P4.3

Are snow accumulation forecasts generally overdone?

Bruce Rose*, Joseph P. Koval and Eric Floehr

I. Introduction

As a leading provider of media-based weather information, The Weather Channel is acutely aware that snowfall forecasts of amount, duration, and severity are closely scrutinized by consumers - as are snow reports or observations of snowfall. Part of this interest is practical (see for example a recent study of snowfall reports from ski resorts: Zinman and Zitzewitz, 2009). But more simply, the casual weather consumer finds snowstorms and winter weather captivating, and snowfall amount is a simple way to categorize or assess the predicted and observed intensity, severity or impact of individual winter storms.

Amongst weather forecasters in the public and commercial domains, there is anecdotal evidence that short-term weather predictions (6-48 hours ahead) of snowfall accumulation are generally overdone or over-predicted in amount and incidence. While there may be some precautionary value in providing worse-case scenarios in winter storm prediction, a persistent or large bias of day to day snowfall prediction would seem less than optimal in serving typical consumers of weather information. Consistently biased or dire snowfall predictions can lead to the same negative consequences associated with other forecasts that exhibit high false-alarms and/or low confirmation bias (see for example, Barnes et. al, 2006); thus, the perceived problem seems worthy of closer study.

a. A quick survey

We asked a number of forecasters at The Weather Channel and elsewhere this same question and received general agreement that snowfall forecasts in the U.S. domain likely are overdone or over-predicted. Some noteworthy comments in response to this question include:

"…I would say yes, simply because most forecasters would rather go on the high side and be wrong instead of the low side....if there is a human component. If we're simply talking models, yes, they're overdone."

"…Basically when it comes to snowfall forecasting we are dependent on the one thing the models and we as human forecasters do the worst on, which is qpf forecasting (.. [I'm somewhat relieved that] people pay far less/if any attention to how much rain we forecast versus how much snow we forecast)."

"… I feel the most critical part of snowfall forecasts don't have anything to do with the actual amount that falls in most situations under 8". My focus as a forecaster is on timing and positioning of the most intense snowfall."

Based on this same expert feedback, there is also sentiment that biased predictions are made over the entire winter season, and occur irrespective of snowfall intensity or amount, geographic region, or forecast provider. Yet there is little or no published documentation of this phenomenon to be found in the literature.

b. Some existing data

The "Northeast Weather Snowfall Forecasting Contest" at http://www.newx-forecasts.com/ (Jim Keller, personal correspondence), a longstanding snowfall forecasting contest, is now in its 11th year. The contest is well organized and well thought out, tracking human-produced snowfall forecasts against observations for 27 locations in the Eastern U.S. We looked at one year of these contest results to see if a pattern of "over-forecasting" was evident in their recorded data.

For the 2008-2009 winter season, there were 8 distinct snowstorms tracked (see **Table 1.1**). If one sums all predicted snow amounts and compares it to the sum of all observed snow amounts, there is little bias (1.03). However, if one organizes the individual storms according to their intensity, an interesting pattern emerges in this data set. We determined an intensity metric called **"Snow per City"** by dividing the total observed snowfall by the number

Storm Number	Fore-casters	Cities with snow	Observed Amount	Predicted Amount	Median	Min	Max	Bias	Snow per City
1	12	14	26.7	42.2	40.0	27.0	75.0	1.58	1.91
2	18	18	119.9	109.9	111.8	54.7	149.1	0.92	6.66
3	11	20	97.7	98.5	95.0	79.0	131.0	1.01	4.89
4	16	14	29.4	46.0	42.8	28.6	78.5	1.57	2.10
5	19	20	58.1	82.5	85.0	57.3	102.0	1.42	2.91
6	13	22	83.3	58.2	53.7	34.9	122.0	0.70	3.78
7	14	25	98.5	106.6	108.1	73.3	133.6	1.08	3.94
8	12	27	169.9	157.6	173.6	61.8	217.5	0.93	6.29
			683.4	701.5	709.9	416.5	1008.7	1.03	4.06

Table 1.1 – Northeast Snowfall Contest results for 2008-2009 season. There were 8 storms with a range of 12-19 forecasters participating in each storm forecast. Cities reporting at least a trace of snow are included in the total observed snowfall. The bias is simply the total forecasted snowfall divided by the total observed snowfall for each storm. Snow per city is the total observed amount divided by cities with snowfall.

