PEAK PERFORMANCE

NWS Office of Climate, Water, and Weather Services
Silver Spring, Maryland

September 2009

In This Issue

Mean or Median: Which Statistic Best Represents Tornado Warning Lead Time? 1
Weather "bits": Information Technology 4
Assessing the Accuracy of Western Region's Precipitation Forecasts 5
Temperature Forecast Opportunities Across The NWS Southern Region 7
Ask Chuck: TAF Verification 9
Aviation Verification- A Central Region Method 10
Did You know? 11
Administration Wants Performance Management Data Put to Better Use (Federal Times Article) 12
On the Road Again 13
New Verification Tool- Sky Cover 17
Contact Information 20

Mean or Median: *Which Statistic Best Represents Tornado Warning Lead Time?*

By Doug Young and Bob Glahn, NWS Headquarters

Recently, the question was asked why the National Weather Service (NWS) uses the average (arithmetic mean) to characterize tornado warning lead time rather than the median. This simple, yet insightful question initiated a flurry of activity to examine these two statistics more closely.

Both the mean and median are measures of central tendency, and each has its strengths and weaknesses in describing the distribution of data. One forecast verification metric never tells the whole story, nor does one characterization of a metric distribution. In the case of lead time calculations, the mean would be the arithmetic average of all lead times. The median, however, would be the lead time that falls in the exact middle if all scores were placed in sequential order from lowest to highest. As a result, it would not matter if a relatively small number of events had exceptionally long lead times, the score would only reflect the more "typical" lead times.

Unfortunately, tornadoes are very difficult to forecast, and even though there has been significant improvement over the past several years; there are still quite a number of tornadic events occurring without warning or before a warning is issued, and therefore have zero lead times. Due to the nature of the median statistic calculation, over the course of a year the relatively large number of zero minute lead times will always skew the median score to the left of the mean (i.e., the median score will always be lower than the mean). **Figure 1** shows the distribution of tornado warning lead time for fiscal year (FY) 2008. After a substantial peak at zero, the distribution for tornado lead times had a slight peak between about

Continued on next page…

Mean or Median: Which Statistic Best Represents Tornado Warning Lead Time? – Continued from Page 1

10 and 20 minutes, with a slight drop off on either side. A more significant drop in lead time occurred beyond 40 minutes. This distribution appears to be fairly representative of tornado lead times over the United States during the course of a year, which characteristically has a median of a few minutes less than the average.

The FY 2008 mean tornado lead time was 13.9 minutes, while the median was only 11.8 minutes. If the median were used instead of the mean, the 2-minute drop in lead time would be significant in NWS Government Performance and Results Act (GPRA) reporting-perhaps being interpreted as a decrease in accuracy, when in actuality it would only be a change in the characterization of the distribution of the metric. More drastic differences, however, would occur in lead time calculations for other small samples (e.g., a particular outbreak or NWS region). The median lead time may actually be zero, while the mean lead time may be 10 minutes or so. If only one statistic is given, the median of zero minutes would not contain as much, or as useful, information as the average.

In addition to the mean lead time, the NWS computes metrics such as the critical success index, probability of detection, and false alarm ratio to track other various components of skill. Usually there is only one or two GPRA measures used for a particular service area, or weather type. GPRA measures have to be simple and easy to understand, essentially without any explanation needed. An "average" is generally well understood, but more likely than not, the median would cause confusion. Furthermore, once established, it is difficult to change a GPRA measure, especially if it were to indicate less "accuracy."

