JARQ 26, 187-191 (1992)

Runoff Analysis in a Catchment Area Covered by Tropical Forests Using a Robust Multiple Regression Model

1. Basic theory

Hajime TANJI and Yoshinobu KITAMURA

Department of Hydraulic Engineering, National Research Institute of Agricultural Engineering (Tsukuba, Ibaraki, 305 Japan)

Abstract
Actual rainfall runoff modeling involves two kinds of procedures; identification of the optimum parameters and selection of suitable data. Although many reports described automatic procedures to obtain the optimum parameters of models, few proposed a suitable procedure of data selection. In these papers, a new multiple regression model which enables to identify the optimum parameters of linear runoff and select suitable data for the modeling of linear runoff is described. This method is mathematically based on the robust statistical theory developed in the past decade. The method is referred to as RFMD (robust fixed maximum discharge) method because it is also based on the general expression of Shiraishi's FMD method of data separation. In this paper, the robust statistical theory was introduced into runoff analysis and tested by using the manipulated data of the Zao dam in Japan. Compared to the FMD method, the RFMD method showed a strong robustness.

Discipline: Irrigation, drainage and reclamation
Additional key words: multiple regression analysis, robust estimation

Introduction

A large number of automatic algorithms is available to identify the optimum parameters of a rainfall runoff model. If hydrological data contain some irregularity, automatic algorithms cannot be used. The algorithms enable to attain a local optimum or cannot lead to a stable solution. Theoretically, it is considered that an optimum model cannot be obtained due to lack of accuracy of data. However, actual rainfall runoff analysis requires the identification of as many as possible runoff models regardless of the accuracy of hydrological data. In this procedure, if a satisfactory model can be identified based on low accuracy data, the model is called "robust". If the accuracy of the data is low and a model can not be identified, model identification must be tried again by using data with a higher accuracy in a limited period of time. Fig. 1 illustrates the procedure of data selection. Even if the parameters of the models are identified by an automatic algorithm, runoff analysis is biased if this procedure of data selection is applied due to the subjectivity of the procedure. In addition, if the data selection procedure is based on a trial and error method, the size of the selection procedure is large. If this procedure of selection is performed by the models themselves automatically, the identification of the models can be more effective and unbiased. During the past decade, a "robust" estimation method, which can omit the biased effect of error distribution, was used for regression analysis. Tanji[4,5] firstly introduced the use of this robust model for rainfall runoff analysis of multiple regression models. In these papers, the model was applied

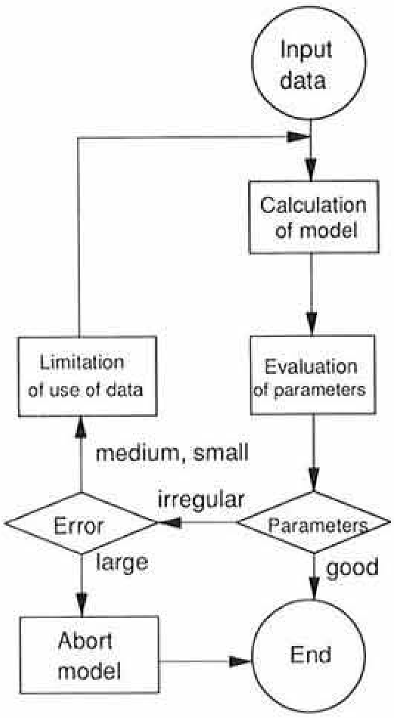

Fig. 1. Outline of model with irregular data

to the rainfall runoff analysis of a catchment covered by tropical forests where the data may contain some unexpected errors.

For the use of a statistical method for rainfall runoff analysis, the usefulness of a model depends on the physical significance of the model. If the physical significance of the statistical parameters can not be interpreted, the model is not useful. Once, Shiraishi et al.[3] and Ito et al.[1] used a multiple regression model for rainfall runoff analysis. At that time, multiple regression analysis was introduced as a general expression of a unit hydrograph method, which enabled to interpret the physical significance of statistical parameters and made this method more valuable. In the use of a robust estimation method for multiple regression analysis, the physical significance of this statistical method must be clearly defined. The author analyzed many kinds of robust estimation methods and found that the M estimation method (Maximum-likelihood-type-estimation) among robust estimation methods was equivalent to one of the general expressions of the FMD method. The biweight method in the M estimation method is homologous to the FMD method from the view point of separation of the linear runoff component. This paper introduced this robust multiple regression method into runoff analysis in relation to the FMD method.