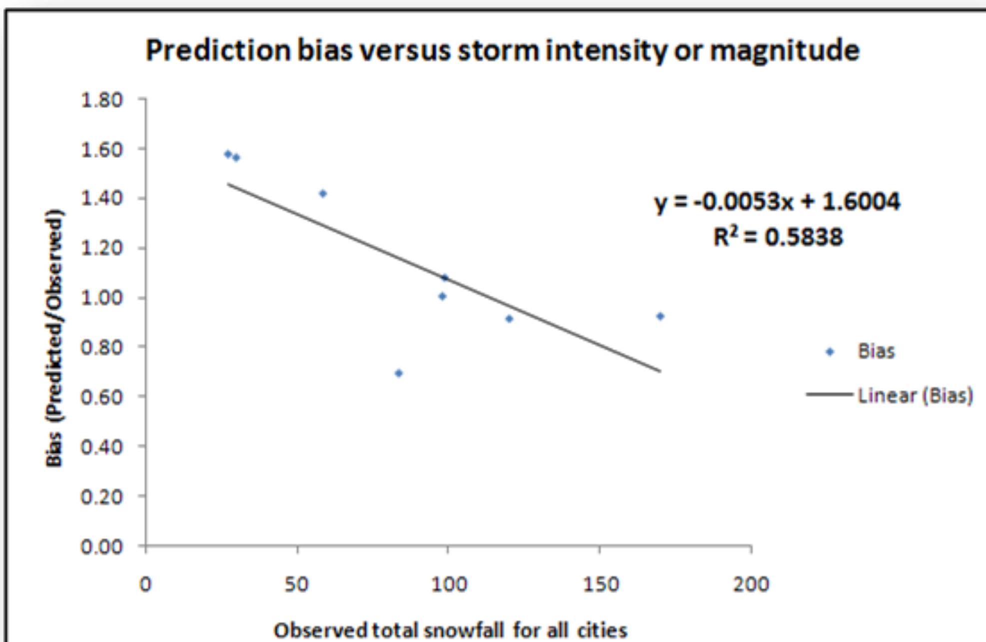

Figure 1.1 – Snow bias for 8 storms predicted by human forecasters at NEWx Snowfall Contest versus the total snowfall occurring for those 8 storms at 27 locations in the Mid-Atlantic and Northeast US. The trend-line is a linear best-fit and the equation of that line is also shown with R^2 statistic.

of cities reporting snow. We then plotted the individual storm snowfall bias against the storm bias. **Figure 1.1** shows this relationship with a best-fit linear trend line. The linear function suggests an inverse relationship between the intensity of an individual snowstorm or snow event with the bias or the tendency for over-prediction of snowfall (R^2 approaching 0.60). More simply, lighter snowfalls are over-predicted while heavier snowfalls are under-predicted.

From this dataset, it appears that snowfall forecasts, from one season's worth of snowstorms (8) predicted by 15-20 amateur forecasters, are overall unbiased. This is counter to the experts' expectations - but with some crucial caveats. With this in mind, we are conducting a more ambitious analysis to see if this preliminary finding will hold true for a larger set of data.

c. Digital forecast verification for the winter of 2009-2010

For the period 01-NOV-2009 through 31-MAR-2010 we are collecting "digital" snowfall forecasts for locations in the Continental United States (CONUS) covering 24-hour forecast periods. In this context, the term "digital" implies a deterministic quantity expressed as an explicit amount (e.g., 4.2" of snowfall), versus a range (e.g., 3-5" of snowfall) or a qualitative description (e.g., "Heavy snowfall expected").

We choose to compare snowfall forecasts for first-order stations (those providing at least once-per-day manual snowfall amount measurements) across CONUS (about 200 stations) against observations for the upcoming winter season. Each evening during this timeframe we extract snowfall and melted quantitative precipitation (QPF) forecasts through the upcoming forecast valid period of 06Z to 06Z from The Weather Channel Global Forecast Center's digital forecast feed (TWC), the National Weather Service's National Digital Forecast Database (NDFD), and the National Center for Environmental Prediction (NCEP) operational NAM/WRF forecasting model (NAM). About 0030Z, these three realtime forecasts are digitally captured and published on a publically hosted web site by one of our authors, Eric Floehr (of Intellovations, LLC website: http://snow.forecastwatch.com). Later the next day (1630Z), the station observations or verifications are amended to that same published record once ground truth can be collected, quality-controlled, and compared to the predictions. **Figure 1.2** is a screenshot of the ForecastWatch web site devoted to this data tracking. Both snowfall and rainfall (QPF) predictions are tracked and simple daily comparisons are made to observations for about 200 points.

This seasonal study should present an interesting comparison of basic skill in snowfall (and QPF) prediction amongst these three broadly available weather forecasts in the public domain, as well as lending objective insight to support or contradict the assertion that these forecasts (TWC, NDFD and NAM) tend to overpredict snowfall. Because of the timing of this contemporary work, only partial results can be presented. In Section II, the methodology of data collection and transformation is thoroughly described. In Section III, preliminary results for the first 57 days of the study period are discussed. Section IV is a brief wrap-up and summary of remaining work and future directions.

II. Methodology

Our methodology is to compare snowfall (and quantitative precipitation or QPF) for NDFD, TWC, and NAM for a 24-hour digital snowfall prediction between 06Z and 06Z of the following day. The forecasts are collected about 0030Z each evening from each provider and are ultimately compared to human observations of snowfall taken between midnight to midnight local time (more on this later). Thus, in the case of NDFD and TWC there is about 6 hours lead time between forecast and verification period and about 12 hours in the case of NAM/WRF (as the model grids employed are based on the operational 18Z run of the NAM).

