4.B1 VERIFICATION OF NATIONAL WEATHER SERVICE WARNINGS USING GEOGRAPHIC INFORMATION SYSTEMS

Ken R. Waters

NOAA/National Weather Service
Honolulu, Hawaii

1. INTRODUCTION

The NOAA/National Weather Service (NWS) is improving the focus of short-term life and property saving warnings with a new "storm-based" warning initiative (Waters, et al, 2005). These short-term warnings include Tornado, Severe Thunderstorm, Flash Flood, and Special Marine Warnings. Historically, these warnings have been issued for individual counties (or parishes). This new "storm-based" methodology will change the official warned area to be a polygon that highlights the primary threat of the weather event rather than an entire county or parish. This new practice will provide more geographic-specific warnings, thus improving the quality of the warning service. It is also expected to result in much smaller numbers of populations having to be warned. Why warn for an entire county when only a small portion of the county is actually at risk?

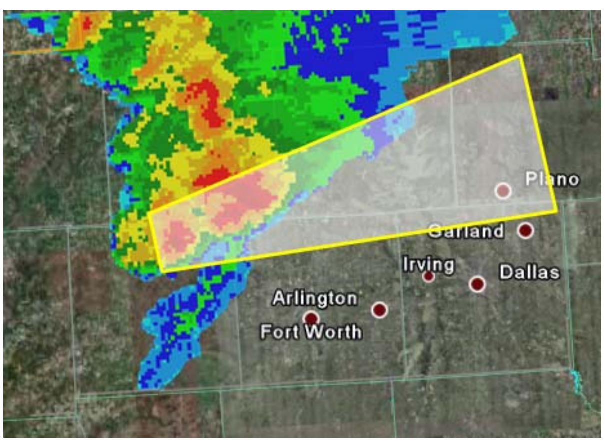

Figure 1. Graphic example of value of using storm-based (polygon) warnings versus county-based warnings (from Ferree et al, 2007).

Corresponding author address: Ken R. Waters, National Weather Service, Pacific Region Headquarters, 737 Bishop St., Ste. 2200, Honolulu HI 96813; e-mail: ken.waters@noaa.gov

The NWS collects data from these warnings and matches specific warnings against actual weather events that have occurred. With these data the NWS can then design quantitative, or verification, measures with which to gauge the quality of service. The existing verification program assumes the warning area to be an entire county. If, for example, a specific weather event (e.g., tornado) occurs anywhere within the warned county during the valid period that the warning is in effect then the warning is considered verified.

This paper addresses how Geographic Information Systems (GIS) technology can assist in calculating verification measures of NWS short-term "storm-based" warnings.

2. LEGACY VERIFICATION

The methods and measures that the NWS has used to verify short-term warnings have changed little in two decades (NWS, 1982). Verification begins with the collection of storm event reports. Initially, these consist of Preliminary Local Storm Reports (LSRs) which are typically sent within a few hours of receipt of the event reports from weather spotters, emergency management officials, and others. These unofficial LSRs are the start for the official Storm Events Database that is submitted (DOC, 1998) by each of the NWS weather forecast offices in a monthly report and archived at the NOAA National Climatic Data Center. The Storm Events Database is used for all official verification measurements.

The three primary measures (NWS, 2005b) are:

- Probability of Detection (POD)
- False Alarm Rate (FAR)
- Lead Time (LT)

Each of these measures can be calculated for a variety of phenomena. For instance, for verification of Severe Thunderstorm Warnings, the criteria (NWS, 2005a) is established to be the occurrence of winds in excess of 50 kts (26 m/s) and/or hail size of ¾ inch (0.019 m) or greater diameter.

For computation of the POD and FAR, each warning is scored in a 2 x 2 specialized contingency table:

	Forecasted?	
Observed?	Yes	No
Yes		
No		

Probability of Detection is event-based, meaning that each confirmed event (e.g., tornado, severe hail, etc.) is checked to see if an appropriate warning was issued and in effect at the time of the event. In order for the event to be considered as a "warned event" the event must pass both the temporal test and spatial test based on counties. Each event is scored as a '1' (warned event) or a '0' (unwarned event) and then the statistics are combined to determine an overall POD. Therefore, a higher POD could indicate an improved service.

False Alarm Rates are computed based on warnings. In other words, for each warning did a warned event, meeting the temporal and spatial tests, occur? The FAR is scored as a '1' for false alarm and '0' for warned event. So, a lower FAR could be used as an indication of improved service.

Lead times are also computed based on warnings. However, computation of LT is only done for warnings that were scored as a warned event. The difference between when the time the warning was issued and the time of the event is scored as the LT.

