

NOAA Technical Memorandum NWS HYDRO-36



---

DETERMINATION OF FLOOD FORECAST EFFECTIVENESS  
BY THE USE OF MEAN FORECAST LEAD TIME

Silver Spring, Md.  
August 1977

---

**noaa**

NATIONAL OCEANIC AND  
ATMOSPHERIC ADMINISTRATION

National Weather  
Service

NOAA TECHNICAL MEMORANDA

National Weather Service, Office of Hydrology Series

The Office of Hydrology (HYDRO) of the National Weather Service (NWS) develops procedures for making river and water supply forecasts, analyzes hydrometeorological data for planning and design criteria for other agencies, and conducts pertinent research and development.

NOAA Technical Memoranda in the NWS HYDRO series facilitate prompt distribution of scientific and technical material by staff members, cooperators, and contractors. Information presented in this series may be preliminary in nature and may be published formally elsewhere at a later date. Publication 1 is in the former series, Weather Bureau Technical Notes (TN); publications 2 to 11 are in the former series, ESSA Technical Memoranda, Weather Bureau Technical Memoranda (WBTM). Beginning with 12, publications are now part of the series, NOAA Technical Memoranda, NWS.

Publications listed below are available from the National Technical Information Service, U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151. Price: \$3.00 paper copy; \$1.45 microfiche. Order by accession number shown in parentheses at end of each entry.

Weather Bureau Technical Notes

TN 44 HYDRO 1 Infrared Radiation from Air to Underlying Surface. Vance A. Myers, May 1966. (PB-170-664)

ESSA Technical Memoranda

WBTM HYDRO 2 Annotated Bibliography of ESSA Publications of Hydrological Interest. J. L. H. Paulhus, February 1967. (Superseded by WBTM HYDRO 8)

WBTM HYDRO 3 The Role of Persistence, Instability, and Moisture in the Intense Rainstorms in Eastern Colorado, June 14-17, 1965. F. K. Schwarz, February 1967. (PB-174-609)

WBTM HYDRO 4 Elements of River Forecasting. Marshall M. Richards and Joseph A. Strahl, October 1967. (Superseded by WBTM HYDRO 9)

WBTM HYDRO 5 Meteorological Estimation of Extreme Precipitation for Spillway Design Floods. Vance A. Myers, October 1967. (PB-177-687)

WBTM HYDRO 6 Annotated Bibliography of ESSA Publications of Hydrometeorological Interest. J. L. H. Paulhus, November 1967. (Superseded by WBTM HYDRO 8)

WBTM HYDRO 7 Meteorology of Major Storms in Western Colorado and Eastern Utah. Robert L. Weaver, January 1968. (PB-177-491)

WBTM HYDRO 8 Annotated Bibliography of ESSA Publications of Hydrometeorological Interest. J. L. H. Paulhus, August 1968. (PB-179-855)

WBTM HYDRO 9 Elements of River Forecasting (Revised). Marshall M. Richards and Joseph A. Strahl, March 1969. (PB-185-969)

WBTM HYDRO 10 Flood Warning Benefit Evaluation - Susquehanna River Basin (Urban Residences). Harold J. Day, March 1970. (PB-190-984)

WBTM HYDRO 11 Joint Probability Method of Tide Frequency Analysis Applied to Atlantic City and Long Beach Island, N.J. Vance A. Myers, April 1970. (PB-192-745)

NOAA Technical Memoranda

NWS HYDRO 12 Direct Search Optimization in Mathematical Modeling and a Watershed Model Application. John C. Monro, April 1971. (COM-71-00616)

NWS HYDRO 13 Time Distribution of Precipitation in 4- to 10-Day Storms--Ohio River Basin. John F. Miller and Ralph H. Frederick, May 1972. (COM-72-11139)

NWS HYDRO 14 National Weather Service River Forecast System Forecast Procedures. December, 1972. COM-73-10517)

(Continued on inside back cover)

NOAA Technical Memorandum NWS HYDRO-36

DETERMINATION OF FLOOD FORECAST EFFECTIVENESS  
BY THE USE OF MEAN FORECAST LEAD TIME

Walter T. Sittner

Silver Spring, Md.  
August 1977

UNITED STATES  
DEPARTMENT OF COMMERCE  
Juanita M. Kreps, Secretary

NATIONAL OCEANIC AND  
ATMOSPHERIC ADMINISTRATION  
Robert M. White, Administrator

National Weather  
Service  
George P. Cressman, Director





CONTENTS

Abstract . . . . . 1

Introduction . . . . . 1

The concept of mean forecast lead time . . . . . 2

Method of computation . . . . . 3

The effect of forecast accuracy . . . . . 5

Forecast revisions and refinements . . . . . 7

Summary of computational rules . . . . . 8

The effect of timing errors . . . . . 9

The use of forecast precipitation in river forecasts . . . . . 11

Maximum attainable values of mean forecast lead time . . . . . 12

Applications and use of mean forecast lead time . . . . . 19

References . . . . . 22



DETERMINATION OF FLOOD FORECAST EFFECTIVENESS BY THE USE OF  
MEAN FORECAST LEAD TIME

Walter T. Sittner  
Hydrologic Research Laboratory  
Office of Hydrology  
National Weather Service, NOAA

ABSTRACT. A method of evaluating flood forecasts is presented. The method expresses the value, to the user, of a series of forecasts that relate to a single flood event. The timeliness of the forecasts and their accuracy are combined into a single numerical score, expressed in units of time, and termed "mean forecast lead time." The score reflects the manner in which a particular combination of timeliness and accuracy affects the user. The system purports to produce a measure of forecast effectiveness that is physically meaningful and that is more closely related to economic benefit than are error statistics based on the difference between forecast and observed hydrographs.

The relationships between the evaluation system and various forecasting problems and practices are discussed.

INTRODUCTION

It is generally agreed that a necessary adjunct to any forecasting operation is some type of evaluation program. The objective of such a program may be the monitoring of the forecast operation itself, the data collection effort, or the dissemination system. The objective may be the evaluation of new forecasting techniques and/or equipment, or it may be the determination of the value to the user of the final product of the entire system. Obviously, the technique used must be devised and applied in such a way that it will reflect those factors that one desires to measure.

Most existing evaluation programs are actually verification systems. That is, they are designed to measure the degree of agreement between the predicted value of a variable and a value that actually occurs at some later time. This is usually accomplished by producing some sort of statistical summary of the differences between predicted and observed values of the variable. The aim of a forecasting system is to make the most accurate forecasts possible consistent with technological and fiscal constraints. Success in this would be indicated by a verification score that is numerically small. The evaluation of a forecast operation in this way has some validity if the objective is to determine the relative accuracy of different forecast models or techniques. If, on the other hand, the aim is to evaluate the effectiveness of the forecast as it relates to the user, such an approach

has serious deficiencies. It is, for example, difficult to compute an error function that is meaningful because a forecast event does not normally consist of one prediction and one event. There are, in fact, usually several predictions made at various times prior to the event, and it is axiomatic that the earlier ones will show a lesser degree of accuracy than those made only a short time before the event occurs. The problem of combining the various errors with their time intervals in such a way as to reach a meaningful conclusion is one that is not easily solved.

In the case of flood forecasting, if forecasts are based only on observed precipitation, rather than predicted precipitation, the crest cannot be predicted until the rain has ended. Unless a storm is very brief, however, a number of preliminary forecasts are usually made during the storm on the basis of precipitation observed up to the time of forecast preparation. The stages called for by such forecasts will then occur as points on the rising limb of the hydrograph rather than as the crest. These forecasts serve a useful purpose by advising the user that the water will be "at least this high," but they are not verifiable by an observed hydrograph. Finally, what is the significance of a specified stage error? If the series of forecasts related to a flood event has an average error, determined in some meaningful way of 0.2 m, is that good or is it bad? Without a great deal of supplementary information, one would not really know.

The evaluation system to be presented is based on an entirely different concept. It results in a score that is, physically, a warning time interval rather than an error function. A superior forecast operation is indicated by a score that is numerically large rather than by one approaching zero.

This measure, called "Mean Forecast Lead Time (MFLT)," is, as the title implies, intended only for use with flood forecasts and not for low water or other types of river predictions. It is intended to indicate the value, available to the users, of a flood forecast or the group of forecasts relating to a flood event. It should be noted at this point that this is the value available to the user. Such value will not actually accrue to him unless he reacts to the forecasts in a suitable manner. MFLT does not reflect the behavioral pattern of the forecast recipient. The potential value of the forecasts is expressed by MFLT without regard to the reason for the forecasts being good or being poor. The effect of sparse or erroneous precipitation reports is indistinguishable from the effect of a poor hydrologic model. The work of a lucky forecaster results in as good an evaluation score as that of a skillful forecaster, but perhaps not as consistently.

#### THE CONCEPT OF MEAN FORECAST LEAD TIME

In concept, MFLT is very simple. A forecast is a statement regarding an event, made prior to the occurrence of the event. It derives its value from the fact that it is made prior to the event. Consequently, that value can be measured by the length of the interval from forecast to occurrence, or "lead time." Obviously, the value of a forecast is diminished if there is not reasonable agreement between it and the future events that actually occur. Therefore, if the concept of using lead time as a measure of forecast

effectiveness is to be valid, there must be provision for an adjustment to the measure when a forecast exhibits a low level of accuracy. MFLT is, in essence, the average warning time provided by a group of forecasts, suitably adjusted for forecast inaccuracy. The basis for the adjustment is the effect of the inaccuracy on the user. What the system attempts to do is to determine the lead time that an accurate forecast would have to have had in order to affect the recipient in the same manner as did the inaccurate forecast that was actually issued and then to use this "equivalent" lead time in the averaging process. MFLT can then be defined as follows:

MFLT is the average warning time that would be provided by a group of error-free forecasts that would have affected the users in the same manner as did the group of forecasts actually issued.

The manner in which lead times are computed and the treatment of forecast errors are discussed in the following sections.

#### METHOD OF COMPUTATION

Lead time is, as previously stated, the interval from forecast issuance to the occurrence of the event. If a flood event involves a series of forecasts, each calling for a successively higher stage, the lead time for each forecast is the interval from the issuance of that forecast to the occurrence of the stage called for. The MFLT is computed by averaging all such intervals.

Figure 1, a relatively simple case, illustrates the basic method. The rise, from a base stage of 0.6 m to a crest of 8.0 m, is caused by 175 mm of rainfall in a 24-hr period as shown. The four 6-hr increments, ending at 1300 and 1900 on day 1 and at 0100 and 0700 on day 2, are 20 mm, 75 mm, 50 mm, and 30 mm. It is assumed in this example that a forecast can be issued 2 hr after the precipitation observations are made. It is further assumed that each forecast is a perfect hydrologic analysis. That is, if the rain were to cease at observation time, the forecast hydrograph would agree exactly with the observed hydrograph. In this example, then, three forecasts were issued, at 2100 on day 1 and at 0300 and 0900 on day 2. (The 20-mm rainfall observed at 1300 on day 1 was not sufficient to warrant the issuance of a flood forecast.) The first forecast was based on the 12-hr rainfall totalling 95 mm, and called for a maximum stage of 4.7 m, which is 0.4 m above flood stage. The second forecast, based on 145 mm of rainfall, called for 7.0 m and the last forecast, based on all 175 mm, correctly predicted the observed crest of 8.0 m. In figure 1, each forecast is plotted with the time of issuance as abscissa and the predicted stage as ordinate. The lead times shown, the intervals from issuance to the occurrence of the stage, are 11.5 hr, 12.3 hr, and 13.0 hr. The MFLT is then the average of the three individual lead times, or 12.3 hr.

There are two ways of thinking of this quantity. Flood stage, 4.3 m in this case, is defined as the lowest stage at which damage occurs. Consequently, the first forecast, 4.7 m, provides a firm warning to those people who become vulnerable between 4.3 and 4.7 m. It also provides an indication of danger to those located slightly above 4.7 m, but with a lesser degree of certainty.

If the appropriate precautionary measures for those in the affected group are of the threshold type, complete evacuation of the property, shutting down a facility, etc., then, this forecast, the first of the three, is the one that provides them with their warning, and it gives them a lead time of 11.5 hr in which to make their preparations. The second forecast, 7.0 m, serves an additional group, those who become affected between approximately 4.7 and 7.0 m, but does little for those who are situated at, say, 7.5 m. The third forecast is the one that provides a warning to them. Thus, MFLT can be thought of as the average warning time given to the community as a whole.

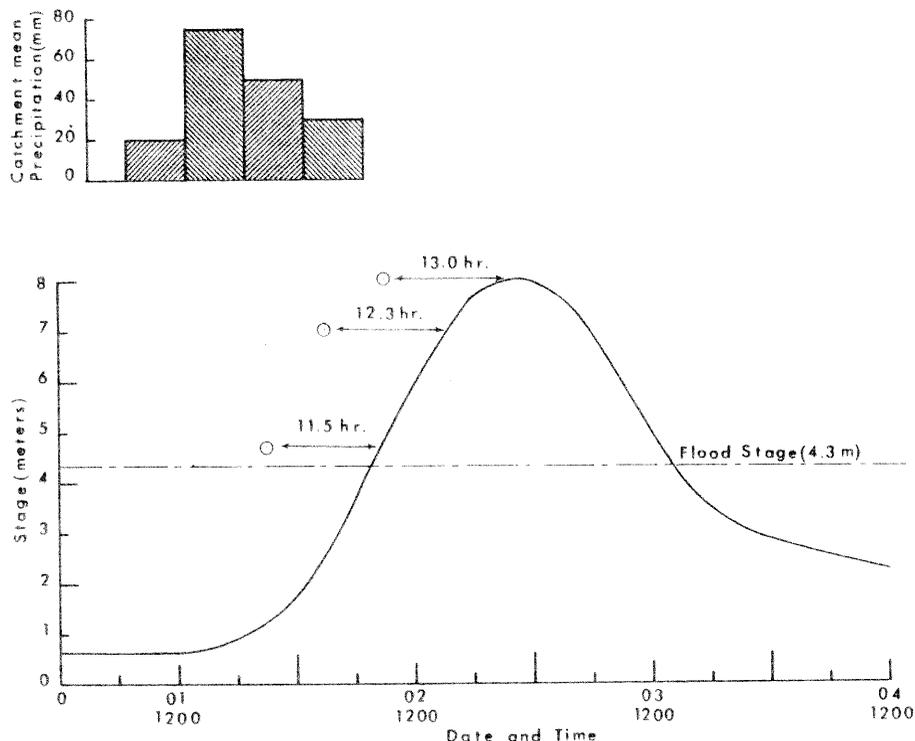


Figure 1

In another light, consider the individual user who becomes vulnerable to flood waters at a certain stage but the extent of whose precautions vary with stage above that initial point. An example might be a store with floor elevation of 5 m and shelves spaced 0.2 m apart. Upon receipt of the 7-m forecast, the owner will remove the contents of the first 10 or 12 shelves. When the 8-m forecast is received 6 hr later, he will empty the top shelves. While the 8-m forecast is the one that defined the ultimate action to be taken, the earlier forecast was also of value since it enabled him to empty the first 10 shelves and stay ahead of the rise. Thus, for this type of user, MFLT is a measure of the average value, to one individual, of all of the forecasts that call for stages within his range of interest.

It might be noted at this point that MFLT, computed as described, is not the true average warning time for all flood plain occupants. That is, in the example of figure 1, an occupant situated at 4.3 m (flood stage) has not had 11.5 hr warning, but only 10.5 hr, since the first forecast calling for

a stage of 4.7 m was issued at 2100 on day 1 and flood stage was reached at 0730 on day 2. Then, the average warning time for all occupants located between 4.0 m and 4.3 m was not 11.5 hr but  $(10.5+11.5)/2$ , or 11.0 hr. Similarly, individuals situated slightly above 4.7 m received their first firm warning from the second forecast, issued only 5.5 hr prior to the inundation at this level. Then, the entire group between 4.7 m and 7.0 m had an average of  $(5.5+12.3)/2$ , or 8.9 hr, warning time. If it were to be assumed that the vertical distribution of forecast users was uniform, the true average warning time for all individuals involved would be 9.5 hr rather than 12.3 hr as previously determined. This discrepancy, although systematic in nature, is not considered serious enough to warrant the added complexity of integrating lead times along the rising limb of the hydrograph. Doing so would probably not produce a much better determination of MFLT since it is not likely that the vertical distribution of economic interests would actually be uniform through the range of flooding. Furthermore, as was pointed out earlier, a forecast stating that river levels will reach a certain stage does serve to at least alert individuals located at slightly higher stages to the possibility that they too will be flooded. In addition, it will be shown later that there are computational procedures for MFLT that, in some circumstances, tend to minimize the effect of not performing such an integration.

#### THE EFFECT OF FORECAST ACCURACY

It was pointed out earlier that the use of lead time as a measure of forecast effectiveness must either presuppose a suitable degree of accuracy or must provide for an adjustment to the measure when such accuracy is not achieved. Since the former supposition would not be realistic, the latter provision must be made.

