

CENTRAL REGION TECHNICAL ATTACHMENT 98-03

Objective Verification Program for Hydrologic Service Area River Forecasts and Flood Warnings

George E. Marshall
National Weather Service Office
Jackson, Kentucky

Noreen Schwein
NWS Central Region Headquarters
Hydrologic Services Division
Kansas City, Missouri

I. OVERVIEW

Currently, there is no national river flood verification requirement or program in use at National Weather Service (NWS) offices. There have been several plans and programs developed over the years, yet the National Weather Service Headquarters and the Office of Hydrology (OH) have not required or implemented a field office verification program. This is probably due to a number of factors such as low temporal resolution of observed river stage data, computer limitations, and significant manual workload. Over the years, the NWS, U.S. Geological Survey (USGS), Corps of Engineers, and others have been able to install and utilize more automated and timely gaging equipment that provides higher temporal resolution of river stage data. Enhanced communications have improved the timely receipt of that data and advanced computer technology has allowed a greater ability to process the data. This article will provide a methodology that uses available observed and forecast data to verify river forecasts by calculating an acceptable window for time to reach flood stage and crest occurrence. It should be noted that it will not be completely accurate given the lack of continuous river gage readings, but is a tool to indicate a modicum level of accuracy. The process can be made to be somewhat automatic, but still requires manual workload.

Flood/Flash Flood Warnings are tracked and verified by the Office of Meteorology (OM). These warnings are verified by receiving confirmation of flooding or flash flooding within a specified county, within a specified valid time period of the warning (NWS 1987). This paper will also address verifying river flood warnings that are issued prior to the river forecast point reaching flood stage (the level at which flooding begins) and the lead time the warning provided. It should be noted here that in certain instances, the initial flood product to be issued is a flood statement instead of a flood warning. Field offices have the option to issue a statement if no threat to life or property is expected. Where "warning" is stated in the following text, it implies the initial product for the event, be it a flood warning or flood statement.

II. NEED FOR LOCAL VERIFICATION

Without a river flood verification program, an office cannot objectively measure performance of the timeliness and accuracy of river forecasts. Such information can help to identify deficiencies in areas such as data resolution, data quality, river model calibration, quantitative precipitation estimations, quantitative precipitation forecasts (QPF), and communications. The purpose of a warning program is to protect the lives and property of people in that service area. Verifying warnings can lead to improvements in that mission. With no objective method to monitor performance, it is difficult to determine deficiencies within the program, and to develop an effective strategy for improvement. Please note, this document does not address Quality Assurance (QA), although QA is a recommended adjunct to a verification program, focusing on the quality of the wording and format of the product.

Immediately upon assumption of Hydrologic Service Area (HSA) responsibility, on March 1, 1997, a widespread flood event occurred in the NWSO Jackson service area. A method of assessing the forecast and warning performance was needed. This verification program was developed to rate the effectiveness of the NWSO Jackson Flood Warning Program. The crest forecasts to be verified are those that were disseminated in the initial warning or statement and may not necessarily match the River Forecast Center (RFC) forecast, although at most times they will. Responsibility of the hydrologic program lies with the NWSFO/NWSO and results should be shared with the appropriate RFCs for use in their forecast program areas.

III. VERIFICATION METHODOLOGY

Record keeping is a necessary part of any warning and verification process. The data required to utilize this program are also needed for monthly hydrologic reports such as the E-3 "Flood Stage Report" (NWS, 1973), therefore no additional data collection is needed. A "Warning Log" (Table 6) was developed to record and track each warning as it is issued. Upon issuance of a warning (or a statement), the date and time of issuance are logged, as well as the forecast time to reach flood stage (if available) and the forecast time of crest. If a hydrograph-type forecast is provided and flood stage is not readily apparent, it can be estimated from the given points on the hydrograph. Similar information is logged for updates or amendments, but at this time, verification for each flood event is calculated for the initial warnings or significant updates to warnings (i.e., the original forecast crest was changed significantly), for individual forecast points. The date and time that flood stage is reached (actual or estimated) is recorded, as well as the date and time the stage falls below flood stage. Finally, the date, time, and level of the highest observed river stage is recorded. The completed Warning Log (Table 6) eases the verification of each initial/significant warning, and provides a convenient record of the entire event. An automated process for this has not yet been developed, but could certainly be a future enhancement.

All possible forecast and observed flood data fall into three categories:

A Hit (H) - Warning Issued and Verified - flooding occurred

A Miss (M) - Warning Issued and Not Verified - no flooding occurred, and

A Missed Event (ME) - flooding occurred, but no warning was issued, or warning was issued after flooding began, or the forecast was out of accepted range in timing or height.