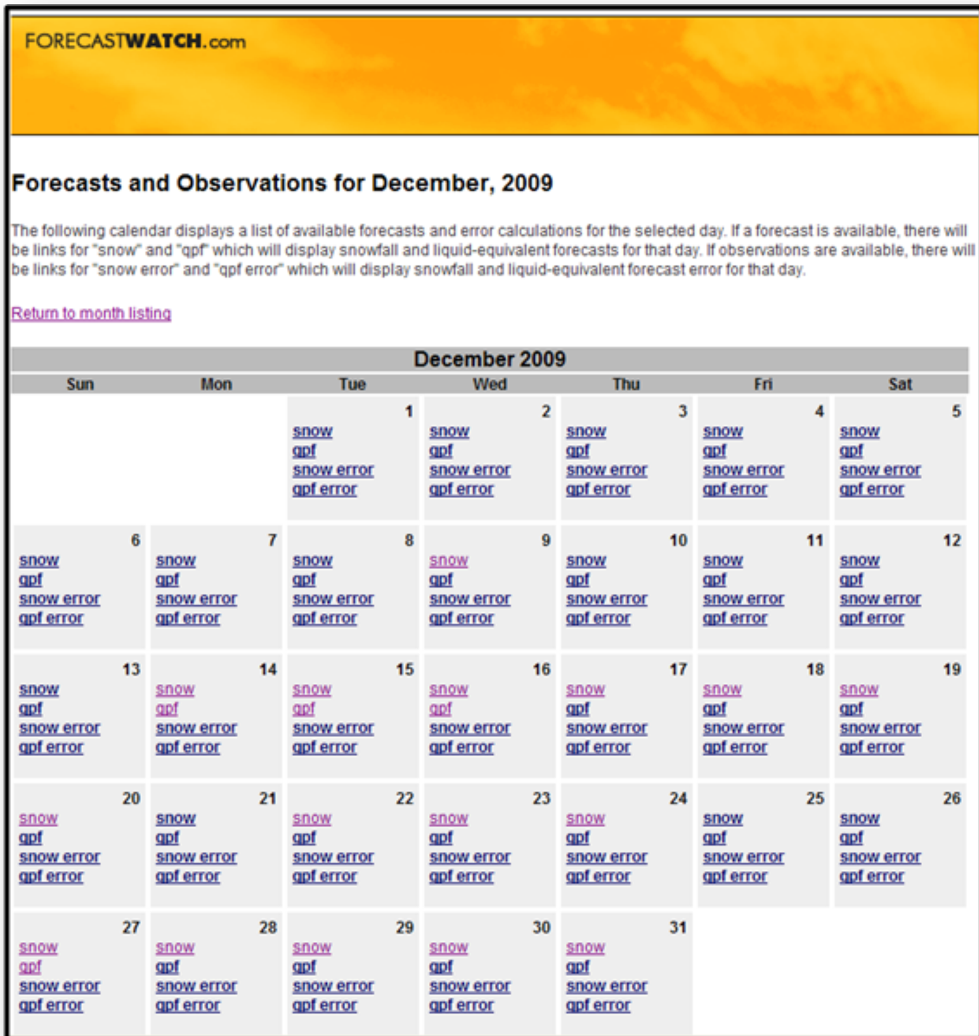

Figure 1.2 – http://snow.forecastwatch.com website showing calendar view of recorded snowfall and QPF predictions, observations, and errors for the month of December, 2009.

Digital snowfall information (in 1/10ths of an inch) is typically not directly consumed by users. More typically, snowfall amounts are aggregated and expressed as a possible accumulation range in worded forecasts or other forecast products. However, digital snowfall forecasts are ideal for verification studies as it removes vagueness or intent of the forecaster, allows for direct comparisons with explicit snowfall measurements, and allows for unambiguous comparison amongst different forecast providers.

a. NDFD data collection

Digital forecasts from NWS NDFD (see for example, Glahn and Ruth, 2003) are extracted for the 200 points studied at 0030Z. Two-dimensional bilinear interpolation is used to estimate the predicted 24-hour snowfall from 06Z to 06Z at the appropriate latitude and longitude of each forecast point.

The NDFD produces a 36-hour digital snowfall forecast on a 5-km grid that is subdivided into 6-hour intervals. The predicted 6-hour snowfall has single decimal precision and is ultimately converted to units of inches of snowfall. Sample extractions from the NDFD grids were compared to NWS forecast products containing either digital snowfall amount or snowfall ranges and showed excellent agreement, so it is thought that little if any error is introduced via this interpolated extraction method.

The NDFD 36-hour snowfall forecasts at 6-hour intervals show good fidelity and availability. In particular, NWS snowfall forecasters seem to take particular care with digital predictions of very light snowfall amounts, as many 24-hour forecasts of a few tenths of an inch or less are routinely found in the digital results and often match well against observed results. One particular forecast from the Marquette Forecast Office (WFO) predicted a 24-hour snowfall of 21.8” on 9-DEC-2009 that was subsequently verified against an actual measured snowfall of 21.6” in the same 24-hour period. A remarkable snowfall forecast of an extreme event.

One peculiarity noticed with the NDFD data is that old 36-hour snowfall forecasts may persist in the database at times. An example of this behavior is the station KPHL for the days 19-DEC-2009 and 20-DEC-2009. Heavy snowfall is predicted for two consecutive days despite a synoptic pattern that suggested strongly that only a single day event was likely to occur. The cause of this is unclear but it appears to be an information processing issue, rather than a forecasting issue. It may also relate to the time of extraction in this study being so close to the 00Z day boundary.

b. TWC data collection

TWC forecasts are transmitted in digital form for about 17,000 points worldwide in XML format to businesses and other partners (see: http://www.weather.com/services/oap.html?from=footer) and can be made available to research and academic interests in XML, FTP, or XOAP formats or data streams.

