Institute of Hydrology

1994/065

Natural Environment Research Council

Natural
Environment
Research
Council

PROJECT FD0404
HYDROGRAPH ESTIMATION PROCEDURES

Comparison of simple conceptual daily rainfall-runoff models

H. A. HOUGHTON-CARR & N.W. ARNELL

Institute of Hydrology
Crowmarsh Gifford
Wallingford
Oxfordshire
OX10 8BB
UK

Tel: 0491 838800
Fax: 0491 832256
Telex: 849365 Hydrol G

December 1994

Executive summary

Bonvoisin & Boorman (1992a) applied 16 simple, conceptual daily rainfall-runoff models to four test catchments to investigate the potential of daily rainfall-runoff modelling as an aid to the transfer of hydrological information between sites.

In this second phase of the study, six of the best of these models are applied to a new and larger dataset comprising 25 catchments covering a range of sizes, topographies, soils and climates. The models have between three and five parameters. Model performance is judged on the ability of the model to reproduce the basic characteristics of the observed flow regime, such as the average annual runoff total and the shape of the flow duration curve, rather than provide a near-exact simulation of the observed response.

The results show that four of the models perform well on the majority of catchments. A range of both objective and subjective measures of fit were applied to the simulation results. For a particular model on a particular catchment, the objective measures, apart from BFI, tend to give the same result; the subjective measures are less consistent. The objective and subjective measures can, but do not necessarily, give the same result. The distinction between models that performed well or badly was fairly consistent. There are certain catchments on which none of the models perform well. On some other catchments, all the models gave good results. These tended to be the baseflow-dominated catchments. The relatively good performance on these catchments was almost certainly because of the smaller variability of their day-to-day flows, and because the faster responding, more impervious catchments were not well represented using a daily time step.

A regression exercise was carried out in order to derive statistical relationships between the model parameters of the four models which performed well and mapped physical and climatic characteristics of the catchments. This exercise gave disappointing results, reflecting the restricted set of variables used in the FSR catchment characteristics dataset, and highlighting the need to extend this dataset to include characteristics which can now be derived by computerised methods.

Future work will involve using model C, the probability-distributed model of Moore (1985), to assess the potential for better estimation of flood frequency through continuous simulation of catchment flows (Spijkers & Naden, 1994). Furthermore, it is likely that the ability of models to fit the observed flow data was constrained by the time step adopted here; for instance, flood peaks may have been modelled better had hourly data rather than daily data been used. Therefore, the future work will also investigate the potential for continuous simulation at a sub-daily time step, as well as the possibilities of using new digital datasets for deriving model parameters.

Contents

1. Introduction

Daily rainfall-runoff models are common in hydrology, and Fleming (1975) reviews many of the existing models. Such models are usually designed with the aim of closely simulating the catchment response, and tend to be used to generate long sequences of synthetic flow data from rainfall data, or to assess the effect of changes to the catchment on the flow regime. Inevitably this often results in a large number of model parameters. However, by relaxing the goodness-of-fit criteria, simpler models, with few parameters, may be used. Rather than providing a near exact simulation of the observed response, these models reproduce the basic characteristics of the flow regime, such as the average annual runoff total and the shape of the flow duration curve.

Such simple, conceptual models have potential for application in flood estimation techniques. The Flood Studies Report (FSR; NERC, 1975) uses an event-based rainfall-runoff model, known as the unit hydrograph and losses model, for estimation of the design flood hydrograph. This model has three parameters: the unit hydrograph time-to-peak which determines how quickly the catchment responds to the effective rainfall input, the percentage runoff which is the ratio of effective rainfall to total rainfall i.e. the proportion of the total rainfall input which becomes response runoff in the river, and finally, the baseflow which represents the flow in the river prior to the event.

The percentage runoff component is composed of standard and dynamic parts; the former representing the normal capacity of the catchment to generate runoff, and the latter representing the variation in runoff dependent on the state of the catchment prior to the storm, and on the storm magnitude. There may also be an urban adjustment. The percentage runoff is the most important component of the unit hydrograph and losses model, yet the most poorly estimated. Thus, a priority in the enhancement of rainfall-runoff modelling for flood estimation is better definition of runoff volumes. Poor representation of catchment storage prior to a rainfall event is one of the main shortcomings in percentage runoff estimation. This can be improved through continuous simulation of runoff from rainfall using a simple, conceptual model. The model parameters control the model output from which the actual percentage runoff can be determined.

This report describes a study, carried out as part of MAFF project FD0404 on hydrograph estimation procedures, to compare several simple, conceptual daily rainfall-runoff models and select one for use in future work. The principal objective was to obtain a simple model with few parameters which could be applied to a large number of gauged catchments, in order to facilitate the transfer of hydrological information between sites, to assist calibration of the HOST (Hydrology Of Soil Types) dataset (Boorman *et al.*, 1994), and to examine the changes in hydrological regime resulting from changes in climate (Arnell & Reynard, 1993).

After this introduction the report is divided into eight sections. The next section considers the background to the work. Section 3 describes the models used in the study and their parameters, and Section 4 gives details of the test catchments plus the data requirements. This is followed by a section considering the methodology. Section 6 describes the results of the study, and Section 7 describes attempts to relate model parameters to catchment and climatic characteristics. Conclusions are drawn in Section 8.

2. Background

This work is the second phase of a study started by Bonvoisin & Boorman (1992a) which investigated the potential of daily rainfall-runoff modelling as an aid to the transfer of hydrological information between sites. Bonvoisin and Boorman applied 16 daily rainfall-runoff models to four test catchments, and assessed the model performance using a specially developed modelling framework (MIMIC; Bonvoisin and Boorman, 1992b). In this phase, the best of the models are developed further, and applied to a new and larger dataset comprising 25 catchments.

MIMIC (Microcomputer-based Interactive package for Model Identification and Calibration) was developed as a framework for daily rainfall-runoff modelling, though in practice any timestep may be used. The user is provided with all the graphical and statistical information necessary for assessment of a rainfall-runoff model and its calibration. The facilities provided by MIMIC include plots of flow hydrographs together with rainfall input, plots of flow duration curves, figures on flow volumes and summary flow statistics. It also allows for both automatic optimisation and manual adjustment of parameters. Whilst the latter facility for manual adjustment incorporates a subjective element into the modelling, the user can rapidly assess the model performance and can gain a good understanding of parameter interaction and sensitivity.

3. Models

Six daily rainfall-runoff models were used in the comparisons, comprising four different models and variants on two of the models. All of the models were simple, conceptual flow generation models with inputs of daily rainfall and daily potential evaporation. Each model consisted of a number of soil moisture stores, with the model parameters controlling the store sizes and the rate of flow from the stores. The models had four basic components:

- *A procedure to determine actual evaporation (AE) from potential evaporation (PE).* The ratio of AE to PE is generally taken to be a function of the water content of one of the soil moisture stores. Some models use a linear function (i.e. a linear decline in evaporation as soil moisture content falls below some maximum), whilst others use a negative exponential function (i.e. the ratio of AE to PE falls slowly at first, but more rapidly as the store empties).

- *A storage accounting procedure to determine the water content of each soil moisture store.* Store content at the end of a timestep is based on the content at the beginning of the step (i.e. at the end of the previous timestep), and on inflow and outflow during the step. The outflow from one store is usually the inflow to another store. Different models have different procedures for determining outflows, usually within prescribed limits e.g. some stores can overflow, whilst others can only drain downwards. Different models have different numbers of stores which may be combined in different ways. In some cases, the store properties may also vary across a catchment.

- *A runoff generation procedure to convert precipitation into runoff.* This is either as direct surface flow or as baseflow. The former is usually through a saturation excess or infiltration excess model, and the latter as a function of the soil moisture store content.

- *A procedure to route the outflow from appropriate soil moisture stores into flow in the river.* This is usually based on a system of linear reservoirs, one from each store.

There are a great many possible combinations of these four basic components, and the following sections consider each of the six models in detail. The model parameters are summarised in Table 3.1 at the end of this section.

3.1 MODEL A

Model A (Figure 3.1) uses a soil moisture deficit (with only one bound of zero), rather than a soil moisture store (with a bound of zero and some maximum). Precipitation is added to the soil decreasing the deficit. Evaporation and subsurface flow occur from the soil increasing the deficit. If the soil becomes saturated the excess precipitation becomes overland flow. The subsurface flow and overland flow are summed and routed through a linear reservoir to become the catchment outflow. The original model on which this is based (Bonvoisin & Boorman, 1992a; model 8) included an additional parameter whereby, if the precipitation exceeded a certain rate, overland flow was generated. However, in practice, this parameter was found to be redundant.

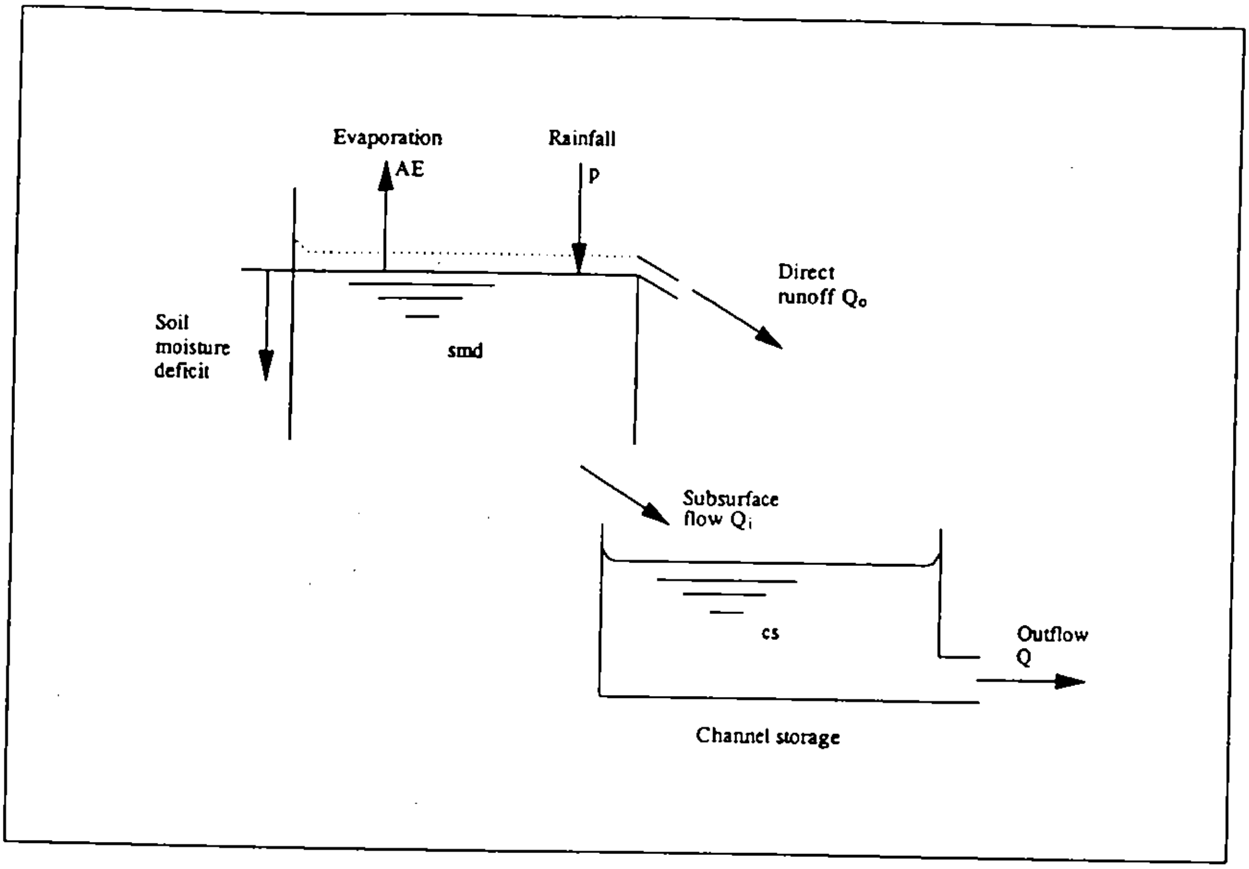


Figure 3.1 Model A

The model has five parameters: an evapotranspiration coefficient *Ca*, subsurface flow coefficients *Cb* and *Cc*, and routing coefficients *Cr1* and *Cr2*. Bonvoisin & Boorman (1992a) tried combining the two routing parameters, but this was found to give poorer results. The model works through the accounting procedure in the following stages:

i. Determine the soil moisture deficit, smd (units expressed as a positive value), after the rainfall, p:

$$smd_t = smd_{t-1} - p_t$$

where: t and t-1 refer to the present and previous days, respectively.

ii. Determine the overland flow, Qo. If the soil moisture deficit is satisfied, the excess rainfall becomes overland flow, and the soil moisture deficit is reset to zero:

$$Qo_t = -smd_t$$
$$smd_t = 0$$

iii. Determine the subsurface flow from the soil, Qi. Subsurface flow from the soil depends on the soil moisture deficit, and also modifies the deficit:

$$Qi_t = Cb * \exp(-Cc * smd_t)$$
$$smd_t = smd_t + Qi_t$$

iv. Determine the evapotranspiration from the soil, AE. The proportion of PE that is satisfied depends on the soil moisture deficit, and again modifies the deficit:

$$AE_t = PE_t * \exp(-Ca * smd_t)$$
$$smd_t = smd_t + AE_t$$

v. Finally, add the subsurface flow and overland flow and route through a linear reservoir to calculate the catchment outflow, Q:

$$Q_t = Q_{t-1} + Cr1 * (Qo_{t-1} + Qi_{t-1} - Q_{t-1}) + Cr2 * (Qo_t - Qo_{t-1} + Qi_t - Qi_{t-1})$$

3.2 MODELS B1 & B2

Both these models are variations on one used by Bonvoisin & Boorman (1992a; model 15). The models (Figure 3.2) comprise two soil moisture stores, the lower with a limiting capacity. Precipitation is added to the upper store, and evapotranspiration occurs from the upper store at potential rate. Infiltration occurs from the upper store to the lower store, limited by the content of the upper store and the capacity of the lower store. Unsatisfied potential evapotranspiration occurs from the lower store at a rate proportional to store content. Runoff occurs from the upper and lower stores in proportion to their content. In model B1 a fraction of the baseflow from the lower store is lost to groundwater, before the remainder is added to the runoff from the upper store to become the catchment outflow. In model B2 the catchment losses to groundwater are assumed to be zero. These models have no channel routing component.

Model B1 has five parameters: the capacity of the lower soil moisture store *smax*, the infiltration rate *Ci*, a runoff coefficient *K1*, a baseflow coefficient *K2*, and a catchment losses coefficient *Cl*. Model B2 has four parameters as *Cl* is set to zero i.e. there are no catchment losses to groundwater. The models work through the accounting procedure in the following stages:

i. Add the rainfall, p, to the upper soil moisture store, su:

$$su_t = su_{t-1} + p_t$$

where: t and t-1 refer to the present and previous days, respectively.

ii. Determine the evapotranspiration from the upper soil moisture store, AE.
(a) If the store content is greater than PE:

$$AE_t = PE_t$$
$$su_t = su_t - AE_t$$

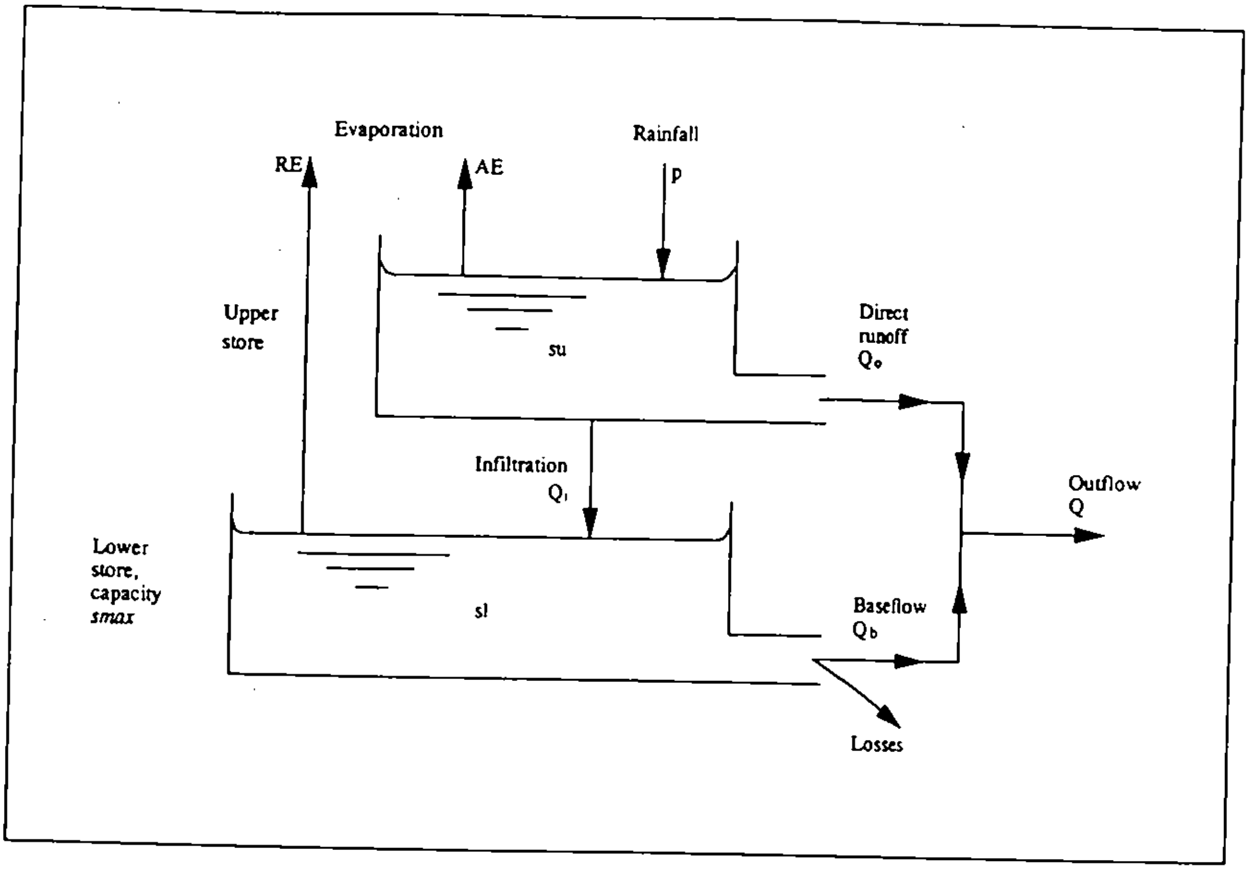


Figure 3.2 *Models B1 and B2*

(b) If the store content is less than PE:

$$AE_t = su_t$$
$$su_t = 0$$

iii. Determine the infiltration, Qi, from the upper soil moisture store to the lower one.
(a) If the upper store content is greater than the daily infiltration rate, *Ci*:

$$Qi_t = Ci$$
$$su_t = su_t - Qi$$

(b) If the upper store content is less than the daily infiltration rate:

$$Qi_t = su_t$$
$$su_t = 0$$

iv. Add the infiltrated soil moisture to the lower soil moisture store, sl:

$$sl_t = sl_{t-1} + Qi_t$$

If the lower soil moisture store is full i.e. the limiting capacity *smax* is reached, moisture remains in the upper store:

$$su_t = su_t + (sl_t - smax)$$
$$sl_t = smax$$

v. Determine the residual evapotranspiration from the lower soil moisture store, RE. Evapotranspiration occurs at a rate proportional to the store content:

$$RE_t = (sl_t / smax) * (PE_t - AE_t)$$

(a) If the lower store content is greater than the residual:

$$AE_t = AE_t + RE_t$$
$$sl_t = sl_t - RE_t$$

(b) If the lower store content is less than the residual:

$$AE_t = AE_t + sl_t$$
$$sl_t = 0$$

vi. A proportion of moisture from the upper store becomes runoff, Qo:

$$Qo_t = K1 * su_t$$
$$su_t = su_t - Qo_t$$

vii. A proportion of moisture from the lower store becomes baseflow, Qb:

$$Qb_t = K2 * sl_t$$
$$sl_t = sl_t - Qb_t$$

viii. In model B2 the catchment losses are assumed zero, but in model B1 a fraction of the baseflow from the lower store is lost:

$$Qb_t = (1.0 - Cl) * Qb_t$$

ix. Finally, the outflows from the upper and lower soil moisture stores are summed to give the catchment outflow, Q:

$$Q_t = Qo_t + Qb_t$$

3.3 MODEL C

Model C (Figure 3.3) is based on Moore's probability-distributed model (PDM; Moore, 1985) which has a soil moisture store, with a capacity varying across the basin, and a groundwater store. The model is being widely used in flood-forecasting (Moore *et al.*, 1990;

Moore & Jones, 1991; Moore, 1993).

The distribution of the soil moisture capacity, c, is represented by the reflected power (or Pareto) distribution:

$$F(c) = 1 - (1 - c / cmax)^{b} \quad \text{for } 0 \le c \le cmax$$

where: *cmax* is the maximum storage capacity at any point within the basin and b is a dimensionless parameter which defines the degree of spatial heterogeneity. The maximum amount of water that can be held in storage in the basin, *smax*, for the reflected power distribution is:

$$smax = \int_0^{cmax} (1 - F(c)) \,.\, dc$$

$$= cmax / (b + 1)$$

In the model, precipitation is added to the soil moisture store, and excess precipitation becomes direct runoff which is routed through two cascading linear reservoirs. Evapotranspiration from the soil moisture store occurs at a rate proportional to store content, as does drainage from the soil moisture store to the groundwater store. Baseflow occurs from the groundwater store and is added to the direct runoff to become the catchment outflow.

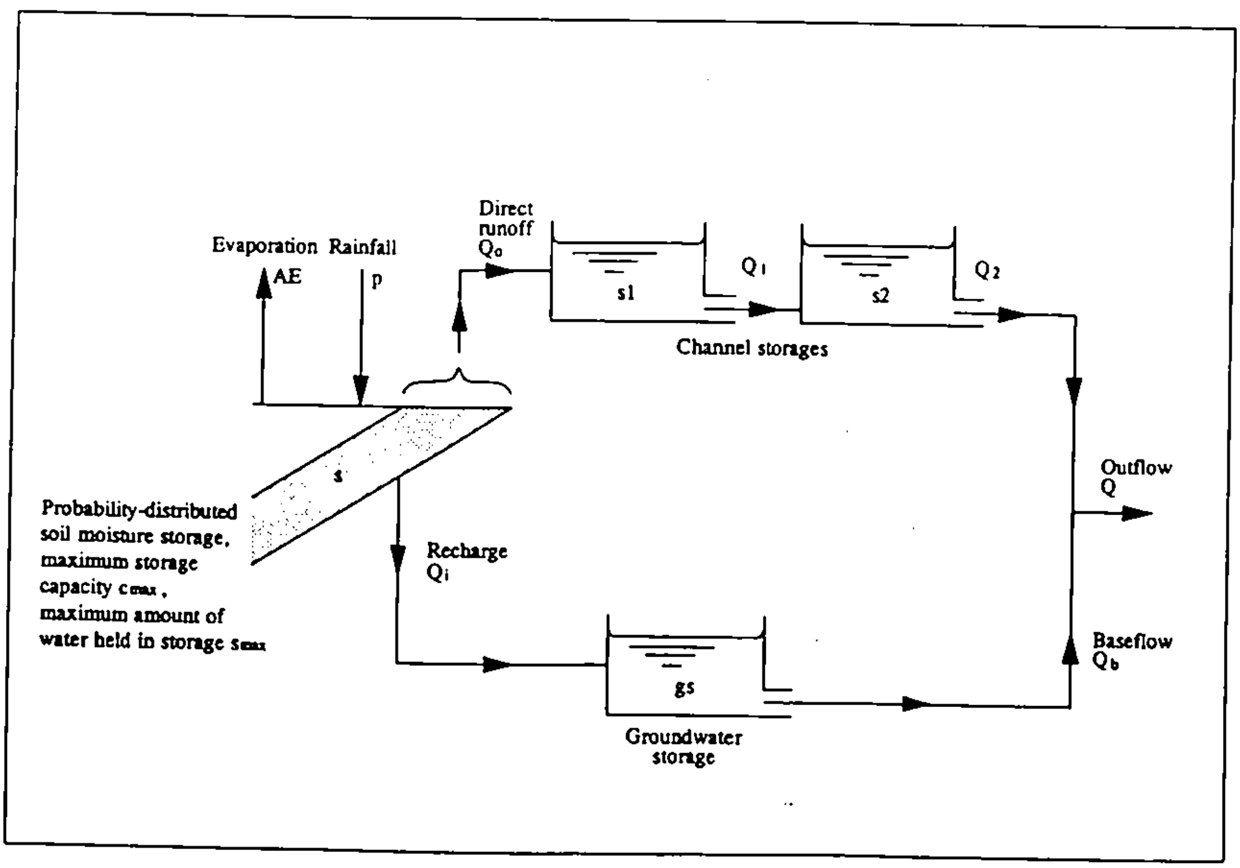


Figure 3.3 Model C

The model has five parameters: the maximum storage capacity at any point within the basin *cmax*, the average maximum amount of water that could be held in storage over the whole basin *smax*, a soil drainage coefficient *Kb*, a groundwater discharge coefficient *Grout*, and a channel routing coefficient *Srout*. *cmax* and *smax* together determine the degree of spatial heterogeneity b as above. The model works through the accounting procedure in the following stages:

i. Determine the evapotranspiration from the soil moisture store, AE. AE is a function of the potential evapotranspiration, PE, and the soil moisture content, s, at the end of the previous timestep:

$$AE_t = \{1 - \exp(-6.68 * s_{t-1} / smax)\} * PE_t$$

where: t and t-1 refer to the present and previous days, respectively. This particular actual evaporation function is taken from Wilmott *et al.* (1985), and assumes that the rate of decline of actual evapotranspiration increases as soil moisture deficit increases (the coefficient of 6.68 is there to ensure that AE $\approx$ PE when s = *smax*, though the actual value of the coefficient is not that important in practice).

ii. Determine the drainage from the soil moisture store to the groundwater store, Qi. Qi is also a function of the soil moisture content at the end of the previous timestep:

$$Qi_t = Kb * s_{t-1} / smax$$

iii. Determine the direct runoff, Qo.

(a) If the precipitation, p, is less than that going to AE and Qi, there is no direct runoff:

$$Qo_t = 0$$

$$s_t = s_{t-1} + (p_t - AE_t - Qi_t)$$

(b) If the precipitation is greater than that going to AE and Qi, direct runoff does occur. The critical capacity, Cc, at the end of the previous timestep, below which all the soil moisture goes to storage is calculated from the reflected power distribution:

$$s_{t-1} = \int_0^{Cc\,t-1} \{1 - F(c)\} \,.\, dc = \int_0^{Cc\,t-1} \{(1 - c / cmax)^b\}.\, dc$$

$$= smax \{1 - (1 - Cc_{t-1} / cmax)^{b+1}\}$$

which yields:

$$Cc_{t-1} = cmax * \{1 - (1 - s_{t-1} / smax)^{1/b+1}\}$$

Therefore the critical capacity at the end of the present timestep is:

$$Cc_t = Cc_{t-1} + (p_t - AE_t - Qi_t)$$

(b1) If Cc is less than *cmax* i.e. the basin is unsaturated, the direct runoff is given by:

$$Qo_t = (p_t - AE_t - Qi_t) - smax * \{(1 - Cc_{t-1} / cmax)^{b+1}) - (1 - Cc_t / cmax)^{b+1}\}$$

$$s_t = s_{t-1} + p_t - AE_t - Qi_t - Qo_t$$

(b2) If Cc is greater than *cmax* i.e. the entire basin has become saturated during the day, the direct runoff is given by:

$$Qo_t = (p_t - AE_t - Qi_t) - (smax - s_{t-1})$$

$$s_t = smax$$

iv. Determine the baseflow from the groundwater store, Qb. Qb is a function of the groundwater store content, gs, at the end of the previous timestep:

$$Qb_t = Grout * gs_{t-1} / 1000$$

$$gs_t = gs_{t-1} + Qi_t - Qb_t$$

v. Route the direct runoff through two cascading linear reservoirs. The reservoirs have the same routing coefficient *Srout*.

For the first reservoir:

$$Q1_t = Srout * (s1_{t-1} + Qo_t)$$

$$s1_t = s1_{t-1} + Qo_t - Q1_t$$

where: s1 is the content of the first reservoir and Q1 is the outflow from the first reservoir.

For the second reservoir:

$$Q2_t = Srout * (s2_{t-1} + Q1_t)$$

$$s2_t = s2_{t-1} + Q1_t - Q2_t$$

where: s2 is the content of the second reservoir and Q2 is the outflow from the second reservoir.

vi. Finally, the outflow from the second linear reservoir is combined with the baseflow to give the catchment outflow, Q:

$$Q_t = Q2_t + Qb_t$$

3.4 MODELS D1 & D2

These models (Figure 3.4), again variations of one used by Bonvoisin & Boorman (1992a;

model 16), use the concept of contributing areas, with a range of soil moisture store capacities from zero to some maximum. The total contents of all the stores translate to a level in the largest store. Precipitation is immediately subject to evapotranspiration at the potential rate, and remaining rainfall is added to the soil moisture stores. Remaining rainfall is also added to a linear channel storage at a rate proportional to the content of the soil moisture stores. Unsatisfied potential evaporation and subsurface flow from the soil moisture stores are again at rates proportional to store content. The subsurface flow is added to the channel storage. The catchment outflow from the channel storage is also at a rate proportional to store content.

The models have three parameters: the capacity of the largest soil moisture store *smax*, a subsurface flow coefficient *K1*, and a routing coefficient *K2*. The difference between the two models lies in the way the subsurface flow from the soil moisture stores is determined. Model D1 has a non-linear dependence on soil moisture, whilst in model D2 the relationship is assumed to be linear. The models work through the accounting procedure in the following stages:

i. Translate the total contents of all the soil moisture stores, scap, to a level, sl, in the largest store:

$$sl_t = smax - 2 * scap_{t-1}$$

where: t and t-1 refer to the present and previous days, respectively.

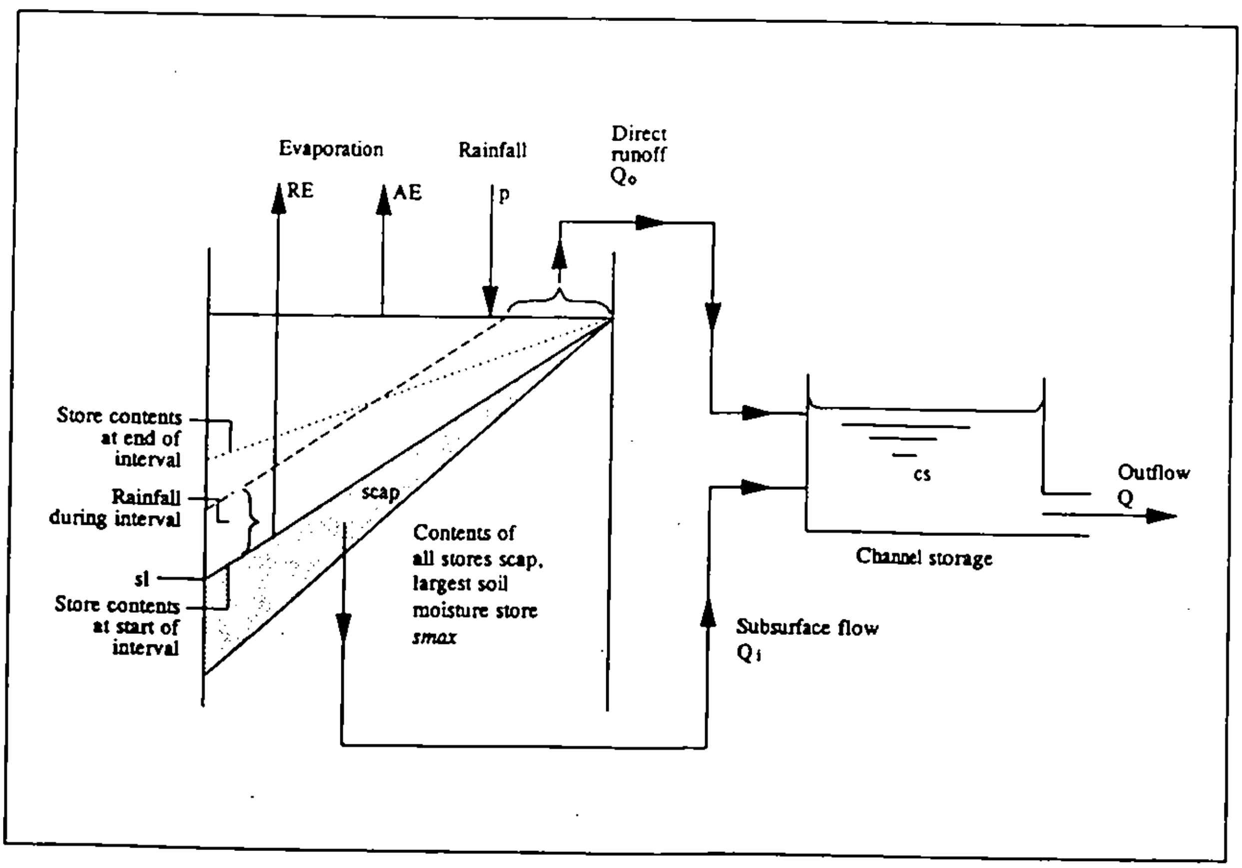


Figure 3.4 *Models D1 and D2*

ii. Determine the evapotranspiration, AE, from the rainfall, p.
(a) If the rainfall is greater than PE:

$$AE_t = PE_t$$
$$RE_t = 0$$
$$p_t = p_t - PE_t$$

(b) If the rainfall is less than PE:

$$AE_t = p_t$$
$$RE_t = PE_t - p_t$$
$$p_t = 0$$

where: RE is the residual evapotranspiration.

iii. Add the remaining rainfall to the soil moisture stores and determine the excess, Qo.
(a) If the soil is saturated all the rainfall is excess:

$$Qo_t = p_t$$

(b1) If the soil is unsaturated and the rainfall is greater than the deficit, the soil will become saturated, reducing the excess:

$$Qo_t = p_t - 0.5 * sl_t$$

(b2) If the soil is unsaturated but rainfall is less than the deficit, the soil will remain unsaturated, reducing the excess further:

$$Qo_t = 0.5 * p_t^2 / sl_t$$

iv. Add the excess to the channel storage, cs:

$$cs_t = cs_{t-1} + Qo_t$$

v. Determine the total content of all the soil moisture stores:

$$scap_t = scap_{t-1} + (p_t - Qo_t)$$

vi. Satisfy the residual evapotranspiration from the soil moisture stores. Evapotranspiration from the soil moisture stores occurs at a rate proportional to content:

$$RE_t = RE_t * (smax - sl_t) / smax$$
$$AE_t = AE_t + RE_t$$
$$scap_t = scap_t - RE_t$$

vii. Determine the subsurface flow, Qi, from the soil moisture stores. Subsurface flow from the soil moisture stores occurs at a rate proportional to content. In model D1 it is proportional to the square of the content:

$$Qi_t = K1 * scap_t^2$$

but in model D2 it is linear:

$$Qi_t = K1 * scap_t$$

(a) If the soil moisture store capacity is greater than the subsurface flow:

$$scap_t = scap_t - Qi_t$$

(b) If the soil moisture store capacity is less than the subsurface flow:

$$Qi_t = scap_t$$
$$scap_t = 0$$

viii. Add the subsurface flow to the channel storage:

$$cs_t = cs_t + Qi_t$$

ix. Finally, route the channel storage to calculate the catchment outflow, Q. The outflow from the channel storage occurs at a rate proportional to content:

$$Q_t = K2 * cs_t$$

(a) If the channel storage is greater than the outflow:

$$cs_t = cs_t - Q_t$$

(b) If the channel storage is less than the outflow:

$$Q_t = cs_t$$
$$cs_t = 0$$

3.5 SUMMARY

Table 3.1 summarises the number and names of the model parameters for each of the six models.

Table 3.1 Summary of model parameters

Model	No. of parameters	Parameters and units				
		1	2	3	4	5
A	5	*Ca* [mm^{-1}]	*Cb* [$mmday^{-1}$]	*Cc* [mm^{-1}]	*Cr1* [-]	*Cr2* [-]
B1	5	*smax* [mm]	*Ci* [$mmday^{-1}$]	*K1* [day^{-1}]	*K2* [day^{-1}]	*Cl* [-]
B2	4	*smax* [mm]	*Ci* [$mmday^{-1}$]	*K1* [day^{-1}]	*K2* [day^{-1}]	
C	5	*cmax* [mm]	*smax* [mm]	*Kb* [day^{-1}]	*Grout* [day^{-1}]	*Srout* [day^{-1}]
D1	3	*smax* [mm]	*K1* [$mm^{-1}day^{-1}$]	*K2* [day^{-1}]		
D2	3	*smax* [mm]	*K1* [day^{-1}]	*K2* [day^{-1}]		

4. Catchments and data

4.1 CATCHMENTS

Twenty-five catchments from around Great Britain, listed in Table 4.1, were used in the model comparison exercise. The locations of the catchments are shown in Figure 4.1. The catchments all have complete flow records for the 10-year period 1980-89, with low or average flows measured to an acceptable accuracy and minimal artificial influences on their flow regimes. They also have good coverage of long-term daily raingauges. The catchments cover a variety of sizes, topographies, soils and climates, and were chosen to represent a wide range of hydrological regimes.

Table 4.1 provides details of physical and climatic characteristics of the catchments as described in the FSR. The catchment areas range from 25 km^2 to 1616 km^2, whilst the topographic indices of mainstream length and channel slope range from 7 km to 117 km, and 0.64 m km^{-1} to 22.20 m km^{-1}, respectively. On more than half of the catchments one of the five WRAP (Winter Rainfall Acceptance Potential) classes dominates, whilst on the others two or three of the classes are found in fairly equal proportions. The climatic index SAAR (Standard Annual Average Rainfall) varies from 595 mm to 2162 mm. Two of the catchments (19001 and 38021) have significant urban fractions, whilst another two (21018 and 40007) have significant areas draining through lakes or reservoirs. Baseflow indices (BFI; Gustard *et al.*, 1992) show the ratio of baseflow to total flow (i.e. the higher the BFI, the less the day-to-day variability in flow) and range from 0.21 to 0.96.

Table 4.1 Catchment characteristics for basins used in this study

Catchment	(gnow)	AREA km²	MSL km	S1085 m/km	S1	S2	S3	S4	S5	SOIL	SAAR mm	URBAN	LAKE	BFI	MORECS
11001	Don at Parkhill(*)	1273.0	116.80	3.44	0.00	0.73	0.10	0.00	0.17	0.35	964	0.00	0.00	0.67	30 31
19001	Almond at Craigiehall(*)	369.0	42.00	5.81	0.00	0.00	0.00	0.80	0.20	0.46	909	0.11	0.04	0.38	57
21018	Lyne Water at Lyne Station(*)	175.0	27.50	6.36	0.00	0.00	0.80	0.02	0.18	0.42	1007	0.00	0.10	0.59	57
24004	Bedburn Beck at Bedburn(*)	74.9	14.00	22.20	0.00	0.00	0.00	0.41	0.59	0.48	950	0.00	0.00	0.46	79
25006	Greta at Rutherford Bridge(*)	86.1	17.89	11.68	0.00	0.00	0.00	0.00	1.00	0.50	1259	0.00	0.00	0.21	85
28008	Dove at Rochester Weir	399.0	48.80	4.59	0.47	0.00	0.01	0.21	0.31	0.32	1020	0.00	0.00	0.61	115
29003	Lud at Louth	55.2	7.45	6.12	0.92	0.00	0.00	0.08	0.00	0.17	729	0.01	0.00	0.90	102
32003	Harpers Brook at Old Mill Bridge	74.3	24.00	3.79	0.00	0.00	0.05	0.95	0.00	0.45	620	0.00	0.00	0.49	127
34004	Wensum at Costessey Mill	561.0	66.20	0.64	0.68	0.00	0.32	0.00	0.00	0.23	668	0.04	0.00	0.73	130
37005	Colne at Lexden	238.2	41.60	1.58	0.01	0.27	0.72	0.00	0.00	0.37	595	0.00	0.00	0.53	153
38021	Turkey Brook at Albany Park	42.2	12.70	5.28	0.00	0.00	0.00	1.00	0.00	0.45	661	0.11	0.00	0.21	161
39008	Thames at Eynsham	1616.2	77.00	1.30	0.39	0.09	0.28	0.24	0.00	0.31	755	0.06	0.00	0.68	148 149
39019	Lambourn at Shaw	234.1	21.30	2.38	0.85	0.01	0.00	0.14	0.00	0.19	737	0.00	0.00	0.96	159
40007	Medway at Chafford Weir	255.1	25.94	2.47	0.00	0.00	0.02	0.98	0.00	0.45	852	0.02	0.18	0.50	173
42003	Lymington at Brockenhurst Park	98.9	14.80	4.67	0.53	0.00	0.03	0.44	0.00	0.29	872	0.00	0.00	0.36	182
43005	Avon at Amesbury	323.7	38.50	1.62	1.00	0.00	0.00	0.00	0.00	0.15	768	0.00	0.00	0.91	169
47001	Tamar at Gunnislake	916.9	68.40	1.75	0.00	0.54	0.00	0.44	0.02	0.37	1240	0.00	0.02	0.46	177
48004	Warleggan at Trengoffe	25.3	10.00	17.48	0.00	0.25	0.00	0.00	0.75	0.45	1512	0.00	0.00	0.72	187
54008	Teme at Tenbury	1134.4	75.90	3.16	0.01	0.84	0.03	0.07	0.05	0.32	878	0.01	0.00	0.57	135
54016	Roden at Rodington	259.0	40.20	0.92	0.50	0.03	0.00	0.47	0.00	0.30	713	0.00	0.00	0.61	124
57004	Cynon at Abercynon(*)	106.0	25.80	7.30	0.00	0.00	0.30	0.00	0.70	0.47	1759	0.04	0.00	0.42	145
58009	Ewenny at Keepers Lodge(*)	62.5	13.05	7.67	0.30	0.30	0.38	0.00	0.02	0.30	1382	0.05	0.00	0.58	155
66011	Conwy at Cwm Llanerch(*)	344.5	29.04	17.20	0.00	0.51	0.00	0.00	0.49	0.40	2162	0.00	0.08	0.29	112
76005	Eden at Temple Sowerby(*)	616.4	56.79	4.05	0.00	0.00	0.00	0.30	0.70	0.49	1216	0.00	0.00	0.37	84
79006	Nith at Drumlanrig(*)	471.0	51.90	3.48	0.00	0.00	0.14	0.06	0.80	0.47	1579	0.00	0.00	0.34	69

where: AREA is catchment area, MSL is mainstream length, S1085 is channel slope, S1-S5 are fractions of catchment in WRAP soil classes 1-5, SOIL is the soil index, SAAR is average annual rainfall over period 1941-70, URBAN is fraction of catchment in urban development, LAKE is fraction of catchment draining through significant lake, BFI is baseflow index and MORECS is Met. Office MORECS square number. For further definition see FSR (NERC, 1975).

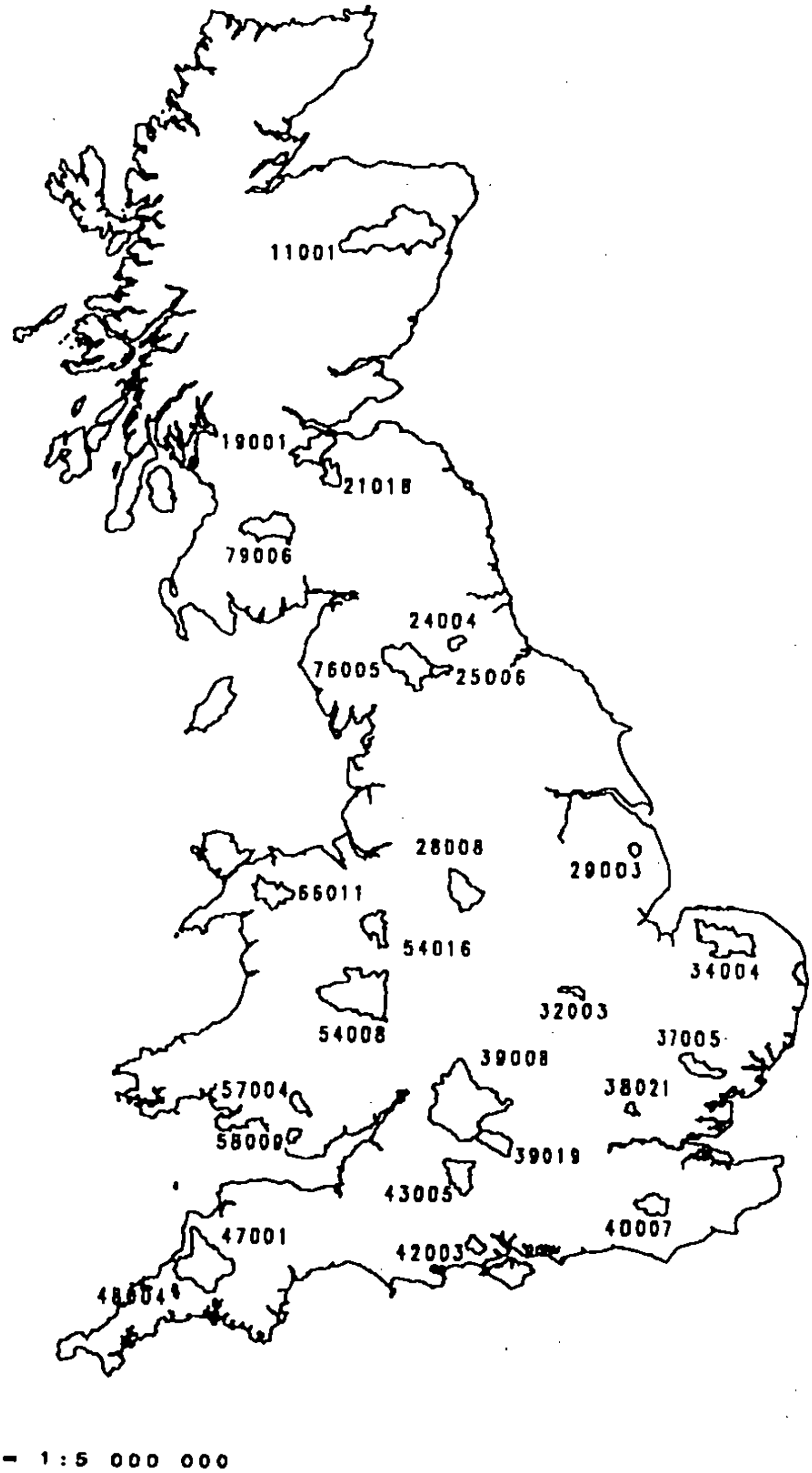


Figure 4.1 *Locations of catchments used in this study*

4.2 DATA

One of the aims of the model comparison study was to use data that were readily available and required minimum processing before application. The data requirements for each of the 25 catchments were daily rainfall from gauges on and near the catchment, daily river flow at the catchment gauging station and daily potential evaporation (PE) for the catchment. A number of catchments, all in the north and west of the UK, were often affected by snowfall and snowmelt. Consequently, it was necessary to adjust the rainfalls of these catchments (indicated by an asterisk after the catchment name and location in Table 4.1), and for this purpose daily temperature data were also collected for these catchments. All data were collected for both a model calibration period and a model validation period. The calibration period spanned 1981 to 1983 with 1980 used as a warm-up year; the validation period spanned 1984 to 1989 with 1983 used as the warm-up year to make maximum use of the data.

Catchment average daily rainfall data were derived from daily point values obtained from the UK Met. Office-supplied rainfall archive at the Institute of Hydrology. The procedure uses the triangle method of Jones (1983), whereby all the working raingauges falling within a quadrilateral bounding the catchment are used to calculate the daily catchment average rainfall. The number of gauges contributing can vary from day to day. This procedure is described in detail in Boorman & Houghton-Carr (1992).

Daily river flow data were extracted from the National Water Archive at the Institute of Hydrology.

PE data were derived from the monthly MORECS (Thompson *et al.*, 1981) dataset for grass. Each of the catchments was assigned to a 40 km by 40 km MORECS box as indicated in Table 4.1. Where a catchment covered two boxes, the area-weighted average PE was used. The monthly box values were applied to the catchments without correction for any differences between the altitudes of the boxes and the altitudes of the catchments. The monthly PE values were converted to daily values by simply dividing by the number of days in the month.

Daily temperature data were obtained from the Met. Office for ten catchments (indicated in Table 4.1). The names and altitudes of the meteorological sites from which the temperature records were collected are listed in Table 4.2, together with the catchment numbers and mean catchment altitudes. The mean catchment altitude was calculated as the mean of the maximum catchment altitude and the altitude of the gauging station. The temperature data were corrected for the mean catchment altitude using a lapse rate of 0.6°C per 100 m, and were then used to adjust the rainfalls for these catchments to allow for snowfall and snowmelt using Harding & Moore's snowmelt model (1988, 1992). The model was not calibrated, and the same parameters were used for all catchments. The parameters used in the snowmelt model were the same as those in the Yorkshire flow forecasting system implementation (Harding & Moore, 1992), and are given in Table 4.3. The model assumes that precipitation falls as snow when the temperature falls below some critical threshold, and that the snowpack begins to melt once another temperature threshold is passed. The melt rate is 4 mm$°C^{-1}day^{-1}$, felt to be appropriate for most UK conditions. The model invokes an areal depletion curve to allow shallow packs to cover only a fraction of the basin area.

Table 4.2 *Temperature records collected for catchments affected by snow*

Catchment	Mean altitude m	Temperature site	Altitude m
11001	452	Inverurie	54
19001	271	Widdybank	513
21018	365	Widdybank	513
24004	320	Widdybank	513
25006	410	Widdybank	513
57004	301	Cilfyndd	194
58009	154	Cilfyndd	194
66011	534	Alwen	335
76005	521	Eskdalemuir	242
79006	389	Eskdalemuir	242

Table 4.3 *Parameters of snowmelt model*

Parameter	Description	Units	Value
Tcrit	Temperature threshold below which precipitation is snow	°C	1.0
Tmelt	Critical temperature above which melt occurs	°C	0.0
Kmelt	Melt factor	mm°C^{-1}day^{-1}	4.0
Sc	Critical water content below which only a proportion of the basin is snow-covered	mm	100.0
S*	Maximum liquid water content as a proportion of total		0.04
K1	Storage time constant for lower orifice		0.15
K2	Storage time constant for upper orifice		0.85

5. Methodology

In the model comparison study, the six daily rainfall-runoff models were applied to each of 25 catchments using the MIMIC software package. Firstly, the model parameters for each of the 25 catchments were estimated for the calibration period (1981 to 1983, with 1980 as a warm-up year). These model parameters were then used with the validation dataset (1984 to 1989, with 1983 as the warm-up year). Model evaluation was based on a comparison of features of the observed and simulated flows during the calibration and validation periods.

5.1 METHODS OF FITTING

The classical approach to fitting a conceptual model to observed data to obtain an optimum parameter set involves minimising an objective function, generally by automatic optimisation techniques e.g. Rosenbrock (1960). The MIMIC software package offers two different objective functions in the automatic optimisation process: least squares and least squares of logarithms (Gan & Burges, 1990). In practice it makes little difference which of the objective functions is used. This is because, in general, most of the problems associated with optimisation are concerned with locating the global minimum in a parameter space that contains numerous local minima. The advantage of daily rainfall-runoff modelling within the MIMIC framework is that the objective function is regarded as a tool to aid fitting and assess the models, rather then the criterion by which fit is judged. Therefore, the objective functions are standardised by dividing by the number of days in the modelling period, in order to make calibration and validation results from each catchment comparable.

The least squares function Obj_1 is given by:

$$Obj_1 = \frac{\sum_{i=1}^{N} (Qobs_i - Qsim_i)^2}{N}$$

where: $Qobs_i$ is the observed flow on day i, $Qsim_i$ is the simulated flow on day i and N is the total number of days. This objective function evaluates the sum of the squares of the residuals, and may give good fits to long periods of low flows but poor fits to higher and more peaky portions of the hydrograph.

The least squares of logarithms function Obj_2 is given by:

$$Obj_2 = \frac{\sum_{i=1}^{N} (\log Qobs_i - \log Qsim_i)^2}{N}$$

This objective function evaluates the sum of the squares of the residuals of the logarithms of the flows and prevents the optimisation becoming biassed towards the largest flows.

For the purpose of calibration i.e. estimation of parameters, all the models were fitted to all the catchments using a two-stage approach. Firstly, starting with reasonably sensible parameter values chosen by the user based on an understanding of the behaviour of both the model and the catchment type, an automatic Rosenbrock (1960) optimisation procedure was invoked minimising the least squares objective function. The second stage involved manual adjustment, by a systematic trial-and-error process, of the optimised parameters where it was felt to be necessary. This was used either to obtain a better visual fit between the observed and simulated flows, or a closer match between the observed and simulated annual runoff totals. Another reason to use manual adjustment was to force a more realistic division between baseflow and surface runoff, as for some of the groundwater-dominated catchments the optimised model parameters sometimes implied that baseflow contributed only a small proportion of the total runoff. Alternatively, manual adjustment was used to adjust the model parameters from unrealistic values produced by the automatic optimisation to more reasonable ones. The final values of the objective function were noted, and the corresponding values of the least squares of logarithms objective function were also evaluated.

Standardising the objective functions enables comparison of the results from the calibration and validation datasets for each catchment; however, these results are not comparable across catchments because the objective functions are not normalised. Therefore, the Nash-Sutcliffe (1970) efficiency criterion was also calculated. This criterion is normalised and has the form:

$$\text{Eff.} = 1.0 - \frac{\sum_{i=1}^{N} (Qobs_i - Qsim_i)^2}{\sum_{i=1}^{N} (Qobs_i - Qbar)^2}$$

where: Qbar is the observed mean daily flow over the N day period. As the optimisation procedure was designed to minimise the objective function, it was also designed to maximise efficiency. Therefore, in this instance, the efficiency is acting as a form of normalised least squares objective function. The efficiency criterion is biassed towards large discharges, but is widely used, and gives an objective indication of model performance. A perfect agreement between the observed and simulated flows yields an efficiency of 1.0, whilst a negative efficiency represents a lack of agreement worse than if the simulated flows were replaced with the observed mean daily flow.

5.2 METHODS OF ASSESSMENT

Model performance was assessed by comparing a selection of both quantitative (objective) criteria and qualitative (subjective) criteria for both the calibration and validation periods, with the aim of searching for consistency in what constitutes a good model.

The first objective criteria considered were the objective functions themselves, and the Nash-Sutcliffe efficiency criterion. Other quantitative measures of model performance were the percentage error in the average annual runoff total {(Δ runoff / Observed runoff) * 100} and the relative error in the BFI {Δ BFI} (where Δ runoff is Simulated runoff minus Observed runoff, and Δ BFI is Simulated BFI minus Observed BFI). Both the signed and absolute errors were examined. The aim was to answer the following questions:

- Does a catchment with a high efficiency have small quantifiable errors?
- Did the models consistently overpredict or underpredict the flow, and how large were the errors regardless of sign?
- Are results consistent between the calibration and validation periods, in particular for the objective functions?

Three subjective criteria were used: comparisons of observed and simulated flow duration curves, observed and simulated monthly flows, and observed and simulated daily flows for two example years (1982 in the calibration period and 1987 in the validation period). The model performance under each criterion for each catchment was judged independently by each of the authors, the highest marks being given to the best performances. The first stage of the analysis of the results investigated whether the authors' judgements were the same. Correlations between the three criteria were also carried out, with a view to resolving the following questions:

- Does a catchment where the monthly flow regime is simulated well also have a well simulated daily flow regime, and does a catchment where the daily flow regime is simulated well also have a well simulated flow duration curve?
- Does a catchment with a well simulated flow duration curve reflect well simulated daily and monthly flows?
- Are results consistent between the calibration and validation periods?

The final part of the analysis of the results involved comparing the different methods of assessment. It seemed unlikely that any single index would be suitable for describing how well a particular model performed, but one of the aims of this part of the study was to determine whether there were certain individual measures, or sets of measures, which could be examined for a definitive assessment of model performance i.e.:

- For a particular model on a particular catchment, did all the objective measures give the same results, and all the subjective measures give the same results, and did those results agree?
- For a particular objective or subjective measure on a particular catchment, was the distinction clear and consistent between models that performed well and models that performed badly?
- Were there certain catchment types on which none of the models performed well, and others on which all of the models gave good results?

The full results from each model are included as appendices. For the most part, the analysis of the results involved evaluating correlations between the various criteria, for both the complete dataset (all six models on all 25 catchments) and each model individually. The figures quoted are correlation coefficients significant at the 95 % level. Visual examination of plots of each criterion against each of the other criteria (referred to later as X-Y plots) was also carried out to ensure that the relationships implied by high correlation coefficients were indeed reasonable.

6. Model assessment

This section describes the model results and assessment in some detail, considering first the quantitative (objective) criteria, and then the qualitative (subjective) criteria. Table 6.1 sets out the naming convention used in these sections. The full model results and optimum parameter sets are given in the appendices, as outlined below. The X-Y plots of each criterion against each of the other criteria are given in Appendix B.1; plots are provided for both the complete dataset and the individual models.

The final sets of model parameters for each of the models are shown in Appendix A.1, together with the objective functions and the Nash-Sutcliffe efficiency criterion calculated over both the calibration and validation periods. Catchments where some manual adjustment took place are indicated by an asterisk, and it should be remembered therefore that the given objective functions and efficiencies are not necessarily the minimum objective functions and the maximum efficiencies. Appendix A.2 shows, for each model, the observed and simulated average annual runoff totals in mm calculated over both the calibration and validation periods, together with the corresponding percentage errors. The match between the observed and simulated average annual runoff totals was considered during manual adjustment of the optimised model parameters. Appendix A.3 shows, for each model, the observed and simulated BFIs calculated over both the calibration and validation periods, together with the corresponding percentage and relative errors. Appendix A.4 shows, for each catchment, histograms of the observed and simulated monthly mean runoffs in mm from each model for the calibration and validation periods. Similarly, Appendix A.5 shows, for each catchment, plots of the observed and simulated daily mean flows in cumecs from each model for two example years: 1982 during the calibration period and 1987 during the validation period. The residual, calculated as simulated minus observed flow, is also shown. Appendix A.6 shows, for each catchment, the observed and simulated flow duration curves from each model for the calibration and validation periods. The matches between the observed and simulated daily flows and the observed and simulated flow duration curves were also considered during manual adjustment of the optimised model parameters.

***Table 6.1** Naming conventions (* highest marks awarded to best fits from consideration of calibration and validation periods together)*

Abbreviation	Meaning
COBJ1	Least squares objective function for calibration period
COBJ2	Least squares of logarithms objective function for calibration period
CEFF	Nash-Sutcliffe efficiency criterion for calibration period
CANRO	Percentage error in average annual runoff total for calibration period
CBFI	Relative error in BFI for calibration period
VOBJ1	Least squares objective function for validation period
VOBJ2	Least squares of logarithms objective function for validation period
VEFF	Nash-Sutcliffe efficiency criterion for validation period
VANRO	Percentage error in average annual runoff total for validation period
VBFI	Relative error in BFI for validation period
MON	Quality of simulated monthly flow regime*
HYD	Quality of simulated daily flow regime*
FDC	Quality of simulated flow duration curve*

6.1 QUANTITATIVE (OBJECTIVE) CRITERIA

The quantitative criteria considered fell into two groupings: firstly, the objective functions which were not normalised, and which could therefore only be compared between the calibration and validation periods for a particular catchment; and secondly, the remaining objective criteria which were normalised, and which could therefore be compared between catchments.

Objective functions

Table 6.2 shows the significant (at the 95 % level) correlation coefficients between the least squares and least squares of logarithms objective functions for the calibration and validation periods for both the complete dataset and the individual models.

Table 6.2 Significant correlations of least squares objective function and least squares of logarithms objective function over calibration and validation periods for each model and total dataset

Model	Criterion	COBJ1	COBJ2	VOBJ1	VOBJ2
A	COBJ1	1.00	-	0.99	-
n=25	COBJ2	-	1.00	-	0.96
	VOBJ1	0.99	-	1.00	-
	VOBJ2	-	0.96	-	1.00
B1	COBJ1	1.00	-	0.98	-
n=25	COBJ2	-	1.00	-	0.94
	VOBJ1	0.98	-	1.00	-
	VOBJ2	-	0.94	-	1.00
B2	COBJ1	1.00	-	0.99	-
n=25	COBJ2	-	1.00	-	0.98
	VOBJ1	0.99	-	1.00	-
	VOBJ2	-	0.98	-	1.00
C	COBJ1	1.00	-	0.99	-
n=25	COBJ2	-	1.00	-	0.48
	VOBJ1	0.99	-	1.00	-
	VOBJ2	-	0.48	-	1.00
D1	COBJ1	1.00	-	0.99	-
n=25	COBJ2	-	1.00	-	0.71
	VOBJ1	0.99	-	1.00	-
	VOBJ2	-	0.71	-	1.00
D2	COBJ1	1.00	0.51	0.99	0.51
n=25	COBJ2	0.51	1.00	0.51	0.97
	VOBJ1	0.99	0.51	1.00	0.51
	VOBJ2	0.51	0.97	0.51	1.00
TOTAL	COBJ1	1.00	0.31	0.99	-
n=150	COBJ2	0.31	1.00	0.29	0.48
	VOBJ1	0.99	0.29	1.00	-
	VOBJ2	-	0.48	-	1.00

The least squares objective functions between the calibration and validation periods were highly correlated (0.99) for both the complete dataset and from considering each model separately. The least squares of logarithms objective functions between the calibration and validation periods were also correlated (0.48) for the complete dataset, but this lower figure reflects the comparatively poor fits for models C (0.48) and D1 (0.71); correlation coefficients for the other four models were greater than 0.94. A significant correlation (0.51) was found between the least squares objective functions and the least squares of logarithms objective functions for model D2, but visual examination of the X-Y plot suggested that this relationship was heavily influenced by a single point, and that there was no relationship between the values from the two objective functions. However, the catchments with the lowest values for both objective functions are the same; these are the baseflow-dominated catchments. The reason for this consistency is presumably because the objective functions do not differ all that much when the range of flows is quite small; the consistency disappears as the range of flows increases.

Because they are normalised, the other three quantitative criteria can be compared between catchments. The Nash-Sutcliffe efficiency criterion can be used to assess model performance in general terms, whilst the percentage error in average annual runoff total and the relative error in BFI are potentially more specific indicators of the quality of the performance. (It was felt that the percentage errors gave a distorted view of the simulated BFI figures e.g. a 0.1 overestimation on an observed BFI of 0.2 gives a considerably higher percentage error than a 0.1 overestimation on an observed BFI of 0.8; therefore, whilst the percentage errors are supplied for completeness, it is the relative errors which are felt more appropriate for comparing model performances). Table 6.3 shows the significant (at the 95 % level) correlation coefficients between these three criteria for the calibration and validation periods for both the complete dataset and the individual models.

<u>Nash-Sutcliffe efficiency criterion</u>

The efficiency criteria between the calibration and validation periods were highly correlated (0.84) for the complete dataset; similar figures were obtained from considering each model separately, though with comparatively poor agreements for models B2 (0.70) and C (0.61). For all the models, there is generally a spread of efficiencies from around 0.50 to 0.90. Particular catchments had consistently low efficiencies, most notably 29003 (negative for model B2), 38021 and 42003 (negative for model D1). Similarly other catchments had consistently high efficiencies, in particular 39019, 43005 and 48004; these are all baseflow-dominated catchments (as is 29003). This suggests that catchments with high efficiencies tend to be those with high baseflows.

Table 6.3 Significant correlations of Nash-Sutcliffe efficiency criterion, percentage error in average annual runoff total and relative error in BFI over calibration and validation periods for each model and total dataset

Model	Criterion	CEFF	CANRO	CBFI	VEFF	VANRO	VBFI
A	CEFF	1.00	−0.70	-	0.82	-	-
n=25	CANRO	−0.70	1.00	-	−0.64	-	-
	CBFI	-	-	1.00	−0.40	-	0.90
	VEFF	0.82	−0.64	−0.40	1.00	-	−0.38
	VANRO	-	-	-	-	1.00	-
	VBFI	-	-	0.90	−0.38	-	1.00
B1	CEFF	1.00	−0.58	-	0.88	−0.48	-
n=25	CANRO	−0.58	1.00	-	−0.43	0.56	-
	CBFI	-	-	1.00	-	-	0.89
	VEFF	0.88	−0.43	-	1.00	−0.47	-
	VANRO	−0.48	0.56	-	−0.47	1.00	-
	VBFI	-	-	0.89	-	-	1.00
B2	CEFF	1.00	-	-	0.70	-	-
n=25	CANRO	-	1.00	-	−0.48	0.50	-
	CBFI	-	-	1.00	-	-	0.87
	VEFF	0.70	−0.48	-	1.00	−0.59	-
	VANRO	-	0.50	-	−0.59	1.00	-
	VBFI	-	-	0.87	-	-	1.00
C	CEFF	1.00	-	-	0.61	-	-
n=25	CANRO	-	1.00	-	-	0.79	-
	CBFI	-	-	1.00	-	-	0.95
	VEFF	0.61	-	-	1.00	-	-
	VANRO	-	0.79	-	-	1.00	-
	VBFI	-	-	0.95	-	-	1.00
D1	CEFF	1.00	−0.45	0.53	0.93	−0.45	0.40
n=25	CANRO	−0.45	1.00	-	−0.44	0.50	-
	CBFI	0.53	-	1.00	0.46	-	0.85
	VEFF	0.93	−0.44	0.46	1.00	−0.44	-
	VANRO	−0.45	0.50	-	−0.44	1.00	−0.48
	VBFI	0.40	-	0.85	-	−0.48	1.00
D2	CEFF	1.00	−0.59	-	0.86	−0.61	-
n=25	CANRO	−0.59	1.00	-	−0.42	-	-
	CBFI	-	-	1.00	-	-	0.90
	VEFF	0.86	−0.42	-	1.00	−0.64	-
	VANRO	−0.61	-	-	−0.64	1.00	-
	VBFI	-	-	0.90	-	-	1.00
TOTAL	CEFF	1.00	−0.44	-	0.84	−0.41	-
n=150	CANRO	−0.44	1.00	-	−0.41	0.44	-
	CBFI	-	-	1.00	-	−0.18	0.90
	VEFF	0.84	−0.41	-	1.00	−0.45	-
	VANRO	−0.41	0.44	−0.18	−0.45	1.00	−0.22
	VBFI	-	-	0.90	-	−0.22	1.00

Percentage error in average annual runoff total

In contrast to the efficiency, the percentage errors in the average annual runoff total between the calibration and validation periods were not particularly highly correlated (0.44) for the complete dataset. Similarly poor figures were obtained from considering each model separately: models A and D2 showed no significant correlation, though model C was as high as 0.79, and the others ranged between 0.50 and 0.56. The range of values of percentage errors was -30.4% (model B1) to 40.8% (model D2) in the calibration period, and -30.2% (model C) to 32.5% (model D2) in the validation period, though the majority of percentage errors were within plus or minus 10%. Except for model B2, the average annual runoff totals tended to be underestimated on the majority of catchments in both the calibration and validation periods. The catchments which tended to consistently have the largest errors were those with low values for the efficiency criterion e.g. 29003, 38021 and 42003, as would be expected, though models B2 and C performed well on 29003. There are several catchments on which some of the models perform well, and others badly e.g. 32003, a low lying impervious catchment on the edge of the East Anglian fenland, has notably low percentage errors with models B1 and D1, but notably high ones with models C and D2. Those catchments which tended to consistently have the smallest errors were 25006, 28008, 34004, 37005, 39019 and 47001; these include several baseflow-dominated catchments.

These results suggest that the catchments with the smallest percentage errors tended to be those with the highest efficiencies, whilst the catchments with the largest percentage errors tended to be those with the lowest efficiencies, and this relationship is confirmed by the correlation coefficients between these two measures in Table 6.6, particularly for model D2 (calibration -0.59, validation -0.64).

Relative error in BFI

Observed BFIs in the calibration period ranged from 0.198 for the responsive, partly-urbanised catchment 38021 to 0.971 for the baseflow-dominated catchment 39019. Observed BFIs during the validation period tended to be slightly higher. The relative errors in the BFI between the calibration and validation periods were highly correlated (0.90) for the complete dataset. Similar figures were obtained from considering each model separately, ranging from 0.85 for model D1 to 0.95 for model C. The range of values of relative errors was -0.180 (model B1) to 0.258 (model A) in the calibration period, and -0.152 (model D2) to 0.256 (model D1) in the validation period, though the majority of the relative errors were within plus or minus 10%. For all the models, BFI tended to be overestimated on the majority of catchments. Over half the catchments produced a poor result with a large relative error for at least one of the models, and the catchments which performed worst for each model varied. Catchments 24004 and 34004 had fairly consistently bad results, though 34004 had a particularly good result for model C. Similarly, nearly half the catchments produced a good result with a small relative error for at least one of the models, and the catchments which performed best for each model varied. Catchments 29003, 39008, 39019, 43005 and 48004 had fairly consistently good results; these include some of the baseflow-dominated catchments.

The BFI results were highly inconsistent across the models, with most of the catchments giving good results with some models, and bad results with the others. Baseflow-dominated catchments showed a mixture of both large and small relative errors, as did some of the more impervious catchments. Although some low correlations exist between the relative errors in the BFIs and the other criteria, these tend not to be consistent across the models.

Analysis of the results enables the questions posed in Section 5 to be answered:

- A catchment with a high efficiency does tend to have a small percentage error in estimation of the average annual runoff total, but the relative error in estimation of the BFI can be either large or small.

- The average annual runoff totals tended to be underestimated on the majority of catchments (except for model B2). The BFIs tended to be overestimated on the majority of catchments.

- The quantitative criteria were significantly and generally highly correlated between the calibration and validation periods. There was no correlation between the two objective functions themselves.

An alternative way of assessing the models is by relative performance i.e. rank, and Table 6.4 shows the means, standard deviations and ranks of all the quantitative criteria for the calibration and validation periods for both the complete dataset and the individual models.

Means and standard deviations of the least squares objective function are found to be slightly lower in the validation period than in the calibration period, whilst for the least squares of logarithms objective function the opposite is the case. Additionally, the model rank order between the means is quite variable both between the calibration and validation periods, and between the two objective functions themselves e.g. the mean for model C ranks top for one objective function, but only fourth for the other, whilst the mean for model B1 is consistently fifth or sixth. Overall, on the basis of the objective functions alone, model C appears to perform best and model B1 worst, with little to choose between the others.

Means and standard deviations of the efficiency tend to be higher in the calibration period than in the validation period. The model rank order between the means is generally consistent between the calibration and validation periods. Overall, on the basis of the efficiency criterion alone, model C appears to perform best, followed by model A. Model B1 performs worst, although model D1 is nearly as bad, with little to choose between the other two. The objective functions and efficiency criterion are therefore reasonably consistent about which models give the best and worst performances.

For models A, B1, C, and D1, the means and standard deviations of the absolute percentage errors in the average annual runoff total tend to be higher in the validation period than in the calibration period; the reverse is true for models B2 and D2. The errors for models A and B1 were less than 5% in the calibration period; all other errors in the calibration period and errors in the validation period ranged between 5% and 10%. The model rank order between the means is fairly consistent between the calibration and validation periods. Overall, on the basis of the absolute percentage errors alone, model A appears to perform best, followed by models B1, B2, C, D1 and D2 in that order.

For half the models (B1, D1 and D2) the means and standard deviations of the absolute relative errors in the BFI tend to be higher in the validation period than in the calibration period; the reverse is true for the other three models (A, B2 and C). For all the models except model A, the relative errors were between 5% and 10%; for model A they were around 11%. The model rank order between the means is consistent between the calibration and validation periods. Overall, on the basis of the absolute relative errors alone, model C performs best, followed by models B2, B1, D1, D2 and A in that order.

Table 6.4 Means, standard deviations and ranks of each model criteria over calibration and validation periods

Model	COBJ1			COBJ2			CEFF			CANRO			CBFI		
	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank
A	1.95	3.31	4	0.20	0.18	1	0.68	0.16	2	4.26	4.45	1	0.11	0.05	6
B1	2.07	3.01	6	0.32	0.24	6	0.64	0.16	6	4.93	6.38	2	0.07	0.06	3
B2	1.92	2.99	3	0.26	0.19	5	0.65	0.19	4	7.59	7.09	5	0.07	0.04	2
C	1.63	2.71	1	0.24	0.18	4	0.75	0.09	1	5.67	5.59	3	0.07	0.05	1
D1	1.83	2.80	2	0.22	0.15	2	0.65	0.22	5	7.22	5.83	4	0.07	0.04	4
D2	2.03	3.32	5	0.22	0.15	3	0.67	0.16	3	8.61	8.87	6	0.08	0.05	5
TOTAL	1.91	2.98	-	0.24	0.19	-	0.67	0.17	-	6.38	6.59	-	0.08	0.05	-

Model	VOBJ1			VOBJ2			VEFF			VANRO			VBFI		
	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank
A	1.97	3.39	6	0.29	0.28	2	0.66	0.14	2	6.06	5.27	2	0.11	0.05	6
B1	1.94	3.09	5	0.43	0.43	5	0.62	0.17	6	6.77	7.12	3	0.07	0.06	3
B2	1.85	2.97	3	0.30	0.24	3	0.66	0.14	3	5.41	6.68	1	0.06	0.05	2
C	1.62	2.81	1	0.33	0.43	4	0.73	0.11	1	7.40	7.28	4	0.05	0.05	1
D1	1.83	2.92	2	0.61	1.93	6	0.62	0.23	5	8.15	6.83	5	0.08	0.06	4
D2	1.90	3.14	4	0.25	0.16	1	0.65	0.18	4	8.34	7.97	6	0.08	0.05	5
TOTAL	1.85	3.01	-	0.37	0.84	-	0.66	0.17	-	7.02	6.87	-	0.08	0.06	-

Summary of model performance on basis of quantitative criteria

Model C ranks top in three of the four quantitative categories: the lowest average objective function, the highest average efficiency, and the smallest average relative errors in the BFI. Model A performed best in the other category, having the smallest average percentage errors in the average annual runoff total, whilst model C only ranked fourth for this criterion. On the basis of the quantitative criteria alone, models A and C perform about equally on average. Model B1 performs worst overall, though it did rank second for the estimation of the annual average runoff totals. There is little to choose between the other three models: B2, D1 and D2.

The quantitative performance criteria suggest that the best model performances tend to be achieved on the baseflow-dominated catchments, with the exception of 29003. The reasons for this include the smaller variability in flow regime of the baseflow-dominated catchments, and the fact that the flashier catchments may be less well represented by a model with a daily timestep. Catchments 29003, 38021 and 42003 show consistently bad results for all the models, and investigation shows that there are specific reasons why these catchments may not be suitable for such a modelling exercise: 29003 is a chalk catchment with a highly unusual anthropogenic flow regime (short-term spikes caused by mill regulation upstream); 38021 is a very responsive partly urbanised catchment; 42003 is a catchment where there may be problems with the observed water balance as comparison of the observed rainfall, potential evaporation and runoff suggests that the rainfall input is too high.

6.2 QUALITATIVE (SUBJECTIVE) CRITERIA

The three qualitative criteria considered were the simulated monthly flows, daily flows and flow duration curves. Model performance for each catchment was judged by the authors, the higher marks being given to the better performances. There was relatively little variation in the quality of the simulated monthly flows, so these were assessed jointly by the authors. However, for the qualities of the simulated daily flows and simulated flow duration curves, it was felt necessary for the authors to make independent assessments. Therefore, the analysis considered whether the authors judgements were the same, before moving on to the standard statistical and correlation analysis, in which the average marks were taken.

Monthly flow regime

All the models simulated the monthly mean flows reasonably well, with similar patterns found in both the calibration and validation periods. There are a few instances of over-estimation and under-estimation, usually in the Spring and Autumn, and there is a tendency for greater errors in the validation period, but differences are small. The joint marks awarded to each model on each catchment for the quality of the simulated monthly flow are given in Table 6.5. Table 6.6 shows the significant (at the 95 % level) correlation coefficients between the quality of the monthly flows and the other quantitative and qualitative criteria for the calibration and validation periods for both the complete dataset and the individual models. There were some significant correlations between the quality of the monthly flows and the other qualitative criteria, but visual examination of the X-Y plots failed to support the implied relationships. However, there were some relationships with the quantitative criteria; namely a positive trend with the efficiency, except for model D1, and a negative trend with the percentage errors in the average annual runoff total. The best results were obtained from the baseflow-dominated catchments, with high efficiencies and small errors. All the models performed badly on 29003, 38021 and 42003, identified in the previous section as potential problem catchments.

Table 6.5 Scores awarded for quality of simulated monthly flow regime

	Model					
Catchment	A	B1	B2	C	D1	D2
11001	3.0	2.5	2.5	3.0	3.0	3.0
19001	3.5	2.5	3.0	3.5	3.5	3.0
21018	4.0	3.0	3.0	2.5	3.0	3.0
24004	3.0	1.5	1.5	2.0	2.0	2.0
25006	3.5	2.5	2.5	3.0	3.0	3.5
28008	3.5	2.5	3.0	3.5	3.0	3.0
29003	1.5	1.5	1.0	1.0	1.5	1.0
32003	2.5	2.5	2.0	1.5	2.5	2.5
34004	4.0	3.5	3.0	4.0	3.5	3.0
37005	3.0	2.0	3.0	3.0	3.5	3.0
38021	1.5	1.5	2.5	3.0	1.5	1.5
39008	3.5	2.0	2.0	3.0	2.5	3.0
39019	4.0	3.0	3.0	4.0	3.0	4.0
40007	3.0	1.5	2.5	3.0	2.5	2.0
42003	1.0	1.0	1.5	1.0	2.0	2.0
43005	3.0	3.0	2.5	3.5	3.5	4.0
47001	4.0	3.5	3.5	3.0	3.0	3.5
48004	2.5	3.0	3.5	3.0	3.5	3.0
54008	3.0	2.0	2.5	2.5	2.5	2.5
54016	3.0	3.0	2.5	3.0	2.5	2.5
57004	3.0	3.0	3.5	2.5	3.0	3.0
58009	2.5	3.0	3.0	2.5	2.5	3.5
66011	3.0	3.0	2.5	3.0	3.0	3.5
76005	2.5	3.0	2.5	2.0	2.0	2.0
79006	3.5	2.5	1.5	2.0	2.5	3.0

Daily flow regime

Table 6.7a shows marks from each of the authors awarded to each model on each catchment for the quality of the daily flows. The average marks show that author 1 (range 2.48 for models B1 and B2 to 2.72 for models C and D1) tended to be more conservative in marking than author 2 (range 1.60 for model B1 to 3.12 for model A) i.e. on average author 1 marked higher than author 2 for poor performances, but lower for high performances. The model rank order on the basis of these marks also varies between the authors. Both agree that models B1 and B2 are worst, but whilst author 2 places model A first, author 1 places it only fourth, upsetting the otherwise near identical rank order. To investigate the differences in marks between the authors further, Table 6.7b shows the correlation coefficients between the marks awarded to each model by each author.

Table 6.6 Significant correlations of qualitative (subjective) criteria for each model and total dataset

Model	Criterion	MON	HYD	FDC	CEFF	CANRO	CBFI	VEFF	VANRO	VBFI
A	MON	1.00	-	-	0.52	−0.45	-	0.43	-	-
n=25	HYD	-	1.00	-	-	-	-	-	-	-
	FDC	-	-	1.00	-	-	-	-	-	-
B1	MON	1.00	-	0.43	0.75	-	-	0.69	−0.57	-
n=25	HYD	-	1.00	−0.59	-	-	-	-	-	-
	FDC	0.43	−0.59	1.00	-	-	-	-	-	0.41
B2	MON	1.00	-	-	0.63	-	-	0.50	−0.48	-
n=25	HYD	-	1.00	-	0.50	-	-	-	-	-
	FDC	-	-	1.00	-	-	-	-	-	-
C	MON	1.00	-	0.49	0.53	−0.47	-	0.58	−0.50	-
n=25	HYD	-	1.00	-	-	-	-	-	-	-
	FDC	0.49	-	1.00	0.41	-	-	-	-	-
D1	MON	1.00	-	0.47	-	−0.42	-	-	−0.50	-
n=25	HYD	-	1.00	−0.45	-	-	-	-	-	-
	FDC	0.47	−0.45	1.00	0.40	-	-	0.42	-	-
D2	MON	1.00	-	-	0.74	−0.47	-	0.53	−0.65	-
n=25	HYD	-	1.00	-	-	-	-	-	-	−0.49
	FDC	-	-	1.00	-	-	-	-	-	-
TOTAL	MON	1.00	0.18	0.41	0.55	−0.34	-	0.46	−0.49	0.19
n=150	HYD	0.18	1.00	-	-	-	-	-	-	-
	FDC	0.41	-	1.00	−0.22	-	-	0.27	−0.25	0.19

Table 6.7a Scores awarded for quality of simulated daily flow regime

	Model and Judge																	
	A			B1			B2			C			D1			D2		
Catchment	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.
11001	1.0	3.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.5	1.0	2.0	1.5	1.0	2.0	1.5
19001	3.0	3.0	3.0	3.0	1.0	2.0	3.0	1.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
21018	2.0	3.0	2.5	2.0	1.0	1.5	2.0	1.0	1.5	3.0	3.0	3.0	2.0	3.0	2.5	3.0	3.0	3.0
24004	3.0	3.0	3.0	3.0	2.0	2.5	2.0	1.0	1.5	3.0	4.0	3.5	3.0	3.0	3.0	3.0	3.0	3.0
25006	3.0	2.0	2.5	4.0	3.0	3.5	3.0	1.0	2.0	3.0	2.0	2.5	3.0	2.0	2.5	4.0	2.0	3.0
28008	3.0	4.0	3.5	3.0	2.0	2.5	4.0	3.0	3.5	4.0	4.0	4.0	3.0	3.0	3.0	3.0	3.0	3.0
29003	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.5	2.0	2.0	2.0	1.0	1.0	1.0
32003	3.0	3.0	3.0	3.0	1.0	2.0	3.0	1.0	2.0	3.0	2.0	2.5	3.0	3.0	3.0	3.0	3.0	3.0
34004	2.0	2.0	2.0	2.0	1.0	1.5	2.0	2.0	2.0	2.0	3.0	2.5	2.0	1.0	1.5	1.0	2.0	1.5
37005	3.0	4.0	3.5	2.0	1.0	1.5	3.0	1.0	2.0	3.0	3.0	3.0	4.0	4.0	4.0	4.0	4.0	4.0
38021	2.0	3.0	2.5	2.0	1.0	1.5	2.0	1.0	1.5	4.0	3.0	3.5	3.0	3.0	3.0	3.0	4.0	3.5
39008	2.0	3.0	2.5	2.0	1.0	1.5	1.0	1.0	1.0	1.0	3.0	2.0	2.0	3.0	2.5	2.0	3.0	2.5
39019	1.0	3.0	2.0	1.0	1.0	1.0	1.0	2.0	1.5	1.0	3.0	2.0	1.0	1.0	1.0	1.0	3.0	2.0
40007	3.0	4.0	3.5	3.0	1.0	2.0	3.0	1.0	2.0	3.0	3.0	3.0	3.0	4.0	3.5	3.0	3.0	3.0
42003	3.0	3.0	3.0	3.0	2.0	2.5	2.0	1.0	1.5	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
43005	2.0	3.0	2.5	2.0	1.0	1.5	2.0	2.0	2.0	1.0	3.0	2.0	2.0	3.0	2.5	1.0	4.0	2.5
47001	3.0	4.0	3.5	3.0	1.0	2.0	4.0	1.0	2.5	4.0	4.0	4.0	3.0	4.0	3.5	4.0	2.0	3.0
48004	2.0	3.0	2.5	2.0	2.0	2.0	3.0	3.0	3.0	3.0	2.0	2.5	3.0	3.0	3.0	2.0	2.0	2.0
54008	2.0	3.0	2.5	2.0	1.0	1.5	1.0	1.0	1.0	3.0	4.0	3.5	2.0	3.0	2.5	3.0	3.0	3.0
54016	3.0	3.0	3.0	2.0	1.0	1.5	1.0	3.0	2.0	2.0	3.0	2.5	3.0	3.0	3.0	2.0	2.0	2.0
57004	3.0	3.0	3.0	3.0	1.0	2.0	3.0	2.0	2.5	3.0	2.0	2.5	3.0	3.0	3.0	3.0	3.0	3.0
58009	3.0	4.0	3.5	2.0	2.0	2.0	3.0	3.0	3.0	3.0	2.0	2.5	3.0	4.0	3.5	3.0	3.0	3.0
66011	4.0	4.0	4.0	4.0	4.0	4.0	4.0	3.0	3.5	3.0	2.0	2.5	4.0	4.0	4.0	4.0	2.0	3.0
76005	3.0	3.0	3.0	2.0	3.0	2.5	3.0	2.0	2.5	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
79006	4.0	4.0	4.0	4.0	3.0	3.5	4.0	3.0	3.5	3.0	2.0	2.5	4.0	4.0	4.0	4.0	2.0	3.0
Average	2.6	3.1	2.8	2.5	1.6	2.0	2.5	1.7	2.1	2.7	2.8	2.8	2.7	3.0	2.8	2.7	2.7	2.7

Notation: J1 is judge 1, J2 is judge 2 and Av. is mean score

Table 6.7b *Correlations between scores awarded by two judges for quality of simulated daily flow regime*

Model	A	B1	B2	C	D1	D2	TOTAL
r	0.58	0.57	-	-	0.73	-	0.67

where: r is correlation coefficient between marks awarded by author 1 and marks awarded by author 2.