In the example of figure 1, each forecast is assumed to be the result of a perfect analysis. Considering the second forecast, this means that had the rain ceased at 0100 on day 2 the crest would have been exactly 7.0 m. Now, consider the effect of an error in that analysis, an error, for example, of +0.3 m. The forecast then would have called for 7.3 m rather than 7.0 m and, in light of subsequent events, would actually have been of greater value to the user. The computed lead time for this forecast would have been from issuance at 0300 to the occurrence of 7.3 m at 1630, or 13.5 hr, rather than the 12.3 hr shown in figure 1. The MFLT for the event would then have been 12.7 hr rather than 12.3 hr. If the error in the analysis had been minus 0.3 m, the forecast would have been 6.7 m, the lead time 11.4 hr, and the MFLT 12.0 hr. The negative error resulted in a less accurate prediction of future events and therefore would have hurt the user rather than helped him. This is reflected in the lower score. As was pointed out earlier, hydrologic analyses based on only a portion of the storm rainfall are not verifiable by the observed hydrograph, and their accuracy can, therefore, not be determined. With the MFLT evaluation system, however, such an accuracy determination is not needed. An inaccuracy in the hydrologic analysis may hurt the user or it may help him. Whatever the effect, it will be automatically reflected in the MFLT score.

The foregoing discussion of the effect of inaccuracy in forecasts made during the storm does not apply to a forecast made after the storm. This last forecast is verifiable; its accuracy may be easily determined; and, if it is in error, the user will be affected adversely whether that error be positive or negative.

Consider the effect of a negative error in the third forecast of figure 1. That is, let the predicted stage be 7.5 m rather than 8.0 m. The lead time for this forecast would then be from 0900 to 1720, or 8.3 hr. Averaging the three lead times for this event would give  $(11.5+12.3+8.3)/3$ , or 10.7 hr. This, however, is not the MFLT because the error in the third forecast has brought a new factor into play; the observed 8.0-m crest has not been predicted. To reflect this, a fourth forecast, which correctly predicted the crest, is assumed to have been issued at the time the crest occurred. That is, the occurrence of an unpredicted stage has the same effect on the user as a correct forecast with zero lead time. The MFLT for this event is then  $(11.5+12.3+8.3+0)/4$ , or 8.0 hr. Thus, in a case where the observed crest is higher than the highest forecast issued, a "low miss" has occurred and its effect on the user is recognized by including one zero lead time in the computation of MFLT.

Now, consider the case in which the last forecast misses the crest but involves an error which is positive rather than negative. For instance, assume that the third forecast of figure 1 had called for 8.5 m, an error of +0.5 m. The definition of lead time for an individual forecast has been given earlier as the interval from issuance to the occurrence of the predicted stage. In this case, the predicted stage, 8.5 m, does not occur, and this definition becomes meaningless. In such circumstances, the lead time is computed in such a way that the positive error has the same effect on the score as a negative error of like magnitude. That is, the lead time is taken as the interval from forecast issuance to the occurrence of a stage that is as far below the crest as the predicted stage is above the crest. Such a stage is 7.5 m, and the lead time is 8.3 hr, the same as for the forecast with an error of -0.5 m.

In the case of the last forecast being a "low miss," the computation of MFLT included one zero lead time because users situated at the 8.0-m level received no warning. In the case of the "high miss," however, they did receive warning. The previous reasoning then would lead to the conclusion that a zero lead time should not be included in the case of the "high miss" and that the MFLT should be determined as  $(11.5+12.3+8.3)/3$ , or 10.7 hr. Note that if this is done the positive error in the last forecast results in a better score than a negative error of the same magnitude. Such an effect may be desirable if one is of the opinion that over-forecasts are preferable to under-forecasts. Whether or not they are is a rather moot point. The author is of the opinion that they are not and that a positive error is just as bad as and no worse than a negative error of the same size. A related consideration is that an evaluation system should be noncorrupting. That is, it must be designed in such a way that it does not encourage the forecaster to develop habits which enhance the scores without consistently improving the service to the user. An evaluation system that treats a positive error more favorably than a negative one will probably cause a

forecaster to shade his predictions upward rather than giving the user his best estimate of what is going to happen. The long-term effect of such a practice is likely to be unsatisfactory. For this reason, it is recommended that when a high miss occurs a zero lead time be included in the MFLT computation just as is done with a low miss. This will cause an over-forecast to reduce the score to exactly the same degree as an under-forecast with the same size error.

This discussion of errors in the final forecast has assumed that if an observed crest is 8.0 m, a forecast of 7.5 or 8.5 m does not predict that crest. In actual practice, things are a bit more complicated. It is normally understood by all concerned that a forecast of 7.5 m is not intended to mean that the crest will be exactly that value. Small differences are expected, and an observed crest of 7.4 m would probably not be considered a "miss." What is involved here is that any forecast, regardless of how stated, actually indicates a range of values rather than one specific value. If a forecast is issued as, say, "7.4 to 7.7 m," then a bracket has been expressed; and if the crest falls within that bracket, it would not be considered a "miss." If a forecast is expressed as a single number, then a bracket has not been expressed, but the implication of a bracket still exists, and it is necessary to establish the size of that "implied bracket" if the principles described above are to be applied. There are many ways in which such brackets might be determined. The bracket might be the range of stage corresponding to a specified percentage change in discharge. Or, for a forecast point with an unstable stage-discharge relationship, it may be a fixed amount reflecting the degree of the instability. It could be based on the slope of the stage-damage curve, or it may be simply a fixed percentage of the stage involved. The important thing is that forecasts be treated as a range of stage rather than as a single value and that the range be established in a manner that is physically meaningful, consistent, and reasonably objective.

#### FORECAST REVISIONS AND REFINEMENTS

In the examples discussed, a forecast event was shown to involve a series of individual forecasts, some made during the storm and one made after the storm. These forecasts differed from each other because each was based on a different rainfall accumulation. In actual practice, several forecasts may be issued after the end of the storm, all based on the final rainfall accumulation, but differing from each other. Such forecasts are based in part on observed discharge data from upstream points or from the point in question. This information permits the forecaster to revise or refine the forecast as the crest approaches. The treatment of such revisions in the evaluation system is, with one exception, no different from the treatment of other forecasts. A revision is an updated issuance, presumably more reliable than the earlier forecasts. The fact that it is issued because of the availability of later and more complete data rather than because of additional rainfall is of no interest to the user. The distinction has no effect on the user and is therefore not considered in the evaluation.

The exception to the foregoing is the case in which a later forecast is not a revision but a refinement. Consider again the example of figure 1. Suppose that the third forecast had been issued as "7.9 to 8.3 m." The bracket of 0.4 m expresses the degree of uncertainty existing at 0900, the

time of issuance. Now, suppose that 5 hr later additional information enables the forecaster to reduce that uncertainty to 0.2 m and issue a forecast at 1400 of "7.9 to 8.1 m." This forecast is of value to the user and should be issued. Its lead time, however, is only 8.0 hr, and, if it were to be included in the computation of the score, the MFLT would be  $(11.5+12.3+13.0+8.0)/4$ , or 11.2 hr, less than the value obtained by using only the first three issuances. Obviously, then, a forecast that is a refinement rather than a revision should not be included in the computation of MFLT. The distinction is made by comparing the bracketed forecasts. If the entire bracket of the later forecast falls within the bracket of the earlier forecast, it is a refinement and not a revision. It should be noted here that it is possible for such a refinement to miss the crest where the earlier forecast with its larger bracket would have been a "hit." In such a case, it is actually a different forecast, one which serves the user poorly, and it should be included in the computation.

#### SUMMARY OF COMPUTATIONAL RULES

An evaluation system must meet certain basic criteria. As explained earlier, it must be noncorrupting. In addition, it must be completely objective and clearly describe the method of determining the score under any conceivable combination of circumstances. The basic concept of MFLT and its major provisions have already been explained in some detail. In this section, is presented a set of computational rules that describe, in a nonambiguous manner, the method of computing MFLT in any type of situation. These rules have been rather carefully thought out and, to the best of the author's knowledge, do constitute an objective procedure that will not yield anomalous results under any circumstances. This is not to imply, however, that they are perfect or even the best rules that might be devised. It is quite possible that in implementing the MFLT system one would wish to change some of them. This should be done with caution since some rather subtle interrelationships exist. Changes should be made only after carefully considering the effect under a wide variety of operational conditions.

$$MFLT = (\sum I)/N$$

where I is the time interval from issuance of forecast to the occurrence of the predicted stage; and

N is the number of forecasts applicable to the event.

The verification bracket is "VB." It is determined, as described earlier, by any objective technique selected by the user. Each forecast is considered to involve a range of stages from the specified stage minus one-half VB to the specified stage plus one-half VB. If a forecast is stated as a range, the "specified stage" is the midpoint of that range.

A "high miss" occurs with each forecast that exceeds the observed crest by more than one-half VB.

A "low miss" occurs when there are no "high misses" and when the last forecast issued is below the observed crest by more than one-half VB.

Subject to the following restrictions, the computation is to include all forecasts that predict the crest or points on the rising limb of the hydrograph.

A. Forecasts for levels below flood stage are not included.

B. If more than one forecast predicts the same stage (entire bracket of a later forecast is within the bracket of an earlier forecast), the later forecast is not included unless it is a miss.

C. If forecasts for a series of points on the rising limb of the hydrograph are issued at the same time, only the one calling for the latest point on the hydrograph is included.

If a "low miss" occurs, one zero "I" value is included in the computation.

For a "high miss," "I" is the time interval from issuance to the occurrence of a stage that is as far below the observed crest as the predicted stage is above it. Note that it is possible for the interval so defined to be negative. Note also that if the forecast exceeds the observed crest by more than the difference between the crest and base stage, the stage described does not exist. In this case, the MFLT for the entire event is zero.

For each "high miss" that is not followed by a forecast that successfully predicts the crest, one zero "I" value is included in the computation.

If flood stage is reached before the first flood forecast is issued, one zero "I" value is included in the computation.

If a secondary rise is involved, the first occurrence of a stage is to be used.

If flood stage is reached and no flood forecasts are issued, this constitutes an event with zero MFLT.

If one or more flood forecasts are issued and flood stage is not reached, the MFLT for the event is computed as described above without regard for the fact that all points on the observed hydrograph are below flood level.

If the computed MFLT for an event is negative, then  $MFLT = 0$ .

Note: The conversion of a negative MFLT to zero expresses the thought that a zero MFLT indicates that the set of forecasts was of no value and that this is the worst possible situation. Quite possibly, a set of forecasts so grossly in error as to result in a negative MFLT might be considered to serve the user in a worse manner than would no forecasts at all. If it is desired to express this concept, then the negative score does have meaning and should be retained.

#### THE EFFECT OF TIMING ERRORS

In the preceding discussion, a point that has not been mentioned is that a flood forecast normally consists of a predicted stage and in addition a prediction of the time at which that stage is to occur. What has been considered thus far is the effect on the user of the combination of predicted

and observed stages with no regard for the fact that there may be a discrepancy between the predicted time of occurrence and the actual time of occurrence.

Such discrepancies, or timing errors, are relatively unimportant and, quite possibly, can be ignored. When a stage above flood level is predicted to occur at a certain time, a user may be expected to take precautions as soon and as fast as is prudently possible. He does not normally delay the start of his measures because the predicted interval is somewhat greater than the time required to complete those measures. Consequently, a timing error of a few hours in an otherwise good forecast is not of great importance.

On the other hand, under conditions of continuing rainfall, timing errors tend to be systematic rather than random. As an illustration of this, consider figure 2. Note that this is the same event as is shown in figure 1. In addition to showing the forecasts plotted at the time of issuance, figure 2 also shows the computed hydrographs (dotted) on which the first two forecasts are based. This shows clearly that the rain which fell after forecast preparation, causing stages to exceed those predicted, also caused the predicted stages to occur considerably earlier.

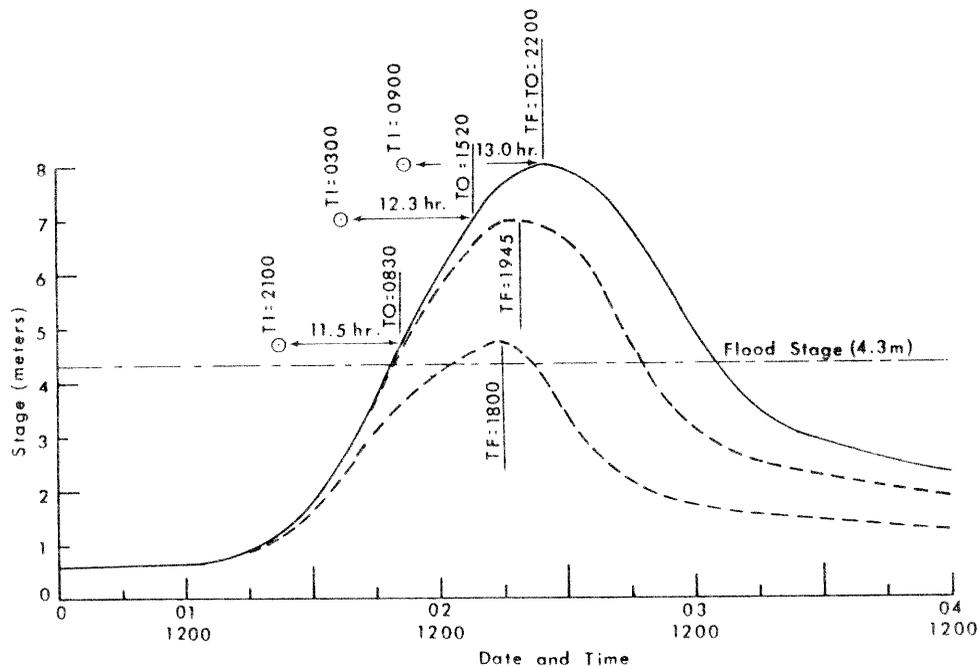


Figure 2

If it is desired to have the MFLT reflect timing errors, this can be accomplished by multiplying the lead time for each forecast by a "Timing Error Factor" (TEF). This factor is equal to unity if the timing error is zero and decreases linearly to zero as the error approaches the interval from forecast issuance to predicted time of occurrence. It is given by:

$$TEF = 1 - \frac{|TF-TO|}{TF-TI}$$

where TI is the time of issuance of the forecast;

TF is the time the stage is predicted to occur; and

TO is the time the stage does occur.

Applying the adjustment to the example of figure 2 gives the following:

Fcst. no.	TI	TF	TO	TF-TO	TF-TI	TEF	I	(TEF)(I)
1	2100(01)	1800	0830	9.5	21.0	0.55	11.5	6.3
2	0300	1945	1520	4.4	16.8	0.74	12.3	9.1
3	0900	2200	2200	0	13.0	1.00	13.0	13.0

The MFLT, adjusted for timing error, is  $(6.3+9.1+13.0)/3$ , or 9.5 hr.

Note that "TEF" can be a negative quantity. If it computes as such, it should be set to zero unless "I" is negative; in which case, "TEF" must always be set to unity.

The TEF adjustment is presented here as an option for possible use with MFLT. As stated above, the effect of timing errors on the user is not felt to be great and the need to reflect those errors in the evaluation score is therefore questionable. It should be borne in mind when considering the use of TEF that MFLT is intended to be a meaningful physical quantity rather than simply an index of the value of forecasts. The inclusion of this adjustment in the computation probably detracts somewhat from the conceptual qualities of MFLT.

#### THE USE OF FORECAST PRECIPITATION IN RIVER FORECASTS

In the examples given, forecast issuances were assumed to be made at some time after the measurement of the precipitation on which the forecast was based. It is often possible to prepare and issue river forecasts before the occurrence of all of the rainfall, basing them, at least in part, on rainfall forecasts. Although this practice has not been treated explicitly in the discussion, the use of predicted precipitation in the preparation of river forecasts has no effect on the method used to evaluate those river forecasts. That is, if perfect precipitation forecasts were always available for a period of 12 hr into the future, the river forecasts issued would be approximately the same as those based on observed precipitation, but they would be issued 12 hr earlier. The users of such forecasts would be receiving 12 hr additional warning time; and the MFLT, computed as described earlier, would be 12 hr greater.

In actual practice, river forecasts that involve predicted precipitation are likely to exhibit a lesser degree of accuracy than those based solely on observed rainfall. There are two reasons for this. First, the application

of precipitation input to a hydrologic model yields a computed hydrograph, not a forecast. The forecast is produced by a human forecaster, based principally on the computed hydrograph, but using in addition observed stages and/or discharges from various river stations, including the one for which the forecast is being prepared. Normally, the forecaster has available the observed hydrograph, at the forecast point, up to the time of forecast preparation, and he uses this to adjust, or "up-date," the computed hydrograph. If, through the use of forecast precipitation, he makes this forecast earlier, a smaller portion of the observed hydrograph is available to him, and the accuracy of the predicted portion of the hydrograph may be expected to suffer.

The second reason is that while observed precipitation values, and the areal means computed from them, usually involve sizeable errors, forecast precipitation generally attains an even lower level of accuracy. This is especially true when the precipitation forecast is made and used before the rainfall begins. Errors in volume, timing, and location of rainfall will all affect the river forecast.

In spite of an expected adverse effect on accuracy, the logic of making use of precipitation forecasts is rather firmly based. Limited studies have shown that, in the contiguous United States, the most frequent duration of runoff producing rainfall is about 12 hr and that, at the end of any 6-hr period within a storm, the probability of the occurrence of additional runoff producing rain is slightly greater than 0.5. That probability is, of course, highest early in the storm and decreases as the storm continues. At the end of the first 6 hr, the probability that the storm is not over is approximately 0.75. It does not drop below 0.5 until the duration has exceeded 24 hr. It may be said then that if river forecasts are being prepared early in a rainfall event the inclusion of additional forecast precipitation is more sound, statistically, than assuming that the storm has ended.