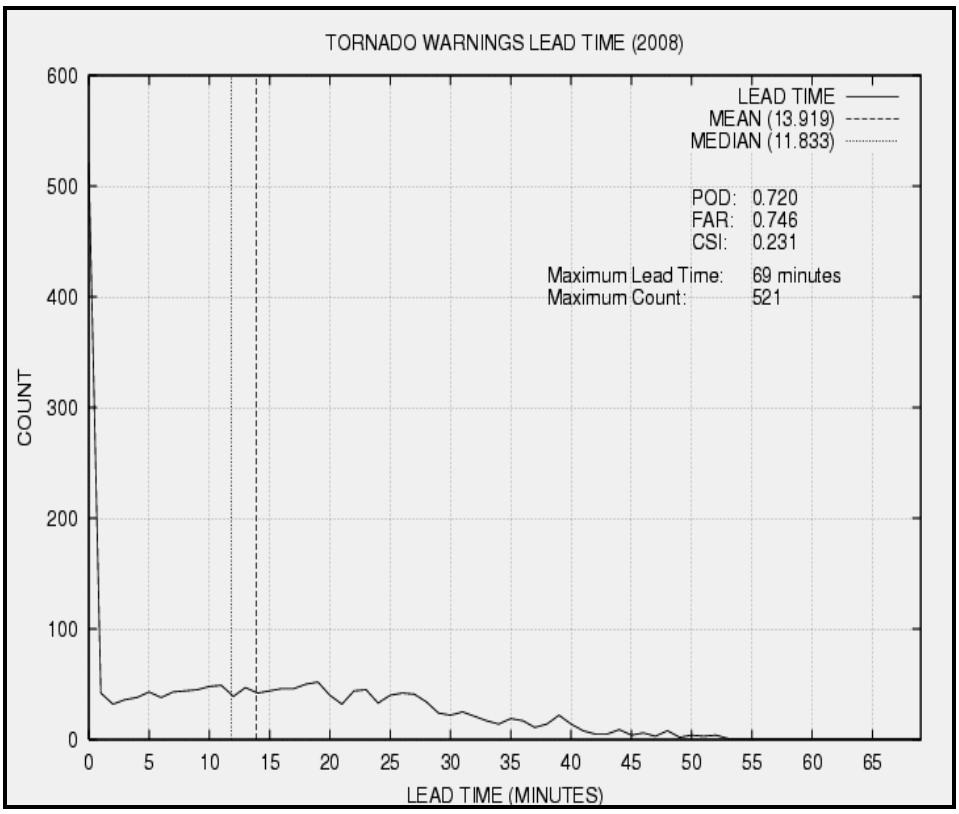


Figure 1: FY 2008 Distribution of Storm-based Tornado Warning Lead Times with mean and median calculations referenced. (Figure courtesy of Scott Scallion, Meteorological Development Laboratory.)

Continued on next page...

Mean or Median: Which Statistic Best Represents Tornado Warning Lead Time? – Continued from Page 2

A final reason for using mean instead of median is the characteristic of greater volatility in the median score until the sample size is sufficiently large. In **Figure 2**, for example, the FY 2009 Tornado Cumulative Median Lead Time score remained below 5 minutes from October through mid February. This low score most likely reflected the difficulty in forecasting sporadic tornado events during the cold season. Scores then jumped in mid February and early April in association with more organized and widespread tornadic activity. However, from the first month of the fiscal year, the mean statistic has painted a more accurate picture of where the final GPRA score will end up in FY 2009. The mean statistic has also remained more stable through the FY.

Overall, neither the mean nor the median fully characterizes the distribution. It is believed, however, that the mean tends to provide a better indication of the lead time for both large and small samples, is more easily understood by the lay person, provides a more accurate and less volatile trend, and ensures long lead times (good service) are not ignored. To be sure that some additional value of the median metric has not been overlooked, the Performance Branch will continue to track it for both tornadoes and flash floods.

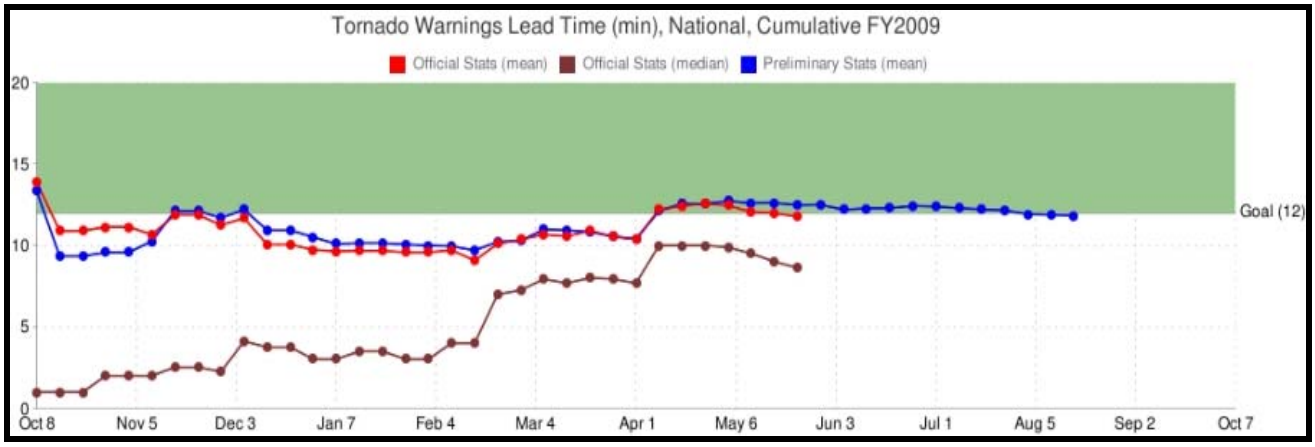


Figure 2: FY 2009 Cumulative Tornado Warning Lead Time Trends. Mean calculation (RED) and Median calculation (Brown) for official statistics. Preliminary mean calculation is also plotted (BLUE). (Figure courtesy of Momchil Georgiev, Performance Branch.)

