P2.18

AN EXPERIMENT IN PROBABILISTIC QUANTITATIVE SNOWFALL FORECASTING

Alan M. Cope* and Mark P. DeLisi
NOAA/National Weather Service Forecast Office, Mount Holly, New Jersey

1. INTRODUCTION

For the past two winter seasons, beginning with the winter of 2001-02, the National Weather Service (NWS) forecast office in Mount Holly, NJ (WFO PHI) has been issuing experimental probabilistic quantitative snowfall forecasts (PQSFs) for five sites in its forecast area. The motivation for this experiment grew out of the "imperfect storm" of March 4-6, 2001, during which large forecast snowfall amounts for the WFO PHI area failed to materialize. Although forecast snow totals for this storm varied for different places at different times, there seemed to be a tendency in the media and the public to focus on the highest forecast snow amounts from any forecast (i.e., the "worst case" scenario). As a result, we felt it worthwhile to try to convey quantitatively the uncertainty in any snow forecast. Also, we felt a quantitative expression of the probability of different snow amounts would help draw attention toward the most likely outcome, and away from the worst case. Even so, we recognize that such a forecast would likely be of most practical value to emergency planners and other more sophisticated users.

Efforts to develop and improve probabilistic quantitative precipitation forecasts have been ongoing for a number of years. A brief review of past efforts and a list of references can be found in Applequist et al. (2002). More specifically, a similar real-time PQSF experiment for snowfall is being conducted at the NWS office in Buffalo, NY (Niziol 2003).

The main focus of our effort was to develop a set of probabilities for different snowfall amounts, given a forecast for a specific snowfall range. For example, if our office forecasted 6 to 10 inches of snow for a location, what is the probability of that place actually receiving at least 4 inches, or 9 inches, or even 24 inches? This paper will describe the process of developing, testing and distributing these probabilities to users.

*Corresponding Author Address:: Alan M. Cope, National Weather Service, 732 Woodlane Road, Mount Holly, NJ 08060; e-mail: Alan.Cope@noaa.gov

2. DEVELOPMENT METHODS AND DATA

The WFO PHI forecast area covers 34 counties in four states (PA, NJ, MD and DE), and includes 40 forecast zones (some counties are sub-divided into two zones). Historically, WFO PHI issues snowfall forecasts as an expected range of accumulations for each zone. Rather than try to develop probabilities for each zone, we decided for this experiment to select five sites that are representative of the different weather regimes within our area, and develop probabilities for those sites. Fig. 1 shows a map of the PHI forecast area with the five sites used for this study. These sites are also listed in Table 1.

As a starting point to develop probabilities, we decided to use the theoretical Gamma probability distribution. This distribution is often used to model precipitation amounts (Wilks 1995). However, we first needed to show that our forecast snow amount ranges were not significantly biased (i.e., that we did not significantly over- or under-forecast the actual snowfall). To do this we looked at the initial snowfall range forecast for 33 winter storm watches issued by our office over six winters, starting with the winter of 1995-96. The mid-range forecast amount for each watch was compared with the observed storm-total snowfall at each of the five sites in Fig 1. From our statistical analysis using the t-test (Wilks 1995) we were not able to show any significant bias in our forecast snowfall amounts for any of the sites. Some results are shown in Table 1. Note the number of cases for each site is less than 33 since no site was included in every watch. The p-values for the bias (last column) are all well above 0.1, which indicates that the bias is probably not significant.

Finding no significant forecast bias, we were able to apply the Gamma distribution to the problem of estimating probabilities for various snowfall amounts, given the mid point of a specific forecast range. To do this, we adjusted the shape parameter of the Gamma distribution such that the 50th percentile of the cumulative distribution fell at the forecast mid-range value. Fig 2a shows an example of the Gamma distribution, where the area under the curve has been shaded above the 50th percentile for a mid-range value of 5 inches. For our purposes, 26

Table 1. Sites used to develop quantitative snowfall probabilities, with results from the initial check for snow amount forecast bias.