As has been pointed out earlier, forecast errors, whatever the cause, may either help or hurt the user. It seems reasonable to conclude, however, that in the long run the effect of an increased error level would be harmful. Thus, the use of forecast precipitation in the preparation of river forecasts may be expected to have two opposite effects on the user: increasing the lead time and decreasing forecast accuracy. MFLT is intended to reflect both effects and express, in a single evaluation score, the net gain or loss to the user.

#### MAXIMUM ATTAINABLE VALUES OF MEAN FORECAST LEAD TIME

If an evaluation system were to be based on forecast verification, it would express a score as some function of forecast errors. A perfect set of forecasts would then result in a score of zero. Even though the attaining of such a condition might well be impossible, it is obvious what the ultimate score, representing the "goal" of the forecast service would be, and the difference between that score and the score for an individual event is equally obvious.

When using MFLT, the situation is somewhat different, since a zero score indicates that the forecasts have no value and a high degree of value is indicated by a numerically large score. The ultimate then, for this type of score, is not obvious but must certainly exist. Since the MFLT pertaining to a specific event has much more of a physical meaning than an error statistic, the need to know the maximum possible value is not crucial. It is probably of value nevertheless to gain some insight into the factors that define the ultimate score and to explore the methods that might be used to evaluate it.

It is quite obvious that MFLT scores resulting from forecasts of floods on large, slowly responding rivers will be much greater than those resulting from forecasts of small, flashy streams. Consequently, with rare exceptions, MFLT scores can be compared only with scores for other events at the same forecast point. Thus, the ultimate, or maximum attainable value of MFLT is unique for a particular forecast point and is a function of the characteristics of the river system and of the storm causing the flood event. The relationships between maximum MFLT and some of these characteristics are explored in this section.

To make these investigations, a hydrologic model calibrated to an existing catchment was used. This catchment, located in the Eastern United States, has an area of 2,120 km<sup>2</sup> and has a concentration time (beginning of runoff to peak discharge) of approximately 30 hr for an event with a runoff duration of 6 hr. Extended low water flow normally is about 4 m<sup>3</sup>/s, which results in a stage of 0.5 m. Flood stage is 4.3 m and results from a discharge of 390 m<sup>3</sup>/s. The maximum stage of record is 10.9 m and involved a discharge of 2,310 m<sup>3</sup>/s.

A series of synthetic storms were applied to the model representing the physical behavior of this catchment. As in any such investigation, it was necessary to make certain assumptions about the synthetic storms. These assumptions were that the precipitation was distributed uniformly in regard to both area and time. That is, the channel response function used to model the hydrograph was one developed from historical storms that presumably involved, on the average, uniform areal distribution of rainfall. Also, each storm, regardless of precipitation volume or duration, had the same rainfall amount in each 6-hr period.

It is quite obvious that if areal distribution of precipitation is not uniform, the response time, and consequently the MFLT, will be affected. For this reason, no attempt has been made to model the effect on MFLT of upstream or downstream concentrations of runoff. What was examined was the effect of variations of the magnitude of the peak stage, antecedent moisture conditions, and storm duration. The effect of these factors on the maximum attainable MFLT is less obvious. In all cases, it was assumed that a forecast was issued at the end of each 6-hr rainfall period, that the time required for data collection and forecast preparation and dissemination was zero. It was further assumed that each forecast represented an error-free hydrologic analysis of the precipitation already fallen.

Figure 3 illustrates the effect on maximum attainable MFLT of the magnitude of the rise. It shows the hydrographs resulting from seven different rainfall

events, all with the same antecedent moisture conditions, all of 18 hr duration and ranging from 94 mm to 210 mm in volume. The peak stages are 4.3 m (flood stage), 5 m, 6 m, 7 m, 8 m, 9 m, and 10 m. Being 18-hr storms, three forecasts can be computed for each, at 6 hr, 12 hr, and 18 hr. The first forecast (6 hr), in all cases, predicts a stage less than 4.3 m and consequently is not involved in the computation of MFLT. The second forecast for each event, issued at 12 hr, is plotted in figure 3. This forecast, in each case, predicts a stage equal to that which would occur if the rain ceased at 12 hr. These stages vary from 2.7 m to 6.8 m. Note, however, that in all seven cases the stage in question occurs at approximately the same time (25.2 hr) on the rising limb of the hydrograph, resulting from the total storm. Thus, the lead time for the second forecast is approximately 13.2 hr in every case.

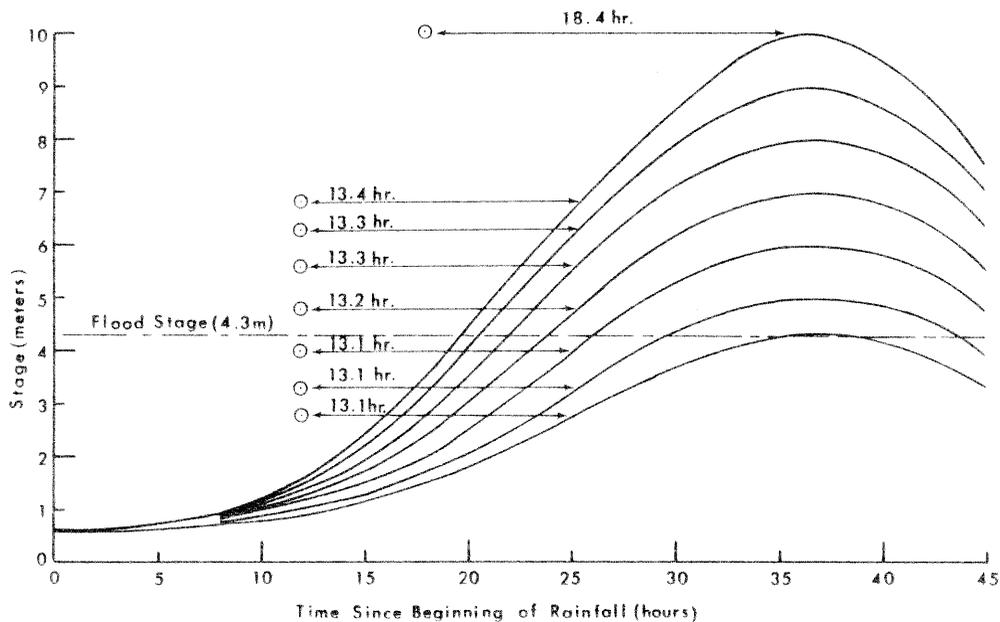


Figure 3

The third forecast in each case is issued at 18 hr and correctly predicts the crest that occurs at approximately 36.6 hr regardless of magnitude. In figure 3, the third forecast is plotted for only the 10-m event. The third forecast then always has a lead time of approximately 18.6 hr. MFLT for the seven events is computed as follows:

Maximum stage	12-hr forecast lead time	18-hr forecast lead time	MFLT
4.3	13.1	18.9	*18.9
5	13.1	18.8	*18.8
6	13.1	18.7	*18.7
7	13.2	18.6	15.9
8	13.3	18.5	15.9
9	13.3	18.4	15.9
10	13.4	18.4	15.9

Note that, for the first three events (marked with asterisks), only the 18-hr forecast is included in the computation, since the 12-hr forecast, as shown in figure 3, predicts a stage below flood.

What this example shows then is that as long as the event involves the same number of flood forecasts, the MFLT is virtually independent of magnitude. For the three smaller events, where the 12-hr forecast is not included, the MFLT is slightly larger and a discontinuity exists between the two groups. It may be noted also that the 6-hr forecast for each event has a lead time of approximately 11.2 hr and that, if a rise were large enough that this forecast were to predict a stage above flood, the MFLT would be approximately equal to the mean of 11.2, 13.4, and 18.4, or 14.3 hr, creating another discontinuity in the relationship between flood magnitude and maximum attainable MFLT. It should also be noted that in the case of an extremely large event, the flood stage might be reached less than 6 hr after the start of rainfall and before the first forecast could be issued. Under these circumstances, one zero lead time would be included in the computation of MFLT and another discontinuity would occur. These discontinuities result from the existence of a flood stage, a stage below which there is no damage and above which damage does occur. Flood stage is a discontinuity of nature (or of man's works) and it is therefore not inappropriate that discontinuities should exist in a computed quantity (MFLT) which recognizes the existence of a flood stage.

The discontinuities noted are small compared to the lead times involved. Furthermore, there are large ranges of flood magnitude within which the maximum attainable MFLT is virtually constant. Keeping in mind the effect of the original assumptions made in the analysis, it might be concluded that the maximum attainable MFLT for a forecast point is largely independent of the magnitude of a flood event.

A second set of simulations explores the effect of antecedent moisture conditions on MFLT. It is obvious of course that for a given rainfall event the magnitude of the resultant rise will be highly dependent upon antecedent moisture. What is being investigated here, however, is the effect on the forecast operation and on the forecast user of varied moisture conditions antecedent to river rises of the same magnitude. These events must then involve differing amounts of precipitation.

Figure 4 illustrates the effect. Two hydrographs are shown, both reaching a crest stage of 8.0 m. Each was simulated by applying to the catchment model an 18-hr period of rainfall, beginning at time zero and distributed equally among three 6-hr periods. The solid graph, labeled "wet," was based on the assumption of wet antecedent conditions and involved 85 mm of rainfall. The dotted graph, labeled "dry," was based on the assumption of dry antecedent conditions and required 225 mm of rainfall to simulate the same 8.0-m crest stage.

Note that the "dry" hydrograph occurs approximately 3.7 hr later than the "wet" hydrograph and has a slightly steeper rise. The reason for this is that the dry antecedent conditions result in high infiltration rates in the early part of the storm, causing the center of mass of runoff to occur later

than it does with wet antecedent conditions. The effect is to give the hydrograph the characteristics of one which would result from a rainfall event having a shorter duration and occurring somewhat later in time.

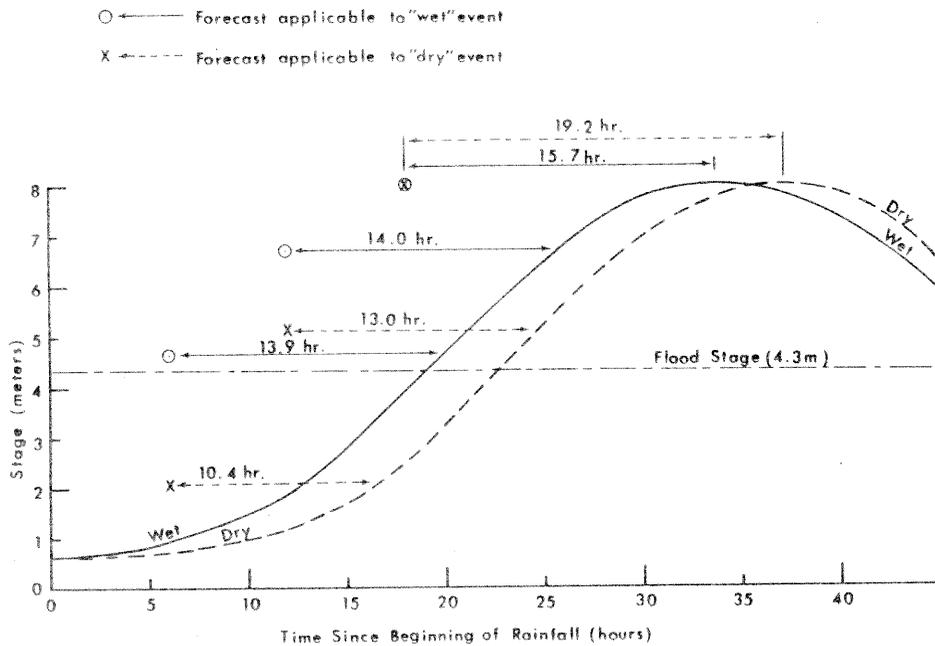


Figure 4

It might appear at first thought that since dry antecedent conditions tend to delay the rise that this would result in more lead time being available to the forecast user and in higher MFLT scores. Actually, it does not work out quite this way. Figure 4 also shows the set of three forecasts applicable to each event. Considering only the third forecast in each set, issued at 18 hr, the lead time for the "dry" event is in fact 3.5 hr greater than for the "wet" event, almost equal to the average time displacement between the two hydrographs. Other, earlier forecasts are also involved, however; and it is clear that the first two forecasts for the "wet" event, which predict stages of 4.6 m and 6.7 m, respectively, are a much better indication of the final hydrograph than are the two preliminary forecasts for the "dry" event, which call for stages of 2.0 m and 5.1 m. What has happened is that the antecedent conditions that cause the "wet" event hydrograph to occur earlier in time also cause the preliminary forecasts to be better indicators of subsequent events. The 6-hr and 12-hr forecasts for the "wet" event are based on one-third and two-thirds of the total precipitation and involve 28 percent and 63 percent respectively of the total computed runoff. In the case of the "dry" event, these two forecasts are still based on one-third and two-thirds of the rainfall, but involve only 7 percent and 39 percent of the runoff.

The MFLT for the "wet" event is computed from the lead times for all three forecasts and is equal to 14.5 hr. For the "dry" event, the first forecast is not included since it calls for a stage below flood level and the MFLT

is 16.1 hr. Thus, the difference between the two MFLT scores is 1.6 hr, less than half of the time displacement between the two hydrographs.

The conclusion is then that variations in initial moisture conditions affect the response of the river and the forecast operation in two ways, which, when related to the forecast user, tend to compensate. The maximum attainable MFLT is therefore not significantly affected by antecedent moisture conditions.

Figures 5, 6, and 7 illustrate the effect of storm duration on maximum attainable MFLT. They show three simulations, all attaining a crest stage of 8.0 m, but resulting from precipitation durations of 6, 18, and 36 hr. The same antecedent soil moisture conditions were assumed for all three events.

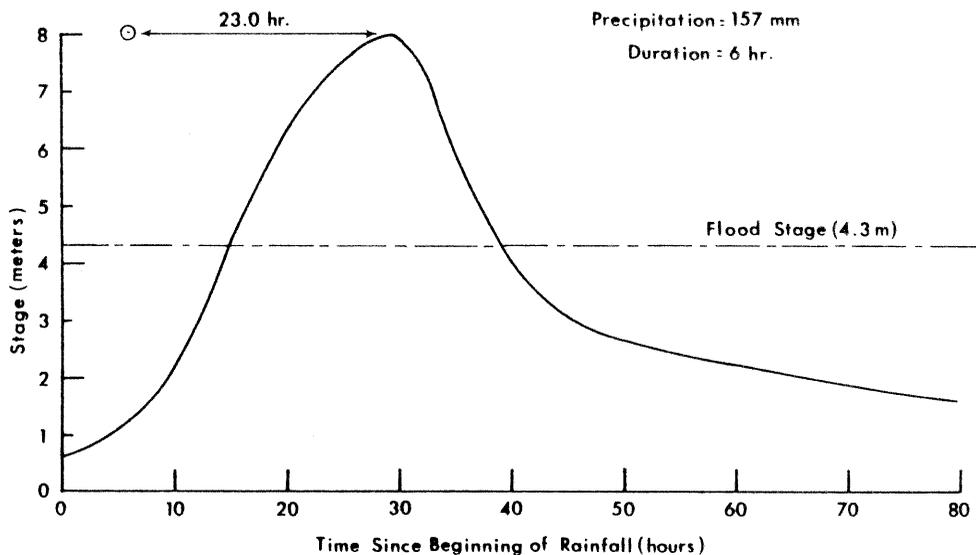


Figure 5

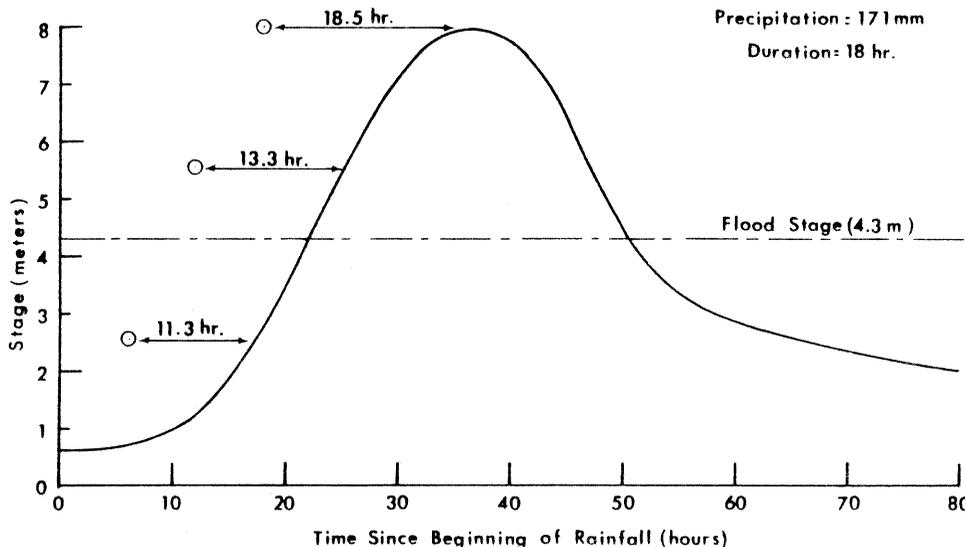


Figure 6

Note that in every case the set of forecasts consists of one final forecast, which predicts the crest and is issued at the end of the rainfall, and a number of preliminary forecasts, the number being equal to one less than the number of 6-hr rainfall periods. The lead time for the final forecast is, in each case, the interval from cessation of the rainfall to occurrence of the crest. In the 6-hr duration event, the crest occurs at 29.0 hr. The 18-hr storm produces a crest at 36.5 hr, and, when the duration is 36 hr, crest time is 50.2 hr. Thus, a 6-hr increment of duration causes the final forecast to be issued 6 hr later, but delays the crest by only 4.2 hr on the average, and, as duration increases, the lead time provided by the final forecast decreases.

