Journal of Hydrology **10** (1970) 282–290; © *North-Holland Publishing Co., Amsterdam*

RIVER FLOW FORECASTING THROUGH CONCEPTUAL MODELS PART I – A DISCUSSION OF PRINCIPLES*

J. E. NASH and J. V. SUTCLIFFE

Dept. of Engineering, University College, Galway and Institute of Hydrology, Howbery Park, Wallingford, Berks., U.K.

Abstract: The principles governing the application of the conceptual model technique to river flow forecasting are discussed. The necessity for a systematic approach to the development and testing of the model is explained and some preliminary ideas suggested.

Introduction

The problem of determining river flows from rainfall, evaporation, and other factors, occupies a central place in the technology of applied hydrology. It is not only the essential problem of flood forecasting but also arises, for example, in predicting the effects of proposed works on a catchment on the flow regime at the outfall. Most studies of representative and experimental basins and a large part of the effort of classical hydrology have been directed towards obtaining at least a partial solution of this problem. Routine flood forecasting requires, in addition to a model or estimate of the operation of transforming input data into discharge, a method of continuous correction of the forecast from the observed error of earlier forecasts. The two requirements are separable and may be studied separately. The present study is confined to the development of an adequate model of the transformation of rainfall and other input data into discharge.

Despite the attention which this problem has attracted over many years, the present position is far from satisfactory. Few hydrologists would confidently compute the discharge hydrograph from rainfall data and the physical description of the catchment. Nevertheless this is a practical problem which must often be faced by practising engineers. Although it would be extremely rash to think that a general answer is near or that techniques will soon be ready to produce accurate forecasts for specific catchments, recent experience

* This is the first of a series of papers which it is hoped to publish from time to time reporting the results of the continuing work in this field of the Institute of Hydrology, Wallingford, Berkshire, U.K.

with conceptual models of the runoff process has shown promise of considerable progress. In this field the pioneering work of Linsley and Crawford[1]) has been comparable in importance to the earlier work of Linsley and others[2]) in the development of classical hydrological methods. Many different models of the runoff process are now being developed. The results obtained are not always presented in a manner which makes possible a judgement of the relative efficiency of these models, nor does there appear to be any general agreement on the method of developing and testing a model for a given catchment or group of catchments. It is intended to set out in this paper, tentatively, as a basis for discussion and amendment, a systematic approach towards developing, testing, and modifying a model for a set of catchments with the development of a forecasting technique for an ungauged member of the set as a long term objective. These preliminary ideas will be modified by experience on the experimental catchments maintained by the Institute of Hydrology and by the views and experience of our colleagues in different countries. It is hoped to encourage a discussion of the general principles by which the conceptual model technique may be put to best use in this difficult but intriguing problem.

The empirical or analytical approach

The process linking rainfall and river flow is a deterministic one, in that it is governed by definite physical laws which by and large are known. It might, therefore, seem that solution of the problem in any specific case involves only the application of these laws to the measured rainfall and the boundary conditions – the physical description of the catchment and the initial distribution within it. However, many hydrologists consider this impractical. The deterrent is the complexity of the boundary conditions rather than any essential difficulty in the physical laws. Some simplification of the boundary conditions seems necessary, even now when high speed electronic computers are available.

As there is little point in applying exact laws to approximate boundary conditions, this, and the limited ranges of the variables encountered, suggest the use of simplified empirical relations. The fact that a basin is not a random assembly of different parts, but a geomorphological system whose parts are related to each other by a long common history, encourages the hope that simplified concepts may be found adequate to describe the operation of the basin in converting rainfall to runoff. If in addition the relation between this operation and the physical features of the catchment can be recognised, the operation of even an ungauged catchment might be forecast from a study of these features.

It is not clear how far this empirical approach should be taken. Some empiricism is unavoidable; few would quarrel with the use of Manning's equation for channel resistance, instead of a more exact treatment through the Navier-Stokes equations. Traditional hydrological methods have however tended to be very empirical. Little use has been made of established physical laws; instead an empirical, analytical, or in current jargon "parametric" approach has been adopted. The essence of this approach is the study of the conversion of rainfall into river flow not by synthesis of physical laws and boundary conditions, but as contained implicitly in the rainfall and runoff records. Hydrologists accept that this method cannot provide exact solutions. This does not distress them; exact answers are rarely needed. They are more concerned with the risk of extrapolating to extreme events not sampled in the records and in the difficulty in extending relations from gauged to ungauged basins.

