Journal of Hydrology, 36 (1978) 375—381

[3]
Short Communications

A CRITERION OF EFFICIENCY FOR RAINFALL-RUNOFF MODELS

M. GARRICK, C. CUNNANE and J.E. NASH

Department of Engineering Hydrology, University College, Galway (Ireland)

(Received May 20, 1977; accepted for publication September 26, 1977)

ABSTRACT

Garrick, M., Cunnane, C. and Nash, J.E., 1978. A criterion of efficiency for rainfall—runoff models. J. Hydrol., 36: 375—381.

Two criteria are proposed for expressing the efficiency of conceptual rainfall—discharge models. The first assesses the model as a means of converting the input factors into discharge by comparison with a forecast based only on the seasonal regime of the river. The second expresses the efficiency of the model under the assumption that it is to be used with an updating procedure to provide a forecast of discharge over a prescribed lead time.

INTRODUCTION

Experience in the current decade has indicated that is is surprisingly easy to develop simple conceptual models which, when suitable values of the parameters are chosen, can reasonably well simulate the rainfall—discharge relationship in a given catchment. One consequence of this experience is to bring in question the utility of the more elaborate models which seek to represent explicitly each of the several parts and paths of the earthbound portion of the hydrologic cycle. The recently published report of W.M.O. (1975) on the intercomparison of rainfall—runoff models could be considered as lending support to the view that, when the optimum values of the parameters have been chosen, simple models are as effective as the more elaborate ones. It may therefore be assumed that such conceptual models will continue, for some time at least, to provide the basis for forecasting schemes.

The W.M.O. report also indicates the need to develop objective criteria of fit which can be used to compare the performance, not only of several models on the same catchment, or of the same model on different catchments, but also of different models on different catchments.

If such criteria could be used to enable the significance of independent parts of a model to be assessed and if they could be applied to models with and without updating procedures, so much the better.

CRITERIA OF EFFICIENCY

A criterion based on "least-squares" fitting expressing the proportion of the original sum of squares of the discharge which is accounted for by the model has been proposed (Nash and Sutcliffe, 1970).

If q represents the series of actual discharges at prescribed intervals and q' the corresponding discharges computed by the model, the residual sum of squares is defined as:

$$F_1^2 = \Sigma(q - q')^2 \tag{1}$$

which may be compared with the initial sum of squares:

$$F_0^2 = \Sigma(q - \overline{q})^2 \tag{2}$$

where $\overline{q}$ is the mean of the observed discharges. F_0^2 expresses the scatter associated with the crudest possible prediction, namely the mean observed discharge $\overline{q}$, and the efficiency of the model is then defined as:

$$R^2 = (F_0^2 - F_1^2)/F_0^2 \tag{3}$$

This criterion has indeed been used, but has been criticised on the grounds that even poor models produce relatively high values, (80 or 90%), and the best models do not produce values which, on first examination, are impressively higher. This objection could be met by taking a suitable function of R^2 as defined (the recriprocal of the difference between R^2 and unity, i.e. F_0^2/F_1^2 would serve). There are, however, at least two, more valid, objections to the measure defined by eq. 3.

1st criticism. The alternative or "no model" forecast is unnecessarily primitive.

The conceptual model provides a means, whereby, information on rainfall and evaporation, and sometimes other input factors, is used to produce a forecast of discharge. A measure of efficiency of the model, therefore, should express in some way the improvement in the forecast obtained by taking these factors into account through the model, compared with the best forecast that could be made without them. R^2 as defined in eq.3 is a measure of model efficiency compared with a "no model" forecast:

$$q_t = \overline{q} \tag{4}$$

It is difficult, however, to conceive of circumstances which would warrant such a primitive forecast. Ordinarily, previous records on the catchment would indicate at least some seasonal variation which could be used in issuing a forecast in the absence of any better methods. Because the base F_0^2 in terms of which the model efficiency is expressed through eq. 3 corresponds to an unduly primitive state of knowledge, the criterion is too favourable to any model being tested, hence the usually observed high values of R^2 even for quite mediocre models.

2nd criticism. No distinction is made between different kinds of error.

If the errors of a rainfall—runoff model are purely random, uncorrelated quantities, there is nothing, short of restructuring the model, which can be done about them. In a sense, the model is complete. Neither is there any possibility of reducing the errors through an updating procedure. If the errors are large, this will be reflected in a low value of the efficiency.

