C4

Proc. 4th IFAC Symp. on
Identification and System Parameter Estimation, Rajbman (ed.)

Tbilisi, Georgia, USSR, 1976.

C8

4.6

IDENTIFICATION OF A RAINFALL-RUNOFF PROCESS

Application to the Semois River (Belgium).

B. LORENT, Fondation Universitaire Luxembourgeoise,
B - 6700 ARLON, Belgium.

M. GEVERS, Louvain University, Bâtiment Maxwell,
B - 1348 LOUVAIN-LA-NEUVE, Belgium.

ABSTRACT : This paper describes a model used to forecast or to simulate the river flow discharge at the outlet of a basin. The model inputs are the measured rainfalls and the estimated potential evapotranspiration. The model structure includes three submodels. The first submodel computes a net rain from the measured rainfalls and the estimated evapotranspiration; it includes the main non linearities of the system. The second submodel simulates the base flow. The third computes the predicted surface runoff by means of a stochastic linear difference equation model. For the studied river basin, the model operates with daily mean values of rainfall, evapotranspiration and river flow.

I. INTRODUCTION.

I.1. The modeling of the rainfall-runoff process of a river basin is of practical interest for a variety of purposes. Among these :

- flood forecasting associated with alarm systems and actions on regulation dams;
- low water forecasting allowing judicious action in water resources management;
- check on the accuracy of dubious discharge data;
- computation of missing data.

The model presented in this paper, has been identified with data from the Semois river basin (Belgium), that is a tributary of the Meuse. This basin has a surface area of 1 235 km2 and a lag of 2 or 3 days.

I.2. The simulation model approximates the dynamics of the " river basin system " whose input is the whole meteorological environment. This global input is characterized by two main factors : the rain and the potential evapotranspiration. These are the inputs of our model. The model output is the flow discharge at the outlet of the basin. For the Semois River model, the output is the average daily flow discharge at Member, outlet of the Belgian river basin of this river.

I.3. Amounts of precipitated water and potential evapotranspiration can vary from one place to another in the basin. However, for the studied basin, daily mean rainfall and potential evapotranspiration have proved to be efficient model inputs. Models using local rainfall measures have also been identified to verify this point.

I.4. The model is identified using only the abovementioned input-output data records, without assuming any knowledge of the physical characteristics of the basin.

This paper is not concerned with the estimation of the potential evapotranspiration, ETP. For the Semois River basin, we have used estimated values of ETP computed by the I.R.M. (" Institut Royal Météorologique " of Belgium), who have also provided the rainfall and flow discharge data.

II. GENERAL DESCRIPTION OF THE MODEL.

II.1. The structure of this rainfall-runoff model has already been described in [4].

With k indicating the k-th day, we call : Q(k), the mean discharge at the outlet of the basin; PB(k), the mean rainfall over the basin; ETP(k), the mean potential evapotranspiration.

We suppose that the discharge Q(k) can be written :

$$Q(k) = R(k) + B(k) \quad (1)$$

where : R(k) is the surface runoff discharge;
B(k) is the groundwater flow discharge.

II.2. A first non linear submodel, whose inputs are PB(k) and ETP(k), computes a mean net rain PN(k) and a drainage (percolation) term D(k), by using a surface storage function S(k). The net rain PN(k) is the input of a second linear submodel, whose output is the

estimated surface runoff $\hat{R}(k)$. The drainage term D(k) is used as input for a third linear submodel whose output is the estimated base flow $\hat{B}(k)$. The model structure is illustrated by figure 1. The model parameters of the 3 submodels are globally identified by minimizing the mean square error between the predicted flow discharge $\hat{Q}(k)$ and the measured value Q(k).

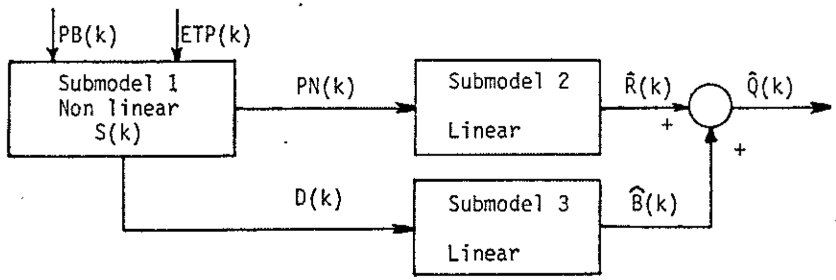

Figure 1.