All of these scores have been computed based on counties (or parishes). If a tornado warning is issued for a county and a confirmed touchdown occurs anywhere in that county while the warning is in effect then the event is considered a warned event.

Some of these measures have been selected to be used in measuring overall performance of the NWS as part of the Government Performance Results Act (GPRA) of 1993 (OMB, 1993). One such example is the POD (or "accuracy") of tornado warnings, which is one of 14 official GPRA measures used by the U.S. government to measure agency performance.

3. POLYGON VERIFICATION

With the new storm-based (polygon) method implementation it becomes necessary to make changes to the existing verification system. No longer will measurements be based on counties. Rather, the official warning will be based on a geographic polygon that focuses on the predicted location of severe weather. For several years the NWS forecast offices have included polygon definitions as part of their severe warnings. However, this information has been little-used (Waters, 2004) since the official warning has continued to be county-based. All of this will change, however, once storm-based warnings become official. This will have implications on a number of NWS programs including dissemination and verification (Ferree, Looney, and Waters, 2007).

The existing verification system is based on a relational database of tables of warnings and events that is labor-intensive to maintain. The advent of storm-based warnings requires a new system be designed---one that is geographically-based. Ideally, this system will be based on GIS technology. This technology is already quite sophisticated, allowing complicated analyses based on geoprocessing tools that would otherwise be difficult to develop.

The same legacy performance measures (POD, FAR, etc.) can be recomputed in terms of the storm-based, or polygon, method. It is important to note, however, that the two sets of measures will not be comparable. Storm-based warnings will often be much smaller than typical-sized counties. This is analogous to the increased difficulty of an archer aiming at a smaller target. For this reason it makes sense to look at new ways to measure service quality in addition to adaptation of the legacy measures.

4. METHODOLOGY

Two sets of primary data were needed for this study. First of all, all short-term warnings had to be archived and databased. Relevant portions of each warning were extracted. These included issue times, expiration times, Universal Geographic Codes (UGCs) indicating which counties were warned, and, most importantly, the defined polygons for the warnings. The individual warning polygons were added to a spatial dataset of shapefiles for each warning type. Polygon definitions of these warnings were collected beginning in January 2003. Over 150,000 warnings have been archived since 2003. There were also some quality control issues that needed to be addressed before using these data. These included issues with erroneous polygon definitions and warnings with missing polygons.

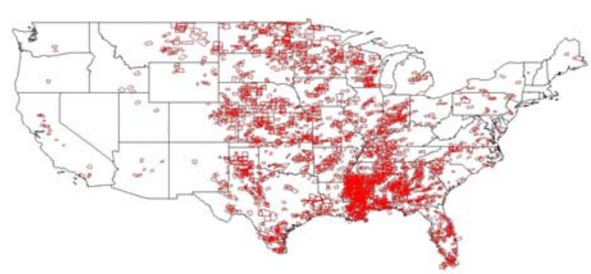

Figure 2. Plot of all tornado warnings issued for the year 2005. Note the great concentration of tornado warnings near Louisiana and Mississippi. Some of that was due to the very active hurricane season.

The other set of data needed were the event reports. These are derived from the official Storm Events Database and converted into spatial point shapefiles. Reports were broken into three categories: tornado, hail, and wind. For each report care was taken to add table attributes for the event time and magnitude (e.g., Fujita scale f-number or hail diameter size).

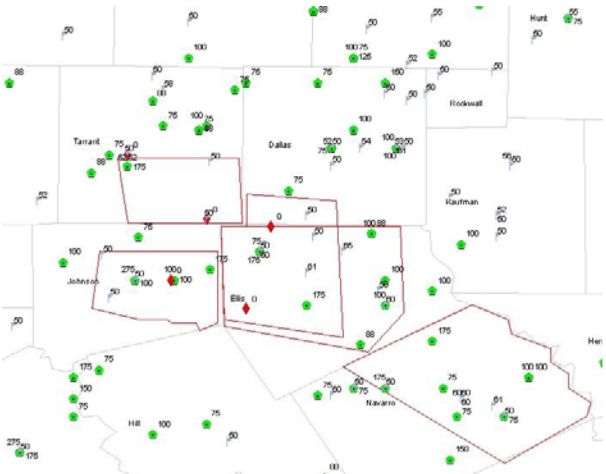

Figure 3. Example of overlaying event reports over warnings. Red outlines are tornado warnings. Red diamonds show locations of tornado events. Green symbols indicate hail reports (labeled with size in hundredths of an inch diameter) and grey flags indicate significant wind reports with estimated wind speed in knots.

These two datasets could then be easily brought into a GIS platform since both were already converted into GIS shapefiles. GIS provides the tremendous advantage of user interaction. It is easy to zoom in or out, pan, change data display attributes, or even add other display themes of interest such as highways or county boundaries.