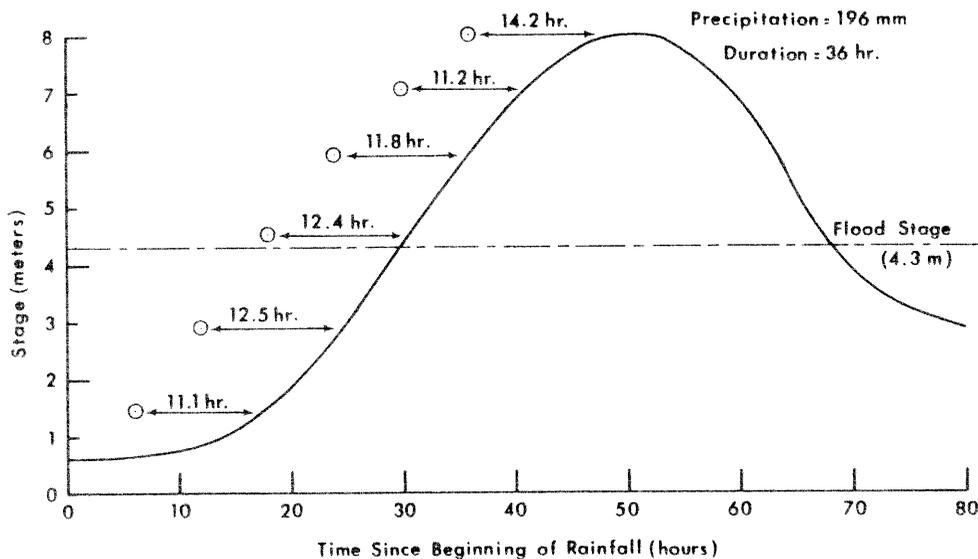


Figure 7

The MFLT is composed of the lead time of the final forecast and that of all preliminary forecasts that predict stages above flood level. Figure 7 shows that the lead times for the five preliminary forecasts do not vary greatly from one another but are somewhat less than the final forecast lead time, 2.4 hr less in this case. This difference is typical and results from the characteristic, concave downward shape of the portion of the hydrograph just prior to the crest. The lead time for the final forecast has been shown to decrease with increasing duration. The lead times for the preliminary forecasts are smaller than for the final forecast, and, the longer the duration, the more of them are included in the average. Thus, as storm duration increases, the maximum attainable value of MFLT decreases. The MFLT values for the three events shown are 23.0 hr, 15.9 hr, and 12.4 hr.

The conclusion then is that for specified areal and temporal distributions of rainfall the maximum attainable MFLT for a particular forecast point is largely independent of all storm characteristics, except duration. The relationship may be derived and, for the catchment used in this study, is shown in figure 8. The plotted points represent six simulations, with durations ranging from 6 to 36 hr, and including the three that are shown

in figures 5, 6, and 7. Beyond 36 hr, the curve is assumed to approach, as a lower limit, a value equal to the average lead time for preliminary forecasts.

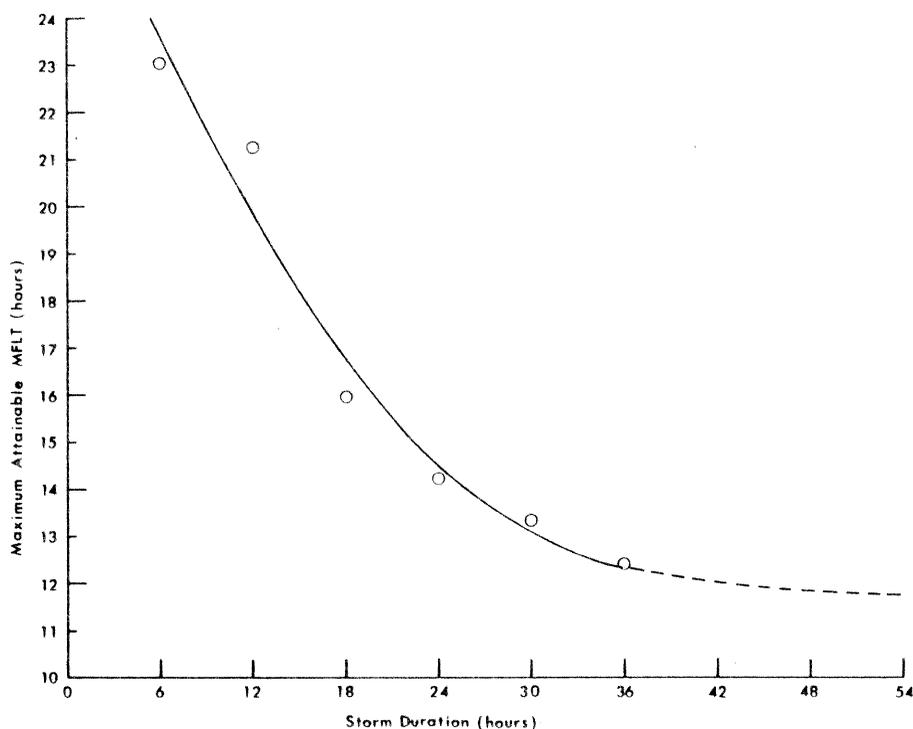


Figure 8

These studies have been based on a catchment analysis. If a similar analysis were to be made for a downstream point in a large river system, it seems logical to conclude that the results would be the same if the same initial assumptions were made. It must be noted, however, that the assumption of uniform areal distribution of rainfall becomes less realistic for larger drainage areas.

#### APPLICATIONS AND USE OF MEAN FORECAST LEAD TIME

In the "Introduction," it was pointed out that there were numerous purposes that a verification or an evaluation program might serve and that the technique used for computing the score should be related to the use which is to be made of it.

Verification of operationally produced forecasts is often suggested as a means of ascertaining the simulation accuracy of a hydrologic model relative to that of other models. The author feels that neither MFLT nor any operational verification system is suitable for this purpose. Such an effort would not only involve the preparation of two or more forecasts (one for each model) in real time, but would also suffer from having the results partially obscured by the high noise level commonly present in real time data. In addition, the collection of sufficient verification data on which to base

a meaningful inter-comparison would take many years. The determination of the relative accuracy of hydrologic models is best accomplished by laboratory testing with the use of historical data. The data must be selected to represent an appropriate scope of hydrological and climatological conditions and experience; the calibration and simulation must be performed in such a way as to minimize the effect of all factors other than model accuracy; and the results must be analyzed rather thoroughly to relate type and magnitude of errors to model characteristics. Such investigations were described by Sittner (1969, 1973, and 1976) and by WMO (1975). The futility of attempting to acquire such information in real time is evident in these references.

MFLT is intended instead to be a means of monitoring the forecast service as a whole. A forecast operation consists of many things. It involves not only hydrologic and hydraulic models, rain gages, and computers but also administrative policies in regard to staffing matters, hours of operation of field offices, data collection by manual or automatic means, delegation of authority, and numerous other details of the operation. Since, at any point in time, a number of these many aspects are likely to be undergoing change, it is necessary for management to constantly check the effect on the final product of all of the parts of the forecast service in combination. MFLT is intended to be a tool that serves this purpose.

In making a management evaluation of a forecasting service, the ultimate question usually becomes "Does the service produce monetary benefits which exceed the cost of operating the service?" or "If a change in the service is made, does that change increase or decrease the benefit and by how much?" Since potential flood damages are so great compared to the cost of operating a forecasting service, a favorable benefit-cost ratio can usually be assumed. Evaluating the effect of a change in the service in comparison to the cost of making the change however is not nearly so simple. A value-measuring system, to be ideal for these purposes, would have to express an evaluation score in dollars or other monetary units.

A measure of the monetary savings that accrue from a forecast service does not reflect the value of lives that may have been saved by flood warnings, unless one is able to place a monetary value on a human life. This is difficult and most would agree that the saving of lives and the reduction of property damage are noncommensurate objectives. In the present context, however, this does not present a great problem. The situation in which a person loses his life due to suddenly rising water and in which his life could have been saved by a flood warning is typically a "flash flood" situation. While warnings are provided for such situations and while lives are saved as a result, the severe time constraints involved usually dictate a type of forecast operation somewhat different from that which has been discussed. That is, a flash flood forecast, or warning, is likely to be a single qualitative statement indicating a yes-or-no situation rather than a series of quantitative stage forecasts defining a hydrograph. MFLT is a procedure for evaluating the type of forecast event that occurs on larger, more slowly responding streams rather than a flash flood event.

It is true of course that people lose their lives in floods of the type being dealt with herein. Experience shows however that almost invariably the victim is aware of the flood condition and meets his fate as the result of an injudicious action on his part. Obviously, then, better flood warnings do not prevent this type of tragedy, and, unless public education is to be considered a part of the forecast service, it is feasible to evaluate the service solely on the basis of the value of the reduction in property damage.

Numerous investigators have studied the economic effect of flood warnings. Day (1969) prepared a simulation model of the response to, and economic effect of, floods and flood warnings in residential areas. Assuming "reliable warnings," he evaluated flood losses under conditions of no warning and of a number of different warning intervals. He thus recognized and implicitly presented the relationship between lead time and economic benefit. Day assumed "reliable warnings" and thus did not address himself to the effect of excessive error in the forecasts. MFLT is, as was stated earlier, the average warning time provided by a group of forecasts, suitably adjusted for forecast inaccuracy. The methods and rationale for making the inaccuracy adjustments have been discussed in some detail to clarify the definition of MFLT given earlier. That is:

MFLT is the average warning time that would be provided by a group of error-free forecasts that would have affected the users in the same manner as did the group of forecasts actually issued.

By this definition, there clearly exists a relationship between the monetary value of forecasts and MFLT, whether the forecasts be of the idealized "suitable accuracy" type or of the type often encountered in actual operations.

Sniedovich et al. (1975) studied the human response aspects of flood forecast evaluation and attempted to relate behavior patterns to economic benefit using a systems approach. They too expressed the concept of lead time being a major input to an economic evaluation model and, in fact, had access to an earlier, unpublished description of MFLT prepared by the author.

Sniedovich's work, and that of others, indicates that the present state of the art lacks understanding of a number of processes essential to a complete modelling of the forecast-response-economic system. If these processes are ever successfully analyzed, the model that computes economic benefit as an output may well utilize MFLT as an input. In the meantime, MFLT is presented as a realistic and meaningful evaluation procedure.

## REFERENCES

- Day, H. J., 1969: Benefit evaluation of a flood warning system to urban residences in the Susquehanna River Basin. U.S. Dept. of Commerce, Environmental Science Services Administration, Purchase Order No. 8-11917.
- Sittner, W. T., Schauss, C. E., and Monro, J. C., 1969: Continuous hydrograph synthesis with an A.P.I. type hydrologic model. Water Resources Research, Vol. 5, No. 5.
- Sittner, W. T., 1973: Modernization of National Weather Service river forecasting techniques. Water Resources Bulletin, Vol. 9, No. 4, American Water Resources Association.
- Sittner, W. T., 1976: WMO project on intercomparison of conceptual models used in hydrological forecasting. Hydrological Sciences Bulletin, Vol. XXI, No. 1, report by WMO to the International Symposium on Mathematical Models.
- Sniedovich, M., Fischer, M., Uhl, V., and Davis, D. R., 1975: The evaluation of flood forecasting-response systems: A decision theoretic approach. National Oceanic and Atmospheric Administration. Contract No. 3-35108.
- World Meteorological Organization, 1975: Intercomparison of conceptual models used in operational hydrological forecasting. Operational Hydrology Report No. 7.

(Continued from inside front cover)

- NWS HYDRO 15 Time Distribution of Precipitation in 4- to 10-Day Storms--Arkansas-Canadian River Basins. Ralph H. Frederick, June 1973. (COM-73-11169)
- NWS HYDRO 16 A Dynamic Model of Stage-Discharge Relations Affected by Changing Discharge. D. L. Fread, December 1973. Revised, September 1976.
- NWS HYDRO 17 National Weather Service River Forecast System--Snow Accumulation and Ablation Model. Eric A. Anderson, November 1973. (COM-74-10728)
- NWS HYDRO 18 Numerical Properties of Implicit Four-Point Finite Difference Equations of Unsteady Flow. D. L. Fread, March 1974.
- NWS HYDRO 19 Storm Tide Frequency Analysis for the Coast of Georgia. Francis P. Ho, September 1974. (COM-74-11746/AS)
- NWS HYDRO 20 Storm Tide Frequency for the Gulf Coast of Florida From Cape San Blas to St. Petersburg Beach. Francis P. Ho and Robert J. Tracey, April 1975. (COM-75-10901/AS)
- NWS HYDRO 21 Storm Tide Frequency Analysis for the Coast of North Carolina, South of Cape Lookout. Francis P. Ho and Robert J. Tracey, May 1975. (COM-75-11000/AS)
- NWS HYDRO 22 Annotated Bibliography of NOAA Publications of Hydrometeorological Interest. John F. Miller, May 1975.
- NWS HYDRO 23 Storm Tide Frequency Analysis for the Coast of Puerto Rico. Francis P. Ho, May 1975. (COM-11001/AS)
- NWS HYDRO 24 The Flood of April 1974 in Southern Mississippi and Southeastern Louisiana. Edwin H. Chin, August 1975.
- NWS HYDRO 25 The Use of a Multizone Hydrologic Model With Distributed Rainfall and Distributed Parameters in the National Weather Service River Forecast System. David J. Morris, August 1975.
- NWS HYDRO 26 Moisture Source for Three Extreme Local Rainfalls in the Southern Intermountain Region. E. Marshall Hansen, October 1975.
- NWS HYDRO 27 Storm Tide Frequency Analysis for the Coast of North Carolina, North of Cape Lookout. Francis P. Ho and Robert J. Tracey. November 1975.
- NWS HYDRO 28 Flood Damage Reduction Potential of River Forecast Services in the Connecticut River Basin. Harold J. Day and Kwang K. Lee, February 1976. (PB-256758)
- NWS HYDRO 29 Water Available for Runoff for 4- to 15-Days Duration in the Snake River Basin in Idaho. Ralph H. Frederick and Robert J. Tracey, June 1976. (PB-258-427)
- NWS HYDRO 30 Meteor Burst Communication System--Alaska Winter Field Test Program. Henry S. Sante-ford, March 1976. (PB-260-449)
- NWS HYDRO 31 Catchment Modeling and Initial Parameter Estimation for the National Weather Service River Forecast System. Eugene L. Peck, June 1976. (PB-264154)
- NWS HYDRO 32 Storm Tide Frequency Analysis for the Open Coast of Virginia, Maryland, and Delaware. Francis P. Ho, Robert J. Tracey, Vance A. Myers, and Narmalee S. Foat, August 1976. (PB-261-969)
- NWS HYDRO 33 Greatest Known Areal Storm Rainfall Depths for the Contiguous United States. Albert P. Shipe and John T. Riedel, December 1976.
- NWS HYDRO 34 Annotated Bibliography of NOAA Publications of Hydrometeorological Interest. John F. Miller, December 1976.
- NWS HYDRO 35 Five- to 60-Minute Precipitation Frequency for the Eastern and Central United States. Ralph H. Frederick, Vance A. Myers, and Eugene P. Auciello, June 1977.

## NOAA SCIENTIFIC AND TECHNICAL PUBLICATIONS

NOAA, the *National Oceanic and Atmospheric Administration*, was established as part of the Department of Commerce on October 3, 1970. The mission responsibilities of NOAA are to monitor and predict the state of the solid Earth, the oceans and their living resources, the atmosphere, and the space environment of the Earth, and to assess the socioeconomic impact of natural and technological changes in the environment.

The six Major Line Components of NOAA regularly produce various types of scientific and technical information in the following kinds of publications:

**PROFESSIONAL PAPERS** — Important definitive research results, major techniques, and special investigations.

**TECHNICAL REPORTS**—Journal quality with extensive details, mathematical developments, or data listings.

**TECHNICAL MEMORANDUMS** — Reports of preliminary, partial, or negative research or technology results, interim instructions, and the like.

**CONTRACT AND GRANT REPORTS**—Reports prepared by contractors or grantees under NOAA sponsorship.

**TECHNICAL SERVICE PUBLICATIONS**—These are publications containing data, observations, instructions, etc. A partial listing: data serials; prediction and outlook periodicals; technical manuals, training papers, planning reports, and information serials; and miscellaneous technical publications.

**ATLAS**—Analysed data generally presented in the form of maps showing distribution of rainfall, chemical and physical conditions of oceans and atmosphere, distribution of fishes and marine mammals, ionospheric conditions, etc.