Significant correlations are obtained for models A, B1 and D1 (0.58, 0.57 and 0.73 respectively), suggesting that even though the actual assigned marks may have differed between the authors, there was reasonable agreement on the quality of performance i.e. on the best performances and the worst performances. However, for models B2, C and D2 there were no significant correlations, suggesting that the authors' judgments on what differences between the simulated and observed hydrographs were or were not acceptable sometimes differed quite considerably. For example, model B2 tends to miss occasional peaks and produce hydrographs that rise and fall too sharply compared to the observed flows; for five particular catchments (19001, 32003, 37005, 40007 and 47001) author 1 felt that the overall model performance justified good marks, but author 2 felt these features were quite unacceptable and gave poor marks. Conversely, model C can produce extremely smoothed hydrographs which fail to reproduce the details of the observed flows even though they bear some resemblance to the basic shape; for three catchments (39008, 39019 and 43005) author 1 felt the representation of the observed hydrographs was not adequate and gave poor marks, but author 2 felt the overall model performance justified good marks. These observations go to illustrate the subjective nature of interpreting hydrological model performance.

All the models simulated the daily flows less well than the monthly flows; the models did produce regimes that differed from catchment to catchment, but details tended to be poorly simulated. The extent to which they managed to reproduce the general regime varied considerably both between models and between catchments for a particular model. For all the models there was little obvious visual difference in performance between the example calibration and validation years; for some catchments the fit was rather worse in the validation year (e.g. model A on 21018), whilst for others it was rather better (e.g. model C on 54008). All the models nearly always underestimated high flow peaks; some peaks were missed entirely, but in most cases peaks existed but were too small, and the underestimation of high flow peaks was particularly bad in Summer and Autumn. This may be due to the models failing to simulate the runoff generation process adequately: too little of the catchment may be assumed to be saturated and able to respond rapidly to rainfall, or alternatively, it may be caused by problems with the input rainfall which is assumed to be falling evenly throughout the day over the whole catchment i.e. in reality the rainfall may be localised in one part of the catchment, or may have fallen in just a few hours. In practice, the answer is probably a combination of both these factors. Additionally, the high frequency day-to-day fluctuations were not reproduced at all well as the models miss what rapid runoff occurs, usually from the limited saturated area adjacent to the river channel. Furthermore, poor model performances with respect to both timing and magnitude on catchments in the north and west may be caused, not by inadequacies in the models, but by errors in the adjustment of estimated catchment rainfalls for snowmelt. Table 6.6 shows the significant (at the 95 % level) correlation coefficients between the average quality of the daily flows and the other criteria for the calibration and validation periods for both the complete dataset and the individual models. There were some significant correlations between the quality of the daily flows and the other criteria, but again visual examination of the X-Y plots failed to support the implied relationships.

Although there was considerable variation in performance both between models and between catchments for a particular model, there were as usual some catchments on which the models tended to perform consistently well or consistently badly, as judged by the authors. The models performed badly on 11001, 29003, 34004 and 39019, and well on 28008, 37005, 47001, 66011 and 79006. It is difficult to find a property common to the catchments within the two groups but which distinguishes between the groups. For example, 28008, 37005 and 47001 were all noted as having particularly low percentage errors in the average annual runoff totals, as were 34004 and 39019; 47001 was also noted as having a particularly well simulated monthly flow, as were 34004 and 39019; 29003 had a notably low efficiency whilst 39019 had a notably high one; 34004 had particularly high relative errors in the BFIs whilst 39019 had particularly low ones. There is little consistency, and this is most probably a reflection of the subjective nature of the scoring. However, it is clear that it is on the catchments with the higher BFIs (greater than around 0.65) that the models appear to perform worse, and the more quickly responding, impervious catchments on which they appear to perform better, though this is not supported by any significant correlations between the quality of the daily flow and the relative errors in BFI as might be expected. This result is against the trend set for the previous criteria, where the models had tended to give the best results on the baseflow-dominated catchments.

Flow duration curve

Table 6.8a shows marks from each of the authors awarded to each model on each catchment for the quality of the flow duration curve. The average marks show that author 1 (range 2.00 for model B1 to 2.72 for model D2) again tended to be more conservative in marking than author 2 (range 1.40 for model B1 to 2.60 for model D2) i.e. on average author 1 marked higher than author 2 for poor performances, though marks for high performances were very similar. The model rank order on the basis of these marks is fairly consistent between authors. Both agree that model D2 is best, followed by models A and D1, then by models B2 and C, and that model B1 is worst. This similarity in marks and rank order is in contrast to that for the daily flows, and Table 6.8b shows the correlation coefficients between the marks awarded to each model by each author.

Table 6.8b *Correlations between scores awarded by two judges for quality of simulated flow duration curve*

Model	A	B1	B2	C	D1	D2	TOTAL
r	0.87	0.71	0.59	0.77	0.64	0.65	0.67

where: r is correlation coefficient between marks awarded by author 1 and marks awarded by author 2.

Significant correlations are obtained from all the models, ranging from 0.59 for model B2 to 0.87 for model A, confirming the similarity in views on quality of performance. The discrepancy between the authors' judgements of the daily flows, and the similarity between their judgements of the flow duration curves, suggests that assessment of the latter is easier.

Table 6.8a Scores awarded for quality of simulated flow duration curve

	Model and Judge																	
	A			B1			B2			C			D1			D2		
Catchment	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.	J1	J2	Av.
11001	3.0	3.0	3.0	3.0	2.0	2.5	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	2.0	2.5
19001	3.0	3.0	3.0	2.0	1.0	1.5	2.0	1.0	1.5	3.0	2.0	2.5	2.0	3.0	2.5	3.0	3.0	3.0
21018	3.0	3.0	3.0	3.0	1.0	2.0	3.0	1.0	2.0	1.0	1.0	1.0	3.0	3.0	3.0	4.0	3.0	3.5
24004	1.0	2.0	1.5	1.0	1.0	1.0	2.0	1.0	1.5	1.0	2.0	1.5	1.0	2.0	1.5	2.0	2.0	2.0
25006	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.5	1.0	2.0	1.5	3.0	2.0	2.5	3.0	2.0	2.5
28008	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0
29003	1.0	1.0	1.0	2.0	1.0	1.5	3.0	2.0	2.5	3.0	2.0	2.5	2.0	2.0	2.0	3.0	3.0	3.0
32003	2.0	2.0	2.0	2.0	1.0	1.5	2.0	1.0	1.5	1.0	1.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0
34004	4.0	3.0	3.5	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	4.0	3.0	3.0	3.0	3.0	3.0	3.0
37005	3.0	3.0	3.0	2.0	2.0	2.0	2.0	1.0	1.5	2.0	2.0	2.0	1.0	2.0	1.5	3.0	4.0	3.5
38021	2.0	1.0	1.5	2.0	1.0	1.5	1.0	1.0	1.0	3.0	3.0	3.0	1.0	2.0	1.5	2.0	2.0	2.0
39008	4.0	4.0	4.0	2.0	1.0	1.5	2.0	1.0	1.5	3.0	2.0	2.5	3.0	3.0	3.0	3.0	3.0	3.0
39019	1.0	2.0	1.5	4.0	3.0	3.5	4.0	3.0	3.5	4.0	3.0	3.5	4.0	3.0	3.5	3.0	3.0	3.0
40007	2.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0	1.5	1.0	2.0	1.5	2.0	3.0	2.5	2.0	2.0	2.0
42003	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.5	1.0	2.0	1.5	2.0	2.0	2.0
43005	3.0	3.0	3.0	4.0	2.0	3.0	4.0	2.0	3.0	4.0	3.0	3.5	3.0	2.0	2.5	4.0	4.0	4.0
47001	3.0	3.0	3.0	2.0	1.0	1.5	2.0	1.0	1.5	3.0	2.0	2.5	3.0	3.0	3.0	3.0	2.0	2.5
48004	1.0	1.0	1.0	2.0	1.0	1.5	1.0	2.0	1.5	1.0	1.0	1.0	2.0	2.0	2.0	2.0	3.0	2.5
54008	1.0	1.0	1.0	2.0	1.0	1.5	2.0	1.0	1.5	1.0	2.0	1.5	1.0	2.0	1.5	2.0	2.0	2.0
54016	4.0	3.0	3.5	3.0	2.0	2.5	3.0	2.0	2.5	1.0	2.0	1.5	3.0	3.0	3.0	3.0	3.0	3.0
57004	2.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0	1.5	3.0	2.0	2.5	3.0	2.0	2.5	3.0	2.0	2.5
58009	3.0	3.0	3.0	2.0	2.0	2.0	3.0	1.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	3.0	3.0	3.0
66011	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	3.0	3.0	3.0	2.0	2.0	2.0
76005	1.0	2.0	1.5	1.0	1.0	1.0	2.0	1.0	1.5	2.0	2.0	2.0	1.0	2.0	1.5	3.0	3.0	3.0
79006	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.5	2.0	2.0	2.0
Average	2.2	2.2	2.2	2.0	1.4	1.7	2.2	1.4	1.8	2.1	2.1	2.1	2.2	2.4	2.3	2.7	2.6	2.7

Notation: J1 is judge 1, J2 is judge 2 and Av. is mean score

When assessing daily flows it is necessary to compare the baseflow, all the peaks and the speed of rise and fall of the hydrographs, whilst for flow duration curves there are just the extreme flows and the middle range of flow to consider. Hence, there is likely to be more agreement between different judges. Quality of the flow duration curve is likely to provide a better qualitative measure of model performance than quality of the monthly flow (too little variation), and quality of the daily flow (too much variation).

Table 6.6 shows the significant (at the 95 % level) correlation coefficients between the average quality of the flow duration curves and the other criteria for the calibration and validation periods for both the complete dataset and the individual models. There were some significant correlations between the quality of the flow duration curves and the other qualitative criteria, but these were not supported by visual examination of the X-Y plots.

For all the models, the simulated flow duration curves in the calibration period tended to be more accurate than those in the validation period. For models C, D1 and D2 the models tended to underestimate both low flows (below the flow exceeded 95% of the time) and high flows (above the flow exceeded 1% of the time), and overestimate intermediate flows (around the flow exceeded 50% of the time). Model A tended to underestimate low flows, but overestimate intermediate and high flows. Models B1 and B2 again performed exceedingly badly on the majority of catchments: the simulated flow duration curves are characterised by a distinct and abrupt bend somewhere in the middle of the curve (usually just on the high flow side of the flow exceeded 50% of the time), and extreme underestimation at the low flow end. The underestimation of low flows for models A, C, D1 and D2 suggests that the models are not simulating the slow release of water during dry periods particularly well. The over- or under-estimation of the high flows for the same models reflects the poor simulation of flood peaks, as already described for the daily flows. The quality of the flow duration curves for models B1 and B2 clearly reflects the poor quality of the simulated daily flows. It is worth noting that whilst models which simulate the day-to-day flow variability of a catchment well, also tend to give a good flow duration curve (e.g. model D2 on 25006), the reverse is not necessarily true i.e. models which produce a good simulated flow duration curve on a particular catchment do not necessarily produce a good daily simulated flow (e.g. model A on 34004). However assessment of the quality of the daily flow appears more subjective than assessment of the quality of the flow duration curve.

The catchments on which the models were noted to perform consistently well, as judged by the authors, were 34004, 39019 and 43005. These are catchments with high efficiencies, low percentage errors in the average annual runoff totals as well as good simulated monthly flows. The latter two catchments, with BFIs greater than 0.90, had low relative errors in the BFIs whilst 34004, with a BFI of 0.73, had a high error. The former two catchments were noted as having badly simulated daily flows, reiterating that the catchments which have a good simulated flow duration curve do not necessarily have a good daily simulated flow. Conversely, there were several catchments on which the models were noted to perform consistently badly, as judged by the authors, namely 24004, 25006, 38021, 42003, 48004, 54008, 66011, 76005 and 79006. This group includes both high efficiency (48004) and low efficiency (42003) catchments, catchments with both low (25006) and high (38021) percentage errors in the average annual runoff totals, catchments with both low (38021) and high (24004) relative errors in the BFIs, and catchments with both good (66011) and bad (54008) simulated daily and monthly flows. It is also worth noting that it is the catchments with the highest BFIs (greater than around 0.70) on which the models appear to perform best, and the more impervious catchments on which they appear to perform worse i.e. the opposite of the relationship for the daily flows, and similar to the results prior to that.

To summarise: the quality of the monthly flow was the worst differentiator between model performance, with all the models assessed to be performing reasonably well. In contrast the quality of the daily flow showed most variation between models, but was so subjective that the two authors sometimes disagreed on what constituted a good or bad model performance. The quality of the flow duration curve showed enough variation, and the authors' judgements agreed enough, to make the criterion more meaningful. However, most users would probably feel more comfortable with a model that simulates day-to-day variation well, rather than a model that simulates the monthly mean flow or the flow duration curve well. The quality of the simulations of the daily flows probably does explain much of the inconsistent BFI results, because calculation of BFI depends on separation of the daily flow hydrograph.

Again, the questions presented in Section 5 can be considered:

- A catchment where the monthly flow regime is simulated well can have, but does not necessarily have, a well simulated daily flow regime. Similarly a catchment where the daily flow regime is simulated well can have, and indeed tends to have, a well simulated flow duration curve.

- However, a catchment with a well simulated flow duration curve does not necessarily have well simulated daily and monthly flows.

- The simulated monthly flow regimes were of similar, good quality in the calibration and validation periods. The simulated daily flow regimes were sometimes better in the calibration period, and sometimes better in the validation period. The simulated flow duration curves tended to be much better in the calibration period.

The models can also be assessed by relative performance i.e. rank, and Table 6.9 shows the means, standard deviations and ranks of all the qualitative criteria for the individual models.

Table 6.9 Means, standard deviations and ranks of each model based on selected qualitative (subjective) criteria

Model	MON			HYD			FDC		
	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank	Mean	St. Dev.	Rank
A	2.98	0.78	1	2.84	0.69	1	2.24	0.89	3
B1	2.48	0.68	6	2.04	0.75	6	1.70	0.72	6
B2	2.54	0.66	5	2.10	0.75	5	1.82	0.71	5
C	2.72	0.79	4	2.76	0.61	3	2.10	0.89	4
D1	2.72	0.60	3	2.84	0.76	2	2.30	0.65	2
D2	2.80	0.74	2	2.70	0.68	4	2.66	0.57	1
TOTAL	2.71	0.72	-	2.55	0.78	-	2.14	0.80	-

All the models simulated the monthly flow regime reasonably well, but, overall, model A performed best, followed by model D2. Model B1 performed worst with model B2 nearly as bad. Models C and D1 scored the same. The results for the simulation of the daily flow regime suggest that models B1 and B2, which do badly on the majority of catchments, are not performing as well as models A, C, D1 and D2, which simulate the observed flow hydrographs to an acceptable standard for nearly all the catchments. Overall, models A and D1 performed best, followed by models C, D2, B2 and B1 in that order. For the simulated flow duration curve, models A, C, D1 and D2 perform to an acceptable standard on more than half the catchments, whilst models B1 and B2 reach this standard only on a small handful of catchments. Overall, model D2 performed best, followed by models D1, A, C, B2 and B1 in that order.

Summary of model performance on basis of qualitative criteria

Model A ranks top in two of the three categories: the highest average scores for the monthly and daily flows. Model D1 also had a very high average score for the daily flows. Model D2 had the highest average score for the flow duration curves. On the basis of the qualitative criteria alone, model A performs best on average, closely followed by models D1 and D2. Models B1 and B2 are worst.

The baseflow-dominated catchments, i.e. the catchments with the least variable flows, tend to have the better monthly flows and flow duration curves, but the worst daily flows. However, the inability of the models to reproduce the entire daily flow hydrographs properly on the baseflow-dominated catchments is probably the result of a combination of factors: it is likely that the models fail to simulate the runoff generation process properly, particularly the slow release of water during dry periods, and additionally there may be problems with the input rainfall (spatially, temporally or through conversion to and from snow). These reasons would also account for the failure of the models to simulate the larger flow peaks properly. The problem catchments, 29003, 38021 and 42003, continued to perform badly.

6.3 SUMMARY

For a model to be judged to perform well it should closely reproduce the basic properties of the flow regime, characterised by average annual runoff totals, monthly flows, daily flows, flow duration curves and BFIs, on a large number of catchments of varied sizes, topographies, soils and climates. Tables 6.10 and 6.11 summarise the results. Table 6.10 shows the model ranks for each performance criterion, together with the overall rank, whilst Table 6.11 lists the catchments and performance criteria, and highlights those catchments which were noted as having consistently particularly good or bad results.

Table 6.10 Assessment of model performance

Model	Quantitative rank				Qualitative rank			Overall rank
	OBJ1	EFF	ANRO	BFI	MON	HYD	FDC	
A	3	2	1	6	1	1	3	1
B1	6	6	2	3	6	6	6	6
B2	5	3	3	2	5	5	5	5
C	1	1	4	1	4	3	4	1
D1	2	5	5	4	3	2	2	3
D2	3	6	5	2	2	4	1	3

Table 6.11 Assessment of catchment performance

Catchment	Quantitative rank				Qualitative rank		
	OBJ1	EFF	ANRO	BFI	MON	HYD	FDC
11001	-	-	-	-	-	BAD	-
19001	-	-	-	-	-	-	-
21018	-	-	-	-	GOOD	-	-
24004	-	-	-	BAD	-	-	BAD
25006	BAD	-	GOOD	-	-	-	BAD
28008	-	-	GOOD	-	-	GOOD	-
29003	GOOD	BAD	BAD	GOOD	BAD	BAD	-
32003	-	-	-	-	-	-	-
34004	GOOD	-	GOOD	BAD	GOOD	BAD	GOOD
37005	-	-	GOOD	-	-	GOOD	-
38021	BAD	BAD	BAD	-	BAD	-	BAD
39008	-	-	-	GOOD	-	-	-
39019	GOOD	GOOD	GOOD	GOOD	GOOD	BAD	GOOD
40007	-	-	-	-	-	-	-
42003	BAD	BAD	BAD	-	BAD	-	BAD
43005	GOOD	GOOD	-	GOOD	GOOD	-	GOOD
47001	-	-	GOOD	-	GOOD	GOOD	-
48004	-	GOOD	-	GOOD	-	-	BAD
54008	-	-	-	-	-	-	BAD
54016	-	-	-	-	-	-	-
57004	BAD	-	-	-	-	-	-
58009	-	-	-	-	-	-	-
66011	BAD	-	-	-	-	GOOD	BAD
76005	BAD	-	-	-	-	-	BAD
79006	BAD	-	-	-	-	GOOD	BAD

Of the quantitative measures, the Nash-Sutcliffe efficiency criterion and the percentage errors in the average annual runoff total tended to agree with each other, whilst the relative errors in the BFIs tended to be far too inconsistent to be of real use. Of the qualitative criteria, the quality of the simulated monthly flow did not discriminate enough between model performances, though in general agreed with the efficiency criterion and the percentage errors in the average annual runoff total. The quality of the simulated daily flow was too subjective a measure, with the authors disagreeing on what constituted a good or bad performance.

Marks for the quality of the simulated flow duration curve were more consistent. No model performed well in all of the assessment categories. Models B1 and B2 were clearly inappropriate, but, of the other four, where one model performed well, all tended to.

Deciding which of the performance criteria are the most important is difficult. For instance, does it matter that a model may tend to simulate daily flows well but flow duration curves badly, or that another model may tend to give low quantifiable errors, but simulate qualitative criteria poorly? Ultimately the relative importance of the performance criteria will depend on what the model will be used for. Therefore, which is considered the best model will vary i.e. model D2 gives by far the best simulated flow duration curves, but model A has the best simulated daily flows, and model C has the highest average efficiency. The minimum requirement for a good model performance is probably a high efficiency and a well simulated daily flow; well simulated average annual runoff totals, monthly flows and flow duration curves should be a natural consequence. Overall, the 5-parameter models A and C performed best and models B1 and B2 worst (five and four parameters respectively); the 3-parameter models D1 and D2 also performed acceptably in many of the categories.

To resolve the outstanding points from Section 5, it can be concluded that

- For a particular model on a particular catchment, the objective measures apart from the BFIs (and also excluding the objective functions) tend to give the same result; the subjective measures are less consistent. The objective and subjective measures can, but do not necessarily, give the same result.

- The distinction between models that performed well or badly was fairly consistent. Models B1 and B2 were earmarked as inappropriate at a relatively early stage, and this was confirmed after consideration of the quality of the simulated daily flows and flow duration curves.

- There are certain catchments on which none of the models perform well, in particular 29003, 38021 and 42003. On some others, all the models gave good results (except sometimes for simulation of the daily flow regime); these tended to be the baseflow-dominated catchments.

The next section briefly considers the relationships between the model parameters and measured descriptive indices of the catchments, in order to assess the likely ease of parameter estimation at ungauged sites.

7 Model parameter estimation

Bonvoisin and Boorman (1992a) found that with results from only four catchments it was not possible to derive statistical relationships between model parameters and mapped physical or climatic characteristics of the catchments. It is necessary to do this in order to extend the use of such models to ungauged catchments, and this section describes the brief investigation carried out to see if such relationships could be derived. Models A, C, D1 and D2, all of which performed adequately on the majority of catchments, were used. Correlation coefficients between the model parameters and selected FSR catchment characteristics were determined. Because of the large range of values for most of the variables, both the catchment characteristics and the model parameters were transformed by taking common logarithms (i.e. to base 10). Since the LAKE and URBAN terms often have zero values, 1.0 was added before taking the logarithm; hence for these variables, if the original term was zero, the transformed term will also be zero.

The correlations of the transformed catchment characteristics are given in Table 7.1. The figures quoted are significant at the 95% level. The most significant correlations are between catchment area and mainstream length (0.94), and between the SOIL index and BFI (-0.72). Channel slope is correlated with area (-0.58), mainstream length (-0.60), and SAAR (0.60). These findings, from the set of 25 catchments used here, reflect those from the much larger FSR dataset (FSR I, 4(312), Table 4.8). LAKE and URBAN are not significantly correlated with any variables, and were not used further in the regression analysis because so few of the catchments in the dataset were affected (only nine out of 25 were at all urbanised and just three of those were more than 5% urbanised; only five out of 25 had significant lakes).

Table 7.1 Significant correlations of log-transformed catchment characteristics from 25 catchments used in study

	AREA	MSL	S1085	SOIL	SAAR	URBAN	LAKE	BFI
AREA	1.00	0.94	-0.58	-	-	-	-	-
MSL	0.94	1.00	-0.60	-	-	-	-	-
S1085	-0.58	-0.60	1.00	0.47	0.60	-	-	-0.41
SOIL	-	-	0.47	1.00	0.43	-	-	-0.72
SAAR	-	-	0.60	0.43	1.00	-	-	-
URBAN	-	-	-	-	-	1.00	-	-
LAKE	-	-	-	-	-	-	1.00	-
BFI	-	-	-0.41	-0.72	-	-	-	1.00

Although correlations were initially carried out on the complete set of 25 catchments, further analyses were done, omitting the three catchments which performed consistently badly (i.e. 29003, 38021 and 42003). Tables 7.2-7.5 show the correlation matrices for each model parameter, the catchment characteristics and the other model parameters, for both the complete dataset and the reduced dataset. Again only correlation coefficients significant at the 95 % level are quoted. In the following sections the relationships between the model parameters and catchment characteristics are grouped by parameter function.

Table 7.2a *Significant correlations of log-transformed model parameters and log-transformed catchment characteristics for model A*

	AREA	MSL	S1085	SOIL	SAAR	BFI	*Ca*	*Cb*	*Cc*	*Cr1*	*Cr2*
Ca	-	-	-	-	-	-0.40	1.00	-	0.53	-	-
	[-]	[-]	[-]	[0.54]	[-]	[-0.58]	[1.00]	[-]	[0.67]	[-]	[-]
Cb	-	-	-	-	0.44	-	-	1.00	-	-	0.52
	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[1.00]	[-]	[-]	[0.62]
Cc	-	-	-	0.67	-	-0.70	0.53	-	1.00	-	-
	[-]	[-]	[-]	[0.70]	[-]	[-0.61]	[0.67]	[-]	[1.00]	[-]	[-]
Cr1	-	-	-	-	-	-	-	-	-	1.00	-
	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[1.00]	[-]
Cr2	-	-	-	0.42	0.44	-0.54	-	0.52	-	-	1.00
	[-]	[-]	[-]	[0.50]	[0.51]	[-0.66]	[-]	[0.62]	[-]	[-]	[1.00]

where: top figure in each pair refers to 25-catchment dataset and figures in brackets refer to the reduced dataset of 22 catchments

Table 7.2b *Significant correlations of log-transformed model parameters and log-transformed catchment characteristics for model C*

	AREA	MSL	S1085	SOIL	SAAR	BFI	*cmax*	*smax*	*Kb*	*Grout*	*Srout*
cmax	-	-	-	-	-	-	1.00	-	-	-	-
	[-]	[-]	[-]	[-]	[-]	[-]	[1.00]	[-]	[-]	[-]	[-]
smax	-	0.42	-	-0.42	-0.49	0.45	-	1.00	-	-	-
	[-]	[-]	[-]	[-]	[-0.51]	[0.55]	[-]	[1.00]	[-]	[-]	[-0.48]
Kb	-	-	0.55	-	0.76	-	-	-	1.00	-	-
	[-]	[-]	[0.59]	[-]	[0.84]	[-]	[-]	[-]	[1.00]	[-]	[-]
Grout	-	-	-	-	-	-	-	-	-	1.00	-
	[-0.51]	[-]	[0.47]	[0.53]	[-]	[-]	[-]	[-]	[-]	[1.00]	[-0.50]
Srout	-	-	0.50	0.68	0.41	-0.61	-	-	-	-	1.00
	[-]	[-]	[0.50]	[0.82]	[0.48]	[-0.71]	[-]	[-0.48]	[-]	[-0.50]	[1.00]

where: top figure in each pair refers to 25-catchment dataset and figures in brackets refer to the reduced dataset of 22 catchments

Table 7.2c *Significant correlations of log-transformed model parameters and log-transformed catchment characteristics for model D1*

	AREA	MSL	S1085	SOIL	SAAR	BFI	*smax*	*K1*	*K2*
smax	-	-	-	-	-0.60	-	1.00	-0.91	-
	[-]	[-]	[-0.47]	[-]	[-0.66]	[0.59]	[1.00]	[-0.88]	[-0.52]
K1	-	-	-	-	0.72	-	-0.91	1.00	-
	[-]	[-]	[0.52]	[-]	[0.78]	[-]	[-0.88]	[1.00]	[0.58]
K2	-	-	-	0.82	0.49	-0.60	-	-	1.00
	[-]	[-]	[0.46]	[0.80]	[0.48]	[-0.63]	[-0.52]	[0.58]	[1.00]

where: top figure in each pair refers to 25-catchment dataset and figures in brackets refer to the reduced dataset of 22 catchments

Table 7.2d *Significant correlations of log-transformed model parameters and log-transformed catchment characteristics for model D2*

	AREA	MSL	S1085	SOIL	SAAR	BFI	*smax*	*K1*	*K2*
smax	-	-	-0.40	-	-0.63	-	1.00	-0.78	-0.47
	[-]	[-]	[-0.43]	[-]	[-0.67]	[0.51]	[1.00]	[-0.77]	[-0.54]
K1	-	-	-	-	0.57	-	-0.78	1.00	0.51
	[-]	[-]	[-]	[-]	[0.57]	[-]	[-0.77]	[1.00]	[0.64]
K2	-	-	0.41	0.76	0.45	-0.69	-0.47	0.51	1.00
	[-]	[-]	[0.46]	[0.74]	[0.46]	[-0.72]	[-0.54]	[0.64]	[1.00]

where: top figure in each pair refers to 25-catchment dataset and figures in brackets refer to the reduced dataset of 22 catchments

7.1 PARAMETERS CONCERNED WITH EVAPORATION

The only parameter solely concerned with determining actual evaporation from potential evaporation is the evapotranspiration coefficient *Ca* in model A. Although other models have parameters used in the procedure to determine actual evaporation e.g. the maximum amount of storage in the basin *smax* in model C, these parameters are essentially storage accounting parameters, and are considered in Section 7.2.

The model A evapotranspiration coefficient *Ca* is negatively correlated with BFI for the complete dataset, and this relationship becomes stronger for the 22-catchment dataset; a positive correlation with the SOIL index also becomes significant for the reduced dataset. There is also a relationship between *Ca* and the subsurface flow coefficient *Cc*.

Ca determines the amount of potential evapotranspiration that is satisfied i.e. the larger the value of *Ca*, the less evapotranspiration takes place. The relationships suggest that the dominant control on the value of *Ca* is the soil permeability and the way it controls the relationship between water movement and soil moisture, as seen in the SOIL index. A low soil permeability (i.e. high SOIL index) is associated with high values of *Ca* and *Cc* i.e. low evaporation loss and low subsurface flow. This is then consistent with a low value of BFI.

7.2 PARAMETERS CONCERNED WITH STORAGE ACCOUNTING

The models have as many as three parameters concerned with storage accounting to determine the water content of each soil moisture store. Some of these parameters also have supplementary roles in determining actual evaporation from potential evaporation. Model A has the two subsurface flow coefficients *Cb* and *Cc*, model C has the maximum storage capacity at any point within the basin *cmax*, the average maximum amount of water that could be held in storage over the whole basin *smax* and the soil drainage coefficient *Kb*, and models D1 and D2 have the capacity of the largest soil moisture store *smax* and the subsurface flow coefficient *K1*.

Taking model A, *Cb* is positively correlated with SAAR for the complete dataset only, whilst *Cc* is correlated with both the SOIL index (positive) and BFI (negative) for both datasets. There are also relationships between *Cb* and the routing coefficient *Cr2*, and between *Cc* and the evapotranspiration coefficient *Ca*. The relationships for *Cc* are consistent with those for *Ca*, and with the interpretation of soil permeability as the dominant control. *Cb* controls the rate of subsurface flow rather than its variation with soil moisture content, and the relationship with SAAR suggests that this is higher in wetter catchments. Other catchment characteristics which might be related to this parameter, such as hillslope topography (P. Broadhurst, pers.comm.), are not represented in the FSR catchment characteristic dataset.

In model C, the maximum storage capacity at any point within the basin *cmax* is not significantly correlated with any of the catchment characteristics or with any of the other model parameters for either dataset. The average maximum amount of water that can be held in storage over the whole basin *smax* is correlated with the mainstream length (positive), the SOIL index (negative), SAAR (negative) and BFI (positive) for the complete dataset. However, with the 22-catchment dataset the first two of these relationships become non-significant, whilst the latter two strengthen, and a relationship with the channel routing coefficient *Srout* becomes significant. The soil drainage coefficient *Kb* is positively correlated with channel slope and SAAR, the relationships again strengthening with the 22-catchment dataset. In the case of *smax*, the relationships suggest that the soil moisture storage of the basin is low in areas of high rainfall, impermeable soils, low baseflows and short mainstream lengths. These are typically small upland catchments. The lack of correlation between *cmax* and any of the catchment characteristics used may be interpreted in terms of the lack of detailed soil information represented in the SOIL index, either in terms of the maximum soil moisture storage at a point or the within-catchment distribution of soil moisture storages. *Kb* is the rate of lateral subsurface drainage which is seen here to be dominated by slope angle, in this case represented by both channel slope and SAAR.

For model D1, the capacity of the largest soil moisture store *smax* is negatively correlated with SAAR and the subsurface flow coefficient *K1*, but for the reduced dataset there are also significant correlations with slope (negative), BFI (positive) and the routing coefficient *K2*. For model D2, *smax* is correlated with more of these variables for the complete dataset i.e. SAAR (negative), slope (negative), *K1* and *K2*; and with BFI (positive) for the 22-catchment dataset only. As well as the relationship with *smax*, *K1* is also positively correlated with SAAR for both models. With the 22-catchment dataset for model D1, *K1* is also positively correlated with slope and *K2*. For model D2 there is no correlation between *K1* and slope, but the correlation with *K2* is also significant for the complete dataset. The similarity of relationships between models D1 and D2 is gratifying, although the details vary slightly. In both cases, higher soil moisture storage capacities are associated with drier catchments of lower slopes and higher baseflows. This is in agreement with the way in which soil properties vary across the UK, and indeed may reflect an uplands/lowlands divide. The larger soil moisture capacity also has a lower drainage rate which is consistent with the lower slope angles of lowland catchments.

7.3 PARAMETERS CONCERNED WITH RUNOFF GENERATION AND ROUTING

All of the models have at least one parameter concerned with the generation of runoff or with the routing of that runoff to the catchment outfall. Model A has the two routing coefficients *Cr1* and *Cr2*, Model C has the groundwater discharge coefficient *Grout* and the channel routing coefficient *Srout*, and models D1 and D2 have the routing coefficient *K2*.

The model A routing coefficient *Cr1* is not significantly correlated with any of the catchment characteristics, but the other model A routing coefficient *Cr2* is positively correlated with the SOIL index and SAAR, and negatively correlated with BFI, the relationships strengthening with the 22-catchment dataset. *Cr2* is also correlated with the subsurface flow coefficient *Cb*. Catchment characteristics associated with routing are limited within the FSR catchment characteristics dataset to the mainstream length, and are clearly insufficient to demonstrate a link with these model parameters. However, it may be suggested that the positive relationship between *Cr2* and both the SOIL index and SAAR implies that *Cr2* might be related to drainage density i.e. the higher the drainage density, the greater the value of *Cr2*. This might be tested with more recently-derived catchment characteristics (e.g. Naden & Polarski, 1990).

The model C groundwater discharge coefficient *Grout* is not significantly correlated with any of the catchment characteristics or with any of the other model parameters for the complete dataset, but for the 22-catchment dataset *Grout* is correlated with area (negative), slope (positive) and the SOIL index (positive). The model C channel routing coefficient *Srout* is positively correlated with slope, SAAR, and the SOIL index, and negatively correlated with BFI. Additionally for the 22-catchment dataset there is the significant correlation between *Srout* and *Grout*, and *Srout* is also correlated with the maximum amount of storage in the basin *smax*. Neither groundwater routing not channel routing is well represented in the FSR catchment characteristics dataset. However, the relationships again point to a possible link with drainage density, and a link to slope angle in the case of *Srout*.

For both models D1 and D2, the routing coefficient *K2* is positively correlated with SAAR and the SOIL index, and negatively correlated with BFI, for both datasets. Model D2 also has *K2* positively correlated with slope and negatively correlated with the capacity of the largest soil moisture store *smax*, but for model D1 the relationships are only significant for the 22-catchment dataset. Again, *K2* is seen to have possible links with drainage density and slope angle.

The consistency in the relationships between parameters which describe routing in each of the different models and catchment characteristics is very encouraging, suggesting that these parameters are indeed playing a similar role.

7.4 SUMMARY

This section has gone some way to showing that, from the significant correlations between some of the model parameters and catchment characteristics, it might be possible to derive regression equations for estimation of those model parameters at ungauged sites. However, other model parameters seem quite unrelated to the variables used, and their estimation might be difficult. Furthermore, some of the model parameters were significantly and highly correlated with other model parameters, which is not a particularly good feature of a model, and some further refinement of model structure and parameters may be necessary.

The relative dominance of the upland/lowland divide (indexed through SAAR) and permeability characteristics (indexed through the SOIL index and BFI) compared to the unimportance of the topographical indices such as area, stream length and slope was disappointing. However, these results reflect the restricted set of variables used in the FSR catchment characteristics dataset. Another potentially important characteristic, stream frequency, was not included in this part of the study due to the difficulties of abstracting it manually for very large catchments, but another more discriminating channel variable might be total network length and, hence, drainage density.

As a consequence of these disappointing results, the subsequent regression exercise has not been carried out as part of this study. Other reasons contributed to this decision: firstly, there would ideally be a larger dataset; additionally, there needs to be a more formal way of identifying potential problem catchments; and finally, with continuing advances in digital mapping and information technology, there are possibilities for deriving catchment characteristics by computerised methods which might replace the manually-derived FSR catchment characteristics (Naden & Polarski, 1990). Such digital data sets include the digital terrain model (DTM) for the UK (Institute of Hydrology; Morris *et al.*, 1990), the digitised UK river network (from 1:50K OS maps), the HOST (Hydrology Of Soil Types) classification of soils (Boorman *et al.*, 1994), and the land cover dataset (Fuller, 1993). Future work will make effective use of these datasets in order to fulfil the ultimate objectives of the project.

8 Conclusions

In this study six simple conceptual daily rainfall-runoff models with three to five parameters were applied to 25 UK catchments covering a variety of sizes, topographies, soils and climates. Firstly, the model parameters for each of the 25 catchments were estimated for the calibration period (1981 to 1983, with 1980 as a warm-up year). Rather than use a completely automatic optimisation routine for calibration, a more subjective calibration system was employed. Automatic optimisation from a reasonably sensible set of starting parameter values was followed by manual adjustment of the optimised parameters to obtain better fits as judged by the user. These model parameters were then used with a validation dataset from 1984 to 1989 (with 1983 as the warm-up year).

Model evaluation was based on a comparison of observed and simulated flows during the calibration and validation periods. Model performance was judged on the ability of the model to reproduce the basic characteristics of the flow, such as average annual runoff totals, monthly flow, daily flow, flow duration curve and baseflow index, rather than to provide a near-exact simulation of the observed response.

Examination of the results shows that the quantitative criteria were significantly and generally highly correlated between the calibration and validation periods, though there was no correlation between the two objective functions themselves. A catchment with a high efficiency tends to have a small percentage error in the estimated average annual runoff total, but the relative error in estimating the BFI can be either large or small. The average annual runoff totals tend to be underestimated, whilst the BFIs tend to be overestimated. The simulated monthly flow regimes were of similar, good, quality in the calibration and validation periods. The simulated daily flow regimes were sometimes better in the calibration period, and sometimes better in the validation period. The simulated flow duration curves tended to be much better in the calibration period. A catchment where the monthly flow regime is simulated well can have, but does not necessarily have, a well simulated daily flow regime. Similarly a catchment where the daily flow regime is simulated well can have, and indeed tends to have, a well simulated flow duration curve. However a catchment with a well simulated flow duration curve does not necessarily have well simulated daily and monthly flows.

The results show that four of the models perform well on the majority of catchments. The best performances were from the 5-parameter models A (Bonvoisin & Boorman, 1992a; model 8) and C (Moore, 1985; probability-distributed model PDM). The 3-parameter models D1 and D2 (Bonvoisin & Boorman, 1992a; model 16) also performed well in many of the categories, and model D2 gave the best simulated flow duration curves. Models B1 and B2 (Bonvoisin & Boorman, 1992a; model 15), with five and four parameters respectively, were clearly inappropriate, due to their large quantifiable errors and their poor performances in the qualitative categories.

The overall conclusions to be drawn from this study are that for a particular model on a particular catchment, the objective measures, apart from the BFIs, tend to give the same result; the subjective measures are less consistent. The objective and subjective measures can, but do not necessarily, give the same result. The distinction between models that performed well or badly was fairly consistent. Models B1 and B2 were earmarked as inappropriate at a relatively early stage, and this was confirmed after consideration of the quality of the simulated daily flows and flow duration curves. There are certain catchments on which none of the models perform well, in particular 29003, 38021 and 42003. On some others, all the models gave good results (except sometimes for simulation of the daily flow regime); these tended to be the baseflow-dominated catchments. The relatively good performance of the baseflow-dominated catchments was almost certainly because of the smaller variability of their day-to-day flows, and because the faster responding, more impervious catchments were not well represented by a daily time step. The more impervious catchments had better simulations of the daily flow regime.

It is unlikely that a single generally applicable simple conceptual model is attainable, as there will always be catchments on which the model does not perform well. The criteria which may be examined for a definitive assessment of model performance vary with the ultimate purpose of the model, but a high efficiency and a well simulated daily flow should indicate a generally good model performance. It is not possible to define model performance solely on the basis of objective criteria, as on a particular catchment a model may give a high efficiency and small quantifiable errors, but have a poorly simulated daily flow regime. Similarly it is not possible to define model performance solely on the basis of subjective criteria, as a well simulated daily flow may imply a well simulated monthly flow and a well simulated flow duration curve, but not only do different users have different perceptions of what constitutes a good or bad fit, it is also almost always necessary to be able to quantify the fit through a criteria such as efficiency in order to compare models.

The regression exercise carried out in order to derive statistical relationships between model parameters and mapped physical and climatic characteristics of the catchments gave disappointing results, despite eliminating the three catchments on which none of the models performed well. The results reflect the restricted set of variables used in the FSR catchment characteristics dataset, and the need to extend this dataset to include characteristics which can now be derived by computerised methods.

Future work will involve using model C to assess the potential for better estimation of flood frequency through continuous simulation of catchment flows (Spijkers & Naden, 1994). It is likely that the ability of the model to fit the observed flow data in this study was constrained by the time step adopted; for instance, flood peaks may have been modelled better had hourly data rather than daily data been used. Therefore, future work will also investigate the potential for continuous simulation at a sub-daily time step, as well as the possibilities of making use of newly available digital datasets.

Acknowledgements

This study was commissioned by the Flood and Coastal Defence Division of the Ministry of Agriculture, Fisheries and Food Flood Defence Division (Project FD0404).

References

Arnell, N.W. & Reynard, N.S. 1993. Impact of climate change on river flow regimes in the United Kingdom. Institute of Hydrology Report to DOE, Wallingford.

Bonvoisin, N.J. & Boorman, D.B. 1992a. Daily rainfall-runoff modelling as an aid to the transfer of hydrological parameters. Institute of Hydrology Report to MAFF, Wallingford.

Bonvoisin, N.J. & Boorman, D.B. 1992b. MIMIC User's Manual (version 1.3). Institute of Hydrology, Wallingford.

Boorman, D.B., Gannon, B., Gustard, A., Hollis, J. & Lilly, A. 1994. Hydrological Aspects of the HOST Classification of Soils. Report to MAFF. Institute of Hydrology, Wallingford.

Boorman, D.B. & Houghton-Carr, H.A. 1992. UK Representative Basin and Flood Event Database Manual. Institute of Hydrology Report to MAFF, Wallingford.

Fleming, G. 1975. Computer simulation techniques. Elsevier.

Fuller, R.M. 1993. The Land Cover Map of Great Britain. Earth Space Review 2, 13-18.

Gan, T.Y & Burges, S.J. 1990. An Assessment of a Conceptual Rainfall-Runoff Model's Ability to Represent the Dynamics of Small Hypothetical Catchments: 1. Models, Model Properties and Experimental Design. Water. Resour. Res. 26(7), 1595-1604.

Gustard, A., Bullock, A. & Dixon, J.M. 1992. Low Flow Estimation in the United Kingdom. Institute of Hydrology Report No. 108, Wallingford.

Harding, R.J. & Moore, R.J. 1988. Assessment of snowmelt models for use in the Severn-Trent flood forecasting system. Report to Severn-Trent Water Authority. Institute of Hydrology, Wallingford.

Harding, R.J. & Moore, R.J. 1992. PACK: A Pragmatic Snowmelt Model for Real-time Use. Institute of Hydrology Report to MAFF, Wallingford.

Jones, S.B. 1983. The estimation of catchment average point rainfall profiles. Institute of Hydrology Report No. 79, Wallingford.

Moore, R.J. 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrol. Sci. J. 30(2), 263-297.

Moore, R.J., Jones, D.A., Bird, P.B. & Cottingham, M.C. 1990. A basin-wide flow forecasting system for real-time flood warning, river control and water management. International Conference on River Flood Hydraulics, 17-20 September, Wallingford, UK.

Moore, R.J. & Jones, D.A. 1991. A river flow forecasting system for region-wide application. MAFF Conference of River and Coastal Engineers, 8-10 July 1991, Loughborough, UK.

Moore, R.J. 1993. Real-time flood forecasting systems: perspectives and prospects. UK-Hungarian Workshop on Flood Defence, 6-10 September 1993, Budapest, Hungary.

Morris, D.G., Flavin, R.W. & Moore, R.V. 1990. A digital terrain model for hydrology. 4th International Symposium on Spatial Data Handling, Zurich, 250-262.

Naden, P.S. & Polarski, M. 1990. Derivation of river network variables from digitised data and their use in flood estimation. Institute of Hydrology Report to MAFF, Wallingford.

Nash, J.E. & Sutcliffe, J.V. 1970. River flow forecasting through conceptual models: 1. A discussion of principles. J. Hydrology, 10, 282-290.

Natural Environment Research Council (NERC). 1975. Flood Studies Report (5 vol.). NERC, London.

Rosenbrock, H.H. 1960. An automatic method of finding the greatest or least value of a function. Computer J., 3, 175-184.

Spijkers, T.M.G. & Naden, P.S. 1994. Continuous rainfall-runoff modelling for flood estimation : initial thoughts and data requirements. Institute of Hydrology Report to MAFF, Wallingford.

Thompson, N., Barrie, I.A. & Ayles, M. 1981. The Meteorological Office Rainfall and Evaporation Calculation System (MORECS). Hydrological Memorandum 45, Met. Off., Bracknell.

Wilmott, C.J., Rowe, C.M. & Mintz, Y. 1985. Climatology of the terrestrial seasonal water balance. J. Climatol. 5, 589-606.

Appendix A

A.1 Model parameters, objective functions and efficiency criteria

	Catchment	Model parameters					Calibration period			Validation period		
		Ca [mm^{-1}] x10^{3}	*Cb* [mm d^{-1}]	*Cc* [mm^{-1}] x10^{2}	*Cr1* [-] x10^{2}	*Cr2* [-]	COBJ1	COBJ2	CEFF	VOBJ1	VOBJ2	VEFF
*	11001	2.651	18.391	3.030	0.979	0.650	0.372	0.077	0.707	0.461	0.114	0.694
*	19001	3.453	14.602	6.803	0.169	0.800	1.524	0.194	0.698	1.639	0.217	0.682
*	21018	0.881	12.544	4.504	0.928	0.650	0.603	0.081	0.772	0.754	0.097	0.727
	24004	7.743	12.535	8.071	3.325	0.546	1.031	0.208	0.736	1.251	0.474	0.653
*	25006	6.163	10.086	9.711	2.361	0.800	7.372	0.391	0.645	9.144	0.597	0.441
	28008	4.274	30.300	5.197	1.374	0.676	0.590	0.072	0.761	0.592	0.109	0.740
*	29003	3.257	9.625	2.500	1.512	0.800	0.108	0.208	0.577	0.072	0.248	0.645
	32003	2.670	10.873	9.276	0.490	0.731	0.431	0.223	0.637	0.348	0.286	0.532
	34004	3.082	10.010	3.789	0.325	0.630	0.095	0.088	0.729	0.100	0.082	0.665
*	37005	2.570	10.723	10.000	0.500	0.550	0.214	0.297	0.465	0.205	0.338	0.489
*	38021	1.371	6.362	15.000	0.050	0.800	0.645	0.907	0.311	0.568	1.443	0.352
	39008	1.412	48.279	3.782	1.648	0.668	0.094	0.084	0.535	0.090	0.177	0.814
	39019	0.477	1.875	0.935	1.071	0.302	0.005	0.012	0.922	0.015	0.062	0.739
*	40007	1.357	40.390	6.050	0.558	0.750	0.763	0.174	0.626	1.163	0.297	0.518
*	42003	3.234	19.300	5.500	0.012	0.600	0.623	0.397	0.505	0.632	0.466	0.481
*	43005	0.826	25.485	1.000	1.411	0.650	0.054	0.025	0.837	0.039	0.037	0.855
	47001	1.421	28.598	3.666	0.505	0.935	1.296	0.096	0.839	1.215	0.136	0.813
	48004	1.421	21.090	2.602	2.329	0.606	0.369	0.098	0.906	0.464	0.248	0.864
*	54008	0.885	20.750	2.313	1.000	0.950	0.806	0.281	0.579	0.626	0.342	0.617
*	54016	2.350	7.969	4.955	0.135	0.700	0.146	0.088	0.668	0.105	0.136	0.707
*	57004	4.398	34.700	2.729	0.050	0.900	4.507	0.130	0.849	4.614	0.187	0.828
*	58009	2.973	21.217	4.783	0.990	0.650	1.260	0.054	0.857	1.448	0.073	0.826
*	66011	1.978	25.541	4.500	0.990	1.000	14.410	0.226	0.757	14.620	0.336	0.696
*	76005	2.167	10.443	4.081	1.073	0.950	5.988	0.272	0.346	4.126	0.450	0.443
*	79006	6.006	24.648	4.500	0.000	0.889	5.560	0.247	0.730	5.063	0.173	0.696

Model parameters, objective functions and efficiency criteria for model A.

	Catchment	Model parameters					Calibration period			Validation period		
		smax [mm]	*Ci* [mm d^{-1}]	*K1* [d^{-1}]	*K2* [d^{-1}]	*Cl* [-]	COBJ1	COBJ2	CEFF	VOBJ1	VOBJ2	VEFF
	11001	128.663	31.846	0.153	0.022	0.107	0.469	0.142	0.630	0.525	0.153	0.651
	19001	63.804	23.770	0.402	0.024	0.262	1.499	0.217	0.704	1.803	0.280	0.651
	21018	154.352	27.956	0.186	0.009	0.091	0.851	0.120	0.678	0.813	0.135	0.706
	24004	51.450	33.954	0.499	0.153	0.242	1.526	0.850	0.609	1.729	1.692	0.521
*	25006	23.510	23.532	0.593	0.100	0.237	7.900	0.615	0.619	8.465	1.211	0.482
	28008	84.420	22.649	0.160	0.017	0.000	0.720	0.152	0.709	0.760	0.151	0.666
*	29003	244.521	5.467	0.026	0.003	0.000	0.179	0.210	0.297	0.130	0.171	0.359
	32003	105.274	28.277	0.341	0.005	0.299	0.445	0.262	0.626	0.432	0.291	0.418
*	34004	131.961	7.500	0.060	0.008	0.054	0.108	0.106	0.692	0.115	0.095	0.616
*	37005	106.373	7.623	0.042	0.005	0.265	0.247	0.213	0.383	0.223	0.265	0.446
*	38021	85.000	26.970	0.137	0.005	0.771	0.473	0.710	0.495	0.504	0.966	0.425
*	39008	165.211	29.708	0.050	0.001	0.000	0.142	0.398	0.750	0.143	0.425	0.706
	39019	455.461	16.650	0.006	0.002	0.075	0.016	0.029	0.759	0.016	0.042	0.728
*	40007	43.906	36.491	0.339	0.100	0.453	1.120	0.910	0.451	0.929	1.253	0.615
*	42003	59.163	24.674	0.151	0.025	0.556	0.880	0.494	0.301	1.023	0.503	0.159
	43005	192.429	29.552	0.022	0.006	0.000	0.062	0.062	0.813	0.044	0.058	0.839
*	47001	129.242	25.503	0.254	0.006	0.074	1.469	0.224	0.817	1.274	0.193	0.804
	48004	148.020	10.448	0.075	0.012	0.000	0.505	0.086	0.872	0.522	0.144	0.844
*	54008	161.350	26.934	0.090	0.002	0.000	0.731	0.308	0.618	0.508	0.344	0.689
	54016	122.493	23.697	0.065	0.006	0.255	0.152	0.099	0.654	0.138	0.130	0.613
*	57004	57.488	39.170	0.499	0.008	0.209	6.426	0.492	0.788	4.354	0.395	0.837
*	58009	125.912	9.453	0.100	0.025	0.000	4.529	0.130	0.581	3.055	0.122	0.633
	66011	44.905	20.575	0.587	0.152	0.165	12.004	0.395	0.798	13.530	0.608	0.718
*	76005	54.899	10.931	0.335	0.090	0.301	3.976	0.347	0.566	3.231	0.601	0.564
	79006	34.246	29.994	0.422	0.115	0.181	5.366	0.423	0.748	4.210	0.464	0.747

Model parameters, objective functions and efficiency criteria for model B1.

Catchment	smax [mm]	Cl [mm d⁻¹]	K1 [d⁻¹]	K2 [d⁻¹]	COBJ1	COBJ2	CEFF	VOBJ1	VOBJ2	VEFF
	Model parameters				Calibration period			Validation period		
11001	135.541	6.771	0.134	0.018	0.408	0.132	0.679	0.492	0.161	0.673
* 19001	75.379	23.219	0.344	0.010	1.608	0.206	0.682	1.637	0.245	0.683
* 21018	198.676	12.217	0.165	0.004	0.929	0.139	0.649	0.851	0.152	0.692
* 24004	115.520	4.377	0.200	0.005	1.463	0.262	0.625	1.404	0.371	0.611
* 25006	71.122	10.135	0.473	0.005	7.506	0.538	0.638	7.646	0.691	0.532
28008	119.282	3.136	0.130	0.019	0.702	0.126	0.716	0.728	0.145	0.680
* 29003	200.000	15.000	0.200	0.008	0.257	0.182	-0.009	0.085	0.141	0.579
* 32003	133.851	28.193	0.271	0.002	0.493	0.284	0.585	0.393	0.277	0.471
* 34004	128.426	3.561	0.040	0.008	0.103	0.108	0.707	0.105	0.091	0.648
37005	187.169	28.171	0.149	0.001	0.186	0.194	0.533	0.187	0.277	0.535
38021	278.873	5.469	0.113	0.000	0.566	0.858	0.395	0.439	0.986	0.500
* 39008	172.034	29.405	0.040	0.001	0.126	0.325	0.778	0.116	0.355	0.761
39019	436.724	11.127	0.005	0.002	0.014	0.028	0.786	0.015	0.045	0.743
40007	220.472	7.039	0.211	0.002	0.772	0.233	0.622	0.666	0.279	0.724
42003	296.266	2.528	0.091	0.000	0.765	0.633	0.393	0.896	0.838	0.264
43005	194.293	29.550	0.022	0.005	0.062	0.061	0.813	0.044	0.057	0.839
* 47001	131.022	26.275	0.254	0.005	1.526	0.232	0.810	1.326	0.202	0.796
* 48004	139.381	9.187	0.074	0.025	0.552	0.138	0.860	0.461	0.113	0.862
54008	171.882	27.187	0.079	0.002	0.177	0.280	0.625	0.502	0.322	0.693
54016	141.818	24.162	0.054	0.004	0.164	0.112	0.627	0.138	0.127	0.613
57004	124.933	12.399	0.368	0.027	5.372	0.153	0.823	4.760	0.217	0.822
58009	133.049	20.861	0.531	0.024	1.649	0.124	0.848	2.217	0.121	0.734
66011	49.790	20.213	0.583	0.126	12.442	0.396	0.790	12.950	0.483	0.731
76005	153.703	4.323	0.243	0.003	4.675	0.403	0.489	3.774	0.437	0.490
79006	39.490	29.879	0.417	0.085	5.560	0.360	0.739	4.494	0.323	0.730

Model parameters, objective functions and efficiency criteria for model B2.

Catchment	Model parameters					Calibration period			Validation period		
	cmax [mm]	*smax* [mm]	*Kb* [d^{-1}]	*Grout* [d^{-1}]	*Srout* [d^{-1}]	COBJ1	COBJ2	CEFF	VOBJ1	VOBJ2	VEFF
11001	279.605	136.790	2.537	2.761	0.472	0.276	0.075	0.782	0.457	0.122	0.696
19001	125.091	82.609	0.900	1.401	0.656	1.228	0.216	0.757	1.358	0.193	0.737
21018	219.581	138.321	1.959	32.936	0.611	0.687	0.156	0.740	0.585	0.182	0.788
24004	580.357	129.737	1.912	55.076	0.726	1.179	0.227	0.698	1.244	0.333	0.655
25006	5863.000	123.160	1.384	77.115	0.818	6.426	0.437	0.690	6.740	0.480	0.588
28008	180.285	85.514	1.531	2.935	0.486	0.538	0.082	0.782	0.579	0.096	0.746
29003	502.660	353.709	3.753	2.768	0.651	0.091	0.128	0.643	0.110	0.200	0.455
32003	185.968	163.592	0.236	331.897	0.669	0.363	0.356	0.694	0.415	0.458	0.441
34004	128.551	88.870	0.882	1.122	0.311	0.055	0.064	0.842	0.086	0.063	0.706
37005	187.919	130.397	0.264	1.774	0.460	0.157	0.218	0.606	0.109	0.251	0.730
38021	171.533	112.243	0.523	0.009	0.633	0.270	0.720	0.711	0.229	0.622	0.738
39008	220.298	147.041	0.685	5.774	0.253	0.115	0.204	0.798	0.092	0.267	0.812
39019	336.535	252.609	0.516	0.546	0.034	0.007	0.012	0.900	0.007	0.027	0.874
40007	316.505	182.498	0.567	2.542	0.560	0.509	0.129	0.743	0.547	0.189	0.773
42003	919.838	366.097	0.681	999.990	0.541	0.399	0.575	0.683	0.340	0.608	0.720
43005	205.556	155.267	1.500	1.841	0.132	0.056	0.044	0.830	0.041	0.041	0.849
47001	140.257	102.825	1.147	4.594	0.522	1.022	0.184	0.873	1.029	0.156	0.842
48004	505.202	201.044	5.191	34.148	0.552	0.532	0.343	0.865	0.513	2.223	0.846
54008	273.338	140.706	0.583	11.812	0.341	0.777	0.345	0.594	0.333	0.222	0.797
54016	273.372	171.276	0.629	15.823	0.393	0.118	0.118	0.731	0.076	0.138	0.789
57004	219.651	107.868	3.514	7.340	0.705	4.182	0.160	0.862	4.166	0.190	0.844
58009	141.753	100.421	2.628	3.831	0.760	1.206	0.089	0.888	1.770	0.092	0.788
66011	156.181	57.164	3.093	5.252	0.778	11.614	0.405	0.804	12.490	0.412	0.740
76005	6675.269	142.065	2.265	4.755	0.618	3.858	0.382	0.578	3.225	0.392	0.565
79006	141.169	33.221	3.438	4.524	0.698	5.076	0.418	0.761	4.007	0.356	0.760

Model parameters, objective functions and efficiency criteria for model C.

Catchment	Model parameters $smax$ [mm] $\times 10^{-2}$	$K1$ [$mm^{-1}d^{-1}$] $\times 10^{2}$	$K2$ [d^{-1}]	Calibration period COBJ1	COBJ2	CEFF	Validation period VOBJ1	VOBJ2	VEFF
11001	4.166	0.009	0.502	0.442	0.121	0.651	0.460	0.134	0.695
19001	2.829	0.006	0.499	1.410	0.201	0.721	1.623	0.230	0.685
21018	3.795	0.005	0.327	0.842	0.110	0.681	0.772	0.117	0.721
24004	2.454	0.008	0.390	1.495	0.165	0.617	1.327	0.233	0.632
25006	3.438	0.001	0.548	7.203	0.367	0.653	7.483	0.486	0.542
28008	3.149	0.009	0.362	0.712	0.136	0.712	0.795	0.147	0.651
29003	2.331	0.011	0.026	0.097	0.105	0.618	0.051	0.112	0.750
32003	5.053	0.001	0.508	0.419	0.284	0.647	0.414	0.299	0.443
34004	4.257	0.002	0.166	0.102	0.131	0.710	0.134	0.104	0.553
* 37005	5.000	0.001	0.500	0.319	0.223	0.201	0.349	0.318	0.133
* 38021	10.750	0.000	0.273	0.489	0.693	0.478	0.560	9.844	0.362
39008	4.515	0.001	0.103	0.166	0.175	0.708	0.148	0.175	0.696
39019	11.209	0.000	0.006	0.008	0.019	0.885	0.009	0.043	0.838
40007	6.008	0.001	0.322	0.798	0.143	0.609	0.716	0.169	0.703
* 42003	10.000	0.000	0.300	1.289	0.487	-0.024	1.243	0.486	-0.022
43005	3.095	0.004	0.019	0.038	0.056	0.887	0.030	0.048	0.889
47001	3.550	0.004	0.289	1.157	0.133	0.856	1.080	0.118	0.834
48004	4.851	0.012	0.682	0.413	0.112	0.895	0.413	0.105	0.879
54008	4.976	0.000	0.096	0.733	0.260	0.617	0.499	0.358	0.695
* 54016	6.975	0.000	0.300	0.342	0.135	0.222	0.233	0.162	0.348
57004	2.990	0.038	0.705	4.542	0.186	0.850	4.782	0.228	0.821
58009	3.004	0.018	0.707	1.135	0.186	0.895	1.645	0.157	0.803
66011	1.523	0.159	0.743	11.753	0.317	0.802	12.740	0.325	0.735
76005	3.707	0.001	0.255	4.695	0.363	0.487	3.816	0.481	0.485
79006	1.387	0.210	0.723	5.263	0.344	0.753	4.400	0.275	0.736

Model parameters, objective functions and efficiency criteria for model D1.

	Catchment	Model parameters $smax$ [mm] $\times 10^{-2}$	$K1$ [d^{-1}] $\times 10^{3}$	$K2$ [d^{-1}]	Calibration period COBJ1	COBJ2	CEFF	Validation period VOBJ1	VOBJ2	VEFF
	11001	3.283	1.847	0.462	0.457	0.152	0.639	0.497	0.183	0.670
*	19001	2.414	0.500	0.456	1.543	0.178	0.695	1.584	0.196	0.693
*	21018	3.024	0.500	0.305	1.111	0.124	0.580	0.874	0.105	0.684
*	24004	2.090	0.500	0.366	1.601	0.235	0.590	1.464	0.323	0.594
*	25006	2.346	0.250	0.542	7.221	0.433	0.652	7.503	0.557	0.541
	28008	2.291	1.737	0.334	0.711	0.111	0.712	0.799	0.123	0.649
*	29003	2.555	0.500	0.039	0.108	0.134	0.575	0.064	0.128	0.683
*	32003	8.168	0.050	0.423	0.435	0.257	0.634	0.495	0.359	0.334
	34004	3.656	0.513	0.154	0.101	0.097	0.712	0.135	0.098	0.548
	37005	3.630	0.137	0.220	0.169	0.147	0.577	0.174	0.219	0.567
*	38021	4.358	0.020	0.330	0.556	0.556	0.407	0.449	0.572	0.488
	39008	3.841	0.060	0.045	0.112	0.135	0.803	0.104	0.170	0.786
*	39019	10.000	0.003	0.006	0.008	0.019	0.883	0.008	0.036	0.857
*	40007	6.562	0.100	0.319	0.847	0.228	0.585	0.746	0.286	0.691
*	42003	5.932	0.025	0.135	1.023	0.539	0.188	1.100	0.569	0.096
	43005	3.280	0.477	0.023	0.033	0.026	0.901	0.028	0.027	0.898
*	47001	3.055	0.250	0.284	1.439	0.219	0.821	1.275	0.200	0.804
*	48004	3.063	1.000	0.084	0.551	0.078	0.860	0.622	0.106	0.817
	54008	4.761	0.080	0.098	0.732	0.257	0.617	0.495	0.321	0.697
	54016	3.533	0.146	0.071	0.184	0.143	0.581	0.155	0.179	0.565
	57004	2.499	2.328	0.415	5.280	0.135	0.826	4.124	0.193	0.846
	58009	2.533	2.439	0.632	1.261	0.105	0.883	1.827	0.105	0.781
*	66011	1.594	0.700	0.604	14.740	0.461	0.752	14.140	0.506	0.706
	76005	3.669	0.177	0.255	4.691	0.348	0.488	3.818	0.401	0.484
*	79006	1.149	0.900	0.436	5.951	0.347	0.720	4.972	0.304	0.702

Model parameters, objective functions and efficiency criteria for model D2.

Appendix A

A.2 Average annual runoff totals with percentage errors

	Calibration period			Validation period		
Catchment	Obs (mm)	Sim (mm)	CANRO (%)	Obs (mm)	Sim (mm)	VANRO (%)
11001	500	504	0.8	529	529	0.0
19001	553	567	2.5	580	553	-4.7
21018	570	568	-0.4	572	584	2.1
24004	521	518	-0.6	545	512	-6.1
25006	907	909	0.2	834	828	-0.7
28008	694	679	-2.2	619	618	-0.2
29003	303	272	-10.2	277	260	-6.1
32003	199	200	0.5	184	162	-12.0
34004	277	277	0.0	252	249	-1.2
37005	158	139	-12.0	153	140	-8.5
38021	179	201	12.3	164	127	-22.6
39008	298	299	0.3	260	280	7.7
39019	270	270	0.0	225	210	-6.7
40007	390	388	-0.5	382	332	-13.1
42003	338	303	-10.4	286	291	1.7
43005	367	352	-4.1	323	311	-3.7
47001	825	827	0.2	733	744	1.5
48004	997	908	-8.9	980	848	-13.5
54008	457	421	-7.9	398	381	-4.3
54016	255	259	1.6	221	205	-7.2
57004	1512	1570	3.8	1362	1407	3.3
58009	1086	993	-8.6	1046	943	-9.8
66011	1865	1970	5.6	1753	1637	-6.6
76005	770	856	11.2	720	775	7.6
79006	1245	1222	-1.8	1185	1193	0.7

Average annual runoff totals with percentage errors for model A.

	Calibration period			Validation period		
Catchment	Obs (mm)	Sim (mm)	CANRO (%)	Obs (mm)	Sim (mm)	VANRO (%)
11001	500	496	-0.8	529	501	-5.3
19001	553	547	-1.1	580	507	-12.6
21018	570	577	1.2	572	595	4.0
24004	521	490	-6.0	545	480	-11.9
25006	907	886	-2.3	834	807	-3.2
28008	694	659	-5.0	619	598	-3.4
29003	303	211	-30.4	277	197	-28.9
32003	199	197	-1.0	184	186	1.1
34004	277	289	4.3	252	260	3.2
37005	158	154	-2.5	153	157	2.6
38021	179	199	11.2	164	167	1.8
39008	298	275	-7.7	260	264	1.5
39019	270	258	-4.4	225	219	-2.7
40007	390	345	-11.5	382	327	-14.4
42003	338	335	-0.9	286	351	22.7
43005	367	334	-9.0	323	309	-4.3
47001	825	826	0.1	733	740	1.0
48004	997	925	-7.2	980	856	-12.7
54008	457	417	-8.8	398	396	-0.5
54016	255	254	-0.4	221	212	-4.1
57004	1512	1499	-0.9	1362	1287	-5.5
58009	1086	1047	-3.6	1046	999	-4.5
66011	1865	1865	0.0	1753	1547	-11.8
76005	770	757	-1.7	720	689	-4.3
79006	1245	1230	-1.2	1185	1200	1.3

Average annual runoff totals with percentage errors for model B1.

	Calibration period			Validation period		
Catchment	Obs (mm)	Sim (mm)	CANRO (%)	Obs (mm)	Sim (mm)	VANRO (%)
11001	500	544	8.8	529	547	3.4
19001	553	608	9.9	580	573	-1.2
21018	570	577	1.2	572	596	4.2
24004	521	469	-10.0	545	454	-16.7
25006	907	894	-1.4	834	810	-2.9
28008	694	688	-0.9	619	632	2.1
29003	303	299	-1.3	277	277	0.0
32003	199	212	6.5	184	194	5.4
34004	277	289	4.3	252	262	4.0
37005	158	152	-3.8	153	149	-2.6
38021	179	238	33.0	164	166	1.2
39008	298	271	-9.1	260	259	-0.4
39019	270	270	0.0	225	231	2.7
40007	390	405	3.8	382	358	-6.3
42003	338	399	18.0	286	377	31.8
43005	367	334	-9.0	323	308	-4.6
47001	825	838	1.6	733	752	2.6
48004	997	992	-0.5	980	926	-5.5
54008	457	412	-9.8	398	391	-1.8
54016	255	284	11.4	221	234	5.9
57004	1512	1661	9.9	1362	1434	5.3
58009	1086	1059	-2.5	1046	1010	-3.4
66011	1865	2075	11.3	1753	1747	-0.3
76005	770	861	11.8	720	778	8.1
79006	1245	1368	9.9	1185	1338	12.9

Average annual runoff totals with percentage errors for model B2.

	Calibration period			Validation period		
Catchment	Obs (mm)	Sim (mm)	CANRO (%)	Obs (mm)	Sim (mm)	VANRO (%)
11001	500	514	2.8	529	515	-2.7
19001	553	595	7.6	580	559	-3.6
21018	570	615	7.8	572	625	9.3
24004	521	501	-3.8	545	491	-9.9
25006	907	897	-1.1	834	815	-2.3
28008	694	650	-6.3	619	581	-6.2
29003	303	298	-1.7	277	266	-3.9
32003	199	161	-19.2	184	128	-30.2
34004	277	284	2.7	252	253	0.3
37005	158	158	0.2	153	157	2.6
38021	179	177	-1.0	164	153	-6.9
39008	298	296	-0.6	260	286	10.1
39019	270	268	-0.8	225	222	-1.2
40007	390	392	0.4	382	344	-10.1
42003	338	402	18.9	286	357	24.9
43005	367	350	-4.5	323	316	-2.2
47001	825	840	1.8	733	753	2.7
48004	997	980	-1.7	980	909	-7.2
54008	457	424	-7.2	398	408	2.5
54016	255	264	3.6	221	217	-1.7
57004	1512	1626	7.6	1362	1394	2.3
58009	1086	1013	-6.8	1046	950	-9.2
66011	1865	1989	6.7	1753	1668	-4.9
76005	770	904	17.4	720	830	15.3
79006	1245	1363	9.5	1185	1336	12.7

Average annual runoff totals with percentage errors for model C.

Catchment	Calibration period Obs (mm)	Sim (mm)	CANRO (%)	Validation period Obs (mm)	Sim (mm)	VANRO (%)
11001	500	524	4.8	529	536	1.3
19001	553	600	8.5	580	566	-2.4
21018	570	606	6.3	572	629	10.0
24004	521	490	-6.0	545	473	-13.2
25006	907	863	-4.9	834	780	-6.5
28008	694	651	-6.2	619	592	-4.4
29003	303	269	-11.2	277	250	-9.7
32003	199	207	4.0	184	175	-4.9
34004	277	277	0.0	252	242	-4.0
37005	158	140	-11.4	153	143	-6.5
38021	179	227	26.8	164	115	-29.9
39008	298	288	-3.4	260	279	7.3
39019	270	271	0.4	225	211	-6.2
40007	390	401	2.8	382	349	-8.6
42003	338	391	15.7	286	357	24.8
43005	367	342	-6.8	323	313	-3.1
47001	825	840	1.8	733	760	3.7
48004	997	994	-0.3	980	939	-4.2
54008	457	411	-10.1	398	382	-4.0
54016	255	260	2.0	221	190	-14.0
57004	1512	1679	11.0	1362	1454	6.8
58009	1086	1036	-4.6	1046	988	-5.5
66011	1865	2053	10.1	1753	1726	-1.5
76005	770	852	10.6	720	770	6.9
79006	1245	1381	10.9	1185	1354	14.3

Average annual runoff totals with percentage errors for model D1.

Catchment	Calibration period Obs (mm)	Sim (mm)	CANRO (%)	Validation period Obs (mm)	Sim (mm)	VANRO (%)
11001	500	560	12.0	529	565	6.8
19001	553	592	7.1	580	558	-3.8
21018	570	597	4.7	572	622	8.7
24004	521	481	-7.7	545	464	-14.9
25006	907	787	-13.2	834	778	-6.7
28008	694	684	-1.4	619	626	1.1
29003	303	239	-21.1	277	229	-17.3
32003	199	181	-9.0	184	132	-28.3
34004	277	298	7.6	252	262	4.0
37005	158	160	1.3	153	159	3.9
38021	179	252	40.8	164	184	12.2
39008	298	286	-4.0	260	274	5.4
39019	270	275	1.9	225	216	-4.0
40007	390	388	-0.5	382	332	-13.1
42003	338	418	23.7	286	379	32.5
43005	367	349	-4.9	323	322	-0.3
47001	825	828	0.4	733	745	1.6
48004	997	938	-5.9	980	877	-10.5
54008	457	412	-9.8	398	384	-3.5
54016	255	274	7.5	221	219	-0.9
57004	1512	1655	9.5	1362	1427	4.8
58009	1086	1063	-2.1	1046	1014	-3.1
66011	1865	1955	4.8	1753	1630	-7.0
76005	770	853	10.8	720	772	7.2
79006	1245	1289	3.5	1185	1268	7.0

Average annual runoff totals with percentage errors for model D2.

Appendix A

A.3 BFIs with percentage and relative errors

Catchment	Calibration period Obs	Sim	C%ERR	CBFI	Validation period Obs	Sim	V%ERR	VBFI
11001	0.692	0.794	14.7	0.102	0.694	0.815	17.4	0.121
19001	0.370	0.547	47.8	0.177	0.383	0.574	49.9	0.191
21018	0.567	0.690	21.7	0.123	0.595	0.717	20.5	0.122
24004	0.474	0.629	32.7	0.155	0.511	0.665	30.1	0.154
25006	0.206	0.383	85.9	0.177	0.235	0.415	76.6	0.180
28008	0.630	0.705	11.9	0.075	0.629	0.661	5.1	0.032
29003	0.914	0.829	-9.3	-0.085	0.913	0.864	-5.4	-0.049
32003	0.467	0.614	31.5	0.147	0.446	0.576	29.1	0.130
34004	0.706	0.831	17.7	0.125	0.695	0.821	18.1	0.126
37005	0.500	0.691	38.2	0.191	0.519	0.677	30.4	0.158
38021	0.198	0.303	53.0	0.105	0.233	0.338	45.1	0.105
39008	0.722	0.841	16.5	0.119	0.712	0.806	13.2	0.094
39019	0.971	0.990	2.0	0.019	0.972	0.990	1.9	0.018
40007	0.489	0.614	25.6	0.125	0.477	0.564	18.2	0.087
42003	0.383	0.535	39.7	0.152	0.386	0.484	25.4	0.098
43005	0.918	0.976	6.3	0.058	0.928	0.973	4.8	0.045
47001	0.482	0.573	18.9	0.091	0.485	0.574	18.4	0.089
48004	0.745	0.792	6.3	0.047	0.753	0.787	4.5	0.034
54008	0.551	0.809	46.8	0.258	0.588	0.808	37.4	0.220
54016	0.642	0.745	16.0	0.103	0.651	0.766	17.7	0.115
57004	0.395	0.482	22.0	0.087	0.397	0.549	38.3	0.152
58009	0.544	0.622	14.3	0.078	0.571	0.630	10.3	0.059
66011	0.263	0.292	11.0	0.029	0.296	0.366	23.6	0.070
76005	0.345	0.462	33.9	0.117	0.377	0.529	40.3	0.152
79006	0.322	0.421	30.7	0.099	0.355	0.414	16.6	0.059

BFIs with percentage and relative errors for model A.

Catchment	Calibration period Obs	Sim	C%ERR	CBFI	Validation period Obs	Sim	V%ERR	VBFI
11001	0.692	0.832	20.2	0.140	0.694	0.856	23.3	0.162
19001	0.370	0.455	23.0	0.085	0.383	0.497	29.8	0.114
21018	0.567	0.606	6.9	0.039	0.595	0.637	7.1	0.042
24004	0.474	0.294	-38.0	-0.180	0.511	0.371	-27.4	-0.140
25006	0.206	0.216	4.9	0.010	0.235	0.254	8.1	0.019
28008	0.630	0.642	1.9	0.012	0.629	0.626	-0.5	-0.003
29003	0.914	0.893	-2.3	-0.021	0.913	0.934	2.3	0.021
32003	0.467	0.517	10.7	0.050	0.446	0.530	18.8	0.084
34004	0.706	0.816	15.6	0.110	0.695	0.845	21.6	0.150
37005	0.500	0.743	48.6	0.243	0.519	0.715	37.8	0.196
38021	0.198	0.285	43.9	0.087	0.233	0.287	23.2	0.054
39008	0.722	0.693	-4.0	-0.029	0.712	0.691	-2.9	-0.021
39019	0.971	0.980	0.9	0.009	0.972	0.990	1.9	0.018
40007	0.489	0.449	-8.2	-0.040	0.477	0.381	-20.1	-0.096
42003	0.383	0.490	27.9	0.107	0.386	0.381	-1.3	-0.005
43005	0.918	0.936	2.0	0.018	0.928	0.938	1.1	0.010
47001	0.482	0.391	-18.9	-0.091	0.485	0.395	-18.6	-0.090
48004	0.745	0.717	-3.8	-0.028	0.753	0.734	-2.5	-0.019
54008	0.551	0.605	9.8	0.054	0.588	0.614	4.4	0.026
54016	0.642	0.723	12.6	0.081	0.651	0.794	22.0	0.143
57004	0.395	0.339	-14.2	-0.056	0.397	0.367	-7.6	-0.030
58009	0.544	0.720	32.4	0.176	0.571	0.731	28.0	0.160
66011	0.263	0.286	8.7	0.023	0.296	0.350	18.2	0.054
76005	0.345	0.403	16.8	0.058	0.377	0.521	38.2	0.144
79006	0.322	0.327	1.6	0.005	0.355	0.343	-3.4	-0.012

BFIs with percentage and relative errors for model B1.

Catchment	Calibration period Obs	Sim	C%ERR	CBFI	Validation period Obs	Sim	V%ERR	VBFI
11001	0.692	0.808	16.8	0.116	0.694	0.821	18.3	0.127
19001	0.370	0.384	3.8	0.014	0.383	0.409	6.8	0.026
21018	0.567	0.463	-18.3	-0.104	0.595	0.505	-15.1	-0.090
24004	0.474	0.403	-15.0	-0.071	0.511	0.384	-24.9	-0.127
25006	0.206	0.147	-28.6	-0.059	0.235	0.150	-36.2	-0.085
28008	0.630	0.703	11.6	0.073	0.629	0.662	5.2	0.033
29003	0.914	0.902	-1.3	-0.012	0.913	0.951	4.2	0.038
32003	0.467	0.379	-18.8	-0.088	0.446	0.412	-7.6	-0.034
34004	0.706	0.825	16.9	0.119	0.695	0.824	18.6	0.129
37005	0.500	0.495	-1.0	-0.005	0.519	0.530	2.1	0.011
38021	0.198	0.315	59.1	0.117	0.233	0.273	17.2	0.040
39008	0.722	0.756	4.7	0.034	0.712	0.733	2.9	0.021
39019	0.971	0.990	2.0	0.019	0.972	0.990	1.9	0.018
40007	0.489	0.391	-20.0	-0.098	0.477	0.456	-4.4	-0.021
42003	0.383	0.401	4.7	0.018	0.386	0.357	-7.5	-0.029
43005	0.918	0.931	1.4	0.013	0.928	0.935	0.8	0.007
47001	0.482	0.377	-21.8	-0.105	0.485	0.378	-22.1	-0.107
48004	0.745	0.799	7.2	0.054	0.753	0.790	4.9	0.037
54008	0.551	0.635	15.2	0.084	0.588	0.657	11.7	0.069
54016	0.642	0.750	16.8	0.108	0.651	0.814	25.0	0.163
57004	0.395	0.478	21.0	0.083	0.397	0.504	27.0	0.107
58009	0.544	0.652	19.9	0.108	0.571	0.648	13.5	0.077
66011	0.263	0.338	28.5	0.075	0.296	0.400	35.1	0.104
76005	0.345	0.306	-11.3	-0.039	0.377	0.290	-23.1	-0.087
79006	0.322	0.393	22.0	0.071	0.355	0.386	8.7	0.031

BFIs with percentage and relative errors for model B2.

Catchment	Calibration period Obs	Sim	C%ERR	CBFI	Validation period Obs	Sim	V%ERR	VBFI
11001	0.692	0.752	8.7	0.060	0.694	0.760	9.5	0.066
19001	0.370	0.435	17.6	0.065	0.383	0.440	14.9	0.057
21018	0.567	0.649	14.5	0.082	0.595	0.657	10.4	0.062
24004	0.474	0.530	11.8	0.056	0.511	0.529	3.5	0.018
25006	0.206	0.319	54.9	0.113	0.235	0.321	36.6	0.086
28008	0.630	0.604	-4.1	-0.026	0.629	0.598	-4.9	-0.031
29003	0.914	0.912	-0.2	-0.002	0.913	0.921	0.9	0.008
32003	0.467	0.350	-25.1	-0.117	0.446	0.398	-10.8	-0.048
34004	0.706	0.705	-0.1	-0.001	0.695	0.728	4.7	0.033
37005	0.500	0.408	-18.4	-0.092	0.519	0.429	-17.3	-0.090
38021	0.198	0.100	-49.5	-0.098	0.233	0.136	-41.6	-0.097
39008	0.722	0.737	2.1	0.015	0.712	0.712	0.0	0.000
39019	0.971	1.000	3.0	0.029	0.972	1.000	2.9	0.028
40007	0.489	0.382	-21.9	-0.107	0.477	0.380	-20.3	-0.097
42003	0.383	0.394	2.9	0.011	0.386	0.384	-0.5	-0.002
43005	0.918	0.966	5.2	0.048	0.928	0.949	2.3	0.021
47001	0.482	0.435	-9.8	-0.047	0.485	0.480	-1.0	-0.005
48004	0.745	0.741	-0.5	-0.004	0.753	0.734	-2.5	-0.019
54008	0.551	0.502	-8.9	-0.049	0.588	0.547	-7.0	-0.041
54016	0.642	0.611	-4.8	-0.031	0.651	0.640	-1.7	-0.011
57004	0.395	0.508	28.6	0.113	0.397	0.535	34.8	0.138
58009	0.544	0.581	6.8	0.037	0.571	0.593	3.9	0.022
66011	0.263	0.382	45.2	0.119	0.296	0.438	48.0	0.142
76005	0.345	0.499	44.6	0.154	0.377	0.493	30.8	0.116
79006	0.322	0.482	49.7	0.160	0.355	0.481	35.5	0.126

BFIs with percentage and relative errors for model C.

Catchment	Calibration period Obs	Sim	C%ERR	CBFI	Validation period Obs	Sim	V%ERR	VBFI
11001	0.692	0.834	20.5	0.142	0.694	0.844	21.6	0.150
19001	0.370	0.487	31.6	0.117	0.383	0.532	38.9	0.149
21018	0.567	0.635	12.0	0.068	0.595	0.658	10.6	0.063
24004	0.474	0.536	13.1	0.062	0.511	0.539	5.5	0.028
25006	0.206	0.165	-19.9	-0.041	0.235	0.187	-20.4	-0.048
28008	0.630	0.709	12.5	0.079	0.629	0.710	12.9	0.081
29003	0.914	0.980	7.2	0.066	0.913	0.990	8.4	0.077
32003	0.467	0.651	39.4	0.184	0.446	0.702	57.4	0.256
34004	0.706	0.822	16.4	0.116	0.695	0.851	22.4	0.156
37005	0.500	0.520	4.0	0.020	0.519	0.495	-4.6	-0.024
38021	0.198	0.209	5.6	0.011	0.233	0.122	-47.6	-0.111
39008	0.722	0.798	10.5	0.076	0.712	0.717	0.7	0.005
39019	0.971	0.990	2.0	0.019	0.972	0.990	1.9	0.018
40007	0.489	0.539	10.2	0.050	0.477	0.544	14.0	0.067
42003	0.383	0.292	-23.8	-0.091	0.386	0.300	-22.3	-0.086
43005	0.918	0.977	6.4	0.059	0.928	0.980	5.6	0.052
47001	0.482	0.479	-0.6	-0.003	0.485	0.480	-1.0	-0.005
48004	0.745	0.808	8.5	0.063	0.753	0.785	4.2	0.032
54008	0.551	0.663	20.3	0.112	0.588	0.688	17.0	0.100
54016	0.642	0.567	-11.7	-0.075	0.651	0.682	4.8	0.031
57004	0.395	0.507	28.4	0.112	0.397	0.547	37.8	0.150
58009	0.544	0.577	6.1	0.033	0.571	0.588	3.0	0.017
66011	0.263	0.332	26.2	0.069	0.296	0.417	40.9	0.121
76005	0.345	0.283	-18.0	-0.062	0.377	0.317	-15.9	-0.060
79006	0.322	0.402	24.8	0.080	0.355	0.395	11.3	0.040

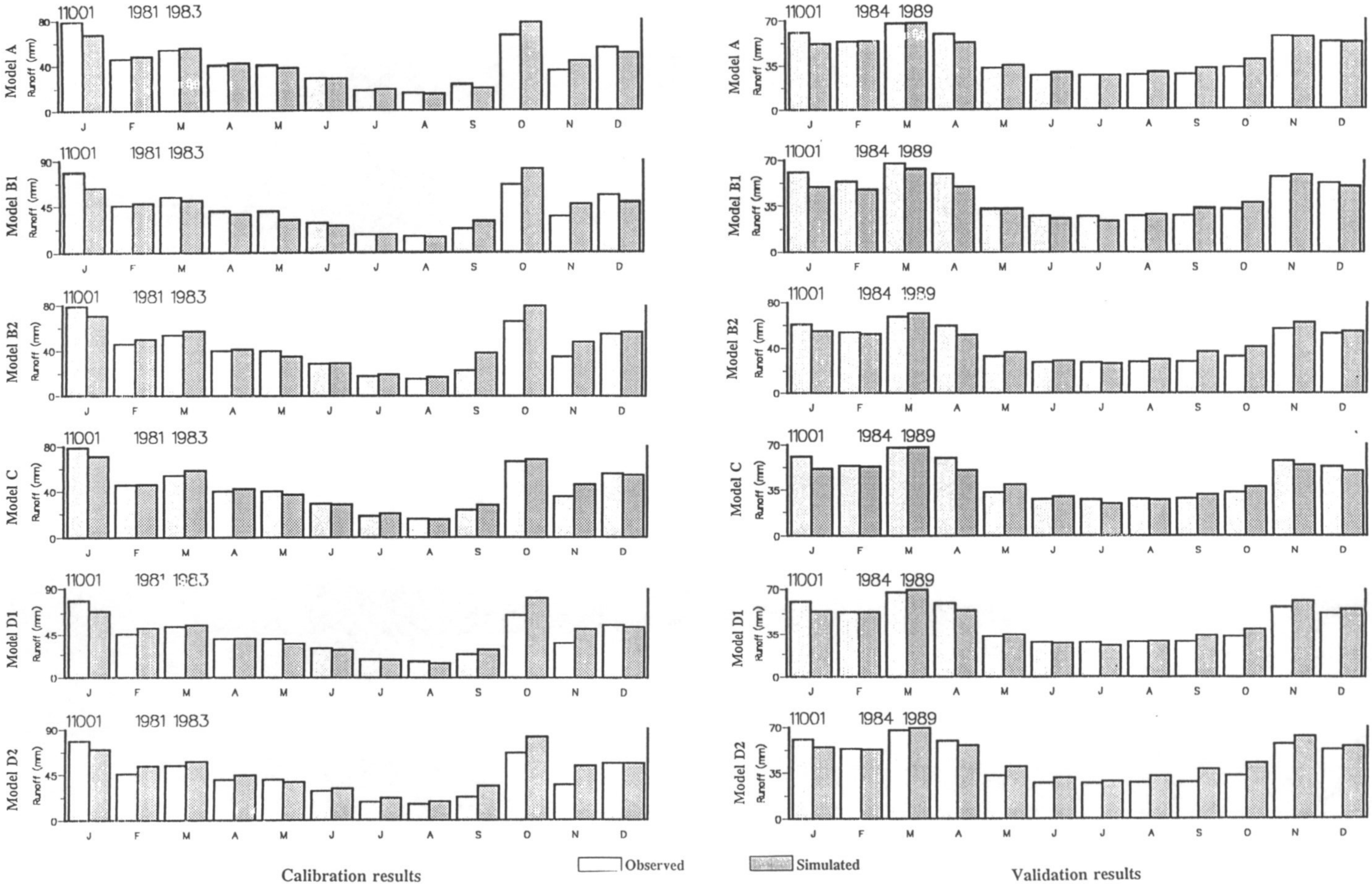
BFIs with percentage and relative errors for model D1.