III. DESCRIPTION OF THE NON-LINEAR SURFACE STORAGE MODEL COMPUTING THE NET RAIN.

III.1. The amount of precipitated water PB(k) can be decomposed as follows :

$$PB(k) = PN(k) + E1(k) + SI(k) \quad (2)$$

where : PN(k) is the net rain, namely the part of PB(k) that runs on the basin surface and reaches the river to form the surface runoff R(k);

E1(k) is the part of the rainfall PB(k) that evaporates during day k;

SI(k) is that part of PB(k) that stays on the soil or that is intercepted by the vegetation.

Later on, this water will either infiltrate or evaporate.

We define the " running coefficient " on day k as :

$$CR(k) = PN(k) / PB(k), \; 0 < CR(k) < 1 \quad (3)$$

The amount of stored water on the basin surface (i.e. vegetation and soil) at the end of day k, is called S(k) and can be decomposed as follows :

$$S(k) = S(k-1) + SI(k) - E2(k) - D(k) \quad (4)$$

where : E2(k) is the part of the stored water that evaporates during day k, directly or through the transpiration mechanism;

D(k) is the amount of water drained to the phreatic surface.

The total amount of evaporated water during day k is :

$$E(k) = E1(k) + E2(k) \quad (5)$$

III.2. In the presented model, we have assumed that it is possible to define PN(k) as a function of PB(k), ETP(k), S(k-1) and some unknown parameters that depend on the physical characteristics of the basin. The difficulty is that this function is highly non linear. We have tried to approximate it considering the following observations :

- the running coefficient increases with the soil moisture;
- for a defined soil moisture, the running coefficient increases with the rainfall intensity and, on the average, with the amount of precipitated water;
- the percolation increases with the amount of water stored on the soil surface.

III.3. Assuming that there is an upper limit Smax for the stored water, we define the " storage deficit " as :

$$DEF(k) = Smax - S(k) \quad (6)$$

Smax is the first unknown parameter of this model.

The model then acts as follows :

A) Computation of PN(k) :

1) If PB(k) ⩾ ETP(k), we chooze : E1(k) = ETP(k)
   If PB(k) < ETP(k), we chooze : E1(k) = PB(k)

2) We define : PE(k) = PB(k) - E1(k) (7)

   SI(k) is then computed by means of the following relationship :

$$SI(k) = DEF(k-1) \, . \left[1 - e^{-PE(k) \, / \, (b.DEF(k-1))} \right] \quad (8)$$

   where b is the second unknown parameter to be adjusted with :
   b > 1, so that : SI(k) < PE(k)

3) PN(k) results finally from (2) :

$$PN(k) = PB(k) - E1(k) - SI(k) \quad (9)$$

B) Computation of D(k) :

1) If S(k-1) + SI(k) ⩾ ETP(k) - E1(k),
   we chooze : E2(k) = ETP(k) - E1(k)
   If S(k-1) + SI(k) < ETP(k) - E1(k),
   we chooze : E2(k) = S(k-1) + SI(k)