Computation of the legacy performance measures, however, can not be done without customized programming. Some GIS software platforms do allow this. For the purpose of this work one such program was chosen: ESRI ArcView and its built-in Visual Basic for Applications (VBA). VBA was chosen due to its ability to interact with all facets of the built-in geoprocessing functions as well as interaction with the GIS user software interface. The use of custom programming allows computation of the sometimes complex performance measures.

Work so far has been limited to tornado performance metrics. In the near future we will be looking at other weather metrics as well, such as accuracy of Severe Thunderstorm Warnings using hail and wind criteria.

The first major programming exercise was to compute tornado warning PODs. To do this, the entire 2005 data base of tornado events was examined. The program looped through each event. With each event another program loop iterated through each warning. For each warning/event pair two tests were done. The first was a temporal test to determine if the event occurred while the warning was in effect. If that test passed then a second, spatial, test was performed. If the event occurred within the specified polygon associated with the warning then a score of ‘1’ was added to a POD attribute field of the warning database. If no warning was found to be in effect at the time of the event then a ‘0’ was scored in that field. The results were summarized and collected by weather forecast office and then also summarized for the entire country to obtain a single POD value.

The second exercise was based on warnings. Each warning was examined and then another loop was performed, looking for any events that occurred while the warning was in effect. Each warning was scored a ‘1’ if a qualifying event occurred and a ‘0’ if no event was found. These values could then be used to compute FAR for each forecast office and for the entire country as a whole. At the same time the LT could be easily calculated and added as a separate field to the warning table.

Size of warnings is also an aspect of interest since it ties in well with the concept of issuing smaller, more focused warnings. Sizes of warnings was computed and compared to the equivalent county areas. The goal here would be to reduce the size of the warning to cover the threatened area and thus reduce the amount of falsely alarmed population that composes the remainder of the county.

5. RESULTS

Probability of Detection for tornadoes was computed for the entire year of 2005. In all, 1340 reported tornado events were initially analyzed. Some errors in the reported locations were identified in the original Storm Events Database. These could only be uncovered through use of GIS after joining the events database to a GIS shapefile of the weather forecast offices. Primarily these errors consisted of events attributed to one forecast office but with the latitude-longitude point actually located outside the office’s warning area.

As expected, the POD using polygon warnings was lower than the associated county-based legacy value. The polygon value showed 57% of events properly warned by polygon whereas using county-based verification 76% of the events were warned for. The only other study available (Browning, 2002) comparing county and polygon verification also found a lower POD (73% for polygon versus 86% for county-based), although this was for just one forecast office and over a 3-month period.

Fig. 4. Graphical depiction of POD (accuracy) of tornado warnings issued in 2005. Green dots signify tornado events that were warned events. Red dots signify tornado events that were not warned for.

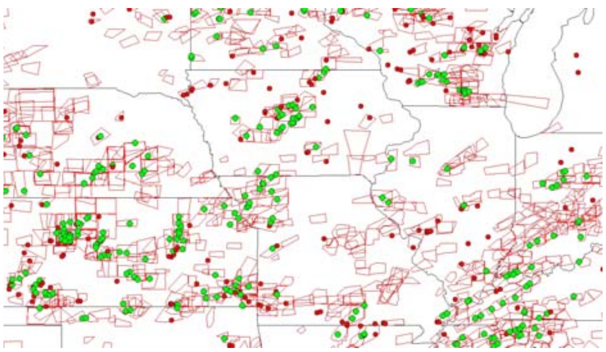

Fig. 5. Close-in view of POD with polygon warnings shown as red outlines.

There are a number of very significant factors that need to be considered when evaluating these numbers. First and foremost, it should be pointed out that the NWS forecasters who issued these warnings were still using county-based methodologies. These methodologies have been developed and trained for over many years. In some cases warnings were more or less shaped to the outline of counties. With the increased awareness of the storm-based warning methodology it is expected that much more attention will be made to the shape of the polygon in the warning.

Another factor to consider is that the verification statistics were collected in terms of the county-based system. Only one confirmed report within a county was sufficient to verify a warning. It's possible that more reports will be reported once the switch is made to polygon-based verification.

One other rather unique factor that needs to be considered is that for the year 2005 there were a large number of hurricane "eyewall" warnings that were issued under the tornado warning. 2005 was a year that broke many records for hurricane frequency in the U.S., including Hurricane Katrina. These warnings tended to be very large in area (as much as 6,800 sq. miles versus the more typical 350 sq. miles) and were used primarily for their dissemination value due to the threat of extreme winds. However, since they were issued as tornado warnings they were scored along with other warnings issued for tornadoes.