Catchment	Calibration period Obs	Sim	C%ERR	CBFI	Validation period Obs	Sim	V%ERR	VBFI
11001	0.692	0.877	26.7	0.185	0.694	0.886	27.7	0.192
19001	0.370	0.351	-5.1	-0.019	0.383	0.387	1.0	0.004
21018	0.567	0.448	-21.0	-0.119	0.595	0.531	-10.8	-0.064
24004	0.474	0.360	-24.1	-0.114	0.511	0.359	-29.7	-0.152
25006	0.206	0.154	-25.2	-0.052	0.235	0.156	-33.6	-0.079
28008	0.630	0.723	14.8	0.093	0.629	0.727	15.6	0.098
29003	0.914	0.900	-1.5	-0.014	0.913	0.938	2.7	0.025
32003	0.467	0.409	-12.4	-0.058	0.446	0.493	10.5	0.047
34004	0.706	0.832	17.8	0.126	0.695	0.871	25.3	0.176
37005	0.500	0.540	8.0	0.040	0.519	0.541	4.2	0.022
38021	0.198	0.184	-7.1	-0.014	0.233	0.170	-27.0	-0.063
39008	0.722	0.884	22.4	0.162	0.712	0.797	11.9	0.085
39019	0.971	0.990	2.0	0.019	0.972	0.990	1.9	0.018
40007	0.489	0.373	-23.7	-0.116	0.477	0.385	-19.3	-0.092
42003	0.383	0.478	24.8	0.095	0.386	0.421	9.1	0.035
43005	0.918	0.973	6.0	0.055	0.928	0.976	5.2	0.048
47001	0.482	0.364	-24.5	-0.118	0.485	0.334	-31.1	-0.151
48004	0.745	0.733	-1.6	-0.012	0.753	0.775	2.9	0.022
54008	0.551	0.647	17.4	0.096	0.588	0.672	14.3	0.084
54016	0.642	0.726	13.1	0.084	0.651	0.775	19.0	0.124
57004	0.395	0.455	15.2	0.060	0.397	0.496	24.9	0.099
58009	0.544	0.606	11.4	0.062	0.571	0.620	8.6	0.049
66011	0.263	0.162	-38.4	-0.101	0.296	0.161	-45.6	-0.135
76005	0.345	0.239	-30.8	-0.106	0.377	0.324	-14.1	-0.053
79006	0.322	0.237	-26.4	-0.085	0.355	0.211	-40.6	-0.144

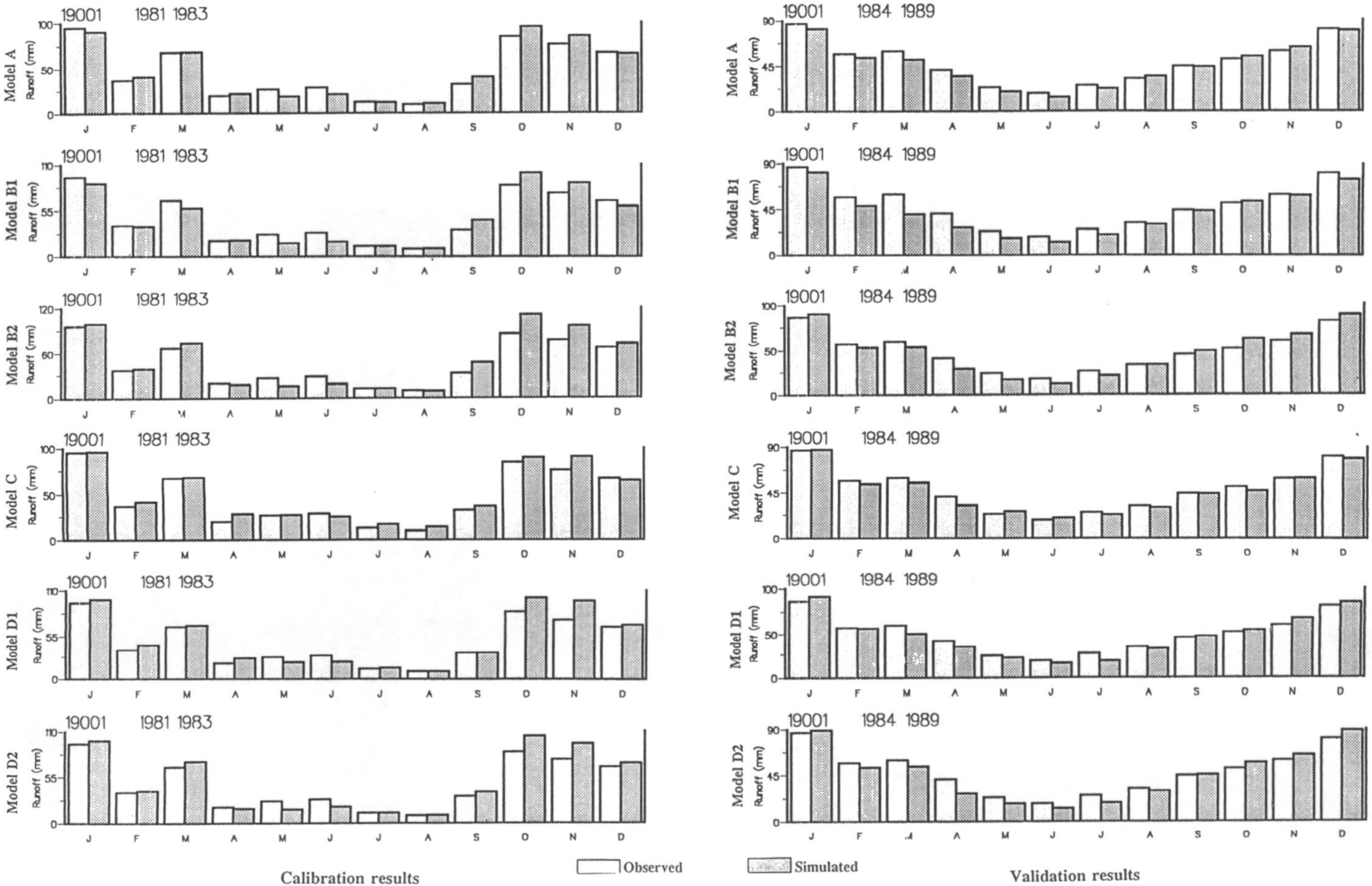
BFIs with percentage and relative errors for model D2.

Appendix A

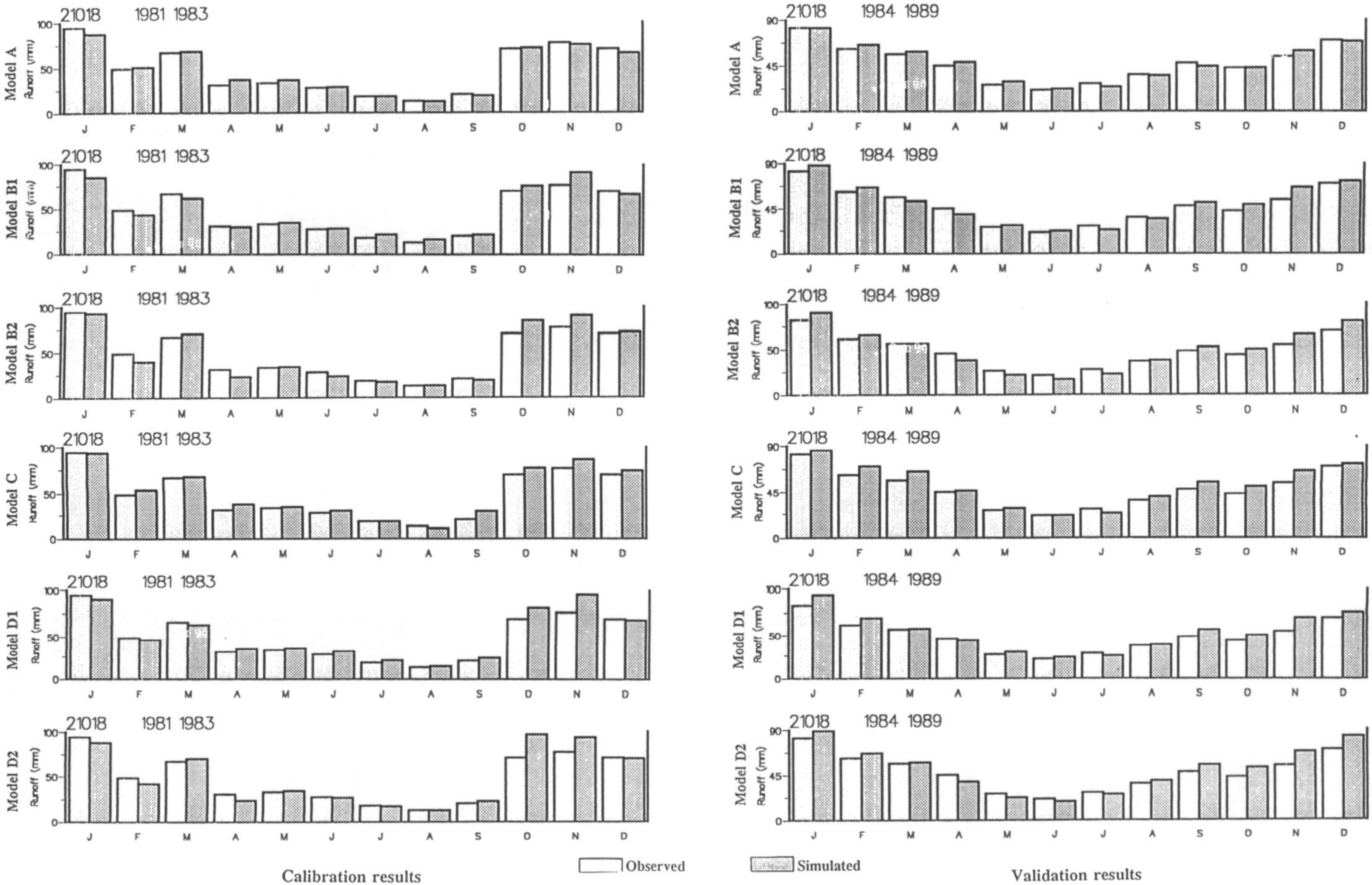
A.4 Histograms of observed and simulated monthly flow regimes



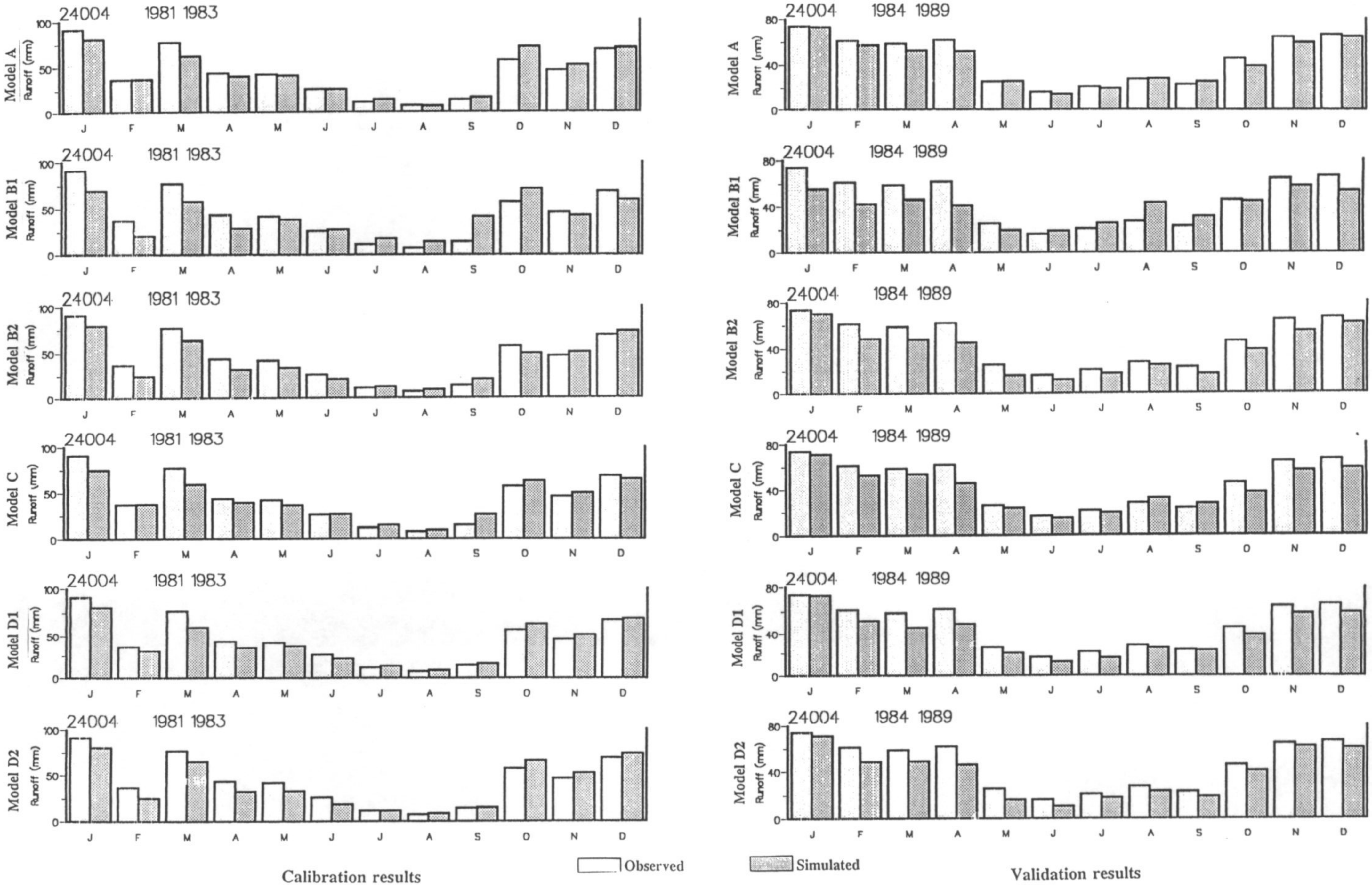
Observed and simulated monthly flow regimes for catchment 11001.



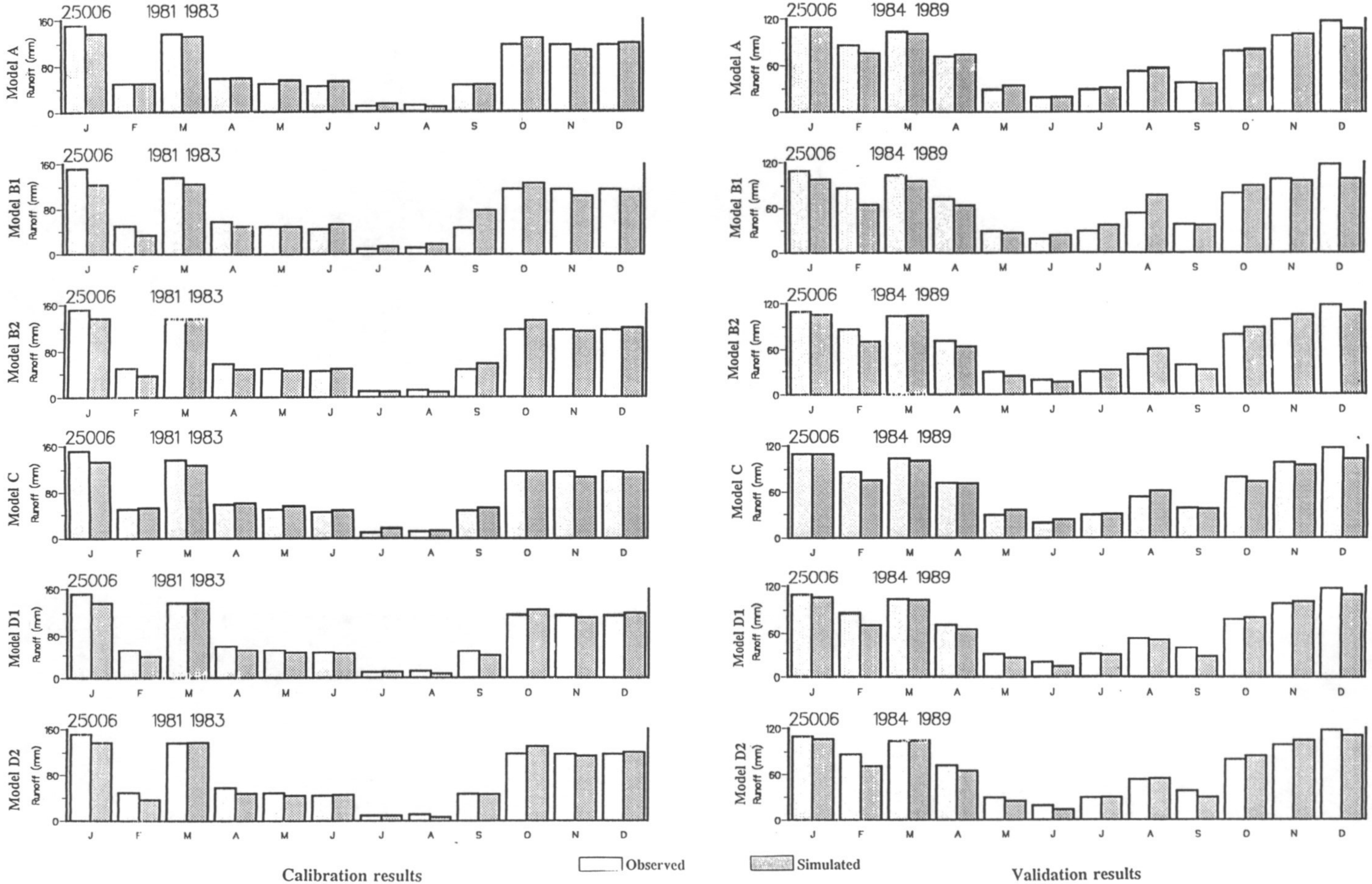
Observed and simulated monthly flow regimes for catchment 1900l.



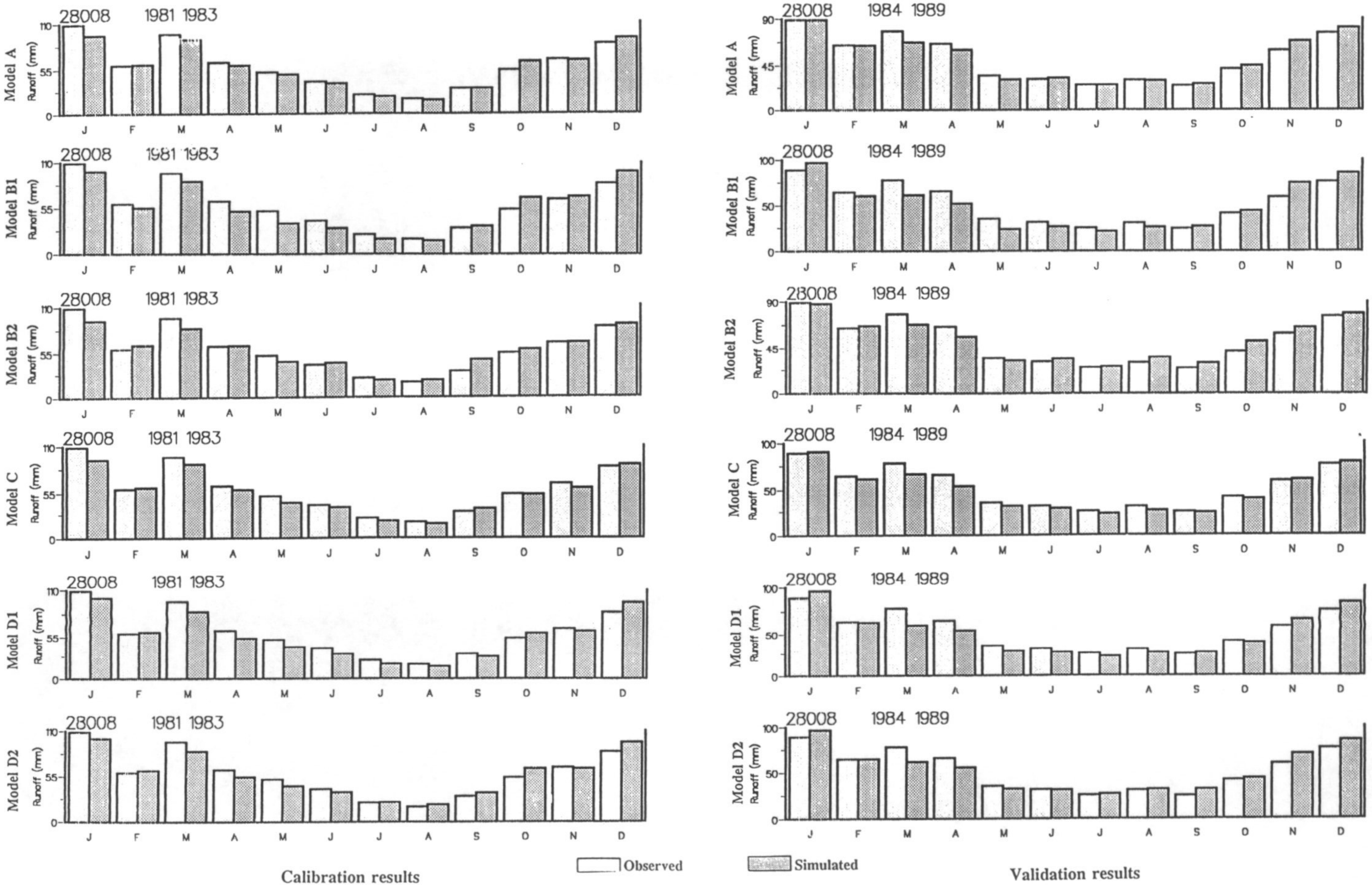
Observed and simulated monthly flow regimes for catchment 21018.



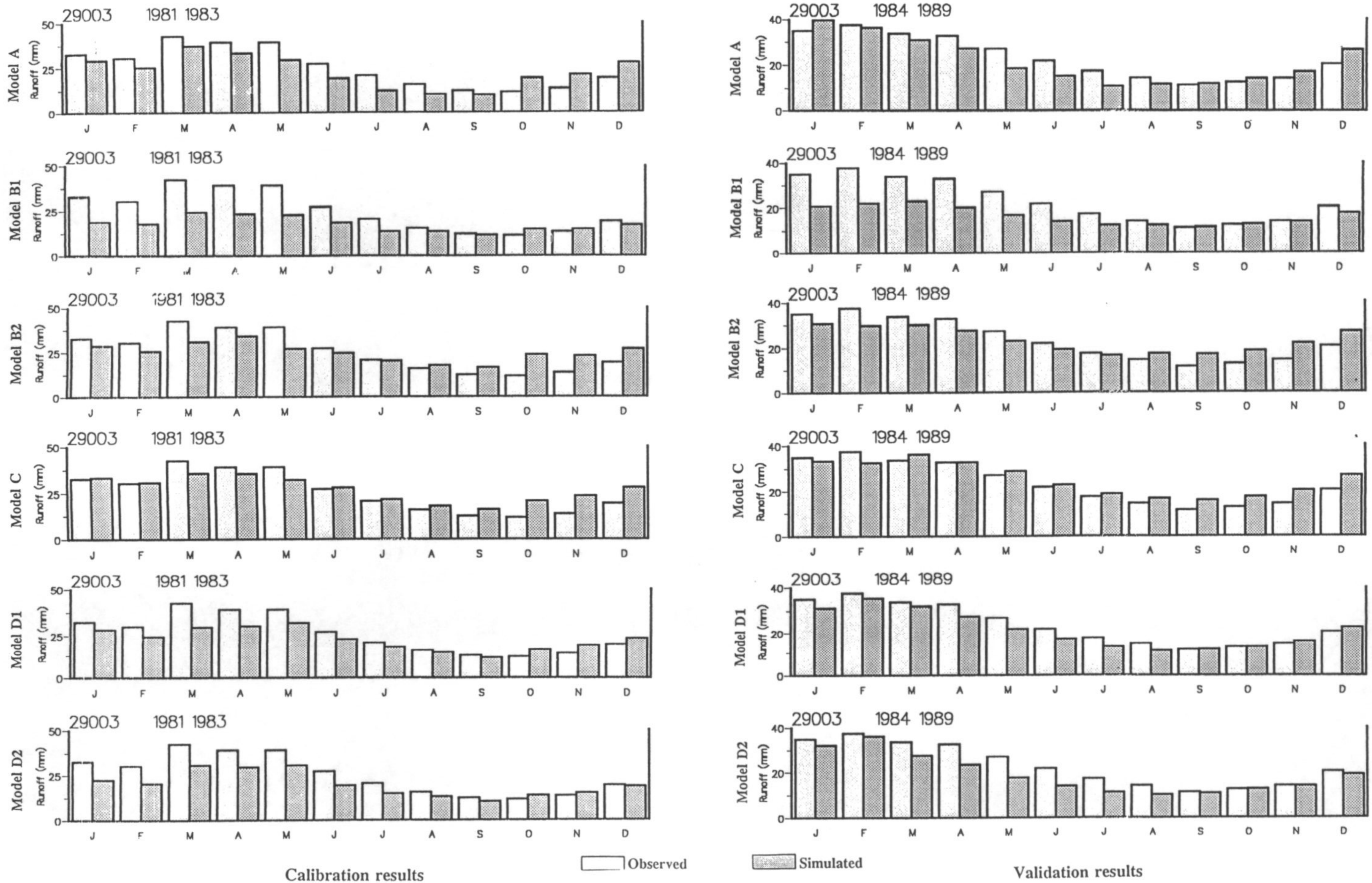
Observed and simulated monthly flow regimes for catchment 24004.



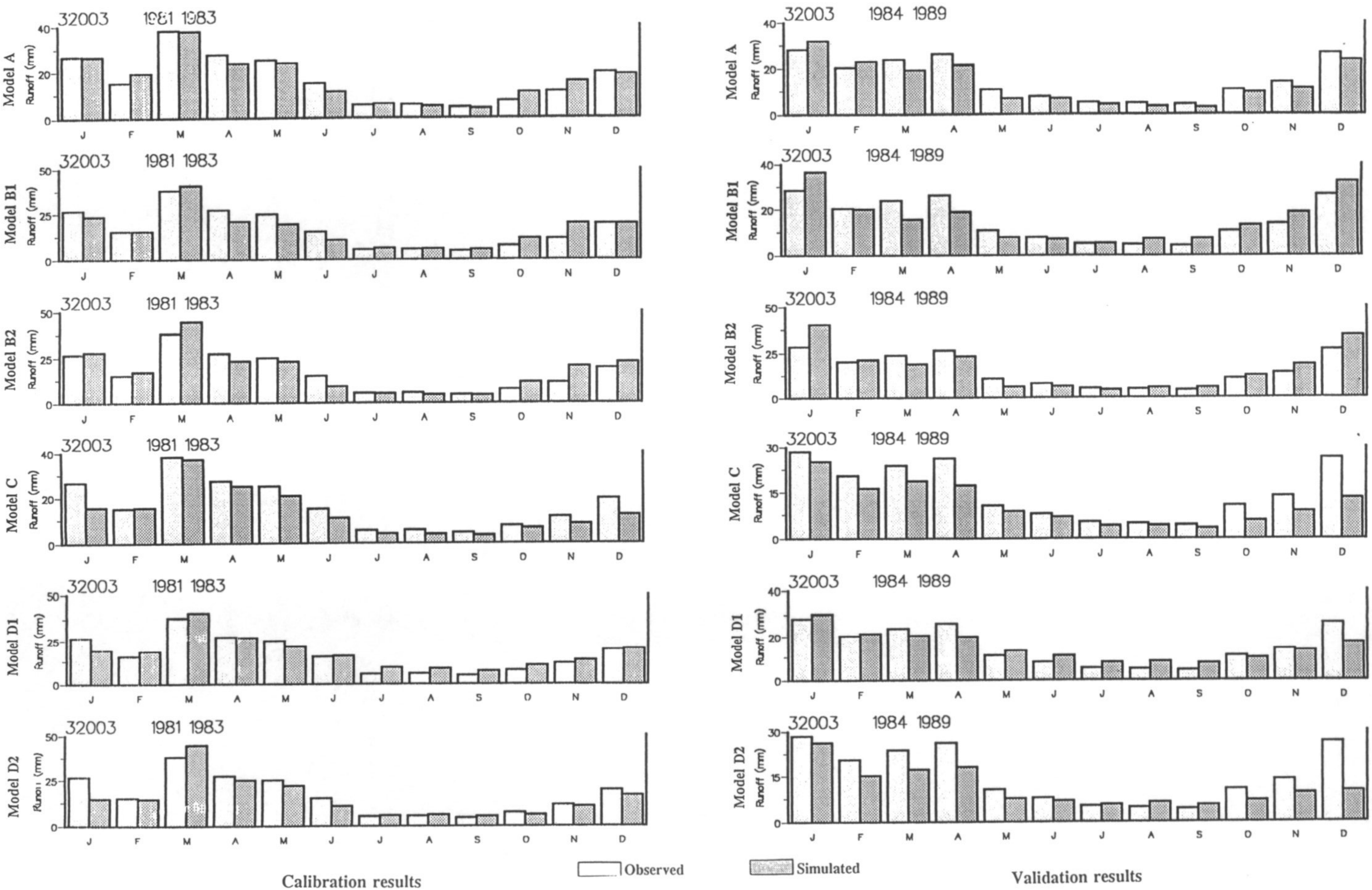
Observed and simulated monthly flow regimes for catchment 25006.



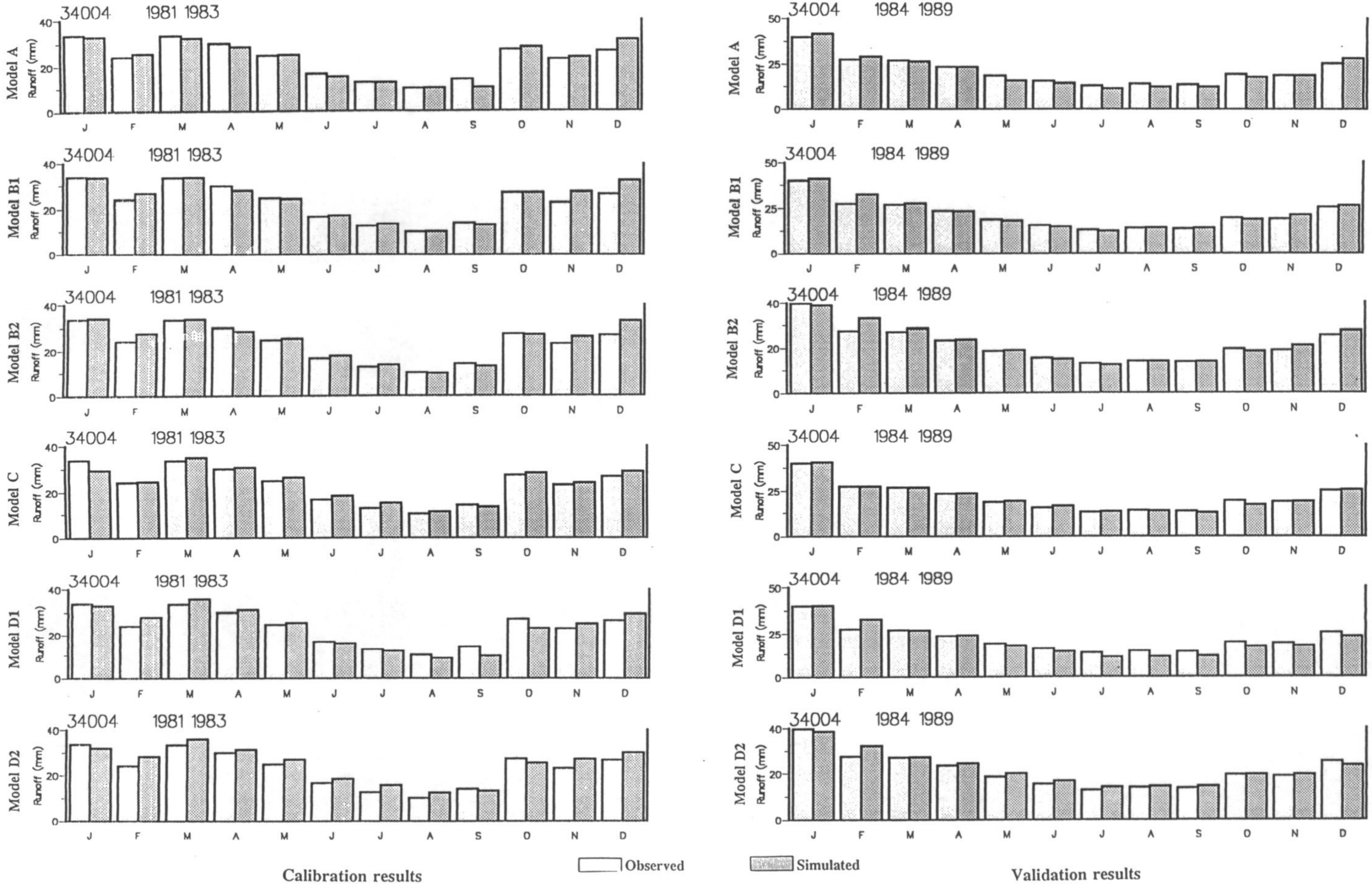
Observed and simulated monthly flow regimes for catchment 28008.



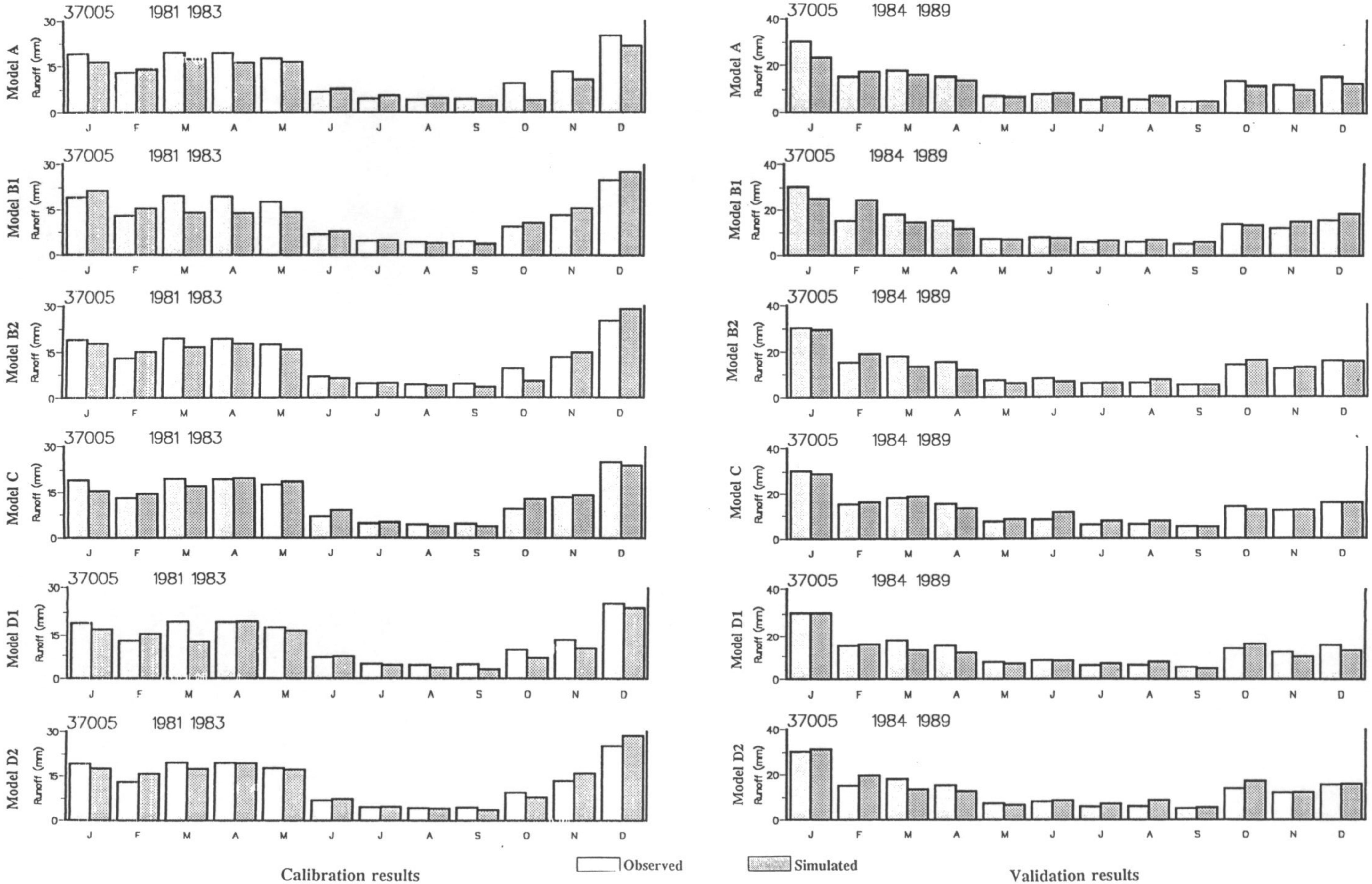
Observed and simulated monthly flow regimes for catchment 29003.



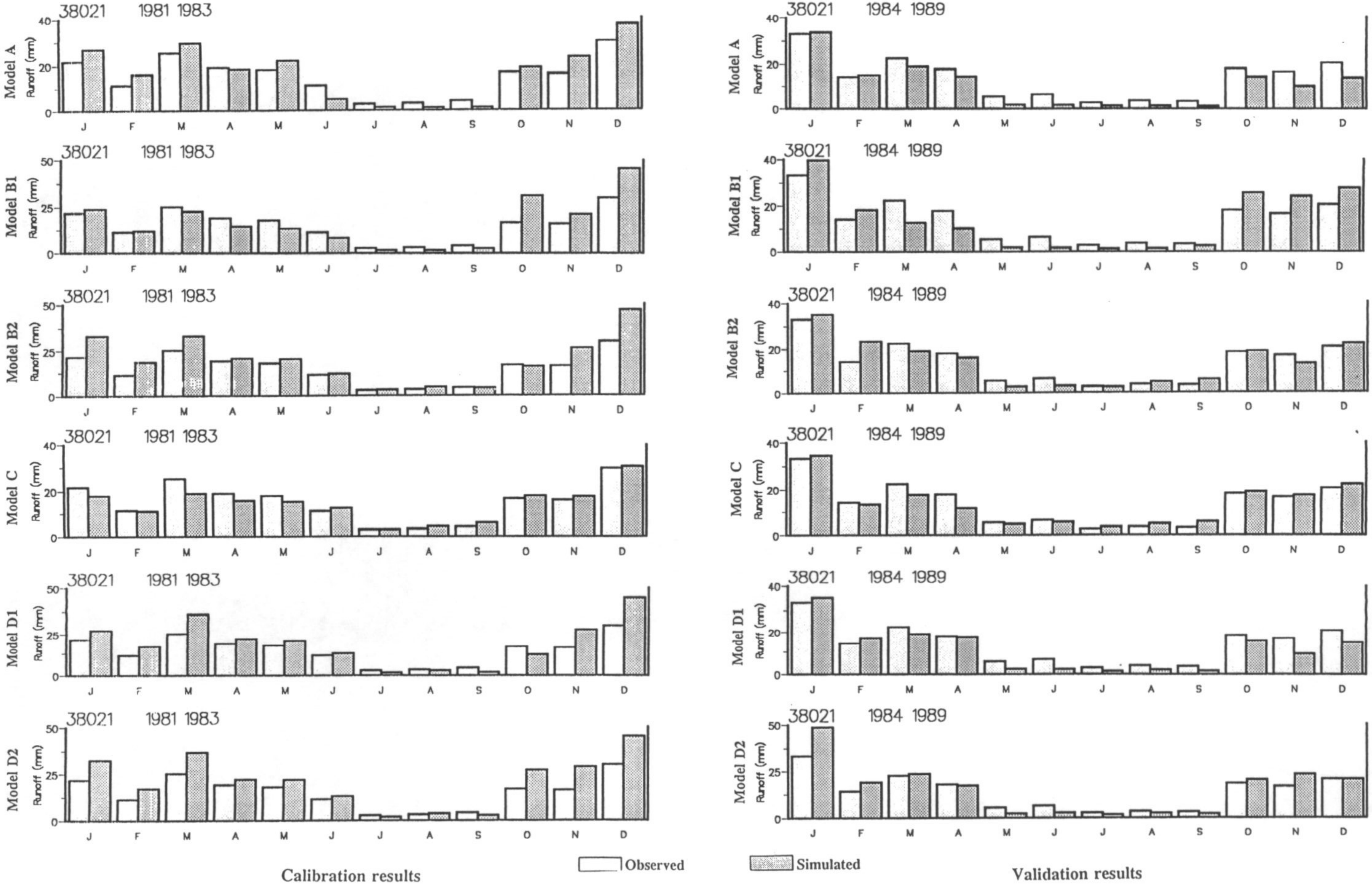
Observed and simulated monthly flow regimes for catchment 32003.



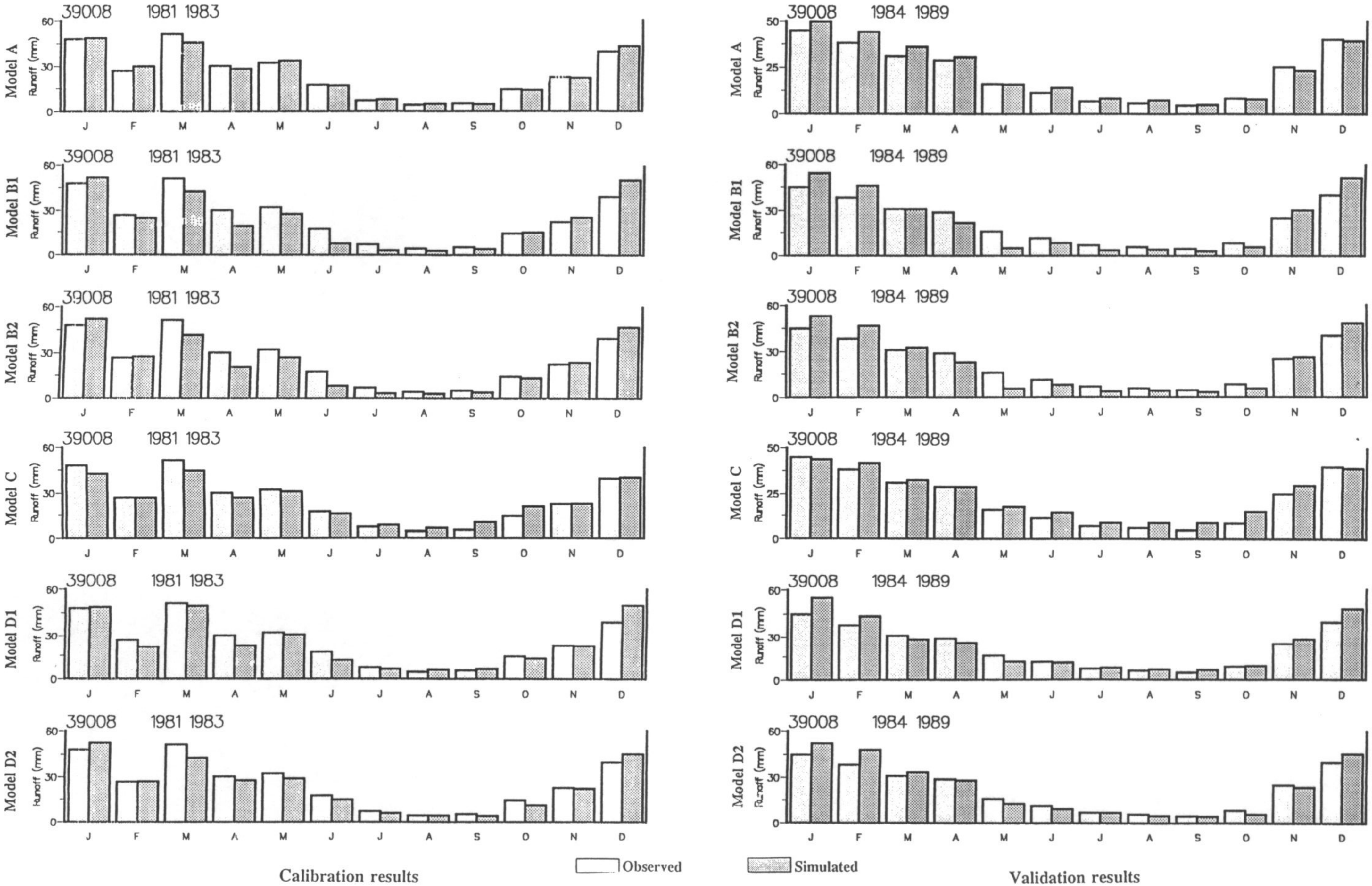
Observed and simulated monthly flow regimes for catchment 34004.



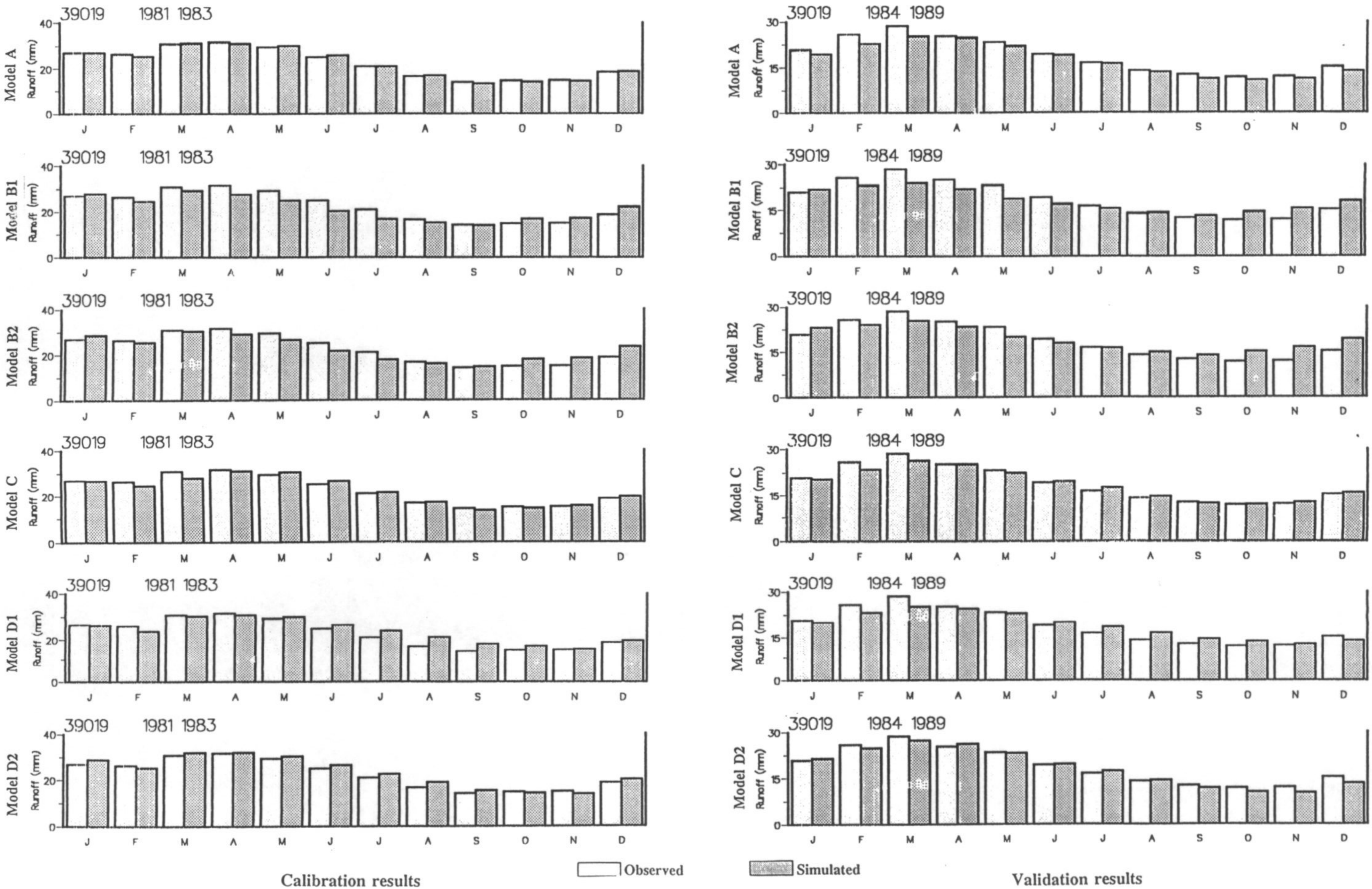
Observed and simulated monthly flow regimes for catchment 37005.



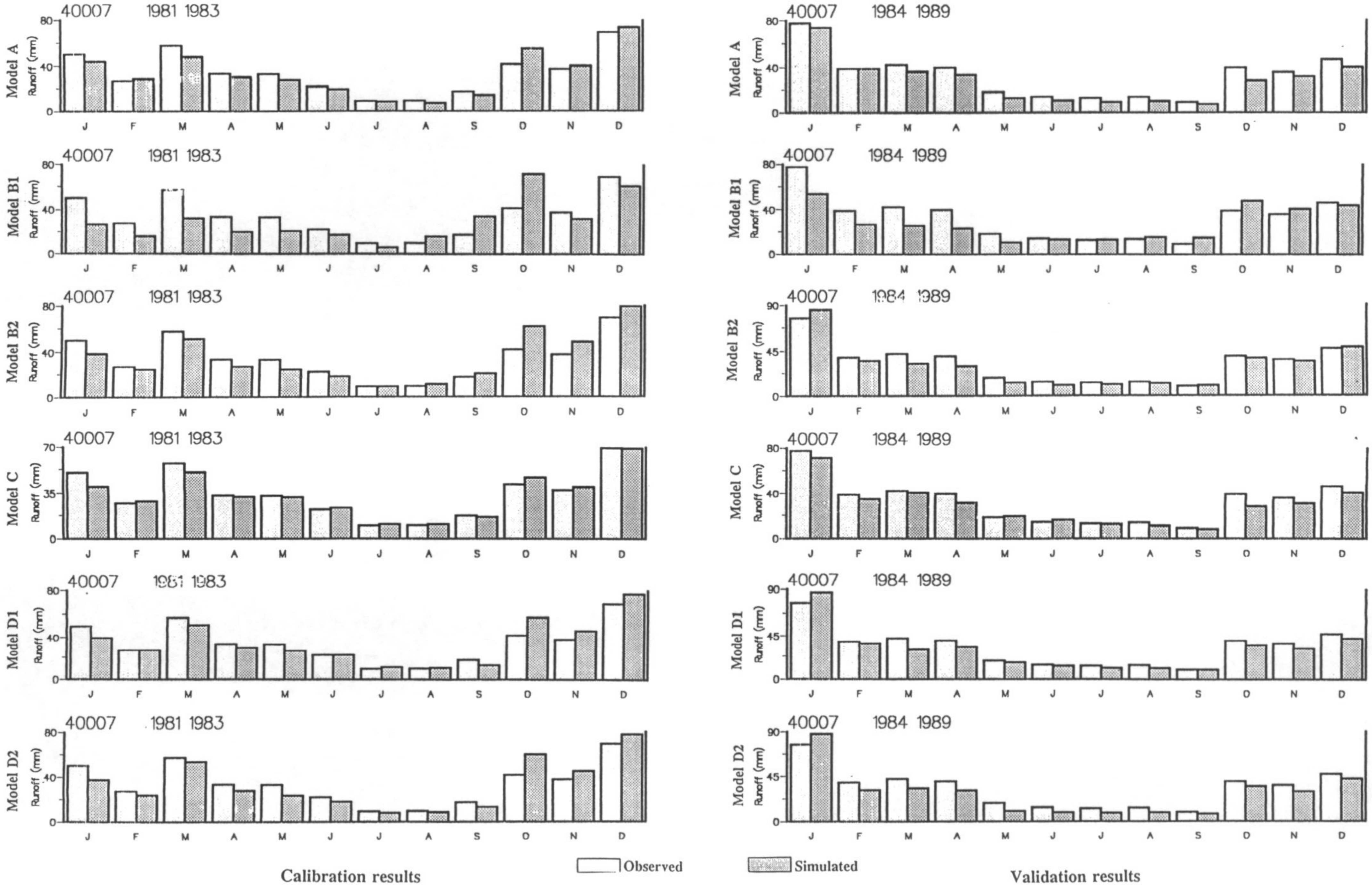
Observed and simulated monthly flow regimes for catchment 38021.



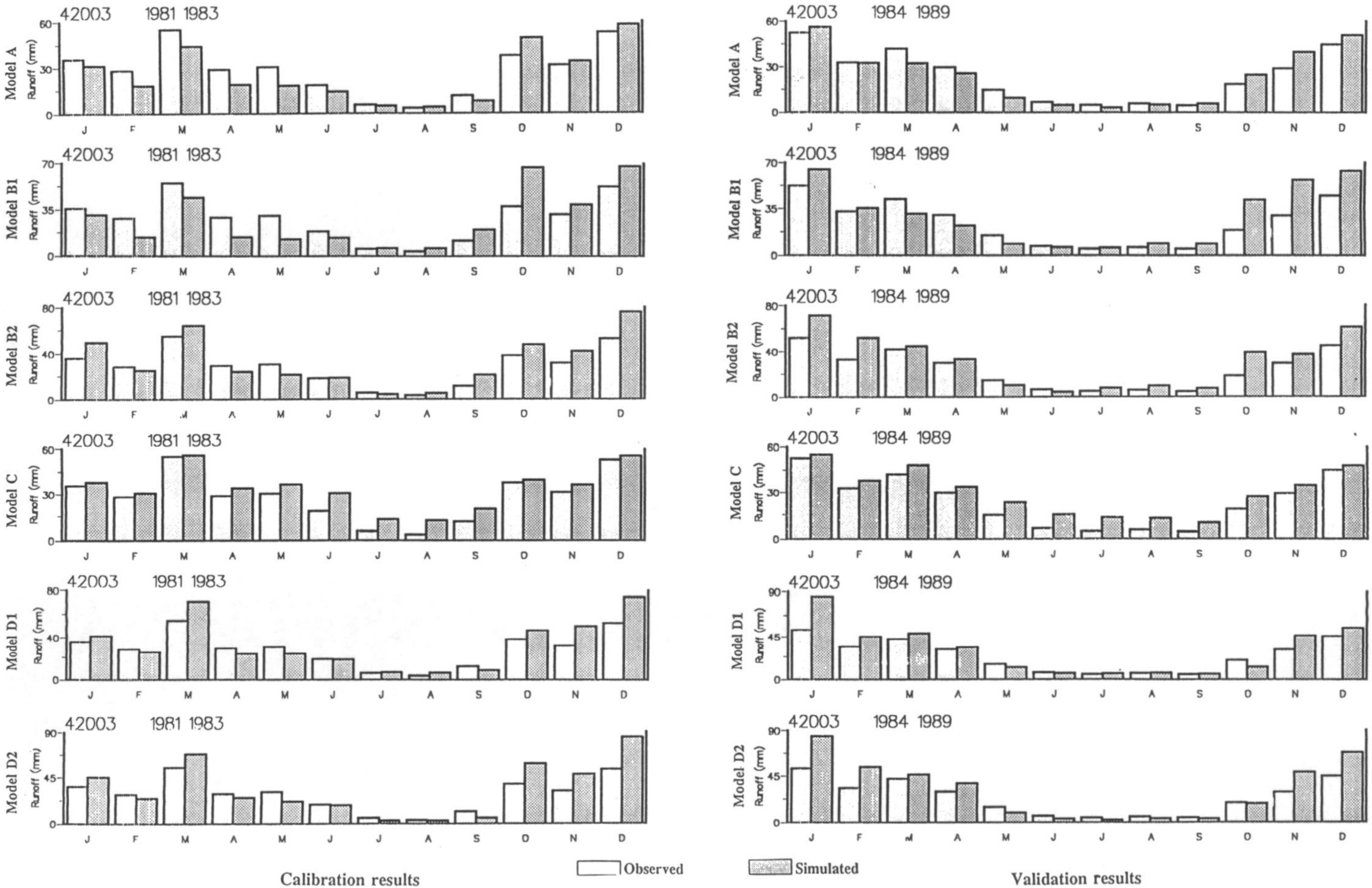
Observed and simulated monthly flow regimes for catchment 39008.



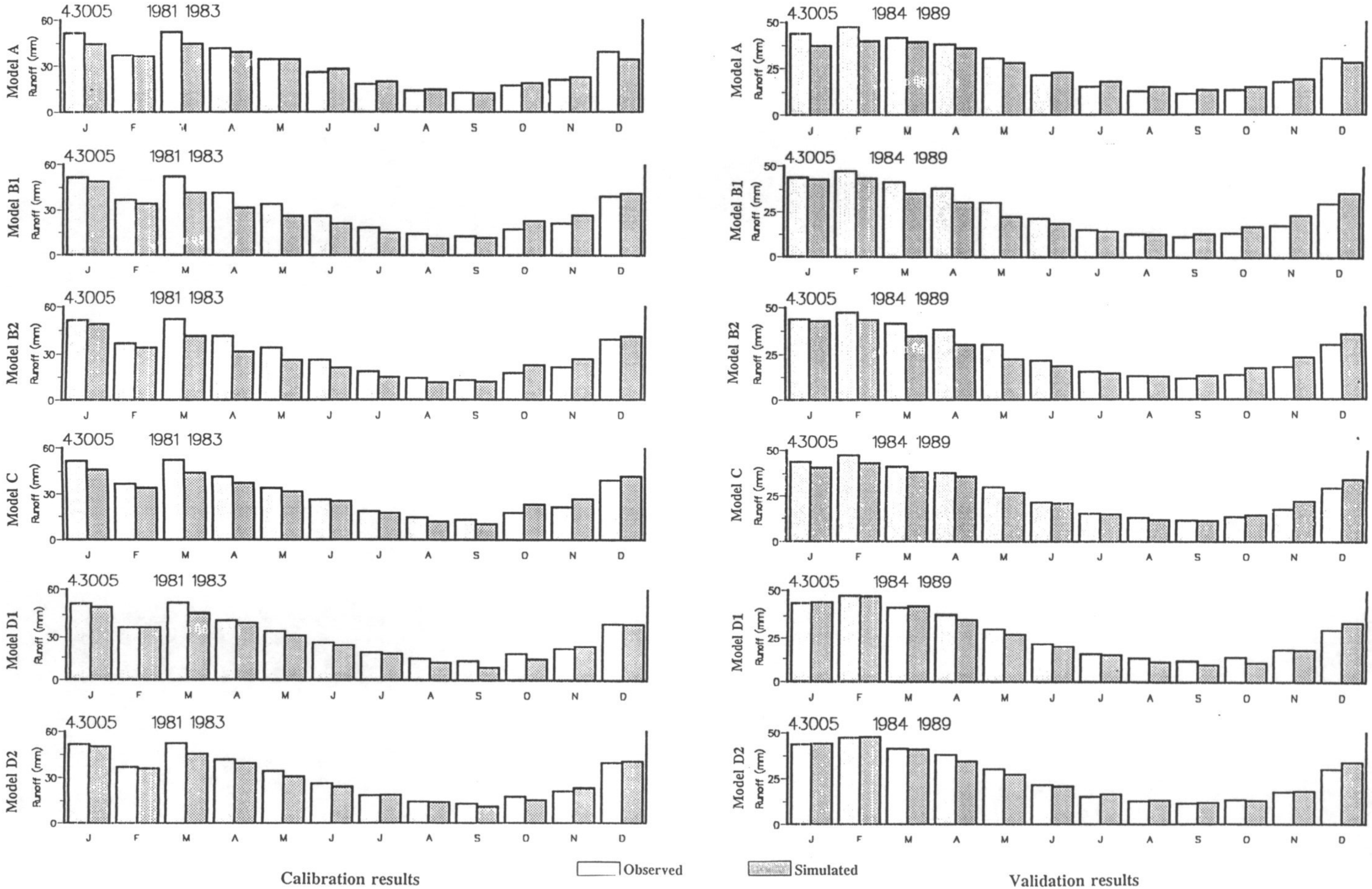
Observed and simulated monthly flow regimes for catchment 39019.



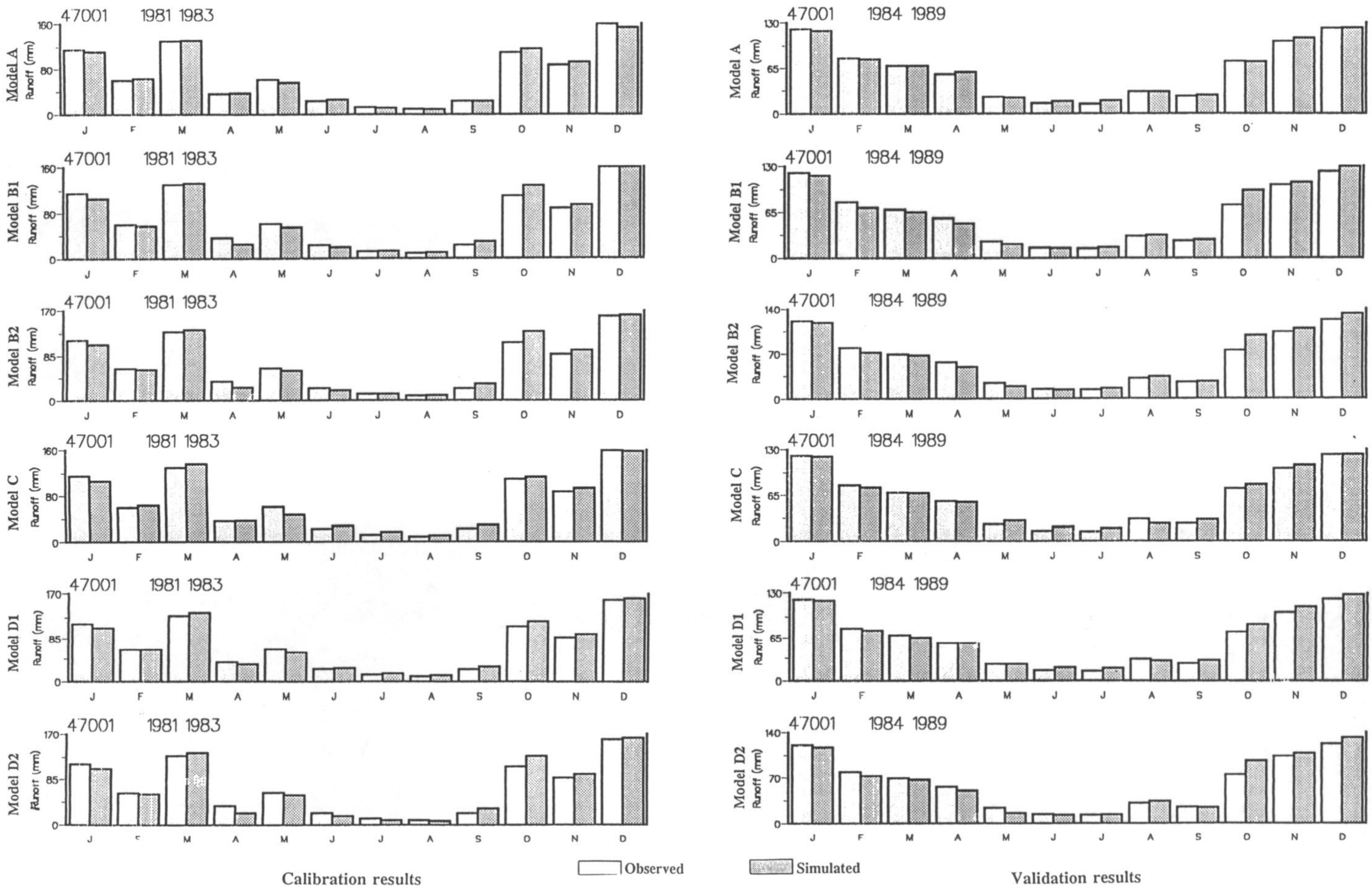
Observed and simulated monthly flow regimes for catchment 40007.



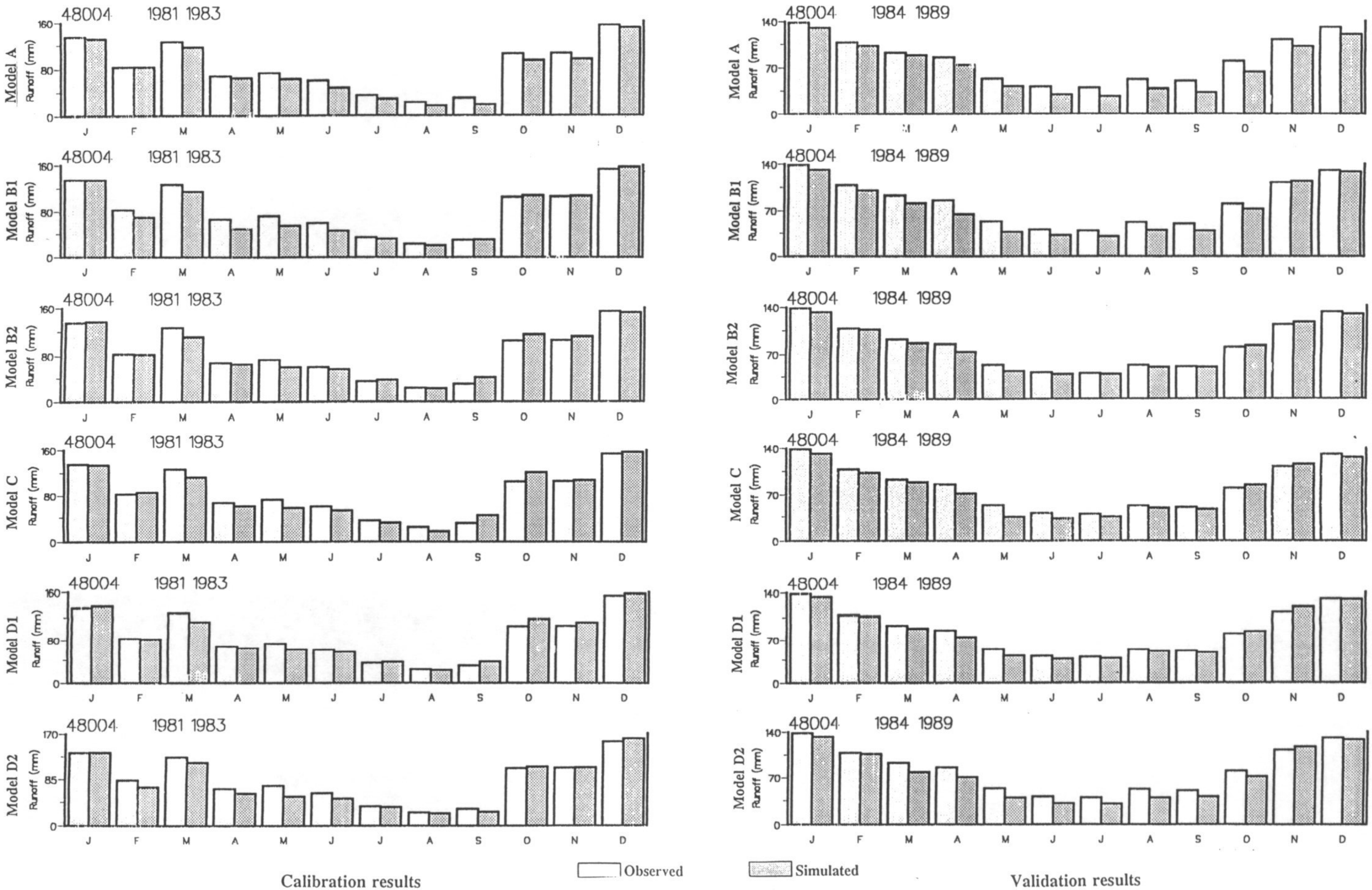
Observed and simulated monthly flow regimes for catchment 42003.



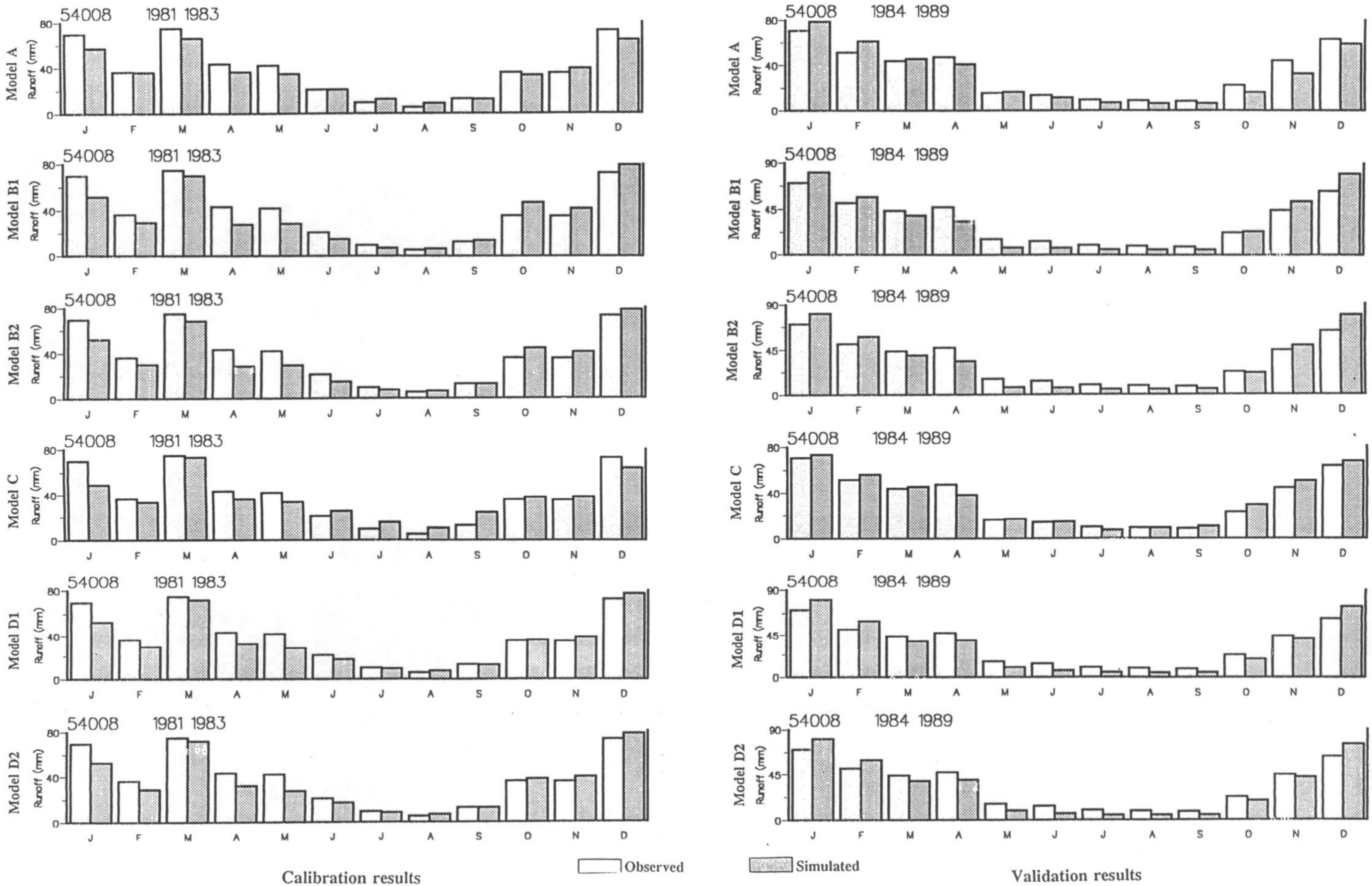
Observed and simulated monthly flow regimes for catchment 43005.



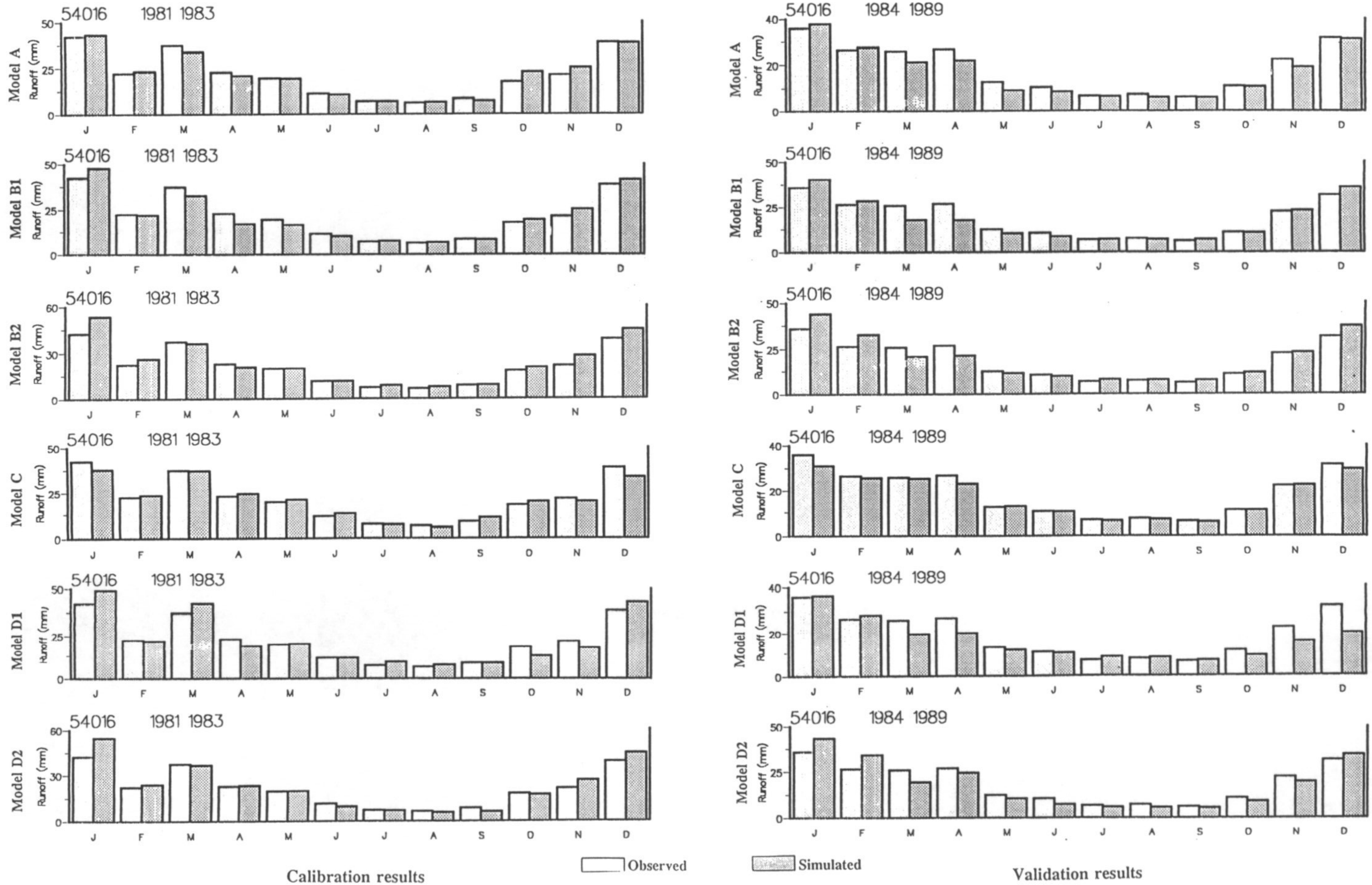
Observed and simulated monthly flow regimes for catchment 47001.



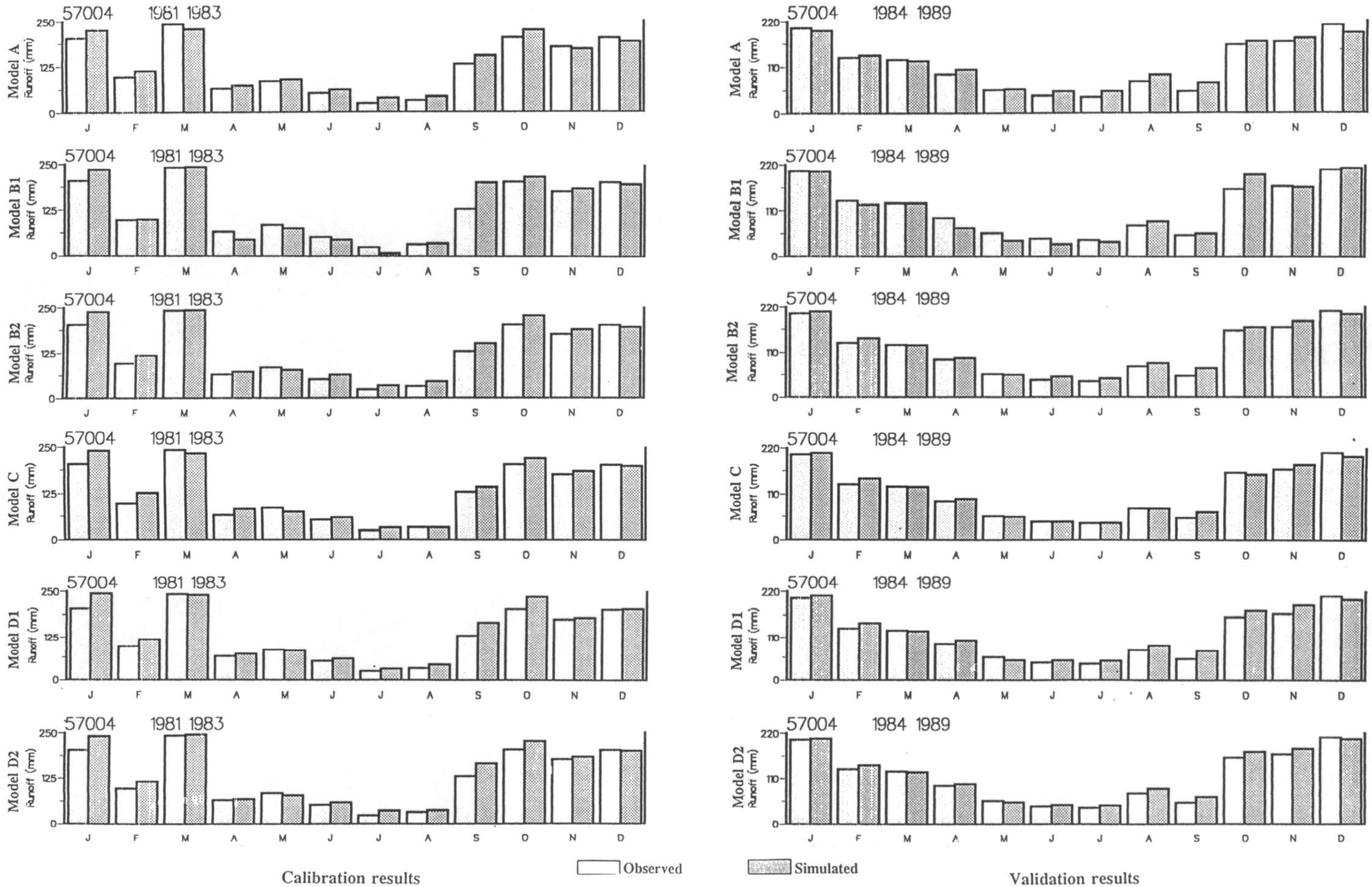
Observed and simulated monthly flow regimes for catchment 48004.



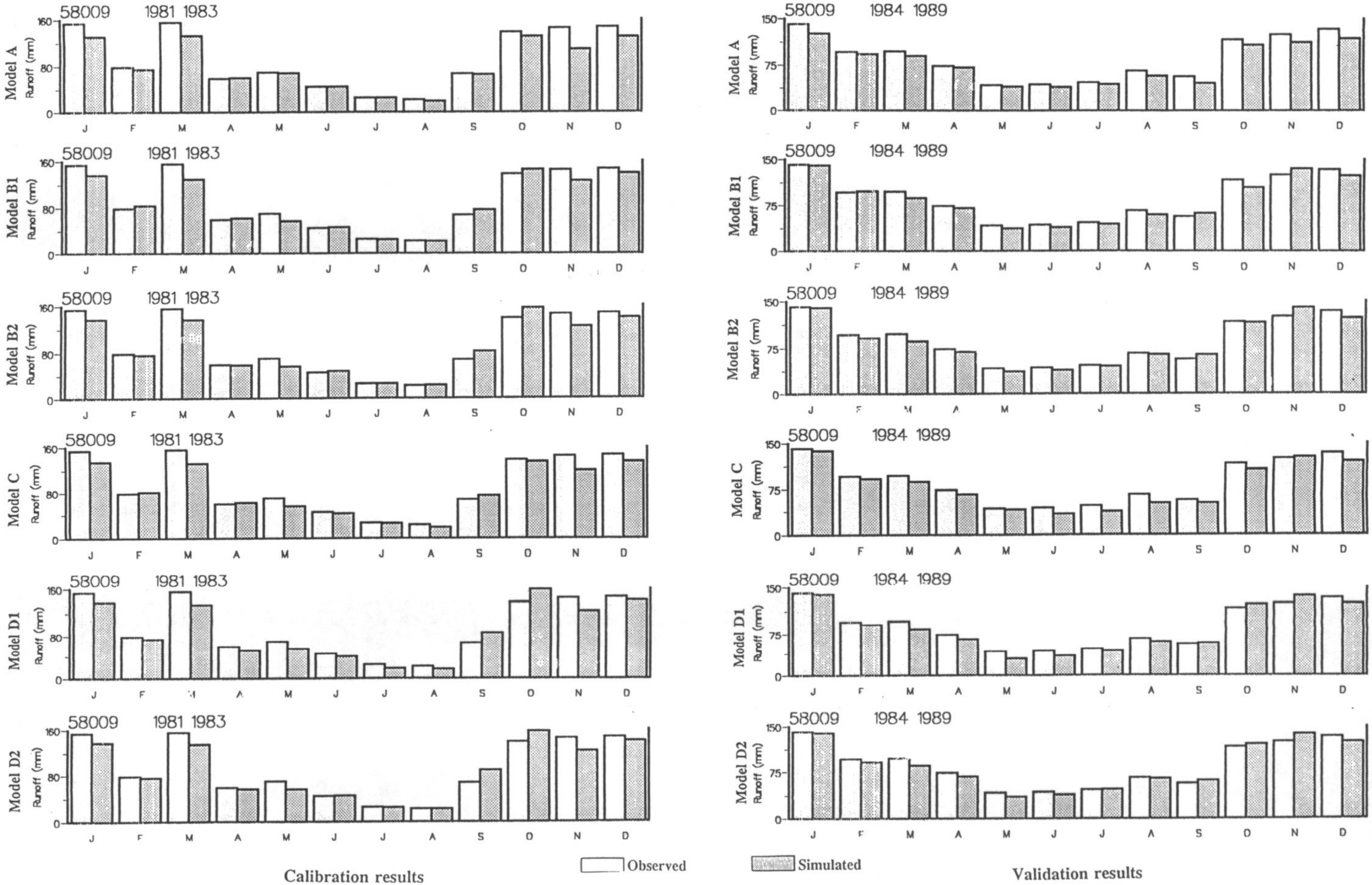
Observed and simulated monthly flow regimes for catchment 54008.



