SEPIK River Flood Forecasting

B. CAPPELAERE[(1)] and M. VIROBO[(2)]

(1) **ORSTOM**, Laboratoire d'Hydrologie, Montpellier, France

(2) **Bureau of Water Resources**, Port Moresby, Papua New Guinea

27 March, 1998

SEPIK River Flood Forecasting

Summary of study performed at ORSTOM's Hydrology Lab.
by B. Cappelaere & M. Virobo
(march 1998)

1. Context of study

The objectives of this study were twofold:

- analyze the possibilities of flood prediction on the Sepik River, and establish a flood forecasting model,
- provide to M. Virobo training with the tools used for model development.

2. Flood forecasting for the Sepik River

2.1 Data and problem analysis

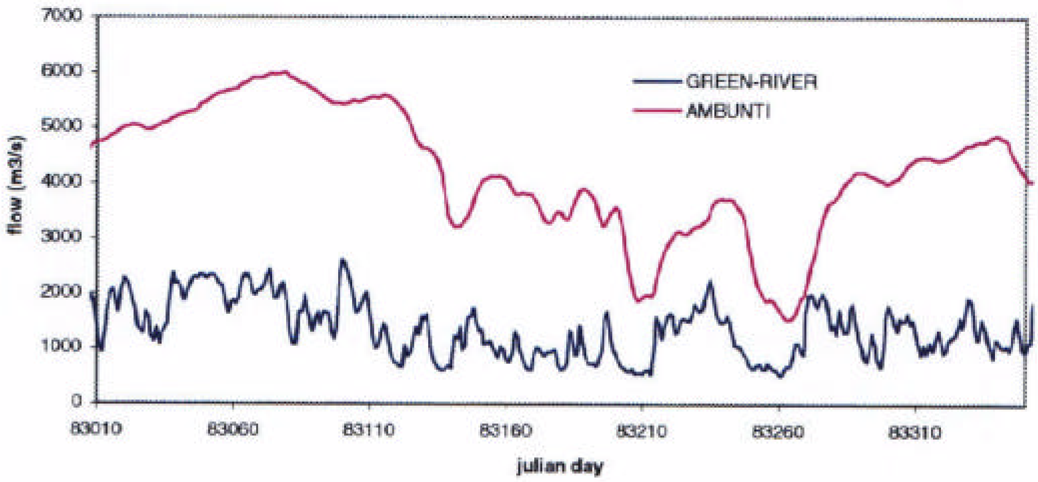
Data is available for the Sepik River at three stations along the main water course, namely Green-River, Ambunti, and Angoram. A flood forecasting study requires to have simultaneous data available for several years of record at the hydrometric stations. This excludes the use of the Angoram station, for which only a very short, recent record is available. On the contrary, concomittant records for Green-River and Ambunti extend over a period of some 28 years (1970 to 1997), although these records are marked by significant gaps with missing data, especially since the late 1980's. The target for this study was therefore defined as issueing of flood forecast at the downstream, namely Ambunti station.

A first step of the work was to analyze the data for possible suspicious or unreliable subsets of data. Hence a small percentage of the overall data set was discarded from our study. The data set includes periods of missing data at one station or the other, or both. The data gaps are managed through a dynamic sample construction procedure in the statistical method used for model development (see §2.2).

Since the watershed controlled by the upstream, Green-River station only accounts for less than a quarter of that at Ambunti (9500 km² against 40922 km²), data representing the intermediate watersheds was looked for. Only the Old-Base-Camp station on the Frieda River could have been a potential candidate, based on situation and length of record, but analysis of the data showed that it could be used together with the two mainstream stations only for very short, non-significant periods, and could not therefore be of any use for our study. Another major station of the Sepik basin, Telefomin, was also discarded as being inappropriate for this purpose.

Hence, the objective is to make a flood forecast model for the Ambunti station, based on data available both at Green-River and Ambunti stations for the preceding days/weeks. Given the very large fraction of the Ambunti watershed that is not controlled by the Green-River station (>3/4), it is essential to combine the data at both stations to produce a reliable forecast at the downstream station. Trustable rating curves are available at both stations over the full record period, so the model variables used are the discharges at the two stations. Analysis of the data showed that daily values provide the most appropriate time-step for our study. No information is available on the geometry and hydraulic characteristics of the Sepik river bed and flood plain, so the model needs to be built though some black-box type of approach, either statistically or conceptually. The fact that the problem has little resemblance with a true flood routing problem can be seen for instance from the comparison of observed hydrographs at the two stations on Figure 1, both in terms of relative discharge magnitude and of signal shape (ie. very distinct wave-lengths).

