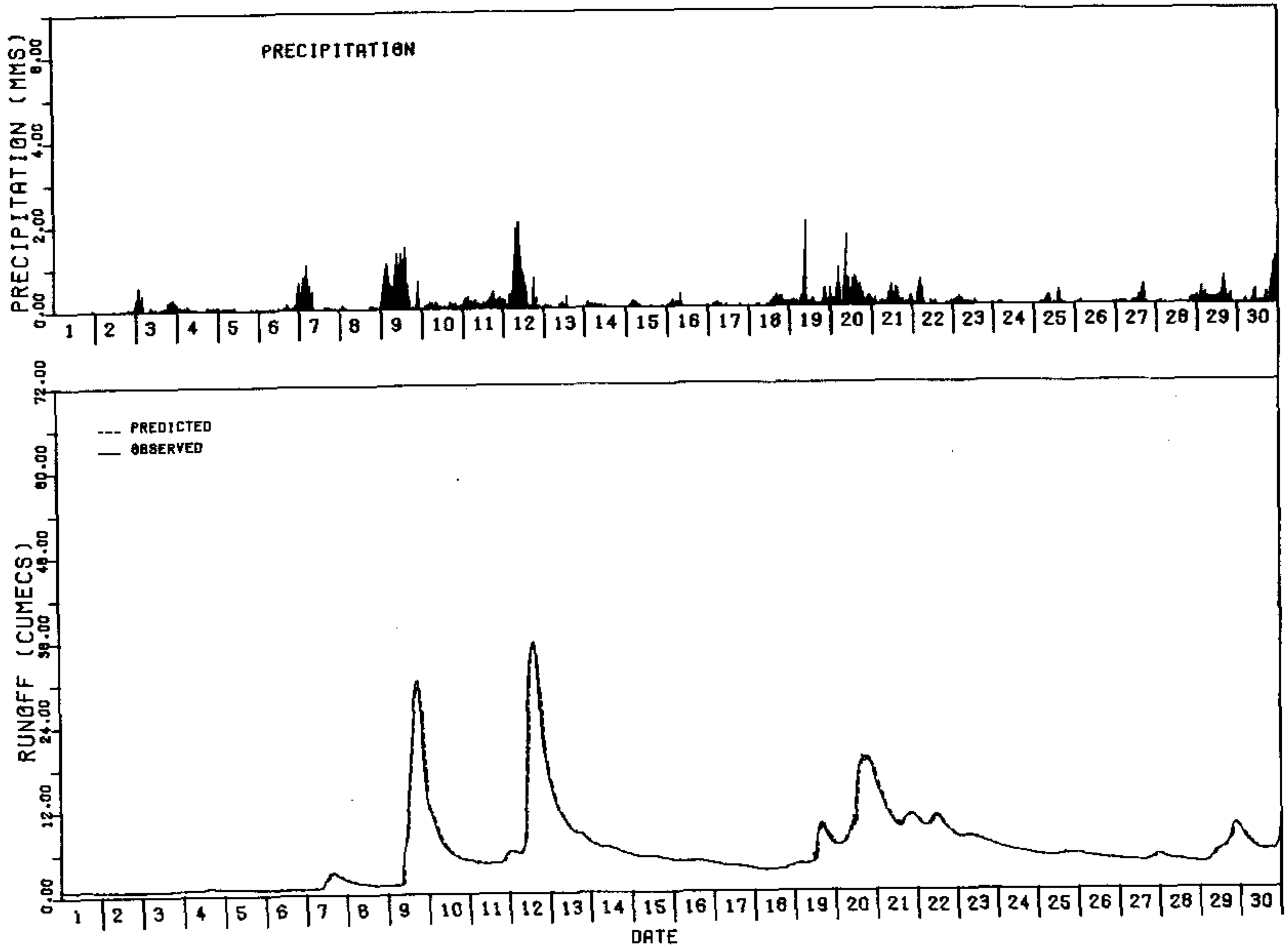


Figure 22

SIMULATED REAL-TIME