In adopting the analytical approach, the research hydrologist need not feel that he is rejecting the scientific method. Such an approach may have a discipline of its own. Indeed the method of postulating a model, testing it, and modifying it may be looked upon as an example of the traditional scientific method, only the context being different from that in which the physicist works. In saying this, however, we must admit that our application of the analytical method in hydrology has not always been very disciplined. Traditional methods of forecasting discharge from rainfall, as distinct from forecasts based on routing hydrographs observed upstream, have tended to divide the problem into (a) forecasting volumes of runoff and (b) forecasting the time distribution of this runoff past the gauging station. The first serious attempt at forecasting volumes was probably the co-axial graphical correlation method developed by Linsley, Kohler and others in the United States[2]). This method relies on establishing empirical relations between the volumes of runoff in single floods and the corresponding volumes and durations of rainfall, indices of previous rainfall, and time of year, as independent variables. Once established, these relations are used to forecast the volume of runoff due to an observed rainfall amount in given circumstances. The distribution of this volume in time is usually attempted through the application of a unit hydrograph.

Most practical techniques of forecasting runoff from rainfall are based on these two methods, neither of which depends on the physical laws of evaporation, soil moisture movement, or fluid mechanics. Recent experience, however, seems to indicate that neither the co-axial graphical technique for forecasting runoff volumes nor the unit hydrograph technique for forecasting the time distribution of discharge appear capable of further evolution as more efficient tools or as aids to understanding the basin's role in converting rainfall to discharge. The principal cause of this sterility lies in the dependence of these techniques on the attempted identification in the hydrograph of at least two components which are usually called storm runoff and baseflow, but which are not adequately defined in terms of their physical origin. It is perhaps arguable whether there are, in fact, any distinct components or whether there is a continuum of different paths by which runoff reaches the streams. In any case the identification of components on the hydrograph is manifestly subjective. All subsequent analyses are also subjective and therefore all conclusions devoid of universal validity. This situation is not improved by the substitution of mathematical for physical definition of the components. The identification of two or three exponentials in the recession of a hydrograph would be difficult even if it were known that the recession consisted of such distinct parts, but the separation is meaningless in the absence of evidence for the number and form of these components.

Because of the base flow separation the observed volume of "storm runoff" in a single event is not the total volume of water contributed to stream flow in that event, but an arbitrary part of it. Because of this arbitrary division the empirical relations found do not usually reflect physical laws (even the principle of conservation of mass) nor has it been found possible to discover the role of soil moisture in determining runoff to the streams. Similarly, the unit hydrograph assumption of a linear time invariant relationship cannot be tested because neither the input (effective rainfall) nor output (storm runoff) are unequivocally defined.

While traditional analytical methods applied to the records of an individual catchment may provide reasonably valid relations, extension to ungauged catchments has not been possible. This extension would require finding links between the physical characteristics of the catchments and the parameters of both the rainfall-runoff conversion and the unit hydrograph. While the links between the unit hydrograph and the catchment have been studied many times, similar studies of the parameters of the volumetric process have not been reported. Further progress would seem to depend on the rejection of the a priori division of hydrographs into ill defined components and on greater emphasis being placed on modelling the real physical processes. Where empirical methods are required they must be more disciplined and systematic.

Rejection of a priori base flow separation introduces difficulties into the traditional analytical approach. The persistence of discharge over long periods when no recharge occurs implies a very slow recession of discharge which makes it difficult to distinguish the discharge associated with each rainfall event. If this analysis cannot be made, the rainfall cannot be divided into "losses" and "effective rainfall" and hence the two parts of the relationship "rainfall to effective rainfall" and "effective rainfall to discharge" can no longer be studied separately but must be treated simultaneously. A model of the process of conversion of rainfall into river flow must be assumed. Its parameters may be evaluated either by physical measurement or by optimisation by successive adjustment until the model reflects as nearly as possible the operation of the basin as shown by the records of rainfall and discharge.

Requirements in a model

If a model were required solely to forecast the flow from a particular basin, it would probably be adequate to specify the model's form and parametric values such that the computed output was a sufficiently close reproduction of the observed output. If the model is also to help us to understand the process of converting rainfall into discharge and the relative importance of different elements in this process, and particularly if it is hoped eventually to use the model for basins without records by establishing relations between the model parameters and basin characteristics, it is essential to obtain some guide to the relative significance of model parts and the accuracy of parametric values. Methods of measuring significance and accuracy of determination must be found which are applicable to complex non-linear models.