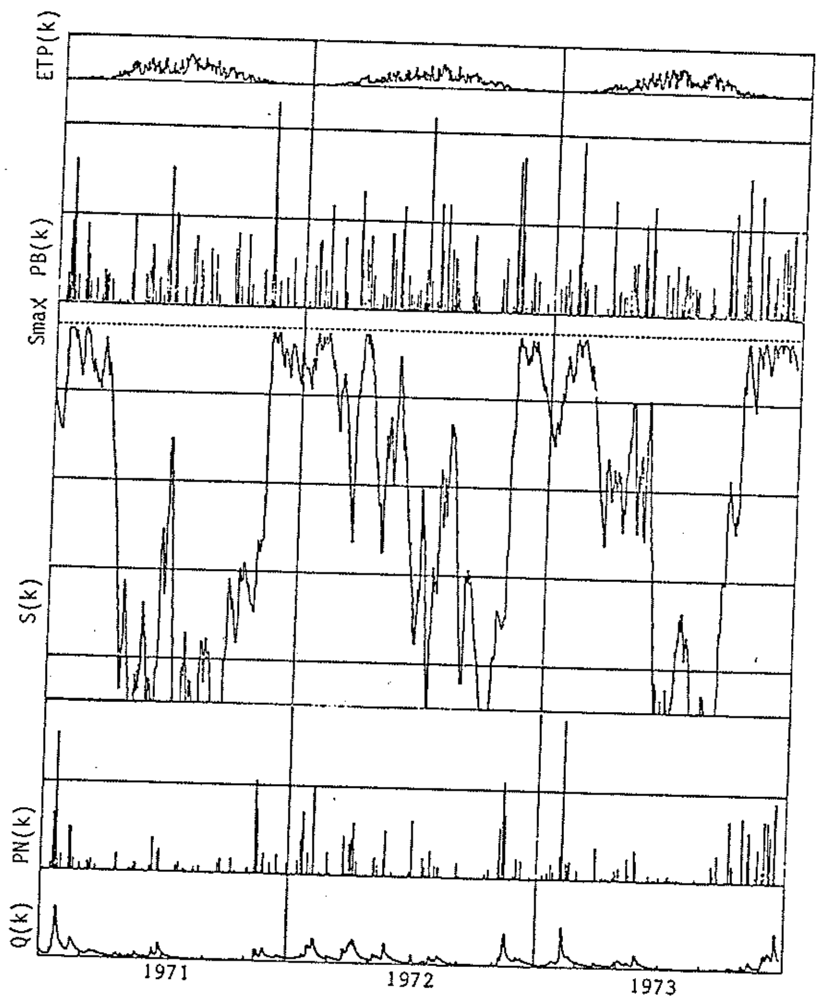

Fig. 2 - Graphic illustration of the efficiency of the model computing the net rain.

2) The percolation term is estimated by :

$$D(k) = (Dmax / Smax) \cdot \left[S(k-1) + SI(k) - E2(k)\right] \qquad (10)$$

Dmax is the third unknown parameter of the model.

3) Finally :

$$S(k) = \left[(Smax - Dmax) / Smax\right] \cdot \left[S(k-1) + SI(k) - E2(k)\right] \qquad (11)$$

III.4. The three parameters, b, Smax and Dmax, depend upon the physical characteristics of the basin. They will be chozen so as to minimize the mean square runoff prediction error of the global model.

IV. DESCRIPTION OF THE GROUNDWATER MODEL.

IV.1. The study of the groundwater depletion curves shows that they can be approximated by :

$$Q(k) - B_o = (Q(k_o) - B_o) \cdot e^{-(k-k_o)/TH} \qquad (12)$$

where k_o is the starting time of the groundwater recession and B_o a limit discharge level characteristic of the considered period. The time k_o is also defined as the time when the value of the surface runoff R(k) is negligible.

Because of this observation, we have decomposed the base flow as follow :

$$\boxed{B(k) = BR(k) + BL(k)} \qquad (13)$$

where BL(k) corresponds to B_o and varies very slowly during the year whereas BR(k) varies faster and decreases during dry weather periods following an exponential law as given by (12).

IV.2. The groundwater phenomena are unknown at the scale of the studied basin and no groundwater model was available. We have therefore decided to simulate the two terms of the base flow with the simplest possible realistic model consistent with the observed data.

Supposing that BR(k) and BL(k) are amounts of water coming from two different groundwater reservoirs filled up by two delayed fractions of the percolation term D(k), we have chozen the following simulation model :

$$BR(k) = \alpha \cdot BR(k-1) + (1-\alpha) \cdot p \cdot D(k-d_r) \qquad (14)$$

$$BL(k) = \beta \cdot BL(k-1) + (1-\beta) \cdot q \cdot D(k-d_l) \qquad (15)$$

α is equal to $e^{-1/TH}$ and can be estimated from the groundwater depletion curves, because during dry weather periods the percolation term $D(k-d_r)$ is equal to zero or negligible.

β, p, q, d_r, d_l are estimated using the discharge data collected when the surface runoff is non existent. However there is very seldom more than one period each year where the Semois riverflow is made up of only BL(k), the long term component of the base flow. For this reason the identification of the model simulating BL(k) is difficult.

V. THE NET RAIN-SURFACE RUNOFF MODEL.

V.1. For a long time, the unitgraph method has been used to compute flood discharges.

If the first model is able to adequately compute the net rain PN(k), then the surface runoff should be linearly related to this net rain :

$$\boxed{R(k) = \sum_{j=0}^{\infty} H(j) \cdot PN(k-j)} \qquad (16)$$

where H(j) is the value, at time j, of the discrete impulse response of the system relating PN(k) to R(k), also called the instantaneous unitgraph.