Data are verified several different ways:

- A. Raw Verification - in the first process verification is based on forecast vs. observed flooding. Observed flooding is defined to begin at flood stage. A Hit would be recorded if a flood warning is issued with some amount of lead time, and the stage does reach flood stage.
- B. Time Phased (Flood Stage) - this process verifies forecast time to reach flood stage with observed time to reach flood stage, if that data is available. A window for the flooding is allowed, in the amount of $2/3$ of the forecast lead time (time from issuance of the warning to forecast time to reach flood stage) with $1/3$ on either side of the projected time from issuance to forecasted time to reach flood stage. For example, if the Warning Point is forecast to go above flood stage in 12 hours, then the window would stretch from 8 to 16 hours. This method, $2/3$ of the forecast lead time, for computing the warning verification window was recently chosen over the original absolute time block method of 3-6 hours, as being more realistic. It allows for a relative time window based upon the time it takes a site to rise to flood stage or crest. A very flashy, rapidly rising-river would have a small window, while a main stem site that takes days to rise would have a correspondingly larger window.

This is a tentative value. As more data are gathered, further research should be done to test the validity of this method. It may need to be adjusted to more accurately measure performance.

- C. Time Phased (Crest) - verification is based on the forecast time to crest versus the time of the highest observed stage. The time window for the crest is computed as in B.

Note: For Time Phased (Flood Stage) in the valid time window, an observation within one foot of flood stage would verify the warning (Schwein, 1996). For Time Phased (Crest) Verification, a forecast within one foot of the highest recorded stage in the valid time window would verify. If the forecast were 12.5 feet, with a flood stage of 12 feet, and the actual crest was 11.8 feet in the appropriate time window, the warning would verify. This allows some leeway where flood stage may not have been reached, but the crest forecast would still be considered a good forecast. This scenario would be recorded as a Miss in A.

Lead Times are necessary for all warnings, in order for emergency managers to take effective protective measures. The warning log provides a quick look at actual lead times. These can be compared to locally determined adequate lead times based upon input from emergency managers, of the time required to take appropriate actions. A lead time error index (LTEI) is not currently part of the verification table such as in Table 3, but could be computed as follows:

$$LTEI = 1 - (LTE/LT) \tag{1}$$

$$LTE = |FLT - LT| \tag{2}$$

where

LT is the actual lead time (time from warning issuance to time of actual crest or flood stage height)

FLT is the forecast lead time (time from warning issuance to time of forecast crest or flood stage height)

and LTE is the error in lead time

An LTEI of 1 would be perfect.

Once the flood event has concluded, all information is available to complete necessary NWS reports, as well as the flood verification statistics. Statistics may be updated if corrections to crest stages are received (e.g., from the USGS). Using the following table, information is entered in the appropriate block, and the verification scores can be computed.

TABLE 1
 FLOOD VERIFICATION MATRIX

OBSERVED		
FORECAST	YES	NO
YES	a.	b.
NO	c.	d.

“Hits” are entered in Block “A”, Misses are entered in block “B”, and “Missed Events” are entered in Block “C”. No entries will be made in Block “D”. These totals are analyzed using standard NWS Severe Weather verification procedures* to produce scores for Probability of Detection (POD), False Alarm Rate (FAR) and Critical Success Index (CSI) (NWS, 1987).

$$POD = a / (a + c) \tag{3}$$

$$FAR = b / (a + b) \tag{4}$$

$$CSI = a/(a + b + c) \tag{5}$$

An expanded version of the above table is actually used to compute all three verification routines, where the formulae are part of the table. Examples can be found in Tables 3, 4, and 5 of the following case studies.

IV. CASE STUDIES

A. 1-6 MAR 1997:

NWSO Jackson assumed HSA responsibility for 33 counties in Eastern Kentucky on March 1, 1997. Between March 1 and March 6, a heavy precipitation event produced flooding in three of the four river basins within the Jackson HSA. Jackson issued flood warnings for 26 forecast points during the period. Table 2 shows that of these, 9 verified (block “A”), 14 missed (block “B”), and 3 were missed events (block “C”).

TABLE 2
 VERIFICATION MATRIX (March 1997 Flood Event)

OBSERVED		
FORECAST	YES	NO
YES	a. 9	b. 14
NO	c. 3	d

$$POD = a / (a + c) = 9 / (9 + 3) = 9 / 12 = .75 \tag{6}$$

$$FAR = b / (a + b) = 14 / (9 + 14) = .61 \tag{7}$$

$$CSI = a/(a + b + c) = 9 / (9 + 14 + 3) = .35 \tag{8}$$

These calculations provide concrete numbers/criteria upon which to base hydrologic program performance. They are criteria with which the NWS personnel are familiar and can be used to identify problems such as areas where more observations are necessary to provide increased warning lead time, or where model improvements are needed.