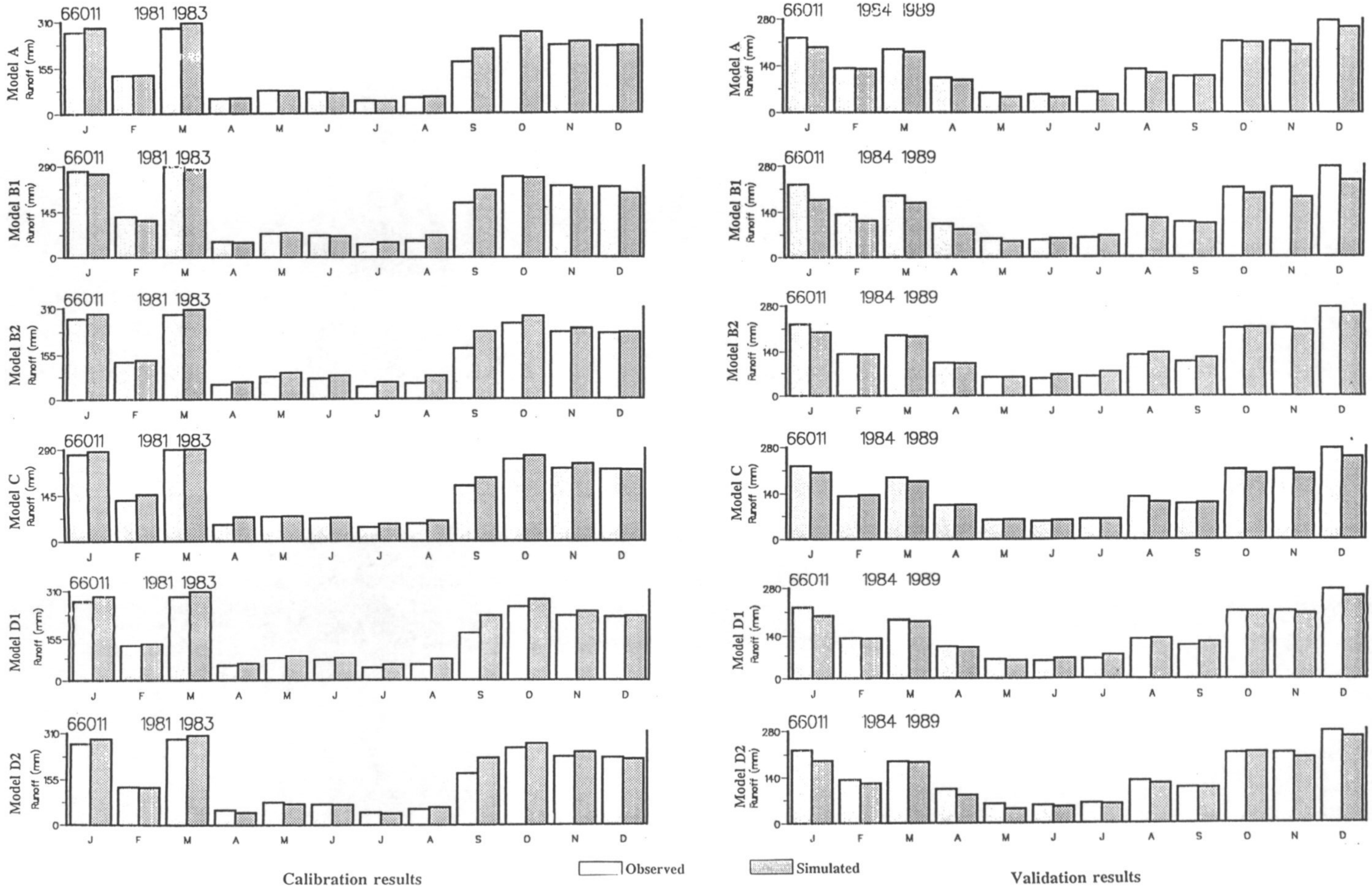
Observed and simulated monthly flow regimes for catchment 54016.



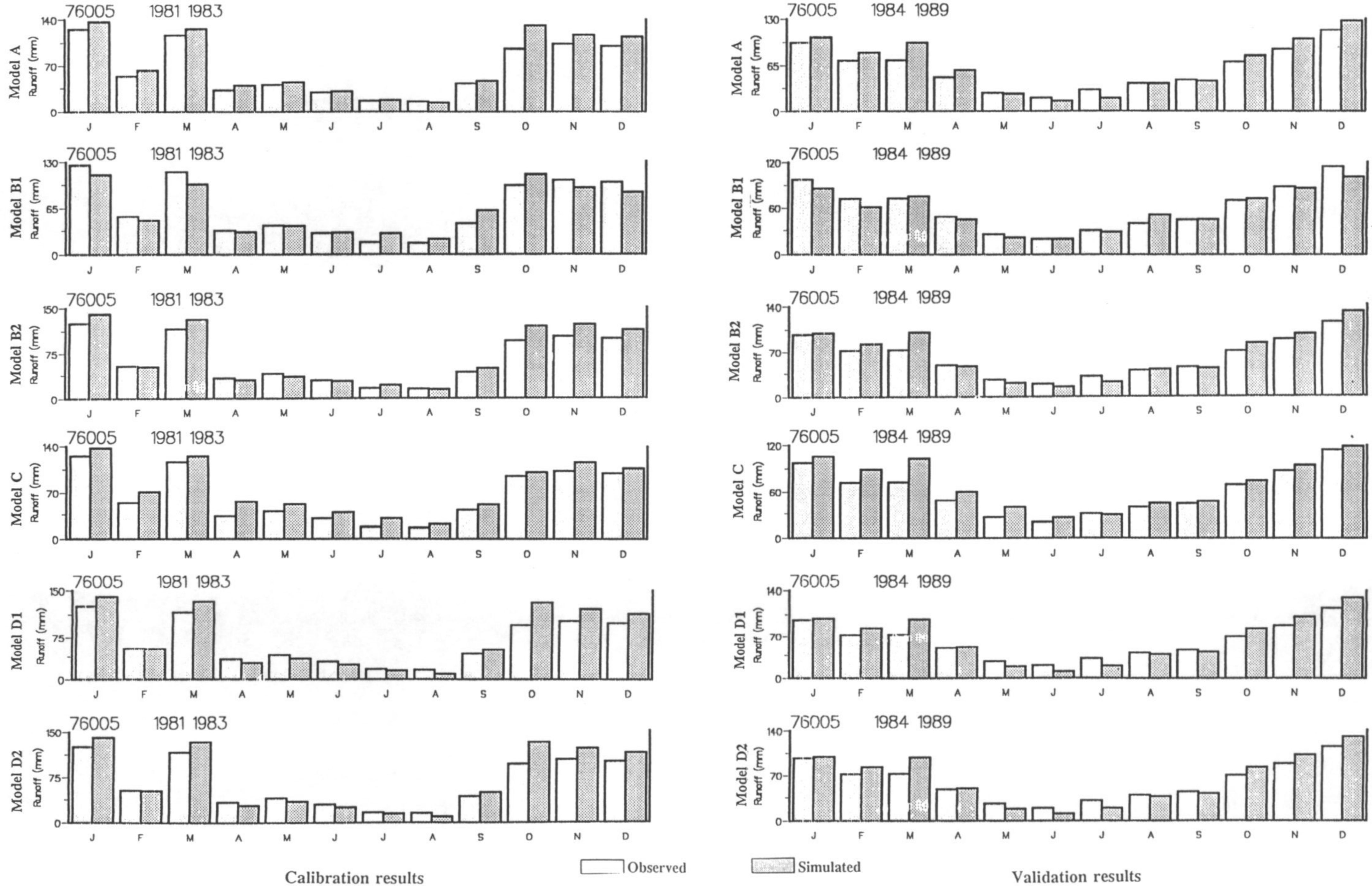
Observed and simulated monthly flow regimes for catchment 57004.



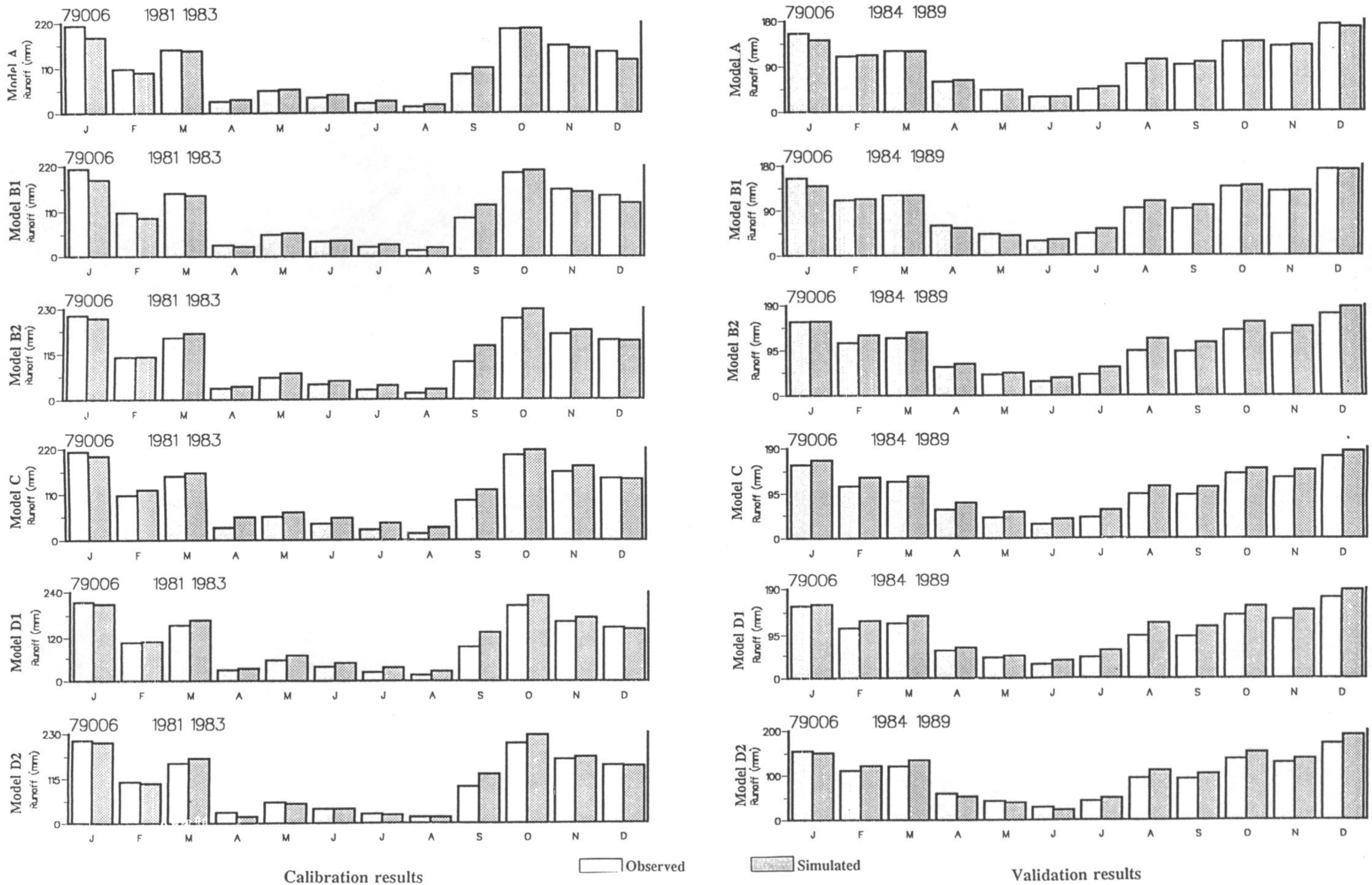
Observed and simulated monthly flow regimes for catchment 58009.



Observed and simulated monthly flow regimes for catchment 66011.



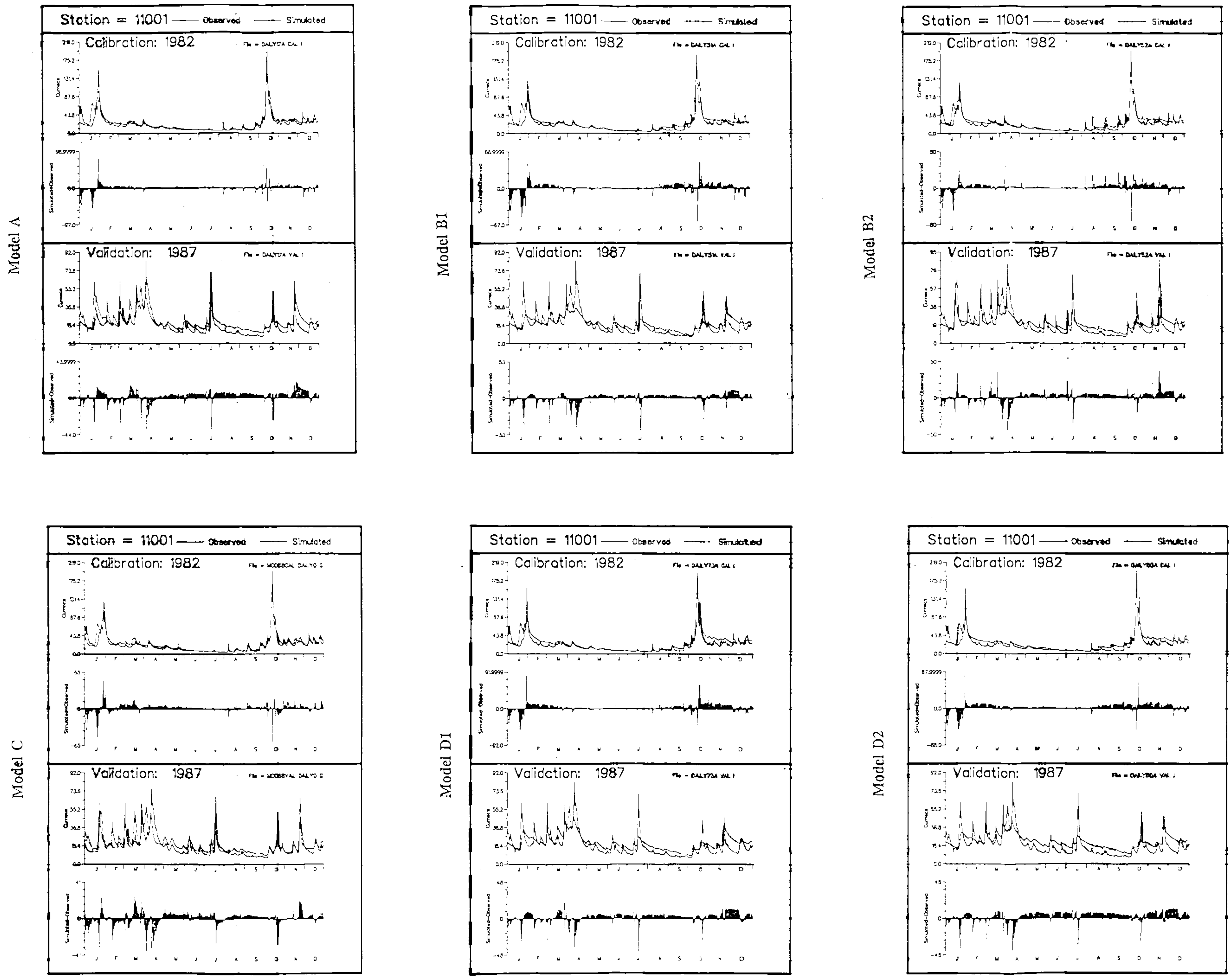
Observed and simulated monthly flow regimes for catchment 76005.



Observed and simulated monthly flow regimes for catchment 79006.

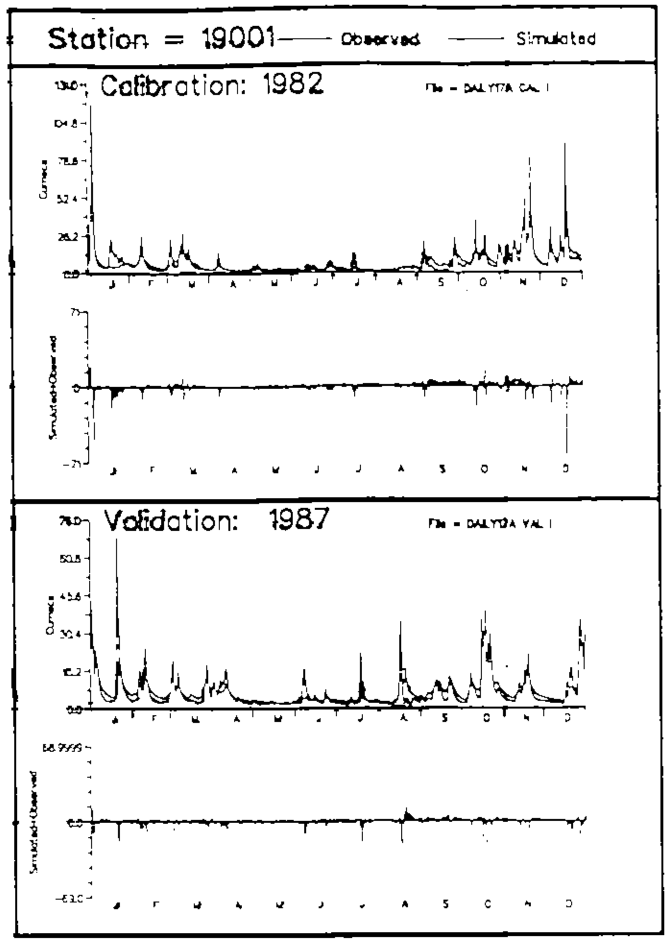
Appendix A

A.5 Plots of observed and simulated daily flow regimes and residuals

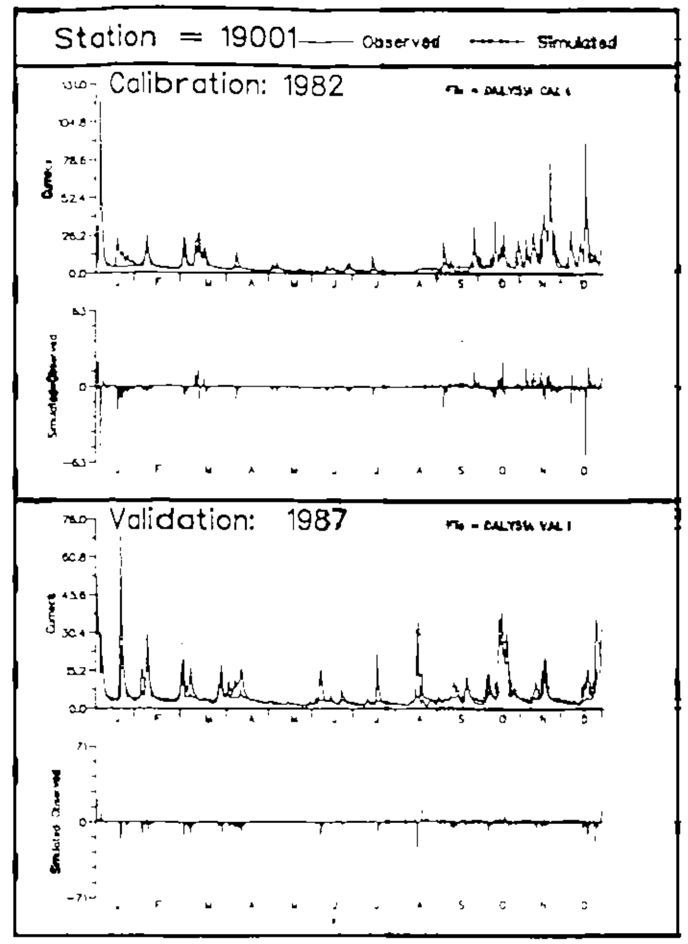


Observed and simulated daily flow regimes for catchment 11001.

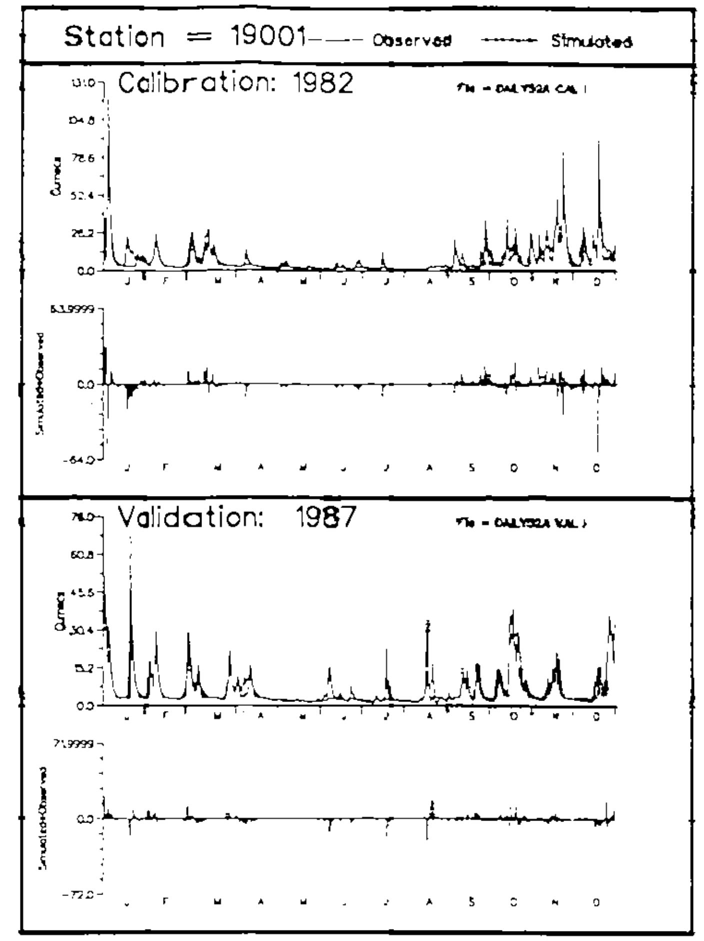
Model A



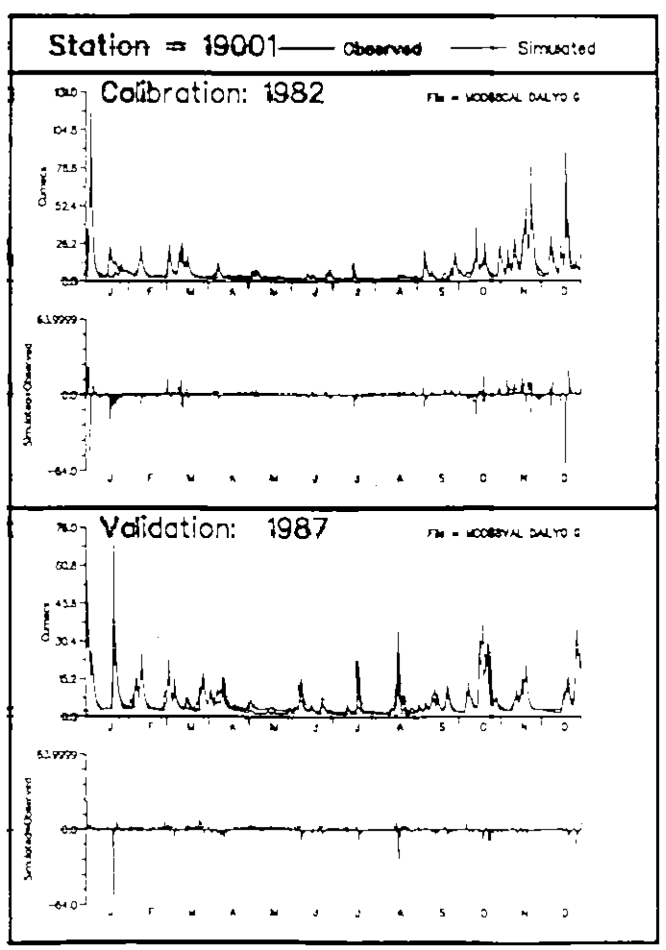
Model B1



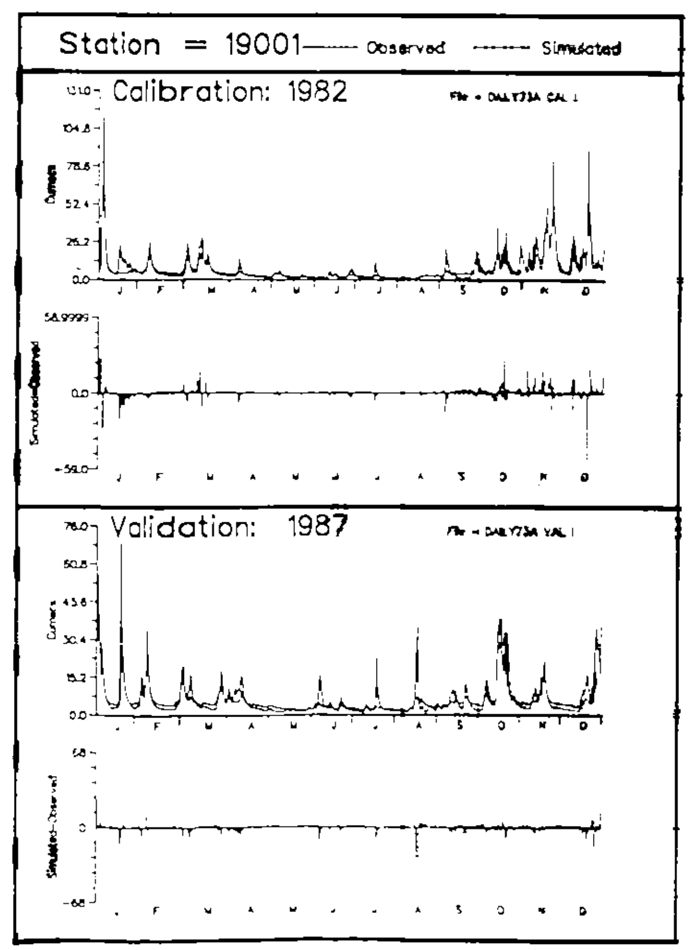
Model B2



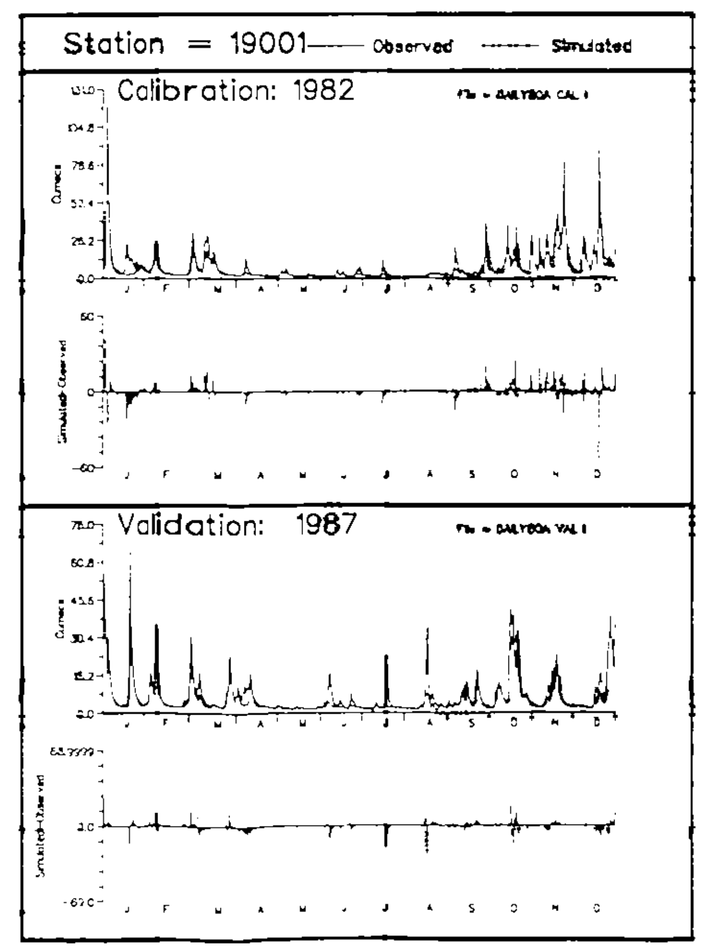
Model C



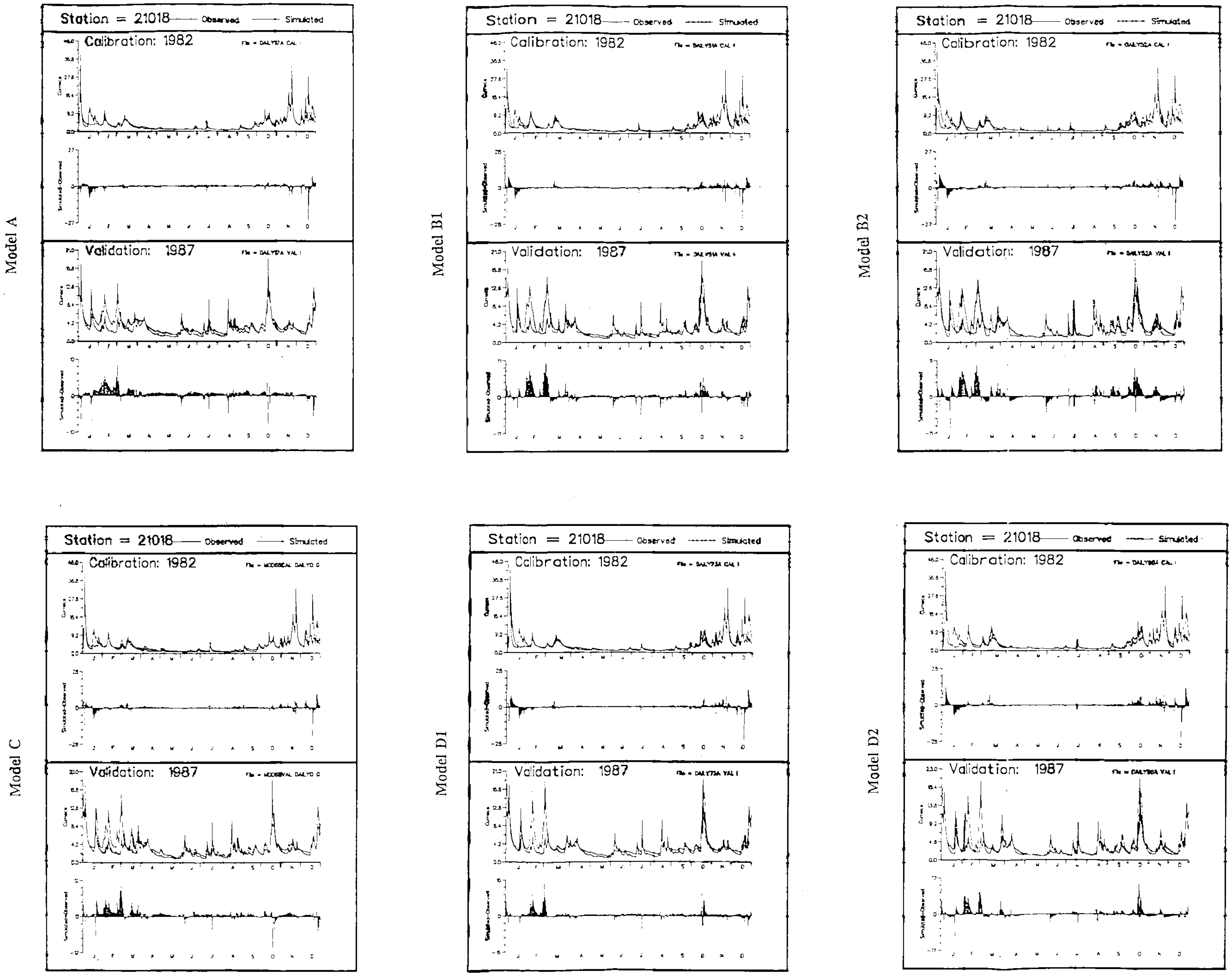
Model D1



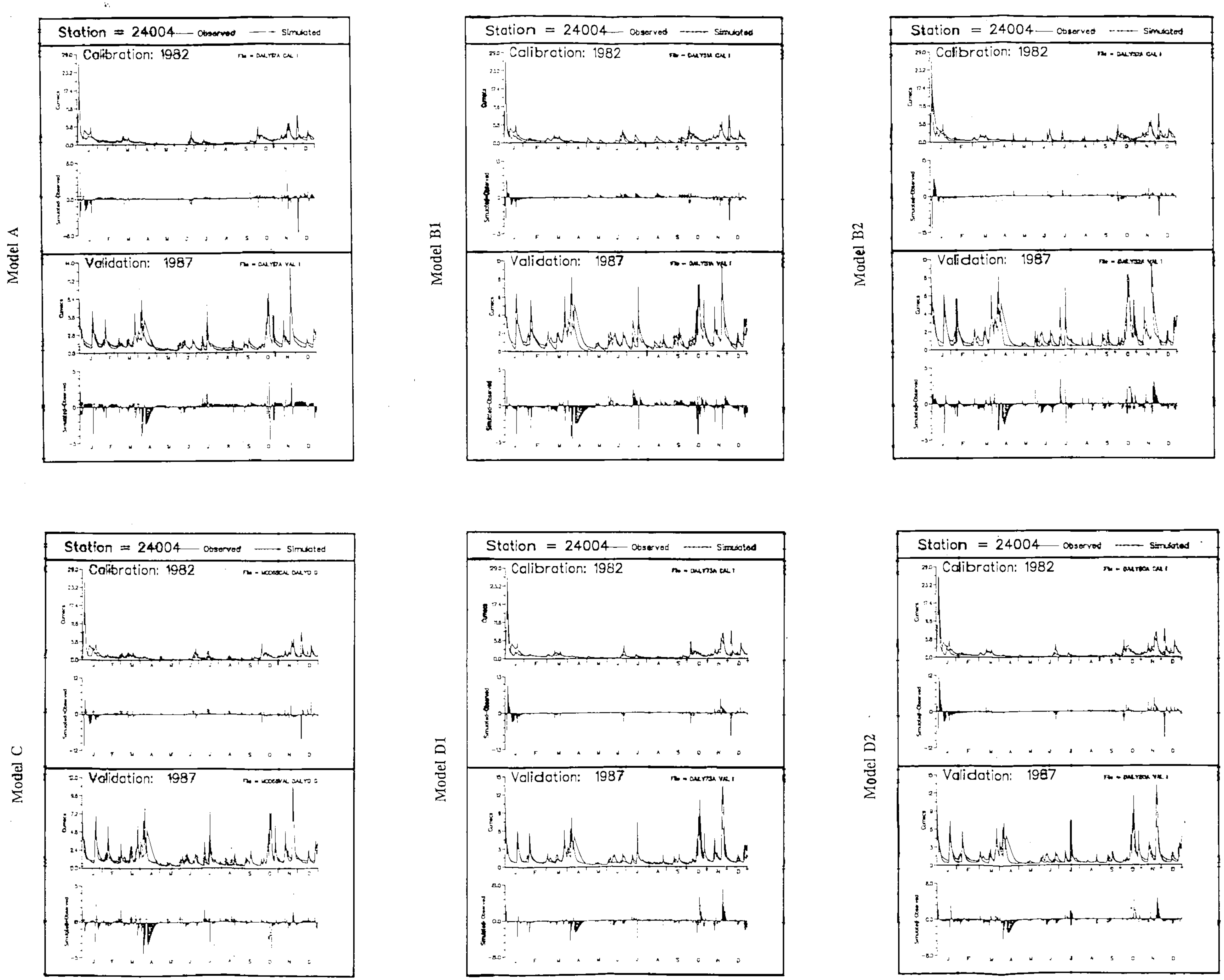
Model D2



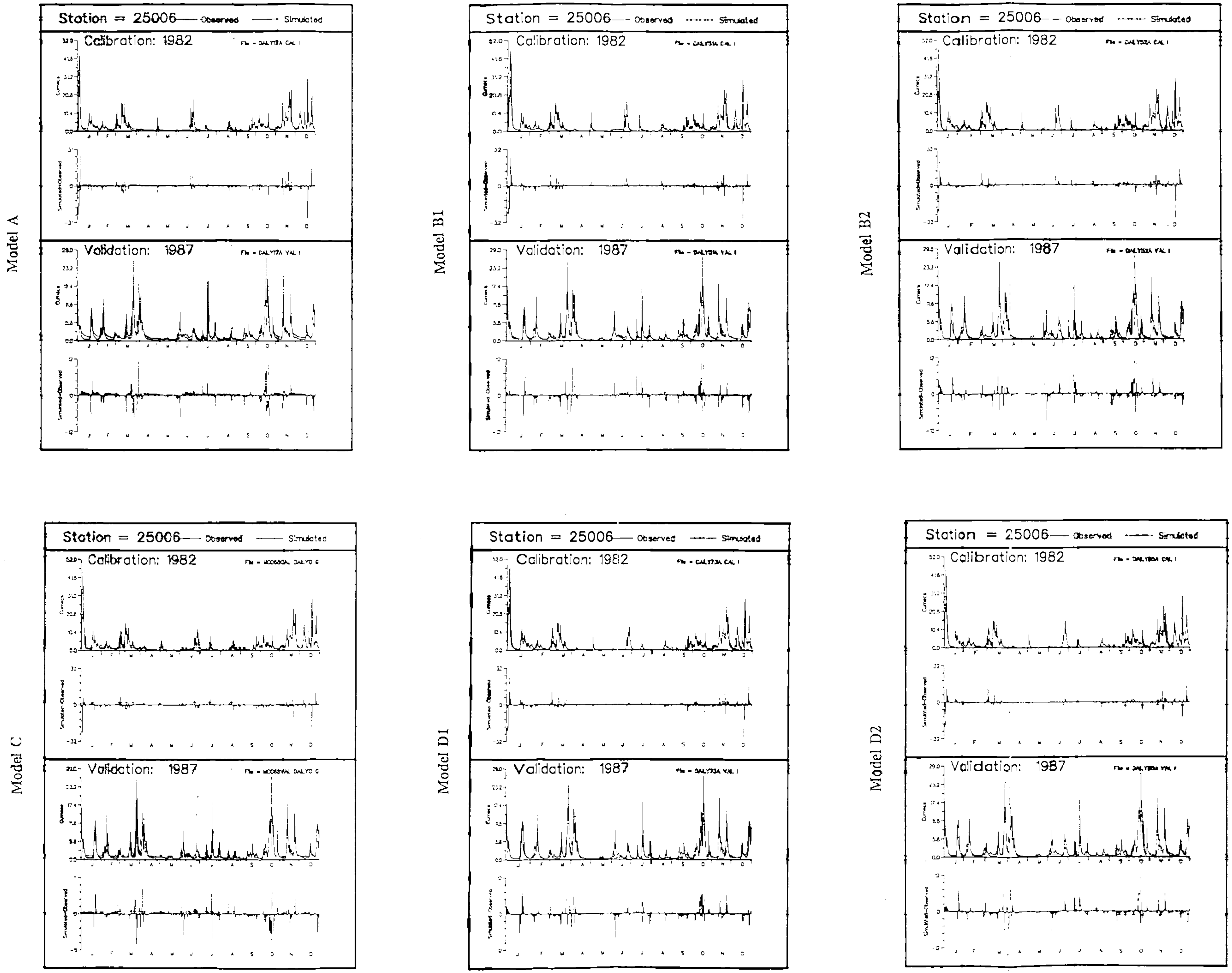
Observed and simulated daily flow regimes for catchment 19001.



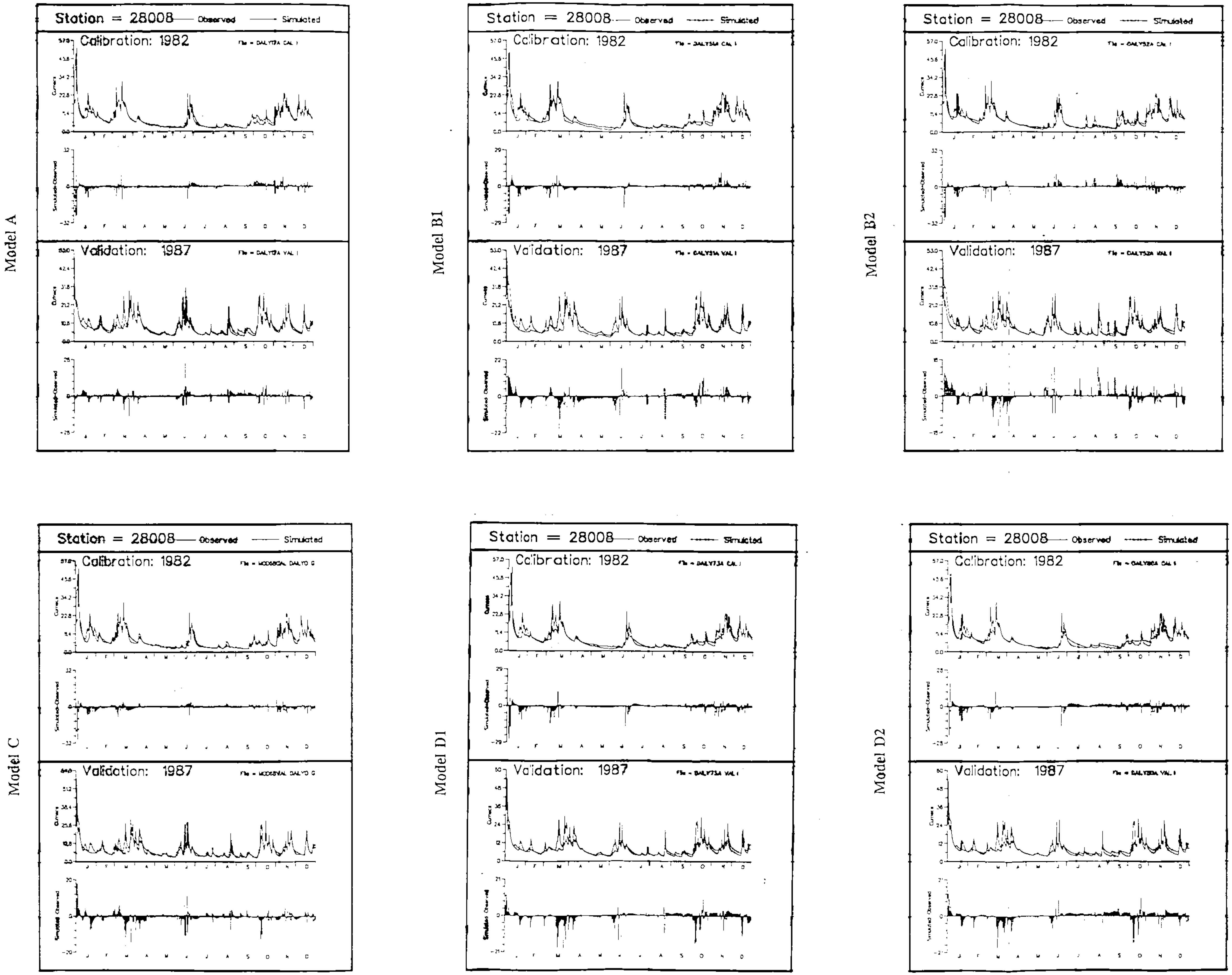
Observed and simulated daily flow regimes for catchment 21018.



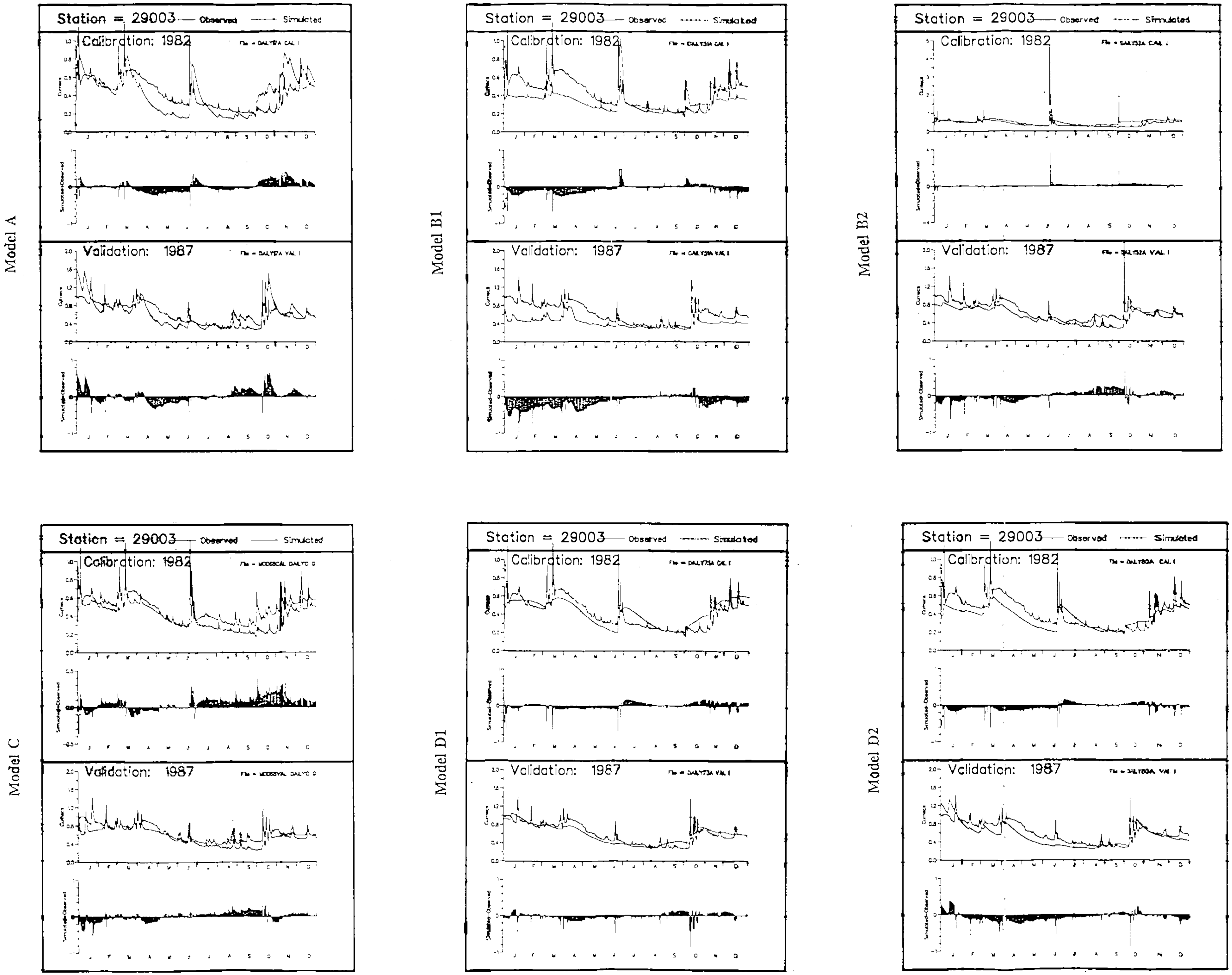
Observed and simulated daily flow regimes for catchment 24004.



Observed and simulated daily flow regimes for catchment 25006.

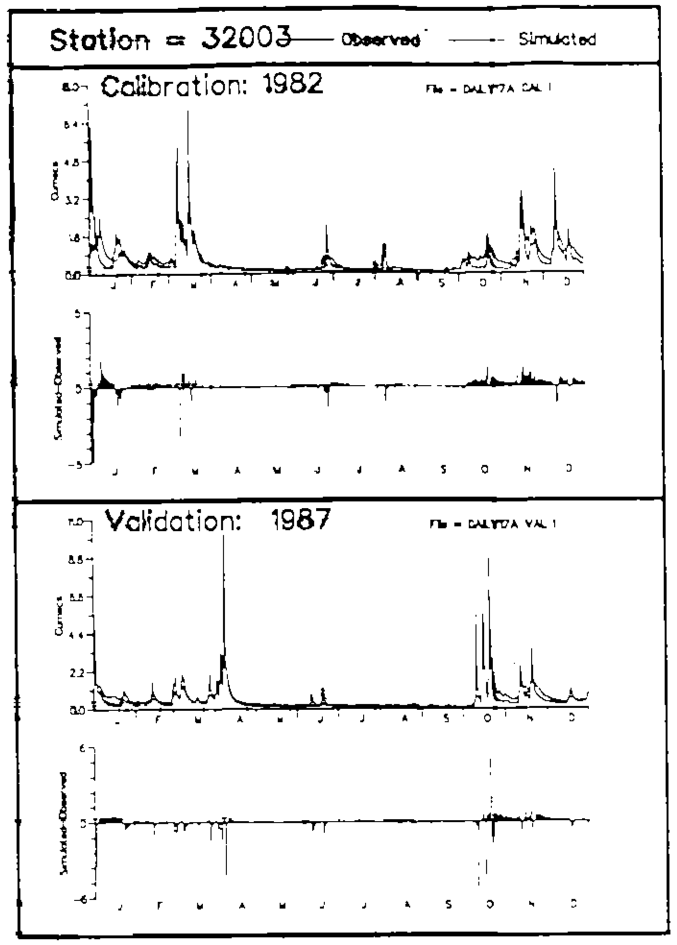


Observed and simulated daily flow regimes for catchment 28008.

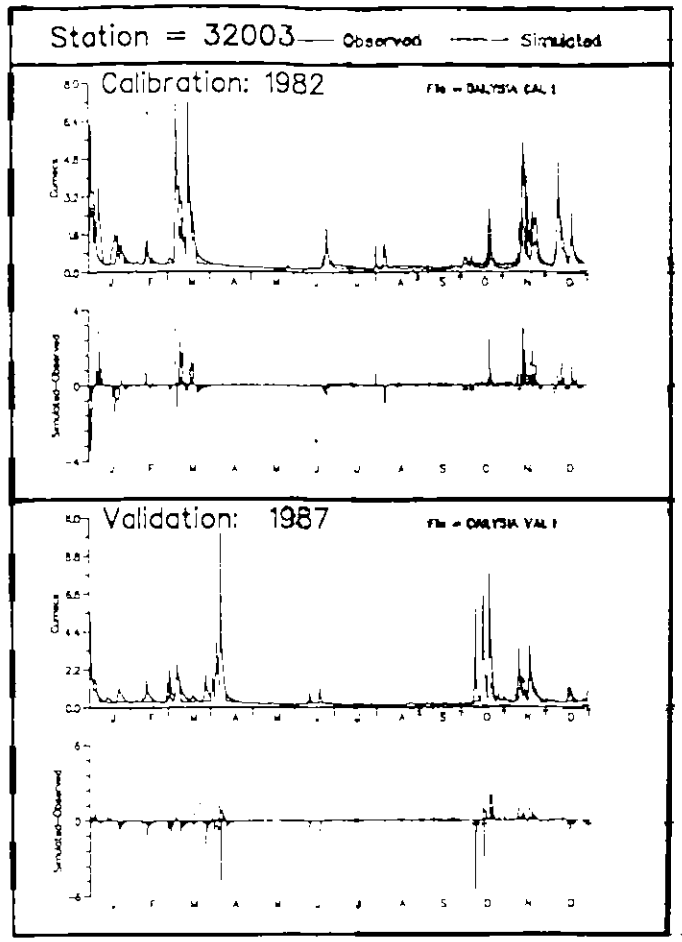


Observed and simulated daily flow regimes for catchment 29003.

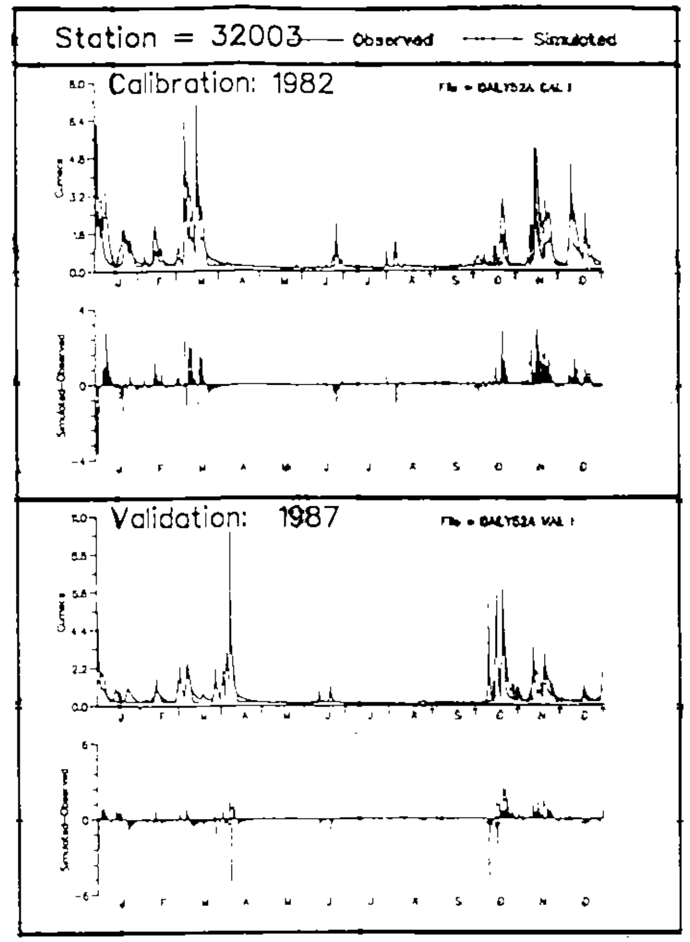
Model A



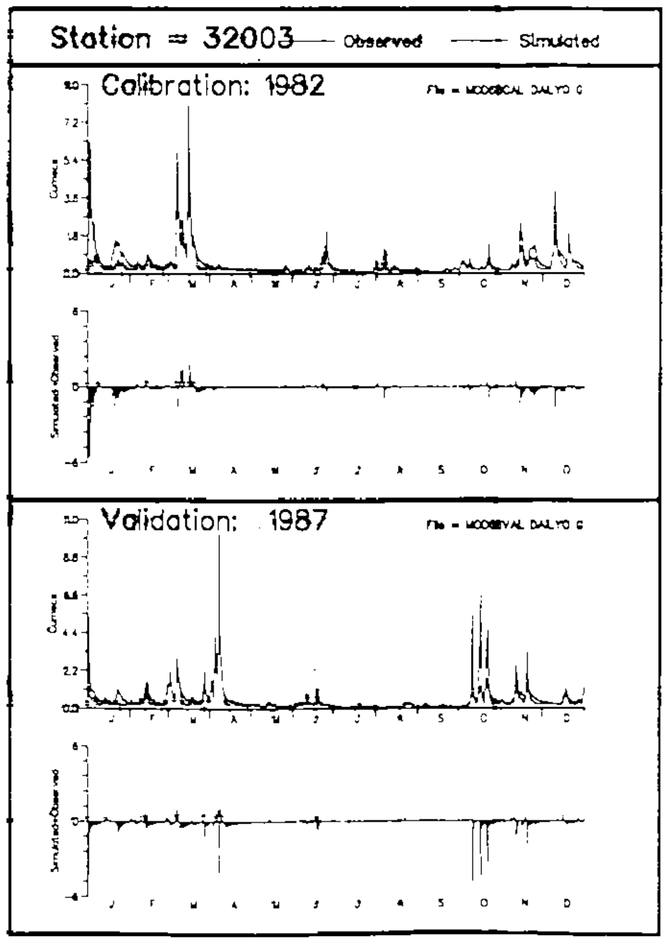
Model B1



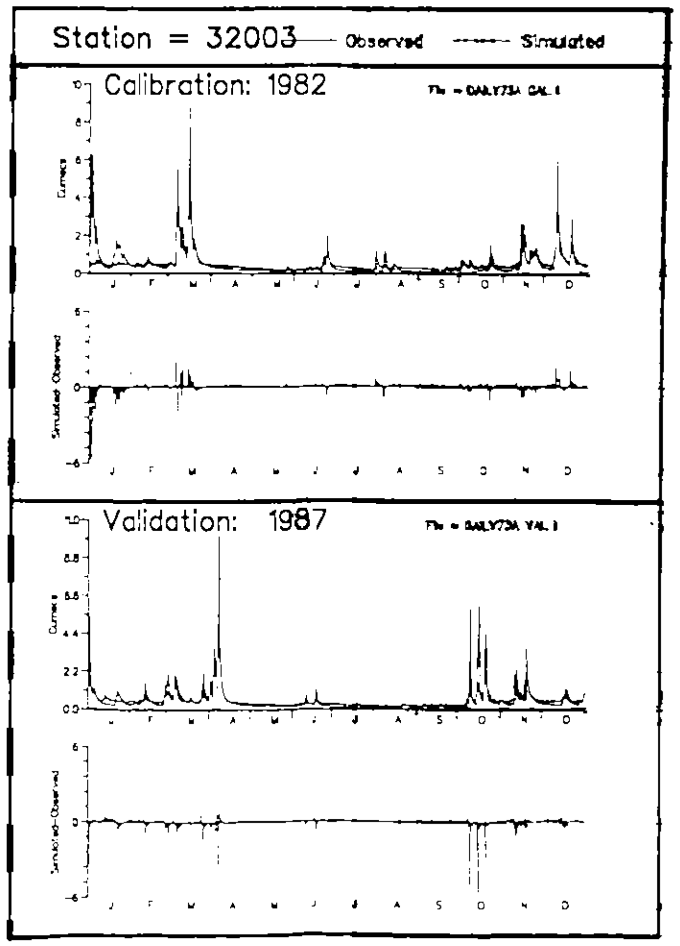
Model B2



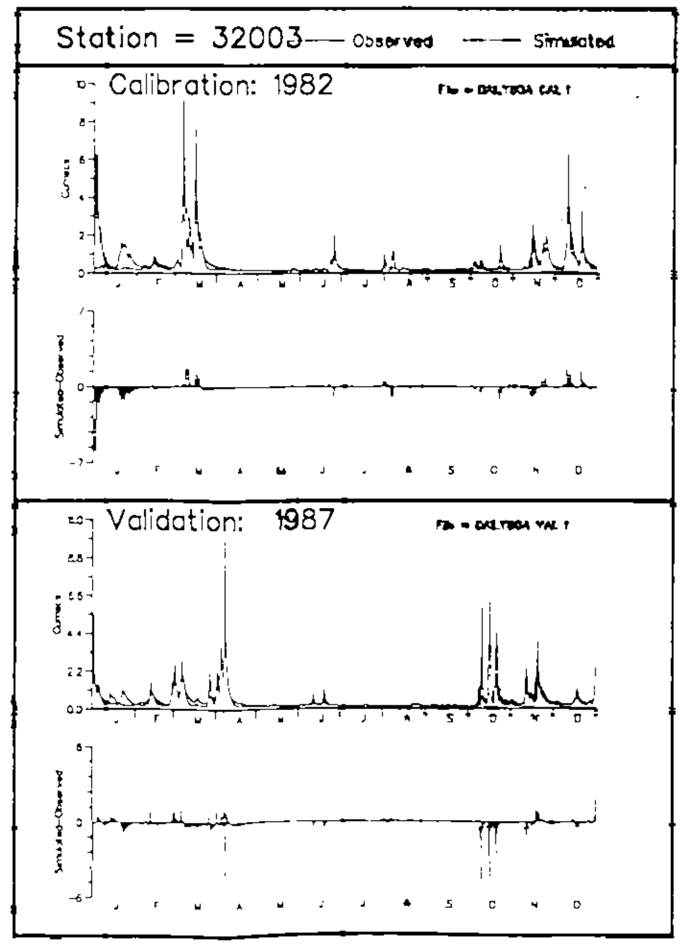
Model C



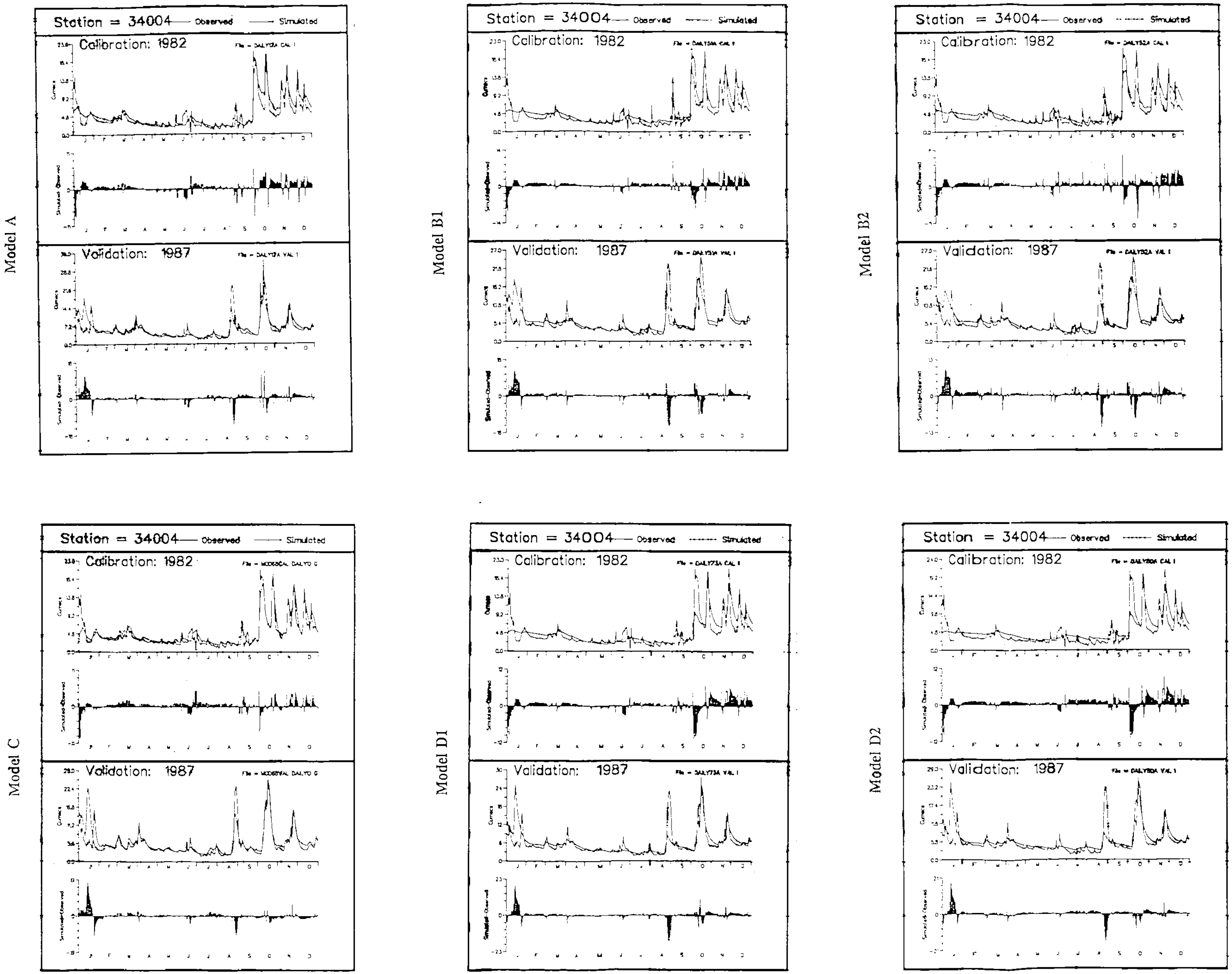
Model D1



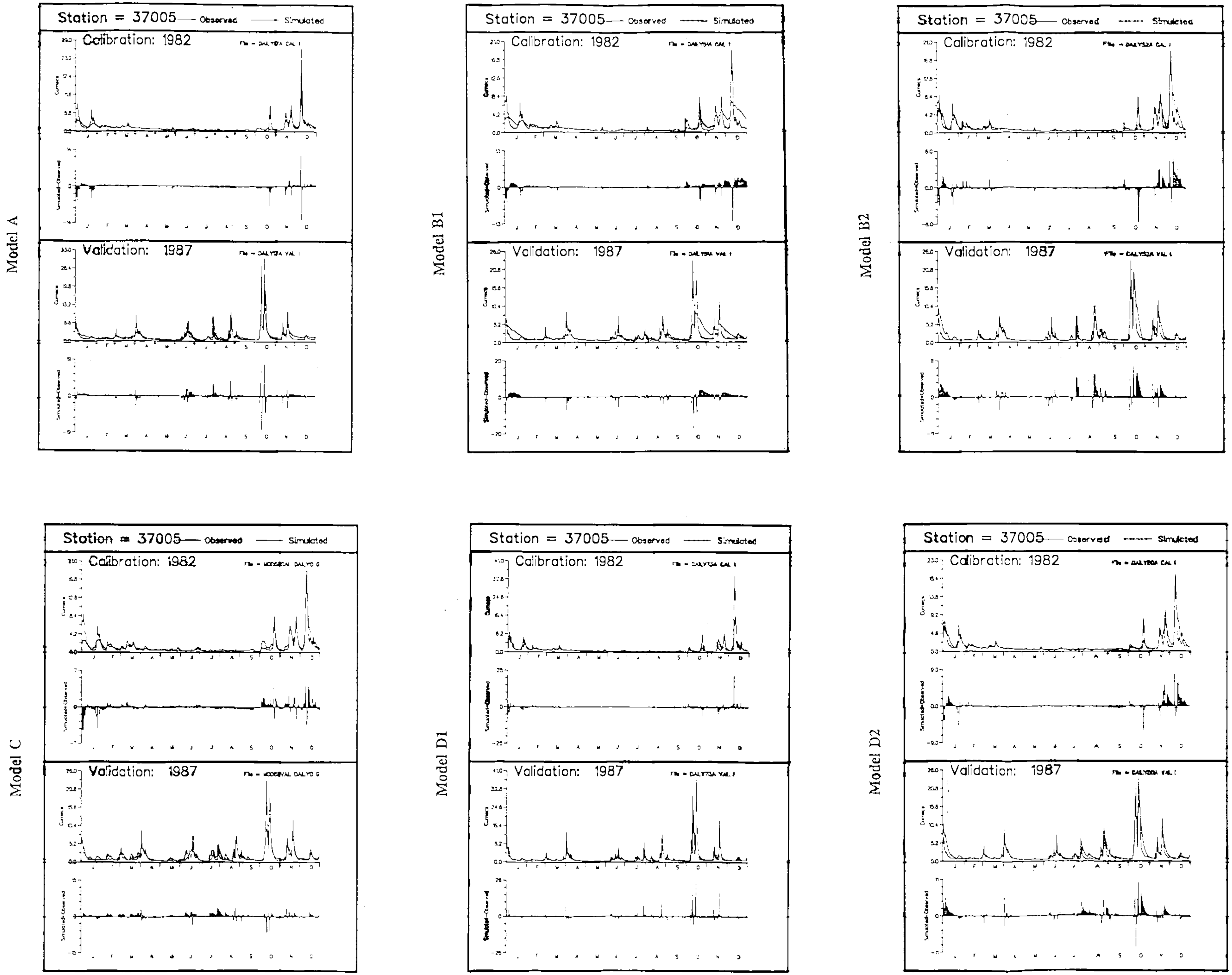
Model D2



Observed and simulated daily flow regimes for catchment 32003.

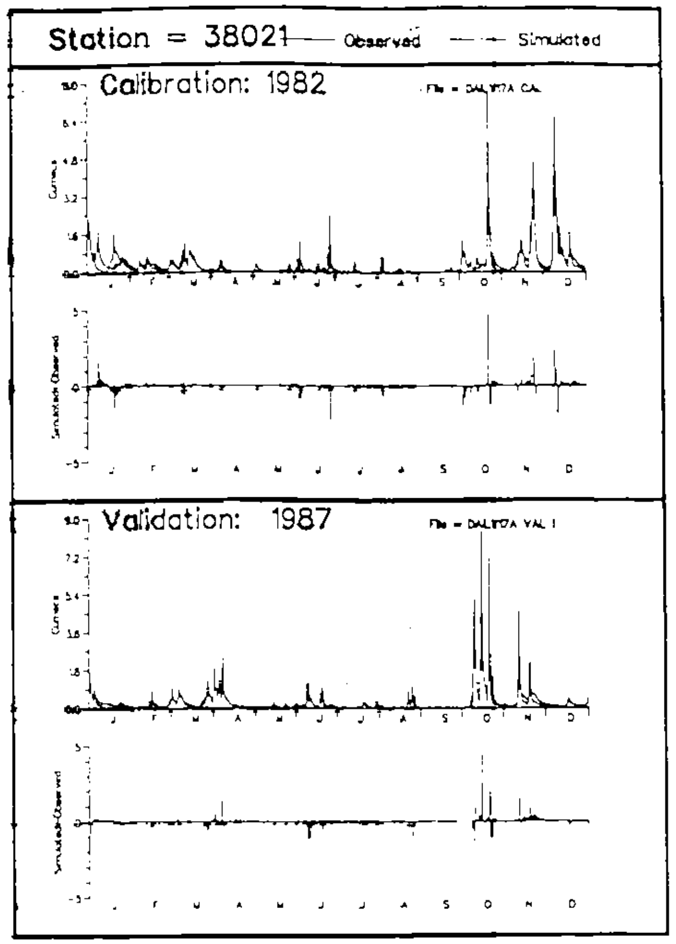


Observed and simulated daily flow regimes for catchment 34004.

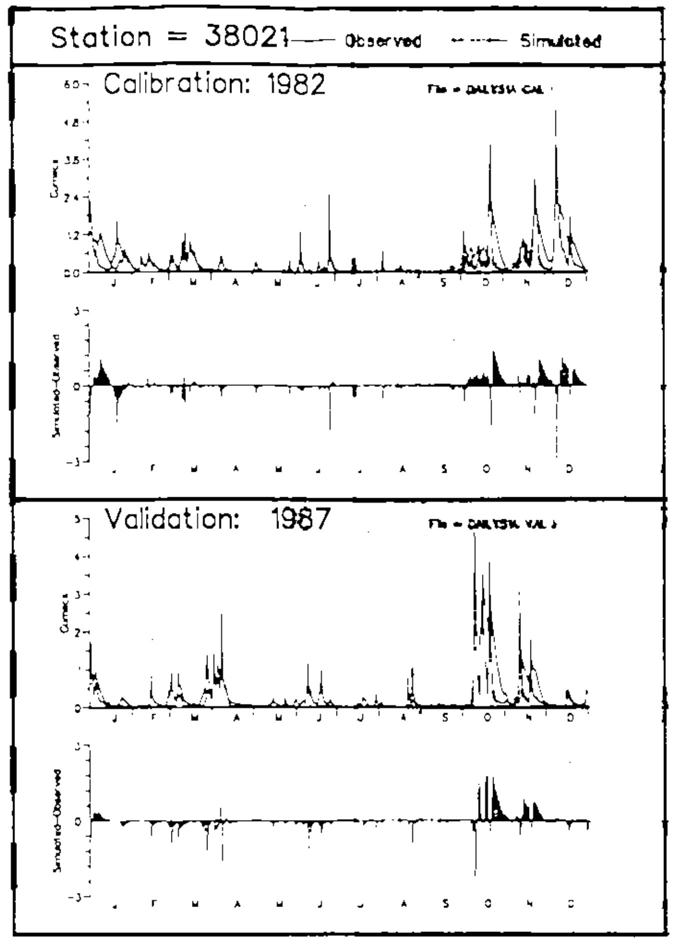


Observed and simulated daily flow regimes for catchment 37005.

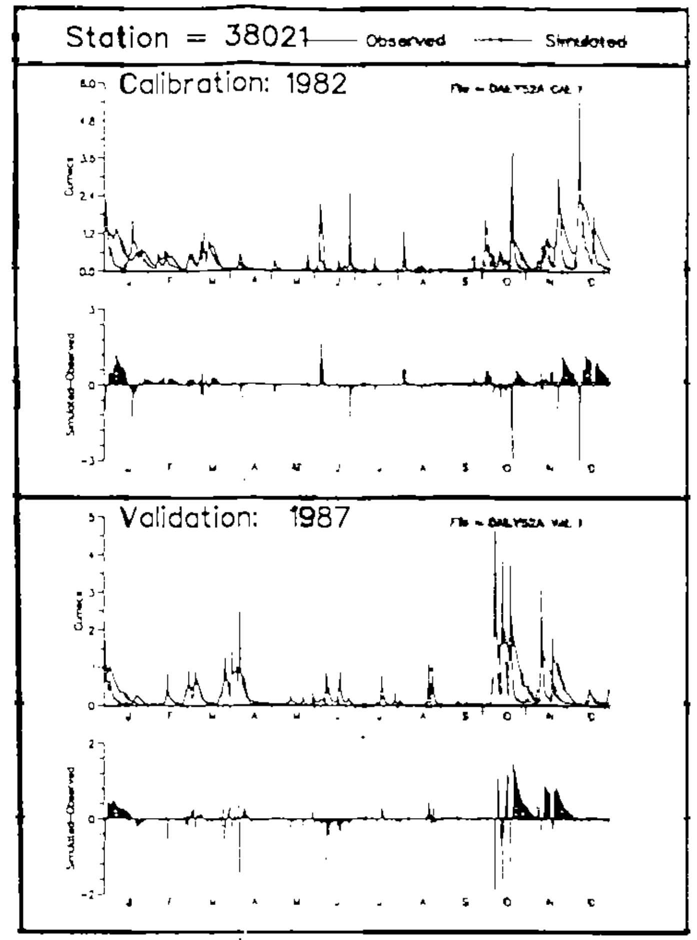
Model A



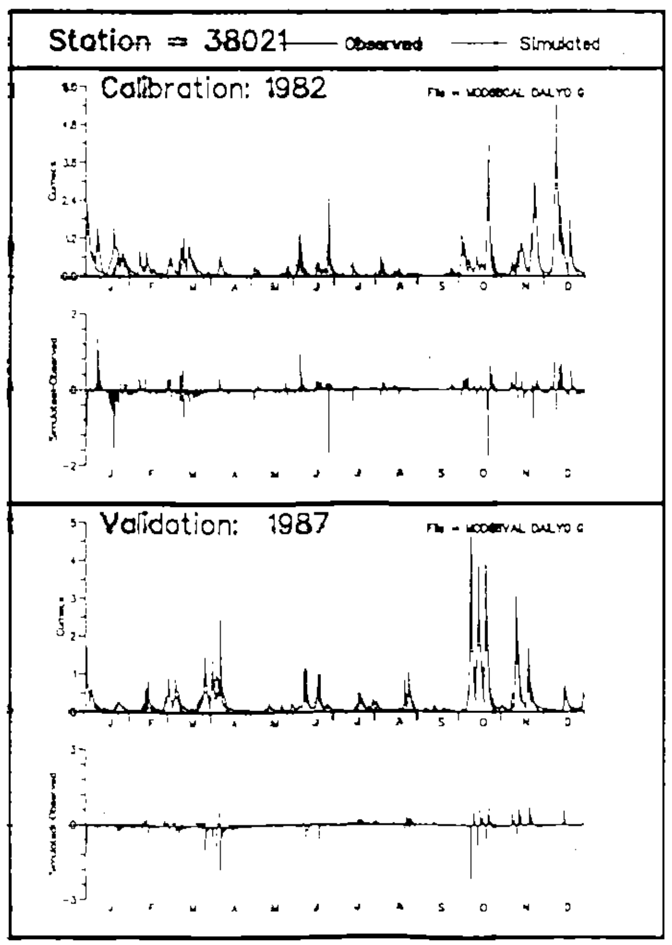
Model B1



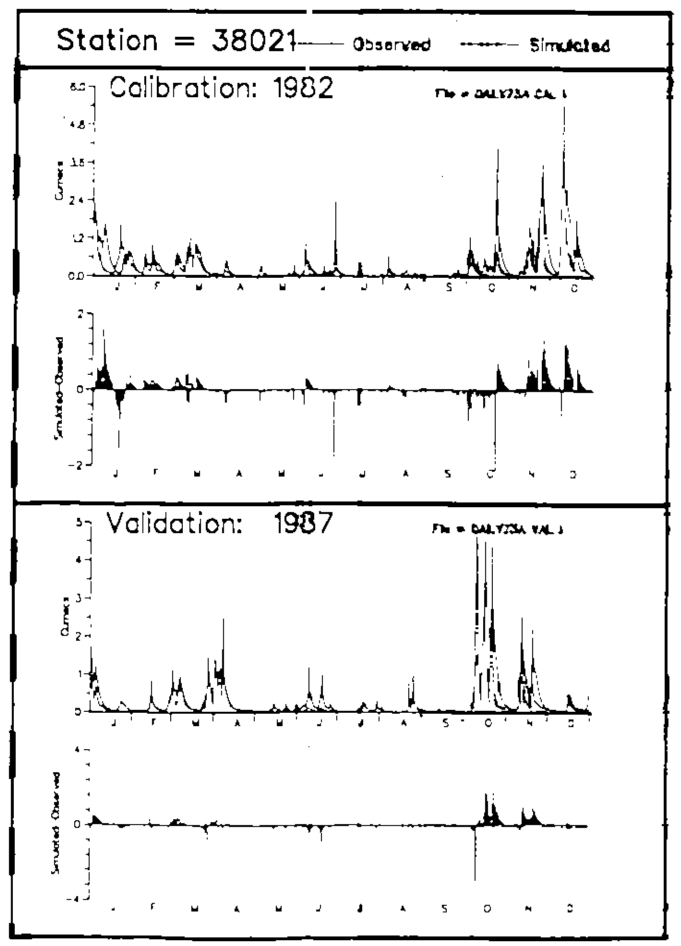
Model B2



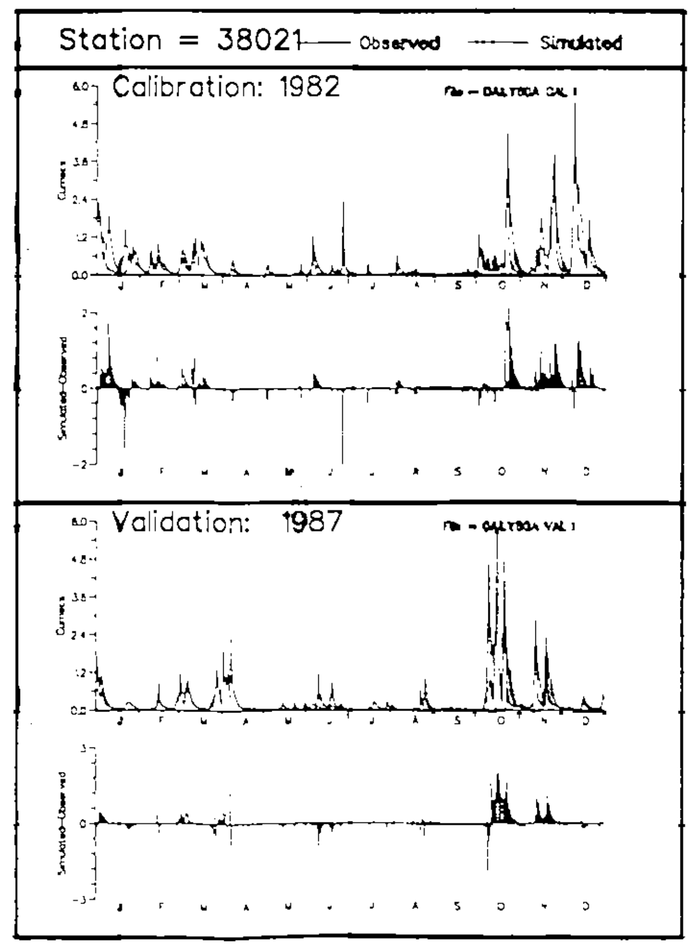
Model C



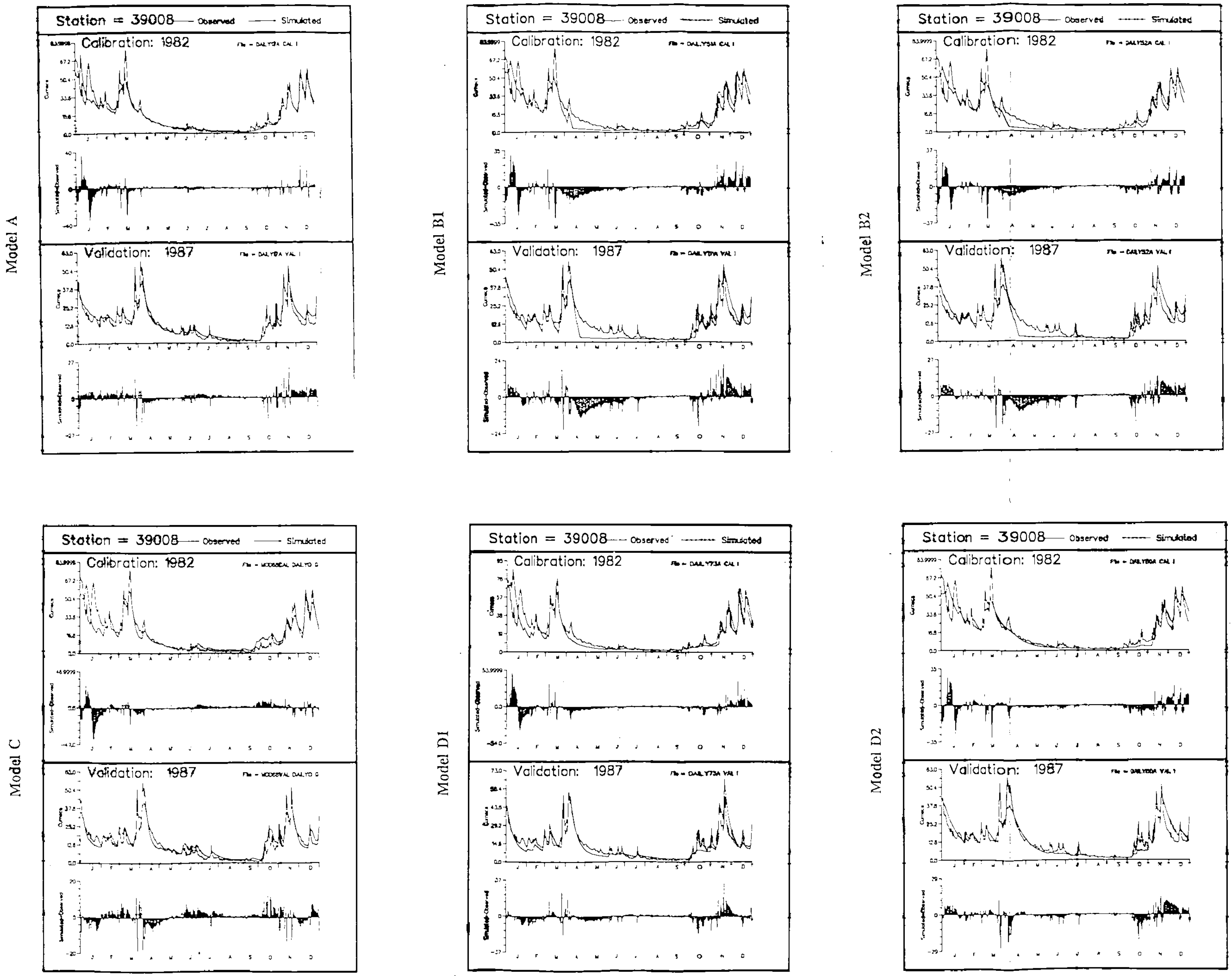
Model D1



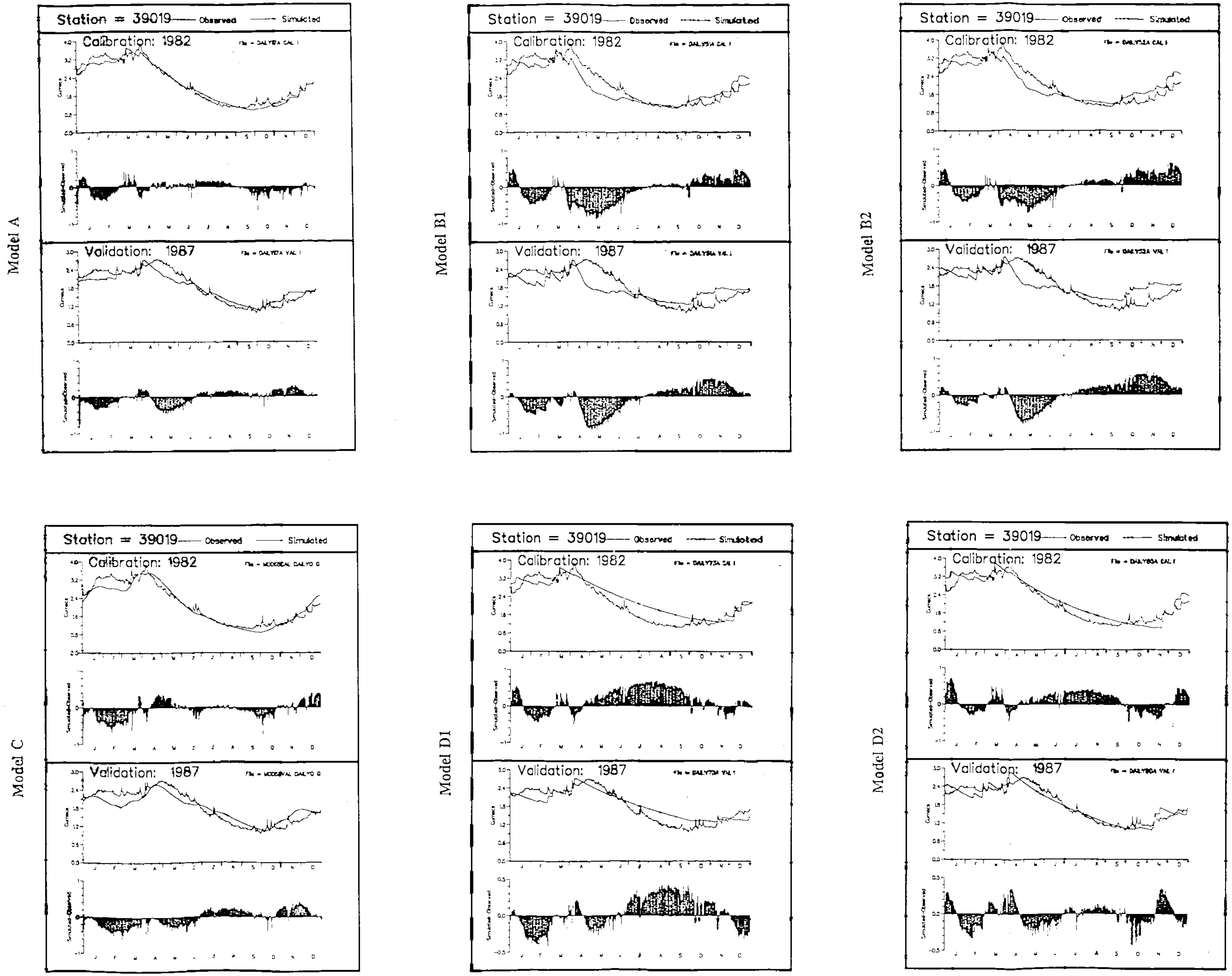
Model D2



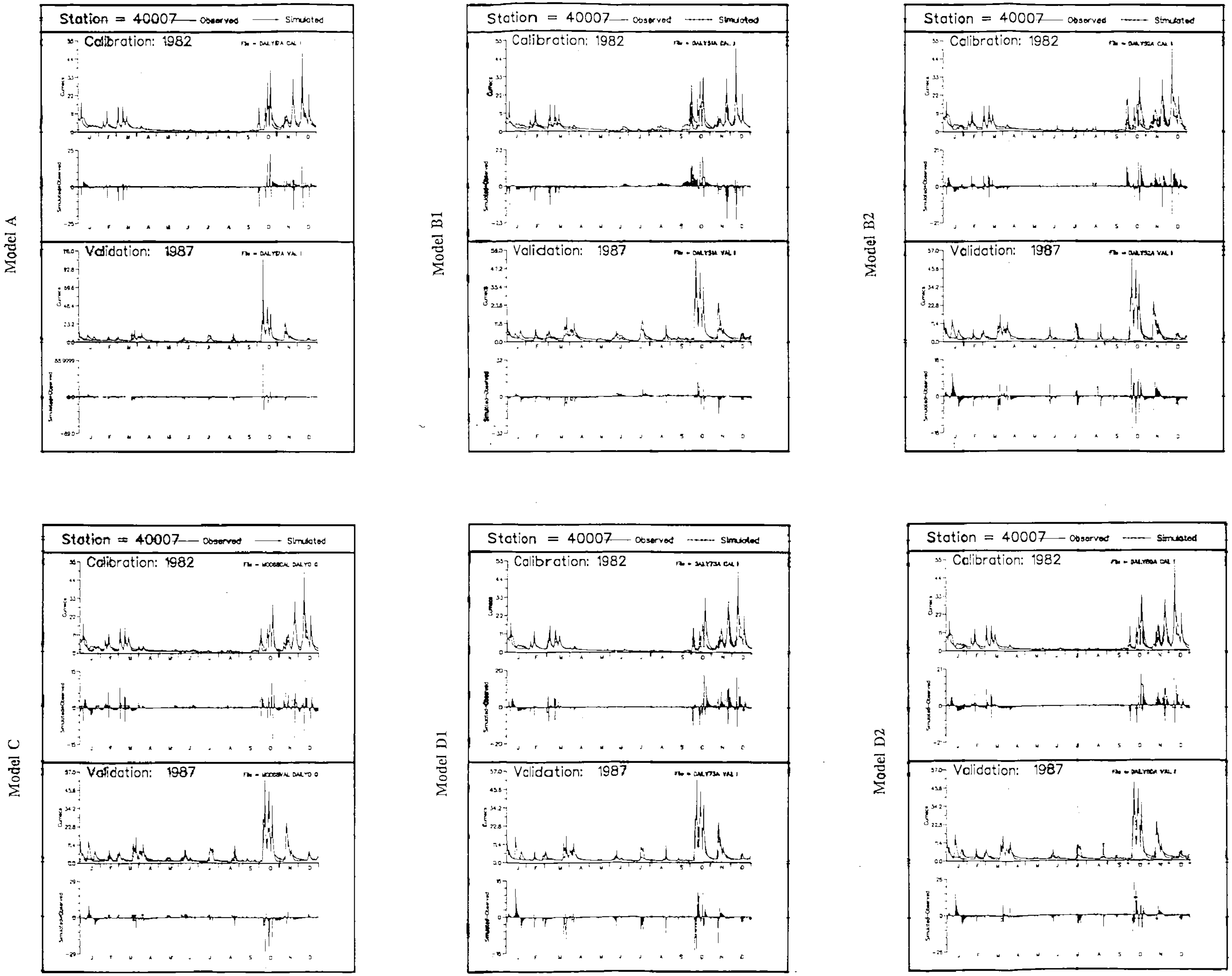
Observed and simulated daily flow regimes for catchment 38021.