This system can also be described by a difference equation, with much fewer parameters :

$$R(k) = \sum_{j=1}^{N} a_j \cdot R(k-j) + \sum_{j=0}^{M} b_j \cdot PN(k-j) \qquad (17)$$

Using the delay operator Z^{-1}, we write :

$$A(Z^{-1}) \cdot R(k) = B(Z^{-1}) \cdot PN(k) \qquad (18)$$

with : $A(Z^{-1}) = 1 - a_1 \cdot Z^{-1} - a_2 \cdot Z^{-2} - \ldots - a_N \cdot Z^{-N}$

$B(Z^{-1}) = b_o + b_1 \cdot Z^{-1} + \ldots + b_M \cdot Z^{-M}$

V.2. A correlated noise term is added to take into account the observation errors and the fact that the first submodel produces only an imperfect estimate of the net rain. The linear model now becomes :

$$\boxed{A(Z^{-1}) \cdot R(k) = B(Z^{-1}) \cdot PN(k) + C(Z^{-1}) \cdot e(k) + C_o} \qquad (19)$$

where : $C(Z^{-1}) = 1 + C_1 . Z^{-1} + \ldots + C_P . Z^{-P}$

and : (H.A.)

$$\begin{cases} E\{e(k)\} = 0 \ , \ E\{e(k).e(k-j)\} = \begin{vmatrix} 0, & j \neq 0 \\ \sigma_e^2, & j = 0 \end{vmatrix} \\ E\{e(k) \ . \ PN(k-j)\} = 0 \ , \forall j \\ E\{e(k) \ . \ R(k-j)\} = 0 \ , \forall j \geq 1 \end{cases}$$

The model can be written :

$$\boxed{R(k) = \underline{\alpha}' \ . \ \underline{Z}(k) + e(k)} \qquad (20)$$

where : $\underline{\alpha}' = \left[a_1 \ a_2 \ \ldots \ a_N \ b_o \ b_1 \ \ldots \ b_M \ C_1 \ C_2 \ \ldots \ C_P \ C_o\right]$

$$\underline{Z}'(k) = \left[R(k-1) \ \ldots \ R(k-N) \ PN(k) \ \ldots \ PN(k-M) \ e(k-1) \ \ldots \ e(k-P) \ 1\right]$$

$\hat{R}(k) = \underline{\alpha}' \ . \ \underline{Z}(k)$ is the predicted value of the surface runoff R(k), knowing the observations $\underline{Z}(k)$.

V.3. The parameter vector $\underline{\alpha}$ is identified using the " Recursive Approximate Maximum Likelihood " algorithm (RAML) described in [6] and [7].

The value of the surface runoff is found by subtracting the simulated base flow from the total river flow discharge. The RAML algorithm can be summarized as follows. Starting with initial values :

$$\underline{a}(o) = \left[o \ \ldots \ o\right]$$

$$\underline{\underline{S}}(o) = s_o \ . \ \underline{\underline{I}}, \text{ where } \underline{\underline{I}} \text{ is a unit } \left[N + M + P + 2, \ N + M + P + 2\right]$$

matrix and s_o is a large positive number.

we have :

$$\underline{\underline{S}}(k) = \underline{\underline{S}}(k-1) - \frac{\underline{\underline{S}}(k-1) \ . \ \hat{\underline{Z}}(k) \ . \ \hat{\underline{Z}}'(k) \ . \ \underline{\underline{S}}(k-1)}{1 + \hat{\underline{Z}}'(k) \ . \ \underline{\underline{S}}(k-1) \ . \ \hat{\underline{Z}}(k)} \qquad (21)$$

$$\hat{\underline{a}}(k) = \hat{\underline{a}}(k-1) + \underline{\underline{S}}(k) \ . \ \hat{\underline{Z}}(k) \ . \ w(k) \qquad (22)$$

where : $\hat{\underline{Z}}(k) = \left[R(k-1) \ \ldots \ R(k-N) \ PN(k) \ \ldots \ PN(k-M) \ \hat{e}(k-1) \ldots \hat{e}(k-P) \ 1\right]$

$$\hat{e}(k) = R(k) - \hat{\underline{a}}'(k) \ . \ \hat{\underline{Z}}(k)$$

$$w(k) = R(k) - \hat{\underline{a}}(k-1) \ . \ \hat{\underline{Z}}(k)$$

If the model is adequate, the parameter vector $\hat{\underline{a}}(k)$ converges to the optimal value $\underline{\alpha}$.

To check the model adequacy, the hypotheses (H.A.) made before have to be verified with the identified parameters.