Figure 1: observed hydrographs at GREEN-RIVER and AMBUNTI for 1983



2.2 Model development

Given the available data and problem characteristics (see §2.1), a linear model appears as best suited for our purpose. This consists, for a given lead-time, in a prediction equation forecasting the discharge at Ambunti as a linear combination of previously observed dsicharges at Green-River and Ambunti. Such linear predictive equations can be obtained through various methods, all having in common to optimize some quality criterion (eg.: mean squared error) expressing the adequation between predicted and observed values for available records.

- ***The statistical approach***

A very powerful method to produce such linear models is the stepwise, progressive regression method (Draper & Smith, 1981). It allows both to select the most appropriate variables and to establish the model using these variables, i.e. determine the linear coefficients to apply to these selected variables.

The **reghyd** computer program developed at Orstom allows to implement this method on time-series of hydrologic data, including significant proportions of missing data. This tool was therefore used for our study, and handed over with full functional capabilities to M. Virobo (see §3.1). The implementation was done for two different lead-times, taken as being both useful in operational situations and compatible with the Sepik River's behavior, namely: 2-days and 5-days. From the whole record of observations, a proportion of about 2/3 of the data was used for model calibration, the remaining 1/3 being used for validation (a calendar year may only be used wholly for calibration or validation, not subdivided between both)

Analysis and comparison of results obtained with the method has led to retaining the following forecasting equations for the two lead times:

- 2-day lead-time: ambunti+2 =A0*ambunti0 +A1*ambunti-1 +A2*ambunti-2
  + G0*green-river0 + G1*green-river-1 + G2*green-river-2
  + G3*green-river-3+ Constant (m3/s)

  where: A0=2.8248, A1=-2.3964, A2=0.5456, G0=0.1622, G1=-0.1542, G2=0.1081, G3=-0.0551, and Constant=22.9043

- 5-day lead-time: ambunti+5 = A0*ambunti0 +A1*ambunti-1 +A2*ambunti-2
  + G0*green-river0 + Constant (m3/s)

  where: A0=3.9042, A1=-4.1807, A2=1.1640, G0=0.2351, Constant=135.4814

(notation: day 0 is the day when the forecast is issued, ambunti-2 designates for instance the observed discharge at Ambunti 2 days before day 0).

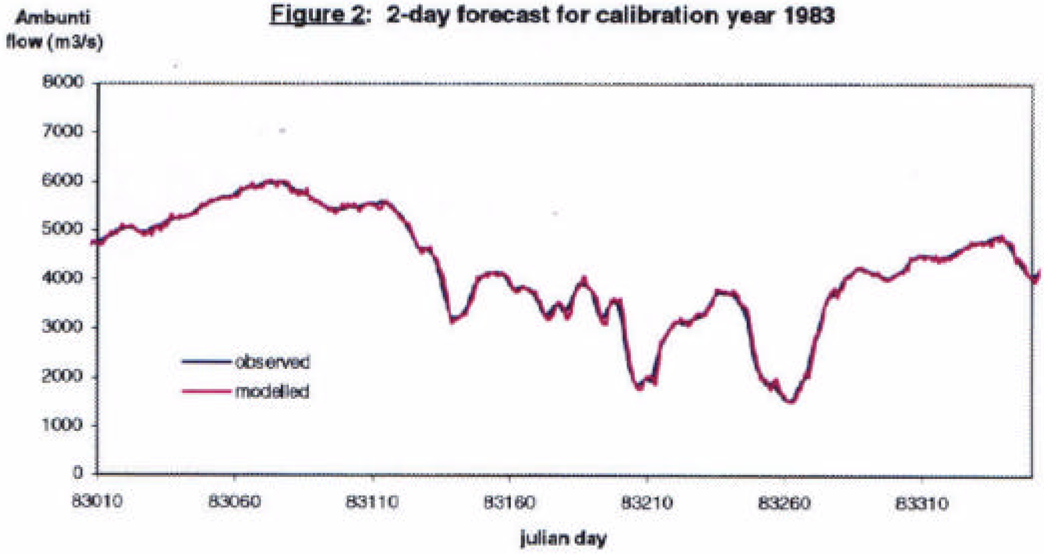
The quality criteria for these two equations are presented in Table 1.