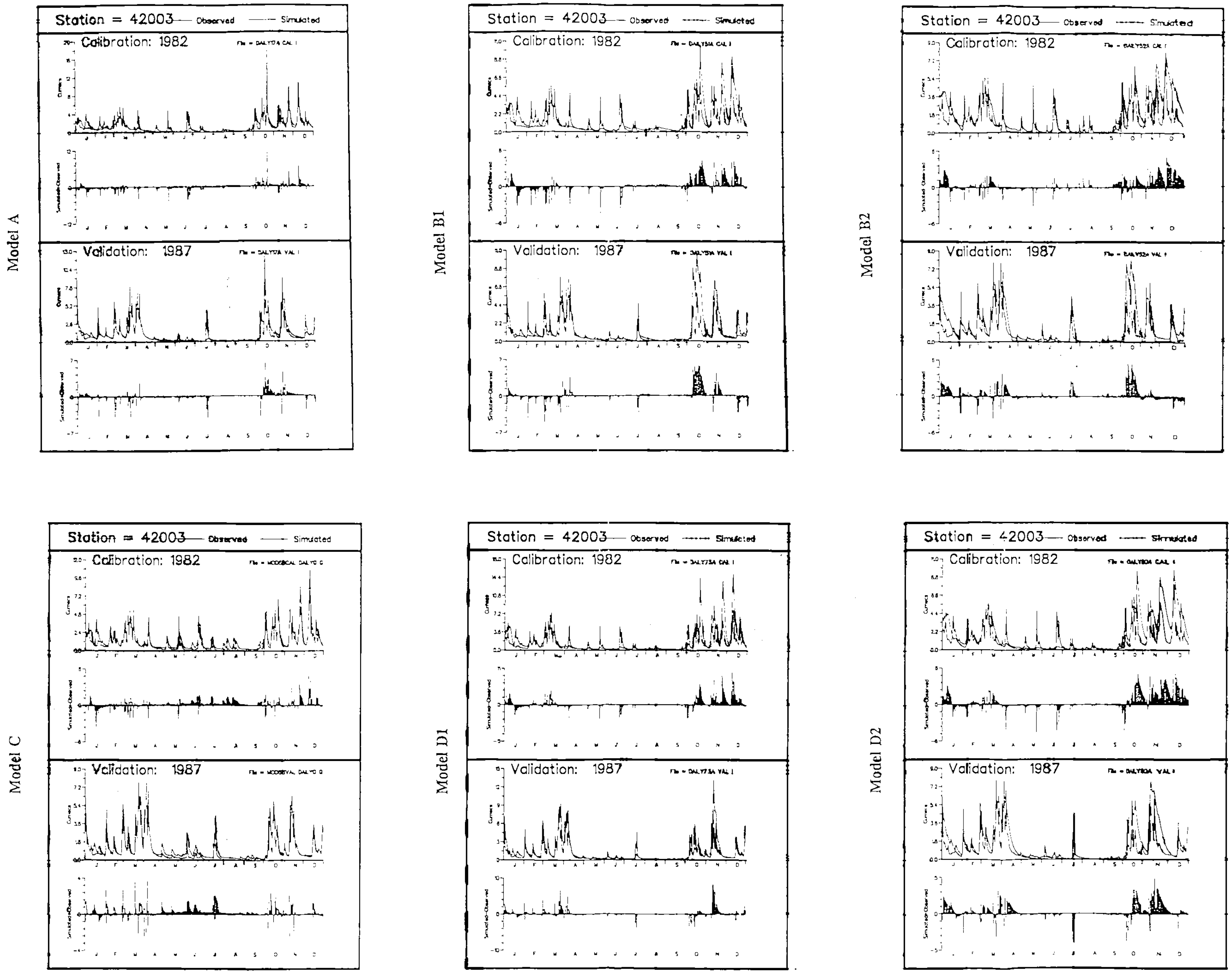
Observed and simulated daily flow regimes for catchment 39008.



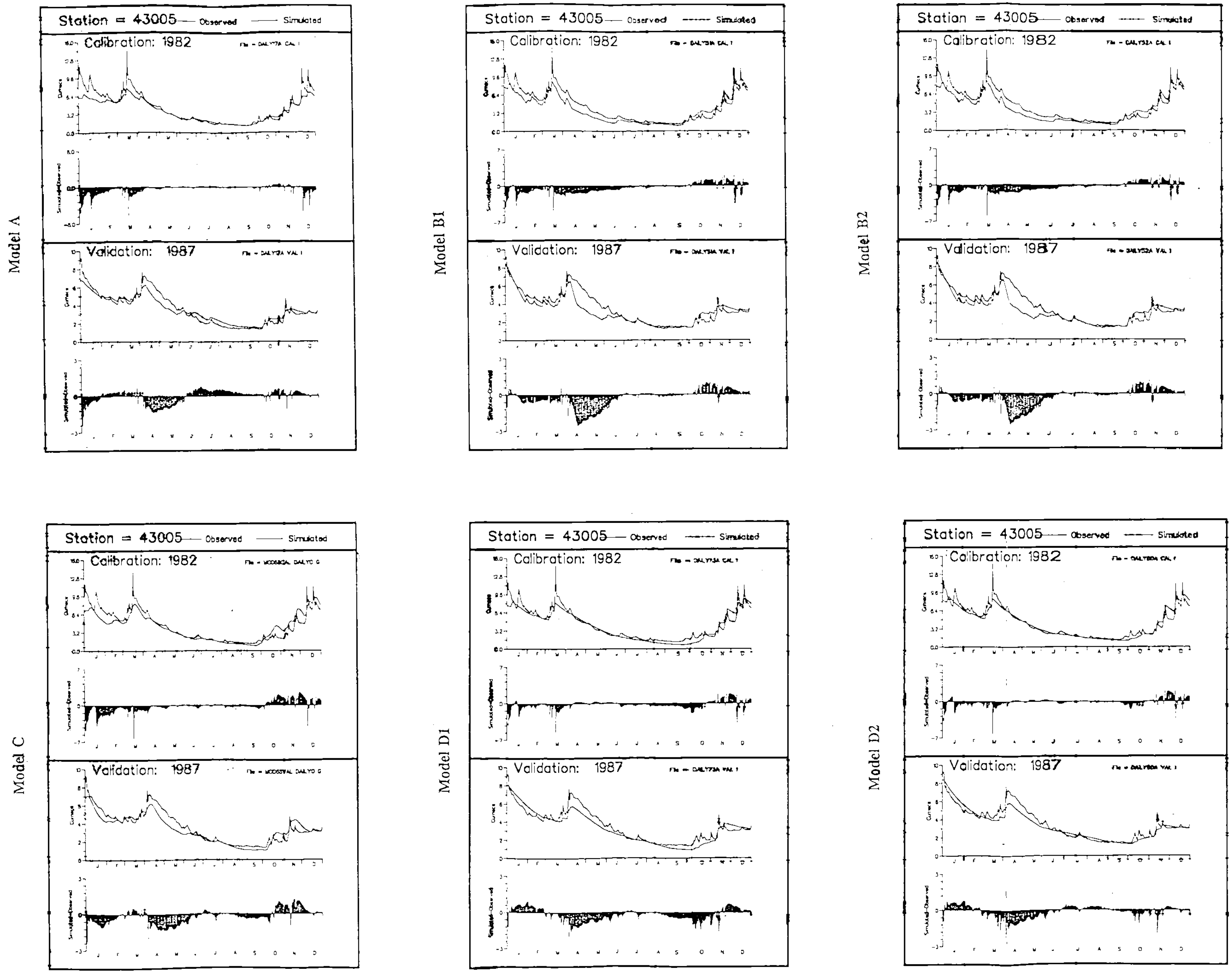
Observed and simulated daily flow regimes for catchment 39019.



Observed and simulated daily flow regimes for catchment 40007.

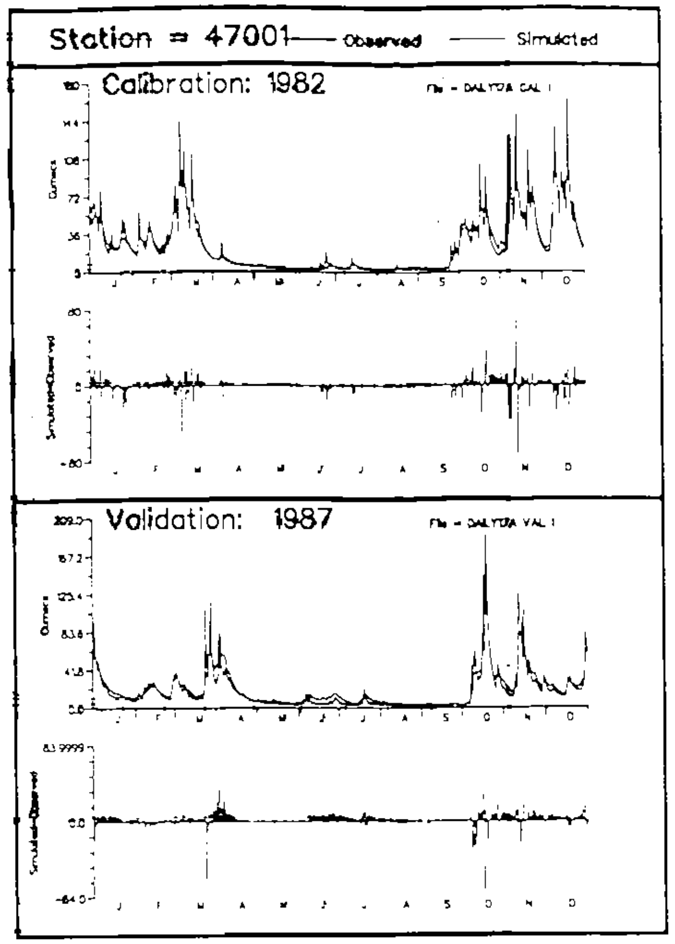


Observed and simulated daily flow regimes for catchment 42003.

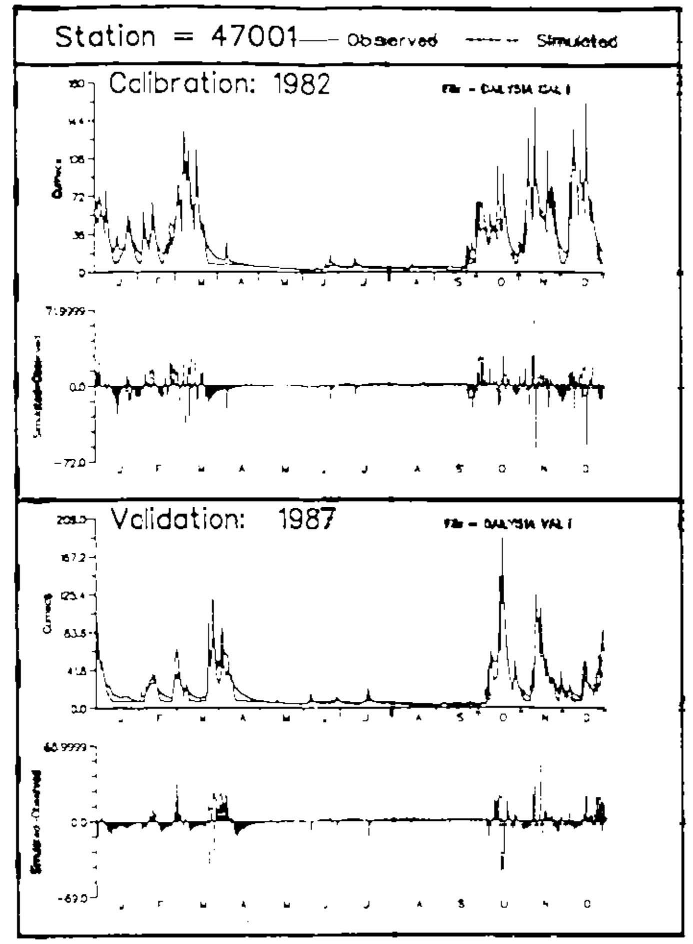


Observed and simulated daily flow regimes for catchment 43005.

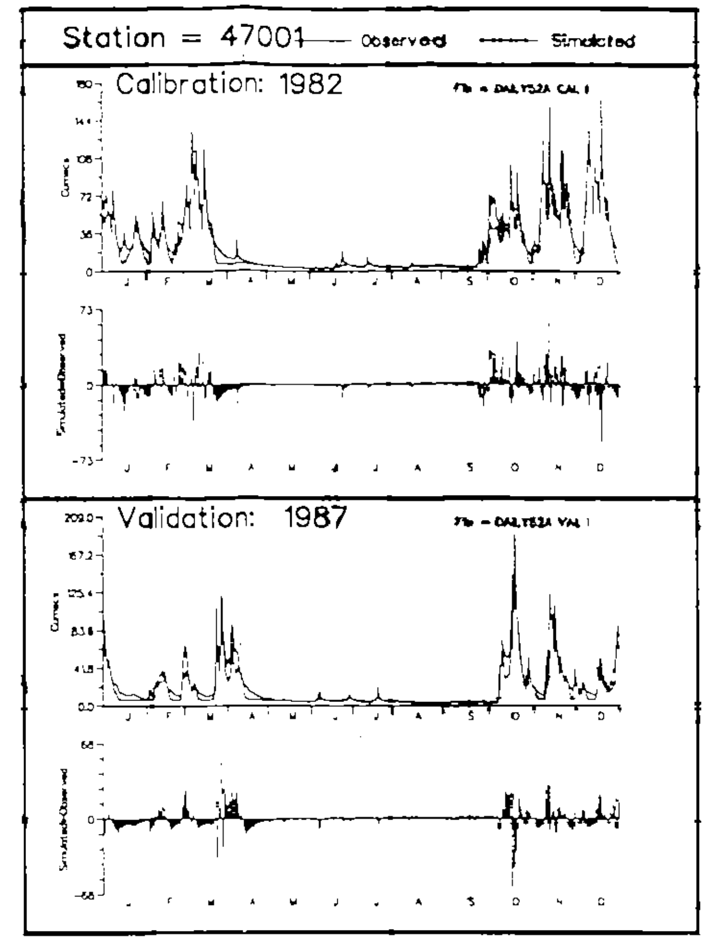
Model A



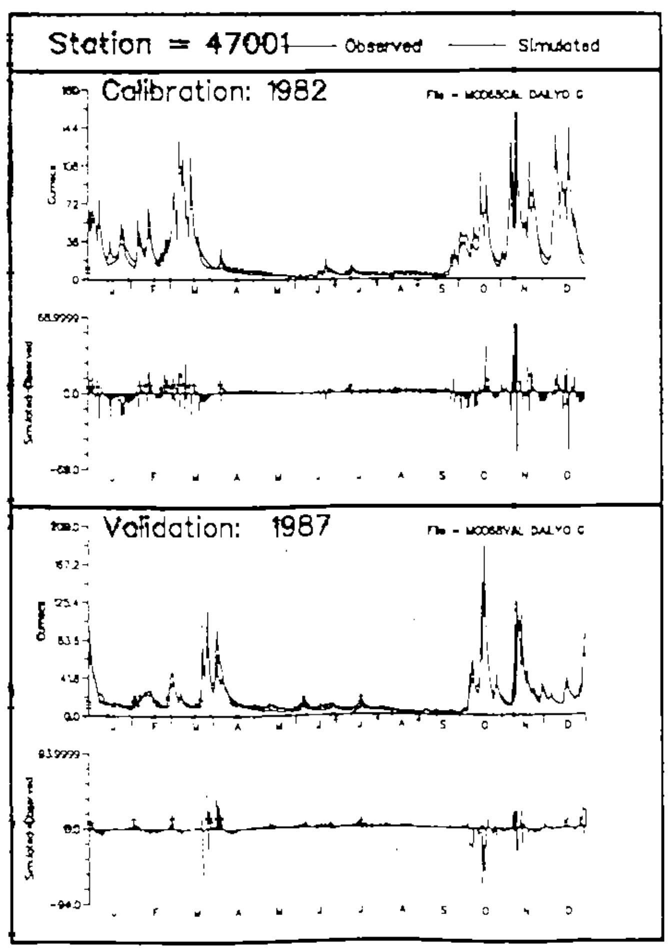
Model B1



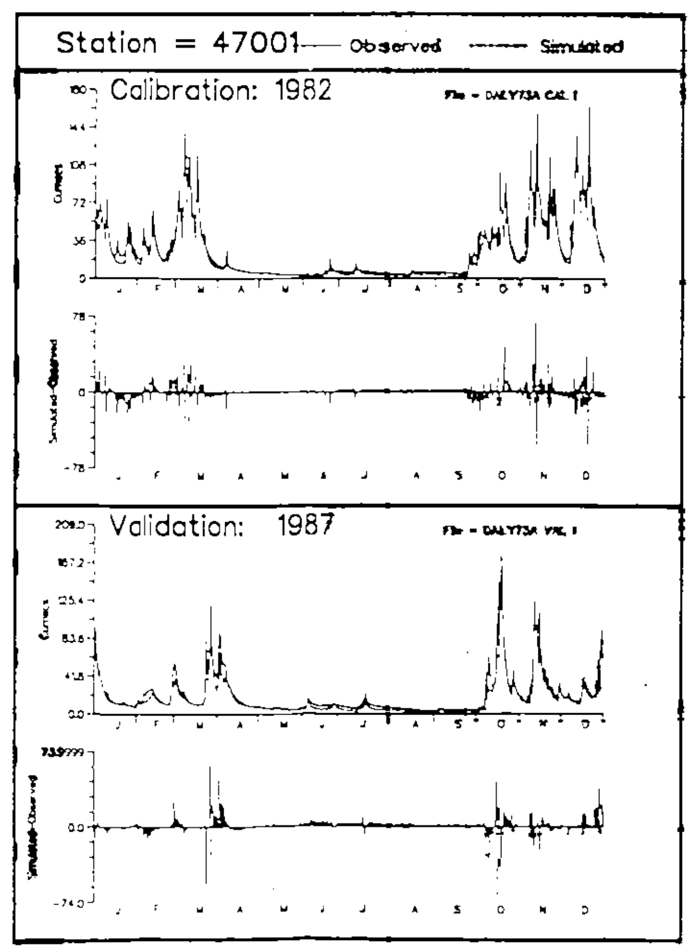
Model B2



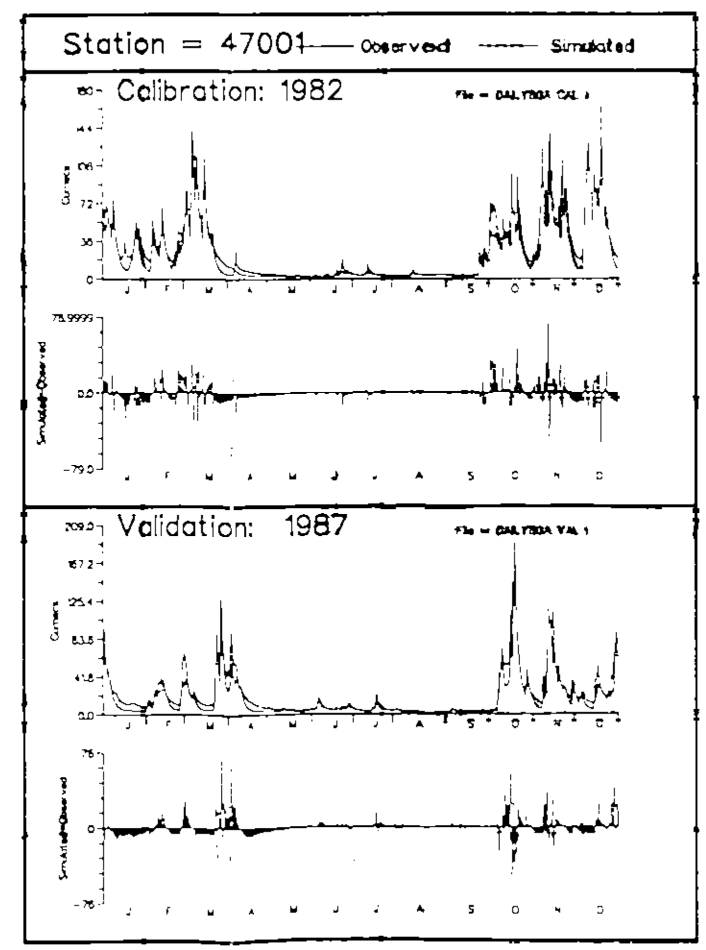
Model C



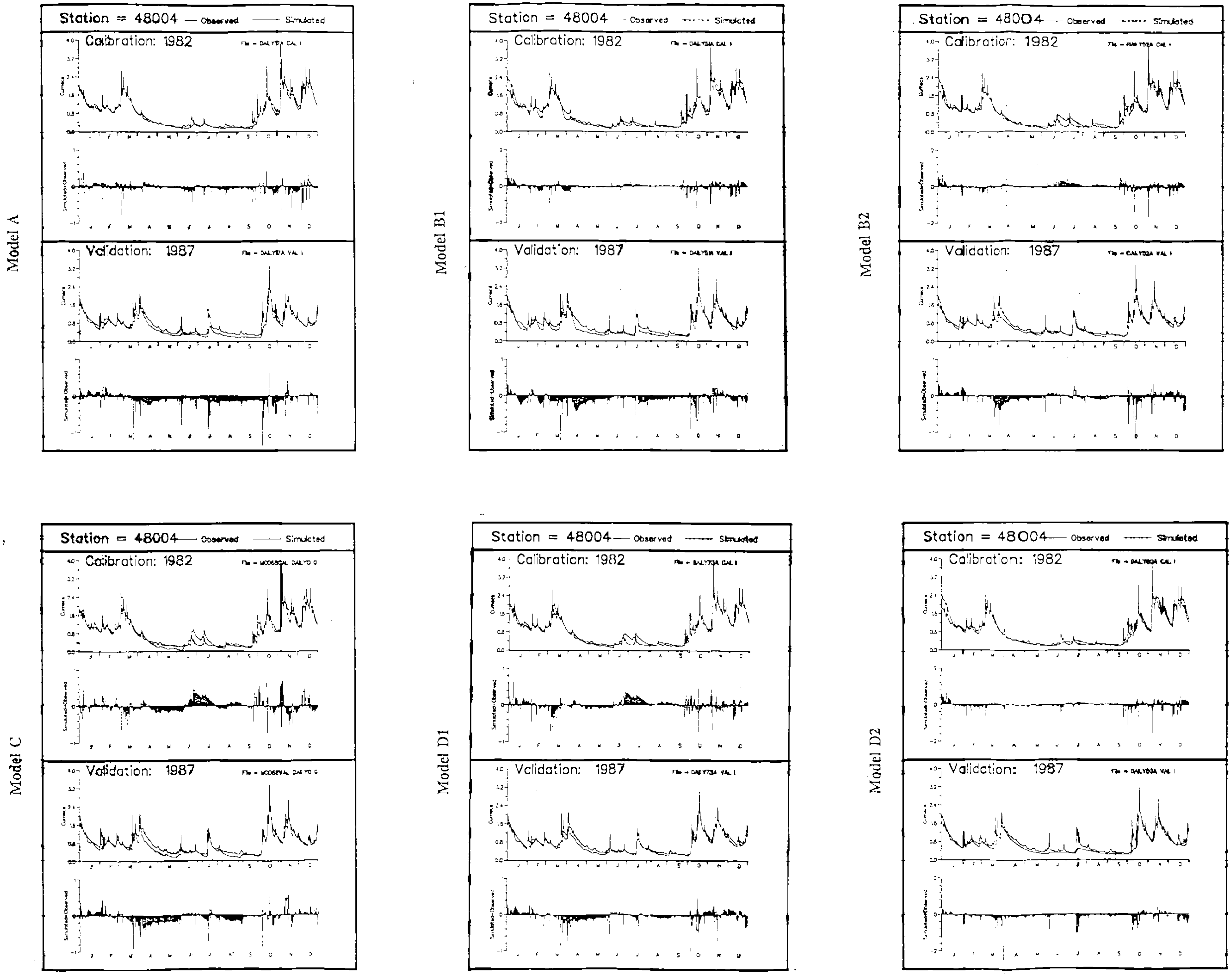
Model D1



Model D2

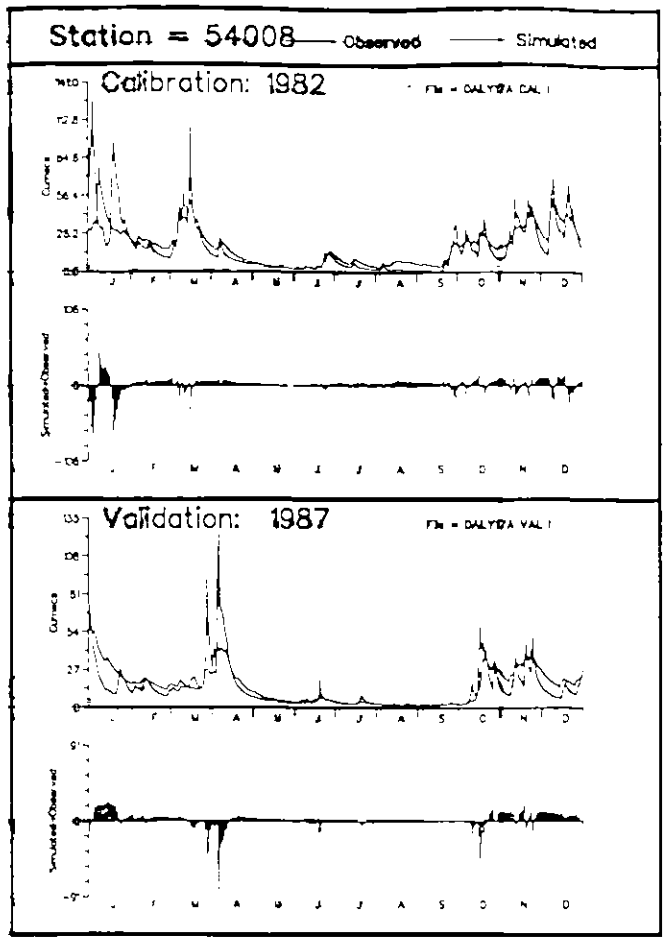


Observed and simulated daily flow regimes for catchment 47001.

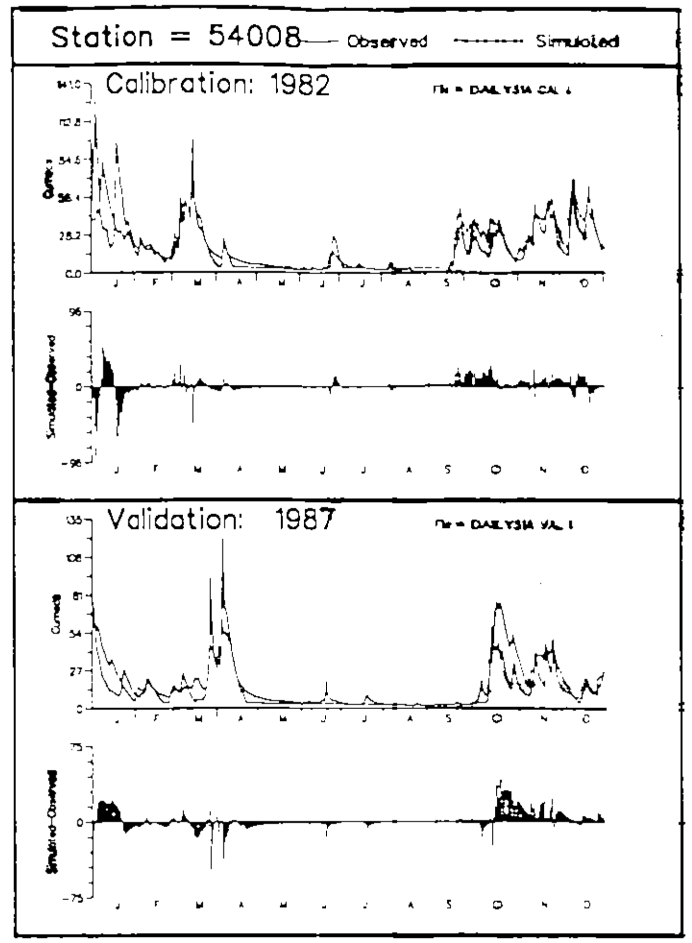


Observed and simulated daily flow regimes for catchment 48004.

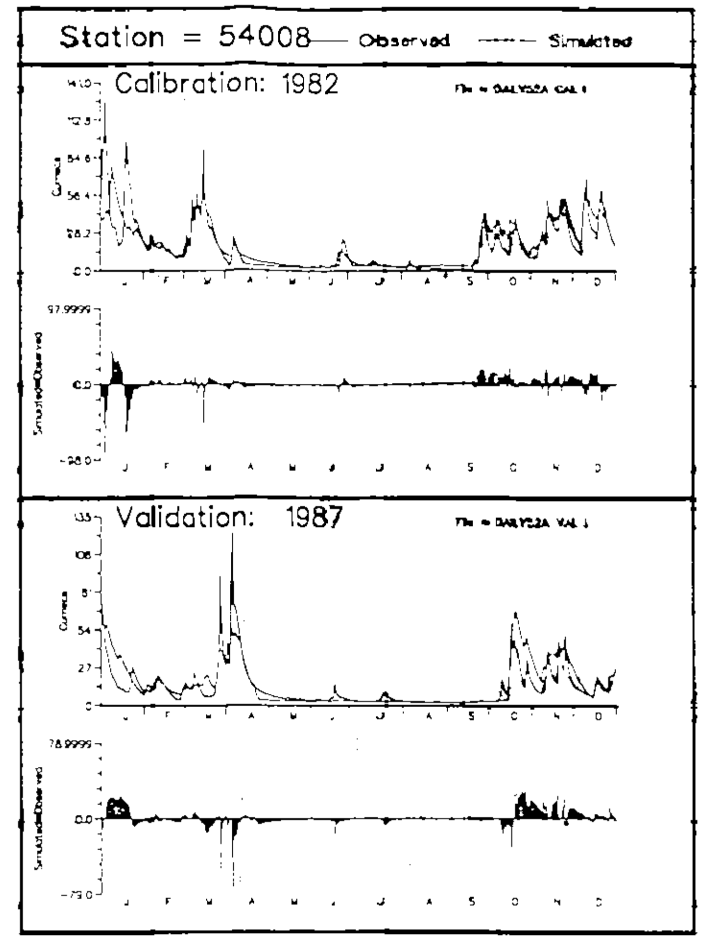
Model A



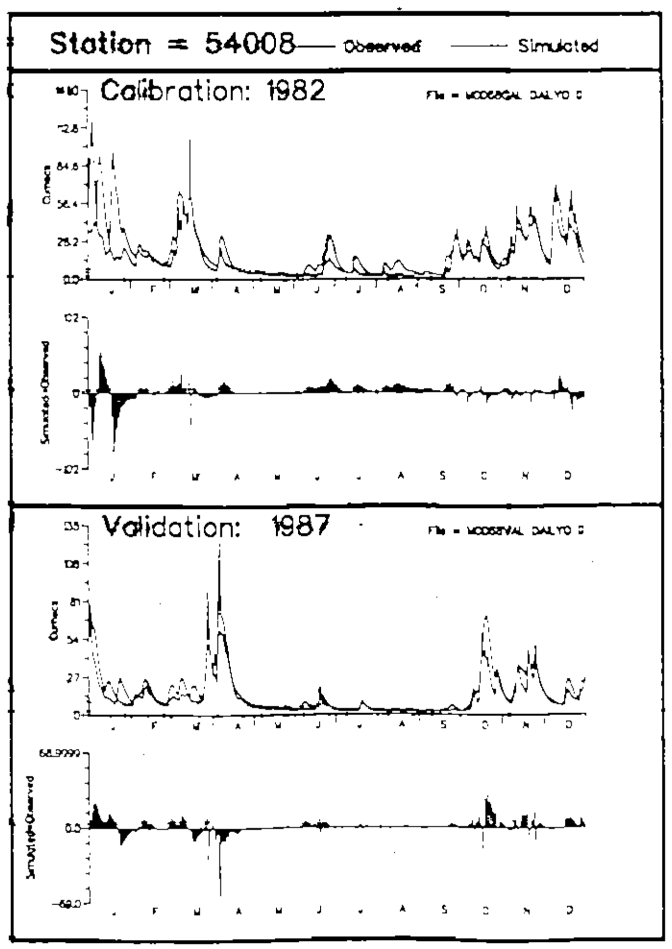
Model B1



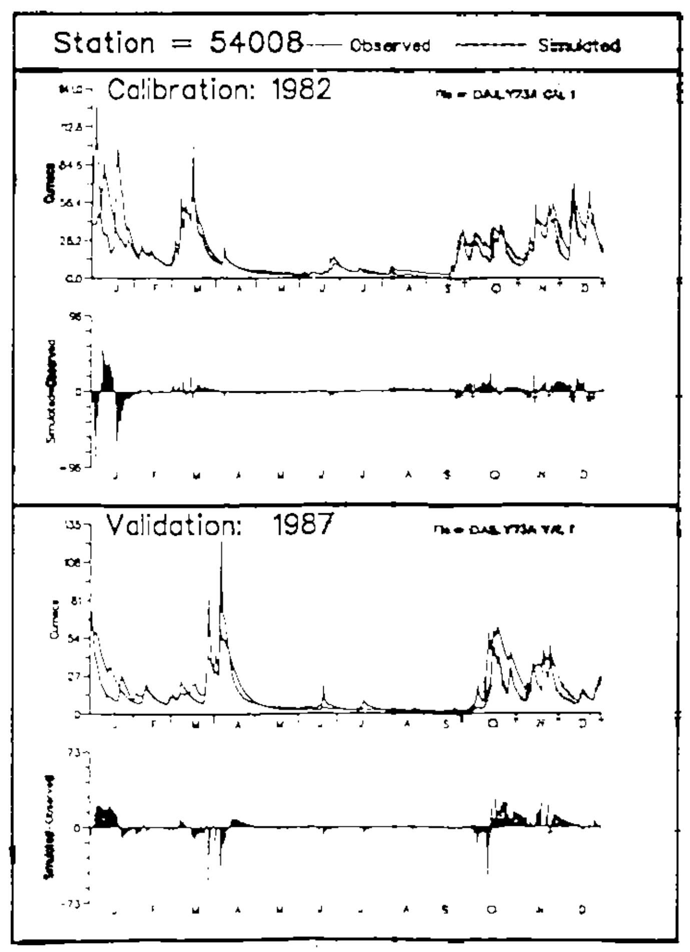
Model B2



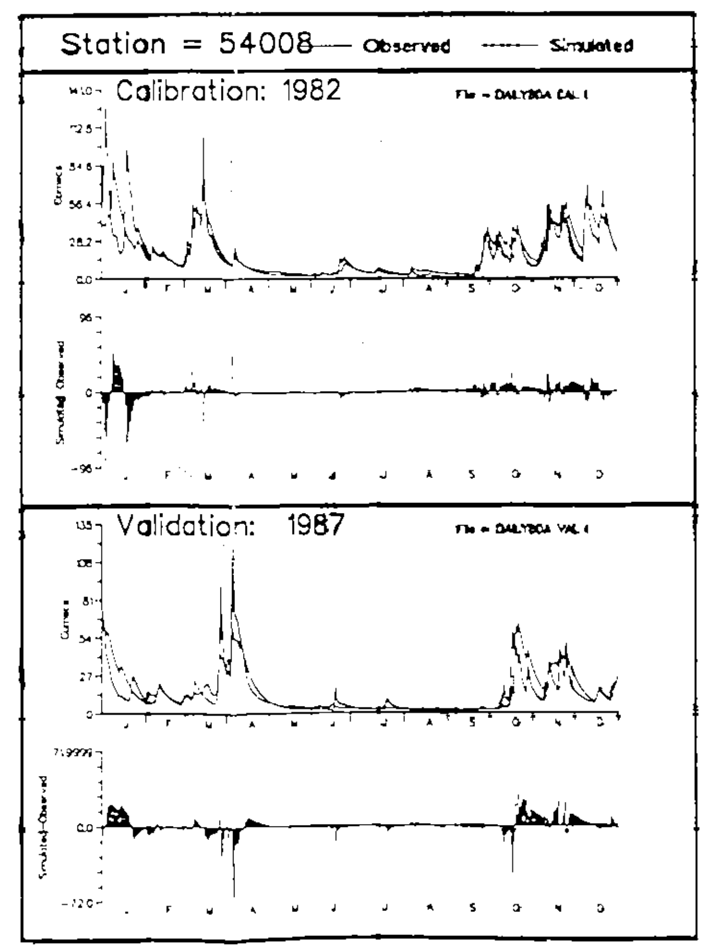
Model C



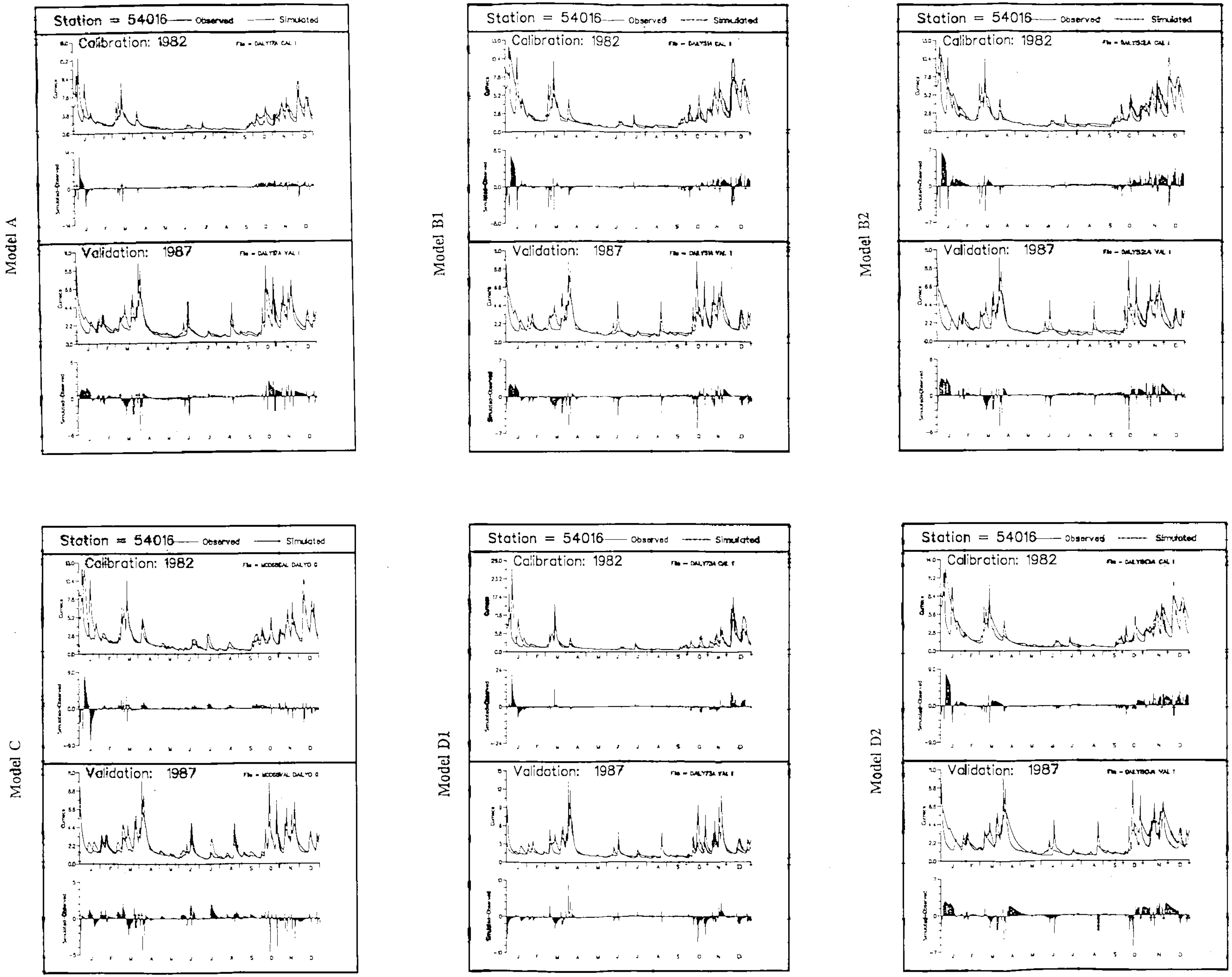
Model D1



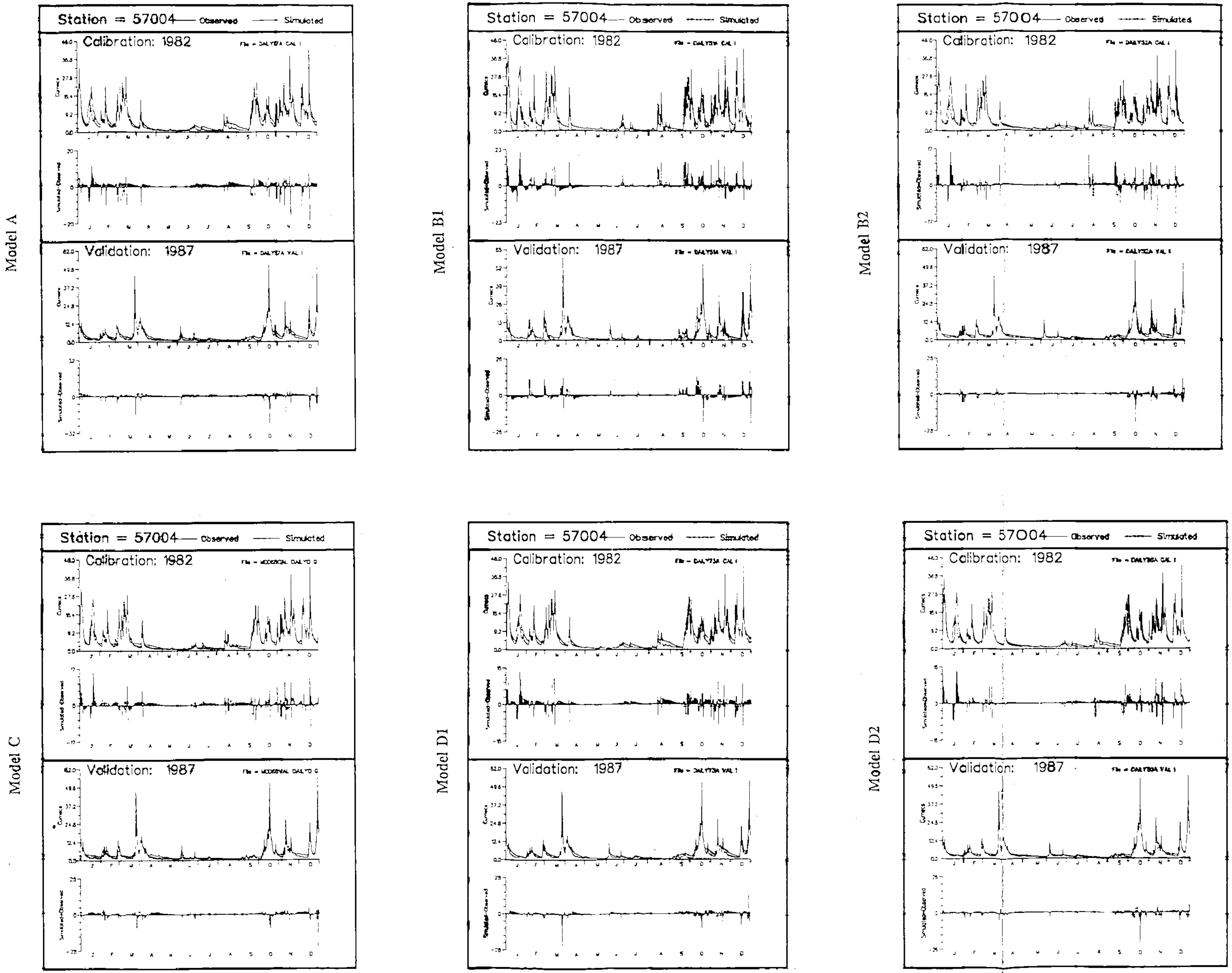
Model D2



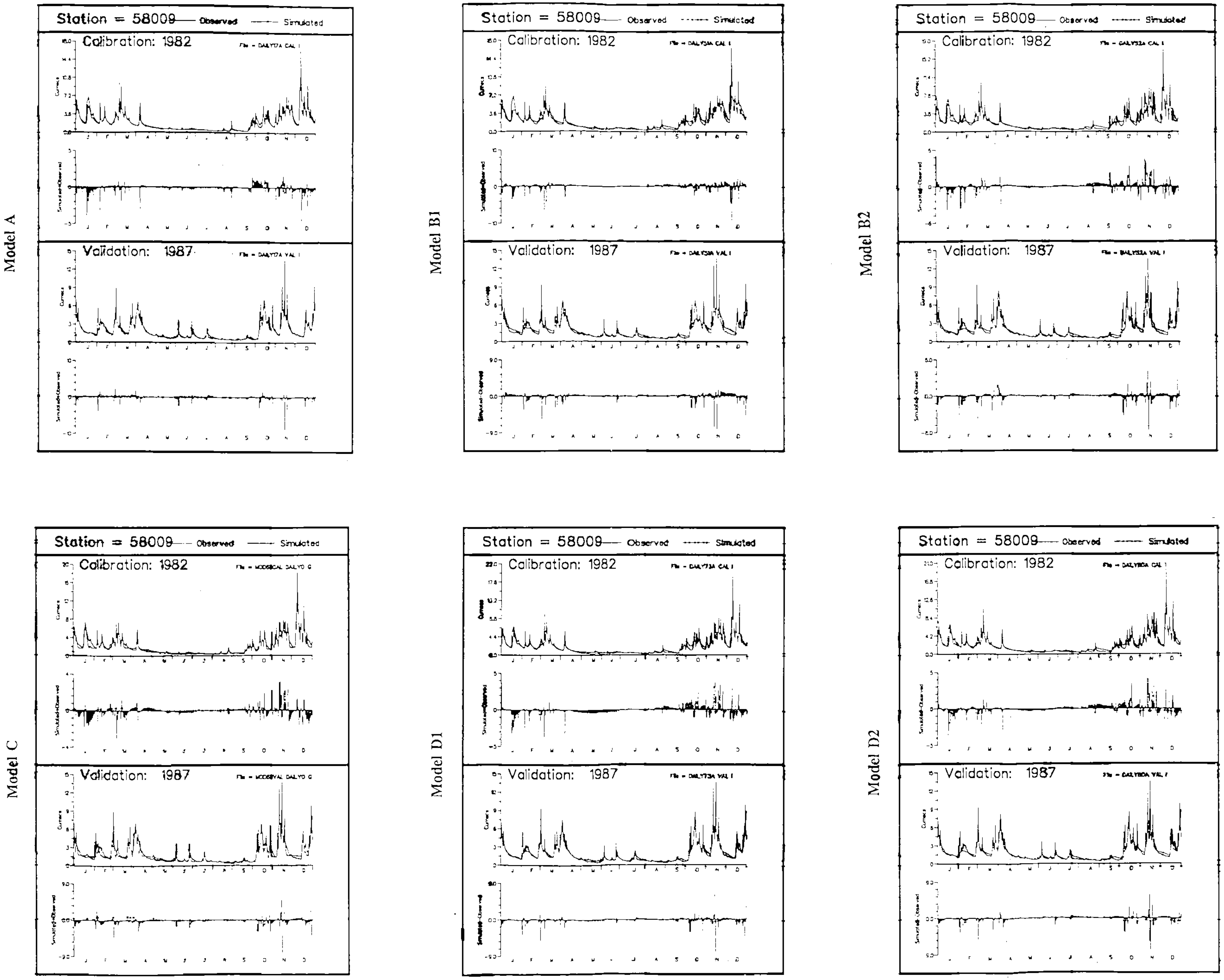
Observed and simulated daily flow regimes for catchment 54008.



Observed and simulated daily flow regimes for catchment 54016.

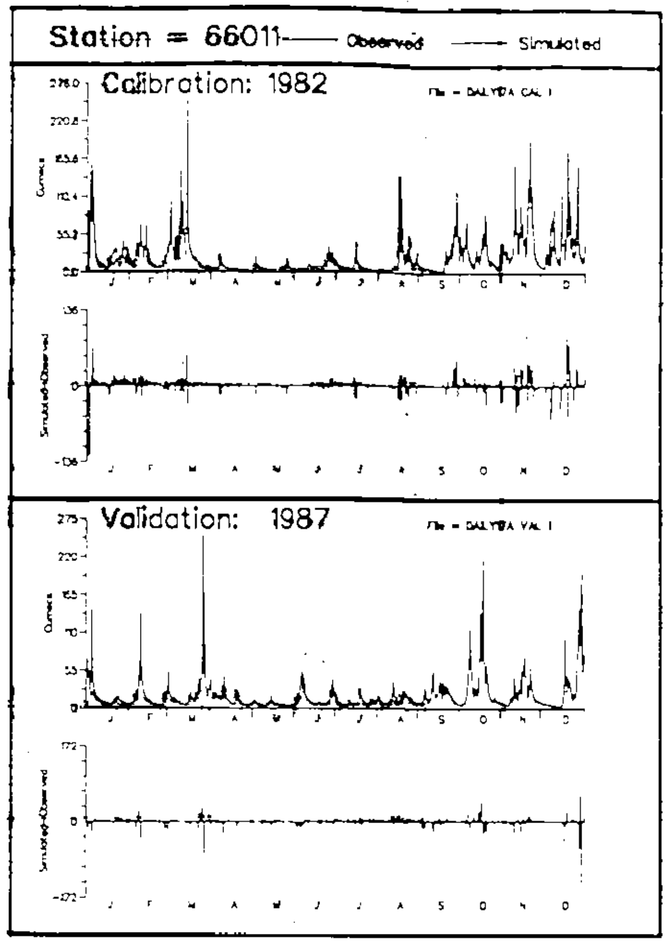


Observed and simulated daily flow regimes for catchment 57004.

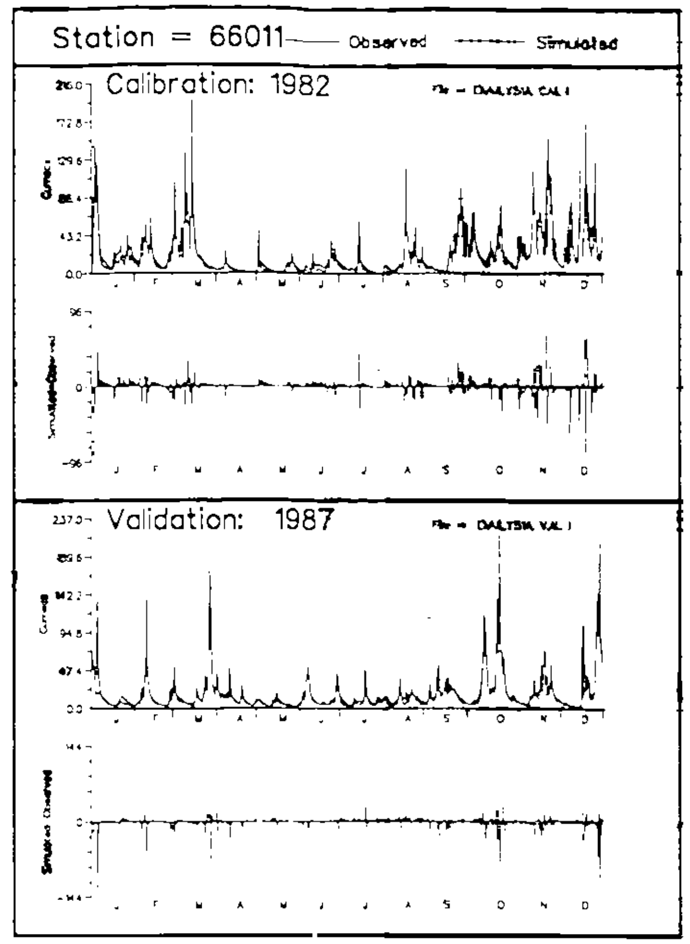


Observed and simulated daily flow regimes for catchment 58009.

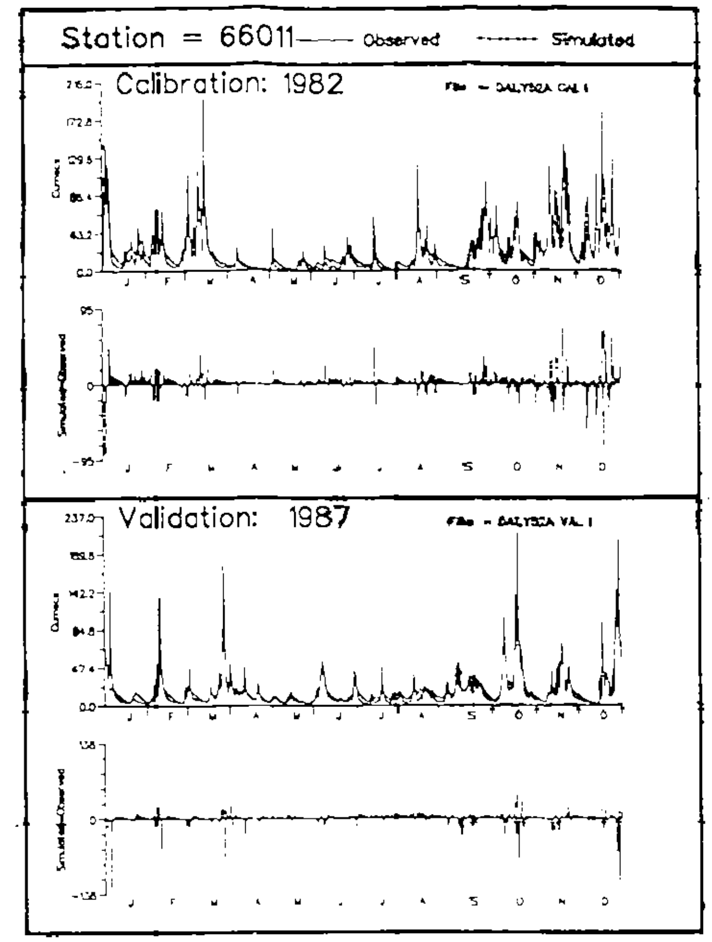
Model A



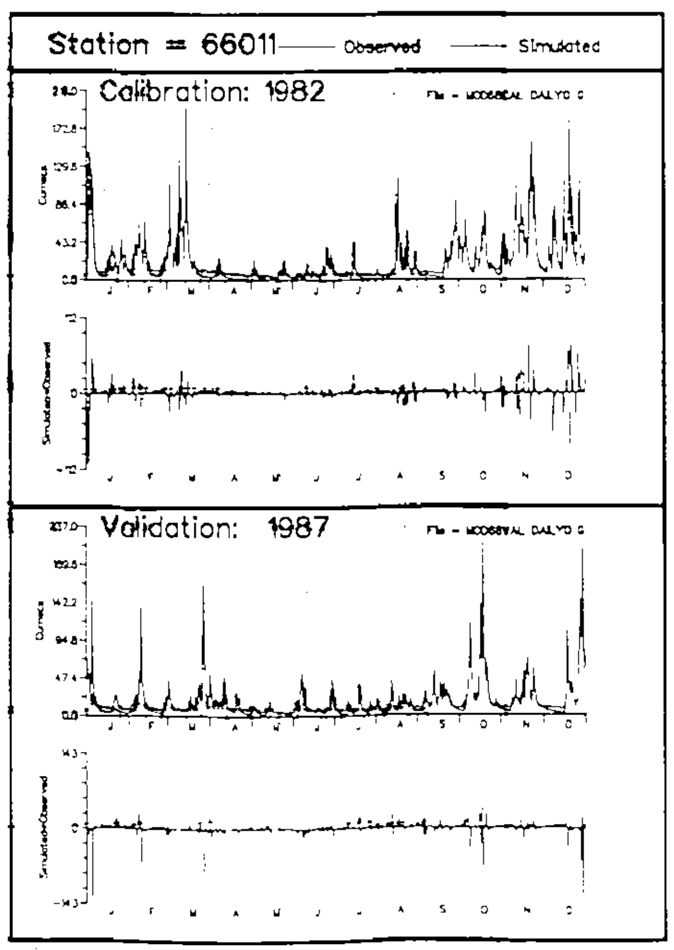
Model B1



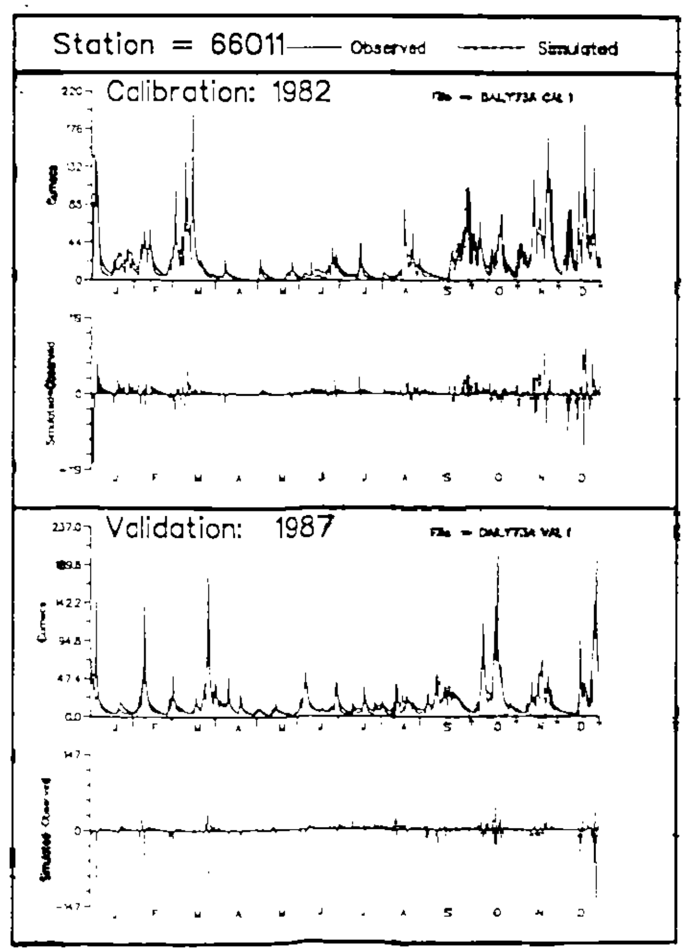
Model B2



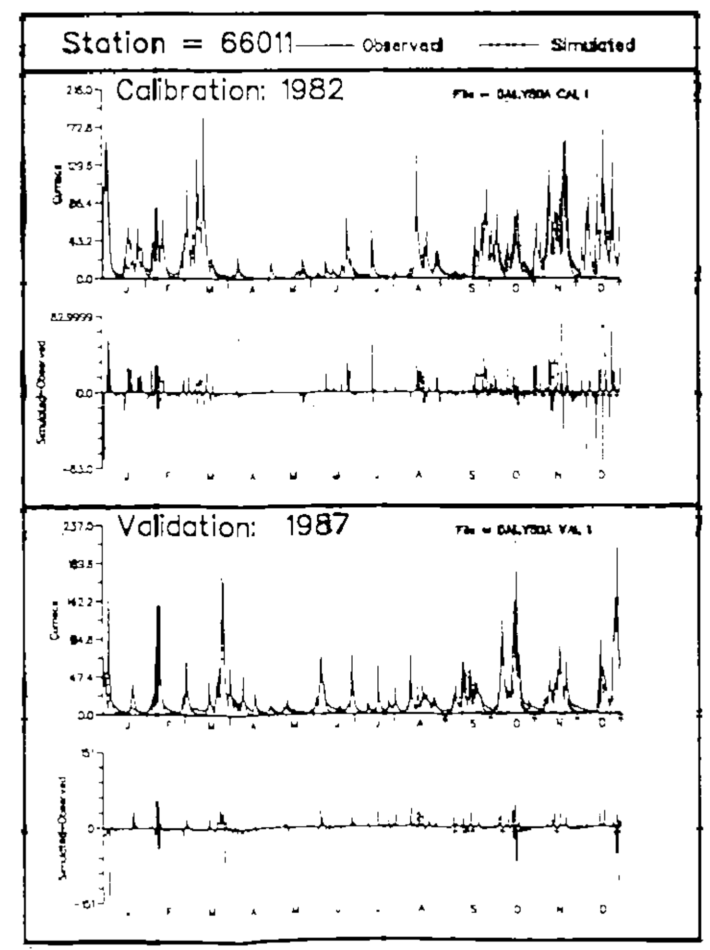
Model C



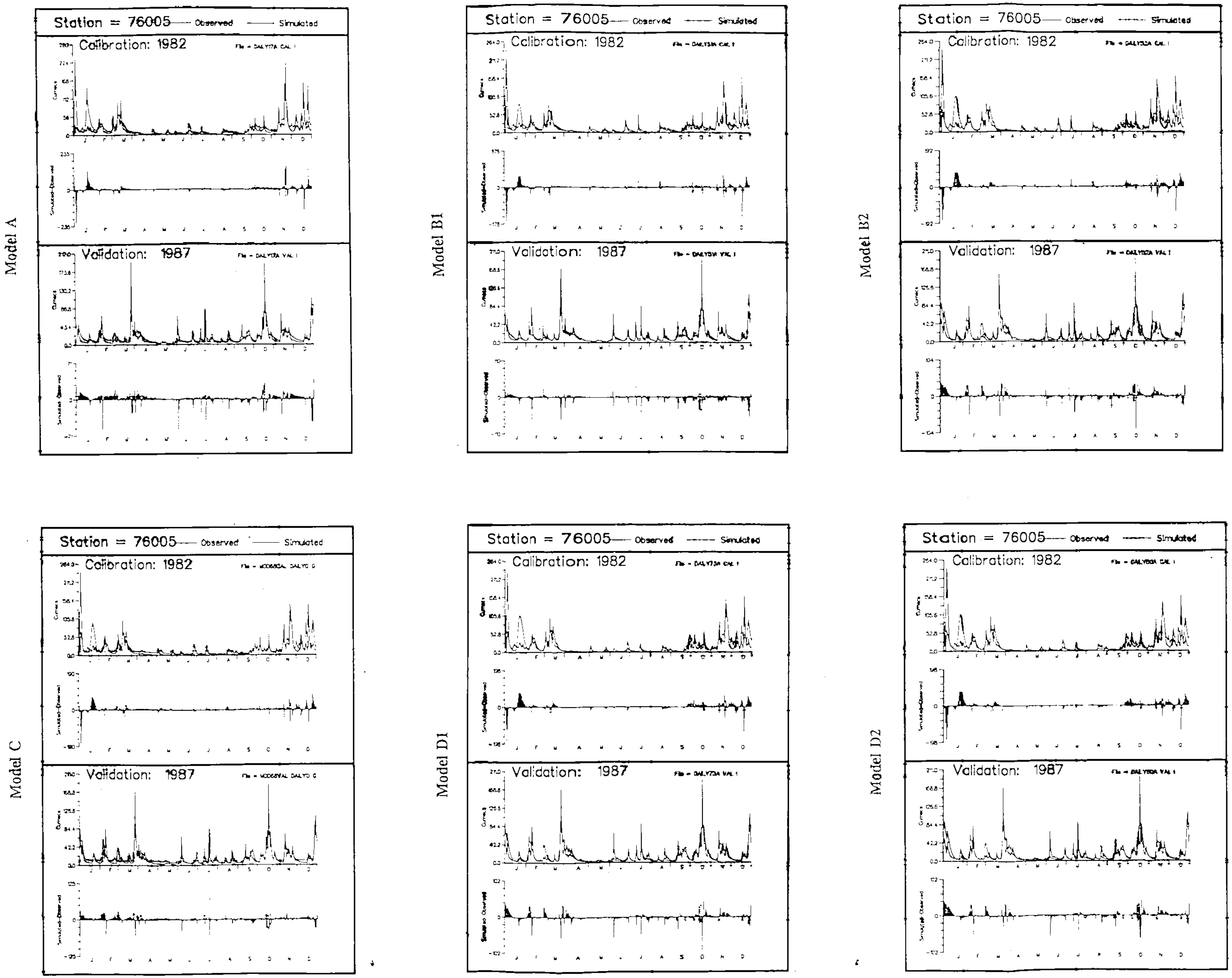
Model D1



Model D2

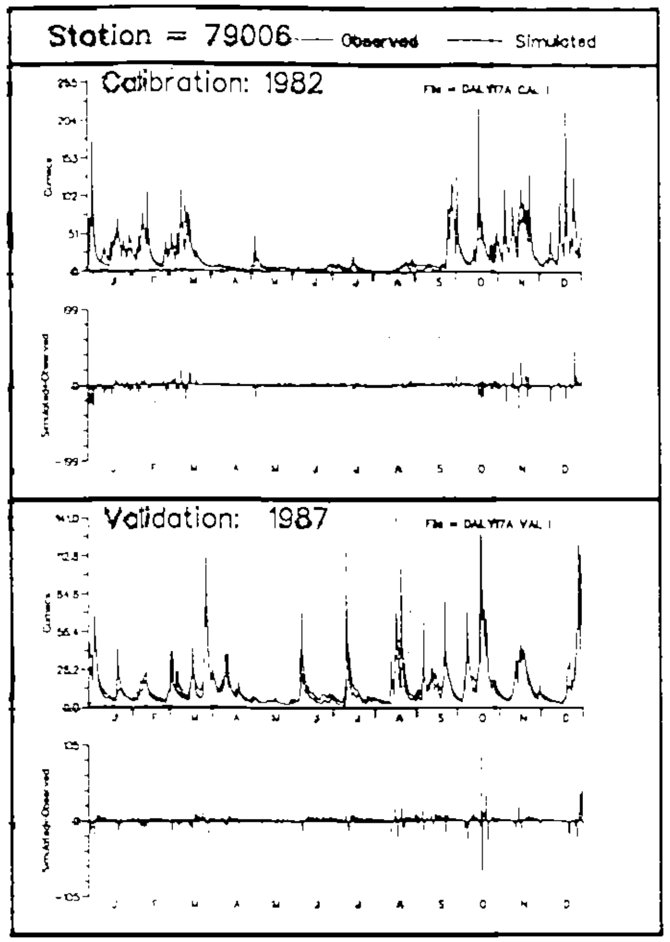


Observed and simulated daily flow regimes for catchment 66011.

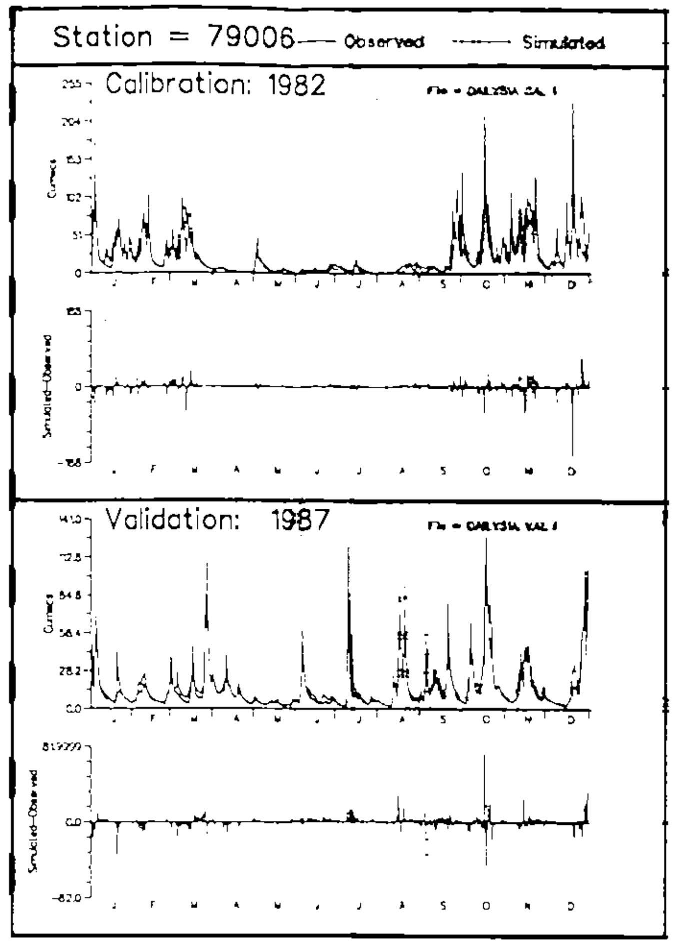


Observed and simulated daily flow regimes for catchment 76005.

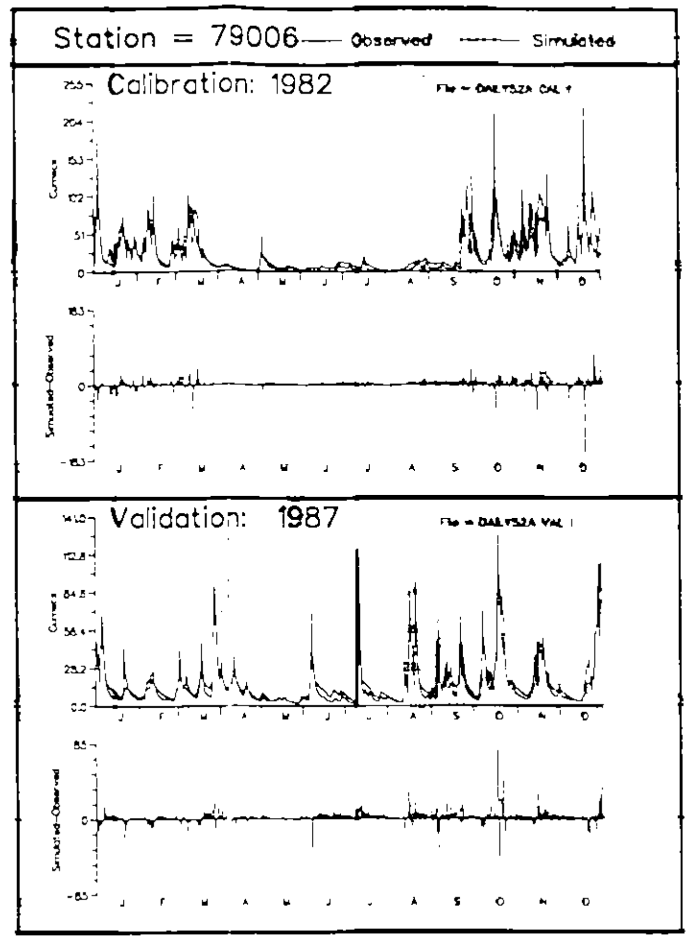
Model A



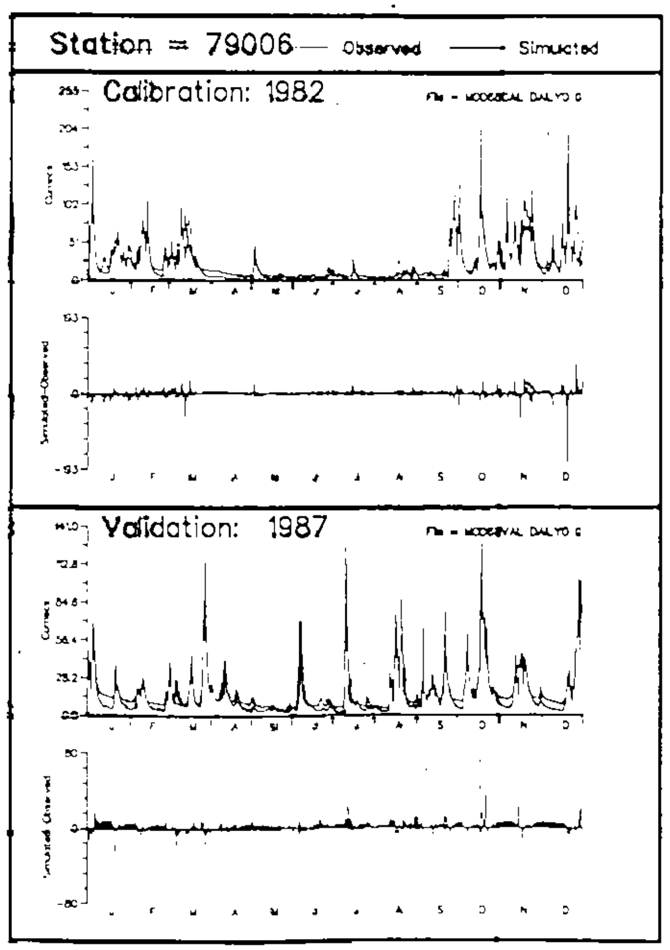
Model B1



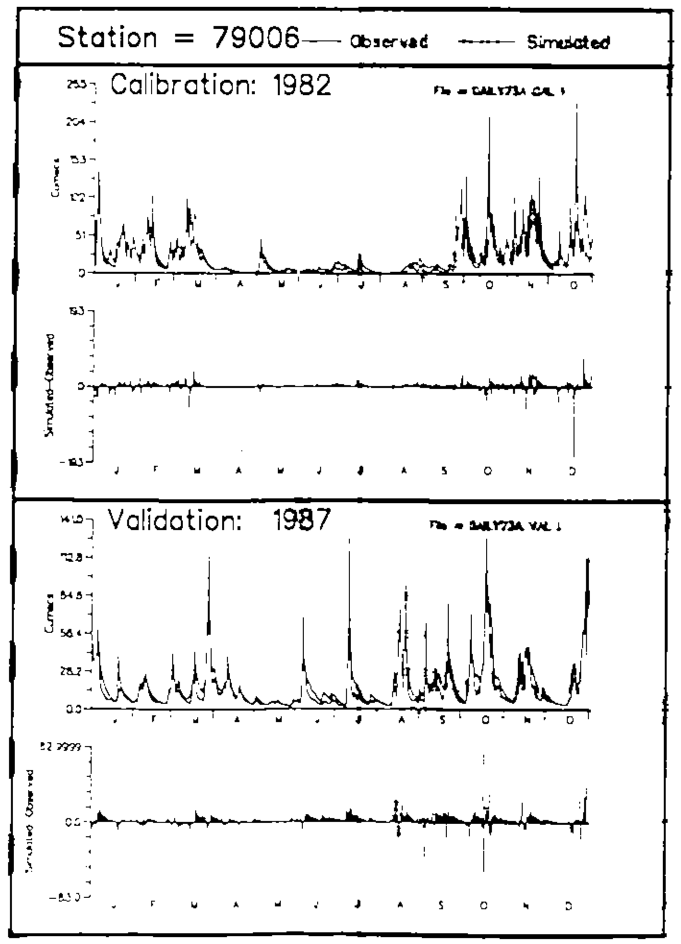
Model B2



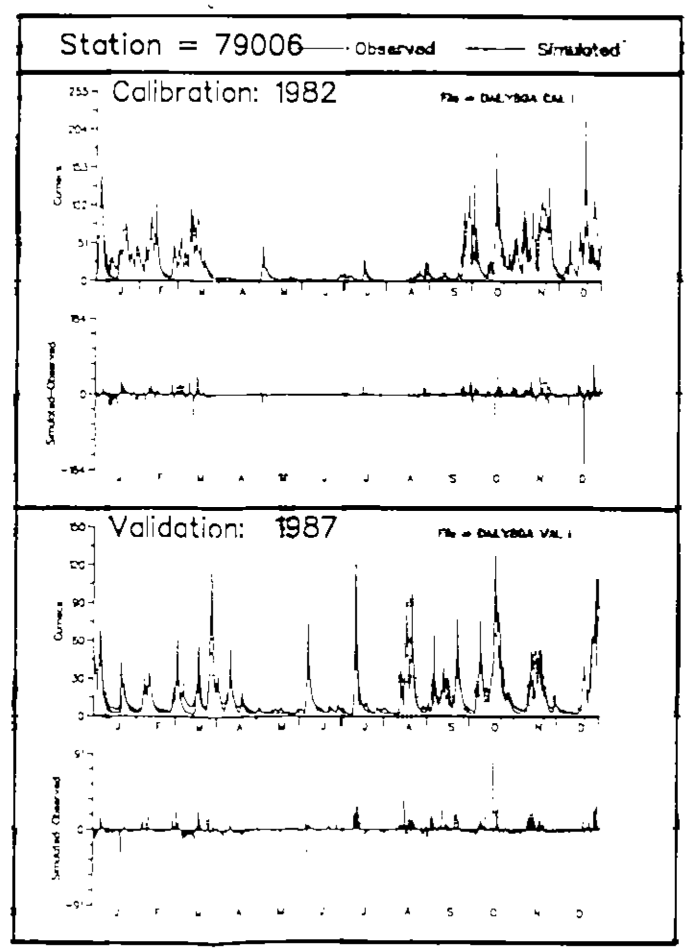
Model C



Model D1



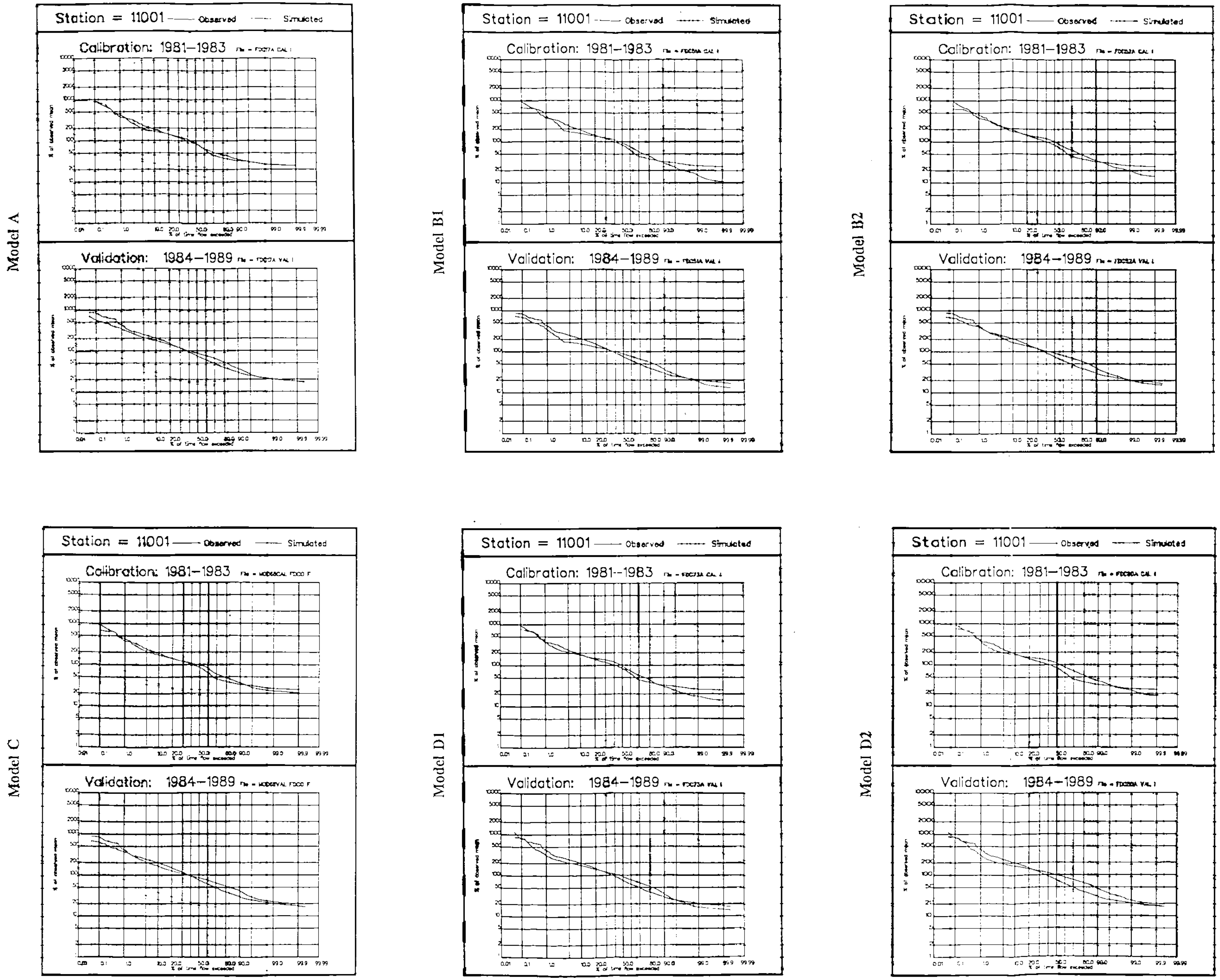
Model D2



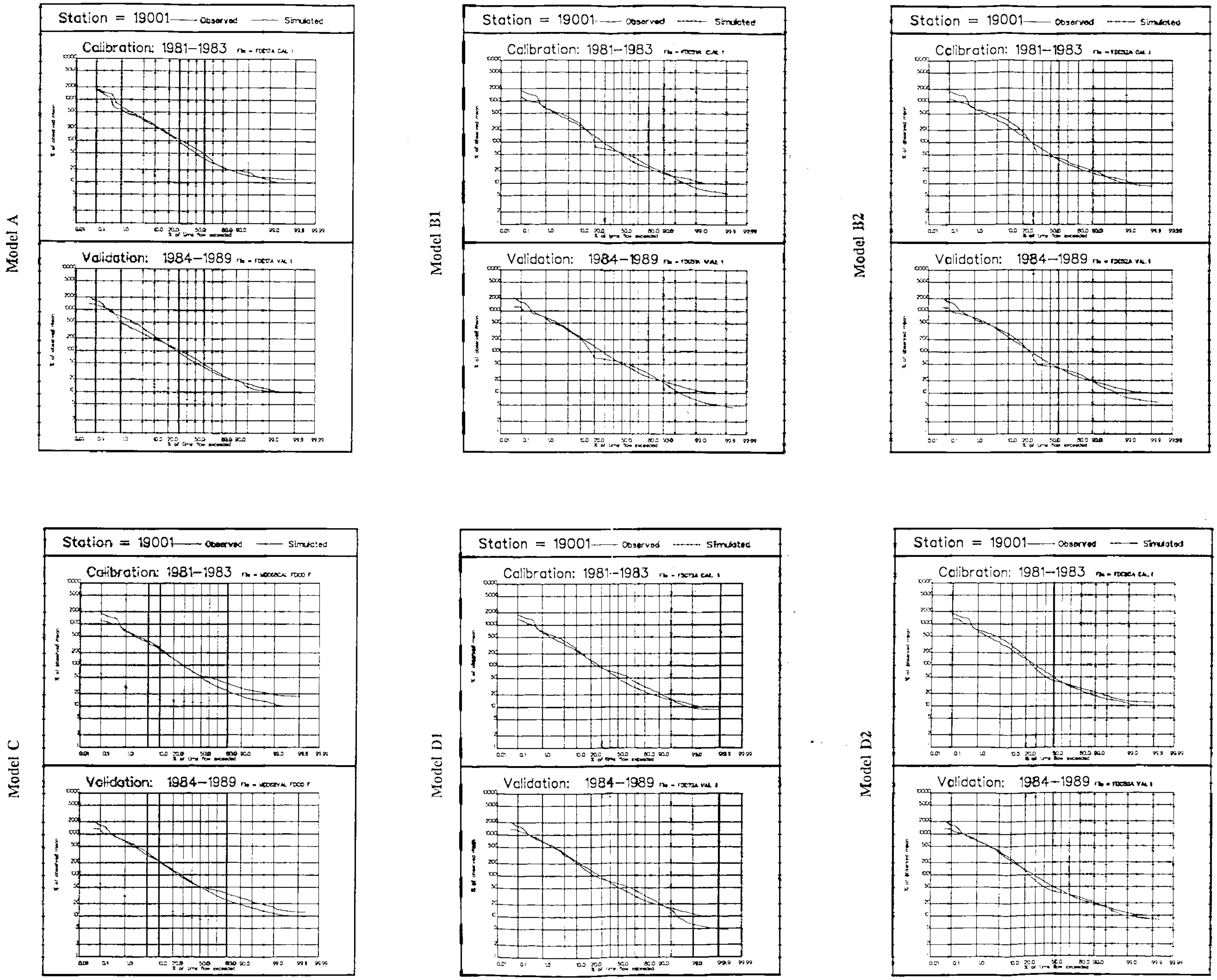
Observed and simulated daily flow regimes for catchment 79006.