The point forecasts include hourly snowfall in single decimal precision for 60 hours into the future. The hourly point forecasts for each of the 200 points between 06Z and 06Z were summed and the results were used to represent the 24-hour snowfall forecast at each point.

c. NAM data collection

In the case of the NAM/WRF model, 24-hour snowfall amount can only be inferred from the model output. This must be done in order to directly compare model results to the National Digital Forecast Database (NDFD) and TWC snow amounts.

This is accomplished by finding a snow to melt-water ratio (snow ratio) to convert a NAM melted snowfall amount to simple snowfall. In our case, we base the snow ratio on 2-m temperature only, using a derived empirical expression of a 3rd order polynomial that was developed experimentally:

$$S_r = -8.268 \times 10^{-4}\, T_{2m}^{3} + 5.502 \times 10^{-2}\, T_{2m}^{2} - 1.714\, T_{2m} + 32.389 \qquad \textbf{Eq. 2.1}$$

Where S_r is a unitless quantity called snow ratio, and T_{2m} is 2-meter temperature in ^{0}F. When snowfall is indicated in the model output, the snow ratio S_r is limited to no smaller than a 5:1 ratio.

Procedurally, each NAM 3-hour melted snowfall precipitation amount is found in a gridded output field called WEASD (see: http://www.emc.ncep.noaa.gov/mmb/rreanl/faq.html for an explanation). For each 3-hour portion of the 24-hour period from 06Z to 06Z, T_{2m} at the end of the 3-hour interval is used as the basis to calculate a 3-hour snow ratio for those periods where snowfall is indicated. An aggregate melted snowfall amount is found by summing 3-hour intervals where melted snowfall amount is non-zero and multiplying this amount by an aggregate snow ratio which is a simple average of those snow ratios calculated for each period where snowfall is indicated. The result yields an estimate of 24-hour snowfall and this quantity is expressed in inches.

The added step of calculating a representative snow ratio for each NAM extracted snowfall amount can certainly add error or uncertainty into the snowfall estimates and this should be noted when comparing directly to TWC or NDFD forecasts where this step is implicitly handled by the forecaster or via a more sophisticated algorithm (e.g. Cox et. al. 2009).

d. Observations

While attempting to build objective MOS forecasts of snowfall amount, Cosgrove and Sfanos (2007) commented that: "Measuring snowfall could be considered a matter of personal style." There is little uniformity in this area and the difficulties measuring and reporting snowfall accumulation are well known and documented (see Carrie Olheiser's presentation at:

http://www.nohrsc.noaa.gov/technology/pdf/Cold_Region_Workshop_CO_20041117.pdf

for an excellent review of the issues). To try and minimize uncertainties in the ground-truth portion of this study we limited snowfall measurements to the Daily Climo points distributed by NWS WFOs under the bulletin heading CDUS41 through CDUS43.

These are first order reporting stations where manual snowfall accumulation measurements are made for the purpose of climatological recordkeeping at least once daily around midnight local time. Within cold climates of the CONUS this amounts to about 200 stations. **Figure 2.2** shows the distribution of these points north of 32N degrees latitude.

The most significant shortcoming of the CDUS41 data is its adherence to midnight to midnight local time conventions. The TWC digital snowfall forecast could be matched for all time zones to midnight through midnight time intervals since it is hourly resolution. However, this was not possible with NDFD or the operational NAM output because of their 6-hour and 3-hour temporal resolution tied to UTC boundaries. Using NDFD as a least common denominator, we determined a universal forecast collection time and forecast interval of 06Z to 06Z for all providers and accepted the mismatch in time between forecast and observation.
This mismatch is as much as 2 hours. It is largest in the Pacific Time zone (where there are few points that report snowfall). The Central time zone matches exactly while Eastern and Mountain Time zones mismatch by one hour. We estimate this overall mismatch between observation and forecast time interval accounts for about 2-5% error or uncertainty in the results, depending on many factors such as the regional distribution of snowfall events. **Figure 2.3** shows the temporal overlap of forecast and observation schematically.

III. Results

For this mid-winter update, we intended to collect all forecasts and observations for the period 1-NOV-2009 through 3-JAN-2010 (64 days). But due to a data outage, All NDFD, NAM, and TWC forecast data were

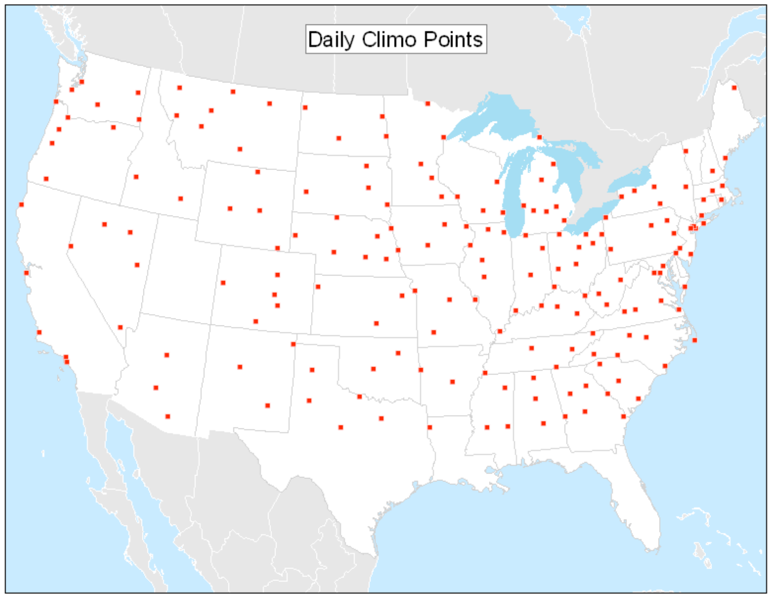

Figure 2.2 – Daily climatological stations managed by NWS Weather Forecast Offices (WFOs) north of 32N latitude in the CONUS.