Site ID	Site Name	Elevation (ft)	No. of Cases.	Mean Abs. Error (in.)	Bias (in.)	p(bias)
MPO	Mount Pocono, PA	1848	27	3.0	-0.04	.9615
ABE	Lehigh Vly Int'l Arpt	375	18	3.8	0.67	.6722
PHL	Philadelphia Int'l Arpt	21	17	3.6	-0.44	.4385
ACY	Atlantic City Int'l Arpt	62	10	3.9	0.71	.6825
GED	Georgetown, DE	42	8	3.7	1.80	.3728
Total			76	3.5	0.36	.5287

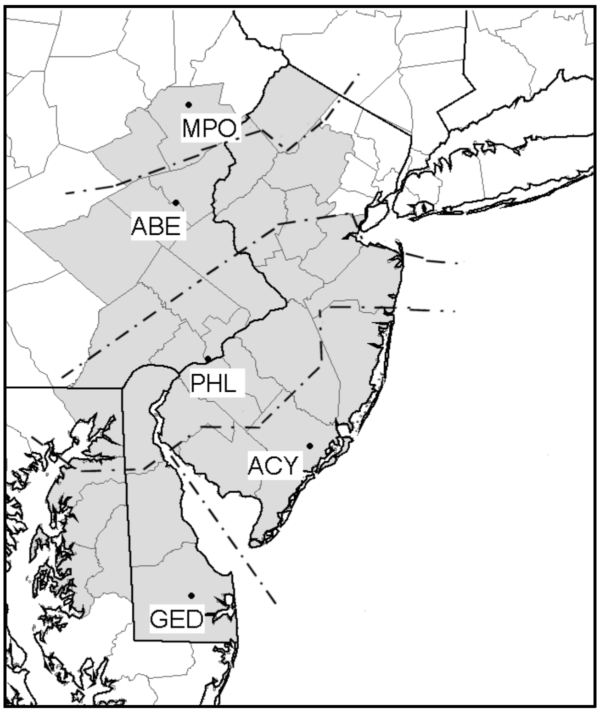

Fig 1. Five sites in the WFO PHI forecast area (shaded region) used for this study. The dot-dash lines indicate approximate areas for snowfall reports used to verify each site.

different Gamma distributions were determined, for mid-range forecast snow amounts of 0 to 7 inches at 0.5 inch increments, and for 8 to 18 inches at 1 inch increments. Each distribution was then converted to a cumulative distribution, as illustrated in Fig. 2b. For any forecast mid-range value, the probability of getting *at least* a given snow amount is then actually 1.0 minus the cumulative probability up to that amount.

Once the theoretical Gamma distributions were chosen, they were compared to the observed distribution of snow amounts for each mid-range forecast value. Again, the observed data came from the six winters beginning with 1995-96. Using Chi-square analysis on class intervals of 0-9 percent, 10-19 percent, etc., we found some significant differences between the Gamma-based cumulative probabilities and the observed cumulative frequencies. Specifically, the low forecast probabilities (< 10 percent) were too low, and the high probabilities (> 90 percent) were too high. To correct this, we subjectively modified the distributions, raising the low probabilities by a few percent and similarly lowering the high probabilities. The modified probabilities were then re-tested with chi-square analysis against observed frequencies, and this time no significant differences could be shown. (The p-value for the chi-squared statistic was .5124.)

The end result of this development was the set of probabilities shown in Table 2. Each row represents a mid-range snow amount forecast. Probabilities for non-integer amounts above 7 inches were interpolated from the rows above and below. Each column represents a snow amount threshold, for which the probability of equaling or exceeding it is given. Note that probabilities are rounded to the nearest 5 percent, and very low (high) probabilities are given as “<5" (“>95"). The values in Table 2 are used for the real-time issuance of the PQSF products.