*Information on availability of NOAA publications can be obtained from:*

**ENVIRONMENTAL SCIENCE INFORMATION CENTER  
ENVIRONMENTAL DATA SERVICE  
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION  
U.S. DEPARTMENT OF COMMERCE**

NOAA Technical Memorandum NWS HYDRO-36



---

DETERMINATION OF FLOOD FORECAST EFFECTIVENESS  
BY THE USE OF MEAN FORECAST LEAD TIME

Silver Spring, Md.  
August 1977

---

**noaa**

NATIONAL OCEANIC AND  
ATMOSPHERIC ADMINISTRATION

National Weather  
Service

NOAA TECHNICAL MEMORANDA

National Weather Service, Office of Hydrology Series

The Office of Hydrology (HYDRO) of the National Weather Service (NWS) develops procedures for making river and water supply forecasts, analyzes hydrometeorological data for planning and design criteria for other agencies, and conducts pertinent research and development.

NOAA Technical Memoranda in the NWS HYDRO series facilitate prompt distribution of scientific and technical material by staff members, cooperators, and contractors. Information presented in this series may be preliminary in nature and may be published formally elsewhere at a later date. Publication 1 is in the former series, Weather Bureau Technical Notes (TN); publications 2 to 11 are in the former series, ESSA Technical Memoranda, Weather Bureau Technical Memoranda (WBTM). Beginning with 12, publications are now part of the series, NOAA Technical Memoranda, NWS.

Publications listed below are available from the National Technical Information Service, U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151. Price: \$3.00 paper copy; \$1.45 microfiche. Order by accession number shown in parentheses at end of each entry.

Weather Bureau Technical Notes

TN 44 HYDRO 1 Infrared Radiation from Air to Underlying Surface. Vance A. Myers, May 1966. (PB-170-664)

ESSA Technical Memoranda

WBTM HYDRO 2 Annotated Bibliography of ESSA Publications of Hydrological Interest. J. L. H. Paulhus, February 1967. (Superseded by WBTM HYDRO 8)

WBTM HYDRO 3 The Role of Persistence, Instability, and Moisture in the Intense Rainstorms in Eastern Colorado, June 14-17, 1965. F. K. Schwarz, February 1967. (PB-174-609)

WBTM HYDRO 4 Elements of River Forecasting. Marshall M. Richards and Joseph A. Strahl, October 1967. (Superseded by WBTM HYDRO 9)

WBTM HYDRO 5 Meteorological Estimation of Extreme Precipitation for Spillway Design Floods. Vance A. Myers, October 1967. (PB-177-687)

WBTM HYDRO 6 Annotated Bibliography of ESSA Publications of Hydrometeorological Interest. J. L. H. Paulhus, November 1967. (Superseded by WBTM HYDRO 8)

WBTM HYDRO 7 Meteorology of Major Storms in Western Colorado and Eastern Utah. Robert L. Weaver, January 1968. (PB-177-491)

WBTM HYDRO 8 Annotated Bibliography of ESSA Publications of Hydrometeorological Interest. J. L. H. Paulhus, August 1968. (PB-179-855)

WBTM HYDRO 9 Elements of River Forecasting (Revised). Marshall M. Richards and Joseph A. Strahl, March 1969. (PB-185-969)

WBTM HYDRO 10 Flood Warning Benefit Evaluation - Susquehanna River Basin (Urban Residences). Harold J. Day, March 1970. (PB-190-984)

WBTM HYDRO 11 Joint Probability Method of Tide Frequency Analysis Applied to Atlantic City and Long Beach Island, N.J. Vance A. Myers, April 1970. (PB-192-745)

NOAA Technical Memoranda

NWS HYDRO 12 Direct Search Optimization in Mathematical Modeling and a Watershed Model Application. John C. Monro, April 1971. (COM-71-00616)

NWS HYDRO 13 Time Distribution of Precipitation in 4- to 10-Day Storms--Ohio River Basin. John F. Miller and Ralph H. Frederick, May 1972. (COM-72-11139)

NWS HYDRO 14 National Weather Service River Forecast System Forecast Procedures. December, 1972. (COM-73-10517)

(Continued on inside back cover)

NOAA Technical Memorandum NWS HYDRO-36

DETERMINATION OF FLOOD FORECAST EFFECTIVENESS  
BY THE USE OF MEAN FORECAST LEAD TIME

Walter T. Sittner

Silver Spring, Md.  
August 1977

UNITED STATES  
DEPARTMENT OF COMMERCE  
Juanita M. Kreps, Secretary

NATIONAL OCEANIC AND  
ATMOSPHERIC ADMINISTRATION  
Robert M. White, Administrator

National Weather  
Service  
George P. Cressman, Director





CONTENTS

Abstract . . . . .	1
Introduction . . . . .	1
The concept of mean forecast lead time . . . . .	2
Method of computation . . . . .	3
The effect of forecast accuracy . . . . .	5
Forecast revisions and refinements . . . . .	7
Summary of computational rules . . . . .	8
The effect of timing errors . . . . .	9
The use of forecast precipitation in river forecasts . . . . .	11
Maximum attainable values of mean forecast lead time . . . . .	12
Applications and use of mean forecast lead time . . . . .	19
References . . . . .	22



DETERMINATION OF FLOOD FORECAST EFFECTIVENESS BY THE USE OF  
MEAN FORECAST LEAD TIME

Walter T. Sittner  
Hydrologic Research Laboratory  
Office of Hydrology  
National Weather Service, NOAA

ABSTRACT. A method of evaluating flood forecasts is presented. The method expresses the value, to the user, of a series of forecasts that relate to a single flood event. The timeliness of the forecasts and their accuracy are combined into a single numerical score, expressed in units of time, and termed "mean forecast lead time." The score reflects the manner in which a particular combination of timeliness and accuracy affects the user. The system purports to produce a measure of forecast effectiveness that is physically meaningful and that is more closely related to economic benefit than are error statistics based on the difference between forecast and observed hydrographs.

The relationships between the evaluation system and various forecasting problems and practices are discussed.

INTRODUCTION

It is generally agreed that a necessary adjunct to any forecasting operation is some type of evaluation program. The objective of such a program may be the monitoring of the forecast operation itself, the data collection effort, or the dissemination system. The objective may be the evaluation of new forecasting techniques and/or equipment, or it may be the determination of the value to the user of the final product of the entire system. Obviously, the technique used must be devised and applied in such a way that it will reflect those factors that one desires to measure.

Most existing evaluation programs are actually verification systems. That is, they are designed to measure the degree of agreement between the predicted value of a variable and a value that actually occurs at some later time. This is usually accomplished by producing some sort of statistical summary of the differences between predicted and observed values of the variable. The aim of a forecasting system is to make the most accurate forecasts possible consistent with technological and fiscal constraints. Success in this would be indicated by a verification score that is numerically small. The evaluation of a forecast operation in this way has some validity if the objective is to determine the relative accuracy of different forecast models or techniques. If, on the other hand, the aim is to evaluate the effectiveness of the forecast as it relates to the user, such an approach

has serious deficiencies. It is, for example, difficult to compute an error function that is meaningful because a forecast event does not normally consist of one prediction and one event. There are, in fact, usually several predictions made at various times prior to the event, and it is axiomatic that the earlier ones will show a lesser degree of accuracy than those made only a short time before the event occurs. The problem of combining the various errors with their time intervals in such a way as to reach a meaningful conclusion is one that is not easily solved.

In the case of flood forecasting, if forecasts are based only on observed precipitation, rather than predicted precipitation, the crest cannot be predicted until the rain has ended. Unless a storm is very brief, however, a number of preliminary forecasts are usually made during the storm on the basis of precipitation observed up to the time of forecast preparation. The stages called for by such forecasts will then occur as points on the rising limb of the hydrograph rather than as the crest. These forecasts serve a useful purpose by advising the user that the water will be "at least this high," but they are not verifiable by an observed hydrograph. Finally, what is the significance of a specified stage error? If the series of forecasts related to a flood event has an average error, determined in some meaningful way of 0.2 m, is that good or is it bad? Without a great deal of supplementary information, one would not really know.

The evaluation system to be presented is based on an entirely different concept. It results in a score that is, physically, a warning time interval rather than an error function. A superior forecast operation is indicated by a score that is numerically large rather than by one approaching zero.

This measure, called "Mean Forecast Lead Time (MFLT)," is, as the title implies, intended only for use with flood forecasts and not for low water or other types of river predictions. It is intended to indicate the value, available to the users, of a flood forecast or the group of forecasts relating to a flood event. It should be noted at this point that this is the value available to the user. Such value will not actually accrue to him unless he reacts to the forecasts in a suitable manner. MFLT does not reflect the behavioral pattern of the forecast recipient. The potential value of the forecasts is expressed by MFLT without regard to the reason for the forecasts being good or being poor. The effect of sparse or erroneous precipitation reports is indistinguishable from the effect of a poor hydrologic model. The work of a lucky forecaster results in as good an evaluation score as that of a skillful forecaster, but perhaps not as consistently.

#### THE CONCEPT OF MEAN FORECAST LEAD TIME

In concept, MFLT is very simple. A forecast is a statement regarding an event, made prior to the occurrence of the event. It derives its value from the fact that it is made prior to the event. Consequently, that value can be measured by the length of the interval from forecast to occurrence, or "lead time." Obviously, the value of a forecast is diminished if there is not reasonable agreement between it and the future events that actually occur. Therefore, if the concept of using lead time as a measure of forecast

effectiveness is to be valid, there must be provision for an adjustment to the measure when a forecast exhibits a low level of accuracy. MFLT is, in essence, the average warning time provided by a group of forecasts, suitably adjusted for forecast inaccuracy. The basis for the adjustment is the effect of the inaccuracy on the user. What the system attempts to do is to determine the lead time that an accurate forecast would have to have had in order to affect the recipient in the same manner as did the inaccurate forecast that was actually issued and then to use this "equivalent" lead time in the averaging process. MFLT can then be defined as follows:

MFLT is the average warning time that would be provided by a group of error-free forecasts that would have affected the users in the same manner as did the group of forecasts actually issued.

The manner in which lead times are computed and the treatment of forecast errors are discussed in the following sections.

#### METHOD OF COMPUTATION

Lead time is, as previously stated, the interval from forecast issuance to the occurrence of the event. If a flood event involves a series of forecasts, each calling for a successively higher stage, the lead time for each forecast is the interval from the issuance of that forecast to the occurrence of the stage called for. The MFLT is computed by averaging all such intervals.

Figure 1, a relatively simple case, illustrates the basic method. The rise, from a base stage of 0.6 m to a crest of 8.0 m, is caused by 175 mm of rainfall in a 24-hr period as shown. The four 6-hr increments, ending at 1300 and 1900 on day 1 and at 0100 and 0700 on day 2, are 20 mm, 75 mm, 50 mm, and 30 mm. It is assumed in this example that a forecast can be issued 2 hr after the precipitation observations are made. It is further assumed that each forecast is a perfect hydrologic analysis. That is, if the rain were to cease at observation time, the forecast hydrograph would agree exactly with the observed hydrograph. In this example, then, three forecasts were issued, at 2100 on day 1 and at 0300 and 0900 on day 2. (The 20-mm rainfall observed at 1300 on day 1 was not sufficient to warrant the issuance of a flood forecast.) The first forecast was based on the 12-hr rainfall totalling 95 mm, and called for a maximum stage of 4.7 m, which is 0.4 m above flood stage. The second forecast, based on 145 mm of rainfall, called for 7.0 m and the last forecast, based on all 175 mm, correctly predicted the observed crest of 8.0 m. In figure 1, each forecast is plotted with the time of issuance as abscissa and the predicted stage as ordinate. The lead times shown, the intervals from issuance to the occurrence of the stage, are 11.5 hr, 12.3 hr, and 13.0 hr. The MFLT is then the average of the three individual lead times, or 12.3 hr.

There are two ways of thinking of this quantity. Flood stage, 4.3 m in this case, is defined as the lowest stage at which damage occurs. Consequently, the first forecast, 4.7 m, provides a firm warning to those people who become vulnerable between 4.3 and 4.7 m. It also provides an indication of danger to those located slightly above 4.7 m, but with a lesser degree of certainty.

If the appropriate precautionary measures for those in the affected group are of the threshold type, complete evacuation of the property, shutting down a facility, etc., then, this forecast, the first of the three, is the one that provides them with their warning, and it gives them a lead time of 11.5 hr in which to make their preparations. The second forecast, 7.0 m, serves an additional group, those who become affected between approximately 4.7 and 7.0 m, but does little for those who are situated at, say, 7.5 m. The third forecast is the one that provides a warning to them. Thus, MFLT can be thought of as the average warning time given to the community as a whole.

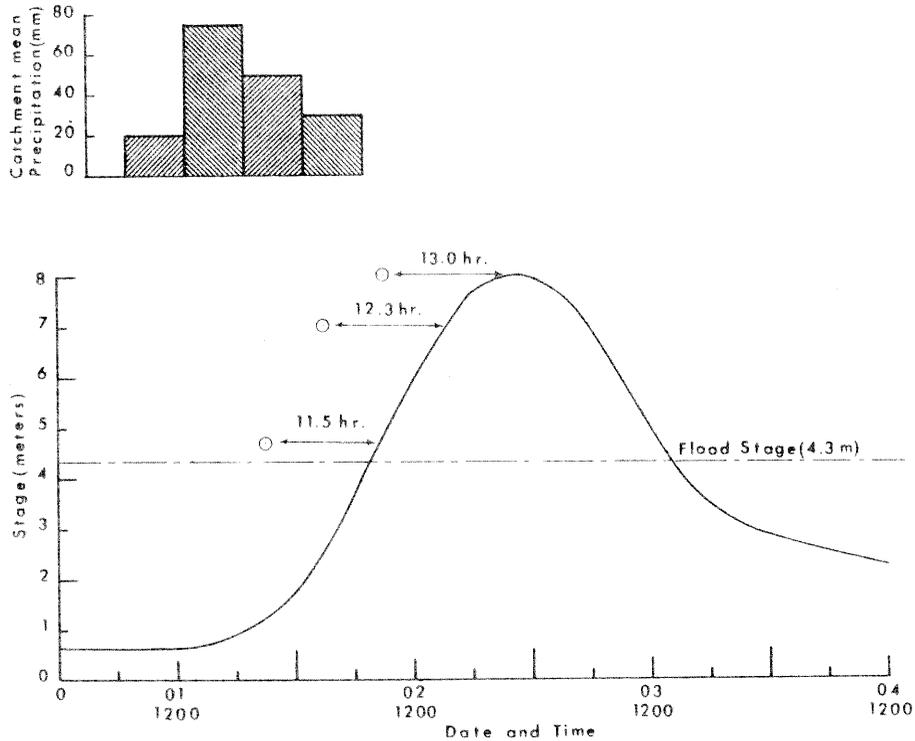


Figure 1

In another light, consider the individual user who becomes vulnerable to flood waters at a certain stage but the extent of whose precautions vary with stage above that initial point. An example might be a store with floor elevation of 5 m and shelves spaced 0.2 m apart. Upon receipt of the 7-m forecast, the owner will remove the contents of the first 10 or 12 shelves. When the 8-m forecast is received 6 hr later, he will empty the top shelves. While the 8-m forecast is the one that defined the ultimate action to be taken, the earlier forecast was also of value since it enabled him to empty the first 10 shelves and stay ahead of the rise. Thus, for this type of user, MFLT is a measure of the average value, to one individual, of all of the forecasts that call for stages within his range of interest.

It might be noted at this point that MFLT, computed as described, is not the true average warning time for all flood plain occupants. That is, in the example of figure 1, an occupant situated at 4.3 m (flood stage) has not had 11.5 hr warning, but only 10.5 hr, since the first forecast calling for

a stage of 4.7 m was issued at 2100 on day 1 and flood stage was reached at 0730 on day 2. Then, the average warning time for all occupants located between 4.0 m and 4.3 m was not 11.5 hr but  $(10.5+11.5)/2$ , or 11.0 hr. Similarly, individuals situated slightly above 4.7 m received their first firm warning from the second forecast, issued only 5.5 hr prior to the inundation at this level. Then, the entire group between 4.7 m and 7.0 m had an average of  $(5.5+12.3)/2$ , or 8.9 hr, warning time. If it were to be assumed that the vertical distribution of forecast users was uniform, the true average warning time for all individuals involved would be 9.5 hr rather than 12.3 hr as previously determined. This discrepancy, although systematic in nature, is not considered serious enough to warrant the added complexity of integrating lead times along the rising limb of the hydrograph. Doing so would probably not produce a much better determination of MFLT since it is not likely that the vertical distribution of economic interests would actually be uniform through the range of flooding. Furthermore, as was pointed out earlier, a forecast stating that river levels will reach a certain stage does serve to at least alert individuals located at slightly higher stages to the possibility that they too will be flooded. In addition, it will be shown later that there are computational procedures for MFLT that, in some circumstances, tend to minimize the effect of not performing such an integration.

#### THE EFFECT OF FORECAST ACCURACY

It was pointed out earlier that the use of lead time as a measure of forecast effectiveness must either presuppose a suitable degree of accuracy or must provide for an adjustment to the measure when such accuracy is not achieved. Since the former supposition would not be realistic, the latter provision must be made.