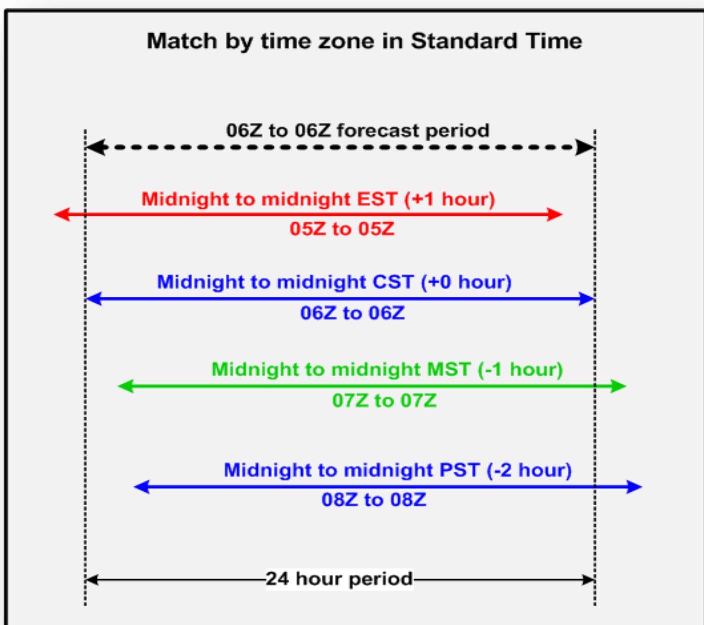

Figure 2.3 – Schematic showing the forecast time interval 06Z to 06Z against the time interval spanning midnight to midnight in Universal Time Coordinate for all time zones (Standard Time) in the CONUS.

collected from 8-NOV through 3-JAN (57 days). The CDUS41 bulletins are quite reliable although about 1-2% are either missing, improperly formatted, or contain miscoded information. We corrected any of these Daily Climo reports that contained obvious typos or formatting errors, and others were simply discarded. This amounts to only a few cases over the two month period.

We therefore stored about 11,000 forecasts and observations. Of these events, 3,329 (30%) contain at least one provider snowfall forecast of at least 0.1", or at least a trace of snowfall observed for the midnight to midnight period for that closely corresponds to that same forecast period. There were 1,328 (11.5 %) cases of observed snowfall of 0.1" or greater, and there were fully 1,110 cases (about 9.5% of total cases) of trace amounts observed. **Figure 3.1** shows a time series of snowfall frequency over the preliminary study period that includes any observed event of at least 0.1" of snowfall in the midnight to midnight observing period. Notable storms occurring so far this winter are annotated along the time series. On an active day as many as 65 of the 200 reporting stations experience at least 0.1" amounts of snowfall which is a sizable proportion.

Because they account for so many of the snowfall events, the treatment of 24-hour trace amounts of snowfall is crucial to any analysis of accuracy or bias. One must decide to allow such reports as confirmation of snowfall occurring or decide to disregard them since they are not considered "measurable" precipitation. In this preliminary analysis we look at the gross skill and bias of the entire time series with and without trace events. When trace events are considered they are assigned a digital accumulation amount of 0.05". However, for finer scale or categorical statistics we treat the Trace events as 0.0" snowfall observed.

a. All snow events

Table 3.1 shows statistics for all events where snowfall was predicted by one of the providers or at least a trace of snowfall was observed. The right-hand side of **Table 3.1** treats Trace amounts as 0.0" snowfall and thus all frequency counts including overall N are changed. In this case, 387 of the 1,110 trace reports are dropped from the second analysis (that is, there are 387 cases where a trace of snowfall is reported and no forecast provider predicts snowfall).

Overall, more snow is predicted than observed. Expressed as a simple ratio (as in Section I) NDFD has the largest bias and TWC the smallest. Threat scores are nearly indistinguishable between NDFD and TWC if trace amounts are treated as measurable snowfall (left-hand side of table), but NDFD lags both TWC and model results in Threat and Equitable Threat Scores if Trace amounts are treated as 0.0" snowfalls.

This result may have more to do with the mechanics of forecast preparation than a measure of forecast skill or accuracy. For example, the TWC forecast preparation allows for weather forecasts of snowfall such as "scattered snow-showers" and "snow flurries" that are also associated with nil quantitative precipitation. This means that in TWC's forecast paradigm, nuisance or light snowfall events with small consequence can be predicted at the same time a QPF forecast of 0.00" is predicted. This may not be the case with NDFD or may be procedurally discouraged with NDFD, and would explain the large swing in objective skill and bias when the two treatments of trace amounts are analyzed as described in **Table 3.1.**

On one hand, it is illogical to predict precipitating weather without predicting an amount of precipitation; but it can also be argued that in wintertime, this is a very common outcome, and may in fact be the best way to capture and describe common wintertime sensible weather. In either case, this behavior in the lower limits of snow amount forecasting highlights a general weakness in the consistency and value of deterministic and digital forecasts that is currently under some debate (e.g., Mass 2003).