OPERATION

67/06 ALWEN AT DRUID NOV72-APR73

DECEMBER

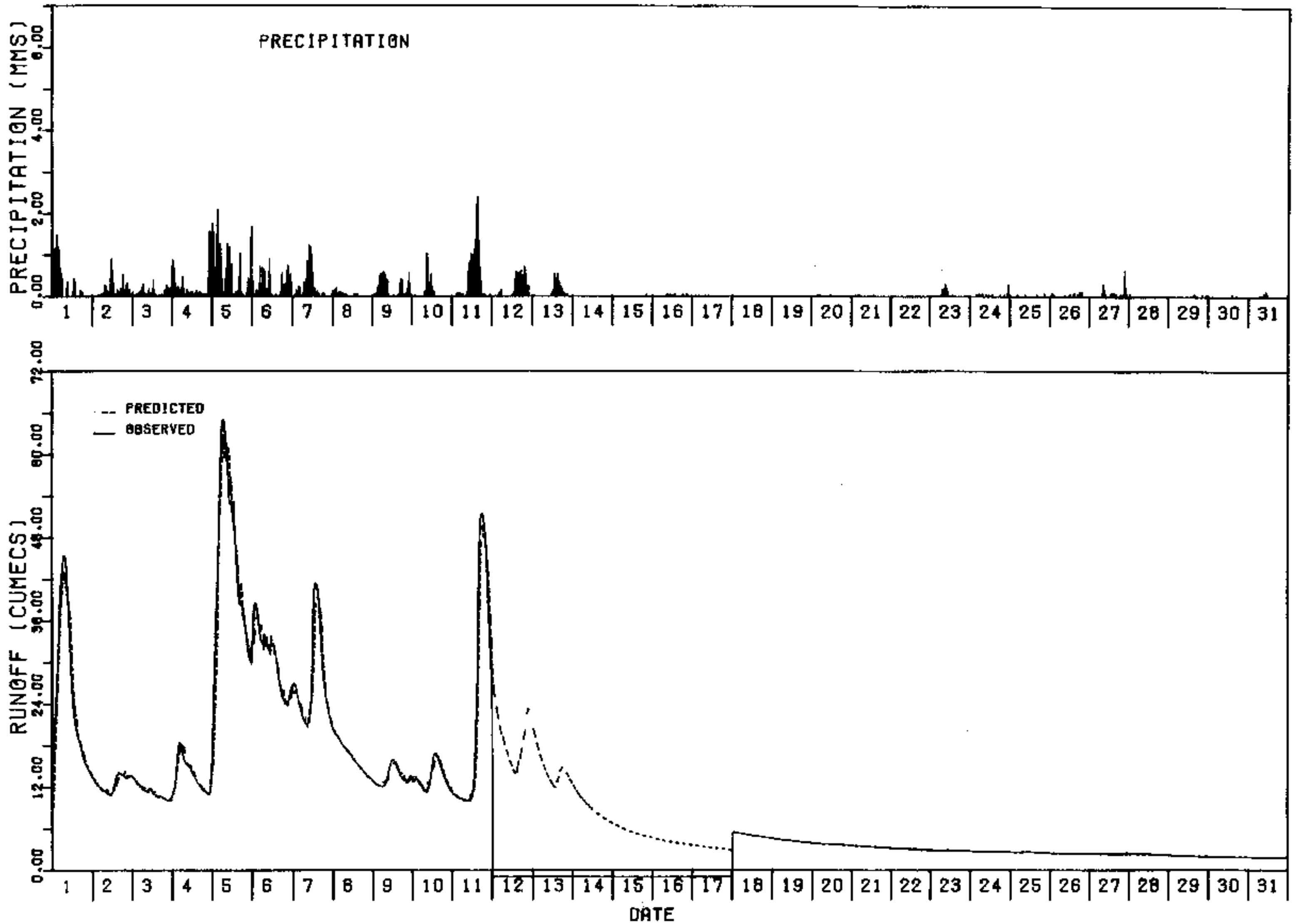
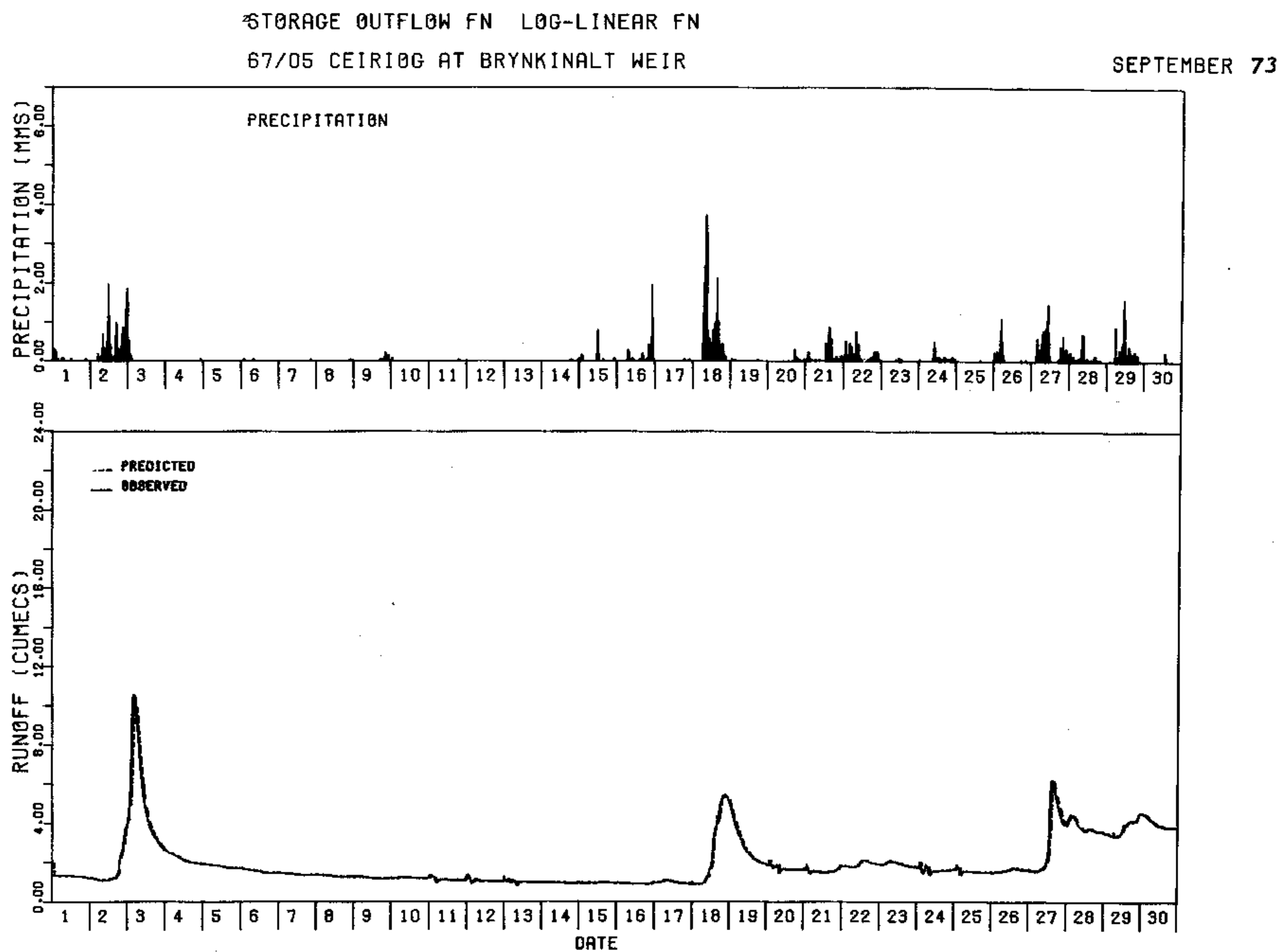