When the errors are not purely random, an analysis of these in the calibration period may indicate constant components or components related to season (e.g., a tendency to overestimate during the dry season) or to input factors, or to the computed output or to any quantity known at the time of making the forecast. Such components, henceforth referred to as type A can be removed by minor adjustments to the model structure.

Components of errors which are related to observed (rather than computed) output cannot be so allowed for. If these are of a persistence nature, the errors already observed at the time of making a forecast can be used to compute the expected value of the error for the day of the forecast. Thus in real-time operational forecasting the persistence components of errors, henceforth referred to as type-B errors, can be partially removed by an updating procedure. If there is no persistence or if it is too weak ot extend over the lead time of the forecast, then the errors are practically random, for that lead time, and are therefore unremovable by an updating procedure.

It is reasonable to assume that when a model is being calibrated for use on a catchment, the user will examine the calibration period errors, quantify the type-A error and suitably amend the model structure, or append an algorithm to it, so that the model predictions q' are free of this error. Henceforth it is assumed that in the evaluation of eq. 1, such corrections have been incorporated in the forecast q'.

The First Modified Criterion is intended to meet the first of these two criticisms. Instead of eq.4, it is proposed to take as the base or "no model" forecast for *any given date* the mean of the observed discharges on that date in the calibration period:

$$q_d = (1/n)\,(q_{d,1} + q_{d,\,2} + \ldots + q_{d,\,n}) \tag{5}$$

where $q_{d,\,1}$, etc., refer to the discharges on date d in the first, second years, etc., of the calibration period.

The residual sum of squares associated with the prediction q_d in the verification period is:

$$F_d^2 = \Sigma(q - q_d)^2 \tag{6}$$

and the modified measure of efficiency of the conceptual model under test becomes:

$$R_1^2 = (F_d^2 - F_1^2)/F_d^2 \tag{7}$$

Such a measure of efficiency would ordinarily be smaller than that defined by eq.3, would be free of bias towards high values on catchments with highly seasonal flow regimes and would properly express the reduction in the sum of squares of error attributable to the conceptual model's use of the input factors.

This is the first modified criterion and it applies to the conceptual model considered as a means of relating the input and output data only, i.e. in the absence of any updating procedure.

The Second Modified Criterion is appropriate when it may be assumed that the model will be used, together with an updating procedure, for real-time forecasting. Thus in evaluating eq.1 the components of the errors which persist over the lead time of the forecast are also suppressed.

If persistence in the errors (e) is observed during the calibration period it may be expressed as:

$$e_t = a_1 e_{t-1} + a_2 e_{t-2} + \ldots + e_n e_{t-n} + \epsilon_t \tag{8}$$

where ϵ is a random component of mean zero.

For a forecast with a lead time of one unit the appropriate correction to q_t' becomes:

$$\hat{e}_t = a_1 e_{t-1} + a_2 e_{t-2} + \ldots + a_n e_{t-n} \tag{9}$$

and the forecast q'' becomes:

$$q_t'' = q_t' + \hat{e}_t$$

The sum of squares of error associated with this updated forecast is reduced to $\Sigma\epsilon^2$. However, the persistence over a lead time of r unit periods is weaker. The relationship may be expressed as:

$$e_t = b_1 e_{t-r} + b_2 e_{t-r-1} + b_3 e_{t-r-2} + \ldots + b_n e_{t-r-n+1} + \gamma_t \tag{10}$$

where the b's may be obtained from the a's by the application of a recursive relationship and γ is the random component.

The sum of squares of errors now becomes:

$$F_2^2 = \Sigma\, \gamma^2 \tag{11}$$

On the assumption that a similar analysis of persistence in the errors associated with the purely seasonal forecast q_d results in a reduction of F_d^2 to F_{du}^2 the second modified criterion becomes:

$$R_2^2 = (F_{du}^2 - F_2^2)/F_{du}^2 \tag{12}$$

This criterion expresses the efficiency of the conceptual model, with its appropriate updating procedure for the prescribed lead time, relative to the purely seasonal prediction with its updating procedure.

THE SSARR MODEL AND THE SANAGA CATCHMENT

For the purpose of demonstration, these criteria were evaluated for the SSARR model on the Sanaga catchment using data provided by W.M.O. This catchment was chosen because its highly seasonal regime would emphasise the distinction between the measures of efficiency now being proposed and that defined by eq.3. The SSARR model was selected because, of all the models tested on the Sanaga catchment in the W.M.O. intercomparison project, it appeared to give the best results.