b. Bias by range of snowfall amount

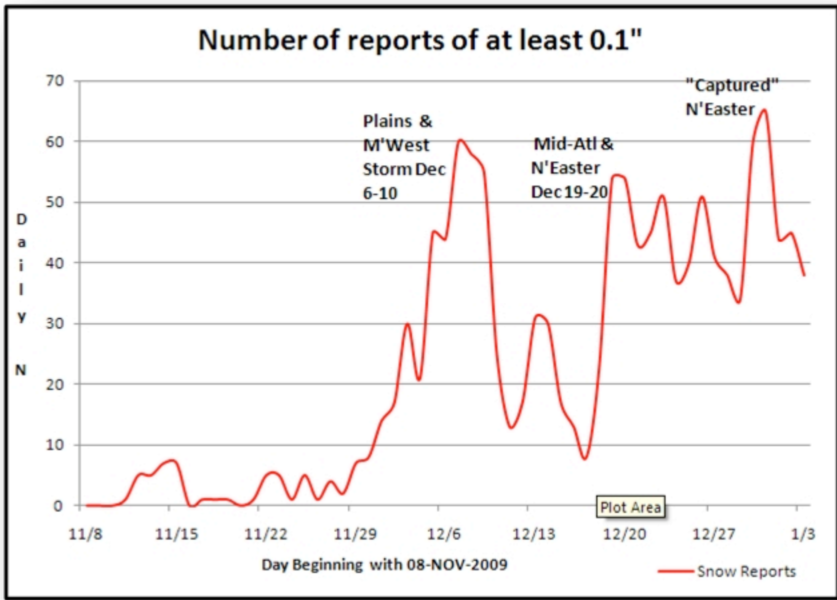

Figure 3.1 – Daily frequency of non-zero snowfall reports for the period 8-NOV-2009 through 3-JAN-2010 from the domain of 200 Daily Climo points used in this study.

	Treating Trace amounts as 0.05" reports				Treating Trace amounts as 0.0" reports		
	NDFD	TWC	NAM		NDFD	TWC	NAM
N	3329	3329	3329		2942	2942	2942
Total snowfall predicted	3298.3	2620.3	2725.8		2792.8	2114.8	2220.3
Total snowfall observed	2475.5	2475.5	2475.5		1970.0	1970.0	1970.0
Ratio/Bias	1.33	1.06	1.10		1.42	1.07	1.13
Correct Rejects	335	484	743		492	714	1153
Hits	1621	1537	1103		1154	1143	889
Misses	718	802	1236		174	185	439
False Alarms	655	506	247		1122	900	461
Threat Score	0.541	0.540	0.427		0.471	0.513	0.497
Equitable Threat Score	0.466	0.469	0.365		0.413	0.461	0.453
Probability of Detection	0.693	0.657	0.427		0.869	0.861	0.497
Bias (contingency)	0.973	0.873	0.577		1.714	1.538	1.017

Table 3.1 – Overall statistics for the period 8-NOV-2009 through 3-JAN-2010. The data is analyzed twice, treating Trace snowfall reports as 0.05" snowfall tallies and also treating trace reports as no snowfall. Since the event list is composed of only those records where one provider or the observation itself is non-zero, the total cases, N, changes as do all of the contingency counts and other metrics.

We took all snowfall events from the study period where at least 0.1" of snowfall was observed, and then further subdivided these cases into light (0.1 – 2.0" snowfall observed), moderate (2.1 – 6.0" of snowfall observed), and heavy (more than 6.0" snowfall observed) events. **Table 3.2** shows statistics for these snowfall categories.

Probability of detection is high for all events with both NDFD and TWC, and is 2 of 3 events with NAM. The gross ratio or snowfall bias is slightly overdone for NDFD and slightly underdone for TWC and NAM. This bias is highest in the light category, especially for NDFD. TWC under-predicts moderate events (2.1 – 6") 7 of 10 times while NDFD is nearly unbiased. However, in the >6" category of high impact events (N=83) TWC ratio is better than NDFD and NAM and is within range over 60% of the time.

c. Scatter-plots

Figure 3.3 show three scatter-plots of each provider forecast versus observation for any case where observed snowfall was at least 0.1". The fourth scatter-plot is a simple average of the three forecasts versus observed. In each case a linear best-fit is computed for the scatter of points (including correlation statistic) and a second straight line representing perfect correspondence between forecast and observed (dotted line) is shown.

The ideal y-intercept should be close to 0.0"; in reality all providers exhibit a positive y-intercept with TWC closest to an unbiased value. The ideal slope should be unity. In reality, each slope is less than 1.0 which is consistent with a scatter where light amounts are over-forecast and heavy amounts are under-forecast. TWC slope comes closest to 1.0 (0.752) while NAM is shallowest at 0.605. The R^2 correlations range from around 0.45 (NDFD and NAM) to 0.60 (TWC and the simple average or composite forecast) which indicates overall reliability in the forecast that may or not include a bias or a differential bias across snowfall categories. In short, a high R^2 occurs in these data when the forecast systematically tracks with the observations in a linear manner – not necessarily when the forecasts are identical to the observed snowfall.

d. Frequency bias

Another simple test of snowfall forecasting bias is performed using simple counts of each provider's forecasts within the light, moderate, and heavy snow categories. These counts are compared to the actual counts over the study period to date. A ratio well over 1.00 would indicate that the forecast provider tends to over-predict snowfall in that category. A ratio well below 1.00 would suggest under-prediction. **Table 3.3** shows these counts and ratios between 8-NOV and 3-JAN.