Figure 23


STORAGE OUTFLOW FN LOG-LINEAR FN
67/05 CEIRIOG AT BRYNKINALT WEIR
SEPTEMBER 73
PRECIPITATION
PRECIPITATION (MMS)
0.00
2.00
4.00
6.00
PREDICTED
OBSERVED
RUNOFF (CUMECS)
0.00
4.00
8.00
12.00
16.00
20.00
24.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
DATE

Appendix 1: ARCHIVING OF RAINFALL AND FLOW DATA AT INSTITUTE OF HYDROLOGY

Flows for nine gauging stations on the Dee and the rainfall data from 67 recording raingauges are stored on a series of magnetic tapes held at the IH Computer Centre. This appendix describes the layout, format, and the steps necessary for the retrieval of the data.

1. Rainfall data: Rainfall data for the 67 Plessey recording raingauges collected for the Dee Weather Radar Project was supplied by the Meteorological Office on two 9-track 1600 b.p.i. tapes DR 1972 and DR 1973. These cover the interval July 1972 - June 1974 in quarter-hourly steps. The organization of data is described in letters by C G Collier in the IH correspondence file HP 27/2/4 dated 29th November 1974 and 8th January 1975. The tapes were written by an IBM machine and the assistance of IH Computer Staff is needed to read them.

2. Flow data: Flows for nine Dee catchment gauging stations were obtained from paper tapes. These paper tapes were produced by processing Fischer and Porter tapes at the Water Data Unit. The data are stored as files 1 ... 9 on the IH Univac tape CO088 (with CO089 as reserve). Files 10 ... 12 on these tapes are the estimated half-hourly catchment rainfalls for the nine catchments.

The files on the tapes CØØ88 and CØØ89 are in the following order:

File Number n	Station Name	Example of Element Name	Duration of Data
1	67/01 Dee at Bala	.BALA73	Feb 70 - Aug 74
2	67/05 Ceiriog at Brynkisalt Weir	.CEIRIOG73	Nov 69 - Aug 74
3	67/06 Alwen at Druid	.ALWEN73	Dec 69 - Aug 74
4	67/13 Hirnant at Plas Rhiwaedog	.HIRNANT73	Jan 70 - Aug 74
5	67/18 (Upper) Dee at New Inn	.NEWINN73	Aug 72 - Aug 74
6	67/15 Dee at Manley Hall	.MH73	Feb 70 - Aug 74
7	67/17 Tryweryn at Llyn Celyn	.TRYWYN73	Sep 70 - Aug 74
8	67/03 Brenig at Pont-Y.Rhuddfa	.BRENIG73	Jan 69 - Aug 74
9	67/10 Gelyn at Cynefail	.GELYN73	Jan 72 - Aug 74
10	Catchment rainfalls for July - December 1972		
11	Catchment rainfalls for Jan - December 1973		
12	Catchment rainfalls for Jan - June 1974		

Reading the nth file from tape to a file NFILE requires the following UNIVAC job control statements:

```
@ASG,UP          NFILE
@MSG             **TAPE CØØ88 WITHOUT RING PLEASE**
@ASG,T           DEE*MT88., 12N, CØØ88
@MOVE            DEE*MT88., i      (where i = n-1)
@COPY,G          DEE*MT88., NFILE
```

If desired, a check on contents of the tape can be made using the following statements to print the directory of element names for each file,

```
@REWIND          DEE*MT88
@USE             TOC., DEE*MT88
@IHLIB5*PROGS    TAPTOC
```

Data for each flow gauging station is divided into yearly sets or elements, normally commencing at 0900 GMT 1st January. The directory for each file holds the names of the elements which are abbreviations of the station name and the year. Within an element, the first line gives a descriptive heading, the next 13 lines cover the 1st day, the second 13 lines the 2nd day, etc. The formats of the data are as follows;

	Line	Format	Data contained
	1	2ØA4	heading for element
	(2	I2,19A4	Day, Month, Year
Day 1	(3 ... 14	8F1Ø.3	96 ¼-hrly flows (cumecs) commencing from 0900 GMT.
	(15	I2,19A4	Day, Month, Year
Day 2	(16 - 27	8F1Ø.3	96 ¼-hrly flows (cumecs)
	etc		

Estimates of catchment rainfalls are arithmetic means of the catches for the gauges operating within, or close to, the boundary of the catchment. Half-hourly estimates are in file 10 for July - December 1972, file 11 for January - December 1973, and file 12 for January - June 1974. These three files are not subdivided into elements. Data for successive half hours are written on successive lines; each line is read with the format (I1Ø, 9F6.2). The first number (integer) is the minute count (from 0000 GMT, 1.1.71) at the start of the half hour interval and the following nine numbers (real) are the estimated catchment rainfalls (mm) for the half hour interval. The catchments are in the following order across the line;

1. Dee at Bala
2. (Upper) Dee at New Inn
3. Ceiriog
4. Alwen at Druid
5. Gelyn
6. Hirnant
7. Dee at Manley Hall
8. Tryweryn at Llyn Celyn
9. Brenig.