Appendix A

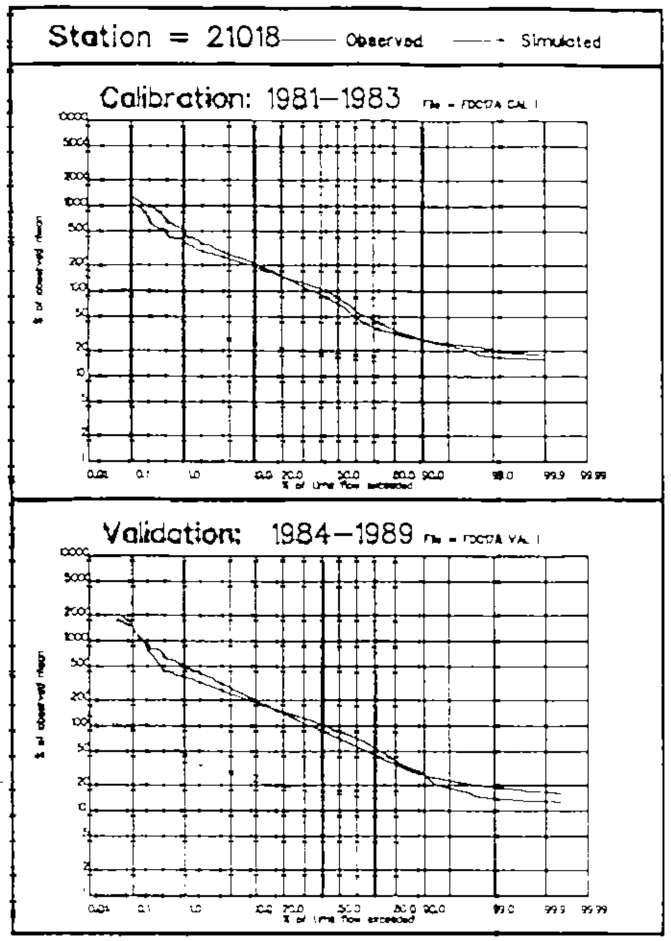
A.6 Plots of observed and simulated flow duration curves



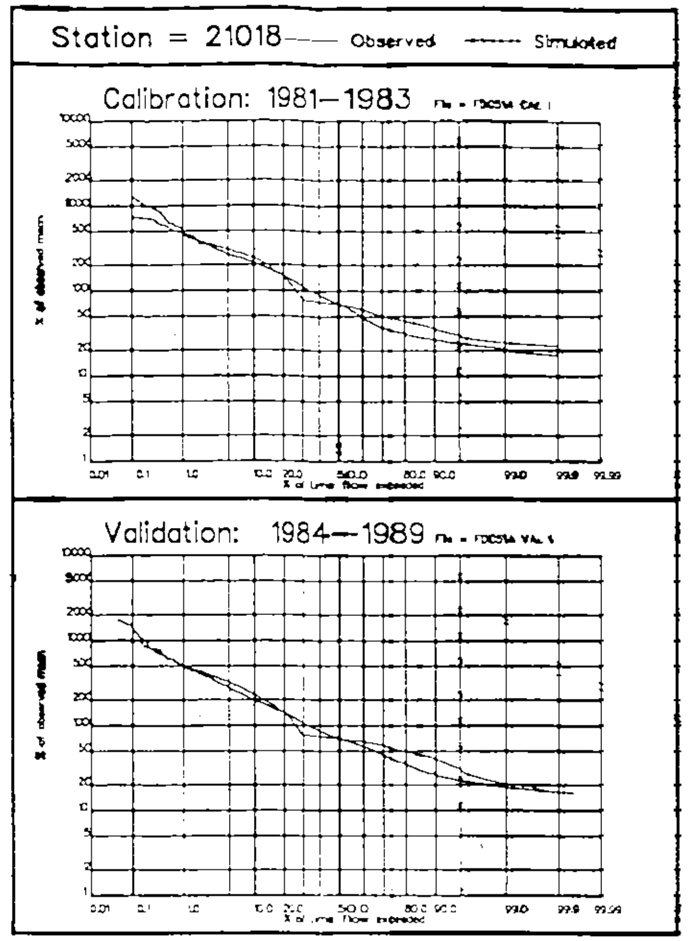
Observed and simulated flow duration curves for catchment 11001.



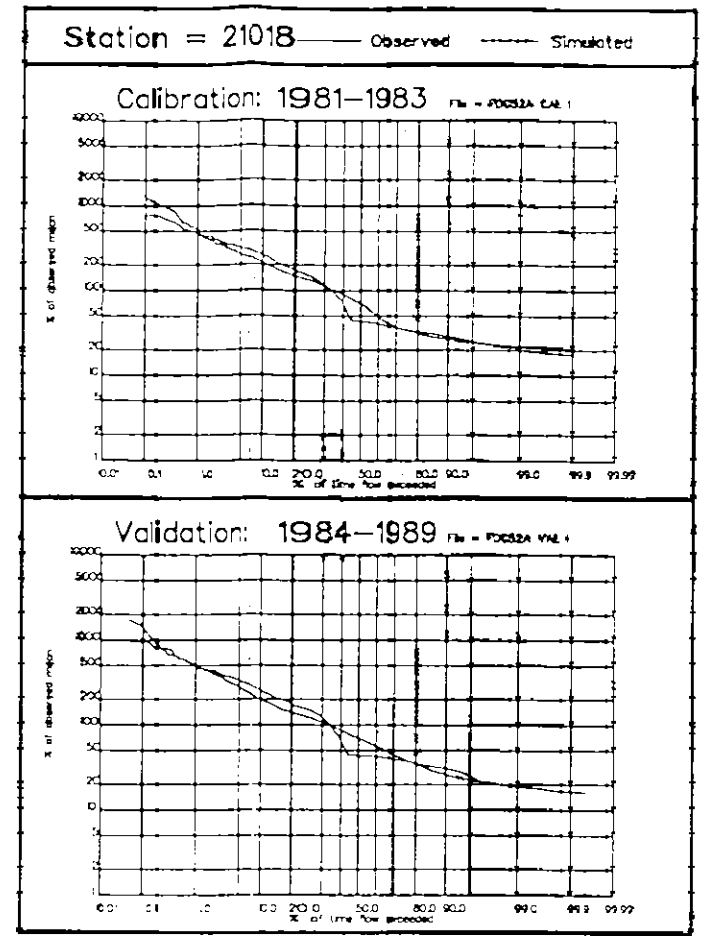
Observed and simulated flow duration curves for catchment 19001.



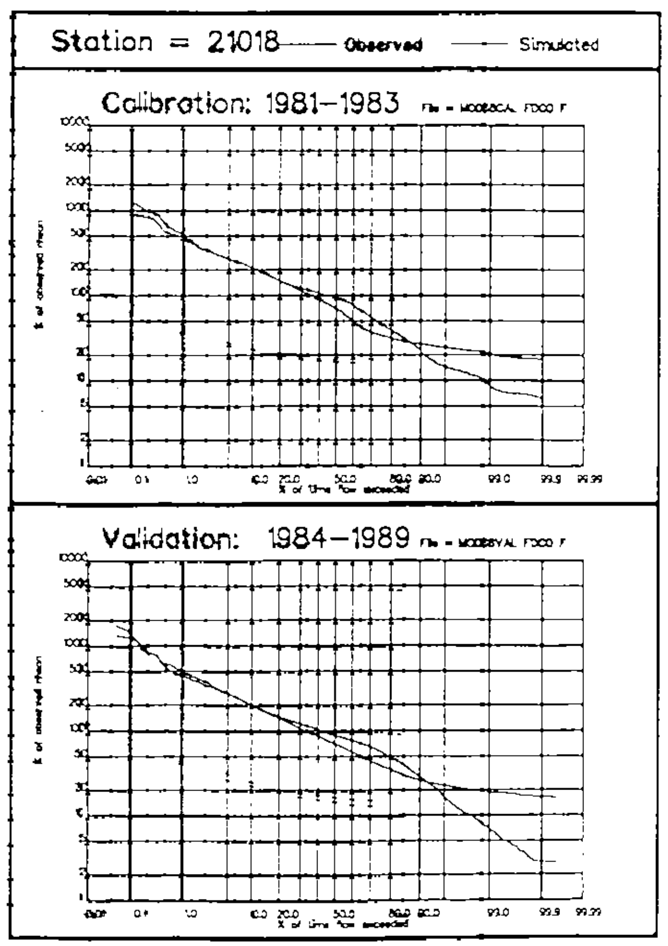
Model A



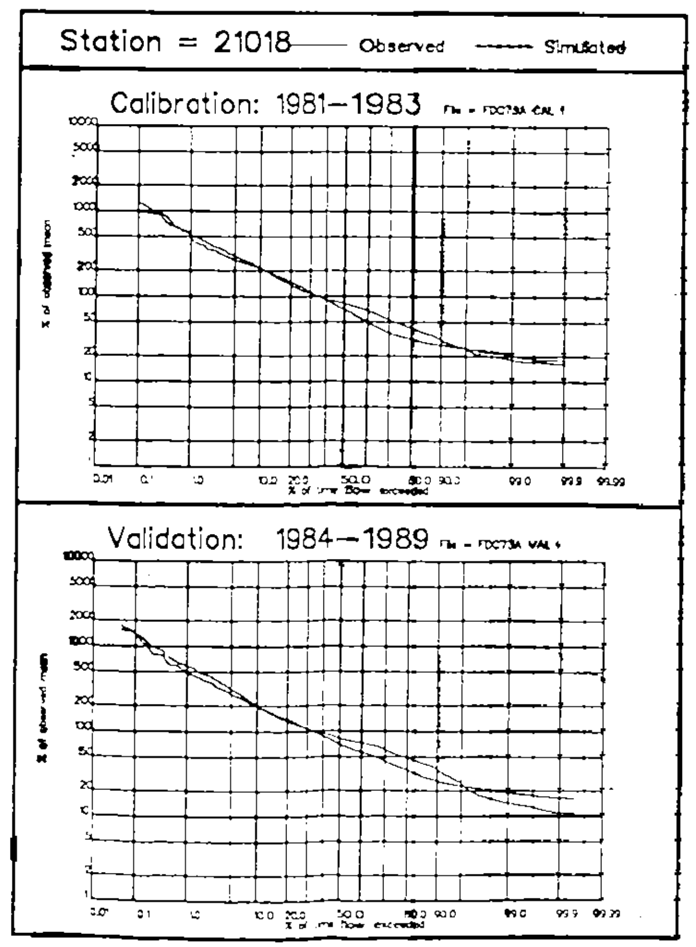
Model B1



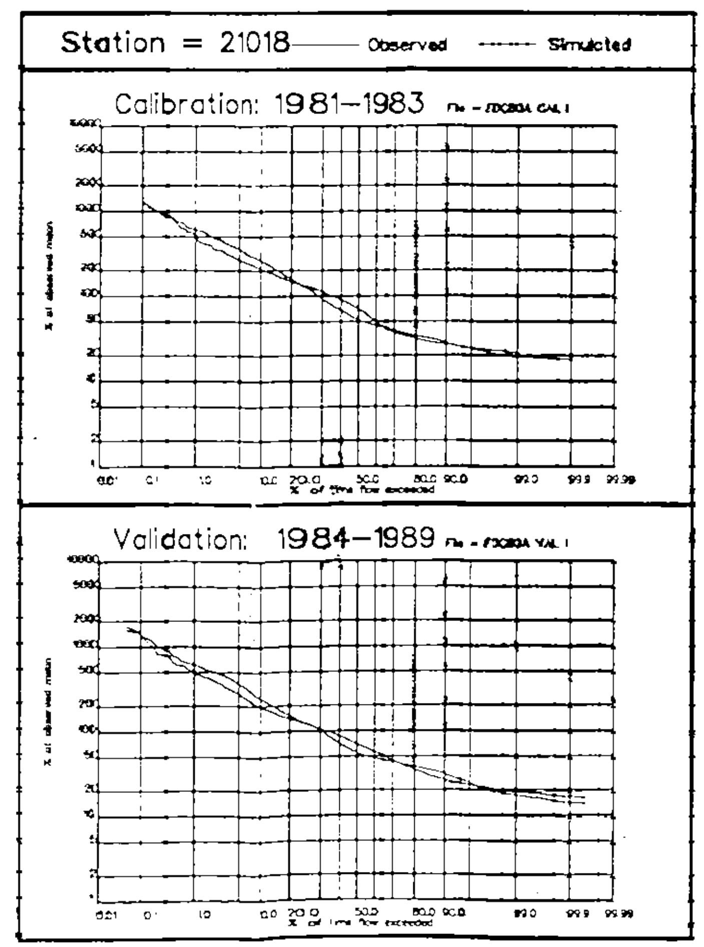
Model B2



Model C

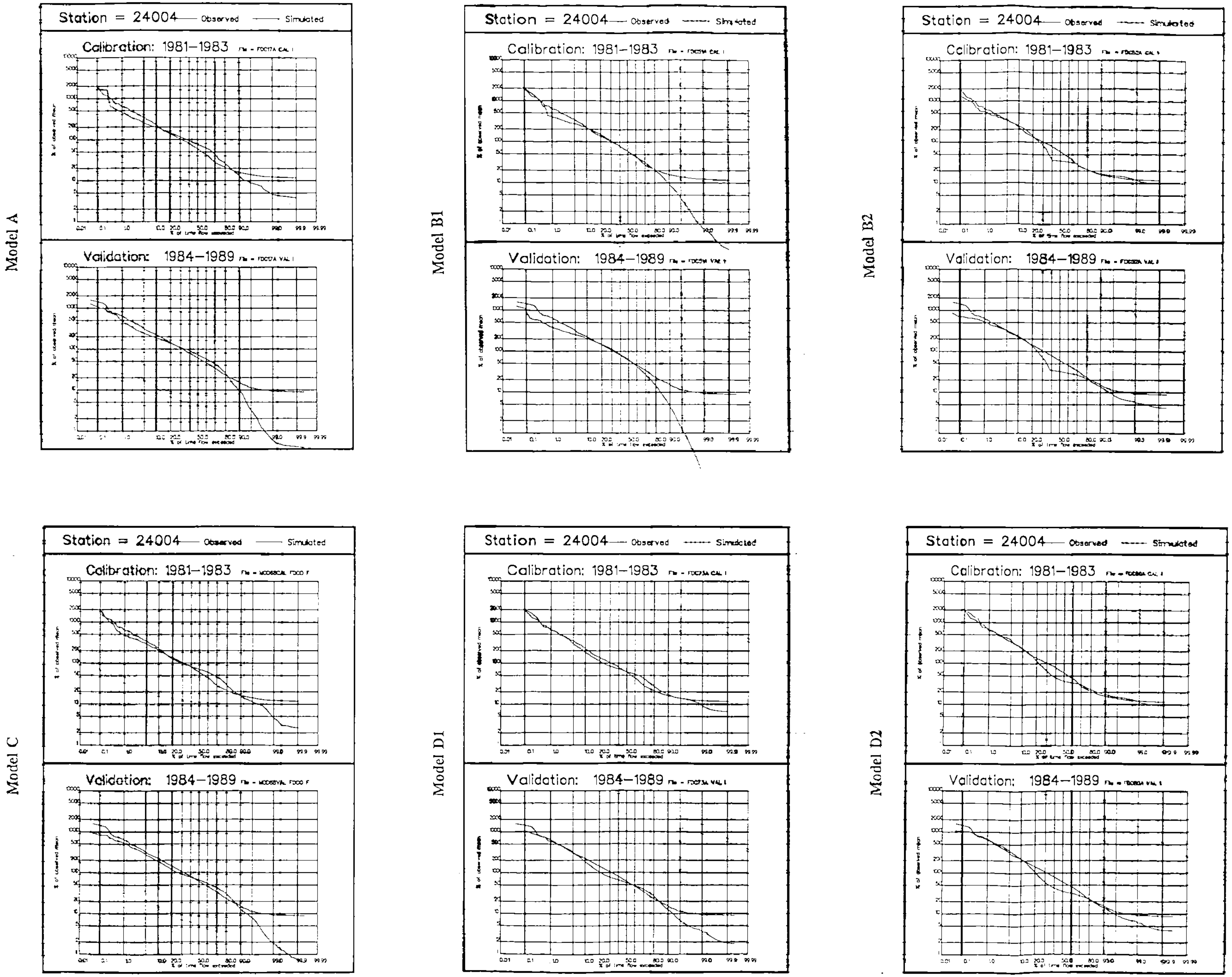


Model D1



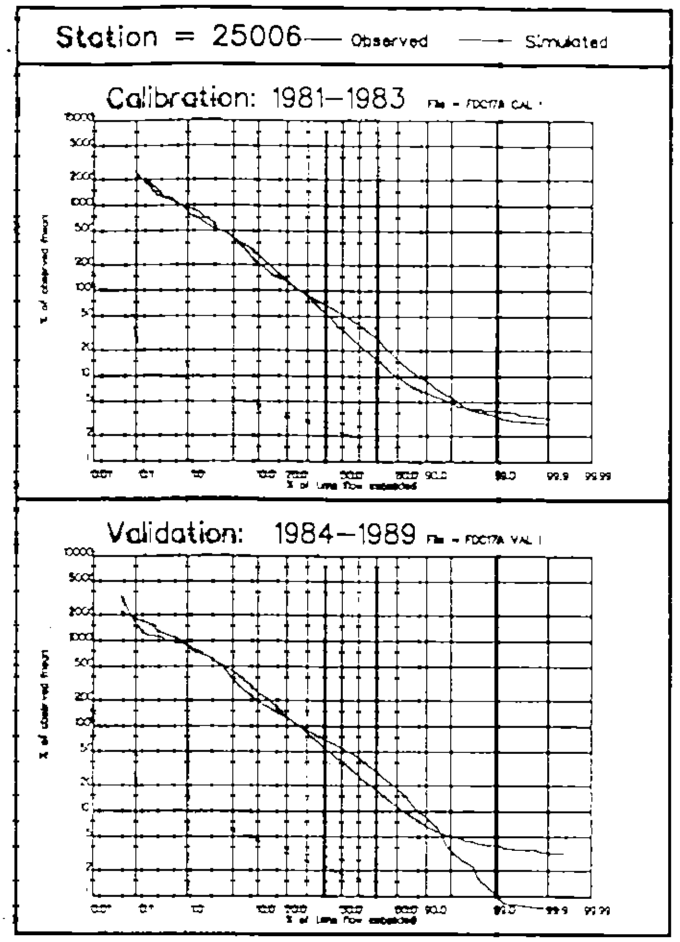
Model D2

Observed and simulated flow duration curves for catchment 21018.

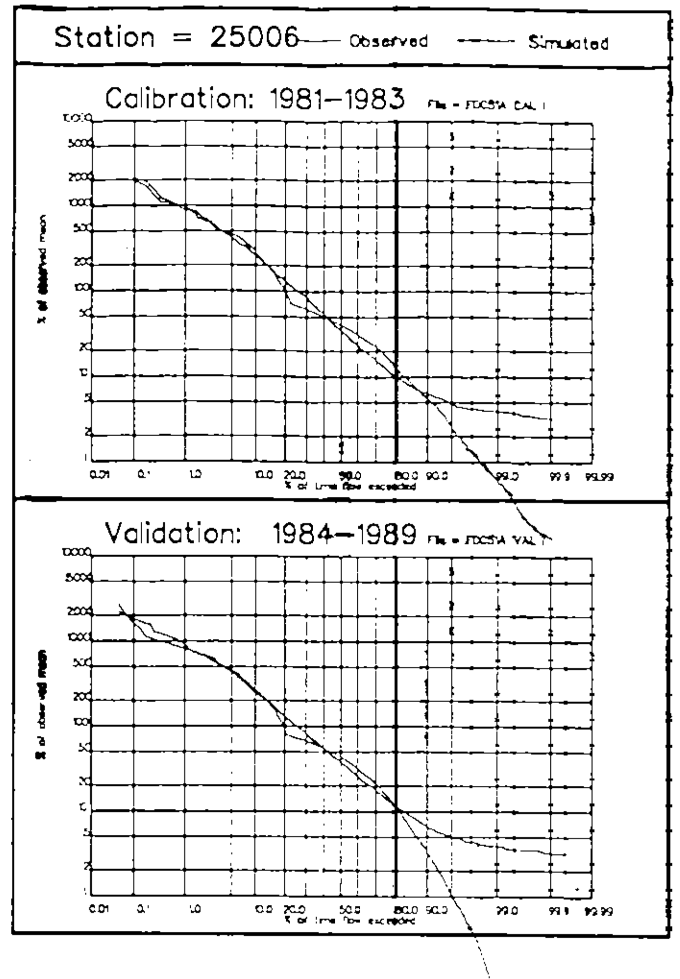


Observed and simulated flow duration curves for catchment 24004.

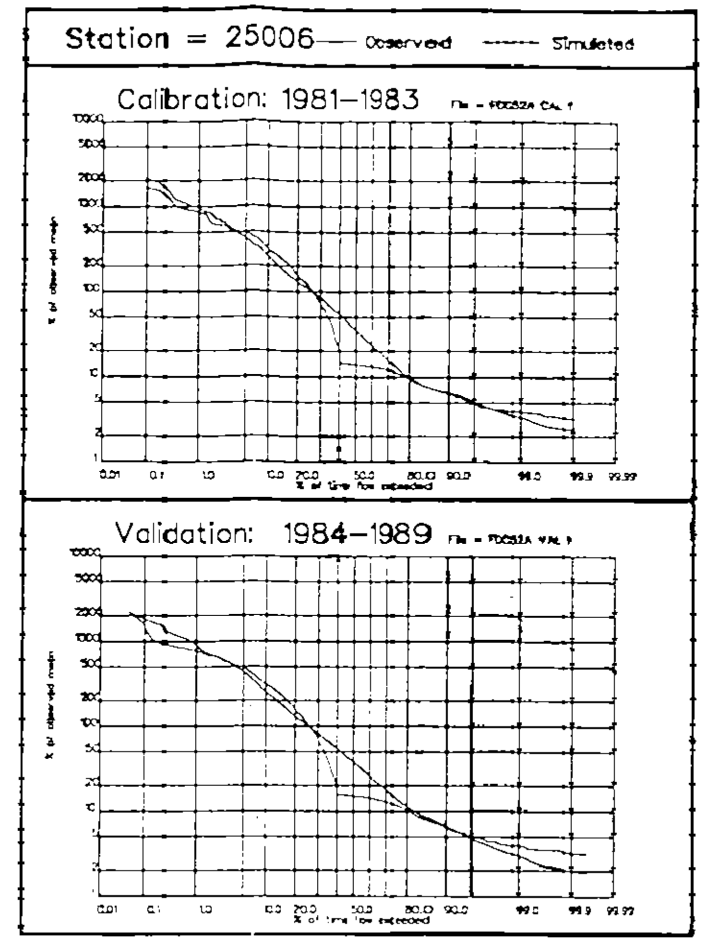
Model A



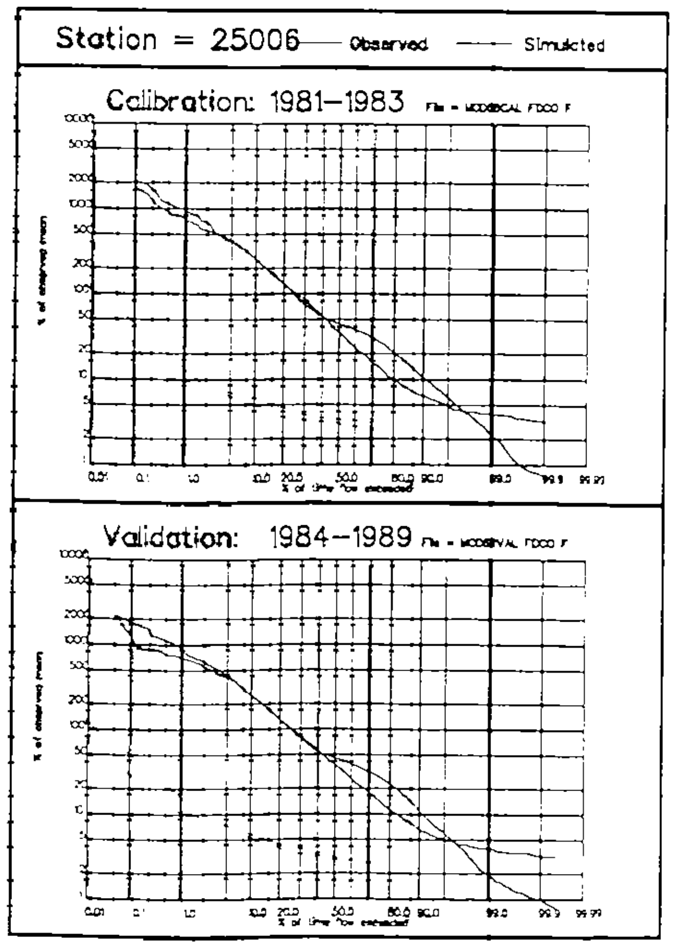
Model B1



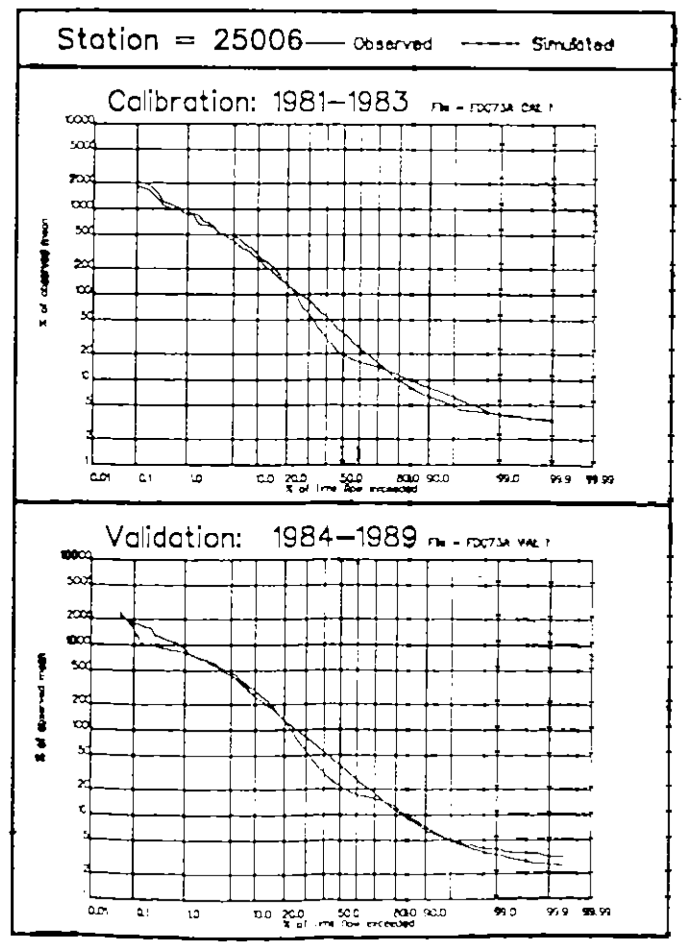
Model B2



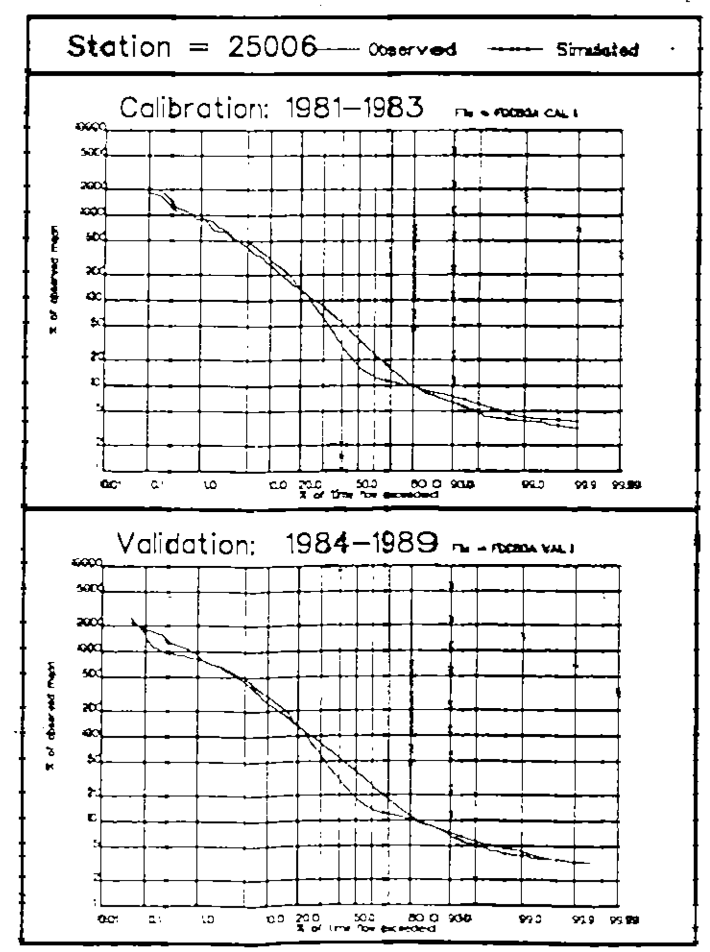
Model C



Model D1

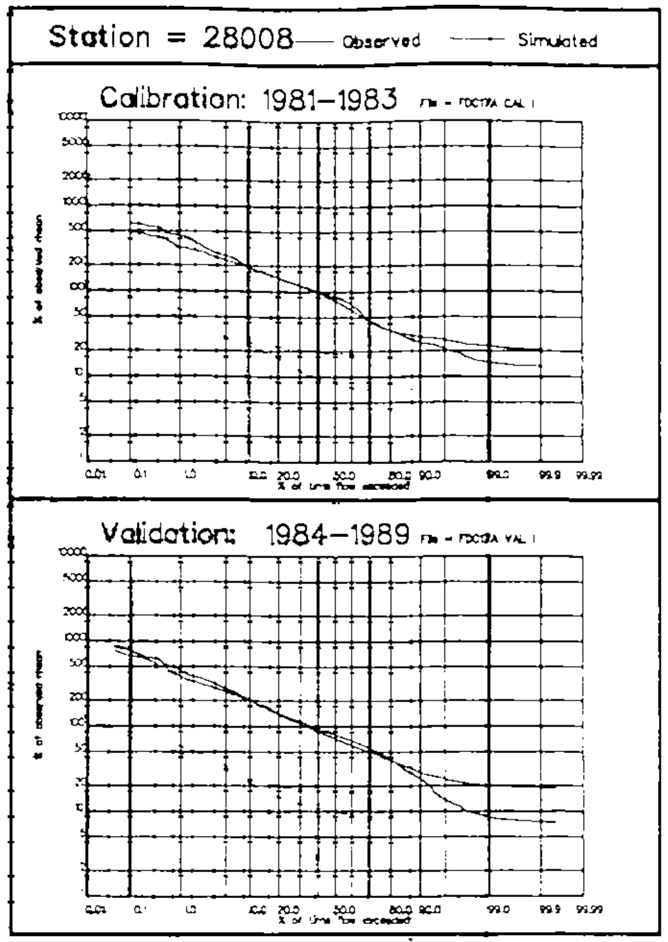


Model D2

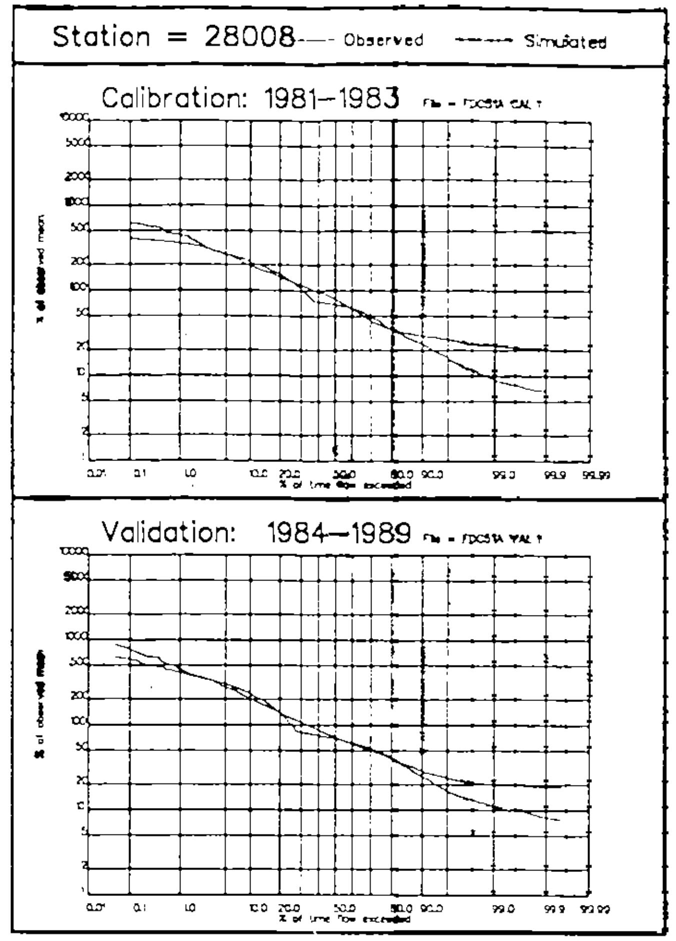


Observed and simulated flow duration curves for catchment 25006.

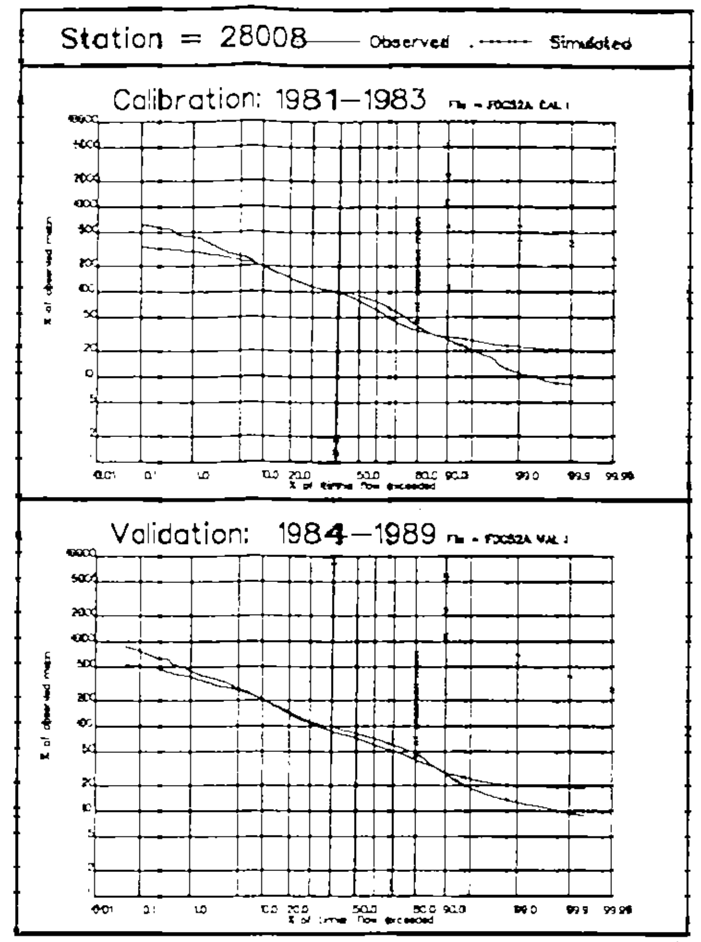
Model A



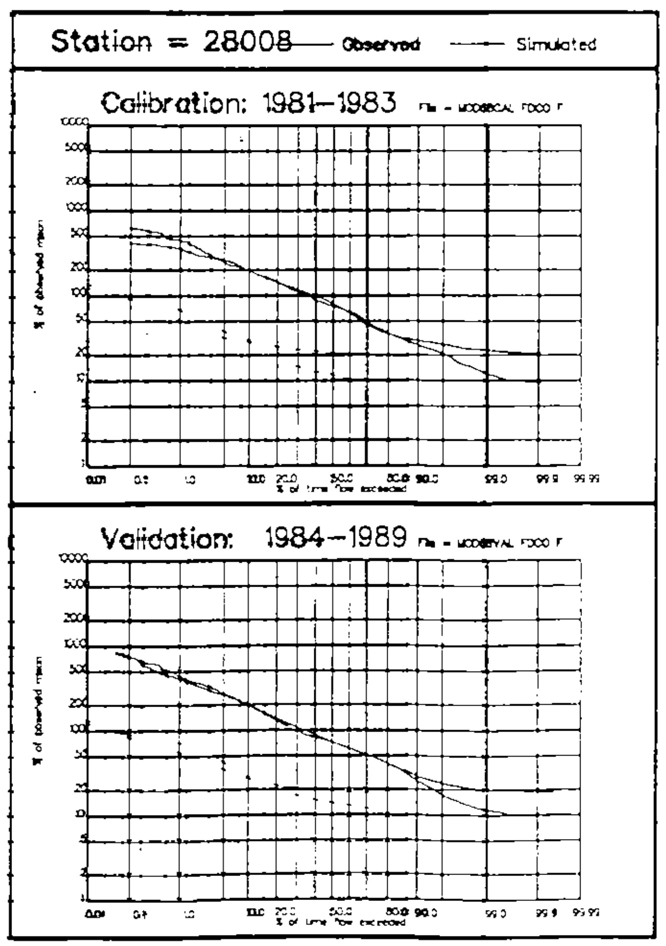
Model B1



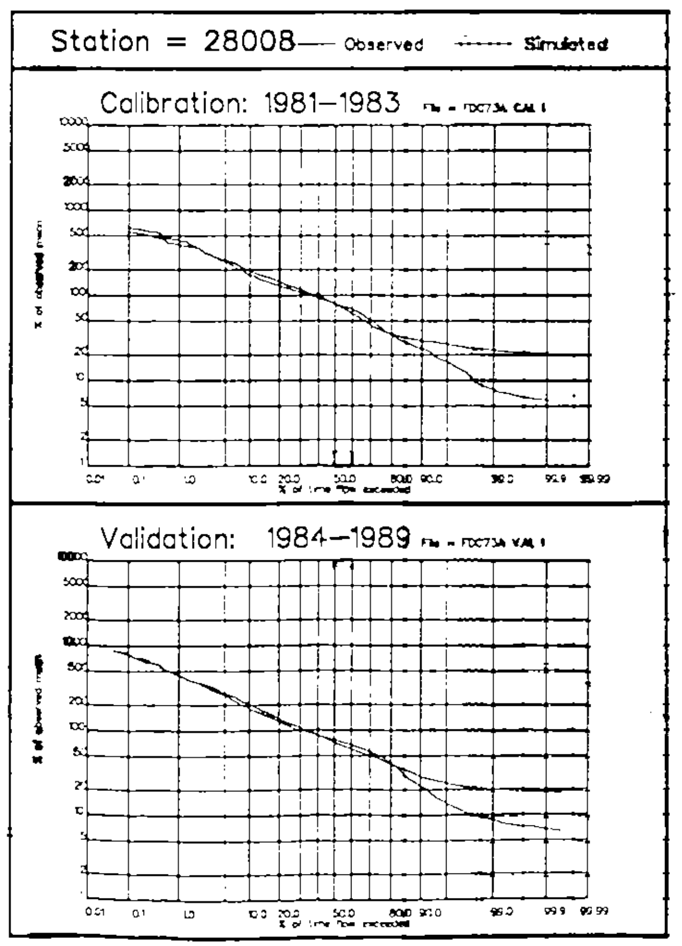
Model B2



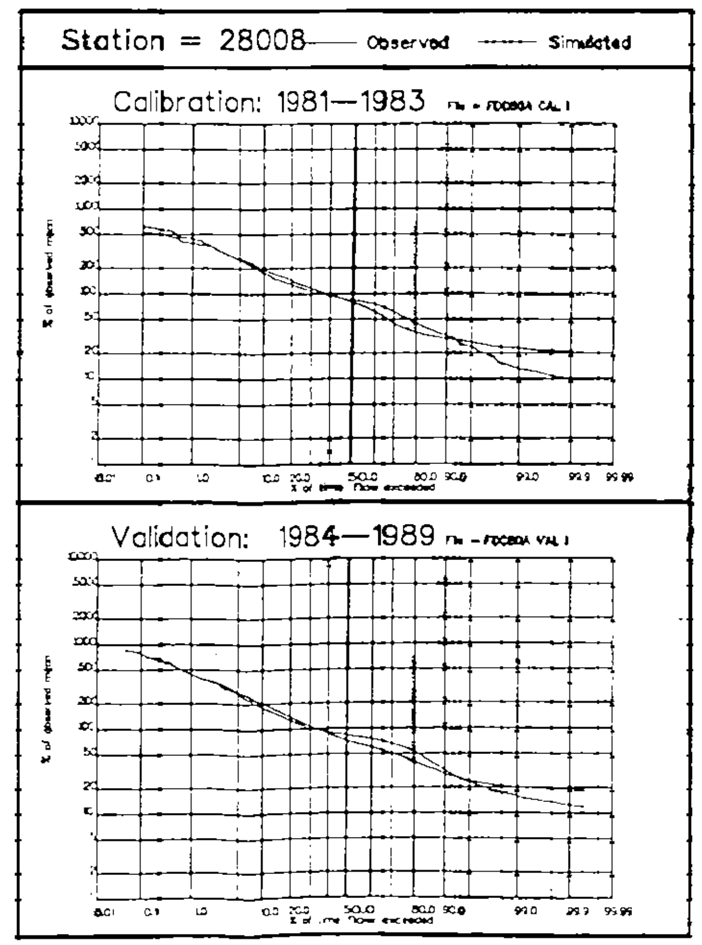
Model C



Model D1

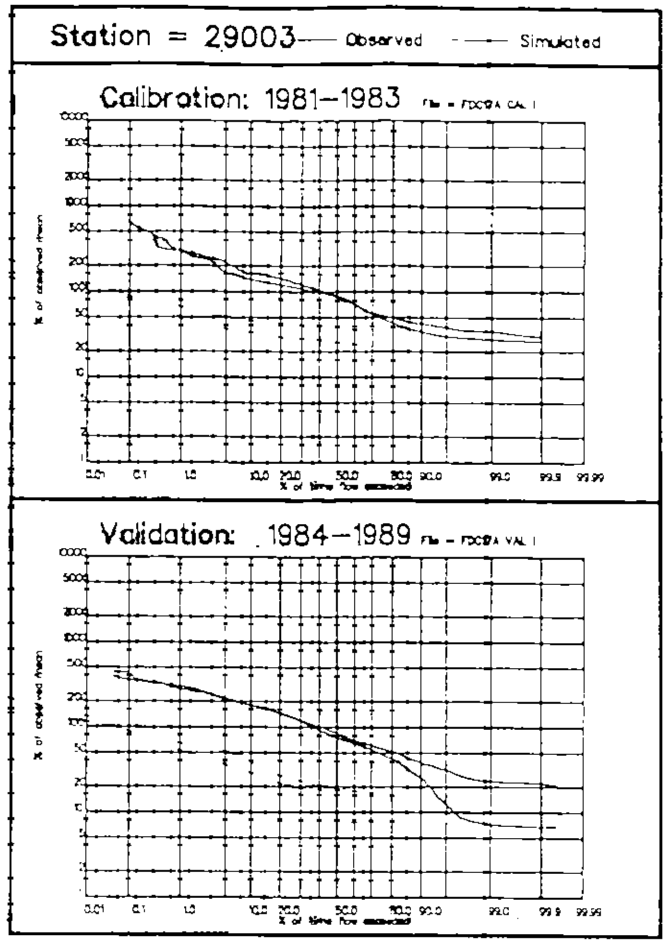


Model D2

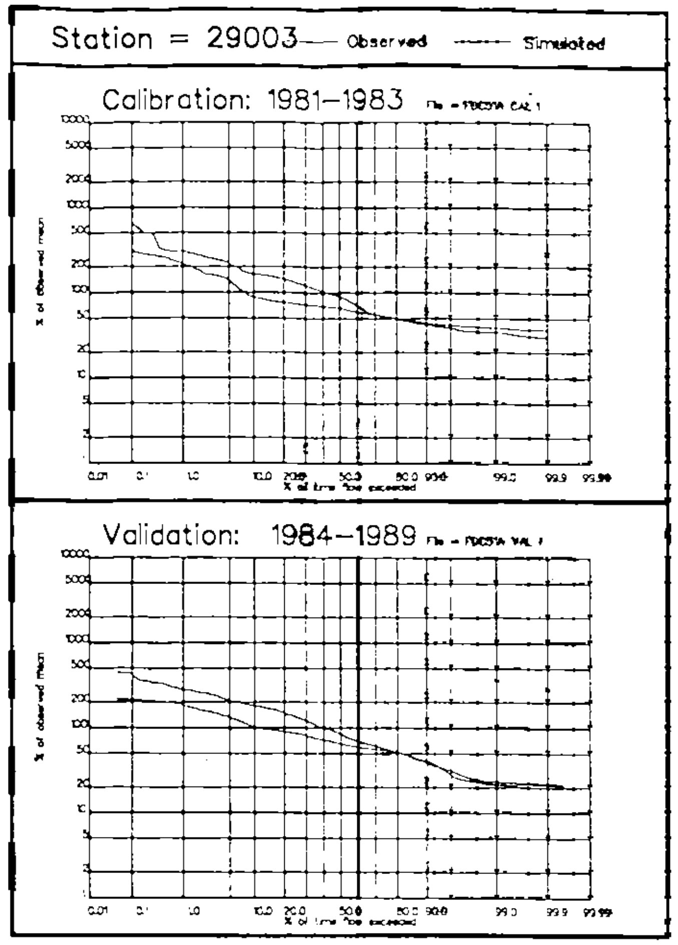


Observed and simulated flow duration curves for catchment 28008.

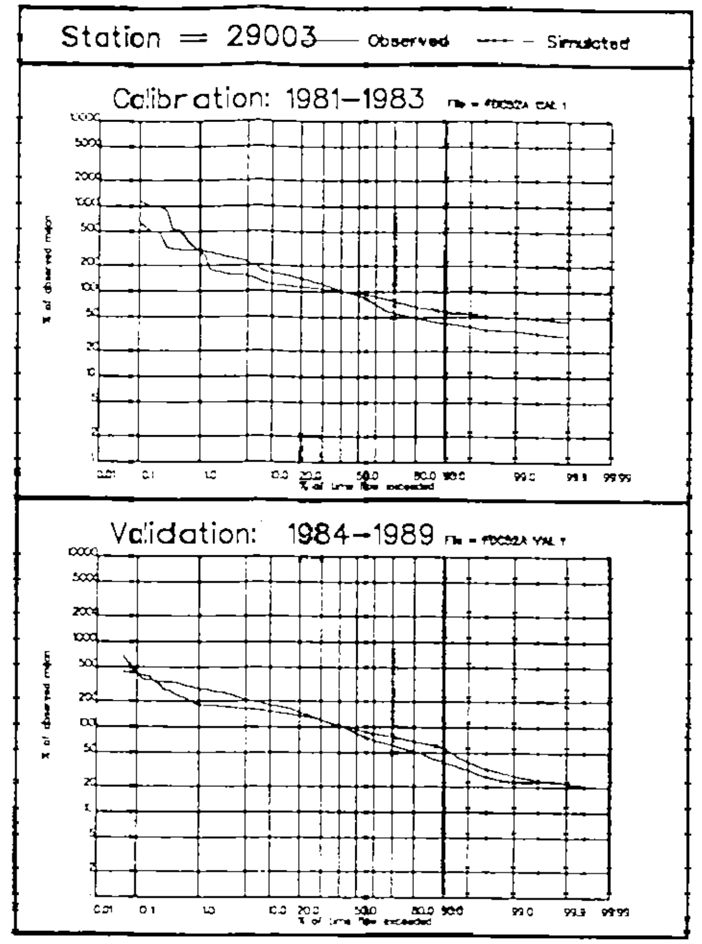
Model A



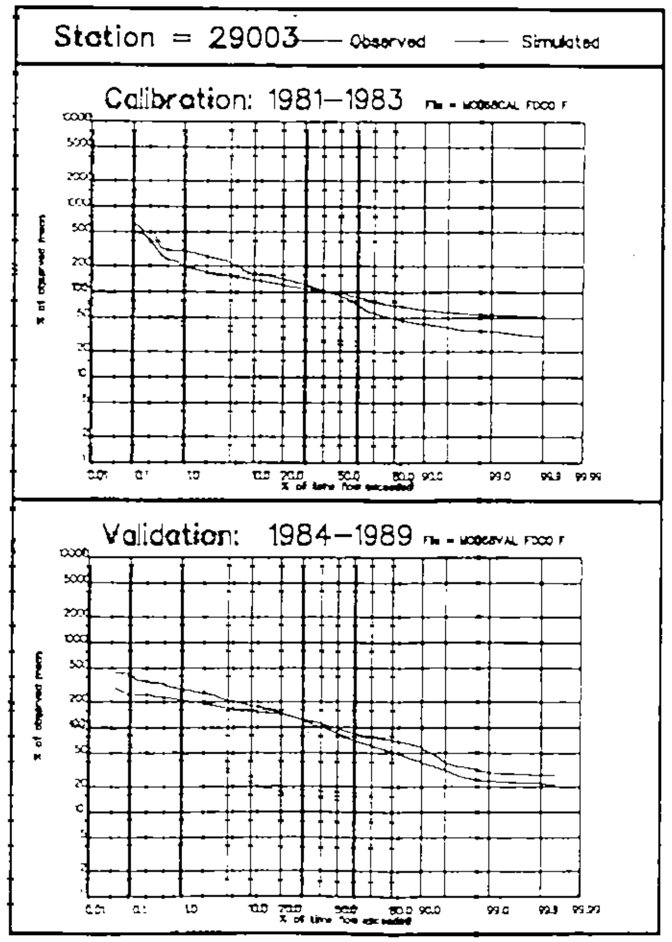
Model B1



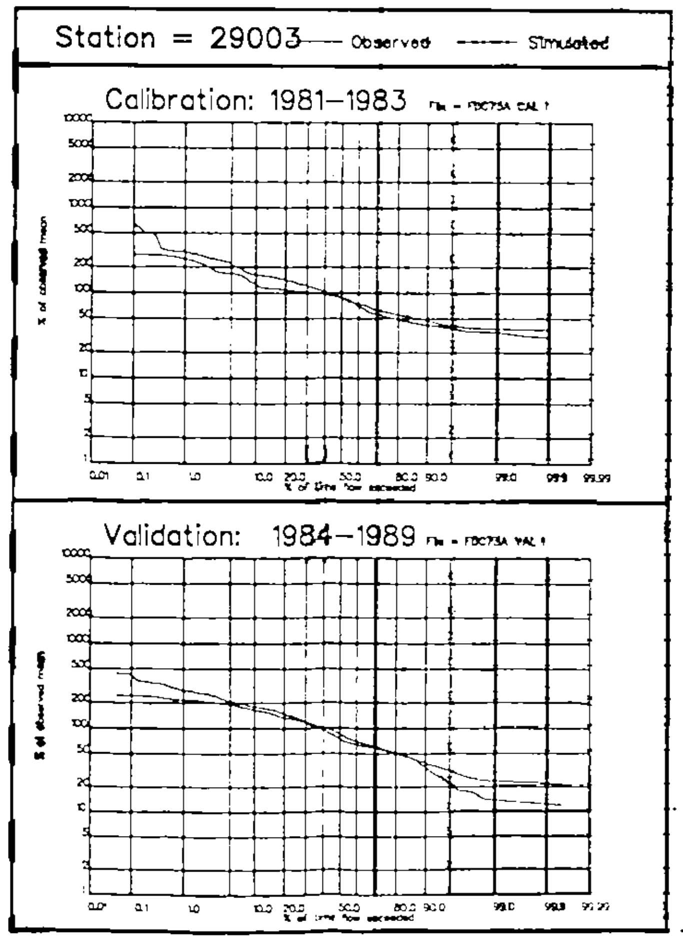
Model B2



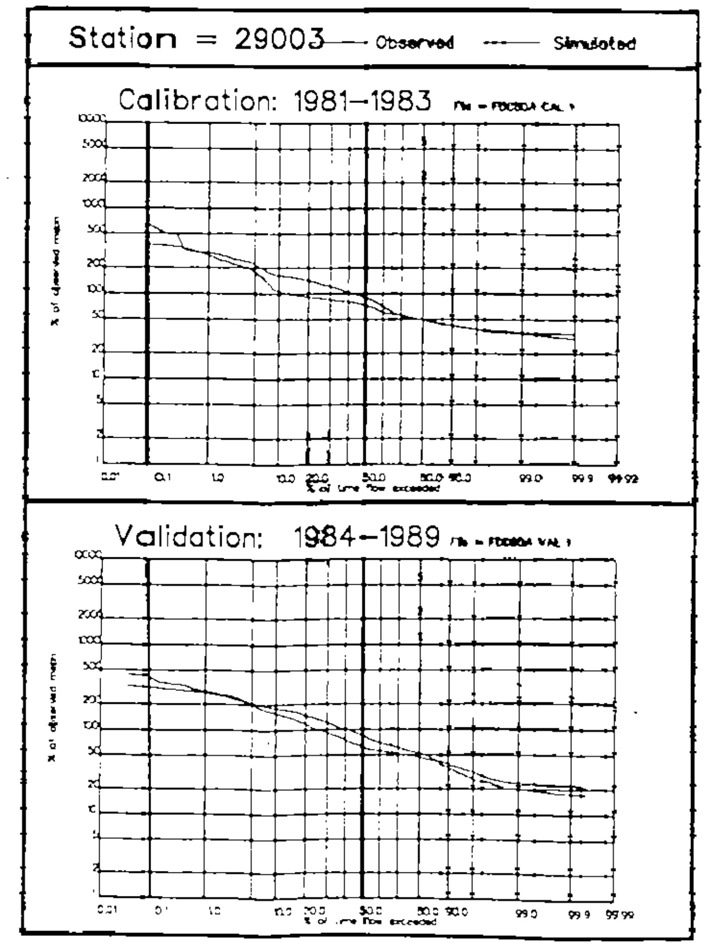
Model C



Model D1

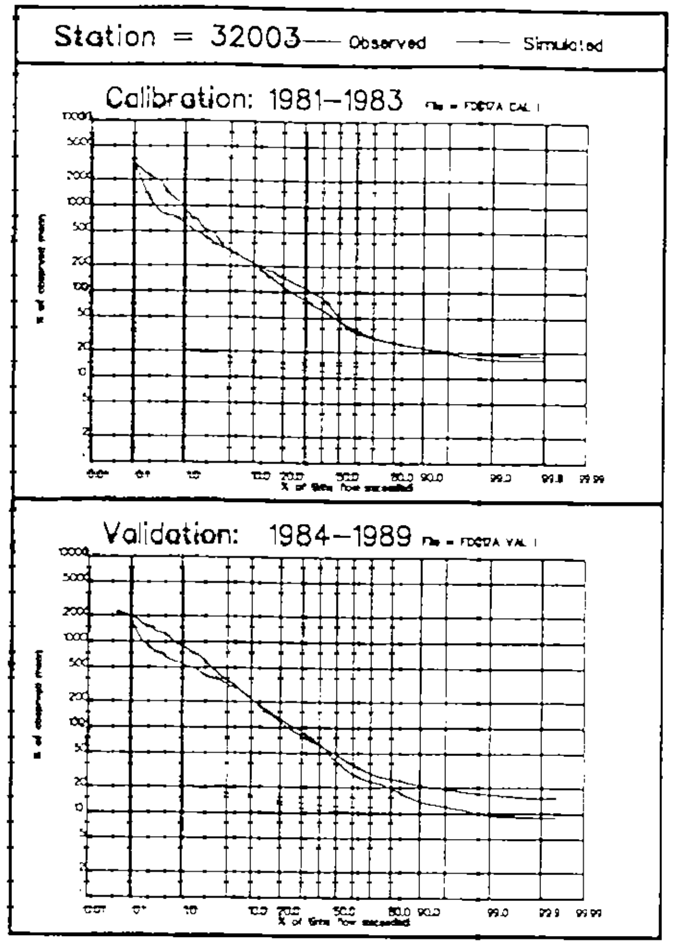


Model D2

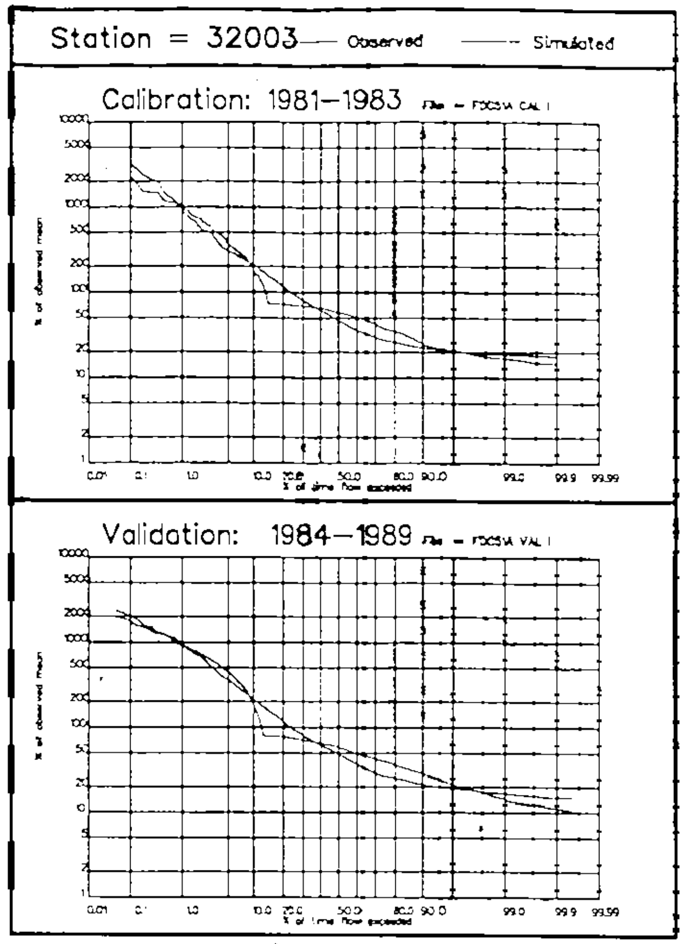


Observed and simulated flow duration curves for catchment 29003.

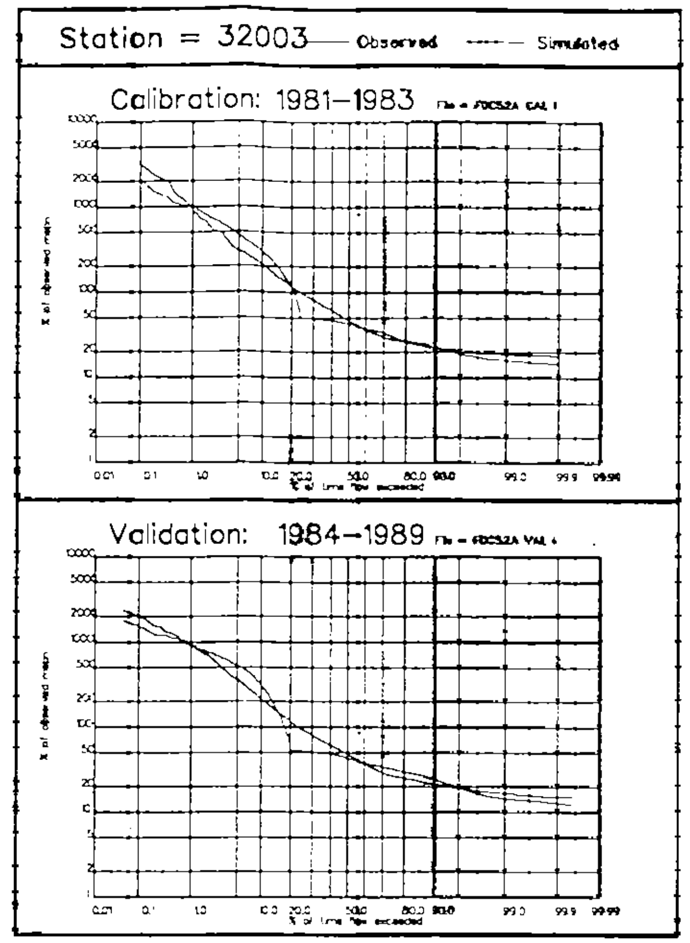
Model A



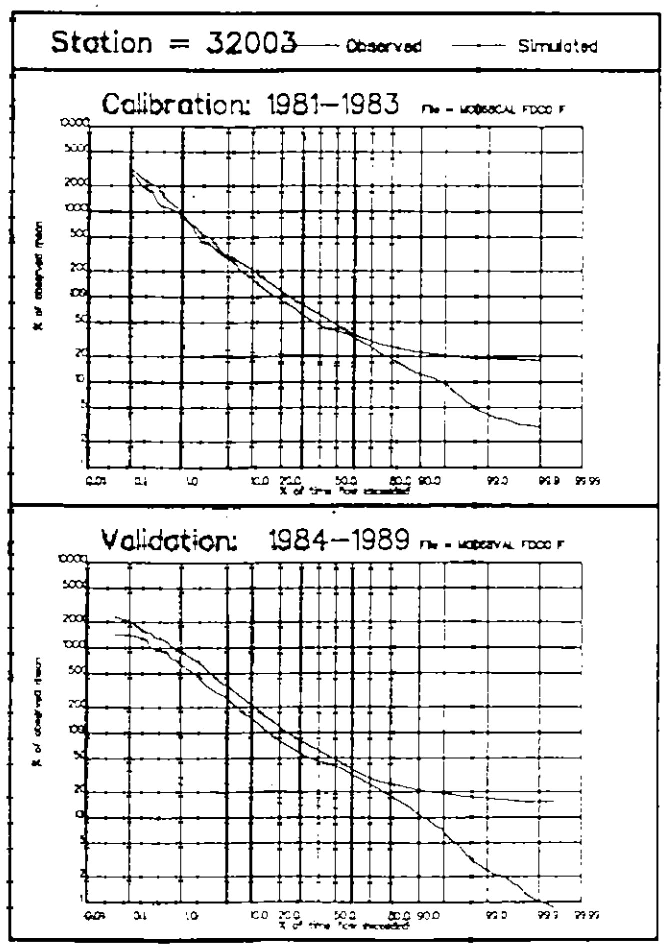
Model B1



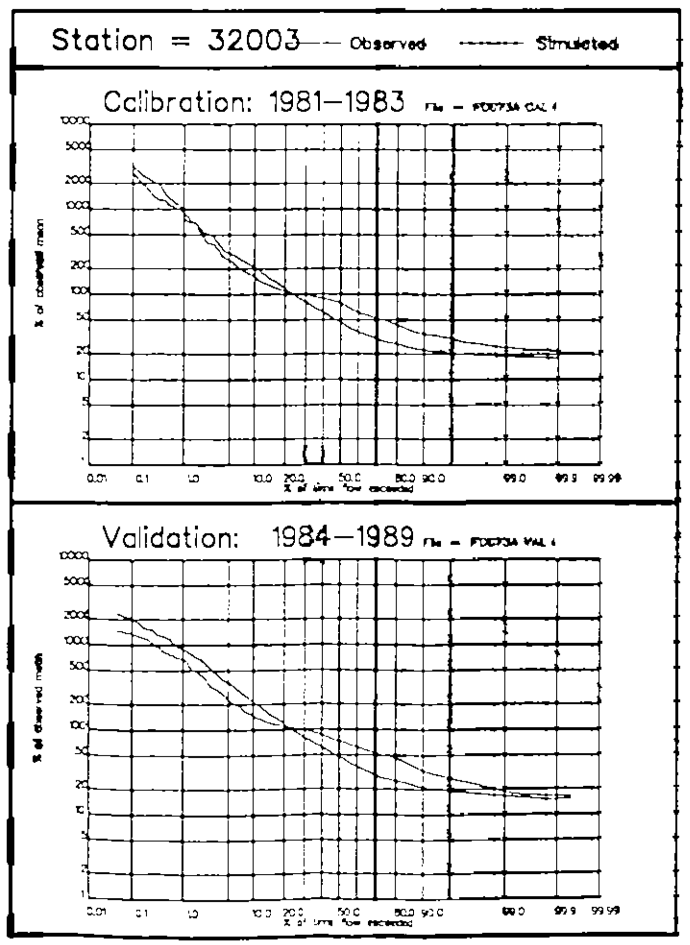
Model B2



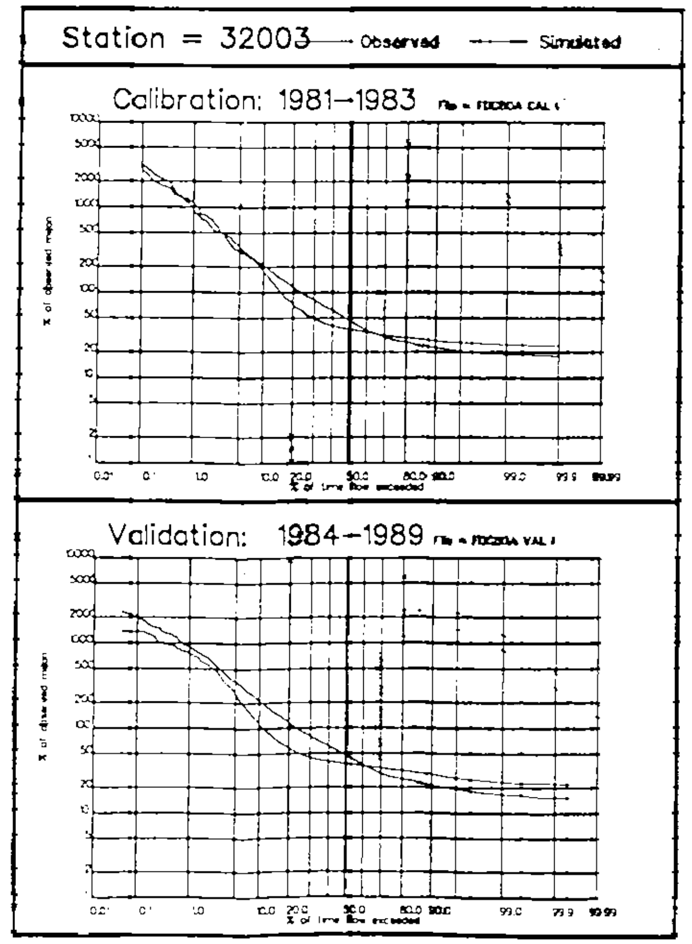
Model C



Model D1

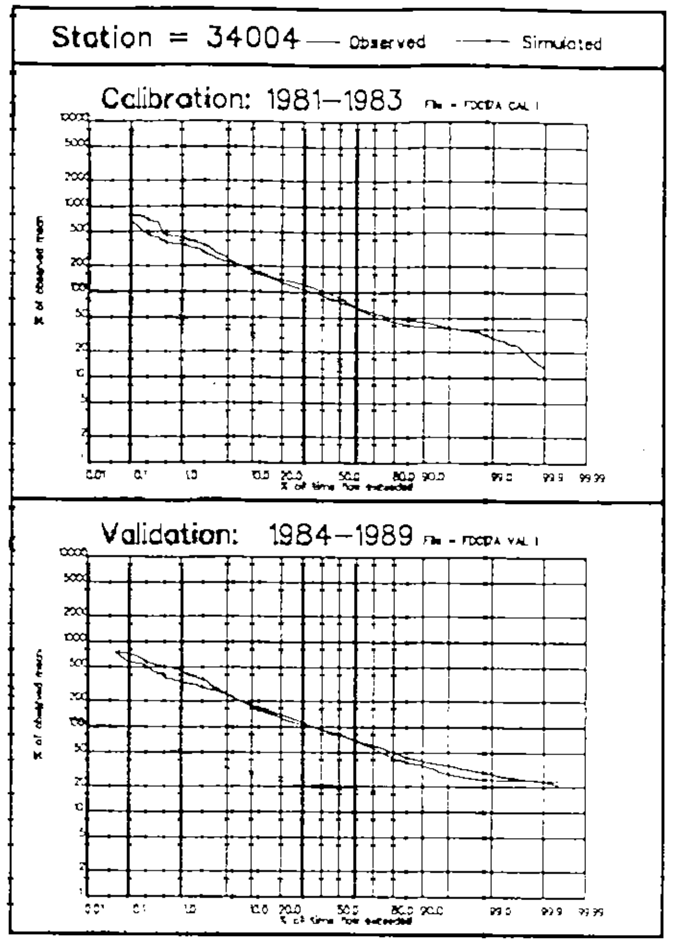


Model D2

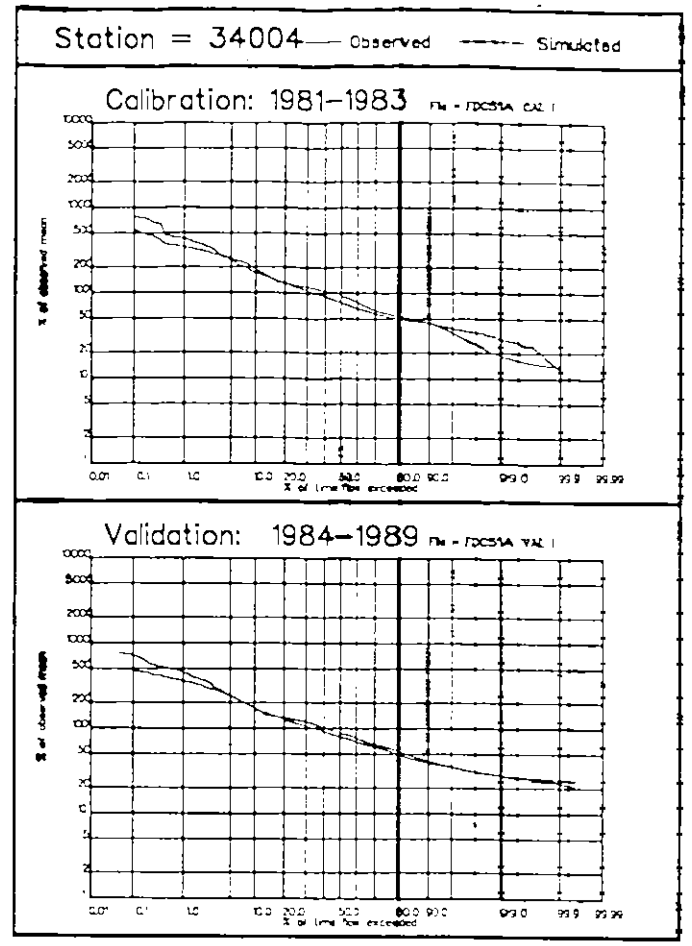


Observed and simulated flow duration curves for catchment 32003.

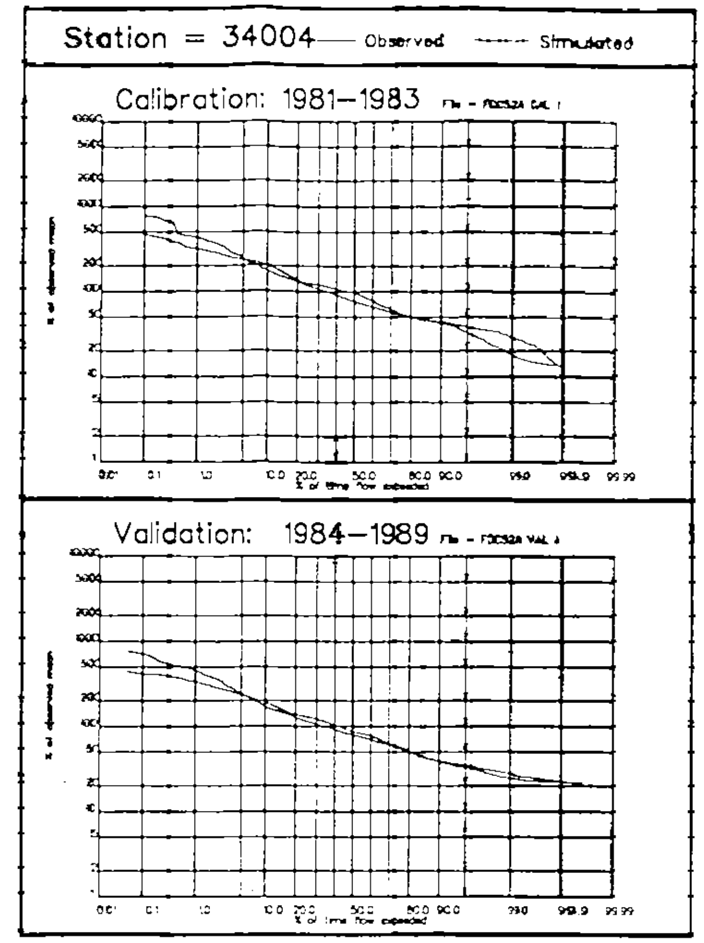
Model A



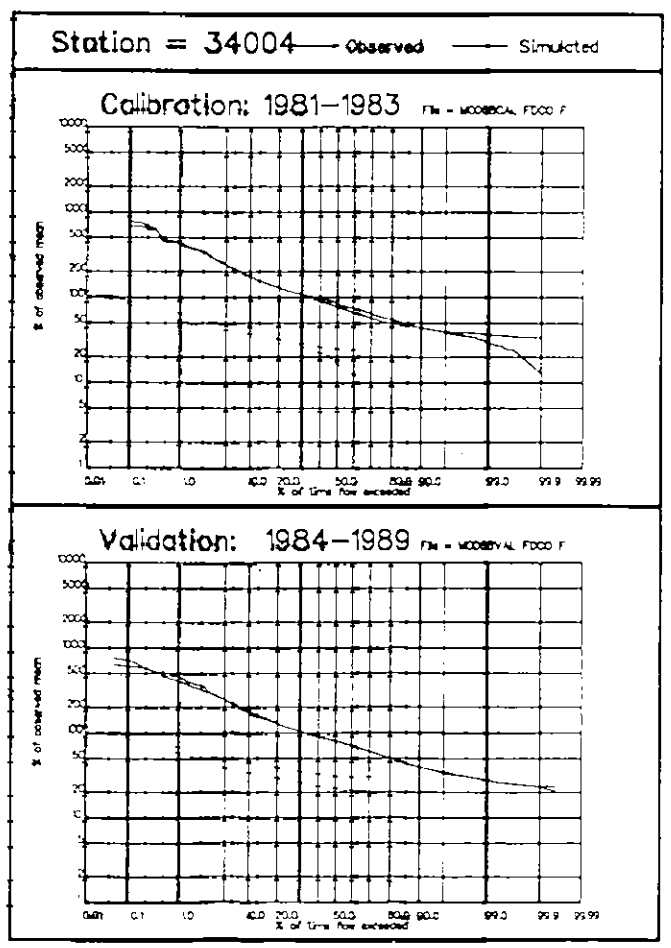
Model B1



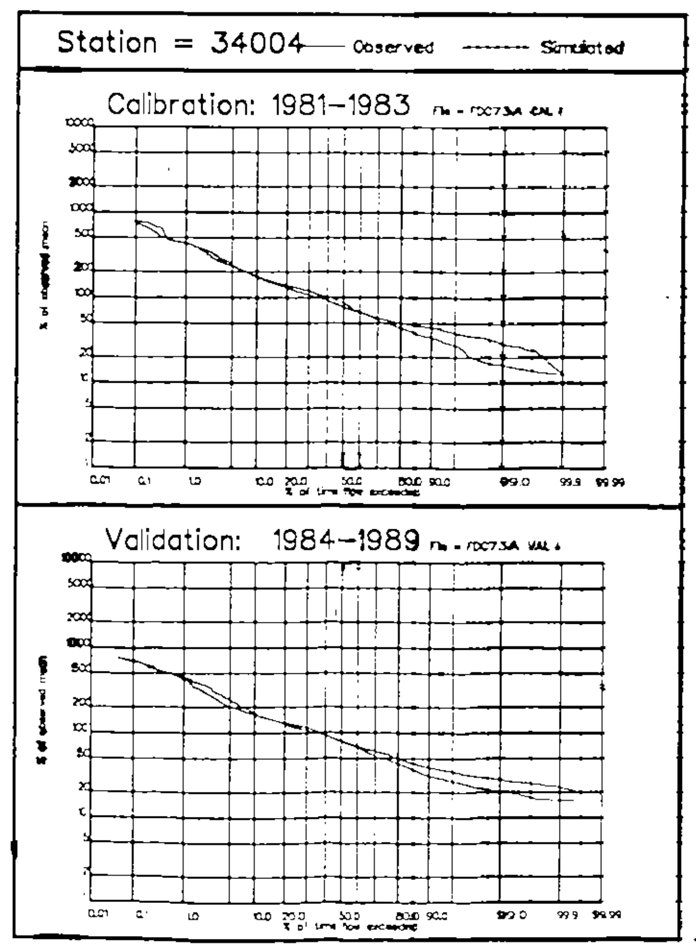
Model B2



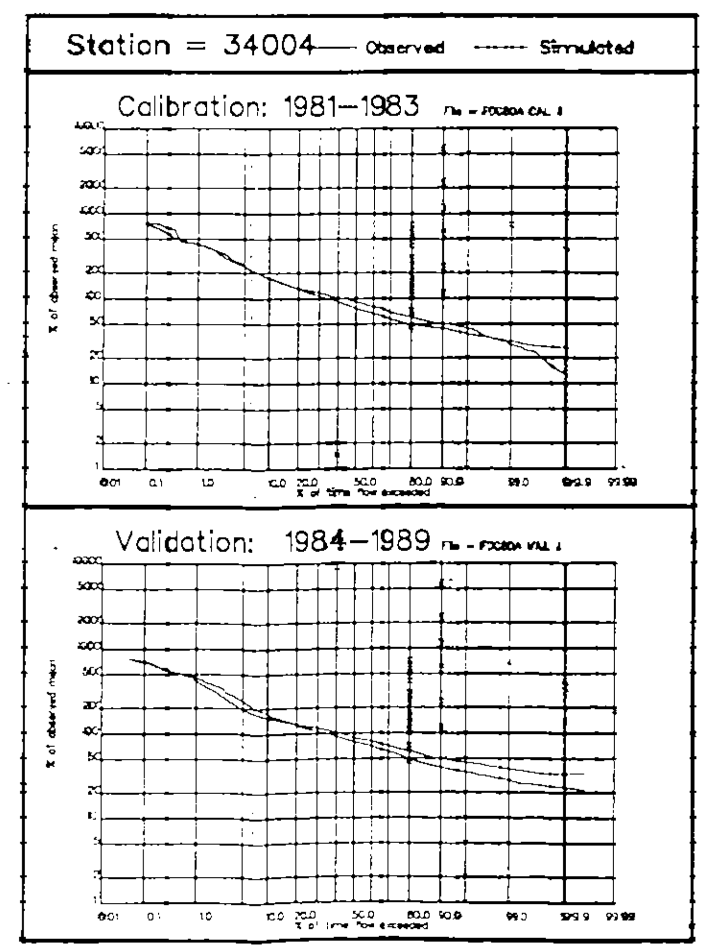
Model C



Model D1

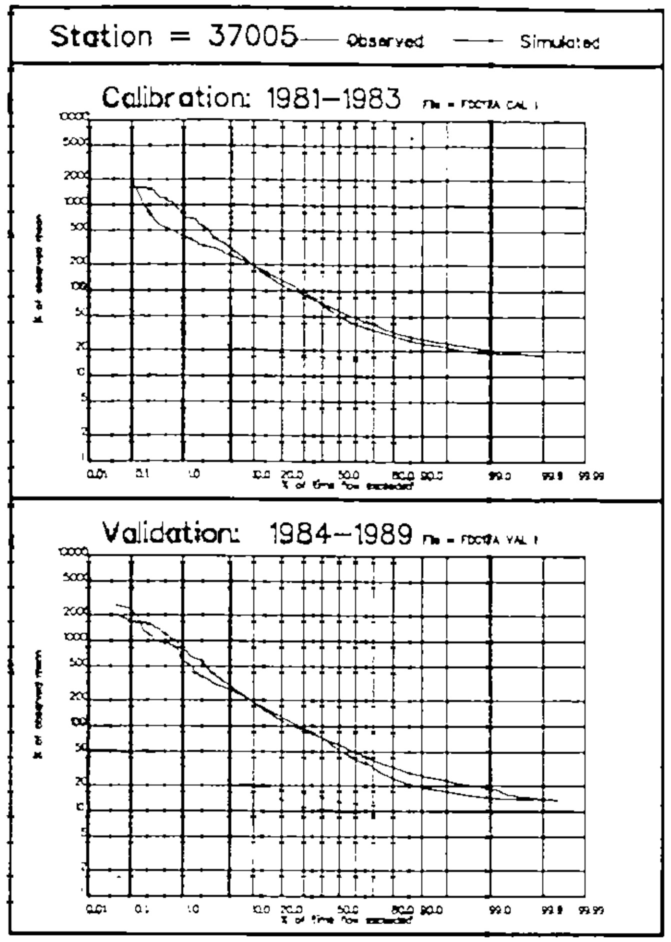


Model D2

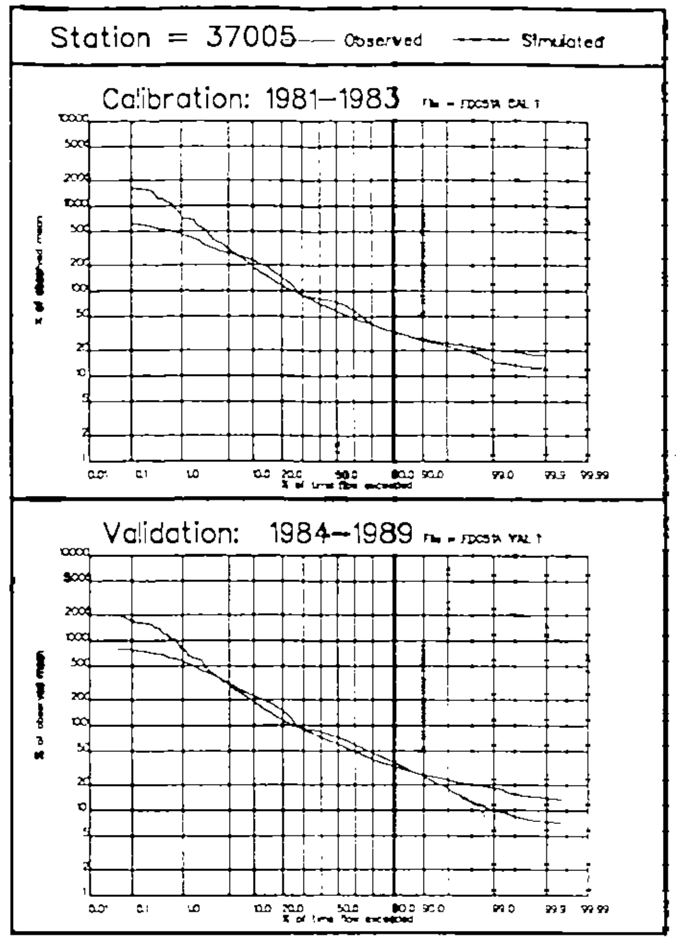


Observed and simulated flow duration curves for catchment 34004.