At the time of this event no verification program was in use. A rough version of this program was developed following this event. The flood warning log described earlier had not been developed. It was the difficulty in tracking

warnings, and the completion of the monthly flood report that led to its creation. The warning log and this program have undergone numerous changes over the last year as this program has developed.

B. 14-20 Apr 1998:

This event was caused by two heavy rainfall episodes, mainly on the Kentucky and Cumberland basins. The first episode produced precipitation totals of around 3 inches. This caused significant rises, and some minor flooding through both basins. The second episode of heavy rain, of generally 4 inches or more, following one day after the first, caused moderate flooding throughout both basins.

Tables 3 through 6 show verification statistics for this event and one warning log for one flood forecast point. All the logs are not included here due to space limitations.

Table 3 shows the results of raw verification: forecast versus observed flooding. Twenty warnings were issued with only two Misses, and two Missed Events. These statistics indicate a high positive result. If this were the only indicator used, it would appear this was an exceptionally well forecast event with a high POD and CSI, and low FAR.

TABLE 3
 RAW VERIFICATION - Based on Warnings Required/Issued/Occurred

OBSERVED		
FORECAST	YES	NO
YES	16	2
NO	2	
POD	0.89	
FAR	0.11	
CSI	0.80	

However looking at the Flood Stage Table (Table 4) it can be seen that there was a problem with forecasting the time to reach flood stage as the POD and CSI are quite low. Reviewing the warning logs, the data showed there was a significant error in timing. After conferring with the RFC, it was concluded that the models were not responding accurately to the amount of rain that had fallen. There could be many reasons for this as there are numerous parameters accounted for in the river models. This is a good example of how to use the data to determine possible problem areas. (Note: There are only 19 warnings in this table. This is due to one of the original 20 warnings being issued for a site that was already above flood stage. A warning for this point was issued during the first rain episode. A new warning was issued for this site for a second crest when it was already above flood stage.)

TABLE 4
 TIMED VERIFICATION - Based upon Time to Reach Flood Stage vs Forecast Time

OBSERVED		
FORECAST	YES	NO
YES	5	2
NO	12	
POD	0.29	
FAR	0.29	
CSI	0.26	

Table 5 shows time phased verification of the Flood Crest forecast. This table indicates that there were problems with the crest forecasts also, but since the statistical calculations take into account both timing and magnitude of the crest, the source of the problem is not readily obvious. Here is where the LTEI would definitely prove beneficial by providing additional information on the timing. (An example calculation of an LTEI for one forecast is given below however, one for each case is not available.) Automating the LTEI in a spread sheet would be very useful. Analyzing the warning logs for this event showed the major problem was with the height of the crest forecast. Of the 14 Missed Events, 6 were due to height alone, 2 for timing alone, and 6 for both height and time.

TABLE 5
 TIMED VERIFICATION - Based upon Time of Crest vs Forecast Time

OBSERVED		
FORECAST	YES	NO
YES	4	2
NO	14	
POD	0.22	
FAR	0.33	
CSI	0.20	

The following table for the forecast/warning point at Fourmile, Kentucky on the Cumberland River shows how information is logged and tracked. The log shows the warning was issued before the fact and provided 8 hours and 37 minutes of actual lead time for a Hit on Flooding Forecast. However the Flood Stage Forecast was a Missed Event because the site went above flood stage outside the forecast time window. The Crest Forecast was also a Missed Event, for both timing and stage. The forecast fell just outside the appropriate time window and was off by 2 feet. The Lead Time Error Index was calculated for this example, for both the Flood Stage and Crest Forecasts, in order to demonstrate the process. The forecast for the Time to Reach Flood Stage was well outside its window giving an LTEI of .56, while the Crest forecast was just outside the time window giving an LTEI of .73.

	Time To Reach Flood Stage	Time To Crest
LT (Lead Time (hh:mm))	8:37	18:52
FLT (Forecast Lead Time)	4:52	13:52
LTE (Lead Time Error)	3:45	5:00

For Time To Reach Flood Stage

$$LTE = |FLT - LT| = |4:52 - 8:37| = |292 \text{ min} - 517 \text{ min}| = 225 \text{ min}$$