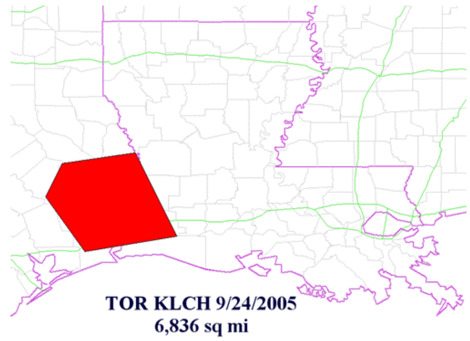

Fig. 6. Example of a hurricane "eyewall warning" issued for Hurricane Rita.

Some initial work was done to examine False Alarm Rates. This work is not complete yet but early results seem to indicate little change in FAR from county-based verification numbers.

An operational test of the "storm-based" concept was conducted from January 2005 through September 2005. 23 of the 122 weather forecast offices volunteered to test the concept. Forecasters were carefully trained to use polygons to define the warning area and to disregard county boundaries as a factor in issuing warnings. The test was a success and resulted in a substantially higher polygon POD for tornado warnings (66% for test offices versus 52% for non-test offices).

Lastly, a comparison was made of area of warnings issued for the year 2004 with the year 2005 (through September). Tornado Warning sizes increased nearly 13%. Again, this is largely attributed to the large number of "eyewall warnings" issued in 2005. Severe Thunderstorm Warnings and Flash Flood Warning area was also analyzed.

	2004	2005 (through Sep 30th)
Tornado	352	397
Severe Thunderstorm	491	499
Flash Flood	910	912

Fig. 7. Comparison of area (square miles) for the years 2004 and 2005.

An experimental measure, the County Area Ratio (CAR), was developed. The CAR was simply a ratio of the area of the warning to the area of the county/counties covered by the warning. Analysis of the first nine months of 2005 showed a tremendous 72% reduction in falsely alarmed area when using polygon warnings.

6. FUTURE PLANS

The Storm-based warning initiative is expected to be implemented in 2008. This initiative is viewed as a huge step forward in the warning service provided by the NWS. Many positive comments have been received from stakeholders looking forward to the full implementation of storm-based warnings.

There remains much potential for the use of GIS to analyze NWS short-term warnings. Many more geoprocessing tools are available and include metrics concerning size of warnings, analysis of population warned, and socio-economic analyses that can be performed. The use of GIS opens up great opportunities to analyze warnings and is absolutely essential to determining geospatial verification measures.

7. ACKNOWLEDGEMENTS

The author thanks Royce Jones and Rod Low of ESRI and is particularly indebted to Jennifer Boulware and Yanlin Ye of the NOAA Coastal Services Center, for their VBA programming expertise. The author also thanks Brent MacAloney from the NWS Verification Branch who supplied several sets of Storm Event data and Robert Bunge from the NWS Office of Chief Information Officer who provided invaluable assistance in helping to move the shapefile creation software onto a national webserver.

8. REFERENCES

Browning, Peter A., and M. Mitchell, 2002: The Advantages of Using Polygons for the Verification of NWS Warnings. 16th *Conference on Probability and Statistics in the Atmospheric Sciences, AMS.*

Ferree, John T., J. Michael Looney, K. Waters, 2007: Storm-based Warnings --- Changes to NWS Warnings for the Digital Age. *35th Conference on Broadcast Meteorology, AMS.*

National Weather Service, 1982: National Verification Plan. *U.S. Dept of Commerce, NOAA, 81pp.*

National Weather Service, 2005: WFO Severe Weather Products Specification. *NWS Instruction 10-511, NOAA.* http://www.nws.noaa.gov/directives/sym/pd01005011curr.pdf

National Weather Service, 2005: Verification Procedures. *NWS Instruction 10-1601, NOAA.* http://www.nws.noaa.gov/directives/sym/pd01016001curr.pdf

Office of Management and Budget, 1993: Government Performance Results Act of 1993, OMB.

U.S. Department of Commerce, Office of Inspector General, 1998: NWS's Verification System for Severe and Hazardous Weather Forecasting Needs Modernization. *Inspection Report #IPE-9255.*

Waters, Ken, 2004: Displaying National Weather Service Warnings in Real-time GIS Formats. *20th International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology.*

Waters, Ken, J.M. Coyne, P.L. Wolf, J.T. Schaefer, J.M. Looney, S. Naglic, B.W. MacAloney II, M.A. Tew, J. Lorens, R.J. Okulski, and N. O. Schwein, 2005: Polygon Weather Warnings—a New Approach for the National Weather Service. *21st International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology.*