In the example of figure 1, each forecast is assumed to be the result of a perfect analysis. Considering the second forecast, this means that had the rain ceased at 0100 on day 2 the crest would have been exactly 7.0 m. Now, consider the effect of an error in that analysis, an error, for example, of +0.3 m. The forecast then would have called for 7.3 m rather than 7.0 m and, in light of subsequent events, would actually have been of greater value to the user. The computed lead time for this forecast would have been from issuance at 0300 to the occurrence of 7.3 m at 1630, or 13.5 hr, rather than the 12.3 hr shown in figure 1. The MFLT for the event would then have been 12.7 hr rather than 12.3 hr. If the error in the analysis had been minus 0.3 m, the forecast would have been 6.7 m, the lead time 11.4 hr, and the MFLT 12.0 hr. The negative error resulted in a less accurate prediction of future events and therefore would have hurt the user rather than helped him. This is reflected in the lower score. As was pointed out earlier, hydrologic analyses based on only a portion of the storm rainfall are not verifiable by the observed hydrograph, and their accuracy can, therefore, not be determined. With the MFLT evaluation system, however, such an accuracy determination is not needed. An inaccuracy in the hydrologic analysis may hurt the user or it may help him. Whatever the effect, it will be automatically reflected in the MFLT score.

The foregoing discussion of the effect of inaccuracy in forecasts made during the storm does not apply to a forecast made after the storm. This last forecast is verifiable; its accuracy may be easily determined; and, if it is in error, the user will be affected adversely whether that error be positive or negative.

Consider the effect of a negative error in the third forecast of figure 1. That is, let the predicted stage be 7.5 m rather than 8.0 m. The lead time for this forecast would then be from 0900 to 1720, or 8.3 hr. Averaging the three lead times for this event would give  $(11.5+12.3+8.3)/3$ , or 10.7 hr. This, however, is not the MFLT because the error in the third forecast has brought a new factor into play; the observed 8.0-m crest has not been predicted. To reflect this, a fourth forecast, which correctly predicted the crest, is assumed to have been issued at the time the crest occurred. That is, the occurrence of an unpredicted stage has the same effect on the user as a correct forecast with zero lead time. The MFLT for this event is then  $(11.5+12.3+8.3+0)/4$ , or 8.0 hr. Thus, in a case where the observed crest is higher than the highest forecast issued, a "low miss" has occurred and its effect on the user is recognized by including one zero lead time in the computation of MFLT.

Now, consider the case in which the last forecast misses the crest but involves an error which is positive rather than negative. For instance, assume that the third forecast of figure 1 had called for 8.5 m, an error of +0.5 m. The definition of lead time for an individual forecast has been given earlier as the interval from issuance to the occurrence of the predicted stage. In this case, the predicted stage, 8.5 m, does not occur, and this definition becomes meaningless. In such circumstances, the lead time is computed in such a way that the positive error has the same effect on the score as a negative error of like magnitude. That is, the lead time is taken as the interval from forecast issuance to the occurrence of a stage that is as far below the crest as the predicted stage is above the crest. Such a stage is 7.5 m, and the lead time is 8.3 hr, the same as for the forecast with an error of -0.5 m.

In the case of the last forecast being a "low miss," the computation of MFLT included one zero lead time because users situated at the 8.0-m level received no warning. In the case of the "high miss," however, they did receive warning. The previous reasoning then would lead to the conclusion that a zero lead time should not be included in the case of the "high miss" and that the MFLT should be determined as  $(11.5+12.3+8.3)/3$ , or 10.7 hr. Note that if this is done the positive error in the last forecast results in a better score than a negative error of the same magnitude. Such an effect may be desirable if one is of the opinion that over-forecasts are preferable to under-forecasts. Whether or not they are is a rather moot point. The author is of the opinion that they are not and that a positive error is just as bad as and no worse than a negative error of the same size. A related consideration is that an evaluation system should be noncorrupting. That is, it must be designed in such a way that it does not encourage the forecaster to develop habits which enhance the scores without consistently improving the service to the user. An evaluation system that treats a positive error more favorably than a negative one will probably cause a

forecaster to shade his predictions upward rather than giving the user his best estimate of what is going to happen. The long-term effect of such a practice is likely to be unsatisfactory. For this reason, it is recommended that when a high miss occurs a zero lead time be included in the MFLT computation just as is done with a low miss. This will cause an over-forecast to reduce the score to exactly the same degree as an under-forecast with the same size error.

This discussion of errors in the final forecast has assumed that if an observed crest is 8.0 m, a forecast of 7.5 or 8.5 m does not predict that crest. In actual practice, things are a bit more complicated. It is normally understood by all concerned that a forecast of 7.5 m is not intended to mean that the crest will be exactly that value. Small differences are expected, and an observed crest of 7.4 m would probably not be considered a "miss." What is involved here is that any forecast, regardless of how stated, actually indicates a range of values rather than one specific value. If a forecast is issued as, say, "7.4 to 7.7 m," then a bracket has been expressed; and if the crest falls within that bracket, it would not be considered a "miss." If a forecast is expressed as a single number, then a bracket has not been expressed, but the implication of a bracket still exists, and it is necessary to establish the size of that "implied bracket" if the principles described above are to be applied. There are many ways in which such brackets might be determined. The bracket might be the range of stage corresponding to a specified percentage change in discharge. Or, for a forecast point with an unstable stage-discharge relationship, it may be a fixed amount reflecting the degree of the instability. It could be based on the slope of the stage-damage curve, or it may be simply a fixed percentage of the stage involved. The important thing is that forecasts be treated as a range of stage rather than as a single value and that the range be established in a manner that is physically meaningful, consistent, and reasonably objective.

#### FORECAST REVISIONS AND REFINEMENTS

In the examples discussed, a forecast event was shown to involve a series of individual forecasts, some made during the storm and one made after the storm. These forecasts differed from each other because each was based on a different rainfall accumulation. In actual practice, several forecasts may be issued after the end of the storm, all based on the final rainfall accumulation, but differing from each other. Such forecasts are based in part on observed discharge data from upstream points or from the point in question. This information permits the forecaster to revise or refine the forecast as the crest approaches. The treatment of such revisions in the evaluation system is, with one exception, no different from the treatment of other forecasts. A revision is an updated issuance, presumably more reliable than the earlier forecasts. The fact that it is issued because of the availability of later and more complete data rather than because of additional rainfall is of no interest to the user. The distinction has no effect on the user and is therefore not considered in the evaluation.

The exception to the foregoing is the case in which a later forecast is not a revision but a refinement. Consider again the example of figure 1. Suppose that the third forecast had been issued as "7.9 to 8.3 m." The bracket of 0.4 m expresses the degree of uncertainty existing at 0900, the

time of issuance. Now, suppose that 5 hr later additional information enables the forecaster to reduce that uncertainty to 0.2 m and issue a forecast at 1400 of "7.9 to 8.1 m." This forecast is of value to the user and should be issued. Its lead time, however, is only 8.0 hr, and, if it were to be included in the computation of the score, the MFLT would be  $(11.5+12.3+13.0+8.0)/4$ , or 11.2 hr, less than the value obtained by using only the first three issuances. Obviously, then, a forecast that is a refinement rather than a revision should not be included in the computation of MFLT. The distinction is made by comparing the bracketed forecasts. If the entire bracket of the later forecast falls within the bracket of the earlier forecast, it is a refinement and not a revision. It should be noted here that it is possible for such a refinement to miss the crest where the earlier forecast with its larger bracket would have been a "hit." In such a case, it is actually a different forecast, one which serves the user poorly, and it should be included in the computation.

#### SUMMARY OF COMPUTATIONAL RULES

An evaluation system must meet certain basic criteria. As explained earlier, it must be noncorrupting. In addition, it must be completely objective and clearly describe the method of determining the score under any conceivable combination of circumstances. The basic concept of MFLT and its major provisions have already been explained in some detail. In this section, is presented a set of computational rules that describe, in a nonambiguous manner, the method of computing MFLT in any type of situation. These rules have been rather carefully thought out and, to the best of the author's knowledge, do constitute an objective procedure that will not yield anomalous results under any circumstances. This is not to imply, however, that they are perfect or even the best rules that might be devised. It is quite possible that in implementing the MFLT system one would wish to change some of them. This should be done with caution since some rather subtle interrelationships exist. Changes should be made only after carefully considering the effect under a wide variety of operational conditions.

$$\text{MFLT} = (\sum I)/N$$

where I is the time interval from issuance of forecast to the occurrence of the predicted stage; and

N is the number of forecasts applicable to the event.

The verification bracket is "VB." It is determined, as described earlier, by any objective technique selected by the user. Each forecast is considered to involve a range of stages from the specified stage minus one-half VB to the specified stage plus one-half VB. If a forecast is stated as a range, the "specified stage" is the midpoint of that range.

A "high miss" occurs with each forecast that exceeds the observed crest by more than one-half VB.

A "low miss" occurs when there are no "high misses" and when the last forecast issued is below the observed crest by more than one-half VB.

Subject to the following restrictions, the computation is to include all forecasts that predict the crest or points on the rising limb of the hydrograph.

- A. Forecasts for levels below flood stage are not included.
- B. If more than one forecast predicts the same stage (entire bracket of a later forecast is within the bracket of an earlier forecast), the later forecast is not included unless it is a miss.
- C. If forecasts for a series of points on the rising limb of the hydrograph are issued at the same time, only the one calling for the latest point on the hydrograph is included.

If a "low miss" occurs, one zero "I" value is included in the computation.

For a "high miss," "I" is the time interval from issuance to the occurrence of a stage that is as far below the observed crest as the predicted stage is above it. Note that it is possible for the interval so defined to be negative. Note also that if the forecast exceeds the observed crest by more than the difference between the crest and base stage, the stage described does not exist. In this case, the MFLT for the entire event is zero.

For each "high miss" that is not followed by a forecast that successfully predicts the crest, one zero "I" value is included in the computation.

If flood stage is reached before the first flood forecast is issued, one zero "I" value is included in the computation.

If a secondary rise is involved, the first occurrence of a stage is to be used.

If flood stage is reached and no flood forecasts are issued, this constitutes an event with zero MFLT.

If one or more flood forecasts are issued and flood stage is not reached, the MFLT for the event is computed as described above without regard for the fact that all points on the observed hydrograph are below flood level.

If the computed MFLT for an event is negative, then  $MFLT = 0$ .

Note: The conversion of a negative MFLT to zero expresses the thought that a zero MFLT indicates that the set of forecasts was of no value and that this is the worst possible situation. Quite possibly, a set of forecasts so grossly in error as to result in a negative MFLT might be considered to serve the user in a worse manner than would no forecasts at all. If it is desired to express this concept, then the negative score does have meaning and should be retained.

#### THE EFFECT OF TIMING ERRORS

In the preceding discussion, a point that has not been mentioned is that a flood forecast normally consists of a predicted stage and in addition a prediction of the time at which that stage is to occur. What has been considered thus far is the effect on the user of the combination of predicted

and observed stages with no regard for the fact that there may be a discrepancy between the predicted time of occurrence and the actual time of occurrence.

Such discrepancies, or timing errors, are relatively unimportant and, quite possibly, can be ignored. When a stage above flood level is predicted to occur at a certain time, a user may be expected to take precautions as soon and as fast as is prudently possible. He does not normally delay the start of his measures because the predicted interval is somewhat greater than the time required to complete those measures. Consequently, a timing error of a few hours in an otherwise good forecast is not of great importance.

On the other hand, under conditions of continuing rainfall, timing errors tend to be systematic rather than random. As an illustration of this, consider figure 2. Note that this is the same event as is shown in figure 1. In addition to showing the forecasts plotted at the time of issuance, figure 2 also shows the computed hydrographs (dotted) on which the first two forecasts are based. This shows clearly that the rain which fell after forecast preparation, causing stages to exceed those predicted, also caused the predicted stages to occur considerably earlier.

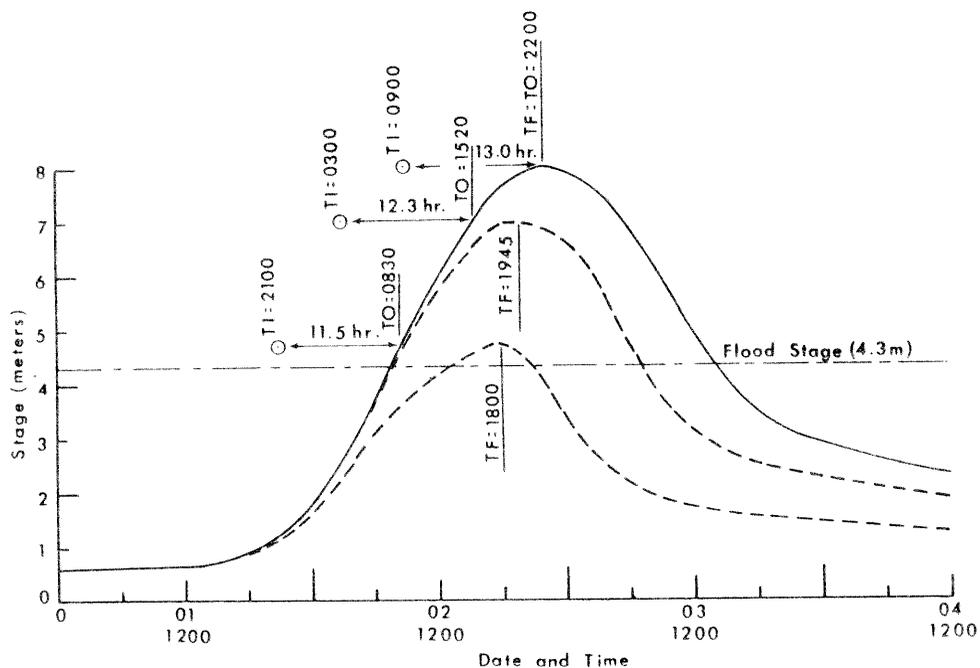


Figure 2

If it is desired to have the MFLT reflect timing errors, this can be accomplished by multiplying the lead time for each forecast by a "Timing Error Factor" (TEF). This factor is equal to unity if the timing error is zero and decreases linearly to zero as the error approaches the interval from forecast issuance to predicted time of occurrence. It is given by:

$$TEF = 1 - \frac{|TF-TO|}{TF-TI}$$

where TI is the time of issuance of the forecast;

TF is the time the stage is predicted to occur; and

TO is the time the stage does occur.

Applying the adjustment to the example of figure 2 gives the following:

Fcst. no.	TI	TF	TO	TF-TO	TF-TI	TEF	I	(TEF)(I)
1	2100(01)	1800	0830	9.5	21.0	0.55	11.5	6.3
2	0300	1945	1520	4.4	16.8	0.74	12.3	9.1
3	0900	2200	2200	0	13.0	1.00	13.0	13.0

The MFLT, adjusted for timing error, is  $(6.3+9.1+13.0)/3$ , or 9.5 hr.

Note that "TEF" can be a negative quantity. If it computes as such, it should be set to zero unless "I" is negative; in which case, "TEF" must always be set to unity.

The TEF adjustment is presented here as an option for possible use with MFLT. As stated above, the effect of timing errors on the user is not felt to be great and the need to reflect those errors in the evaluation score is therefore questionable. It should be borne in mind when considering the use of TEF that MFLT is intended to be a meaningful physical quantity rather than simply an index of the value of forecasts. The inclusion of this adjustment in the computation probably detracts somewhat from the conceptual qualities of MFLT.

#### THE USE OF FORECAST PRECIPITATION IN RIVER FORECASTS

In the examples given, forecast issuances were assumed to be made at some time after the measurement of the precipitation on which the forecast was based. It is often possible to prepare and issue river forecasts before the occurrence of all of the rainfall, basing them, at least in part, on rainfall forecasts. Although this practice has not been treated explicitly in the discussion, the use of predicted precipitation in the preparation of river forecasts has no effect on the method used to evaluate those river forecasts. That is, if perfect precipitation forecasts were always available for a period of 12 hr into the future, the river forecasts issued would be approximately the same as those based on observed precipitation, but they would be issued 12 hr earlier. The users of such forecasts would be receiving 12 hr additional warning time; and the MFLT, computed as described earlier, would be 12 hr greater.

In actual practice, river forecasts that involve predicted precipitation are likely to exhibit a lesser degree of accuracy than those based solely on observed rainfall. There are two reasons for this. First, the application

of precipitation input to a hydrologic model yields a computed hydrograph, not a forecast. The forecast is produced by a human forecaster, based principally on the computed hydrograph, but using in addition observed stages and/or discharges from various river stations, including the one for which the forecast is being prepared. Normally, the forecaster has available the observed hydrograph, at the forecast point, up to the time of forecast preparation, and he uses this to adjust, or "up-date," the computed hydrograph. If, through the use of forecast precipitation, he makes this forecast earlier, a smaller portion of the observed hydrograph is available to him, and the accuracy of the predicted portion of the hydrograph may be expected to suffer.

The second reason is that while observed precipitation values, and the areal means computed from them, usually involve sizeable errors, forecast precipitation generally attains an even lower level of accuracy. This is especially true when the precipitation forecast is made and used before the rainfall begins. Errors in volume, timing, and location of rainfall will all affect the river forecast.

In spite of an expected adverse effect on accuracy, the logic of making use of precipitation forecasts is rather firmly based. Limited studies have shown that, in the contiguous United States, the most frequent duration of runoff producing rainfall is about 12 hr and that, at the end of any 6-hr period within a storm, the probability of the occurrence of additional runoff producing rain is slightly greater than 0.5. That probability is, of course, highest early in the storm and decreases as the storm continues. At the end of the first 6 hr, the probability that the storm is not over is approximately 0.75. It does not drop below 0.5 until the duration has exceeded 24 hr. It may be said then that if river forecasts are being prepared early in a rainfall event the inclusion of additional forecast precipitation is more sound, statistically, than assuming that the storm has ended.