Although simplification of the operation of a basin is necessary, especially in terms of variability over the area, it is desirable that the model should reflect the physical reality as closely as possible. If it is hoped to transfer the model to an ungauged basin the parametric values can be determined only by measuring the physical characteristics of the basin. Therefore the further the operation of the model departs from known physical laws the more tenuous is likely to be the relationship between model parameters and basin characteristics. On the other hand if the model parameters are to be fixed by optimisation or comparison of computed and observed outputs, the more detailed and complex the model the more difficult it becomes to establish the values of the parameters, particularly if these are interdependent. This conflict cannot be resolved entirely, but there should be no unnecessary proliferation of parameters to be optimised and model parts with similar effects should not be combined. If, for example, two separate coefficients were introduced to allow for possible systematic errors in measurement of rainfall and discharge, it would probably be very difficult to evaluate both these coefficients accurately, though either alone might be estimated very precisely.

The requirement of versatility should be added to those of simplicity and lack of duplication. Each additional part of a model must substantially extend the range of application of the whole model. In other words, we are prepared to accept additional parts and hence greater difficulty in determining parametric values only if the increased versatility of the model makes it much more likely to obtain a good fit between observed and computed output.

Isolation of operations within the model

If one or more functions of the model could be isolated and the relevant parameters optimised one would expect to obtain a much higher accuracy of determination than if the model were optimised as a whole. This might be attempted in many ways. For example, a single part of the model or the value of a single parameter may alone determine the actual evaporation for a given input. Such a parameter might be optimised by choosing a value to make the total computed evaporation over several years equal to the total rainfall less the total observed discharge. In general terms, if a function of known value is dependent on a single model parameter and independent of the others, this parameter can be optimised by ensuring that the model correctly evaluates the function. If no such function is dependent on a single parameter, there may be a function of a limited number of parameters which may be similarly used. For instance, the volumes of runoff in individual storms may be determined by a part of the model which is independent of the part associated with the damping effect. In this case the parameters of each part can be optimised separately. Such possibilities must be borne in mind during the initial specification of the model.

Fitting the model

To remove subjectivity in fitting the model to the data or in determining the parametric values, O'Donnell[3]) suggested automatic optimisation. This involves successive changes of parameter values according to some pre-conceived rule or pattern of increments which takes into account the results of previous steps and in particular whether or not a change improved the fitting.

Clearly optimisation needs an index of agreement or disagreement between the observed and computed discharges. Linear regression analysis suggests a sum of squares criterion such as:

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

where F^2 is the index of disagreement and q and q' are the observed and computed discharges at corresponding times. The sum may be taken over all q's at intervals Δt, or at preselected times such as peaks or troughs in the hydrograph. F^2 is analogous to the *residual variance* of a regression analysis.

The *initial variance* F_0^2 defined by

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

where $\bar{q}$ is the mean of the observed q's and the sum is taken as before, may also be defined as the "no model" value of F^2. This enables the *efficiency of a model* to be defined by R^2 (analogous to the coefficient of determination) as the proportion of the initial variance accounted for by that model.

$$R^2 = \frac{F_0^2 - F^2}{F_0^2} \tag{3}$$

The *efficiency of a separable model part* may be judged by the change in R^2 which follows insertion of the part or by the proportion of the residual variance accounted for by its insertion:

$$r^2 = \frac{F_1^2 - F_2^2}{F_1^2} = \frac{R_2^2 - R_1^2}{1 - R_1^2} \tag{4}$$

where the suffixes 1 and 2 denote before and after insertion of the model part under consideration.

It would be useful to have an objective significance test for R^2 and r^2. The standard test for the significance of a correlation coefficient could perhaps be used if the number of degrees of freedom were known. However, some model parts (*e.g.*, those dealing with snow melt) may effect the output in only a few places; hence the number of degrees of freedom associated with such parts is determined by the number of events so affected (less the number of parameters) rather than the number of events in the record. Nevertheless R^2 and r^2 are useful indices for a general interpretation of the efficiency of a model or a part.

The quantity F^2 is a function of the parameter space and, of course, of the input and output. Optimisation involves finding the values of the parameters which minimise F^2. This may be done by a "steepest descent" method, or a search can be conducted in the super space by moving parallel to the parameter axes.