Note that the flow data, particularly for the earlier years 1969 and 1970, may contain occasional errors due to omitting characters when reading paper tapes: these are easily checked on listings of the data. Contiguous flows and rainfalls from this tape were written on to another pair of tapes in a form more suitable for modelling studies. Also, all of the contiguous data and most of the remaining flow data were plotted as a check for gross errors.

3. Combined flow and rainfall data: For the standard period 1 July 1972 - 30 June 1974, the half hourly rainfall and flow data were combined for each station and written into nine files on tape CØ1Ø7 (and tape CØ1Ø8 as reserve). Each file is made up of one element containing all 24 months of data. The Univac control instructions for reading the tape are as given before but with CØ1Ø7 substituted for CØØ88 and MT1Ø7 substituted for MT88 and n to be read from the table below.

For the five subcatchments used in the modelling study the data has been edited into smaller elements. The names of these elements denote the inclusive months and the corresponding files were rewritten on to the tapes CØ1Ø7 and CØ1Ø8 as files 10 - 14. Unfortunately the order of the stations on these tapes differs from the tape CØØ88. Note that the mass storage file into which tape files 10 - 14 are copied should be assigned with at least 180 tracks; this is greater than the IH standard of 128 tracks. The order is as follows:

File Number n	Station Name	Element Name
1	Ceiriog	.CEIR
2	Dee at New Inn	.NEWINN
3	Dee at Bala	.BALA
4	Tryweryn at Llyn Celyn	.TRYWERYN
5	Alwen at Druid	.ALWEN
6	Brenig	.BRENIG
7	Dee at Manley Hall	.MH
8	Hirnant	.HIRNANT

File Number n	Station Name	Element Name
(ctd) 9	Gelyn	.GELYN
10	Ceiriog)	Example of an element name is
11	Alwen)	.NOV72APR73; this element includes
12	Gelyn)	the months from November 1972 to
13	Hirnant)	April 1973.
14	Dee at New Inn)	

The format of the data is as follows:

Line	Format	Information contained
1	20A4	Heading
1st (2	I2,19A4	Day, month, year
day (3 - 14	8F10.3	48 successive pairs of ½-hrly rainfalls (mms) and end-of-interval flows (cumecs), commencing from the interval ending at 0900 GMT.
(15	I2,19A4	Day, month, year
(16 - 27	8F10.3	48 pairs of ½-hrly rainfalls and flows
etc.		

Missing data, either flows or rainfalls, is substituted by minus unity. Invalid days in the modelling elements (files 10-14) are flagged by setting the day number to minus unity: this was done for the Alwen record for the days when the Alwen Reservoir overflowed.

Appendix 2: NOTES ON RETRIEVAL AND USE OF COMPUTER PROGRAMS

This appendix describes the retrieval of the file of the subcatchment model and the time series analysis programs from magnetic tape, the tasks that they perform and the setting-up of input data.

Catchment Model Programs

The model file is stored as the seventh file on tape AØØØ7 (with AØØ46 as reserve) at the IH Computer Centre. The file can be retrieved with the following instructions:

```
@ASG,AX      MODEL
@MSG,W       **TAPE AØØØ7 WITHOUT RING PLEASE**
@ASG,T       FS*MTØ7,12N,AØØØ7
@MOVE        FS*MTØ7,6
@COPY,G      FS*MTØ7,MODEL
@FREE        FS*MTØ7
```

The main program is an extensively modified version of the routines described in J R Douglas's IH Report No 24, but with Lambert's storage - outflow model substituted for the Douglas conceptual model. Despite the modifications, the format for submitting data to the program has been kept as close as possible to Douglas' original. The program is designed to carry out catchment modelling with up to 10416 data points (seven 31-day months of flows and rainfalls with 48 readings per day). All the modelling data is read into the core of the computer at once instead of using blocks of monthly units and reading and re-reading as in the original. Placing all the data into core has given a substantial reduction in the computing time requirements of the optimizing routine.

The tasks the program will perform are

1) determine the "no-model" sum-of-squares

$$F_o = \Sigma\ (q_t - \bar{q})^2$$

2) For an assumed storage-outflow function and specified initial parameter values, determine the parameter values giving the minimum of the sum-of-squares function.

3) For given parameter values estimate the flow hydrograph and draw a CALCOMP plot comparing the estimated with the observed.

4) Within a specified grid of parameter values compute the sum of squares function.