As has been pointed out earlier, forecast errors, whatever the cause, may either help or hurt the user. It seems reasonable to conclude, however, that in the long run the effect of an increased error level would be harmful. Thus, the use of forecast precipitation in the preparation of river forecasts may be expected to have two opposite effects on the user: increasing the lead time and decreasing forecast accuracy. MFLT is intended to reflect both effects and express, in a single evaluation score, the net gain or loss to the user.

#### MAXIMUM ATTAINABLE VALUES OF MEAN FORECAST LEAD TIME

If an evaluation system were to be based on forecast verification, it would express a score as some function of forecast errors. A perfect set of forecasts would then result in a score of zero. Even though the attaining of such a condition might well be impossible, it is obvious what the ultimate score, representing the "goal" of the forecast service would be, and the difference between that score and the score for an individual event is equally obvious.

When using MFLT, the situation is somewhat different, since a zero score indicates that the forecasts have no value and a high degree of value is indicated by a numerically large score. The ultimate then, for this type of score, is not obvious but must certainly exist. Since the MFLT pertaining to a specific event has much more of a physical meaning than an error statistic, the need to know the maximum possible value is not crucial. It is probably of value nevertheless to gain some insight into the factors that define the ultimate score and to explore the methods that might be used to evaluate it.

It is quite obvious that MFLT scores resulting from forecasts of floods on large, slowly responding rivers will be much greater than those resulting from forecasts of small, flashy streams. Consequently, with rare exceptions, MFLT scores can be compared only with scores for other events at the same forecast point. Thus, the ultimate, or maximum attainable value of MFLT is unique for a particular forecast point and is a function of the characteristics of the river system and of the storm causing the flood event. The relationships between maximum MFLT and some of these characteristics are explored in this section.

To make these investigations, a hydrologic model calibrated to an existing catchment was used. This catchment, located in the Eastern United States, has an area of 2,120 km<sup>2</sup> and has a concentration time (beginning of runoff to peak discharge) of approximately 30 hr for an event with a runoff duration of 6 hr. Extended low water flow normally is about 4 m<sup>3</sup>/s, which results in a stage of 0.5 m. Flood stage is 4.3 m and results from a discharge of 390 m<sup>3</sup>/s. The maximum stage of record is 10.9 m and involved a discharge of 2,310 m<sup>3</sup>/s.

A series of synthetic storms were applied to the model representing the physical behavior of this catchment. As in any such investigation, it was necessary to make certain assumptions about the synthetic storms. These assumptions were that the precipitation was distributed uniformly in regard to both area and time. That is, the channel response function used to model the hydrograph was one developed from historical storms that presumably involved, on the average, uniform areal distribution of rainfall. Also, each storm, regardless of precipitation volume or duration, had the same rainfall amount in each 6-hr period.

It is quite obvious that if areal distribution of precipitation is not uniform, the response time, and consequently the MFLT, will be affected. For this reason, no attempt has been made to model the effect on MFLT of upstream or downstream concentrations of runoff. What was examined was the effect of variations of the magnitude of the peak stage, antecedent moisture conditions, and storm duration. The effect of these factors on the maximum attainable MFLT is less obvious. In all cases, it was assumed that a forecast was issued at the end of each 6-hr rainfall period, that the time required for data collection and forecast preparation and dissemination was zero. It was further assumed that each forecast represented an error-free hydrologic analysis of the precipitation already fallen.

Figure 3 illustrates the effect on maximum attainable MFLT of the magnitude of the rise. It shows the hydrographs resulting from seven different rainfall

events, all with the same antecedent moisture conditions, all of 18 hr duration and ranging from 94 mm to 210 mm in volume. The peak stages are 4.3 m (flood stage), 5 m, 6 m, 7 m, 8 m, 9 m, and 10 m. Being 18-hr storms, three forecasts can be computed for each, at 6 hr, 12 hr, and 18 hr. The first forecast (6 hr), in all cases, predicts a stage less than 4.3 m and consequently is not involved in the computation of MFLT. The second forecast for each event, issued at 12 hr, is plotted in figure 3. This forecast, in each case, predicts a stage equal to that which would occur if the rain ceased at 12 hr. These stages vary from 2.7 m to 6.8 m. Note, however, that in all seven cases the stage in question occurs at approximately the same time (25.2 hr) on the rising limb of the hydrograph, resulting from the total storm. Thus, the lead time for the second forecast is approximately 13.2 hr in every case.

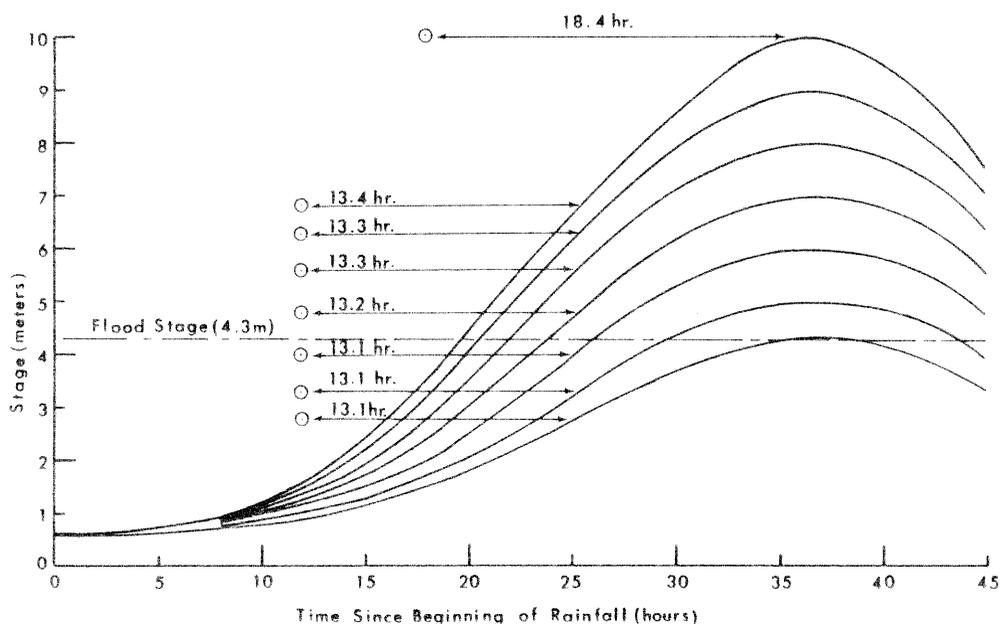


Figure 3

The third forecast in each case is issued at 18 hr and correctly predicts the crest that occurs at approximately 36.6 hr regardless of magnitude. In figure 3, the third forecast is plotted for only the 10-m event. The third forecast then always has a lead time of approximately 18.6 hr. MFLT for the seven events is computed as follows:

Maximum stage	12-hr forecast lead time	18-hr forecast lead time	MFLT
4.3	13.1	18.9	*18.9
5	13.1	18.8	*18.8
6	13.1	18.7	*18.7
7	13.2	18.6	15.9
8	13.3	18.5	15.9
9	13.3	18.4	15.9
10	13.4	18.4	15.9

Note that, for the first three events (marked with asterisks), only the 18-hr forecast is included in the computation, since the 12-hr forecast, as shown in figure 3, predicts a stage below flood.

What this example shows then is that as long as the event involves the same number of flood forecasts, the MFLT is virtually independent of magnitude. For the three smaller events, where the 12-hr forecast is not included, the MFLT is slightly larger and a discontinuity exists between the two groups. It may be noted also that the 6-hr forecast for each event has a lead time of approximately 11.2 hr and that, if a rise were large enough that this forecast were to predict a stage above flood, the MFLT would be approximately equal to the mean of 11.2, 13.4, and 18.4, or 14.3 hr, creating another discontinuity in the relationship between flood magnitude and maximum attainable MFLT. It should also be noted that in the case of an extremely large event, the flood stage might be reached less than 6 hr after the start of rainfall and before the first forecast could be issued. Under these circumstances, one zero lead time would be included in the computation of MFLT and another discontinuity would occur. These discontinuities result from the existence of a flood stage, a stage below which there is no damage and above which damage does occur. Flood stage is a discontinuity of nature (or of man's works) and it is therefore not inappropriate that discontinuities should exist in a computed quantity (MFLT) which recognizes the existence of a flood stage.

The discontinuities noted are small compared to the lead times involved. Furthermore, there are large ranges of flood magnitude within which the maximum attainable MFLT is virtually constant. Keeping in mind the effect of the original assumptions made in the analysis, it might be concluded that the maximum attainable MFLT for a forecast point is largely independent of the magnitude of a flood event.

A second set of simulations explores the effect of antecedent moisture conditions on MFLT. It is obvious of course that for a given rainfall event the magnitude of the resultant rise will be highly dependent upon antecedent moisture. What is being investigated here, however, is the effect on the forecast operation and on the forecast user of varied moisture conditions antecedent to river rises of the same magnitude. These events must then involve differing amounts of precipitation.

Figure 4 illustrates the effect. Two hydrographs are shown, both reaching a crest stage of 8.0 m. Each was simulated by applying to the catchment model an 18-hr period of rainfall, beginning at time zero and distributed equally among three 6-hr periods. The solid graph, labeled "wet," was based on the assumption of wet antecedent conditions and involved 85 mm of rainfall. The dotted graph, labeled "dry," was based on the assumption of dry antecedent conditions and required 225 mm of rainfall to simulate the same 8.0-m crest stage.

Note that the "dry" hydrograph occurs approximately 3.7 hr later than the "wet" hydrograph and has a slightly steeper rise. The reason for this is that the dry antecedent conditions result in high infiltration rates in the early part of the storm, causing the center of mass of runoff to occur later

than it does with wet antecedent conditions. The effect is to give the hydrograph the characteristics of one which would result from a rainfall event having a shorter duration and occurring somewhat later in time.

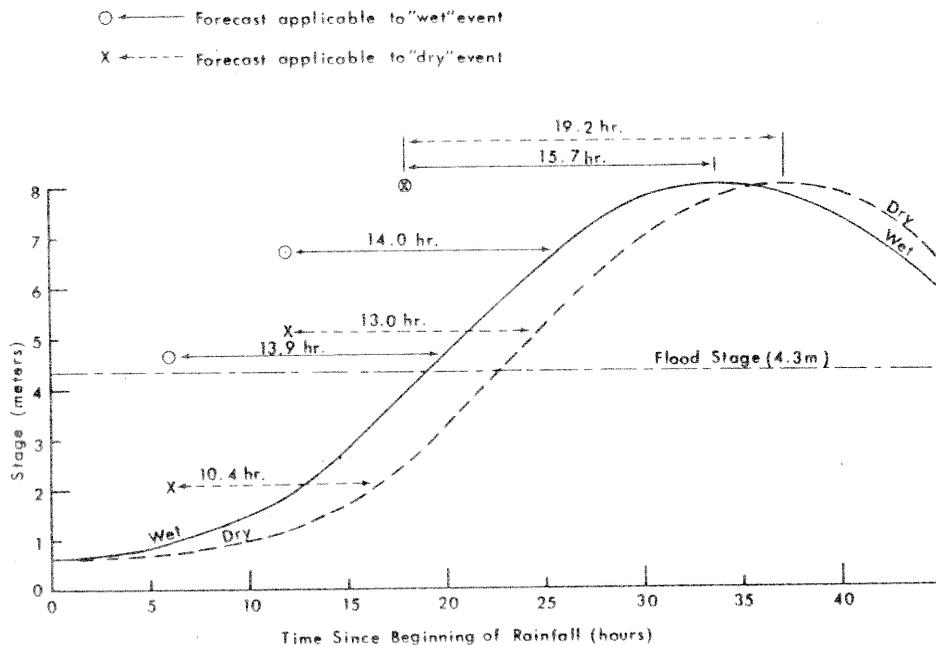


Figure 4

It might appear at first thought that since dry antecedent conditions tend to delay the rise that this would result in more lead time being available to the forecast user and in higher MFLT scores. Actually, it does not work out quite this way. Figure 4 also shows the set of three forecasts applicable to each event. Considering only the third forecast in each set, issued at 18 hr, the lead time for the "dry" event is in fact 3.5 hr greater than for the "wet" event, almost equal to the average time displacement between the two hydrographs. Other, earlier forecasts are also involved, however; and it is clear that the first two forecasts for the "wet" event, which predict stages of 4.6 m and 6.7 m, respectively, are a much better indication of the final hydrograph than are the two preliminary forecasts for the "dry" event, which call for stages of 2.0 m and 5.1 m. What has happened is that the antecedent conditions that cause the "wet" event hydrograph to occur earlier in time also cause the preliminary forecasts to be better indicators of subsequent events. The 6-hr and 12-hr forecasts for the "wet" event are based on one-third and two-thirds of the total precipitation and involve 28 percent and 63 percent respectively of the total computed runoff. In the case of the "dry" event, these two forecasts are still based on one-third and two-thirds of the rainfall, but involve only 7 percent and 39 percent of the runoff.

The MFLT for the "wet" event is computed from the lead times for all three forecasts and is equal to 14.5 hr. For the "dry" event, the first forecast is not included since it calls for a stage below flood level and the MFLT

is 16.1 hr. Thus, the difference between the two MFLT scores is 1.6 hr, less than half of the time displacement between the two hydrographs.

The conclusion is then that variations in initial moisture conditions affect the response of the river and the forecast operation in two ways, which, when related to the forecast user, tend to compensate. The maximum attainable MFLT is therefore not significantly affected by antecedent moisture conditions.

Figures 5, 6, and 7 illustrate the effect of storm duration on maximum attainable MFLT. They show three simulations, all attaining a crest stage of 8.0 m, but resulting from precipitation durations of 6, 18, and 36 hr. The same antecedent soil moisture conditions were assumed for all three events.

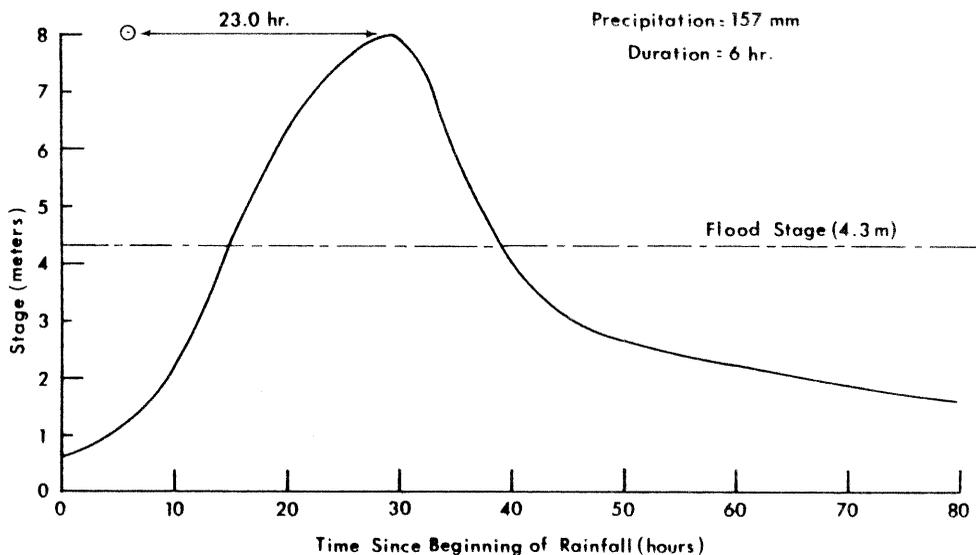


Figure 5

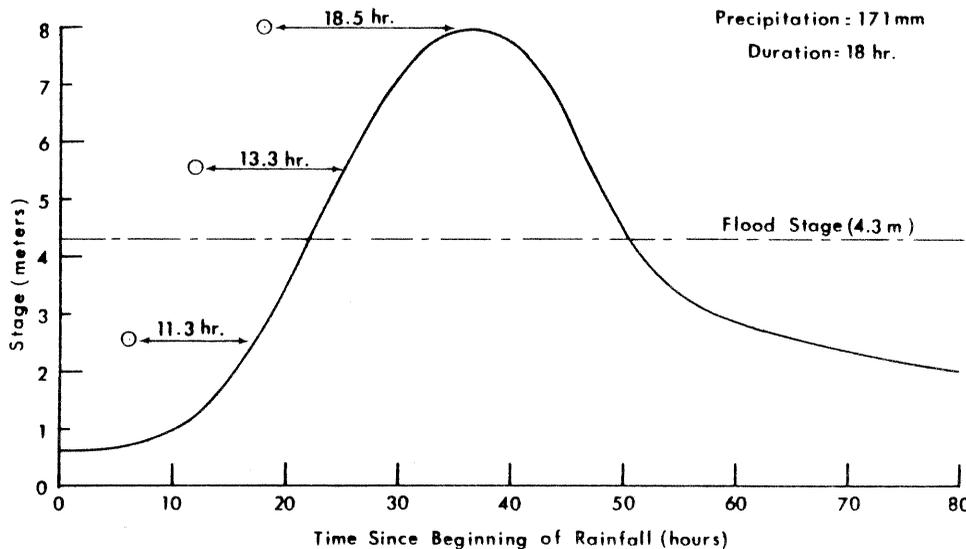


Figure 6

Note that in every case the set of forecasts consists of one final forecast, which predicts the crest and is issued at the end of the rainfall, and a number of preliminary forecasts, the number being equal to one less than the number of 6-hr rainfall periods. The lead time for the final forecast is, in each case, the interval from cessation of the rainfall to occurrence of the crest. In the 6-hr duration event, the crest occurs at 29.0 hr. The 18-hr storm produces a crest at 36.5 hr, and, when the duration is 36 hr, crest time is 50.2 hr. Thus, a 6-hr increment of duration causes the final forecast to be issued 6 hr later, but delays the crest by only 4.2 hr on the average, and, as duration increases, the lead time provided by the final forecast decreases.