$$LTEI = 1 - (LTE / LT) = 1 - (225/517) = .56$$

For Time To Crest

$$LTEI = 1 - (LTE / LT) = 1 - (5:00 / 18:52) = .73$$

TABLE 6

SITE		FLOOD WARNING LOG															
FLOOD STAGE		1. FLOOD				2. FLOOD STAGE				3. CREST							
WARNING ISSUED	DT/TIME	FOMK2	990.0	FLOOD OCCURRED Y/N	LEAD TIME	HM/ME	FORECAST		OBSERVED		FORECAST		OBSERVED		VER +/- 1'		
							DT/TIME	WINDOW	ABOVE	BELOW	DT/TIME	WINDOW	DT/TIME	WINDOW		DT/TIME	WINDOW
04/16/98 2308L*				Y	8:37	H	17/0400L	17/0223L - 17/0537L	17/0745L	18/0600L	ME	992.2	17/1300L	17/0824L - 17/1736L	994.4	17/1800L	ME
17/01/47L							17/0400L					993.4	17/1300L				
17/06/25L							17/0700L					993.2	17/1400L				
17/11/15L												994.1	17/1700L				
04/18/98 2233L*				Y	5:27	H	19/1700L	19/1051L - 19/2309L	19/0400L	20/1630L	ME	994.6	20/0200L	19/1651L - 20/1109L	1004.0	20/0000L	ME
19/02/29L							19/0400L					995.0	19/0900L				
19/05/52L												997.0	19/1100L				
19/12/08L												1005.0	19/1900L				
19/16/36L												1005.0	20/0100L				

* initial warning or significant update.

V. ENHANCEMENTS

There are several places where modifications can be made to provide more utility to this process. While they have not been implemented in the NWSO Jackson Program, they could prove beneficial, at a future time.

Verifying the magnitude of river height within one foot is an arbitrary number which has accepted use over a large number of river forecast points. It does not work well, however, for rivers that have a small slope and fluctuate very little, perhaps only a few feet. For those points, +/- one foot leeway would be too forgiving. For flashy rivers that have a high range of fluctuation, +/- one foot would be too restricting. It has been suggested by Dr. John Schaake of OH (1998) to utilize a flood frequency curve to obtain a "significant range" from a frequent flood (e.g. 2-year flood) to an infrequent flood such as a 10, 50, or 100-year flood. This range could be compared to the difference between the forecast and observed stage to assess the accuracy of the magnitude of the crest forecast.

What we have calculated here are statistics for the HSA as a whole. Producing verification statistics on each river basin would be another enhancement. This could alert an office to a problem in forecasting for a particular basin. The overall office statistics may be quite good, however, verifying each basin individually might show one to score considerably lower than the others. This could be due to guidance and/or input parameters used to make the forecast/warning for that basin. Tracked over a period of time, this could prove fruitful in improving the forecast/warning performance for that area.

Another excellent enhancement would be to create a computer program to automatically track all data and verify/compute all verification statistics. Currently, records are manually logged, and indices (POD, FAR, CSI) are calculated using a WordPerfect Table with preprogrammed formulas. A Flood Verification Program is expected, as a future enhancement, to the AWIPS WFO Hydrologic Forecast System, some time after it is fielded.

VI. SUMMARY

These procedures allow any office to objectively evaluate their flood warning performance. Over time, they provide the potential for an office to recognize problems and construct corrective measures to improve the timeliness and accuracy of flood warnings to better fulfill the NWS mission of saving of lives and property. This program will not provide a direct answer to problems involved with providing adequate warning to the public. It will, however, alert an office to deficiencies in the Flood Warning Program. When there is a problem with either the timing or the height of a forecast, it should be obvious from the collected data and statistics. One or two flood events will not likely provide sufficient data to determine deficiencies, but a recurring problem will become apparent.

What has been presented here is oriented toward NWSO/NWSFO HSAs, however, RFCs can also use some of the verification aspects presented here to verify the timing and magnitude of river forecasts. Additional statistical calculations with regard to the magnitude such as bias, mean error or root mean square error, variance, etc., would probably be more beneficial to the RFC and have not been addressed here.

Copies of the Warning Log and the Verification tables are attached and can be downloaded and saved to WordPerfect 8.0.

VII. ACKNOWLEDGEMENTS

The authors wish to thank Robert Cox of the Missouri Basin River Forecast Center, Preston Leftwich of CRH Scientific Services Division, and Mark Walton of NWSO Grand Rapids, Michigan for their review of this paper.

VIII. REFERENCES

National Weather Service, 1987: Weather Service Operations Manual, C-72, 5.3, Exhibit C-72-3, 12-13.

National Weather Service, 1987: Weather Service Operations Manual, C-72, 6 a - c, 16-17.

National Weather Service, 1973: Weather Service Operations Manual, E-41, 3.2-3.3, 8-9.

Schaake, J., 1998: personal communication

Schwein, N., 1996: The Effect of Quantitative Precipitation Forecasts on River Forecasts. NOAA Technical Memorandum NWS CR-110. DOC, NOAA, NWS, Central Region Headquarters, Scientific Services Division, Kansas City, MO, 39 pp.

AIR FORCE
WEATHER TECHNICAL LIBRARY
151 PATTON AVE. RM.120
ASHEVILLE, NC 28801-5002