Methods of analysis

Basic equations of robust estimation are introduced here based on the report of Koyanagi[2]. The estimation method is limited to the M estimation method among the robust estimation methods, where, measured data of time series $y_j\,(j=1, \cdots, n)$ can be expressed as a function of the parameters $x_i\,(i=1, \cdots, m)$ plus errors ε_j.

$$y_j = f_j(x_1, \cdots, x_m) + \varepsilon_j$$
$$= f_j(\mathrm{X}) + \varepsilon_j. \quad \cdots\cdots (1)$$

In runoff analysis, (X) is a time series of rainfall data and y_j that of discharge. Both dimensions are expressed as mm/d.

Especially, if f_j is expressed as the linear function of g_j, y_j is written as follows:

$$y_j = g_1x_1 + \cdots + g_mx_m. \quad \cdots\cdots (2)$$

The parameter g_1 is called an intercept in a multiple regression method. The sum of partial regression coefficients $g_2, \cdots, g_m$ is referred to as "linear percentage runoff". Here, the most likelihood method is generalized and objective variables are set at X as follows:

$$\sum_{j=1}^{n} \Psi_j(y_j - f_j(\mathrm{X})) = \tilde{M}(\mathrm{X}) = min. \quad \cdots\cdots (3)$$

The least square method is expressed as follows:

$$\Psi_j(y_j - f_j(\mathrm{X})) = \frac{(y_j - f_j(\mathrm{X}))^2}{\sigma_j^2}. \quad \cdots\cdots (4)$$

Now, $v_j = y_j - f_j(\mathrm{X})$ and the rate of $\Psi_j\,(v_j)$ to v_j^2 are set or $\omega_j(v_j)$. Then equation (3) is written as follows:

$$\tilde{M}(\mathrm{X}) = \sum_{j=1}^{n} \frac{(y_j - f_j(\mathrm{X}))^2}{\sigma_j^2} \omega_j(v_j) = min. \quad \cdots\cdots (5)$$

ω_j can be solved iteratively because ω_j is approximately equal to 1 for most of the data. ω_j is called

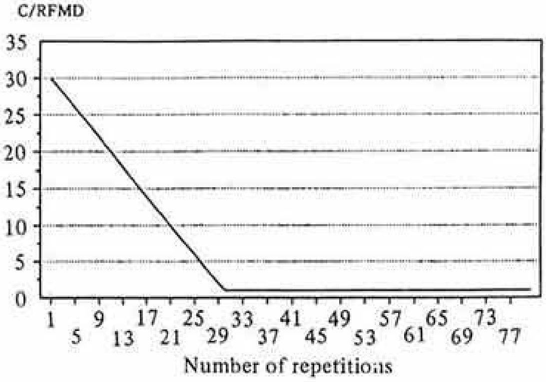

Fig. 2. Temporary RFMD value

an adjustive weight. For the calculation of ω_j, the biweight method in equation (6) was adopted. The value of c in equation (6) decreases with repetitions. The final value of c is the value equivalent to FMD. This final value will be called RFMD value (robust FMD) in this paper.

$$\omega_j(v_j) = \begin{cases} (1-(\frac{v_j}{c})^2)^2 & |v_j| \leq c \\ 0 & |v_j| > c. \end{cases} \quad \cdots\cdots\cdots\cdots (6)$$

The robust regression model constructed by the biweight method will be referred to as an RFMD model in comparison to the Shiraishi's multiple regression model obtained by the FMD method.

For the calculation of the value of ω_j, the repetition time was set at 80. The value of c decreased after 30 repetitions to reach the RFMD value and remained at this value after more than 50 repetitions as shown in Fig. 2.

Relation between FMD and RFMD methods

Fig. 3 outlines the basic concept of the FMD and RFMD methods. The FMD method treats the discharge between zero and FMD as a linear component, while the RFMD method treats the discharge within RFMD from the linear estimated discharge as a linear component. The two methods are common in that the linear component is separable by a certain permissible range. For this reason the c value in equation (7) is called RFMD value. In this interpretation, the RFMD method treats the linear component of runoff as the neighboring discharge of the linear estimated discharge. The FMD method treats the lower part of the discharge as a linear component. This interpretation is more rational than that of the FMD method. The RFMD method is expressed as follows as in the case of the FMD method.