3. USE IN OPERATIONS

Since we developed the snow amount probabilities from forecast amounts tied to the first winter storm watch for any given event, we decided to issue real-time PQSF’s the same way, (i.e., once per forecast snow event, coincident with the first winter storm watch). Also, we felt a probabilistic approach was better suited to the longer lead time and greater uncertainty associated with a watch, compared to a winter storm warning which is issued later as the event becomes imminent. Although the forecasts

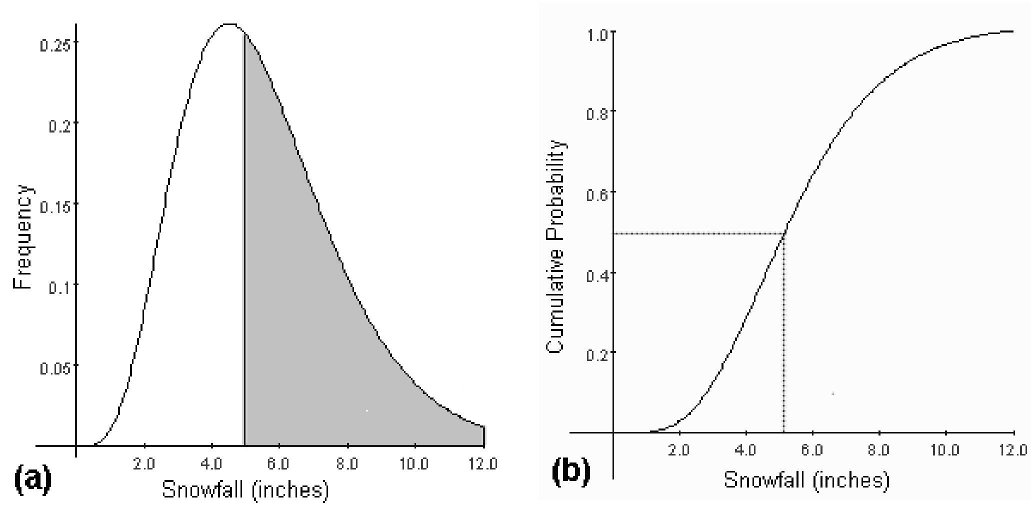

Fig. 2. Illustration of the Gamma probability distribution (a) with the 50th percentile corresponding to an expected 5 inch snowfall; and (b) the same distribution converted to a cumulative probability function.

are issued as public products and are available to anyone, they are primarily intended for more specialized users such as state and local emergency managers.

An example of a real-time PQSF product issued by WFO PHI is shown in Fig 3. The PQSF is issued whenever a watch is issued for at least one zone in the WFO PHI forecast area, and includes probabilities for all 5 forecast points. The product header is "PUBLIC INFORMATION STATEMENT", and it is clearly labeled as "experimental". The product shows snow amount probabilities for the five sites, listed in a table. The data in this table are taken directly from Table 2. Comparing Table 2 with Fig 3, it is seen that in this case the mid-range forecast snow amounts (inches) for MPO, ABE, PHL, ACY and GED were 9.0, 7.0, 1.5, 0.5, and 0.0, respectively. Probabilities are given for all five sites, even if no snow is expected at one or more of the sites (e.g. GED in this case).

These experimental PQSFs have been issued in real time for the past two winter seasons, 2001-02 and 2002-03. The winter of 2001-02 was very mild and dry in the eastern U.S., and a PQSF was issued for only two events. The winter of 2002-03 was much more active, with PQSFs issued for 12 events.