NDFD demonstrates the snowiest forecasts. TWC forecasts are mostly unbiased in the moderate and heavy categories. NAM forecasts are somewhat overdone for heavy and moderate events but show good fidelity with light events – the only case amongst the three where light events are not significantly over-predicted in terms of simple frequency counts. However, it is important to keep in mind that Trace amounts cannot be forecast by any provider, and trace amounts are not considered in the Observation counts. If Trace amounts are considered, the large prediction biases for the light snowfalls become unbiased or under-predicted in all cases.

IV. Summary

Much like the simple analysis of the NEWx Snowfall Contest, the general assertion that snowfall forecasts are over-predicted cannot be confirmed. There is a marked tendency to over-predict light events, but this trend is reversed with moderate and heavy snowfalls. This preliminary set of 57 days worth of data would suggest that measureable snowfalls are often predicted, but may verify less frequently. However, when moderate to heavy snowfall is predicted, measurable snowfall generally does occur and often in amounts greater than that predicted.

Notwithstanding the uncertainties introduced by the transformation of NAM output to actual snowfall amount, the model snowfall estimates are much different from the NDFD and TWC predictions. This suggests that either forecaster guidance is an amalgam of model inputs (such that the NAM snowfall signal is obscured), or that forecasters edit and change model guidance to a large degree.

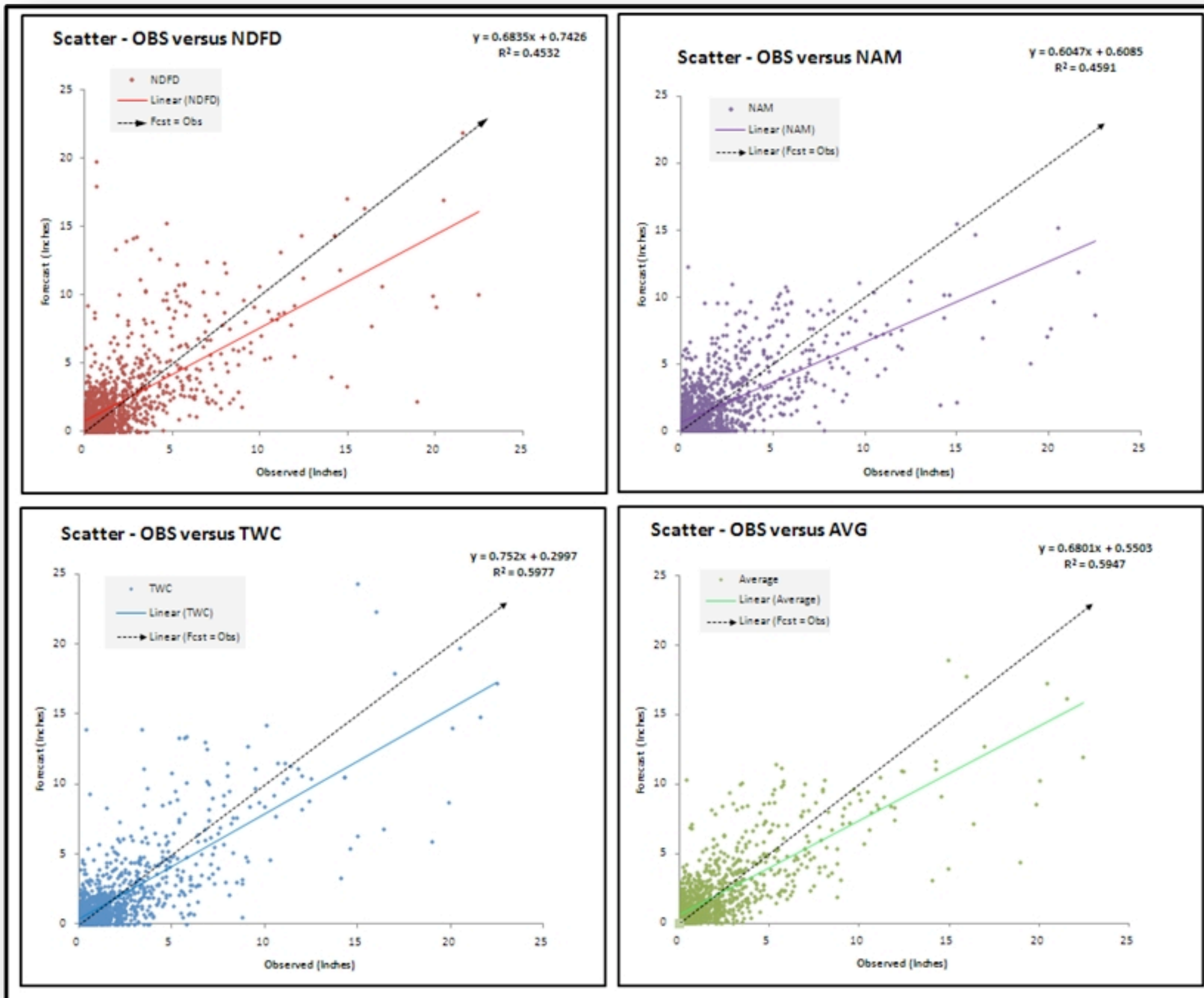

Figure 3.3 – Scatterplot of Observed snowfall versus NDFD predictions for all cases where observed snowfall is at least 0.1" for NDFD, TWC, NAM and a simple average of all three providers. The linear trend-line is a best-fit regression equation shown in the upper right-side of each graph. The dotted line is the plot of x=y which would demonstrate perfect correspondence between forecast and observation.