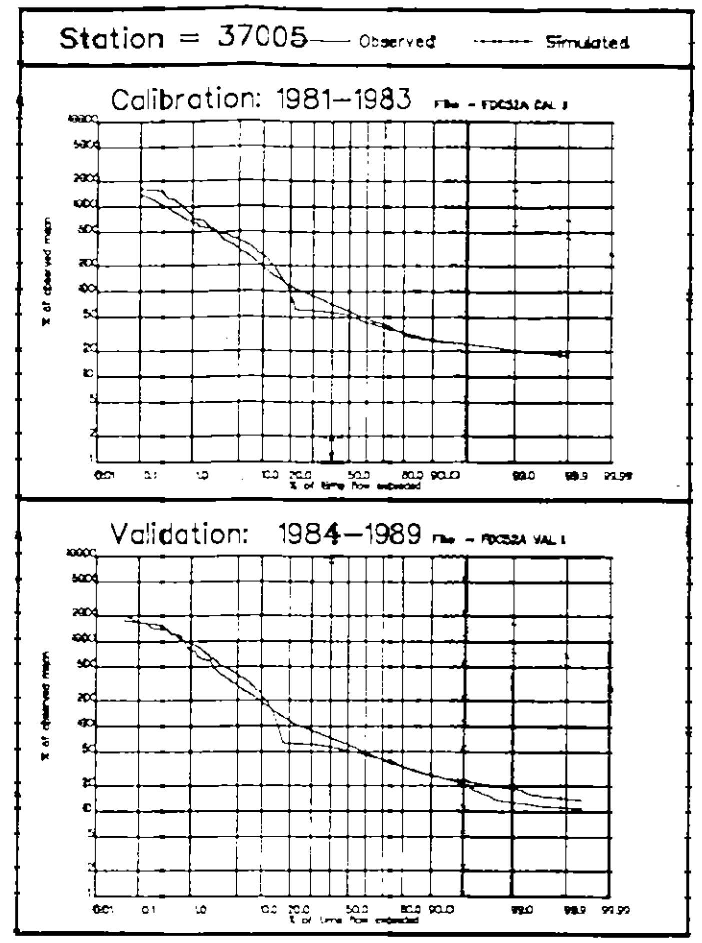
Model A



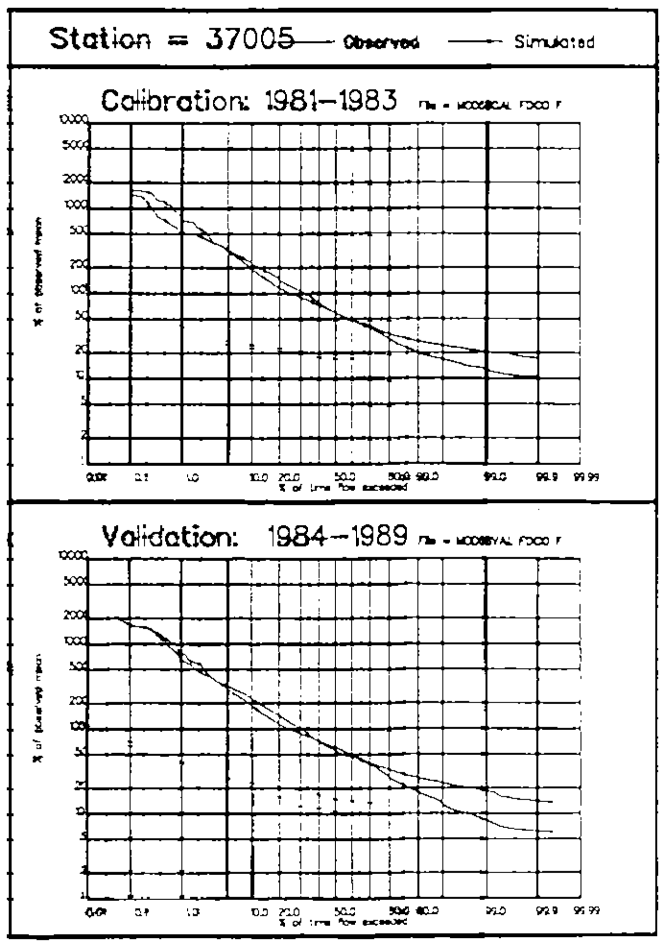
Model B1



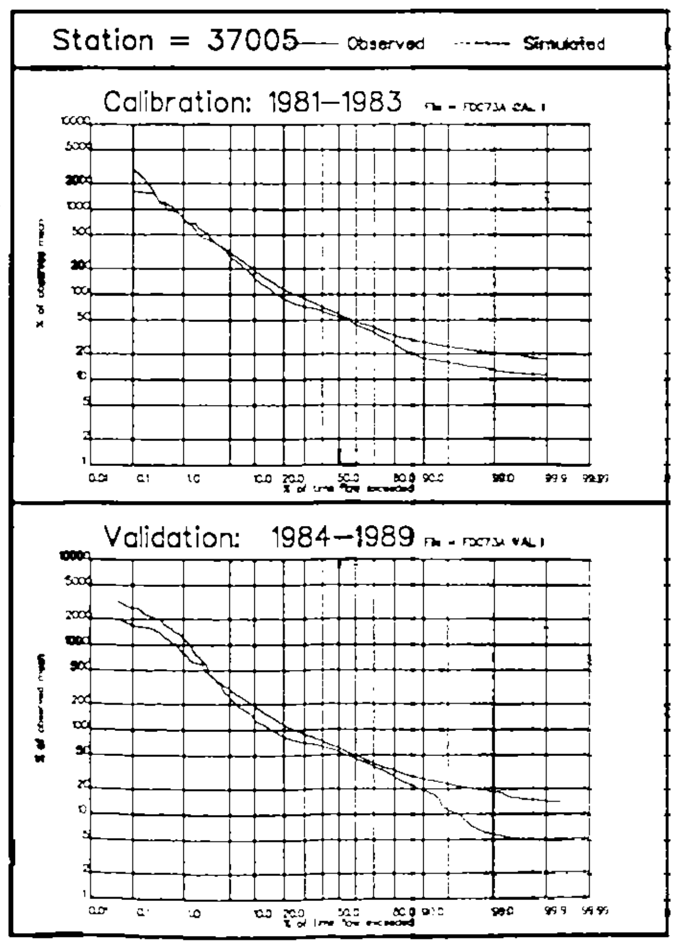
Model B2



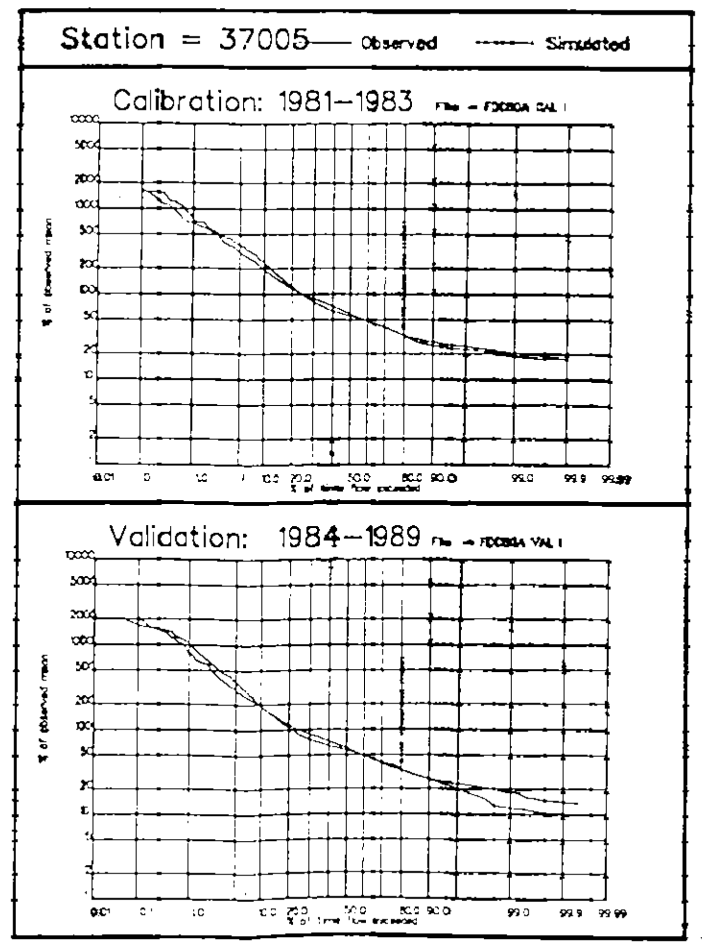
Model C



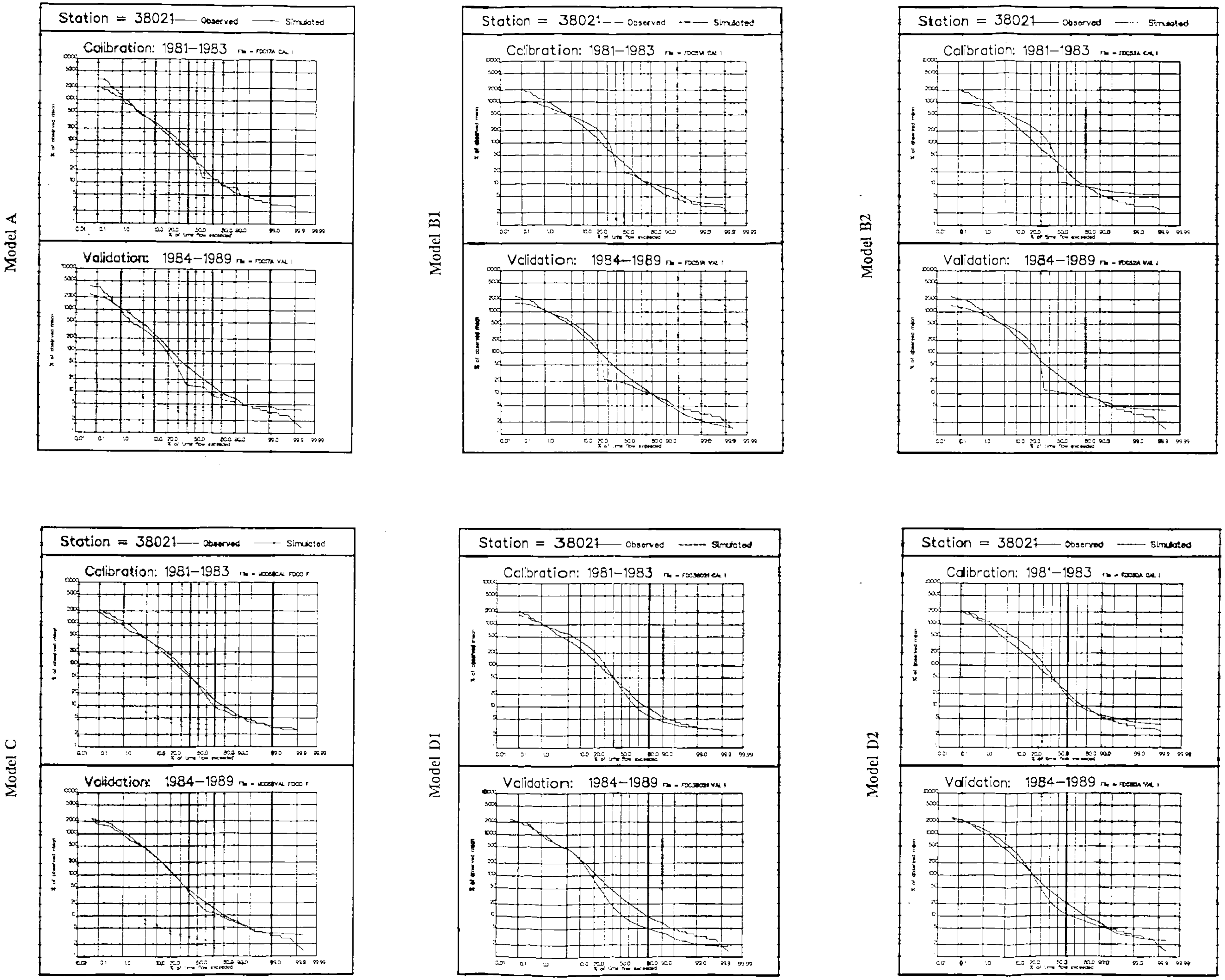
Model D1



Model D2

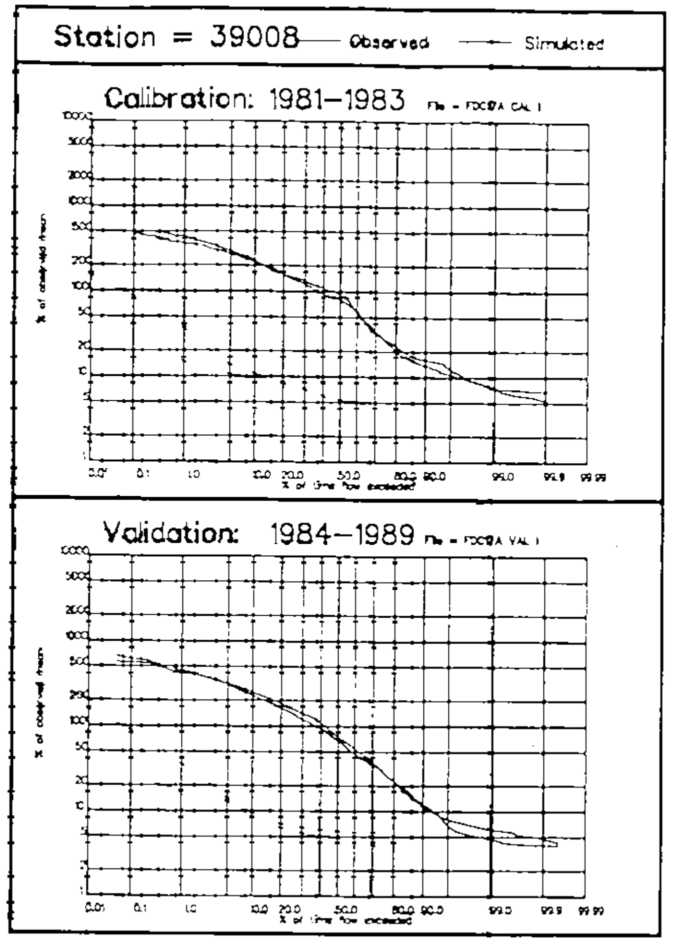


Observed and simulated flow duration curves for catchment 37005.

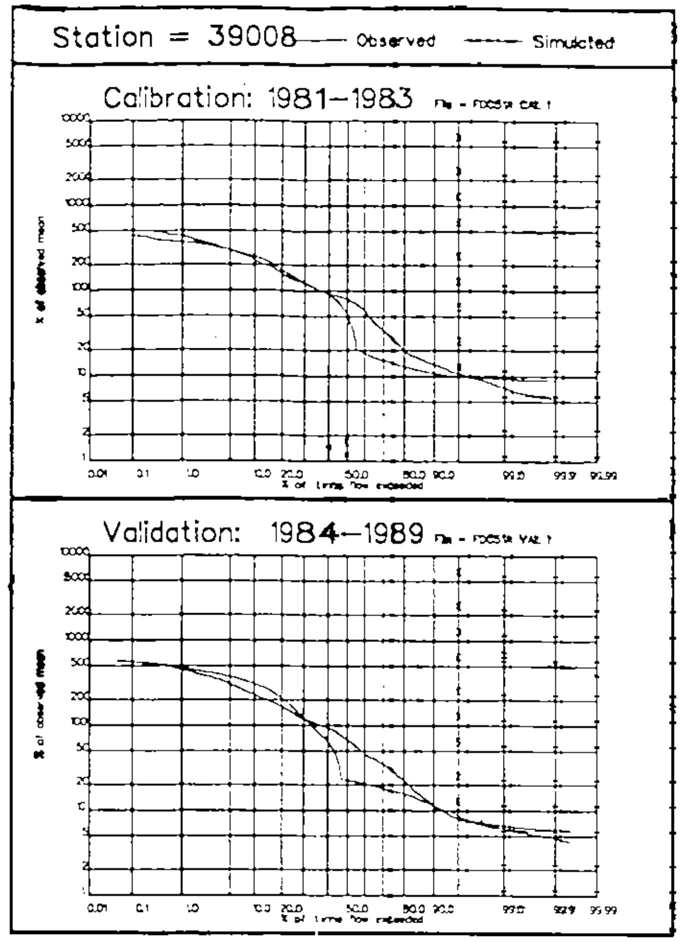


Observed and simulated flow duration curves for catchment 38021.

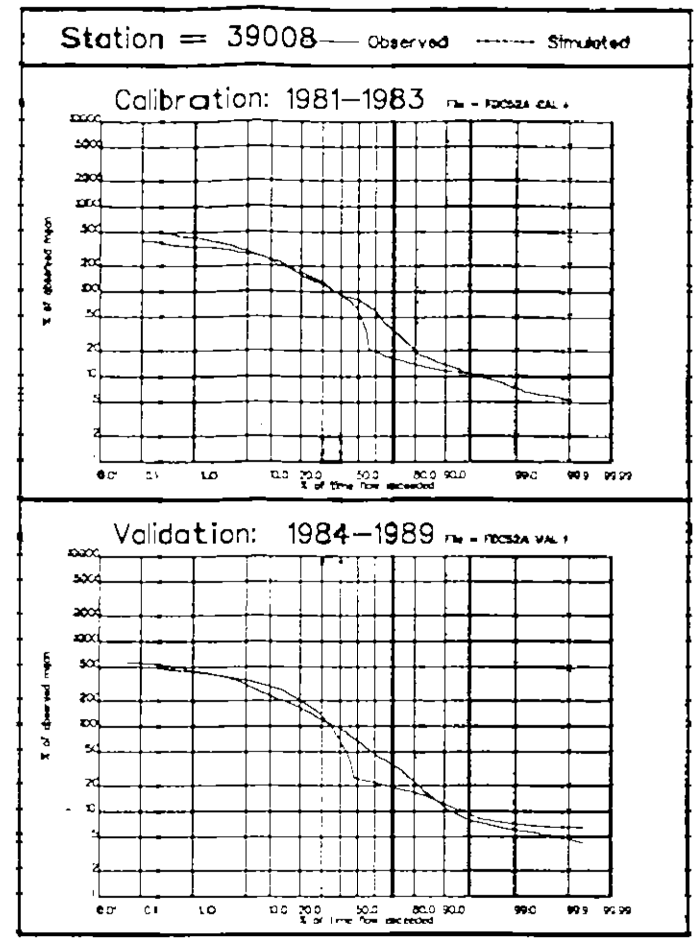
Model A



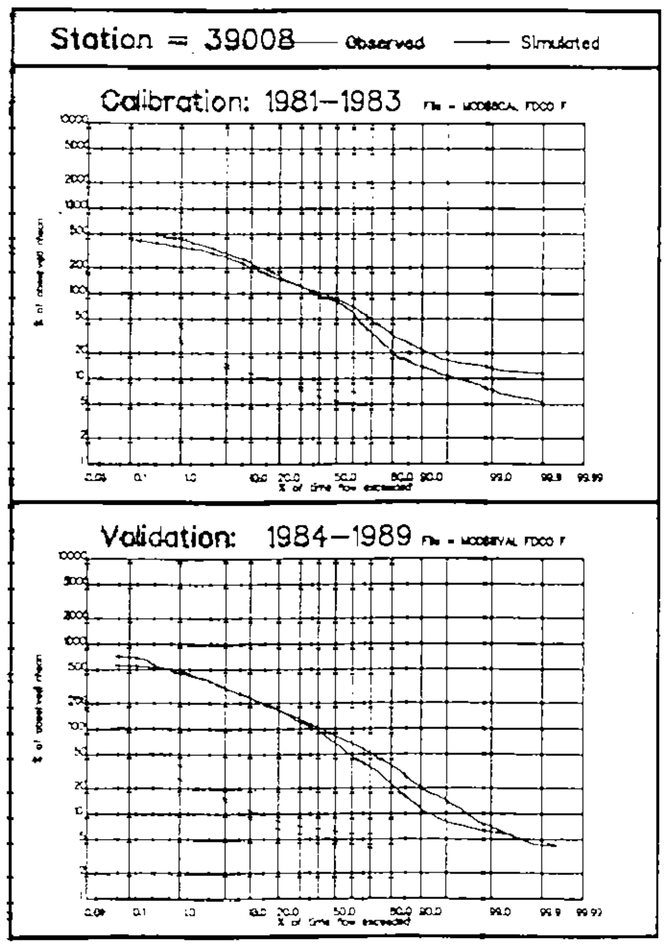
Model B1



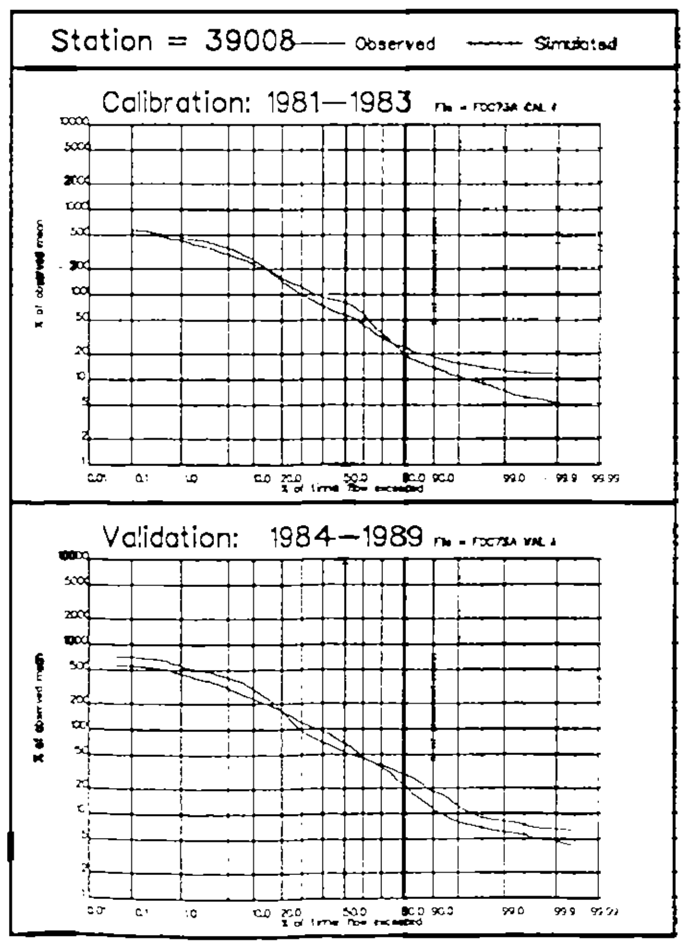
Model B2



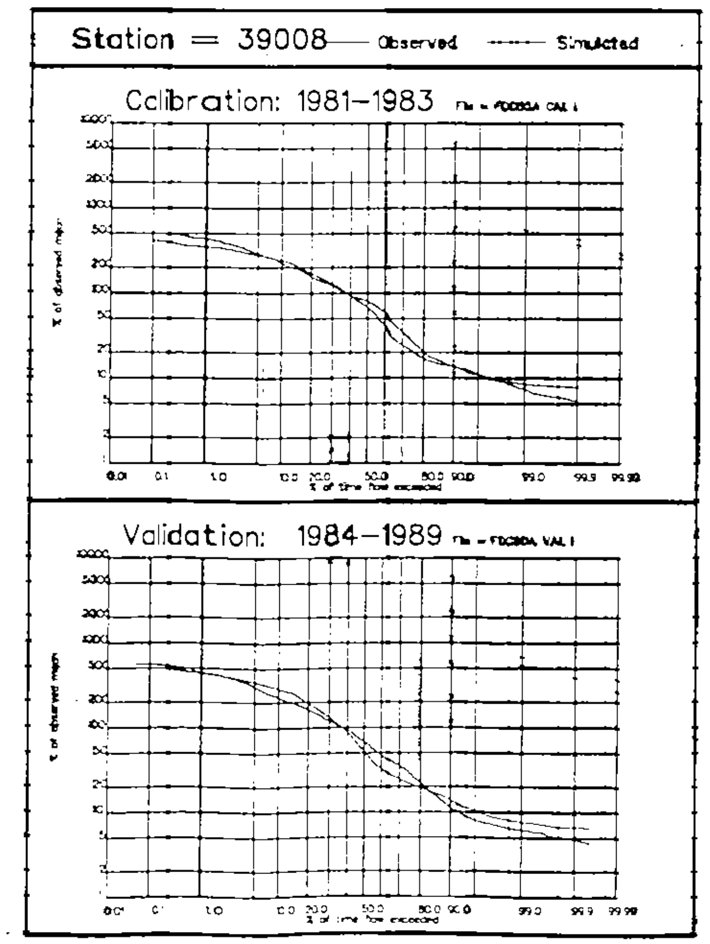
Model C



Model D1

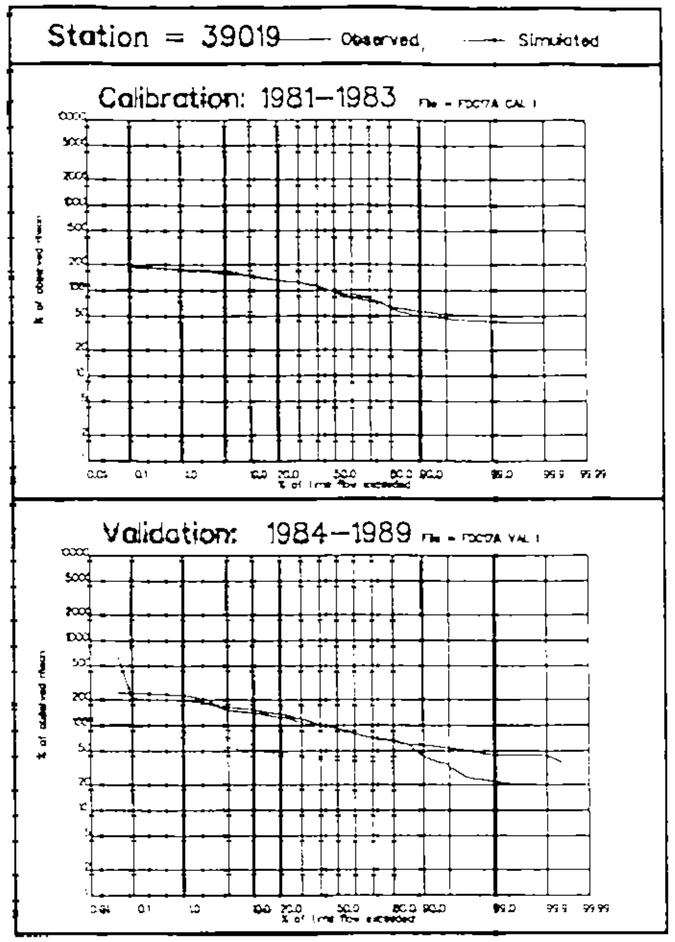


Model D2

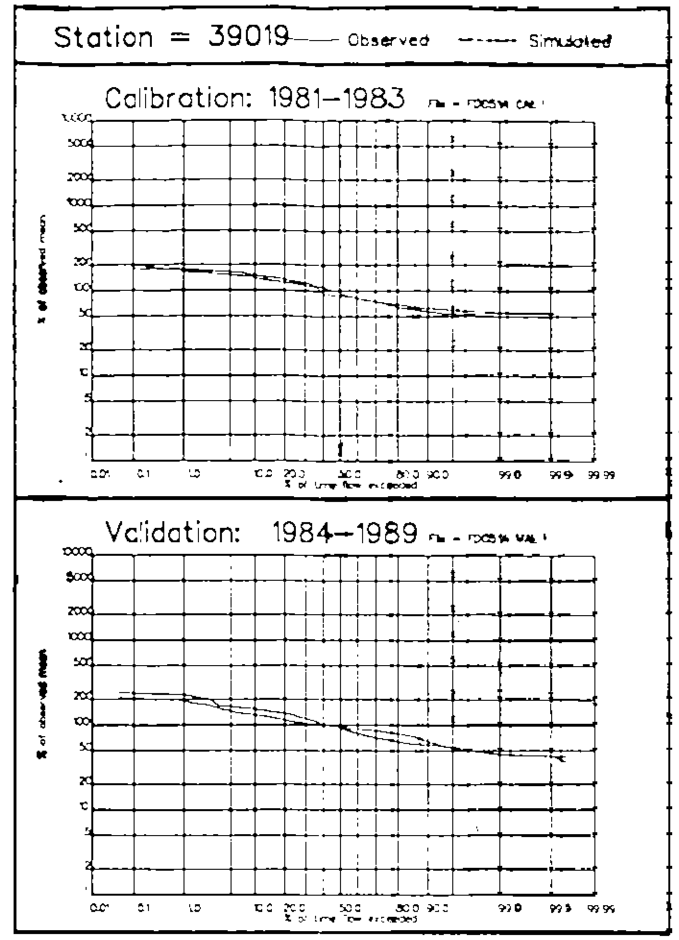


Observed and simulated flow duration curves for catchment 39008.

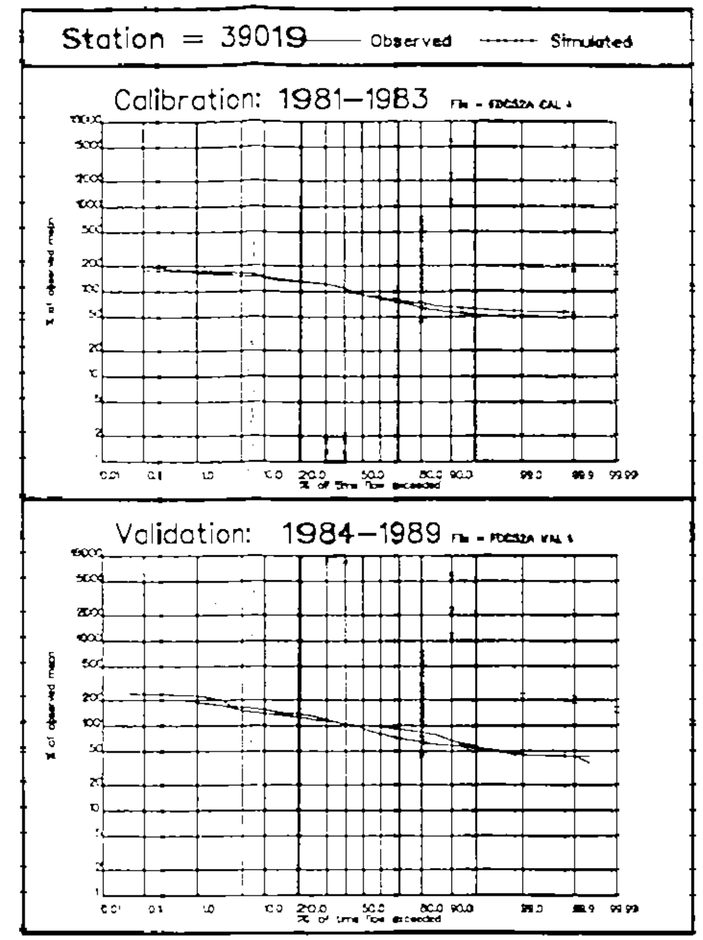
Model A



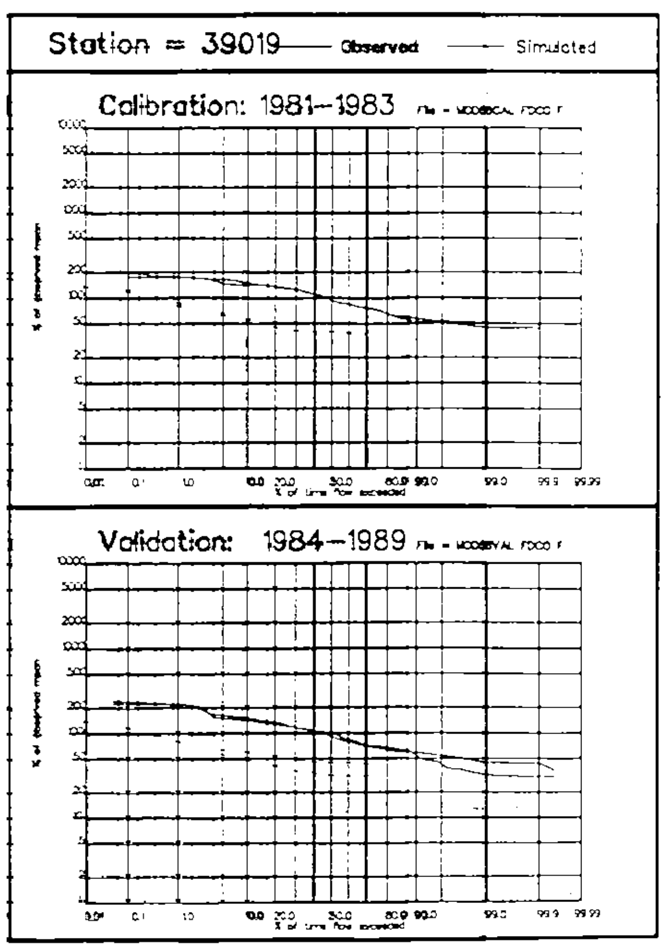
Model B1



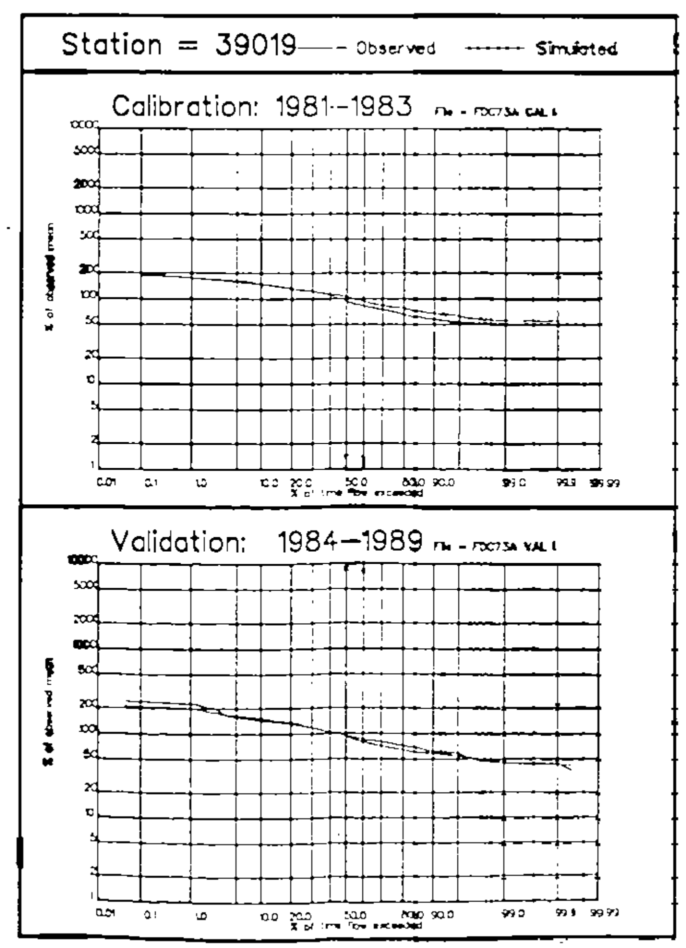
Model B2



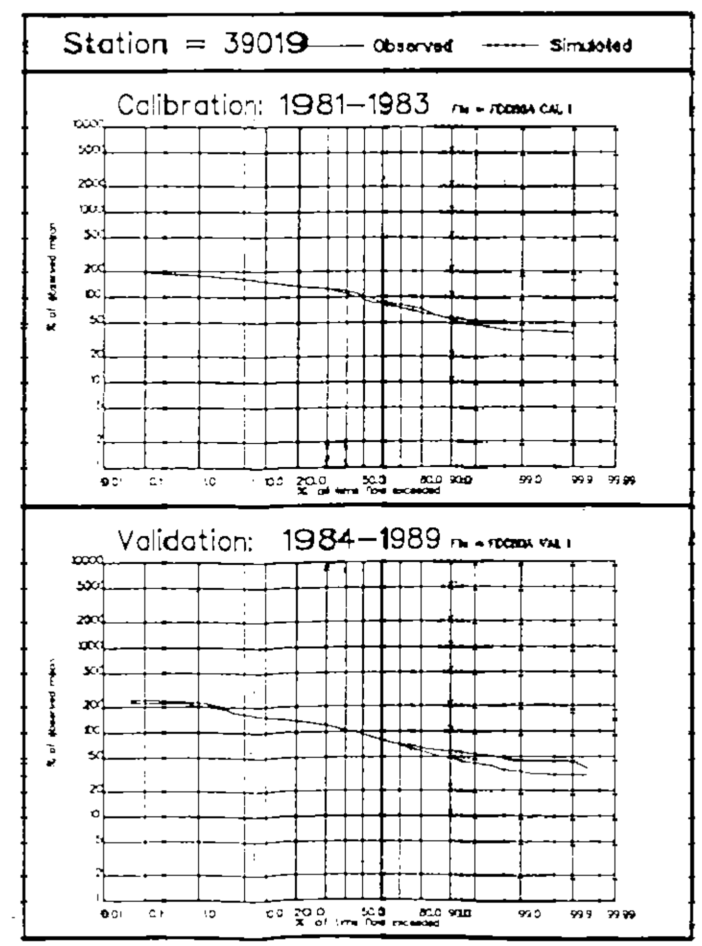
Model C



Model D1

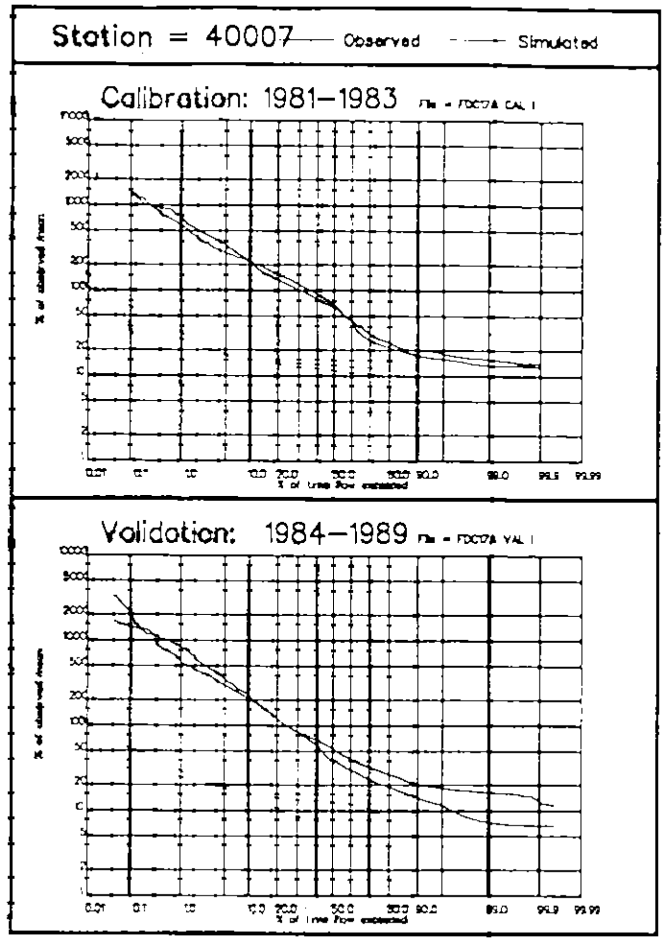


Model D2

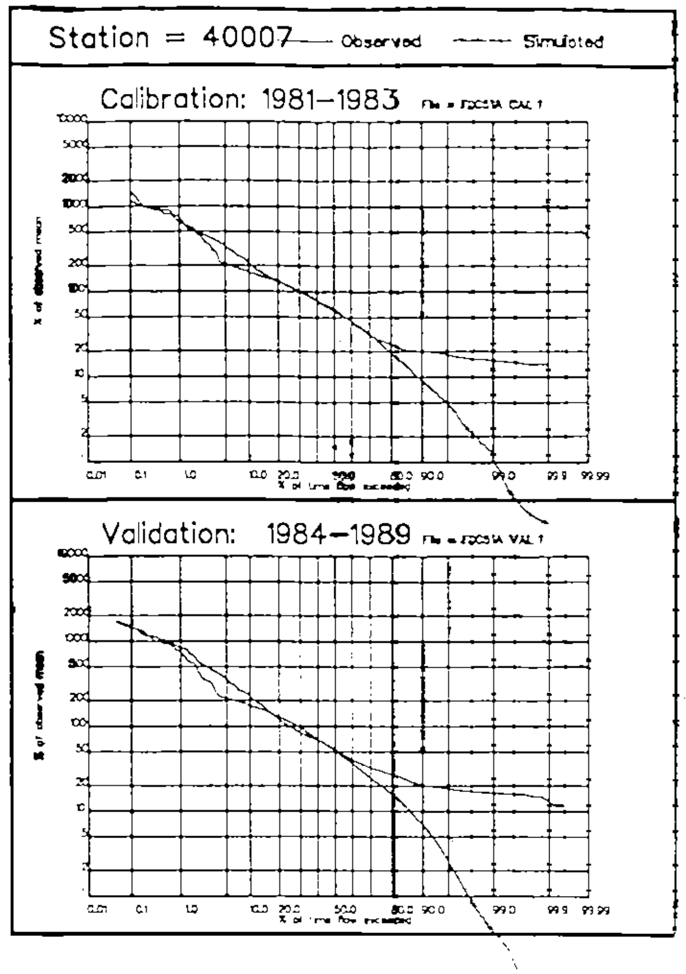


Observed and simulated flow duration curves for catchment 39019.

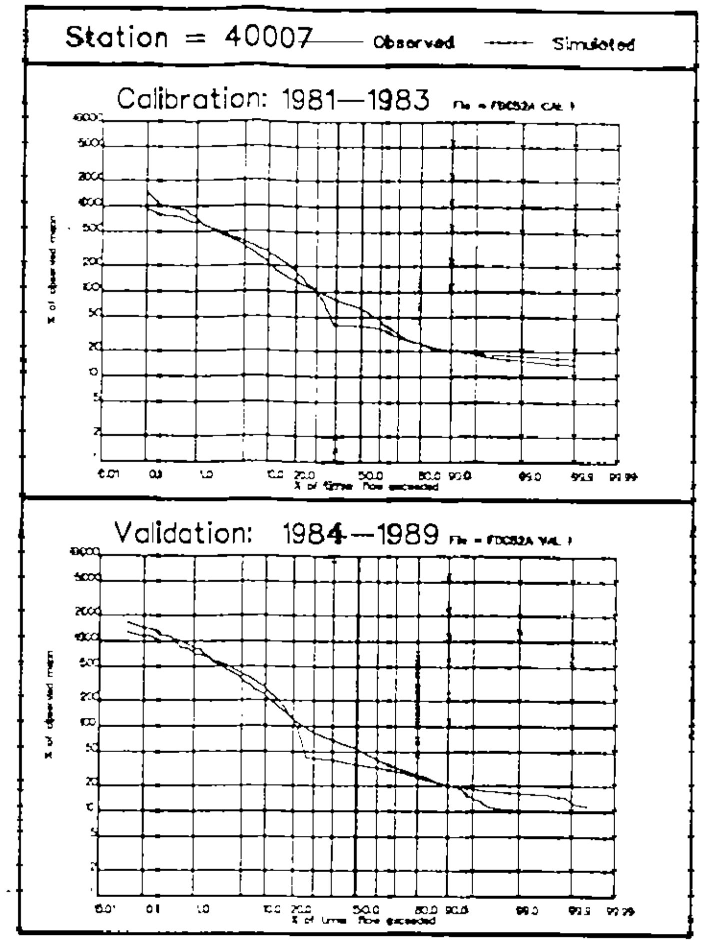
Model A



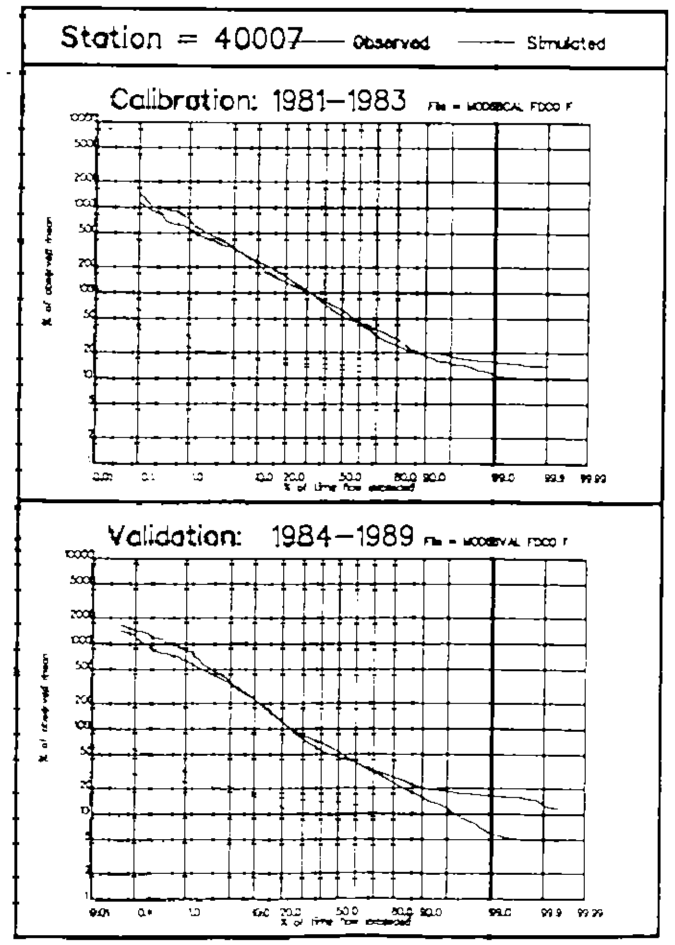
Model B1



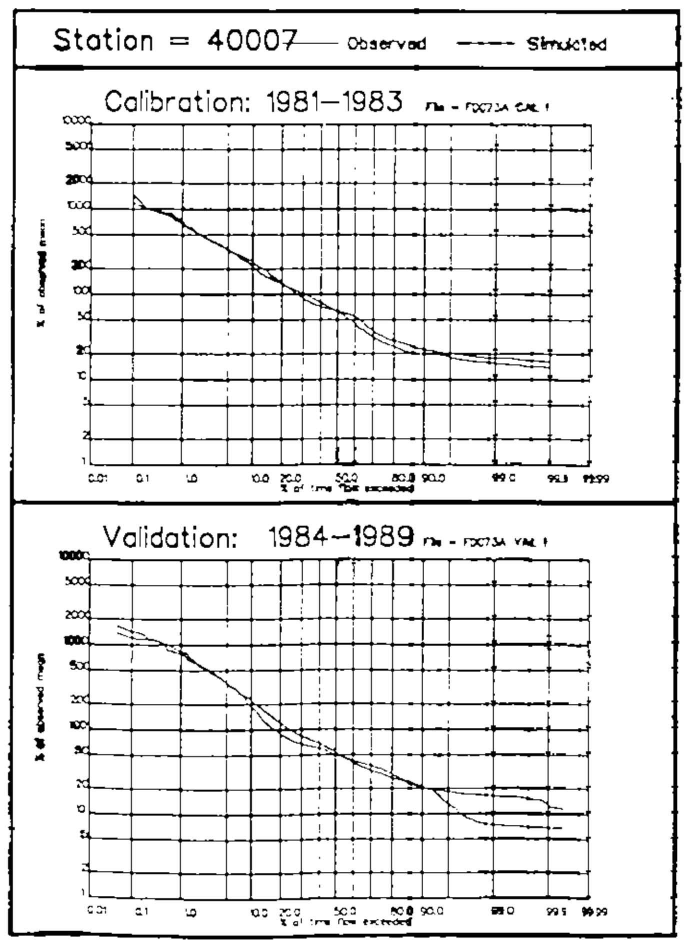
Model B2



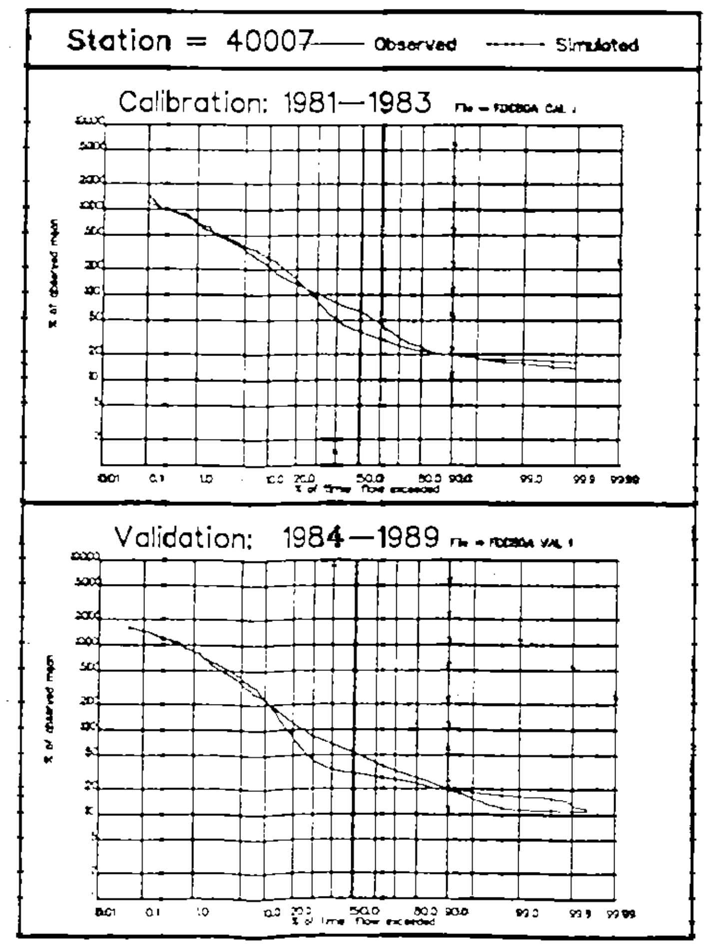
Model C



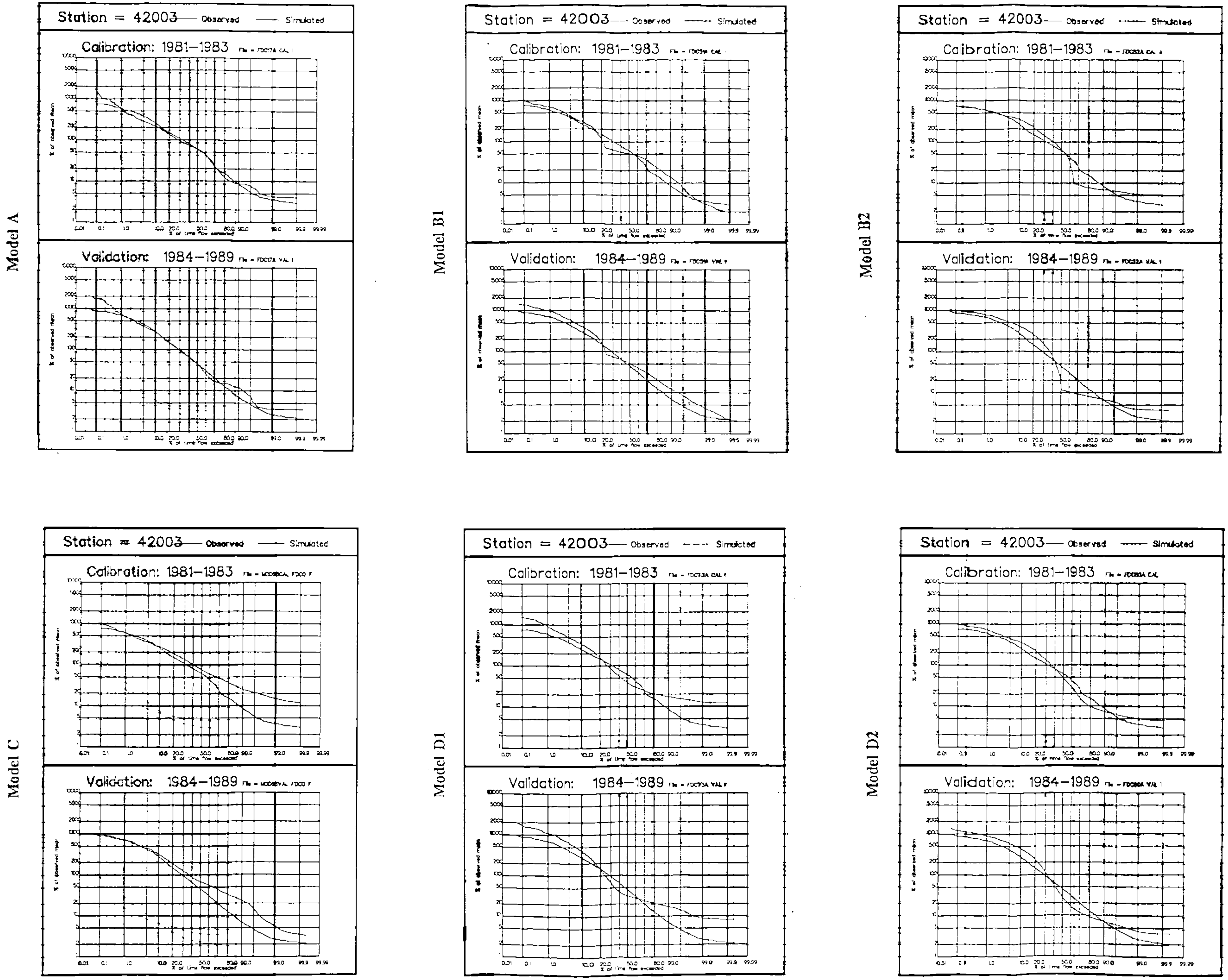
Model D1



Model D2

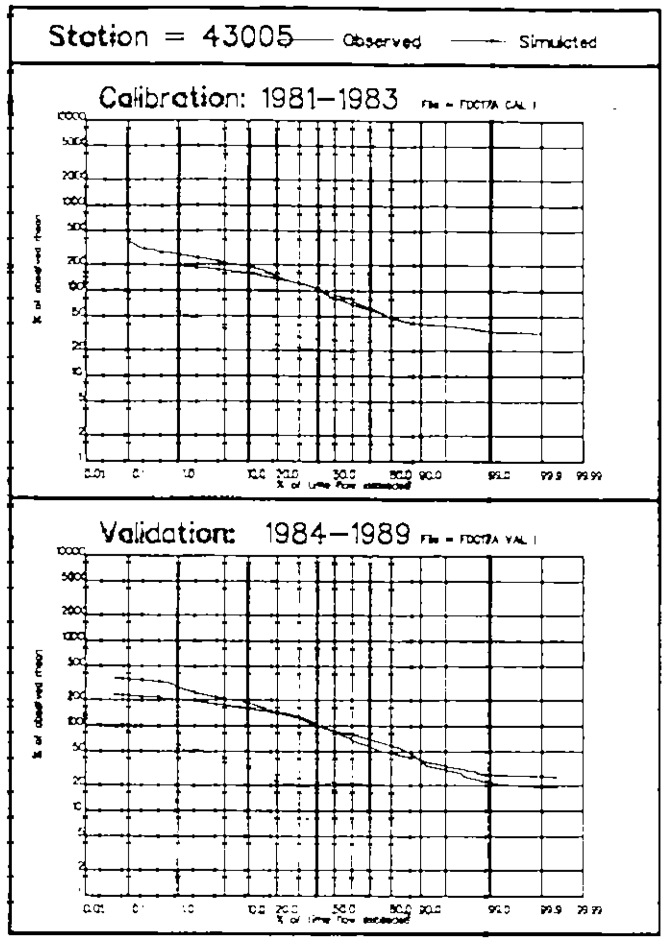


Observed and simulated flow duration curves for catchment 40007.

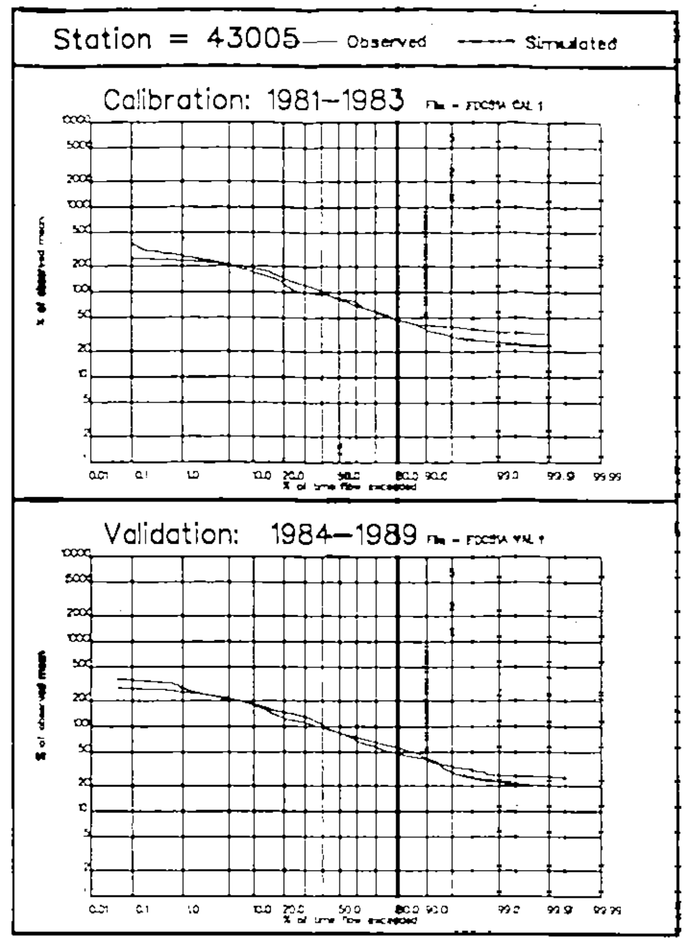


Observed and simulated flow duration curves for catchment 42003.

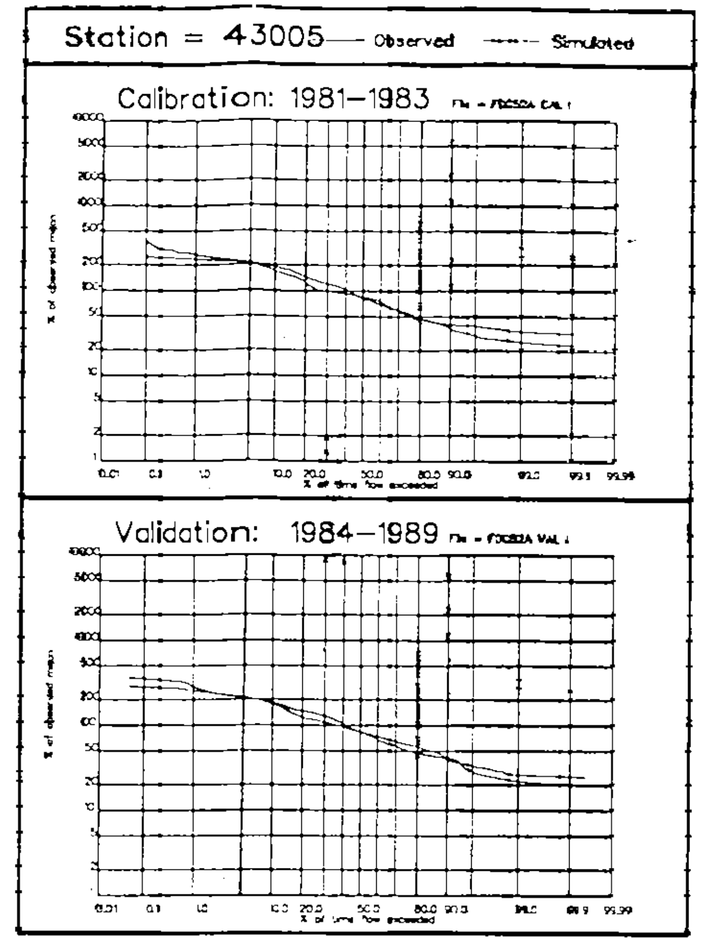
Model A



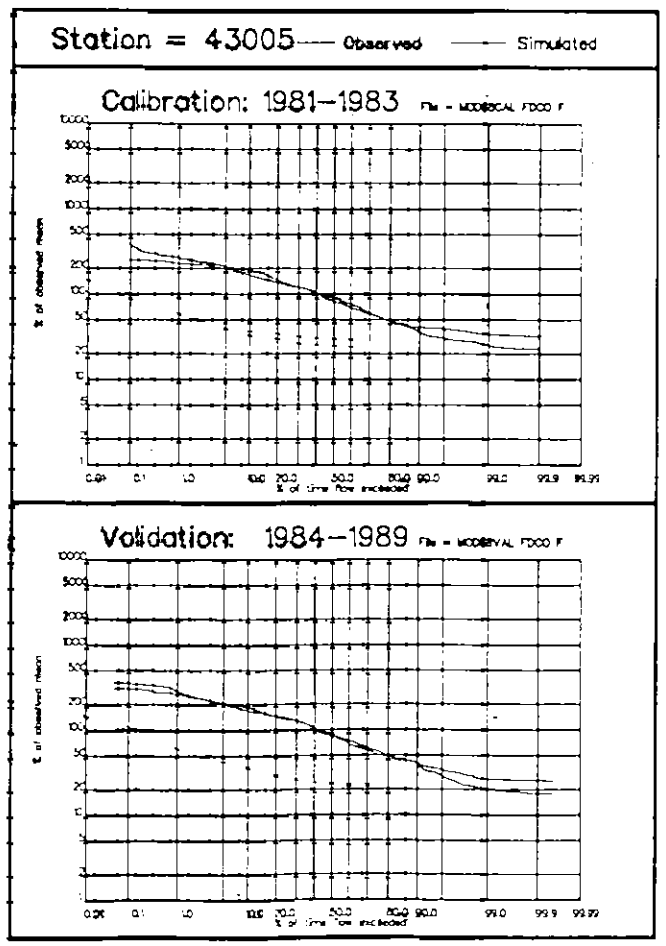
Model B1



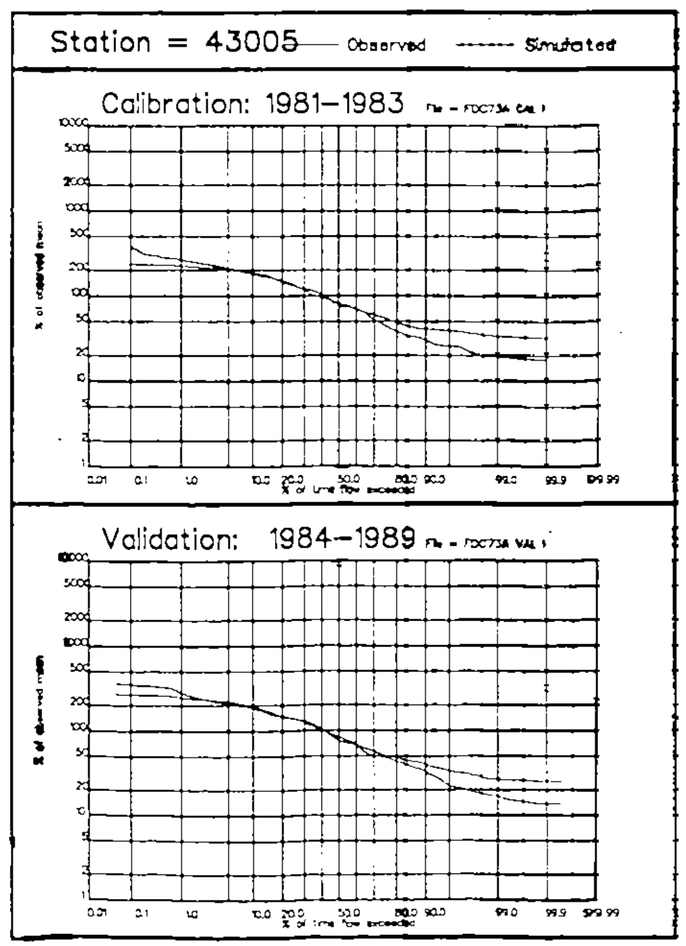
Model B2



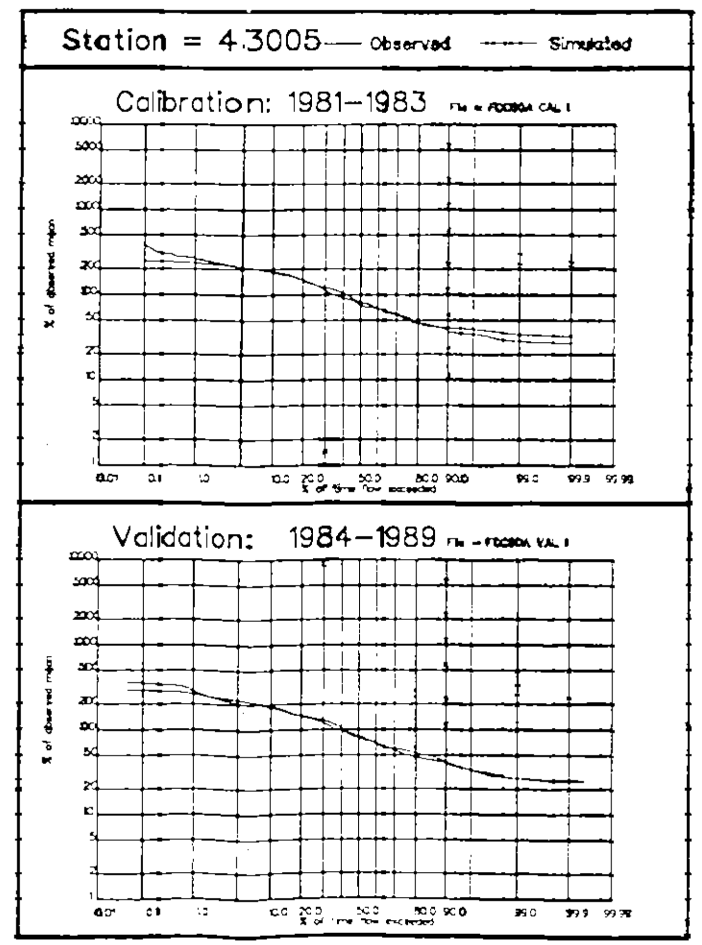
Model C



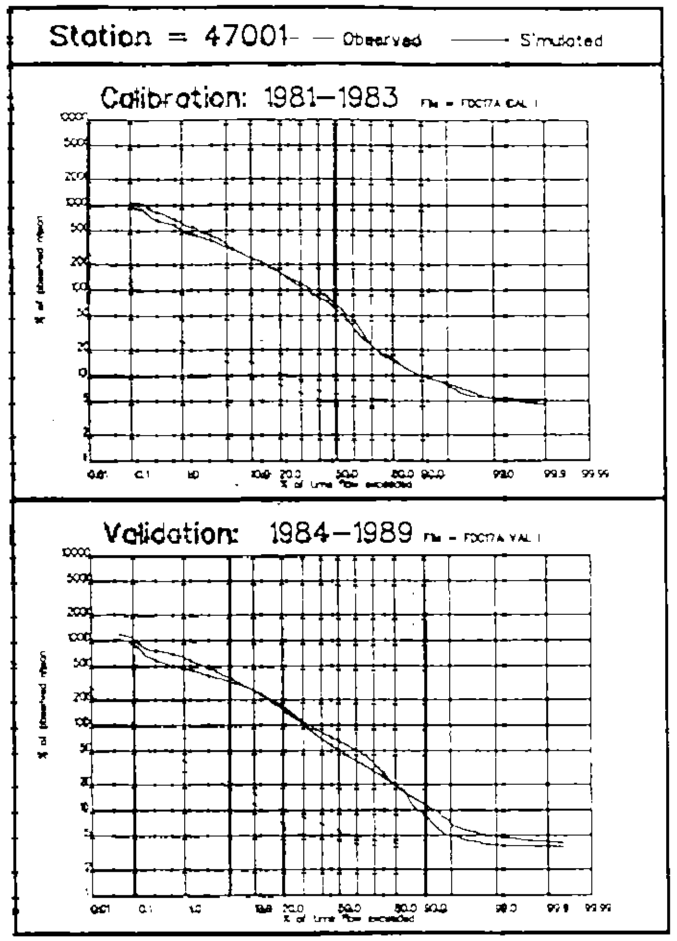
Model D1



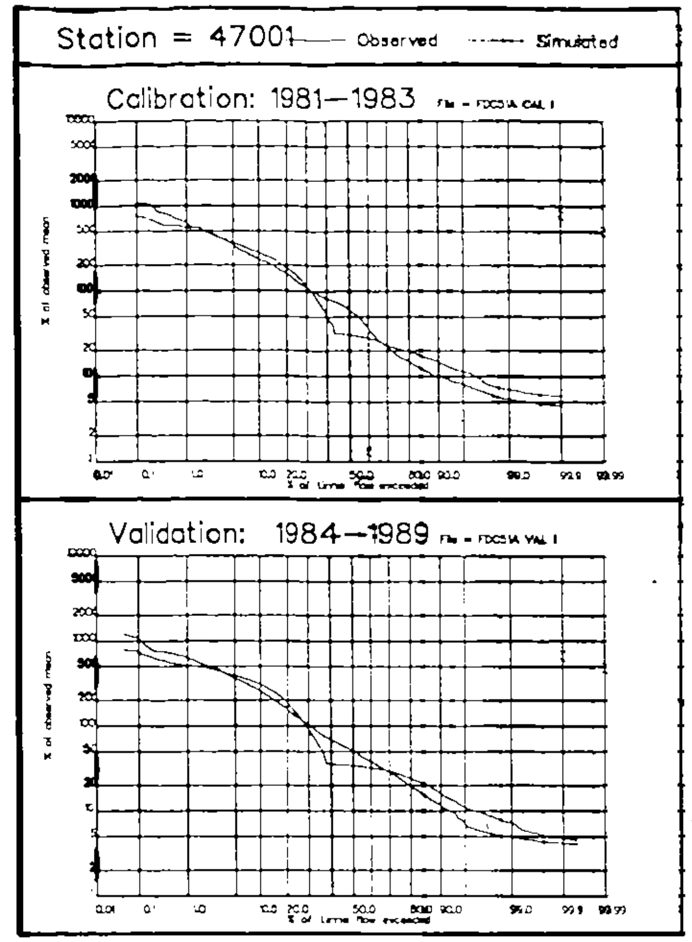
Model D2



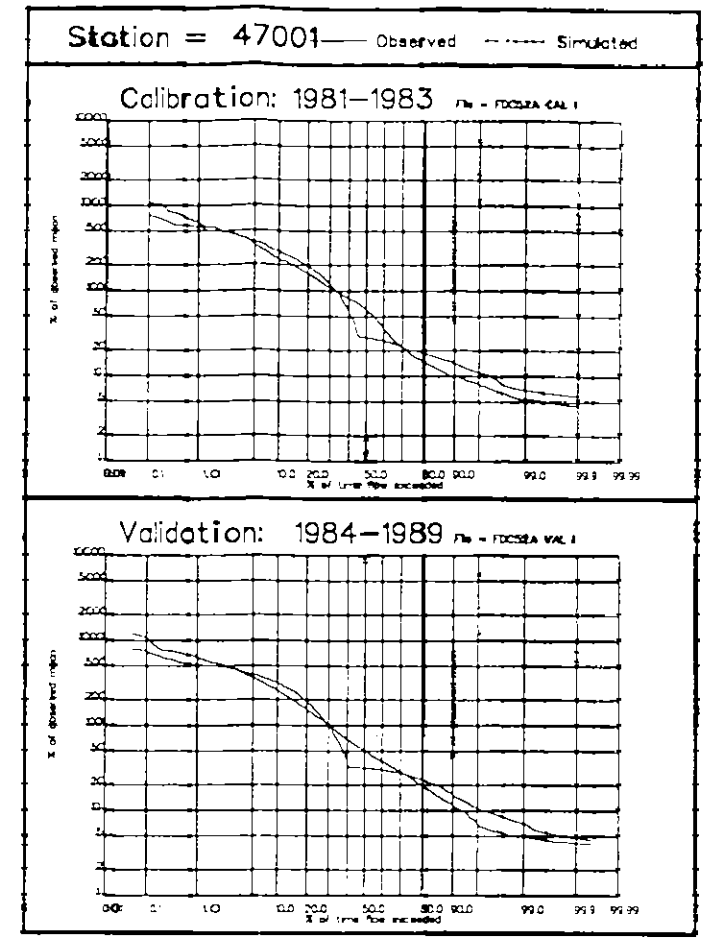
Observed and simulated flow duration curves for catchment 43005.



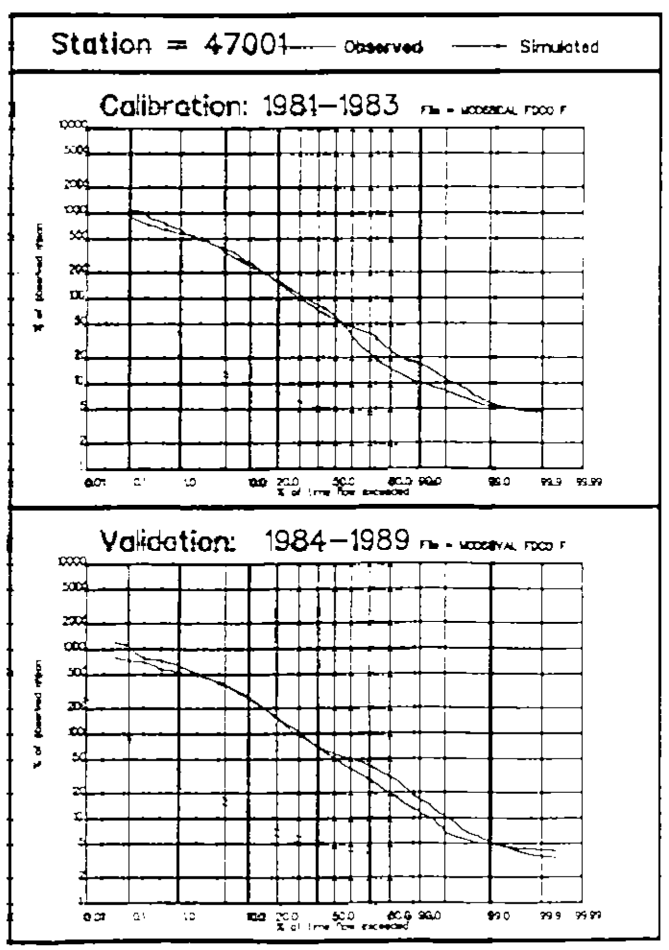
Model A



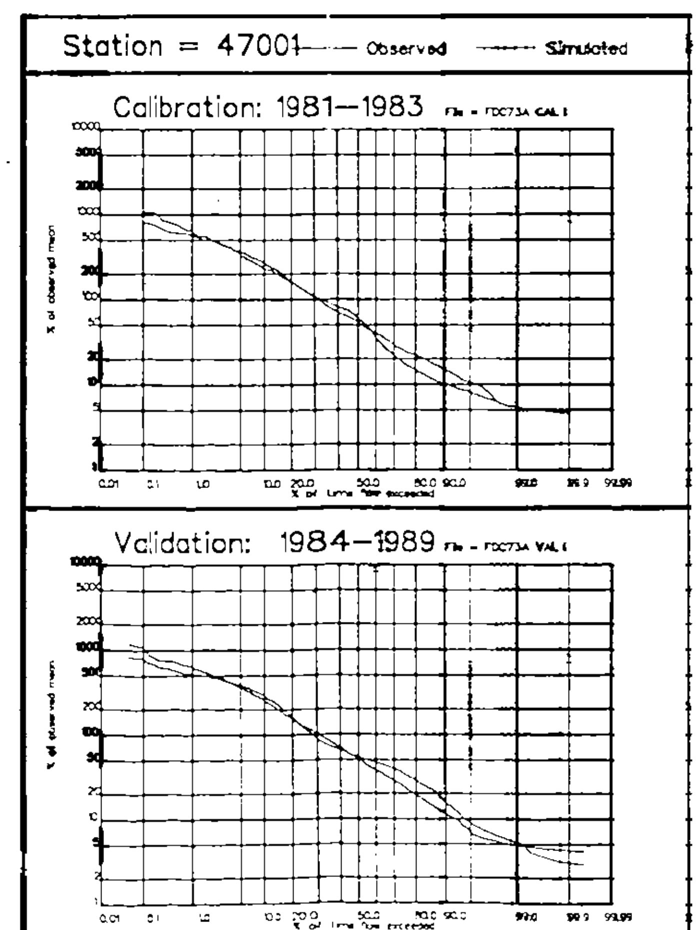
Model B1



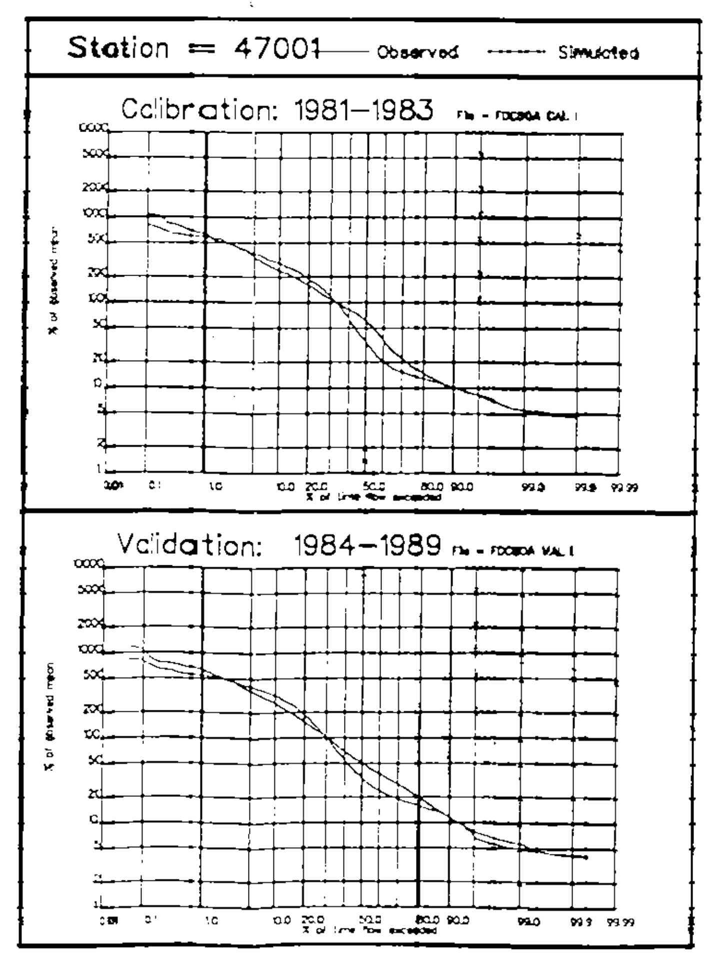
Model B2



Model C



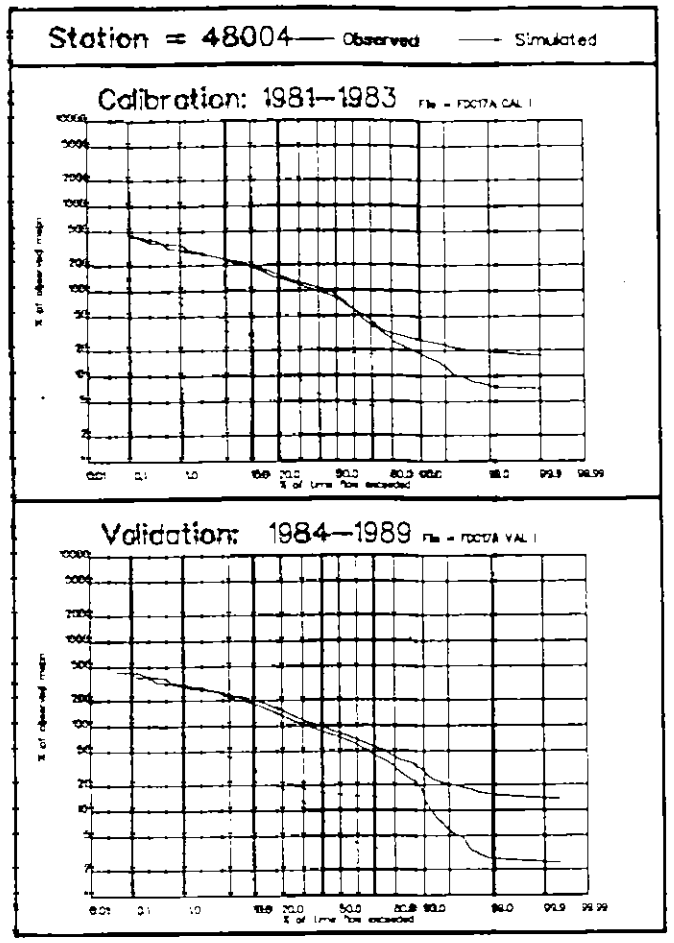
Model D1



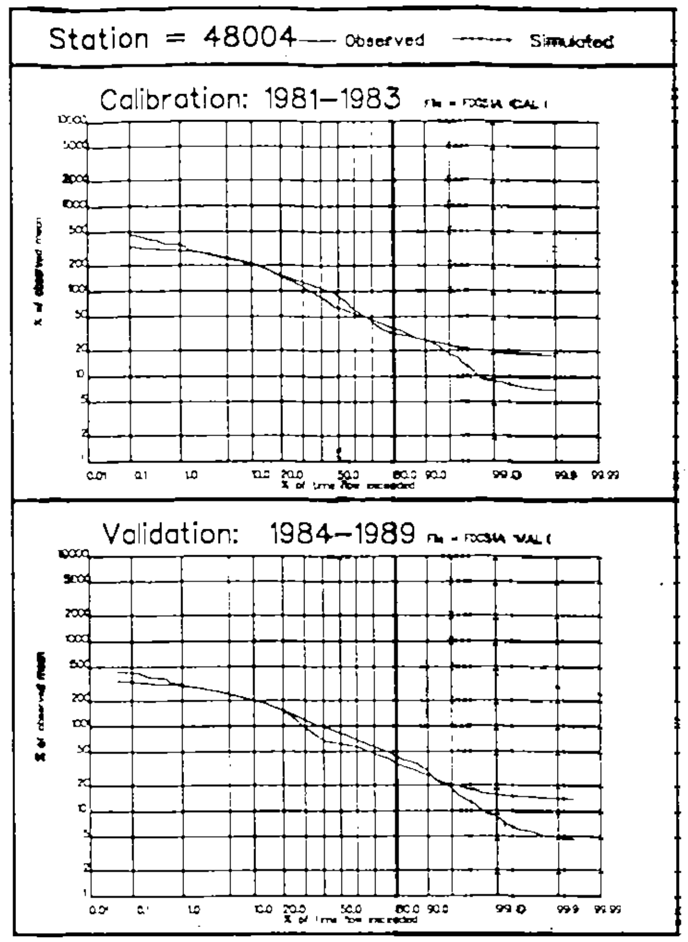
Model D2

Observed and simulated flow duration curves for catchment 47001.