4. VERIFICATION

Although we believe the basis of our PQSF development is sound, it is important to verify any statistically-based forecast on independent data. Ideally, we would like to compare forecast probabilities with observed frequencies at each of the five sites; however, it could take many years to acquire a statistically significant independent sample for a single site. As an alternative, we decided to use snow totals from cooperative observers in the region around each site. Our office typically receives several dozen of these "co-op" reports for any significant winter storm. Each report was assumed to be representative of one of the five sites, according to proximity and/or similar topography. The area associated with each site is shown in Fig 1. This approach is somewhat subjective but it maximizes the number of reports used. Also, we recognize that this method introduces the issue of spatial variability of snowfall, which can sometimes be quite large over just a few counties

Specifically, for each winter storm event and for each site, we counted the number of nearby co-op snowfall reports that equaled or exceeded each of the seven thresholds in Table 2, and then divided by the total number of co-op reports near that site. This gives an observed frequency distribution which can be compared to the forecast probabilities issued in the PQSF text product. Five sites times 14 winter storm events (over two seasons) gives a total of 70 forecast probability/observed frequency distribution pairs for comparison.

Table 2. Adjusted probabilities of equaling or exceeding selected snowfall amounts.

Mid-Range of Forecast (Inches)	Probability of Threshold Snow Amount (percent)						
	>=2in.	>=4in.	>=6in.	>=9in.	>=12in.	>=18in.	>=24in.
0.0	<5	<5	<5	<5	<5	<5	<5
0.5	10	<5	<5	<5	<5	<5	<5
1.0	25	10	<5	<5	<5	<5	<5
1.5	35	15	5	<5	<5	<5	<5
2.0	50	15	5	<5	<5	<5	<5
2.5	60	20	10	<5	<5	<5	<5
3.0	70	30	15	5	<5	<5	<5
3.5	80	40	20	5	<5	<5	<5
4.0	85	50	25	10	5	<5	<5
4.5	85	60	30	10	5	<5	<5
5.0	90	65	35	10	5	<5	<5
5.5	90	75	45	15	5	<5	<5
6.0	95	80	50	20	10	<5	<5
6.5	95	80	55	25	15	<5	<5
7.0	95	85	65	30	15	5	<5
8.0	95	90	75	40	20	5	<5
9.0	95	95	85	50	20	10	<5
10.0	>95	95	90	65	30	10	<5
11.0	>95	95	90	70	40	15	<5
12.0	>95	>95	90	80	50	15	5
13.0	>95	>95	95	85	60	20	5
14.0	>95	>95	95	90	65	25	5
15.0	>95	>95	95	90	75	30	10
16.0	>95	>95	>95	90	85	35	10
17.0	>95	>95	>95	95	90	45	15
18.0	>95	>95	>95	95	90	50	20

Fig. 4 shows the distribution of forecast probability minus observed frequency for selected snow amount thresholds. These are combined values for all sites and all storms. It can be seen that in the vast majority of cases the difference was within +/- 25 percent. Larger differences (i.e., < -0.25 or > 0.25), tend to occur on the positive side for the lower thresholds and on the negative side for larger thresholds. This suggests a slight tendency to over-forecast the lower amounts and under-forecast the higher amounts (except 24 inches). The latter is due primarily to large observed snow amounts from the Feb 16-17, 2003 snow storm. In that event, there were a great many reports of snowfall between 12 and 24 inches, but very few greater than 24 inches. The expected snowfall range with the first watch issuance was around 6 to 10 inches.

Table 3 shows additional verification results, averaged for all 5 sites and all 14 storms. The mean absolute error is generally around 25 percent or less, with greater errors for the lower thresholds. The bias numbers suggest the same pattern described above, i.e. high bias for low amounts and low bias for higher amounts (except 24 in.). The low p-value (right-most column) indicates that the bias for the 2 inch threshold is significant. The zero p-value for bias at the 24-inch threshold appears significant, but is likely just the result of a small but consistently non-zero forecast probability compared to a near-zero observed frequency. Although not statistically significant, the bias for the other higher thresholds does suggest a reluctance to forecast snowfall amounts greater than 12 inches at the early watch stage. Overall, these results suggest that the numbers in Table 2 may benefit from some minor adjustments, especially for the 2-inch threshold.