The program is formed from a main routine called CONTRO. This calls a large subroutine OPTION which in turn calls further subroutines MODEL (not to be confused with the file MODEL) and AUG, and PLOTER or PLOTRE if CALCOMP plots are required. When these routines have been compiled, the program is prepared for execution with the following statements:

```
@PREP       MODEL
@MAP,IS     MODEL.BIN
IN          MODEL.CONTRO
LIB         MODEL
LIB         IHLIBS*PLOT
END
```

Execution requires the following:

```
@MSG        **CALCOMP PLOTS PLEASE**)
@ASG,TJ     GRAPHTAPE,8C            ) only if plotting
@USE        8,GRAPHTAPE             )
@XQT        MODEL.BIN
@ADD,P      MODEL.CONTROLELEM
@ADD,P      DATA.DATA
```

The first @ADD refers the program to an element of data - CONTROLELEM in which is set out parameters describing the model data and parameters controlling the program options. This was set up as an element of a file, but it could be submitted directly (on cards if batch processing) after the @XQT statement and replacing the first @ADD statement. The second @ADD statement refers the program to the appropriate element of hydrological data (.DATA). The organization of this data for the Dee subcatchments and its retrieval from tape CØ1Ø7 is described in Appendix 1.

Two CALCOMP plotting programs are available, both of which plot blocks of monthly flow and precipitation data. One routine called PLOTER draws the observed and computed hydrographs in full lines of different colours. The other, PLOTRE uses one colour and draws the hydrographs as full and dashed lines of different thickness and gives less detail on the axes. The hydrographs shown in the report were drawn by this routine. The routines would require some modification to plot data with frequencies of other than 48 readings per day.

The elements of control data for most of the model runs for summer and winter data for the different subcatchments were stored in the model file and can be identified by the inclusion of the letters CON in the element name. Thus, for example, the element ALCONSUM is the control data for summer data for the Alwen subcatchment. The control data are set out in card format as follows:

Card	Symbol	Format	Columns	Comments
1	JJ(1) ... (10)	10A6	1...60	Description of model
2	JJ(11) ... (20)	10A6	1...60	Name and location of catchment
3	JJ(21) ... (30)	10A6	1...60	Duration of data
4	II(1)	I1	4	Mode of operation: = 1 optimization = 2 prediction = 4 surface mapping = 5 "no model" sum of squares = 6 simulated "real time" operation, recentring at every ½ hour
	II(2)	I1	8	Not used
	II(3)	I1	12	Set = 1
	II(4)	I1	16	Not used
	II(5)	I1	20	Controls graphical output = 0, no graphs = 1, plot with PLOTER = 2, plot with PLOTRE
	II(6)	I4	21...24	For optimization, maximum number of iterations required
	II(7)	I1	28	) Not used
	II(8)	I1	32	)
	II(9)	I1	36	Controls form of storage outflow function used = 0 assumes log-linear function = 1 function type to be specified
5	ISOTYP(1)	10I4	1 ...40	Included only if II(9) = 1 Type of function for each segment of storage-outflow fn = 1 log-linear

Card	Symbol	Format	Columns	Comments
5 (Cont'd)				= 2 linear The number of segments is given by MM(8)
6	MM(1)	I4	1 ... 4	Frequency of source data (readings per day)
	MM(2)	I4	5 ... 8	Desired time interval for model (readings per day)
	MM(3)	I4	9 ...12	Not used
	MM(4)	I4	13 ...16	Number of months to be modelled (less than 8)
	MM(5)	I4	17 ...20	Position in year of starting month (Jan = 1, Feb = 2, etc)
	MM(6)	I4	21 ...24	Number of days to be modelled (less than 218)
	MM(7)	I1	28	= 0, months in calendar Order = 1, months not in order (Order is read into array LL)
	MM(8)	I4	29 ...32	Number of segments in storage-outflow fn (normally one or two)
	MM(9)	I4	33 ...36	)) not used
	MM(10)	I4	37 ...40	)
	LL(1)	I4	41 ...44	LL is used only when MM(7) = 1 Number of 1st month (Jan = 1, etc)
	LL(2) etc	I4	45 ...48	Number of 2nd month
7	RUNRNG	F10.5	1 ...10	Maximum runoff (cumecs) rounded upwards to a multiple of 6 for plotting.
	CAKM	F10.5	11 ...20	Catchment area (Kms^2)
8	N	I4	1 ... 4	Number of parameters to be included in optimization