The W.M.O. data tape provided the observed daily discharge of the Sanaga and the corresponding computed values using the SSARR model during the years 1962—1967 inclusive. The errors observed in this period are not independent because this was the period used in the calibration of the model. In the W.M.O. study, the model was assessed on its ability to reproduce the discharges in an independent period (the years 1968—1969 inclusive) but the observed discharges for this period were not supplied on the data tape in accordance with the original project of W.M.O.

Accordingly, the first four years of the calibration period were chosen as a new calibration period for the seasonal predictions and all analyses were carried out exclusively in this period. The last two years, 1966 and 1967, of the original calibration period were used for verification, i.e. for calculation of F_d^2, F_1^2, etc. This procedure introduced some bias in favour of the SSARR model because these two years had already been used in the calibration of that model.

The terms of eq.3 were first obtained as $F_0^2 = 0.232 \cdot 10^{10}$ and $F_1^2 = 0.250 \cdot 10^9$ yielding an unmodified criterion $R^2 = 89\%$.

For any given date the seasonal prediction was obtained from the mean of the four values on the corresponding dates in the years 1962—1965 slightly smoothed over date. The corresponding sum of squares of errors (differences between the actual values and the seasonal predictions for the years 1966 and 1967) was:

$$F_d^2 = 0.201 \cdot 10^9$$

The sum of squares associated with the SSARR model predictions during the same period was:

$$F_1^2 = 0.250 \cdot 10^9$$

The errors, or differences between the observed and SSARR computed discharges, in the years 1962—1965 were then analysed for seasonal fluctuation and when this was removed from the corresponding errors in 1966 and 1967 the sum of squares of errors was reduced to:

$$F_1^2 = 0.213 \cdot 10^9$$

The first modified criterion thus became:

$$R_1^2 = -0.059$$

The negative value reflects the fact that the seasonal forecast is better in the least-squares sense than that provided by the SSARR model even when the seasonal components of the errors of the latter have been removed.

The errors of the SSARR model in the calibration period, corrected for seasonal variation, indicated a persistence structure:

$$e_t = 1.014\, e_{t-1} - 0.298\, e_{t-2} + 0.212\, e_{t-3} + \epsilon_t$$

which provided a 5-day lead-time correction:

$$\hat{e}_t = 0.624\, e_{t-5} - 0.058\, e_{t-6} + 0.197\, e_{t-7}$$

When this correction was made to the forecast in the verification period the sum of squares of errors was reduced to:

$$F_2^2 = 0.957 \cdot 10^8$$

A similar analysis of the errors of the seasonal forecast in the calibration period indicated a persistence:

$$e_t = 1.101\, e_{t-1} - 0.037\, e_{t-2} - 0.069\, e_{t-3} + 0.009\, e_{t-4} - 0.035\, e_{t-5} + \epsilon_t$$

which provided a 5-day lead-time correction:

$$\hat{e}_t = 1.313\, e_{t-5} - 0.166\, e_{t-6} - 0.125\, e_{t-7} - 0.034\, e_{t-8} - 0.047\, e_{t-9}$$

When this correction was made to the seasonal forecast in the verification period the sum of squares of errors was reduced to

$$F_{du}^2 = 0.725 \cdot 10^8$$

corresponding to a negative value for the second criterion:

$$R_2^2 = -\,0.319$$

indicating that when the seasonal forecast and the SSARR model are compared, each with its appropriate updating procedure, the former gives better results on this catchment.

CONCLUSIONS

The proposed criteria of efficiency, which compare the conceptual model with a merely seasonal forecast based only on date, indicate clearly a result which had not previously been brought to light by the cirteria used in the W.M.O. intercomparison, viz., that on the highly seasonal Sanaga catchment, the SSARR model is less efficient than the simple seasonal prediction. This is true whether the models are compared with or without updating procedures. The result in no way reflects on the SSARR model as a general instrument. Similar indications would probably be obtained for any of the other models used in the intercomparison on this catchment. Less spectacular results would, no doubt, be obtained on catchments with less highly seasonal regimes. The results, however, do indicate the utility of the criteria being proposed.

REFERENCES

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, Part 1: A discussion of principles. J. Hydrol., 10: 282—290.

W.M.O., 1975. Intercomparison of conceptual models used in operational hydrological forecasting. Operational Hydrology Report No. 7, W.M.O., Geneva.