If F^2 contained only one minimum in the parameter space, steepest descent methods would doubtless find it most quickly; where several minima may exist and consequently the whole space must be searched the advantages of such methods are reduced.

The shape of the F^2 surface in the vicinity of the optimum point may be used as an indication of the *stability* (the inverse of the sampling variance) of the optimum value of the parameters. In Fig. 1(a) a cross-section through an optimum and parallel to one axis (X) is shown.

Clearly the greater the radius of curvature at X_{opt} the less well is X defined.

Fig. 1(b) shows the same cross-section through $\partial F^2/\partial X$. The intersection with the X axis is at X_{opt}. The higher the angle of intersection ($\partial^2 F^2/\partial X^2$) the better is the definition of X_{opt}. Therefore the value of the second derivative at X_{opt} is an index of the stability of X_{opt}. This relationship should be developed so that the index becomes a measure of stability.

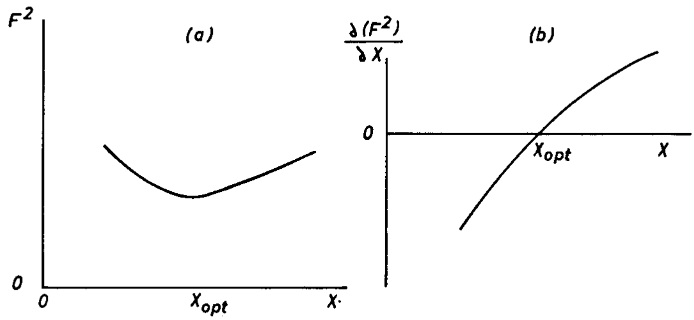

Fig. 1. The sampling variance of X_{opt}.

If the several parameters are mutually independent, then the index of stability of the optimised values may be obtained by considering the second derivative parallel to each axis only. If, however, substantial dependence exists between two or more parameters this is not sufficient. In Fig. 2 the dependence of F^2 on X and Y is indicated by a set of contours. Dependence between X and Y is indicated by the valley in the surface roughly along $X+Y=\text{const}$.

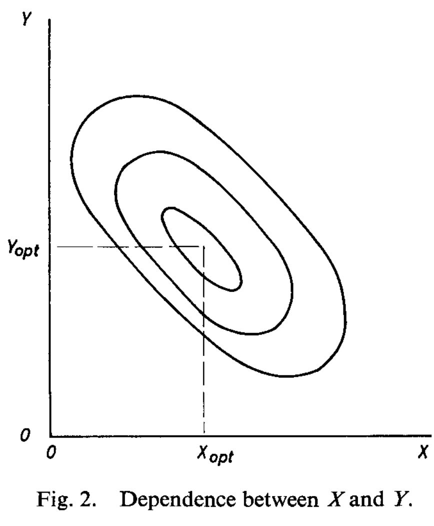

Fig. 2. Dependence between X and Y.

The optimum values X_{opt} and Y_{opt} may indeed be found and cross-sections parallel to the axes might look like those of Fig. 1. A cross-section along the valley, however, would appear very different, indicating that while a function of X_{opt} and Y_{opt} is well defined, the separate values are not. The occurrence of such a relationship could be discovered only by taking the second derivative in all directions through the minimum point of F^2.

Progressive modification

If one accepts that it is desirable to have a simple rather than a complex model, and this is certainly true if it is hoped to obtain stable values of the optimised parameters, then it would seem that a systematic procedure would be as follows:

(1) Assume a simple model, but one which can be elaborated further.
(2) Optimise the parameters and study their stability.
(3) Measure the efficiency R^2.
(4) Modify the model – if possible by the introduction of a new part – repeat (2) and (3), measure r^2 and decide on acceptance or rejection of the modification.
(5) Choose the next modification. A comparative plotting of computed and observed discharge hydrographs may indicate what modification is desirable.
(6) Because all models cannot be arranged in increasing order of complexity it may be necessary to compare two or more models of similar complexity. This may be done by comparing R^2.

References

1) N. H. Crawford and R. K. Linsley, A conceptual model of the hydrologic cycle, Int. Assoc. Sci. Hydr., Publ. No. 63 (1964) 573–587
2) R. K. Linsley, M. A. Kohler, and J. L. H. Paulhus, Applied Hydrology (McGraw-Hill, New York, 1949)
3) T. O'Donnell, Computer evaluation of catchment behaviour and parameters significant in flood hydrology, I.C.E. Symp., River Flood Hydrology (1966) 103–113