Table 1: Quality criteria values obtained for the 2-day and 5-day lead-times with the statistical method (MSE=mean squared error):

- 2-day:	calibration MSE=100 m3/s; validation MSE=116 m3/s; max.error=649 m3/s
- 5-day:	calibration MSE=295 m3/s; validation MSE=322 m3/s; max.error=1304 m3/s

Figures 2 to 5 show the comparison of observed and predicted discharges at the Ambunti station for the two lead times and for either calibrating and validating years.

Figure 2: 2-day forecast for calibration year 1983



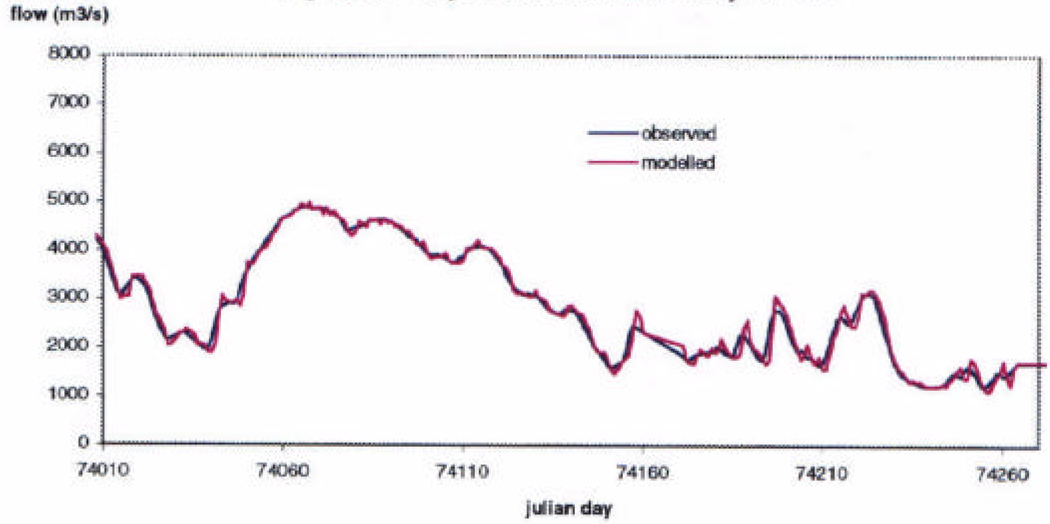


Figure 3: 2-day forecast for validation year 1974

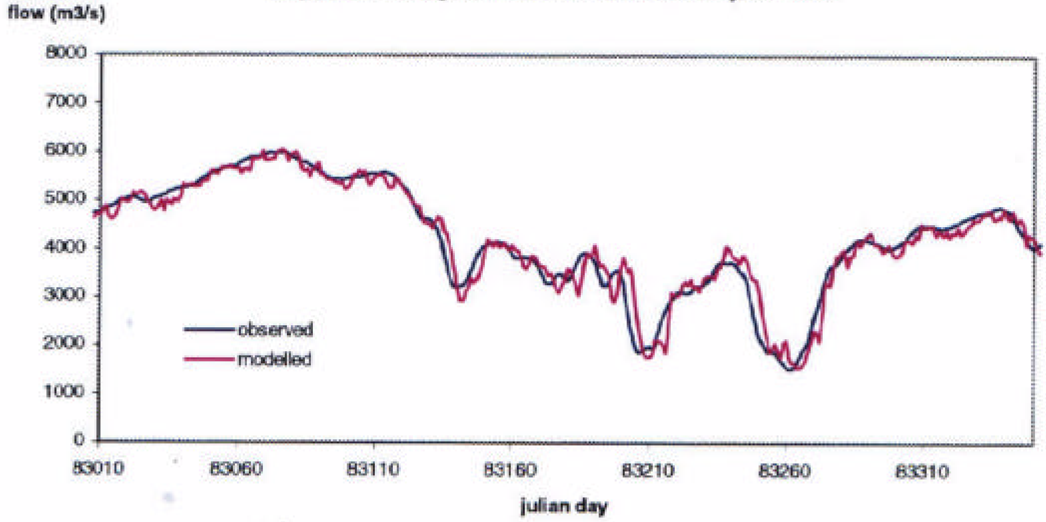
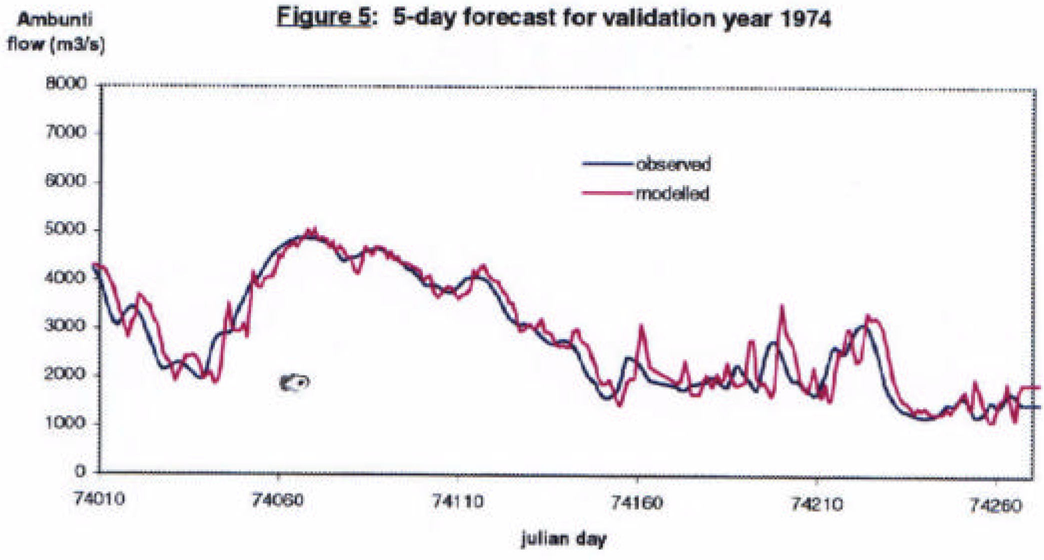


Figure 4: 5-day forecast for calibration year 1983

Figure 5: 5-day forecast for validation year 1974



- ***A hybrid statistical-conceptual approach***

Although the progressive regression method is likely to produce optimal or close-to-optimal models for the given data set, another approach was also attempted combining statistical and conceptual aspects with the aim of producing somewhat smoother predicted hydrographs. The idea behind this method was to have the upstream, Green-River information come in the model through a pre-processing transfer-function type filter, instead of raw daily data. This idea arises from the high variability at Green-River at the daily time-step, compared to Ambunti, and from the physical routing process that, as a first approximation, can be represented by a linear transfer function. Hence, in this method, the transfer function parameters replace the coefficients of daily Green-River discharges as model parameters.

We used for transfer function a log-normal density function, thereby introducing three parameters: two for shape (named T0 and T1, they have the dimension of time, and represent the mean and standard deviation of the function seen as a distribution over time, respectively) and one for magnitude (S, a discharge scaling factor). The model therefore writes:

ambunti+5 = A0*ambunti0 + A1*ambunti-1 + A2*ambunti-2 + Constant
+ Transfer(T0, T1, S ; green-river0, green-river-1, ... green-river-20)

where: A0=1.9546, A1=-0.8391, A2=-0.2358, Constant=177.503 (m3/s)
T0=1 day, T1=2 days, S=0.17764

Calibration was performed under Excel using the optimizing solver with the mean squared error criterion. This Excel worksheet has also been handed over to M. Virobo.

A sample output is provided on Figure 6. It shows that, although the mean squared error criterion (331 m3/s) is less good than for the pure regression model, the predicted hydrograph is significantly smoother, meaning that the daily updating of the forecast is less responsive to day-to-day upstream variations. This property may be preferred for operational purposes.