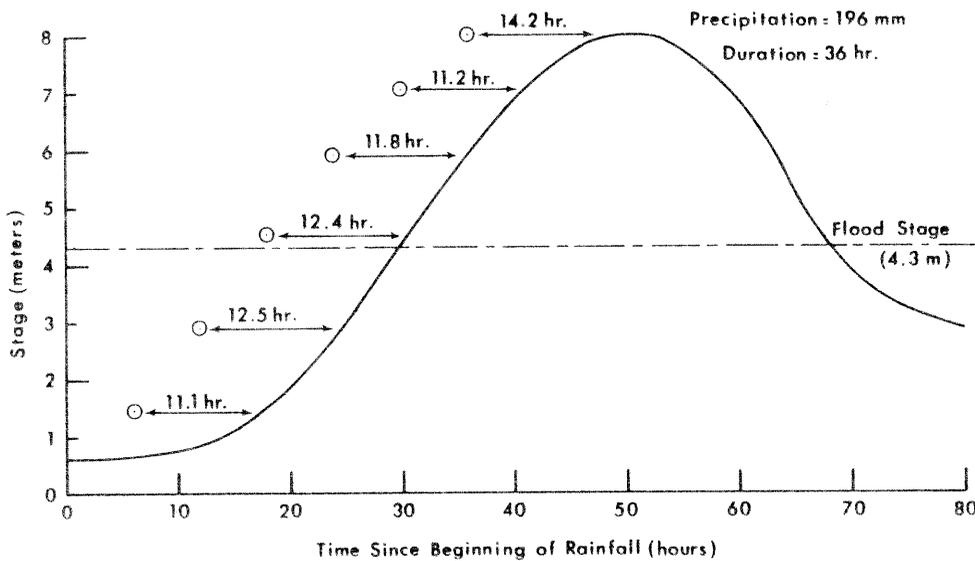


Figure 7

The MFLT is composed of the lead time of the final forecast and that of all preliminary forecasts that predict stages above flood level. Figure 7 shows that the lead times for the five preliminary forecasts do not vary greatly from one another but are somewhat less than the final forecast lead time, 2.4 hr less in this case. This difference is typical and results from the characteristic, concave downward shape of the portion of the hydrograph just prior to the crest. The lead time for the final forecast has been shown to decrease with increasing duration. The lead times for the preliminary forecasts are smaller than for the final forecast, and, the longer the duration, the more of them are included in the average. Thus, as storm duration increases, the maximum attainable value of MFLT decreases. The MFLT values for the three events shown are 23.0 hr, 15.9 hr, and 12.4 hr.

The conclusion then is that for specified areal and temporal distributions of rainfall the maximum attainable MFLT for a particular forecast point is largely independent of all storm characteristics, except duration. The relationship may be derived and, for the catchment used in this study, is shown in figure 8. The plotted points represent six simulations, with durations ranging from 6 to 36 hr, and including the three that are shown

in figures 5, 6, and 7. Beyond 36 hr, the curve is assumed to approach, as a lower limit, a value equal to the average lead time for preliminary forecasts.

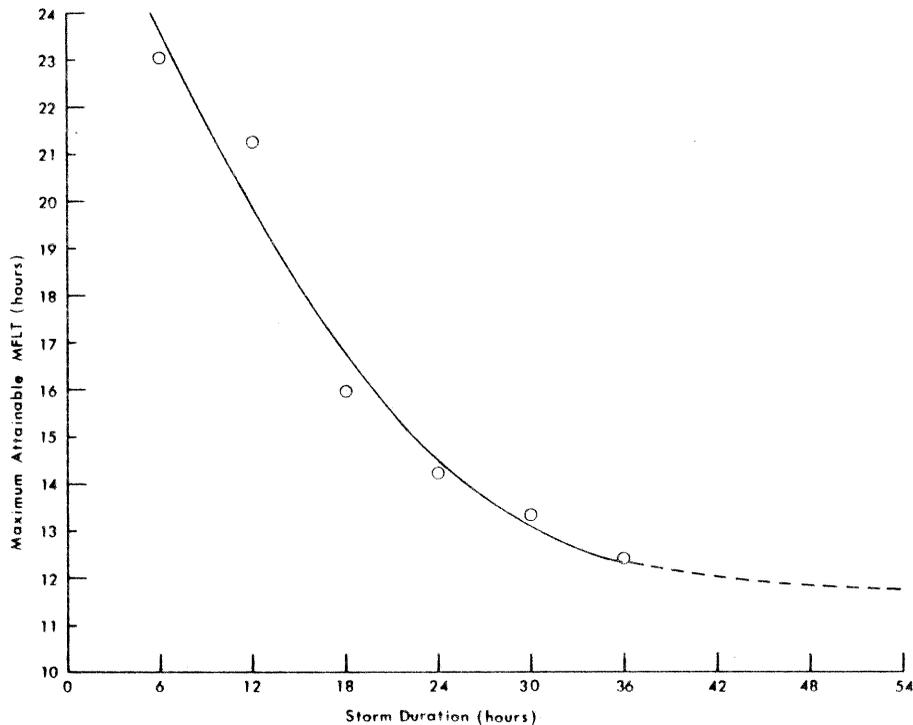


Figure 8

These studies have been based on a catchment analysis. If a similar analysis were to be made for a downstream point in a large river system, it seems logical to conclude that the results would be the same if the same initial assumptions were made. It must be noted, however, that the assumption of uniform areal distribution of rainfall becomes less realistic for larger drainage areas.

#### APPLICATIONS AND USE OF MEAN FORECAST LEAD TIME

In the "Introduction," it was pointed out that there were numerous purposes that a verification or an evaluation program might serve and that the technique used for computing the score should be related to the use which is to be made of it.

Verification of operationally produced forecasts is often suggested as a means of ascertaining the simulation accuracy of a hydrologic model relative to that of other models. The author feels that neither MFLT nor any operational verification system is suitable for this purpose. Such an effort would not only involve the preparation of two or more forecasts (one for each model) in real time, but would also suffer from having the results partially obscured by the high noise level commonly present in real time data. In addition, the collection of sufficient verification data on which to base

a meaningful inter-comparison would take many years. The determination of the relative accuracy of hydrologic models is best accomplished by laboratory testing with the use of historical data. The data must be selected to represent an appropriate scope of hydrological and climatological conditions and experience; the calibration and simulation must be performed in such a way as to minimize the effect of all factors other than model accuracy; and the results must be analyzed rather thoroughly to relate type and magnitude of errors to model characteristics. Such investigations were described by Sittner (1969, 1973, and 1976) and by WMO (1975). The futility of attempting to acquire such information in real time is evident in these references.

MFLT is intended instead to be a means of monitoring the forecast service as a whole. A forecast operation consists of many things. It involves not only hydrologic and hydraulic models, rain gages, and computers but also administrative policies in regard to staffing matters, hours of operation of field offices, data collection by manual or automatic means, delegation of authority, and numerous other details of the operation. Since, at any point in time, a number of these many aspects are likely to be undergoing change, it is necessary for management to constantly check the effect on the final product of all of the parts of the forecast service in combination. MFLT is intended to be a tool that serves this purpose.

In making a management evaluation of a forecasting service, the ultimate question usually becomes "Does the service produce monetary benefits which exceed the cost of operating the service?" or "If a change in the service is made, does that change increase or decrease the benefit and by how much?" Since potential flood damages are so great compared to the cost of operating a forecasting service, a favorable benefit-cost ratio can usually be assumed. Evaluating the effect of a change in the service in comparison to the cost of making the change however is not nearly so simple. A value-measuring system, to be ideal for these purposes, would have to express an evaluation score in dollars or other monetary units.

A measure of the monetary savings that accrue from a forecast service does not reflect the value of lives that may have been saved by flood warnings, unless one is able to place a monetary value on a human life. This is difficult and most would agree that the saving of lives and the reduction of property damage are noncommensurate objectives. In the present context, however, this does not present a great problem. The situation in which a person loses his life due to suddenly rising water and in which his life could have been saved by a flood warning is typically a "flash flood" situation. While warnings are provided for such situations and while lives are saved as a result, the severe time constraints involved usually dictate a type of forecast operation somewhat different from that which has been discussed. That is, a flash flood forecast, or warning, is likely to be a single qualitative statement indicating a yes-or-no situation rather than a series of quantitative stage forecasts defining a hydrograph. MFLT is a procedure for evaluating the type of forecast event that occurs on larger, more slowly responding streams rather than a flash flood event.

It is true of course that people lose their lives in floods of the type being dealt with herein. Experience shows however that almost invariably the victim is aware of the flood condition and meets his fate as the result of an injudicious action on his part. Obviously, then, better flood warnings do not prevent this type of tragedy, and, unless public education is to be considered a part of the forecast service, it is feasible to evaluate the service solely on the basis of the value of the reduction in property damage.

Numerous investigators have studied the economic effect of flood warnings. Day (1969) prepared a simulation model of the response to, and economic effect of, floods and flood warnings in residential areas. Assuming "reliable warnings," he evaluated flood losses under conditions of no warning and of a number of different warning intervals. He thus recognized and implicitly presented the relationship between lead time and economic benefit. Day assumed "reliable warnings" and thus did not address himself to the effect of excessive error in the forecasts. MFLT is, as was stated earlier, the average warning time provided by a group of forecasts, suitably adjusted for forecast inaccuracy. The methods and rationale for making the inaccuracy adjustments have been discussed in some detail to clarify the definition of MFLT given earlier. That is:

MFLT is the average warning time that would be provided by a group of error-free forecasts that would have affected the users in the same manner as did the group of forecasts actually issued.

By this definition, there clearly exists a relationship between the monetary value of forecasts and MFLT, whether the forecasts be of the idealized "suitable accuracy" type or of the type often encountered in actual operations.

Sniedovich et al. (1975) studied the human response aspects of flood forecast evaluation and attempted to relate behavior patterns to economic benefit using a systems approach. They too expressed the concept of lead time being a major input to an economic evaluation model and, in fact, had access to an earlier, unpublished description of MFLT prepared by the author.

Sniedovich's work, and that of others, indicates that the present state of the art lacks understanding of a number of processes essential to a complete modelling of the forecast-response-economic system. If these processes are ever successfully analyzed, the model that computes economic benefit as an output may well utilize MFLT as an input. In the meantime, MFLT is presented as a realistic and meaningful evaluation procedure.

## REFERENCES

- Day, H. J., 1969: Benefit evaluation of a flood warning system to urban residences in the Susquehanna River Basin. U.S. Dept. of Commerce, Environmental Science Services Administration, Purchase Order No. 8-11917.
- Sittner, W. T., Schauss, C. E., and Monro, J. C., 1969: Continuous hydrograph synthesis with an A.P.I. type hydrologic model. Water Resources Research, Vol. 5, No. 5.
- Sittner, W. T., 1973: Modernization of National Weather Service river forecasting techniques. Water Resources Bulletin, Vol. 9, No. 4, American Water Resources Association.
- Sittner, W. T., 1976: WMO project on intercomparison of conceptual models used in hydrological forecasting. Hydrological Sciences Bulletin, Vol. XXI, No. 1, report by WMO to the International Symposium on Mathematical Models.
- Sniedovich, M., Fischer, M., Uhl, V., and Davis, D. R., 1975: The evaluation of flood forecasting-response systems: A decision theoretic approach. National Oceanic and Atmospheric Administration. Contract No. 3-35108.
- World Meteorological Organization, 1975: Intercomparison of conceptual models used in operational hydrological forecasting. Operational Hydrology Report No. 7.

(Continued from inside front cover)

- NWS HYDRO 15 Time Distribution of Precipitation in 4- to 10-Day Storms--Arkansas-Canadian River Basins. Ralph H. Frederick, June 1973. (COM-73-11169)
- NWS HYDRO 16 A Dynamic Model of Stage-Discharge Relations Affected by Changing Discharge. D. L. Fread, December 1973. Revised, September 1976.
- NWS HYDRO 17 National Weather Service River Forecast System--Snow Accumulation and Ablation Model. Eric A. Anderson, November 1973. (COM-74-10728)
- NWS HYDRO 18 Numerical Properties of Implicit Four-Point Finite Difference Equations of Unsteady Flow. D. L. Fread, March 1974.
- NWS HYDRO 19 Storm Tide Frequency Analysis for the Coast of Georgia. Francis P. Ho, September 1974. (COM-74-11746/AS)
- NWS HYDRO 20 Storm Tide Frequency for the Gulf Coast of Florida From Cape San Blas to St. Petersburg Beach. Francis P. Ho and Robert J. Tracey, April 1975. (COM-75-10901/AS)
- NWS HYDRO 21 Storm Tide Frequency Analysis for the Coast of North Carolina, South of Cape Lookout. Francis P. Ho and Robert J. Tracey, May 1975. (COM-75-11000/AS)
- NWS HYDRO 22 Annotated Bibliography of NOAA Publications of Hydrometeorological Interest. John F. Miller, May 1975.
- NWS HYDRO 23 Storm Tide Frequency Analysis for the Coast of Puerto Rico. Francis P. Ho, May 1975. (COM-11001/AS)
- NWS HYDRO 24 The Flood of April 1974 in Southern Mississippi and Southeastern Louisiana. Edwin H. Chin, August 1975.
- NWS HYDRO 25 The Use of a Multizone Hydrologic Model With Distributed Rainfall and Distributed Parameters in the National Weather Service River Forecast System. David J. Morris, August 1975.
- NWS HYDRO 26 Moisture Source for Three Extreme Local Rainfalls in the Southern Intermountain Region. E. Marshall Hansen, October 1975.
- NWS HYDRO 27 Storm Tide Frequency Analysis for the Coast of North Carolina, North of Cape Lookout. Francis P. Ho and Robert J. Tracey. November 1975.
- NWS HYDRO 28 Flood Damage Reduction Potential of River Forecast Services in the Connecticut River Basin. Harold J. Day and Kwang K. Lee, February 1976. (PB-256758)
- NWS HYDRO 29 Water Available for Runoff for 4- to 15-Days Duration in the Snake River Basin in Idaho. Ralph H. Frederick and Robert J. Tracey, June 1976. (PB-258-427)
- NWS HYDRO 30 Meteor Burst Communication System--Alaska Winter Field Test Program. Henry S. Sante-ford, March 1976. (PB-260-449)
- NWS HYDRO 31 Catchment Modeling and Initial Parameter Estimation for the National Weather Service River Forecast System. Eugene L. Peck, June 1976. (PB-264154)
- NWS HYDRO 32 Storm Tide Frequency Analysis for the Open Coast of Virginia, Maryland, and Delaware. Francis P. Ho, Robert J. Tracey, Vance A. Myers, and Narmalee S. Foat, August 1976. (PB-261-969)
- NWS HYDRO 33 Greatest Known Areal Storm Rainfall Depths for the Contiguous United States. Albert P. Shipe and John T. Riedel, December 1976.
- NWS HYDRO 34 Annotated Bibliography of NOAA Publications of Hydrometeorological Interest. John F. Miller, December 1976.
- NWS HYDRO 35 Five- to 60-Minute Precipitation Frequency for the Eastern and Central United States. Ralph H. Frederick, Vance A. Myers, and Eugene P. Auciello, June 1977.

## NOAA SCIENTIFIC AND TECHNICAL PUBLICATIONS

NOAA, the *National Oceanic and Atmospheric Administration*, was established as part of the Department of Commerce on October 3, 1970. The mission responsibilities of NOAA are to monitor and predict the state of the solid Earth, the oceans and their living resources, the atmosphere, and the space environment of the Earth, and to assess the socioeconomic impact of natural and technological changes in the environment.

The six Major Line Components of NOAA regularly produce various types of scientific and technical information in the following kinds of publications:

**PROFESSIONAL PAPERS** — Important definitive research results, major techniques, and special investigations.

**TECHNICAL REPORTS**—Journal quality with extensive details, mathematical developments, or data listings.

**TECHNICAL MEMORANDUMS** — Reports of preliminary, partial, or negative research or technology results, interim instructions, and the like.

**CONTRACT AND GRANT REPORTS**—Reports prepared by contractors or grantees under NOAA sponsorship.

**TECHNICAL SERVICE PUBLICATIONS**—These are publications containing data, observations, instructions, etc. A partial listing: data serials; prediction and outlook periodicals; technical manuals, training papers, planning reports, and information serials; and miscellaneous technical publications.

**ATLAS**—Analysed data generally presented in the form of maps showing distribution of rainfall, chemical and physical conditions of oceans and atmosphere, distribution of fishes and marine mammals, ionospheric conditions, etc.



*Information on availability of NOAA publications can be obtained from:*

**ENVIRONMENTAL SCIENCE INFORMATION CENTER  
ENVIRONMENTAL DATA SERVICE  
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION  
U.S. DEPARTMENT OF COMMERCE**