V.4. The global identification procedure goes as follows. An initial guess is made for the three parameters (b, Smax, Dmax). This allows one to compute PN(k) and D(k). The parameters of the other two submodels are then identified so as to minimize the mean square error between the measured and the predicted river flow. By successive steps, the set of values (b, Smax, Dmax) are adjusted and the other parameters recomputed so as to globally minimize the mean square prediction error. Since the base flow is on the average much smaller than the runoff, the identification procedure can be accelerated by neglecting the base flow model in the first few iterations.

V.5. The identified model (19) can be used to predict the surface runoff a few days ahead, by replacing the net rain inputs by meteorological predictions. It can also be used for simulation, namely to estimate runoff discharge from the only knowledge of the net rain computed from the rainfall and potential evapotranspiration. This can be useful for the following purposes :

- verification of dubious discharge data;
- study of the information given by the computed net rain about the surface runoff;
- study of the future river flow discharge under various possible meteorological conditions.

The simulation model is derived from the prediction model (19) as follows :

$$A(Z^{-1}) \ . \ R^S(k) = B(Z^{-1}) \ . \ PN(k) + C_o \qquad (23)$$

If we call s(k), the simulation error :

$$s(k) = R(k) - R^S(k) \qquad (24)$$

then it is easy to see that the simulation error has the following dynamics :

$$A(Z^{-1}) \ . \ s(k) = C(Z^{-1}) \ . \ e(k) \qquad (25)$$

It can be shown that s(k) is the best (i.e. minimum variance) linear unbiased simulation error.

CONCLUSION.

We have presented an efficient model that is able to forecast river flows and to adequately simulate the runoff with only the main meteorological information, namely the rainfalls and potential evapotranspiration. The model has been identified with data from the Semois river covering a 7 year time span. It has been shown to perform very accurately. The model does not work with snow inputs and for floods caused by snow melt. The difficulty is to find an efficient input for the surface runoff model such as a delayed equivalent net rain.

Another problem not presented in this paper, is the choice of optimal sets of locations for the rainfall measurements. This is the object of our present research and will be presented in a next paper.

REFERENCES.

[1] G. BAZIER et E. PERSOONS, 1975 - " *Application of the LGR system of telemeasurement on flow predicting* ", System simulation in water ressources, ed. G.C. VANSTEENKISTE, 1976, North Holland Publishing company.

[2] F. BULTOT et G.L. DUPRIEZ, 1975 - " *Conceptual hydrological model for an average - sized catchment - area* ", Journal of Hydrology, 29 (1976), 251-292.

[3] G. de MARSILY, 1971 - " *La relation pluie-débit dans le bassin versant de l'Hallue* ". Ecole Nationale des Mines de Paris, Centre d'informatique géologique, Fontainebleau.

[4] B. LORENT, 1975 - " *Modélisation et identification d'une relation pluie-débit. Le modèle " Semois "* ". 7th IFIP Conference on Optimization Techniques, Nice, sept. 1975.

[5] M. ROCHE, 1963 - " *Hydrologie de surface* ". Gauthier-Villars, Paris.

[6] P.C. YOUNG, 1972 - " *Applying parameter estimation to dynamic systems* ", Part. 1, Control Engineering, 1969.

[7] R.L. KASHYAP et A.R. RAO, 1972 - " *Real time recursive prediction of river flows* ", Automatica, vol. 9, pp. 175-183, March 1973.

[8] G.E.P. BOX et G.M. JENKINS, 1970 - " *Time - Series Analysis Forecasting and Control* ", Holden Day.

Identification and System Parameter Estimation, Rajbman (ed.)

4.7

APPLICATIONS OF RECURSIVE ESTIMATION TECHNIQUES TO TIME VARIABLE WATER RESOURCE SYSTEMS

P. G. Whitehead (UK)*

SUMMARY In this paper a flexible methodology is illustrated using the recursive instrumental variable approximate maximum likelihood algorithms of time series analysis to identify parameter variations and characterise time varying water resource systems. The approach is developed for a single input-single output rainfall-runoff process and extended to the multivariable situation for a river water quality system.

1. INTRODUCTION Unlike most industrial plant water resource systems are subject to temporal variations which may be naturally occuring, such as seasonal soil moisture variations, or "man-made" if the physical nature of the system is altered by operational management. In such circumstances it is unlikely that parameters in a model of the system will be stationary and identification of model structure is particularly difficult.

Models of water resource systems have been estimated against data collected during an extensive three year study of the Bedford-Ouse River system in Central-Eastern England [1]. The objectives of the study were to develop and utilise

* Control and systems Group, Department of Engineering, University of Cambridge.
(As from November 1975 - Research Fellow, Centre for Resource and Environmental Studies, Australian National University, Australia).