Card	Symbol	Format	Columns	Comments
8 (ctd)	NN	I4	5 ... 8	Total number of parameters to be read
(8+1)	(PNAMES(J)	A6	1 ... 6	Parameter name
	(KK(J)	I4	7 ... 10	Order of inclusion in optimization if II(1) = 1
(8+NN)	(			Order for surface mapping if II(1) = 4
	(			Otherwise = 0
	YI(J)	F10	11 ... 20	Initial value of parameter
	(BB	F10	21 ... 30	Lower bound) only needed for
	CC	F10.0	31 ... 40	Upper bound) KK(J) = 0
Only for surface mapping |II(1)=4|				The following cards are needed in the order given by KK(J):
(8+NN+1)	(MNI	I2	1,2	Number of values of a parameter to be read for mapping
(8+NN+N)	(PM	F8.0	3 ...10	Parameter values for grid points
			11 ...18	

The ordering of the parameter cards is important: the first card specifies the catchment lag L (hrs), the second and following card give the K values for each specified segment of the storage outflow function. Finally for each segment specified there must follow a card giving the lower bound flow value (mm/hr) for that segment; thus the zero bound is always included, even for only one segment.

Time Series Programs

Two programs were used for the time series analysis of the Alwen errors. They are also placed in the file MODEL.

The first program carries out the univariate stochastic series identification procedure along the lines of program 1 in Box and Jenkins (1970).

This program is stored under the name .TSIDENT (time series identification) and can be implemented with the following statements:

```
@MAP,IS      MODEL.TSIDENTBIN
IN           MODEL.TSIDENT
END
```

```
@XQT        MODEL.TSIDENTBIN
@ADD        MODEL.RESIDS
```

The program produces the mean, variance and skewness of the series, the autocorrelation and partial autocorrelation functions. Data is supplied to the program through the @ADD statement which refers to the element .RESIDS. The element has the following form:

1st line (FORMAT I5)
N, the number of data prints
(maximum of 4000 in this version)
2nd and subsequent lines (FORMAT 8E10.4)
the series of data

The second program estimates for an assumed order of seasonal ARIMA structure for a univariate stochastic series, the AR and MA parameters. The method is the non-linear interative scheme described by Box and Jenkins (1970). The program is written as a main routine TSEST (time series estimation) and the subroutine UNCONS and LAZYB. An IH library matrix inversion routine (GJR) is also used. The program is set up and executed from compiled code with the following statements:

```
@PREP       MODEL
@MAP,IS     MODEL.TSESTBIN
IN          MODEL.TSEST
LIB         MODEL
LIB         IHLIBS*MATHSTAT
END
@XQT        MODEL.TSESTBIN
@ADD        MODEL.TSESTCON
@ADD        MODEL.RESIDS
```

The controlling element TSESTCON has the following structure

Line 1 (Format 8I5)
- 1st element, N, the number of data points
- 2nd element, p, the order of non-seasonal AR
- 3rd element, d, the order of non-seasonal differencing
- 4th element, q, the order of non-seasonal MA
- 5th element, P, the order of seasonal AR
- 6th element, D, the order of seasonal differencing
- 7th element, Q, the order of seasonal MA
- 8th element, NS, the periodicity of the seasons, (only if P and/or $Q \neq 0$)

Line 2 (Format 5F5.1), (only if $p > 0$)
initial estimates of the AR(1)...AR(p) coeffs

Line 3 (Format 5F5.1) (only if $q > 0$)
initial estimates of the MA(1)...MA(q) coeffs

Line 4 (Format 5F5.1) (only if $P > 0$)
initial estimates of the seasonal AR coeffs

Line 5 (Format 5F5.1) (only if Q > O)
initial estimates of the seasonal MA coeffs

The second element .RESIDS is as described above. Further details of these programs are available in McKerchar and Delleur, 1972.

References:

Box G E P and Jenkins G M, 1970, Time Series Analysis, Forecasting and Control, Holden Day, San Francisco, Calif.

McKerchar A I and Delleur J W, 1972, Stochastic Analysis of Monthly Flow Data: Application to Lower Ohio Tributaries. Tech Rep No 26. Purdue University Water Resources Research Center, Lafayette, Indiana.