5. DISCUSSION AND CONCLUSIONS

In this paper we have described an experimental effort to provide NWS customers with more quantitative information about the range of possible snowfall amounts from a forecast winter storm. We have developed a technique based on statistics and climatology that allows operational forecasters to provide, with minimal extra effort, a quantitative estimate of the uncertainty of the expected snow totals. This technique has been applied in real time for snow storms over the past two winter seasons, and has met with some favorable response from emergency managers. Further outreach and familiarization will likely bring about increased use of the product, especially since we can demonstrate some success over the past two winters. We will continue to verify the forecast probabilities over coming winters, and the results from subsequent winters will be used to refine the forecast probabilities.

6. REFERENCES

Applequist, S., G. E. Gahrs, R. L. Pfeffer, and X.-F. Niu, 2002: Comparison of methodologies for probabilistic quantitative precipitation forecasting. Wea. Forecasting., 17, 783–799.

Niziol, T., 2003: Probabilistic Quantitative Snowfall Forecasts. [Available on-line at: http://www.erh.noaa.gov/buf/SpotLES/qpsf1.htm].

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 467 pp.

NOUS41 KPHI 241550
PNSPHI

PUBLIC INFORMATION STATEMENT - EXPERIMENTAL SNOWFALL PROBABILITIES
NATIONAL WEATHER SERVICE MOUNT HOLLY NJ
1050 AM EST TUE DEC 24 2002

...THIS IS AN EXPERIMENTAL PRODUCT FOR WINTER 2002-2003...

THE PROBABILITIES GIVEN BELOW ARE BASED ON A COMBINATION OF STATISTICAL THEORY, OUR SNOWFALL FORECASTS OVER THE PAST FEW WINTERS, AND THE MOST LIKELY SNOWFALL TOTALS FOR THIS POTENTIAL STORM. THE PERIOD OF EXPECTED SNOWFALL IS LATE TONIGHT THROUGH CHRISTMAS DAY.

PROBABILITY (PERCENT) OF REACHING OR EXCEEDING THE FOLLOWING STORM-TOTAL SNOWFALL AMOUNTS (INCHES)

STATION	2IN	4IN	6IN	9IN	12IN	18IN	24IN
MPO	95	95	85	50	20	5	<5
ABE	95	85	65	30	15	5	<5
PHL	35	15	5	<5	<5	<5	<5
ACY	10	<5	<5	<5	<5	<5	<5
GED	<5	<5	<5	<5	<5	<5	<5

PROBABILITIES ARE ROUNDED TO THE NEAREST 5 PERCENT.

STATION IDENTIFIERS:

MPO - MOUNT POCONO, PA (MOUNT POCONO AIRPORT)
ABE - ALLENTOWN, PA (LEHIGH VALLEY INTERNATIONAL AIRPORT)
PHL - PHILADELPHIA, PA (PHILADELPHIA INTERNATIONAL AIRPORT)
ACY - POMONA, NJ (ATLANTIC CITY INTERNATIONAL AIRPORT)
GED - GEORGETOWN, DE (SUSSEX COUNTY AIRPORT)

MORE INFORMATION ON THIS EXPERIMENTAL PRODUCT CAN BE FOUND ON THE NWS MOUNT HOLLY INTERNET HOMEPAGE AT WWW.ERH.NOAA.GOV/ER/PHI.

$$

Fig 3. Example of an experimental probabilistic snow forecast issued in real time.

Table 3. Verification results, combined for all 5 site and all 14 winter storm cases.

Snow Amount Threshold (in)	Forecast Probability	Observed Frequency	Mean Abs. Error	Bias	p(Bias)
>= 2	0.576	0.464	0.234	0.112	0.0073
>= 4	0.403	0.329	0.287	0.074	0.1226
>= 6	0.268	0.205	0.254	0.063	0.1294
>= 9	0.124	0.092	0.124	0.032	0.2039
>= 12	0.062	0.081	0.101	-0.019	0.4602
>= 18	0.027	0.045	0.062	-0.018	0.3172
>= 24	0.020	0.003	0.021	0.017	0.0000