The NAM output tends to be more objective and more conservative in heavy snowfall cases, but less skillful overall compared with the human-intervened forecasts of NDFD and TWC. There appears to be a tendency for human forecasters to add false alarms or bias but to sharpen and improve forecast skill, most probably through decreasing phase and amplitude errors in the 24-hr forecasts.

Melted QPF is also recorded and tracked in this winter study. In this mid-term assessment we did not analyze this QPF data. Partly, this is because it does not help much in answering our central question of snowfall over-prediction. But also we are concerned that melted precipitation observations from the first-order stations may be systematically underreported due to cold weather aspects of gage reporting.

	All Events at least .1" Observed			0.1" - 2.0" Observed			2.1" - 6.0" Observed			> 6.0" Observed		
	NDFD	TWC	NAM	NDFD	TWC	NAM	NDFD	TWC	NAM	NDFD	TWC	NAM
N	1328	1328	1328	975	975	975	270	270	270	83	83	83
Total snowfall predicted	2643.7	2221.4	2274.4	1097.5	786.1	890.4	959.2	772.8	861.0	587.0	662.5	523.0
Total snowfall observed	2424.9	2424.9	2424.9	662.8	662.8	662.8	932.0	932.0	932.0	830.1	830.1	830.1
Ratio/Bias	1.09	0.92	0.94	1.66	1.19	1.34	1.03	0.83	0.92	0.71	0.80	0.63
Percent overpredicted	55%	44%	39%	61%	50%	42%	42%	30%	40%	23%	28%	12%
Percent underpredicted	45%	56%	61%	39%	50%	58%	58%	70%	60%	77%	72%	88%
Hits	1154	1143	889	814	804	561	257	256	246	83	83	82
Misses	174	185	439	161	171	414	13	14	24	0	0	1
Prob of Detection	87%	86%	67%	83%	82%	58%	95%	95%	91%	100%	100%	99%
Within Range	87%	86%	67%	67%	74%	42%	48%	48%	46%	54%	63%	54%

Table 3.2 – Snowfall events of 0.1" observed or greater. Percent over-predicted or under-predicted is the proportion of forecasts that exceed or fall below the observed amount. "Within Range" means the prediction fell within the predetermined snowfall category.

Snowfall	Frequencies				Ratios		
Category	NDFD	TWC	NAM	OBS	NDFD	TWC	NAM
.1 - 2"	2276	2043	1350	1328	1.71	1.54	1.02
2.1" - 6.0"	462	346	443	353	1.31	0.98	1.25
> 6"	104	81	96	83	1.25	0.98	1.16

Table 3.3 – Raw frequency counts of forecasts or observations within predetermined snowfall categories (right-hand side) and the ratio of forecast counts over observation counts for each category. Trace amounts are not used.

We initially believed that such issues as heating elements, evaporation and catchment of snowfall at automated sensors were known problems at ASOS and AWOS sites, but were not problems at the manual reporting sites appearing in the CDUS41-43 station lists. However, at least one expert in this field pointed out that all wintertime melted precipitation reports are confounded by these issues including the CDUS41-43 points. As a result we need to review the current status and accuracy of wintertime gage reports before proceeding with additional analyses of melted QPF.

Our next steps will be to redo the analysis once the entire winter season is completed. We expect the general findings discussed here to hold for the entire season, except perhaps for the high-impact events in the 6+" category of snowfall. We are most interested in further discussion on the treatment of Trace snowfall amounts since this has such a large impact on the objective results, especially considering that over-prediction of light snowfall events contributes the lion's share of the overall snowfall forecasting bias.

[*Corresponding Author Address: Bruce Rose. The Weather Channel. 300 Interstate North Parkway. Atlanta, GA 30339.* brose@weather.com]

Acknowledgements

Thanks to Mr. Jim Geller of NEWx Snowfall Forecast Contest for helpful discussions surrounding their snowfall forecasting contest, its constituents, and the resulting data.

V. References

Barnes, Lindsey R. et. al., 2007: False alarms and close calls: A conceptual model of warning accuracy. Weather and Forecasting. **Vol. 22**, pp. 1140-1147.

Cosgrove, R. L., and B. Sfanos, 2004: Producing MOS snowfall amount forecasts from cooperative observer reports. Preprints, *20th Conference on Weather Analysis and Forecasting*, Seattle, WA, Amer. Meteor. Soc.

Cox, Andy, J. P. Koval and B. Rose, 2009: A scheme to produce high-resolution gridded snowfall estimates over the conterminous United States. Preprints, *23th Conference on Weather Analysis and Forecasting*, Omaha, NE, Amer. Meteor. Soc.

Mass, Clifford F., 2003: IFPS and the future of the National Weather Service. Weather and Forecasting. Forecasters' Forum. Vol. 18, pp. 75-79.

Ruth, David P. and H. Glahn, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.,* **Vol. 84**, pp. 195-201.

Zinman, Jonathan and E. Zitzewitz, 2009: Wintertime for deceptive advertisers?. Working papers, August 2009.