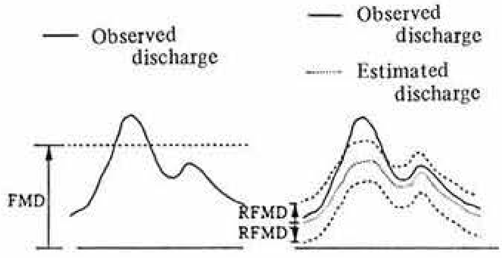

Fig. 3. Concept of FMD and RFMD

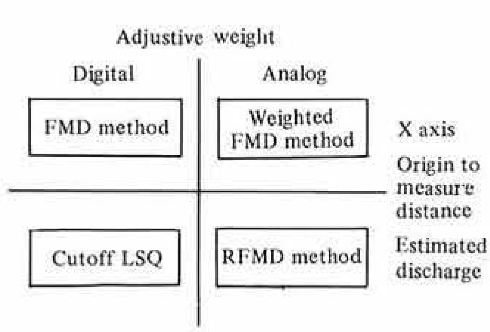

Fig. 4. Relation between FMD and RFMD methods

$$\omega_j(v_j) = \begin{cases} 1 & y_j \leq \text{FMD} \\ 0 & y_j > \text{FMD}. \end{cases} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (7)$$

This equation shows that the relationship between the FMD and RFMD methods is homologous.

A robust regression model takes an adjustive weight even if the datum is located within the range of RFMD. Fig. 4 shows the relationship between the FMD and RFMD methods in taking account of this assumption. From the view point of criteria of linear runoff for the separation of the linear component of runoff, multiple regression analysis can be categorized into the following four kinds: center of the set of the linear component (center of distance) is the X axis or linear estimated discharge, set of the linear data is crisp (the adjustive weight is digital) or fuzzy (the adjustive weight is analog).

This paper considers only the RFMD and FMD methods based on these four kinds of criteria.

Test of robustness

Comparison of the FMD and RFMD methods shows that the main difference appears when a low discharge contains abnormal data. However it is difficult to identify the data in this case. Firstly, the author used the manipulated data of the Zao dam in Yamagata Prefecture in Japan as the check of robustness. Fig. 5 shows the outline of the Zao dam basin (21.0 km^2). Analysis was performed on the data during the period June 5 to November 29, 1971. The data from September 1 to 10 were replaced by a dummy abnormal low discharge and those from September 11 to 20 were replaced by a dummy abnormal high discharge. Both abnormal dummy data were not negligible based on the FMD method.

Fig. 6 shows the estimated discharge calculated by the FMD method for a value of 8 mm/d of FMD.

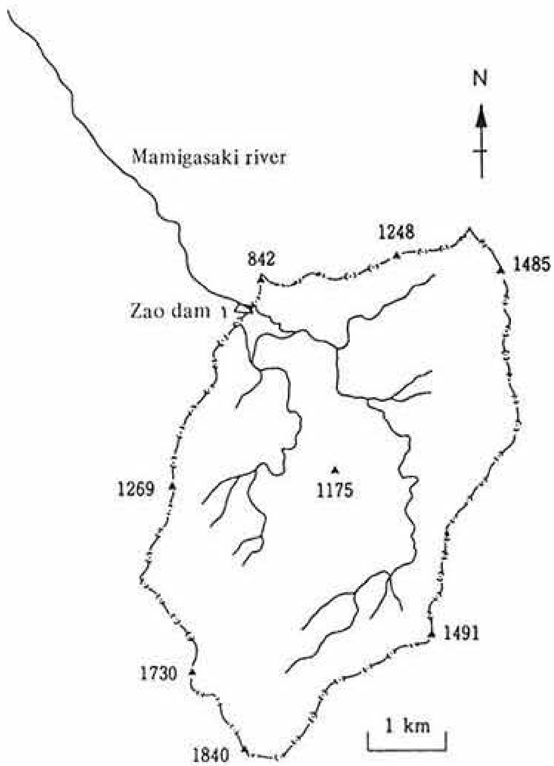

Fig. 5. Outline of the area used for model

Fig. 7 shows the estimated discharge calculated by the RFMD method for a value of 4 mm/d of RFMD. Estimated discharge calculated by the RFMD method was more accurate than that obtained by the FMD method.

Fig. 8 shows the unit hydrographs obtained by the FMD and RFMD methods based on the original data and the data after replacing the accurate data by inaccurate data. Compared with the unit hydrographs based on the original data, the unit hydrograph obtained by the FMD method for the manipulated data was more accurate than that obtained by the RFMD method.