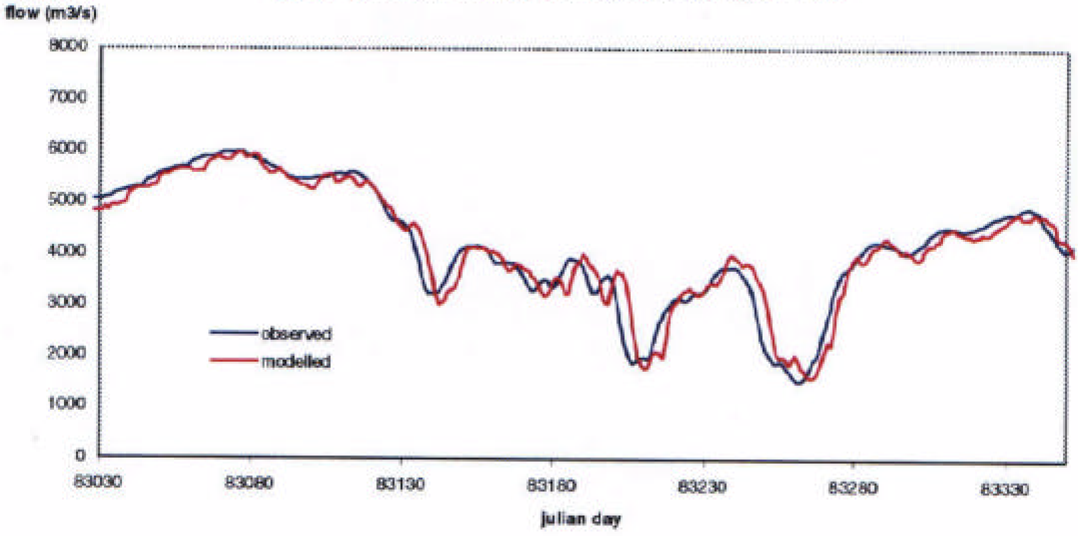


Figure 6: 5-day forecast with hybrid model, year 1983

3. Training with the tools used for model development

The various tools used for model development have been transferred to M. Virobo, as a comprehensive diskette archive for installation on the BWR computers, together with training for proper use.

The archive, containing both software and data files, is organized according to the directory tree structure of Figure 7. Software includes:

- the stepwise regression program (REGHYD) for hydrometric data, used to produce the statistical Sepik forecasting model (see §2.1); since this tool can accept two types of data file formats, either binary or ascii-column, the full data set and control file are provided under the binary format, together with a smaller, sample data set and control file for the ascii-column format;
- the general-purpose stepwise regression program (REGALL) meant to be used for any type of data, not only hydrometric time series, with a sample data file;
- the Excel worksheet EXTRACT5.XLS that contains the hybrid statistical/conceptual model for 5-day lead-time forecasting;
- the raw data file GREENAMB.QJ containing the 1970-1997 daily discharges at Green-River and Ambunti, in ascii-column format.

Figure 7: Content of Software & Data archive

REGHYD		stepwise regression for hydro data
	REGHYD20.EXE	large-memory executable
	REGHYD10.EXE	small-memory executable
SEPIKDAT.BIN		with binary hydro data:
	LAG5.DON	control file
	QJ70.BIN	1970 hydro data
	QJ71.BIN	1971
	.....	....
	QJ96.BIN	1996
	QJ97.BIN	1997
SAMPLDAT.COL		with ascii-column hydro data:
	TEST.DON	control file
	Z.75	sample hydro data (1975)
	Z.76	(1976)
REGALL		general purpose stepwise regression (any data)
	REGALL.EXE	executable
	SAMPLE.DAT	sample ascii-column data
	RESULTS.LST	sample output file
SEPIK.XLS		hybrid statistical-conceptual model
	EXTRACT5.XLS	Excel worksheet for model+data
RAWDATA		raw data for:
	GREENAMB.QJ	1970-97 Green-River & Ambunti daily flows

References:

Danloux J. (1997): *Operational hydrology in Papua New Guinea, Report n° 1: Initial results and recommendations.* Orstom-Noumea Hydrology section, and PNG Bureau of Water Resources.

Draper N. and Smith H. (1981): *Applied Regression Analysis.* Wiley-Interscience.