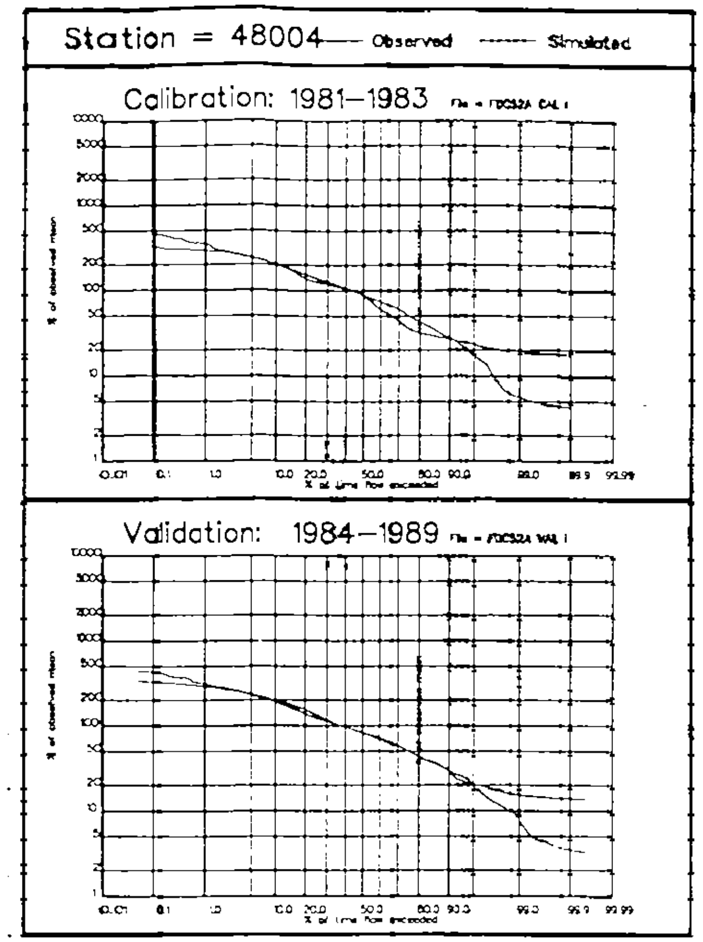
Model A



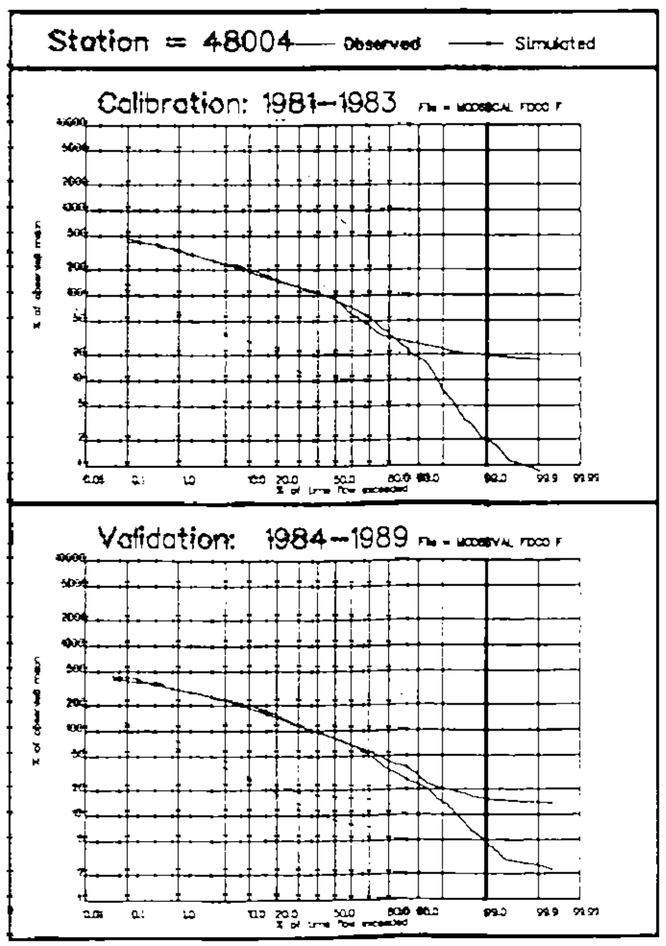
Model B1



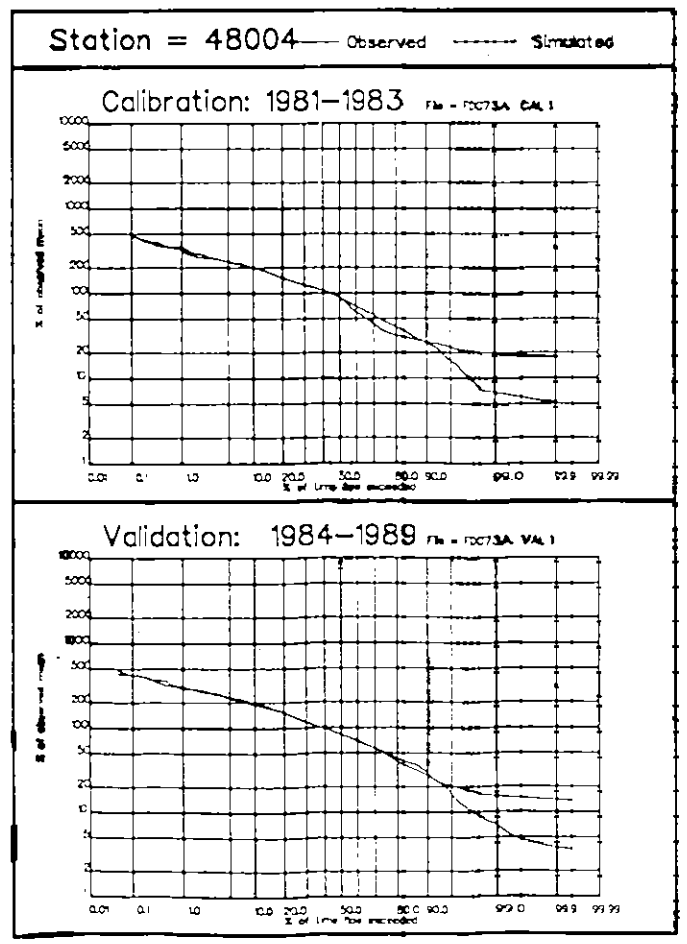
Model B2



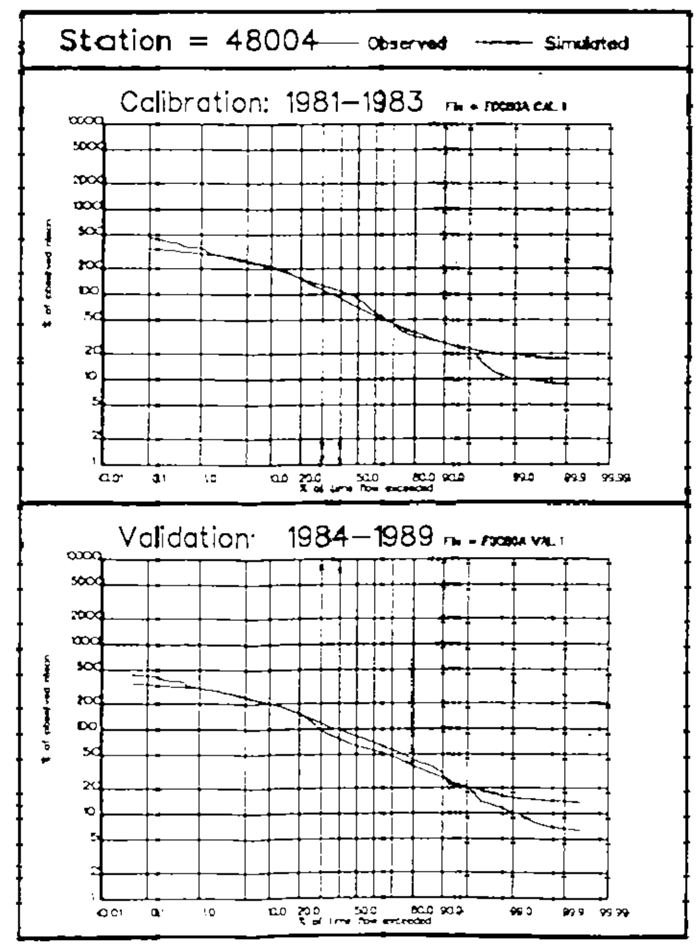
Model C



Model D1

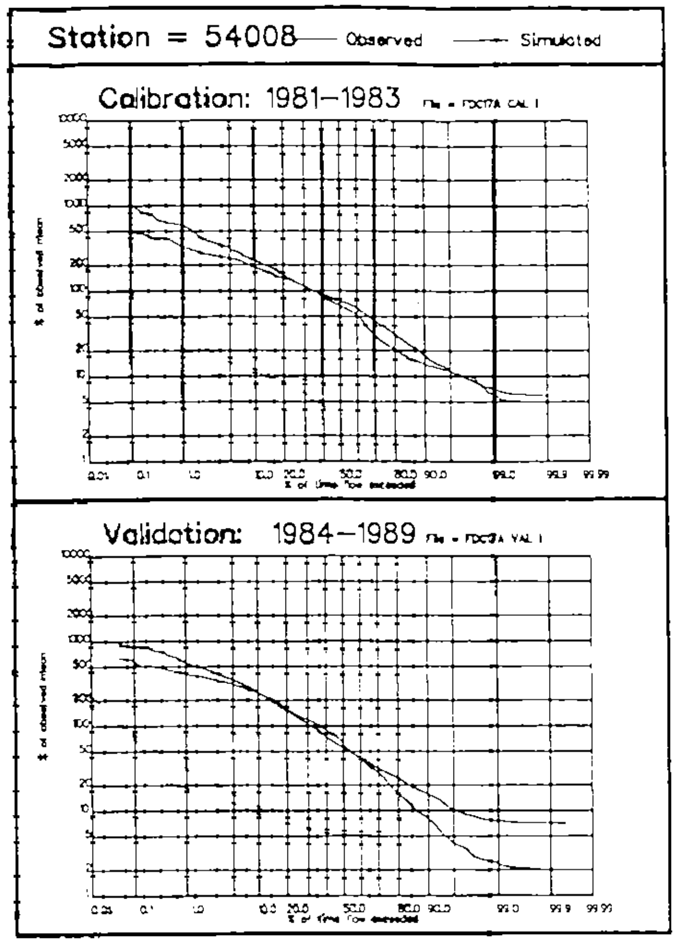


Model D2

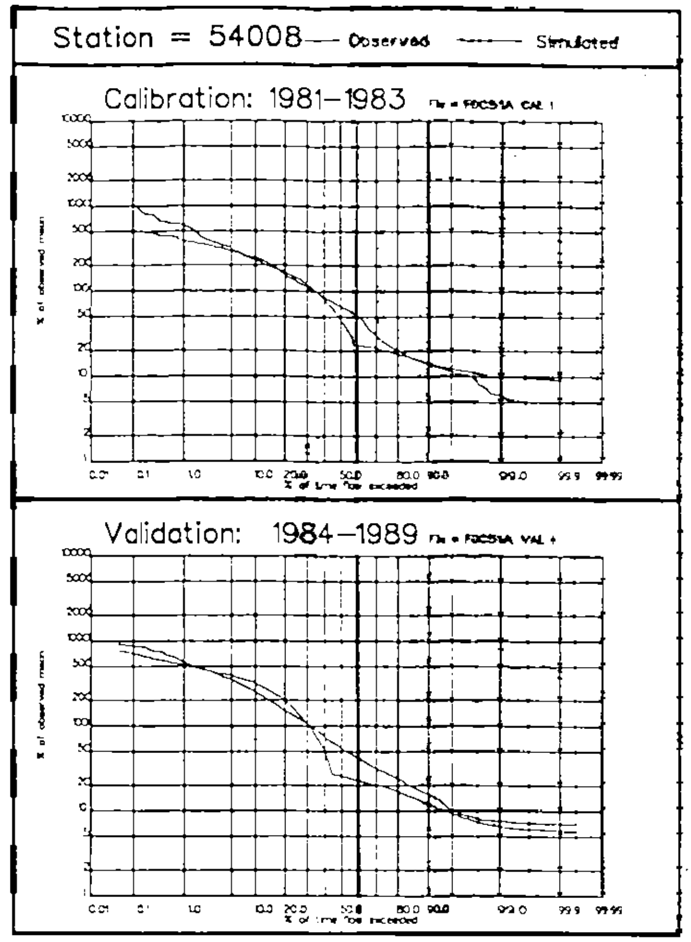


Observed and simulated flow duration curves for catchment 48004.

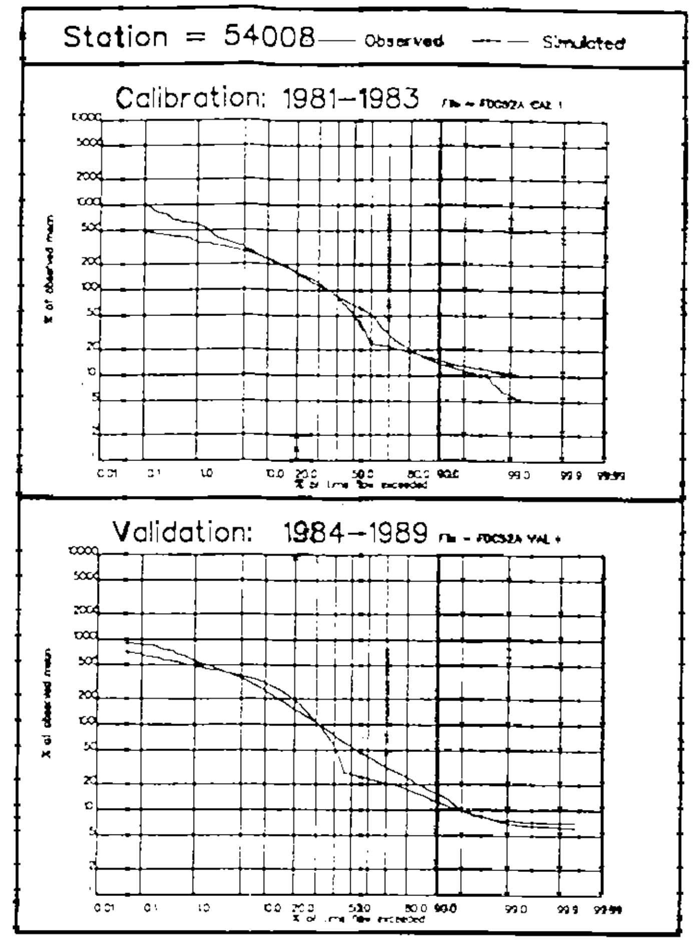
Model A



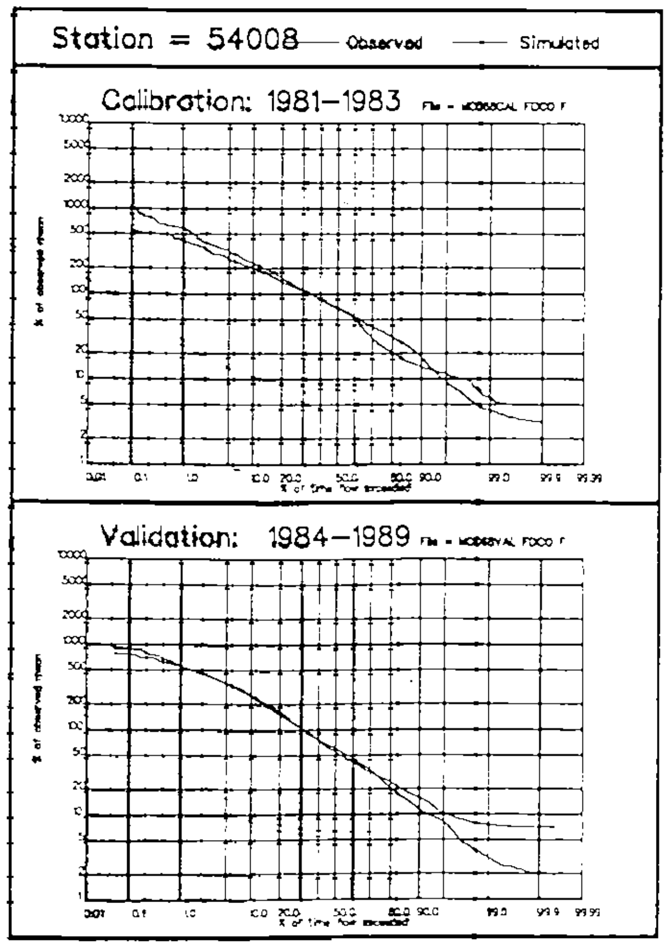
Model B1



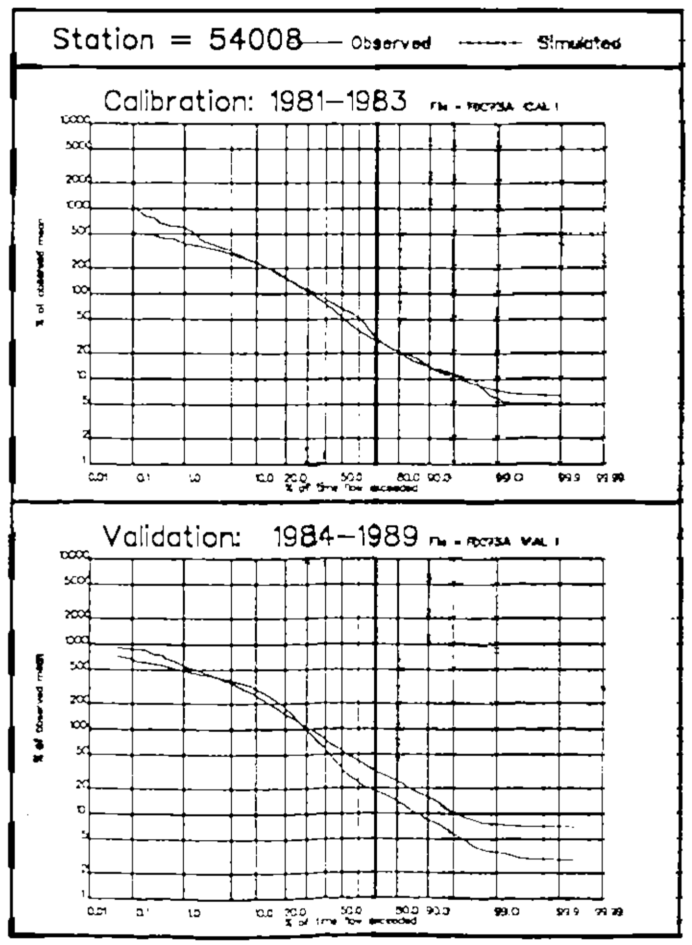
Model B2



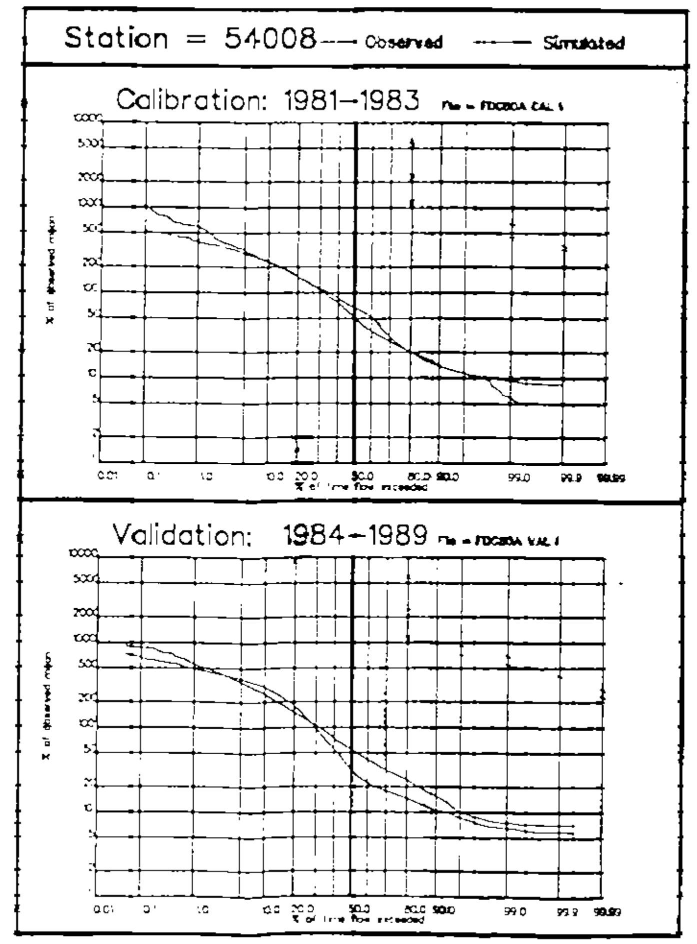
Model C



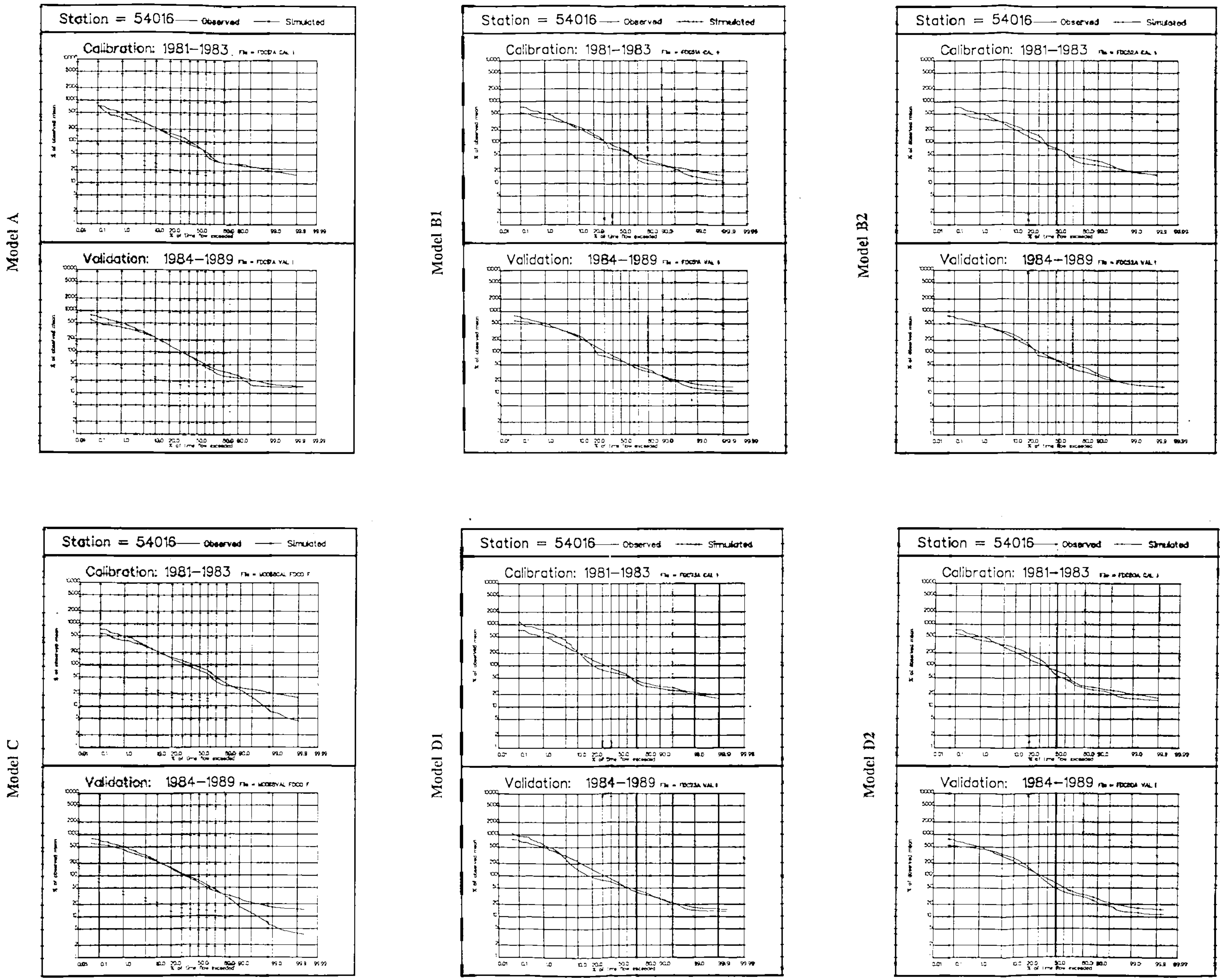
Model D1



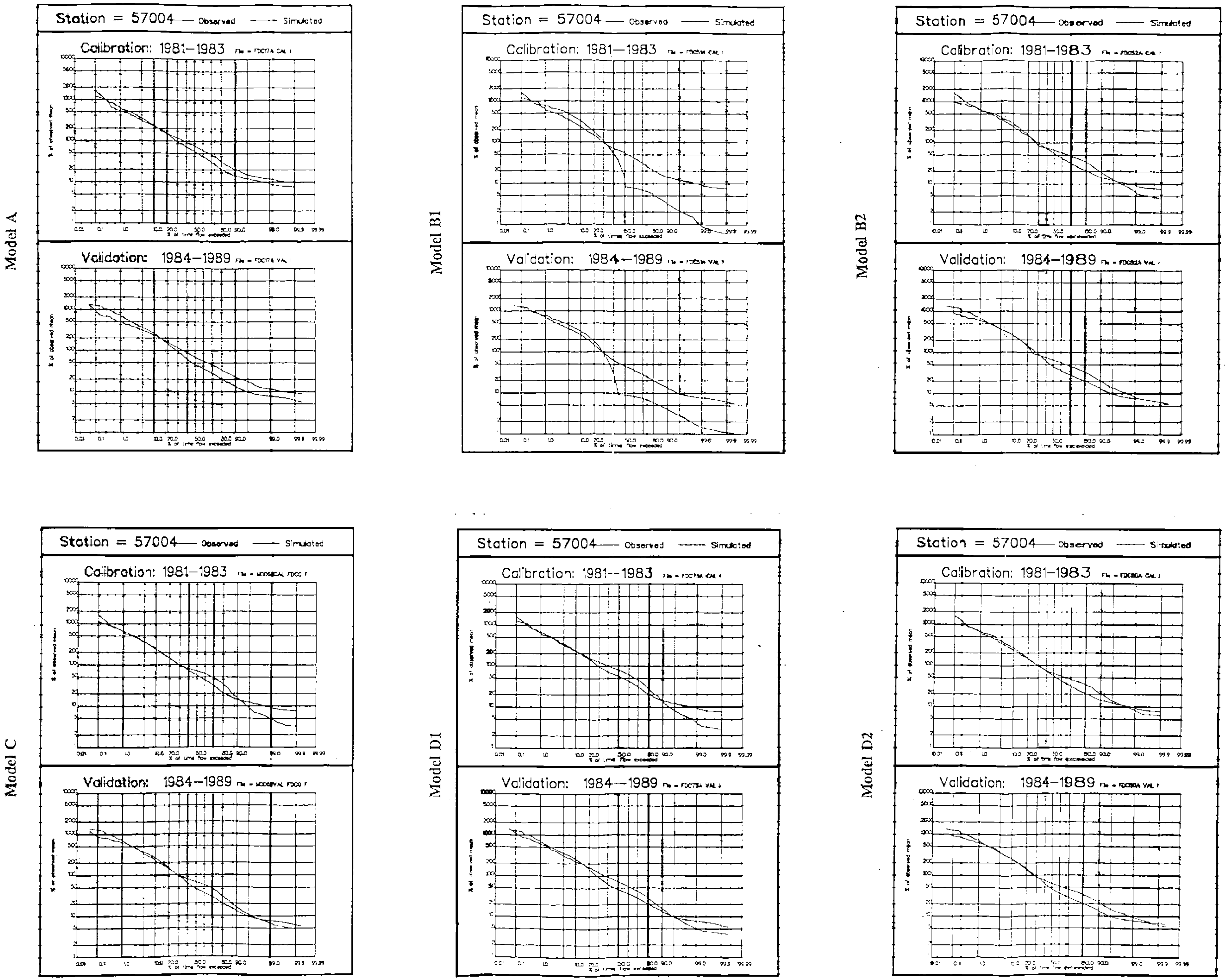
Model D2



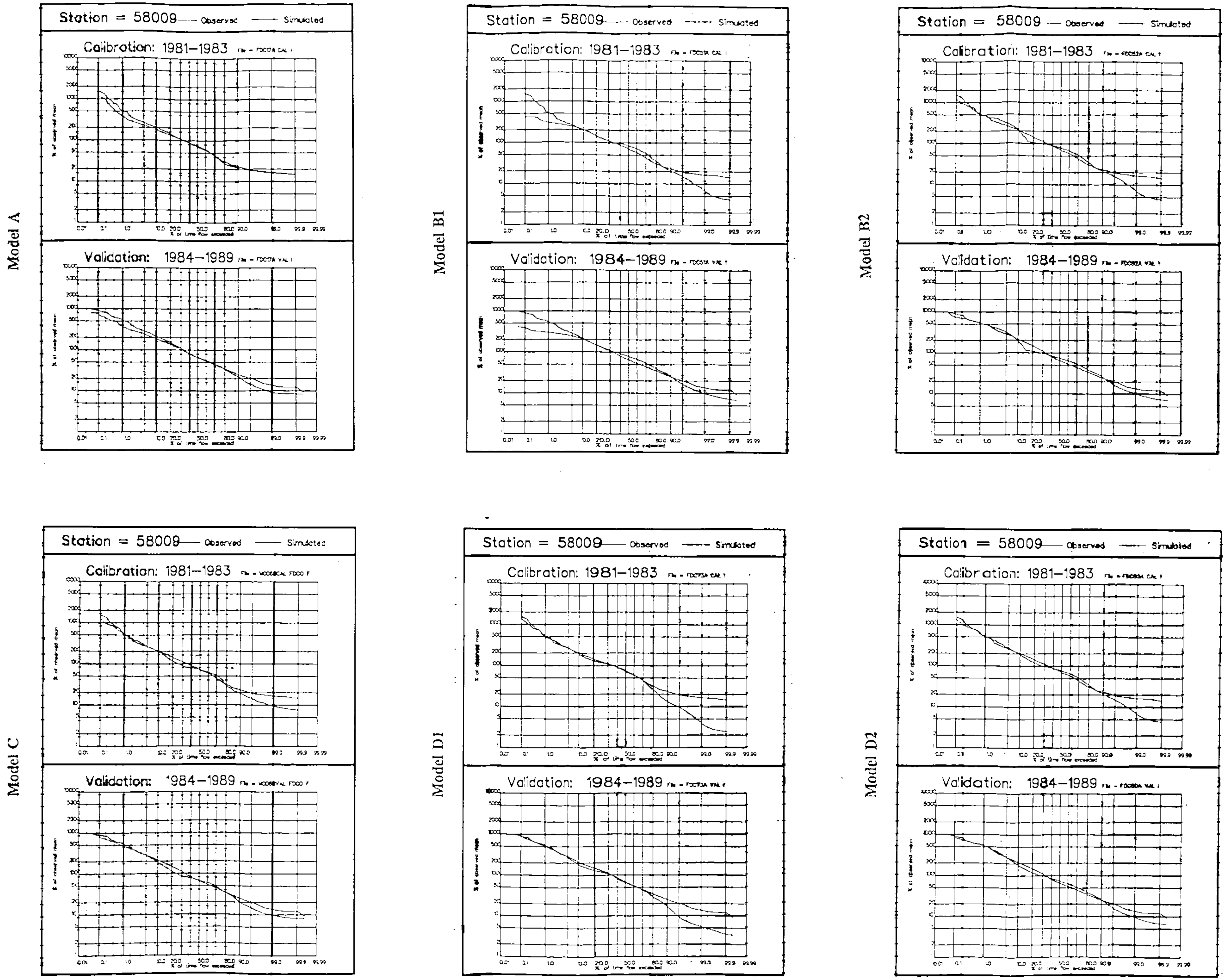
Observed and simulated flow duration curves for catchment 54008.



Observed and simulated flow duration curves for catchment 54016.

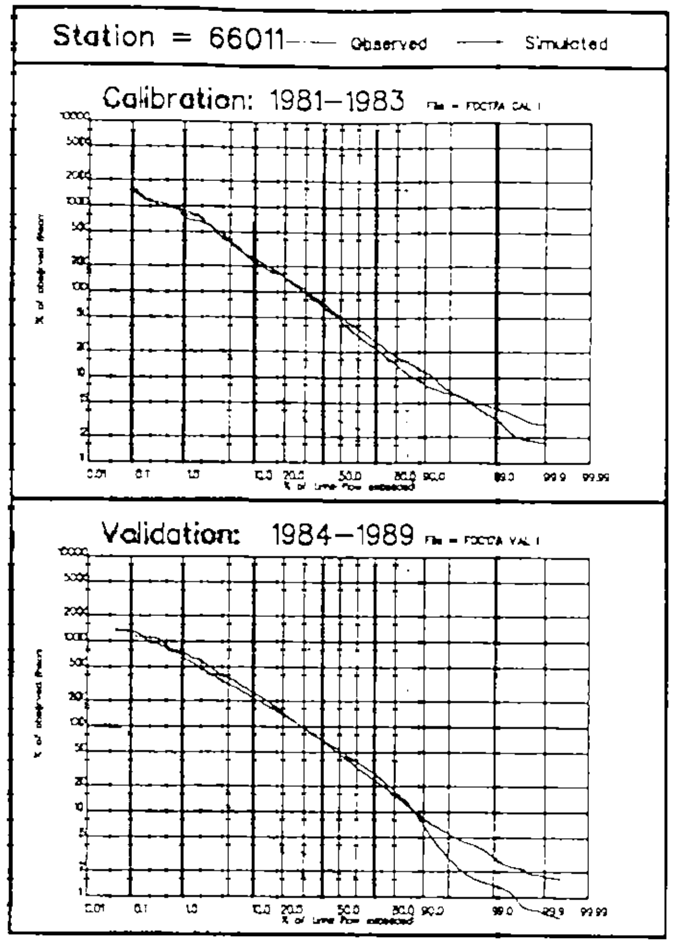


Observed and simulated flow duration curves for catchment 57004.

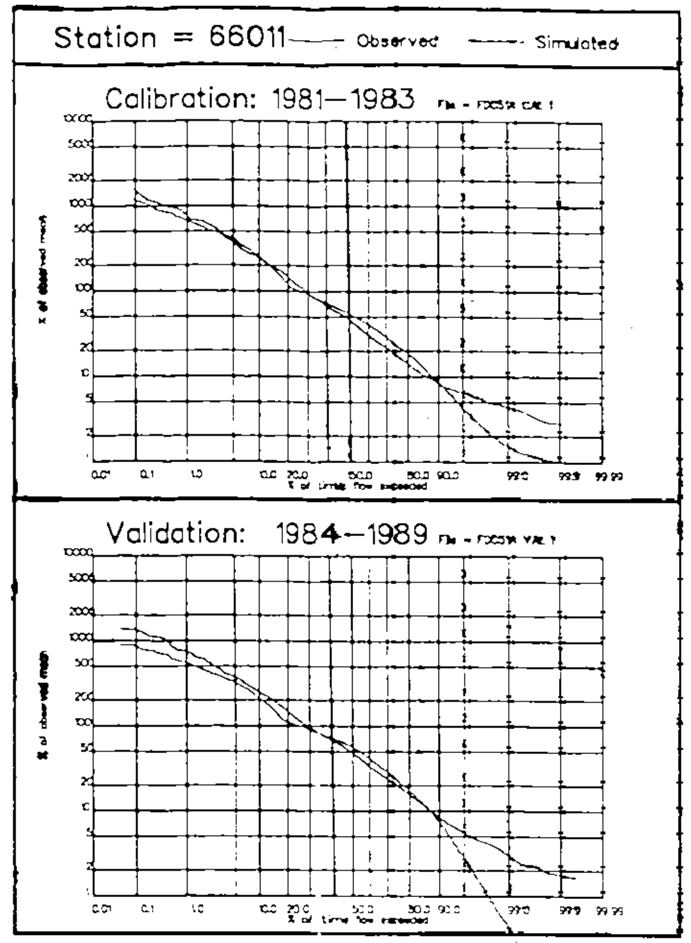


Observed and simulated flow duration curves for catchment 58009.

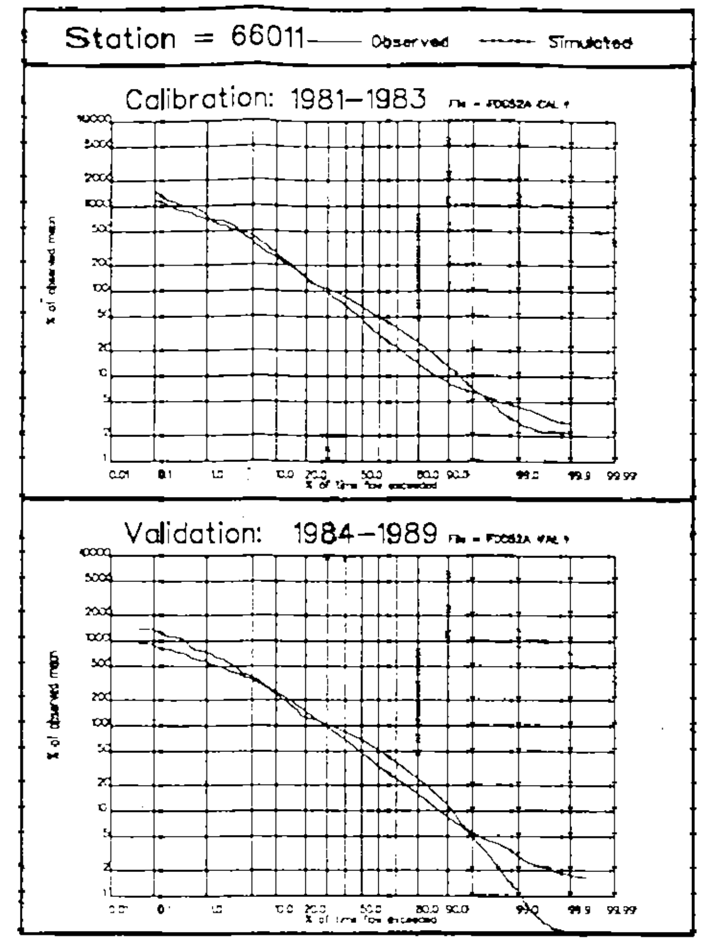
Model A



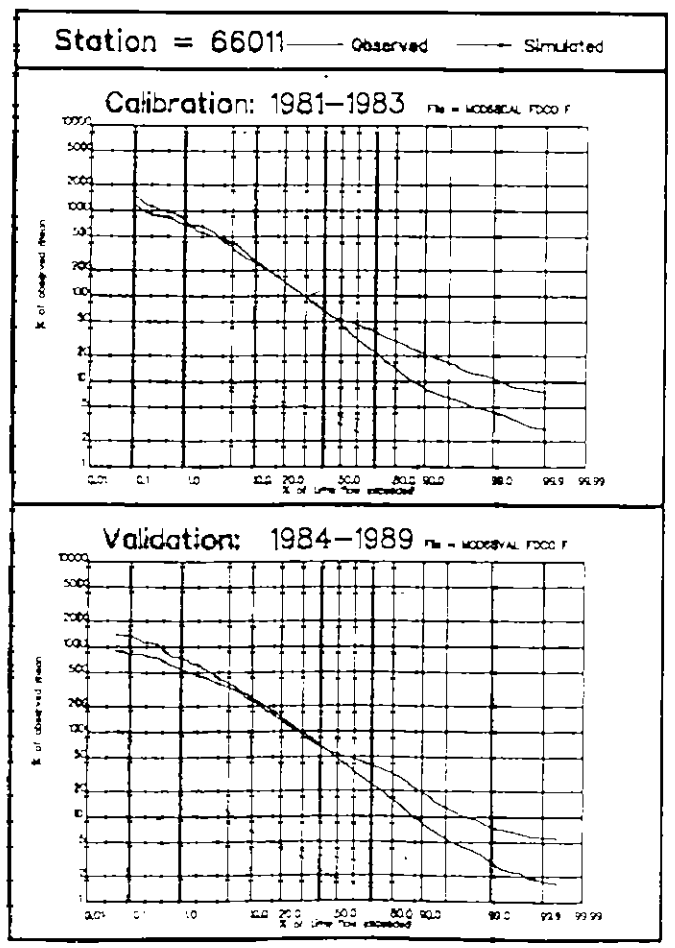
Model B1



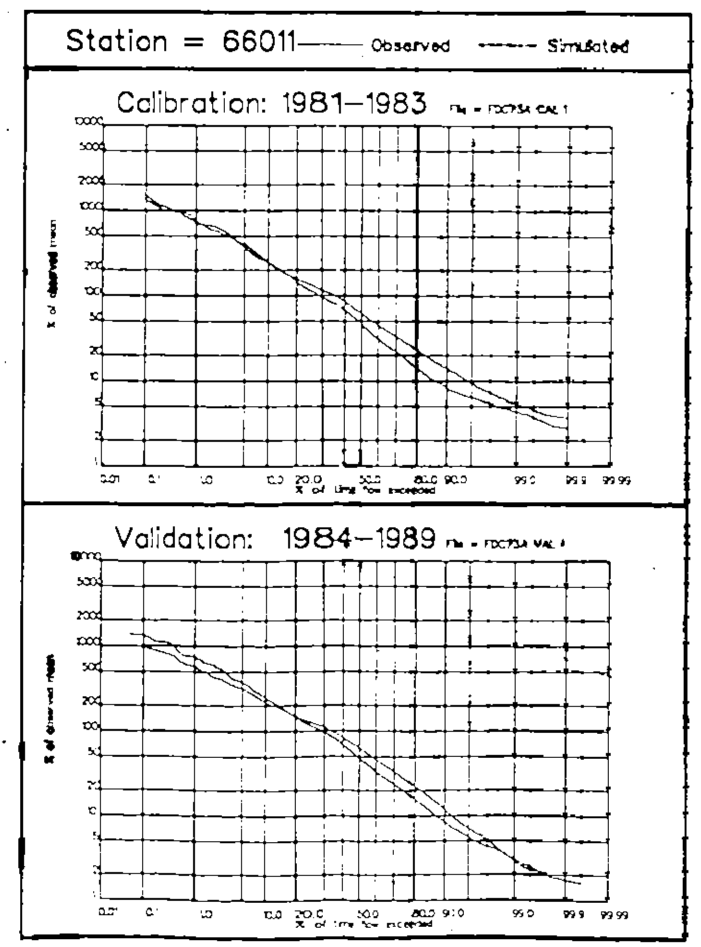
Model B2



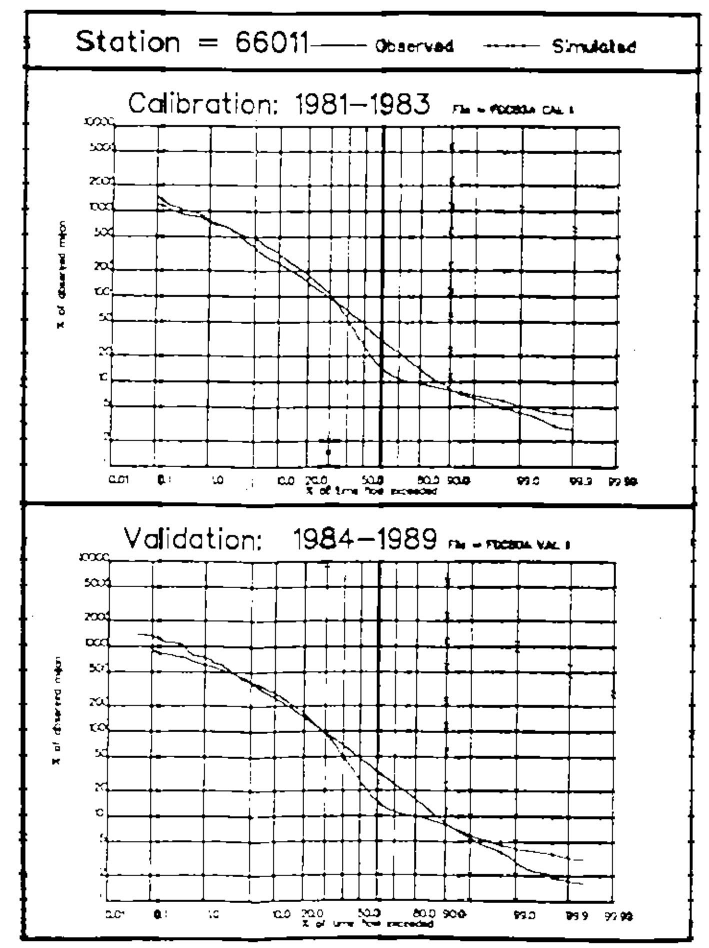
Model C



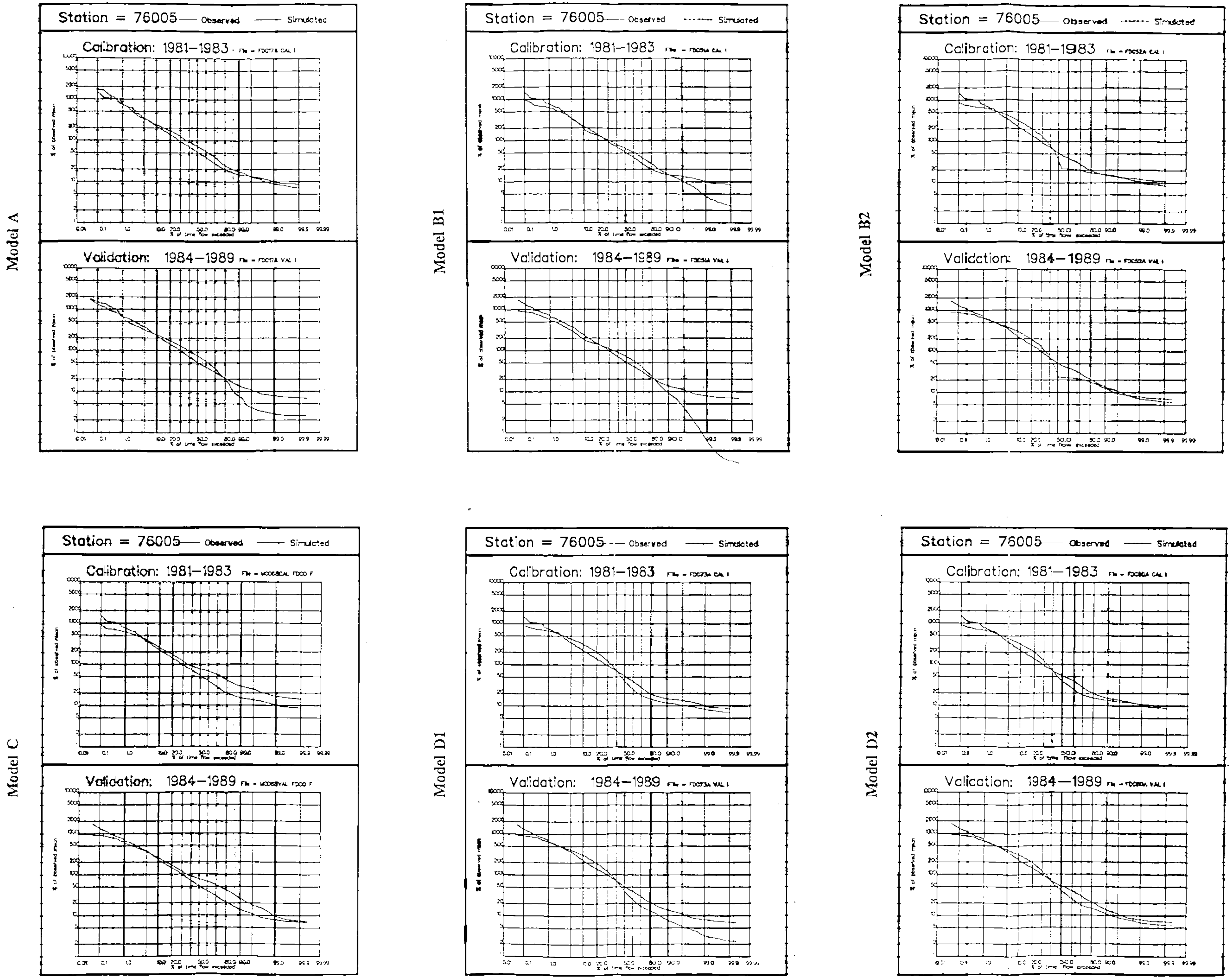
Model D1



Model D2

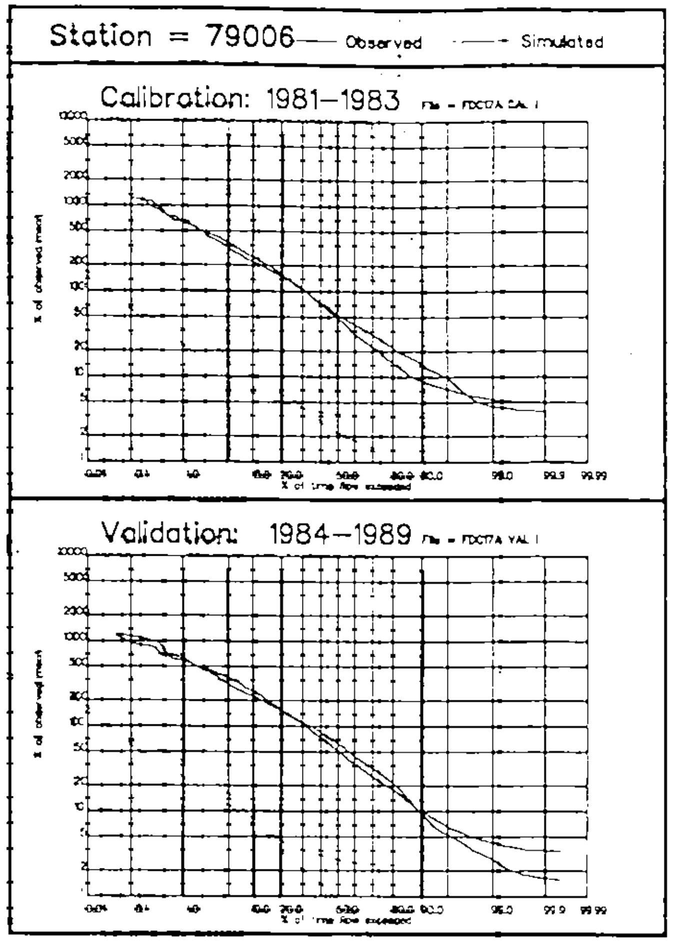


Observed and simulated flow duration curves for catchment 66011.

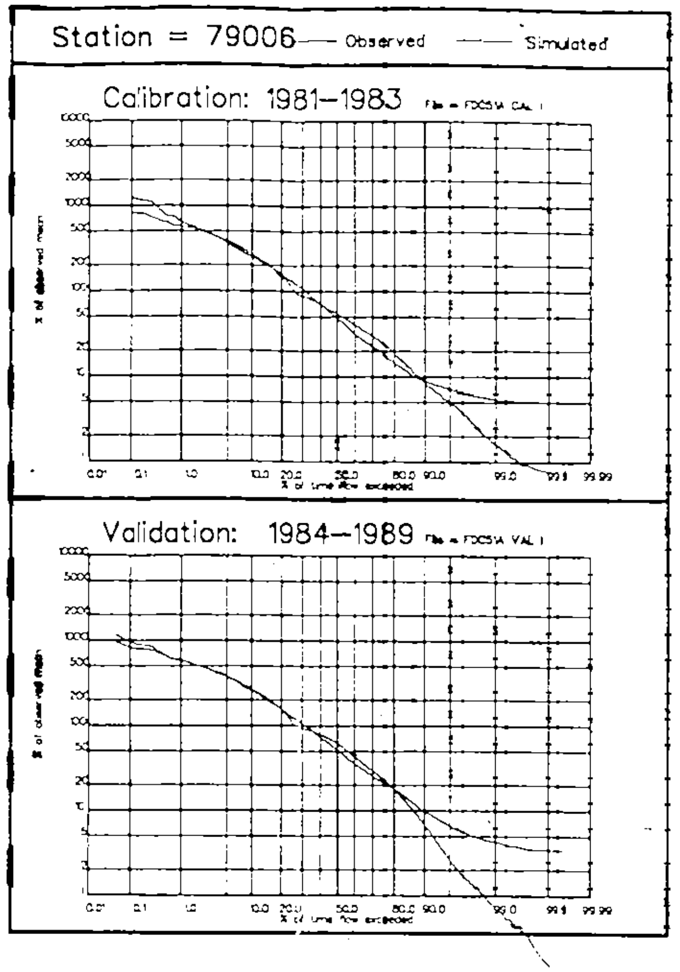


Observed and simulated flow duration curves for catchment 76005.

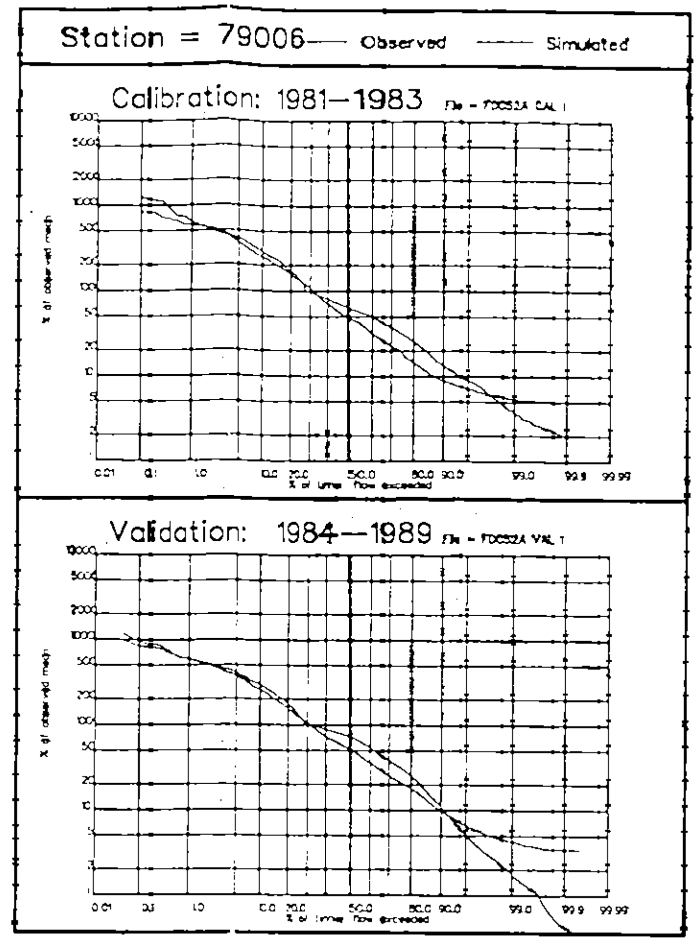
Model A



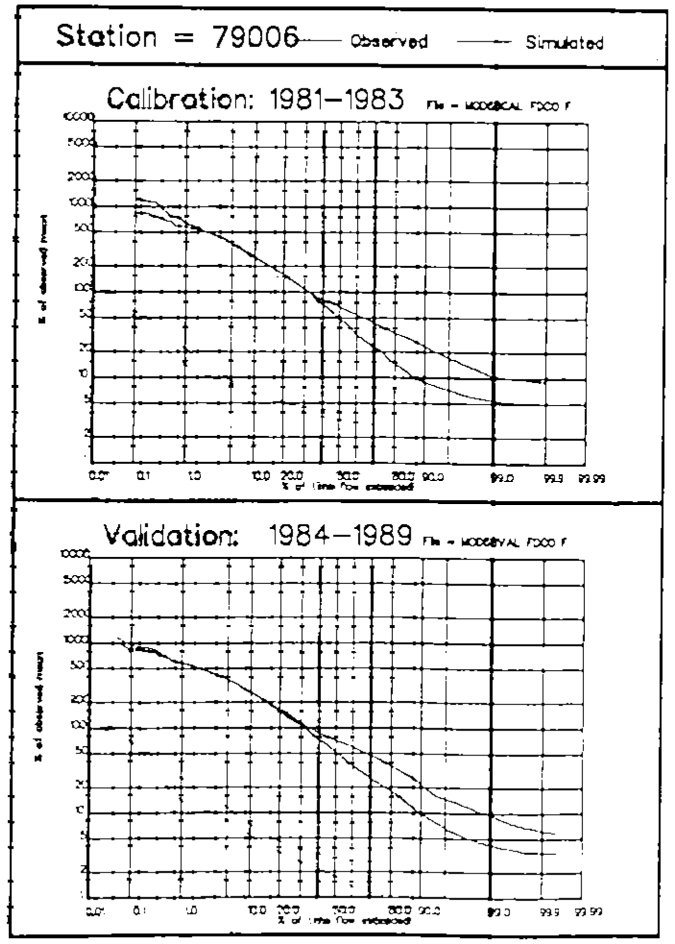
Model B1



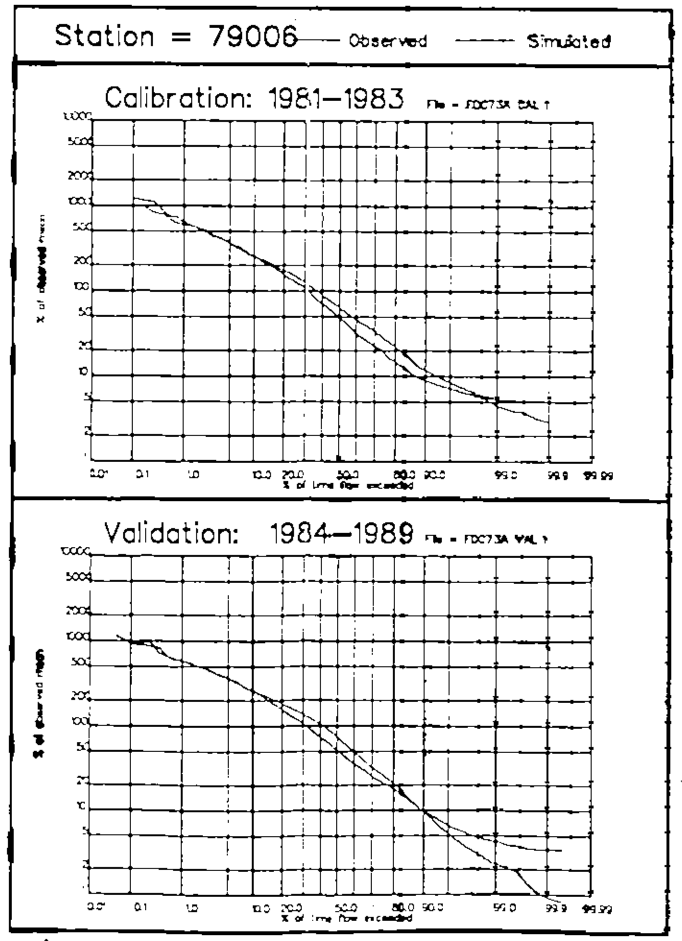
Model B2



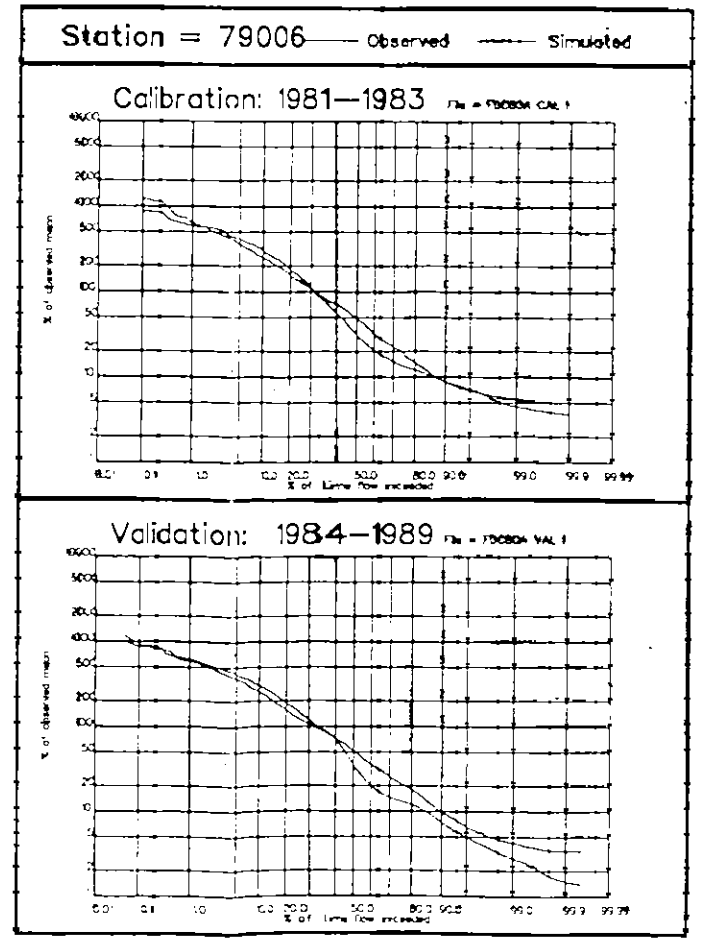
Model C



Model D1



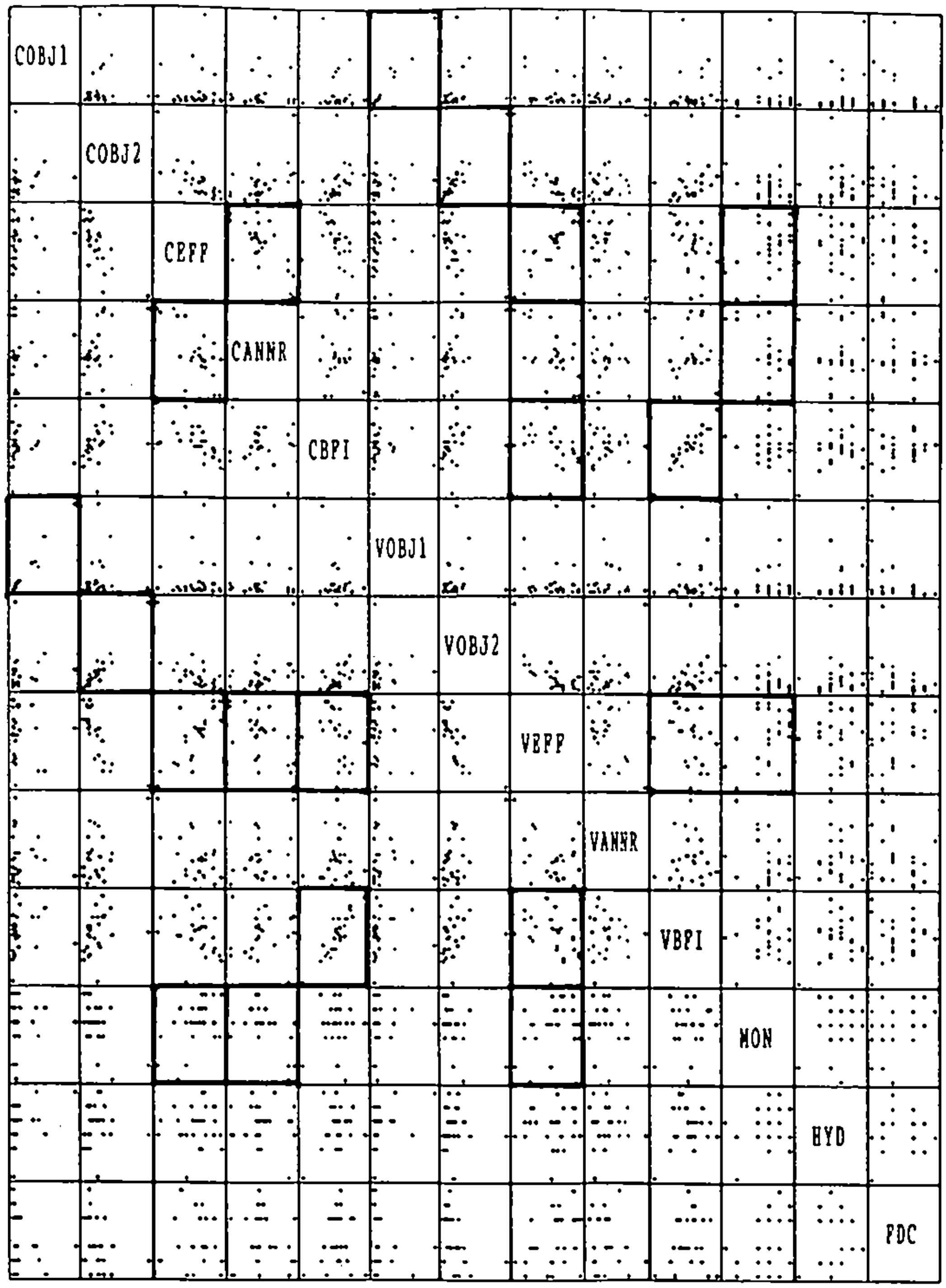
Model D2



Observed and simulated flow duration curves for catchment 79006.

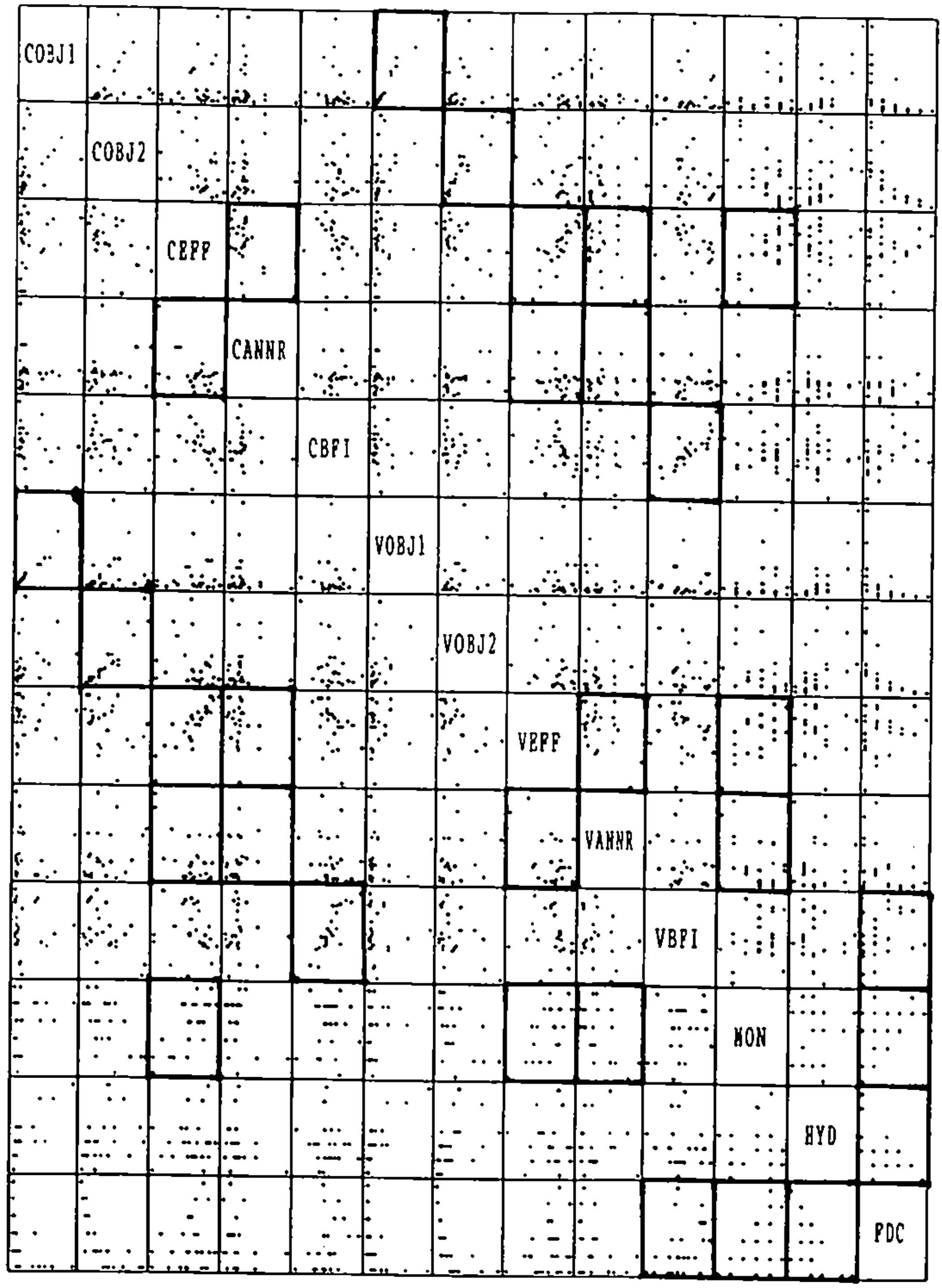
Appendix B

B.1 Comparative plots of model performance assessment criteria



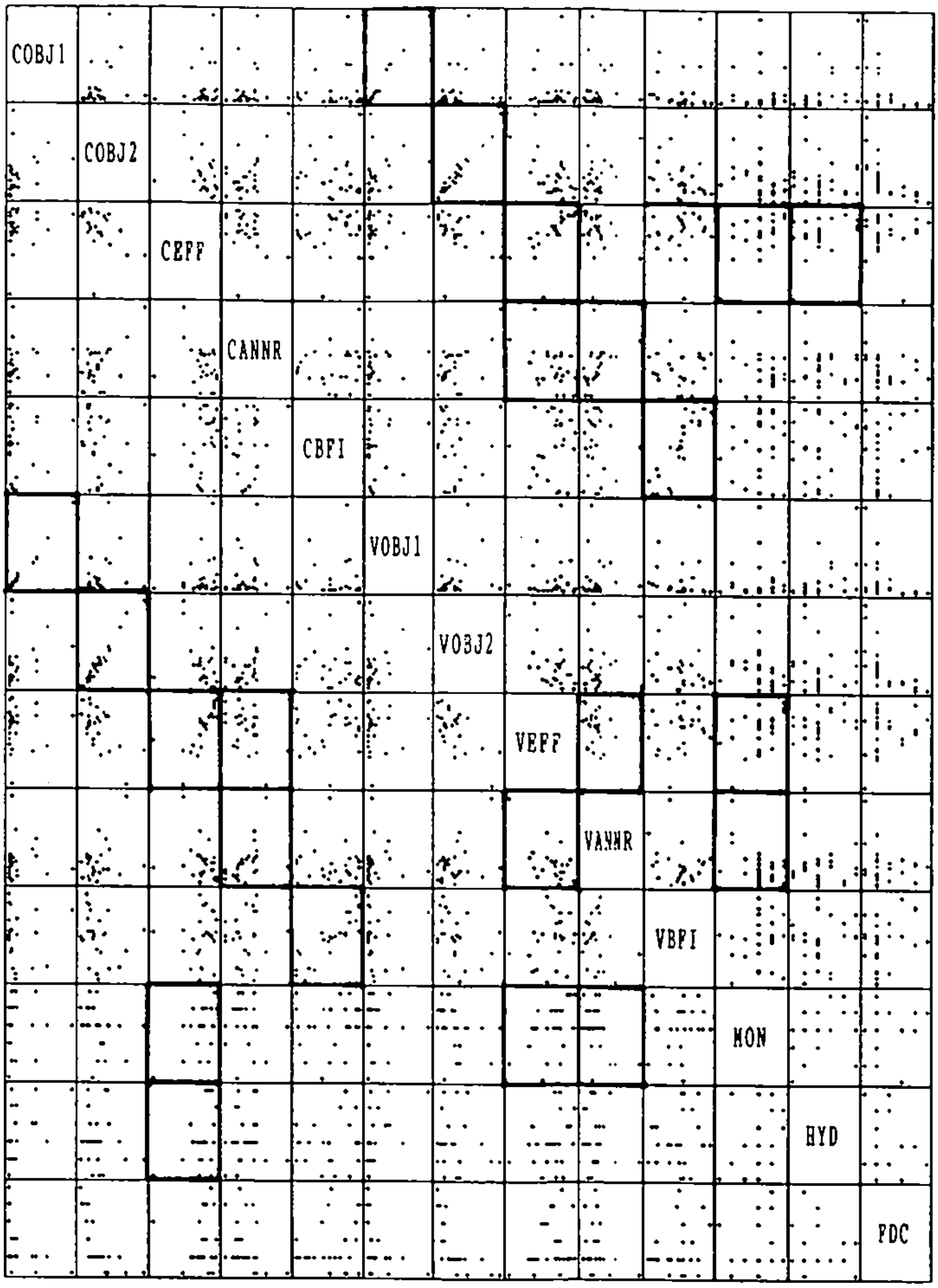
(n=25; highlighted boxes indicate correlations significant at the 95 % level)

Comparative plots of model performance assessment criteria for model A.



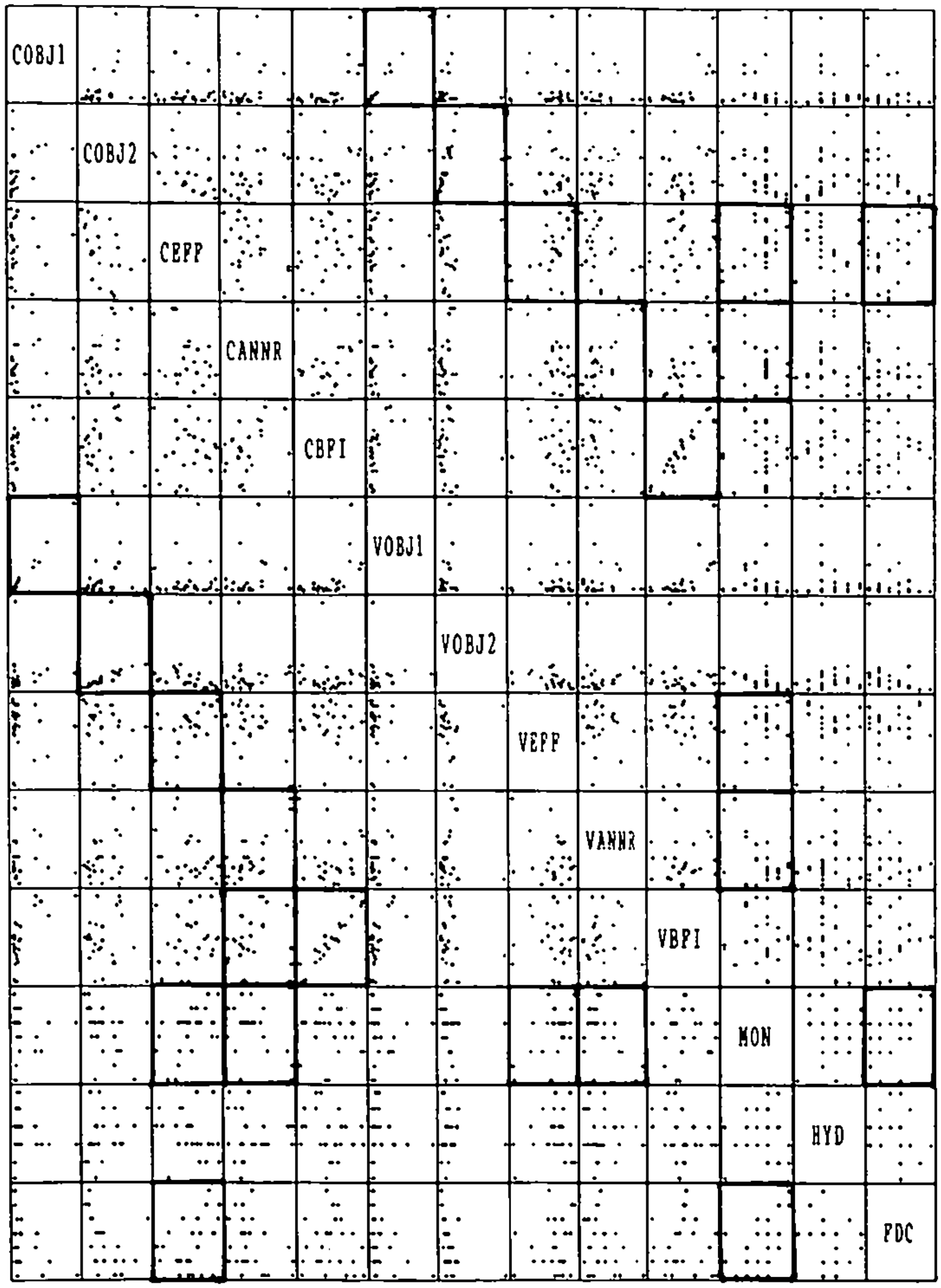
(n=25; highlighted boxes indicate correlations significant at the 95 % level)

Comparative plots of model performance assessment criteria for model B1.



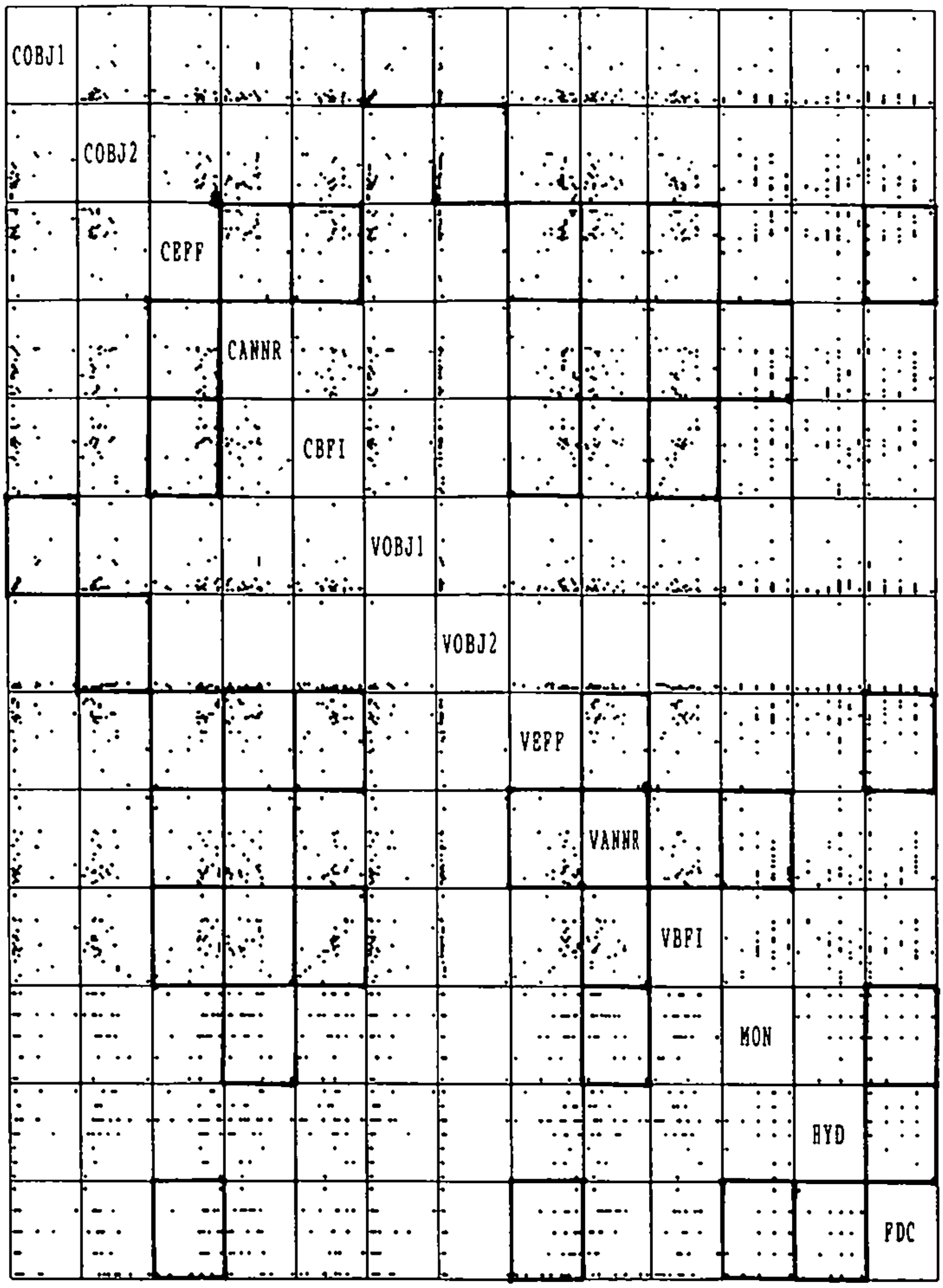
(n=25; highlighted boxes indicate correlations significant at the 95 % level)

Comparative plots of model performance assessment criteria for model B2.



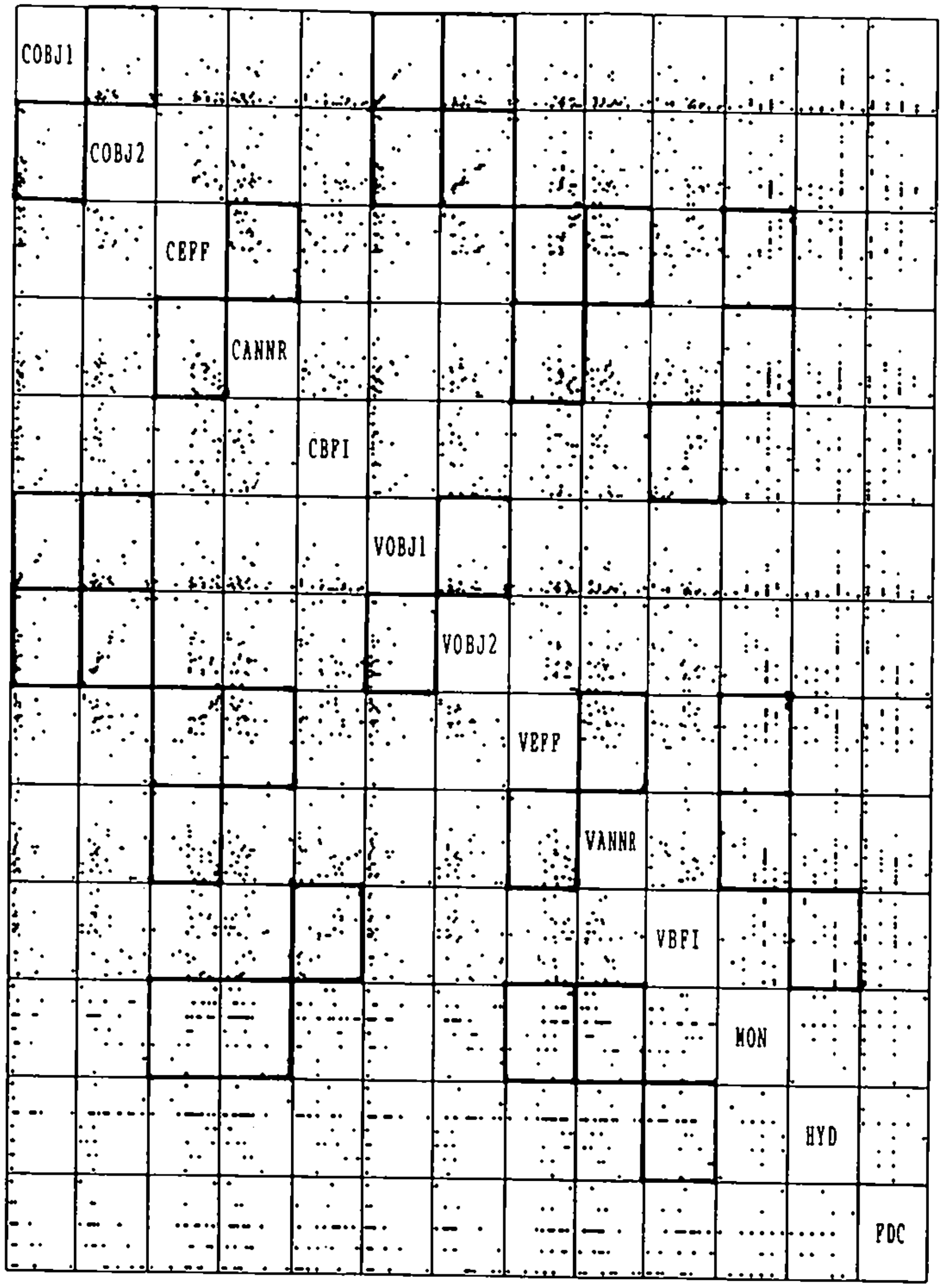
(n=25; highlighted boxes indicate correlations significant at the 95 % level)

Comparative plots of model performance assessment criteria for model C.



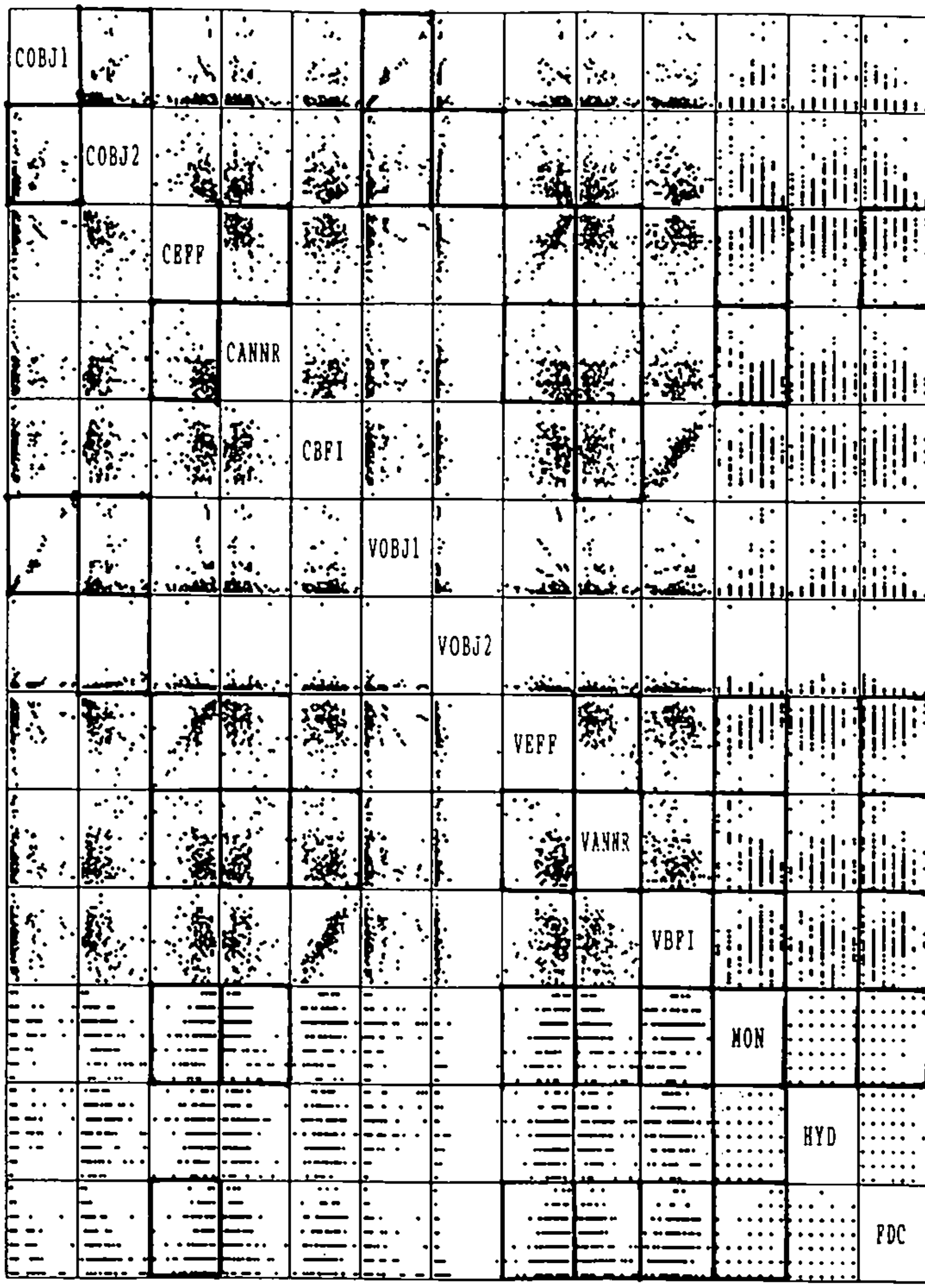
(n=25; highlighted boxes indicate correlations significant at the 95 % level)

Comparative plots of model performance assessment criteria for model D1.



(n=25; highlighted boxes indicate correlations significant at the 95 % level)

Comparative plots of model performance assessment criteria for model D2.



(n=150; highlighted boxes indicate correlations significant at the 95 % level)

Comparative plots of model performance assessment criteria for complete dataset.