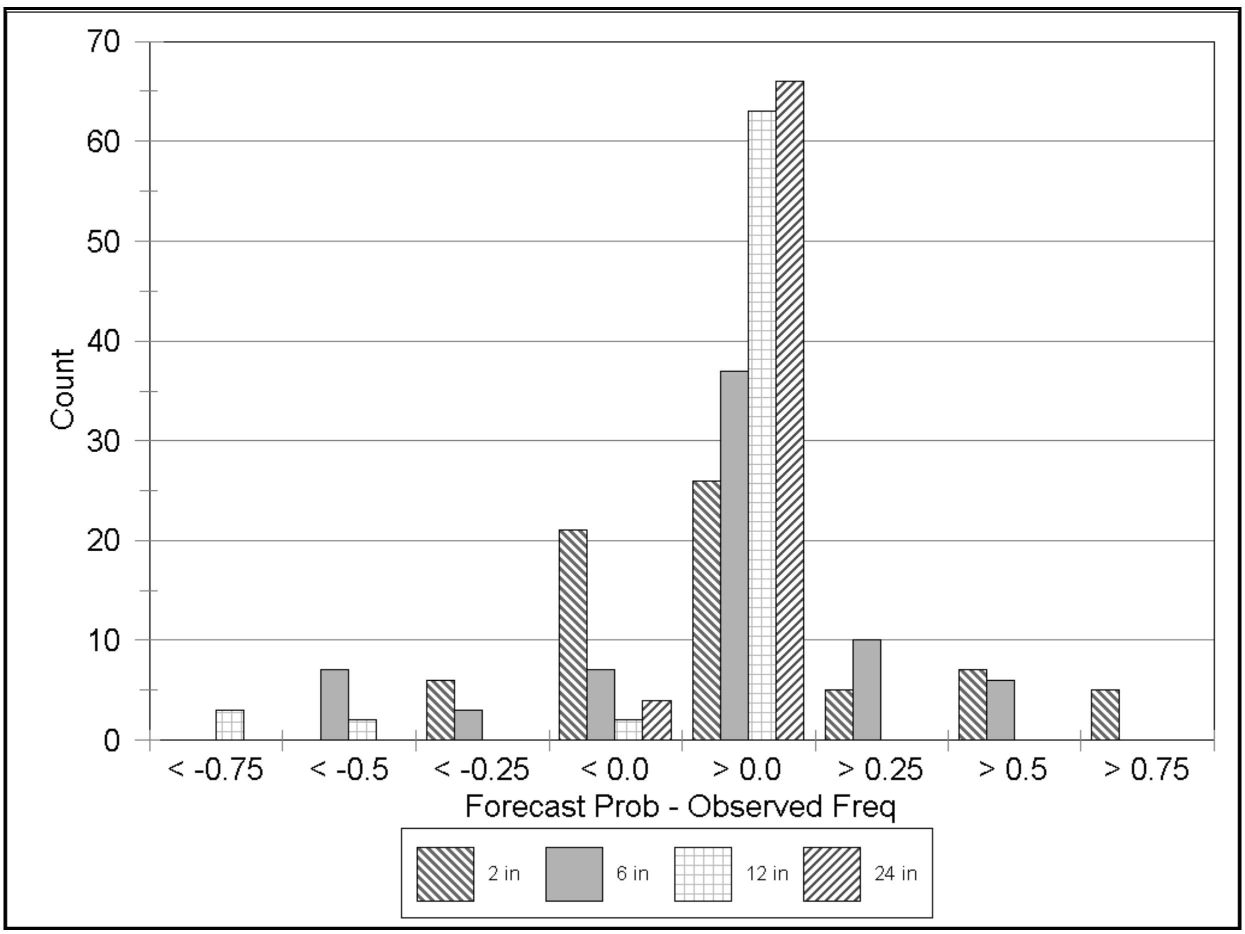

Fig 4. Comparison of forecast probability vs observed frequency for selected snow amount thresholds. Values are averages over all 5 sites and all 14 storms.