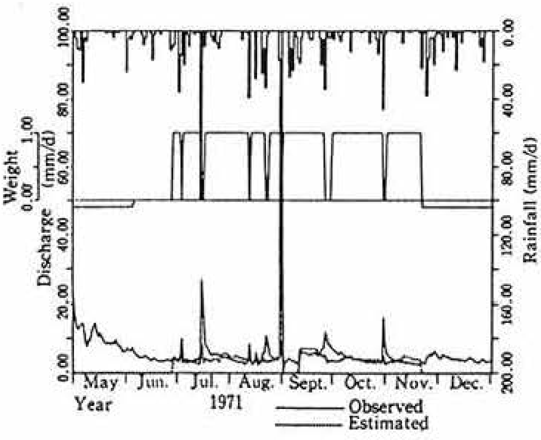

Fig. 6. Comparison of estimated and observed discharges Manipulated data: FMD 8(mm/d).

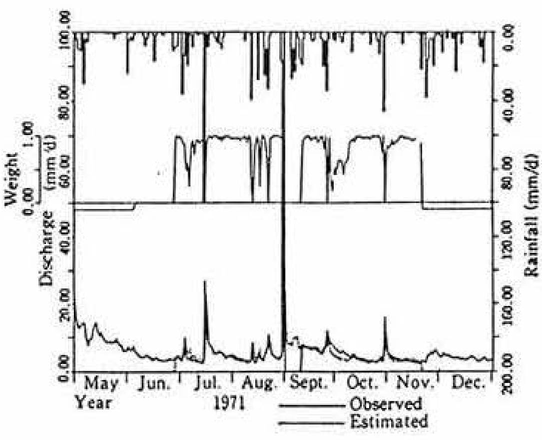

Fig. 7. Comparison of estimated and observed discharges Manipulated data: RFMD 4(mm/d).

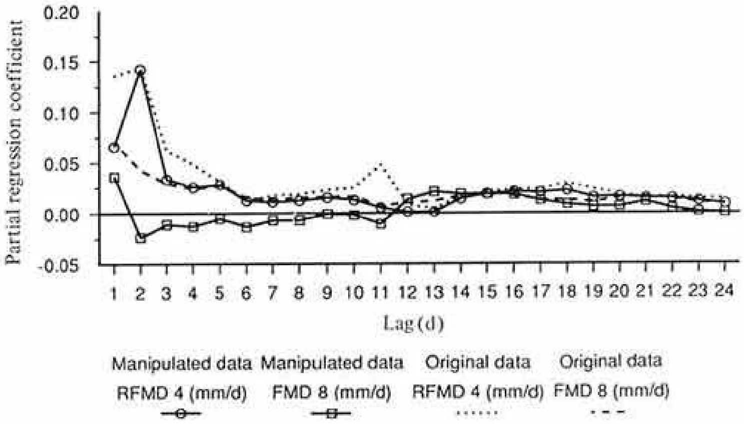

Fig. 8. Statistical unit hydrograph

Conclusion

The authors introduced a new robust method of a rainfall runoff model (RFMD method) which enables to identify the model based on some irregular data. This method was theoretically equivalent to the generalized method of the Shiraishi's FMD method. This method was applied to a dam basin in Japan by using the manipulated data. The results confirmed the strong robustness of this method.

References

1) Ito, Y., Shiraishi, H. & Oonishi, R. (1980): Numerical estimation of return flow in river basin. *JARQ*, **14**, 24-30.

2) Koyanagi, Y. (1978): Robust estimation and its application to data analysis. *Operations Research*, **23**, 299-304 [In Japanese].

3) Shiraishi, H., Oonishi, R. & Ito, Y. (1976): Nonlinear approach to rainfall-runoff process by multiple regression analysis — Application of multiple regressions analysis to hydrologic studies by basin sytem (1) —. *Trans. Jpn. Soc. Irrig. Drain. Reclam. Eng.*, **63**, 43-49 [In Japanese with English summary].

4) Tanji, H. (1990): Runoff analysis with the method of robust regression. *Trans. Jpn. Soc. Irrig. Drain. Reclam. Eng.*, **150**, 67-73 [In Japanese with English summary].

5) Tanji, H. & Kitamura, Y. (1991): Runoff analysis of tropical forest area by a robust regression model. *In* Proceedings of 1991 Annual Conference, Japan Society of Hydrology and Water Resources, 226-229 [In Japanese].

(Received for publication, Nov. 8, 1991)