USING FEEDBACK THAT WORKS!

"When performance is measured, *performance improves*. When performance is measured and reported back, *the rate of improvement accelerates*."

Information Managers Face-to-Face Conference Held at Central Region Headquarters in June

NWS Headquarters Office of the Chief Information Officer (OCIO) and Central Region Headquarters (CRH) hosted the Information Managers *Face-to-Face* Conference (IMC) on June 16-19, 2009. There were several great technical presentations given by all regions. Many innovative and intuitive things are going on throughout NWS! My thoughts:

Microsoft SCCM2007 – The inventory and catalog capability are very powerful, with almost endless capabilities in generating reports. Package/ advertisement distribution is simple and truly an enterprise solution in updating and patching both server and client systems **(Figure 1)**.

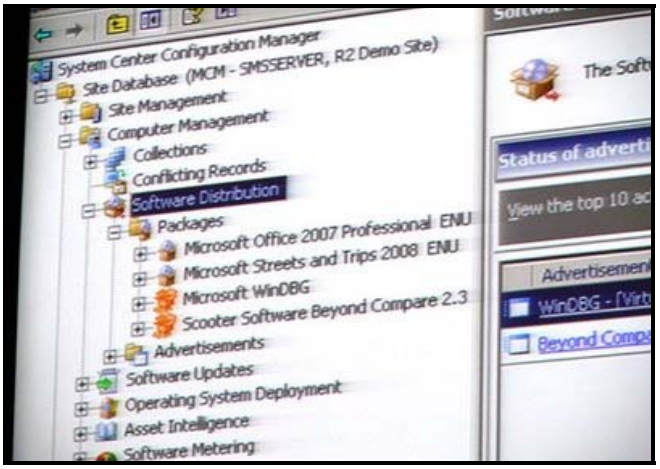


Figure 1: Microsoft Systems Center Configuration Manager 2007 (SCCM2007)

NWS National Active Directory – A must if the end-state is to transitively thread all regions into one corporate forest! If one added Common Access Card (CAC) and Public Key Infrastructure (PKI) enabling into Active Directory, a true *single sign on* environment could exist.

Microsoft Enterprise Equipment Purchase – FY10 will hopefully bring the migration to Microsoft's next generation operating system, Office 2007 and powerful hardware to match. My old system is a workhorse! Yes, it's old, large footprint, heavy, slow when I multi-task it; I could go on, but can you imagine the ROI after all these years and it still delivers.

NWS Headquarters Office of the Chief Information Officer Announces Authorizing Official Policy Change!

The announcement: Authorizing Official (AO) authority will be delegated to the Director level for all NOAA/NWS standard systems. The decision is based on National Institute of Standards and Technology (NIST) guidance, where the standard system's AO must lie within the acquisition cycle of the system. The NWS Chief Information Officer (CIO) is not part, nor owns, the ability to control acquisitions for systems not directly under the OCIO control- the *funding line*, as I refer to it!

Windows 7 – Test driven it yet?

The expected official release date of the MS Win 7 Release Candidate 1 (RC1) is October 22, 2009. How many out there have installed and tested the newest Microsoft operating system?

Continued on next page…

Weather Bits– Continued from Page 4

Figure 2: Microsoft Windows 7 Release Candidate 1 (RCI)

My Windows 7 experience started with the infamous release of the Beta, only 1 million downloads and license keys, and an entire weekend trying to get the image. RC1 was a little easier in grabbing the image and key. In less than 30 minutes, RC1 was installed—nothing missed or corrupt. I have not had any errors, not one, since the RC1 install. Windows 7 **(Figures 2 and 3)** feels

Figure 3: Microsoft Windows 7 Release Candidate 1 (RCI)

lighter in footprint, faster in speed, and more appealing esthetically, compared to the early days of Vista.

Which version do you purchase? Yes, multiple options are available, as with Vista.

Anyone out there with a development experience, I encourage you to email me and share the experience! ●

ASSESSING THE ACCURACY OF WESTERN REGION'S PRECIPITATION FORECASTS

By Colin Baxter, Western Region Headquarters

Verification efforts this summer in Western Region (WR) have focused on the quality of quantitative precipitation (QPF) and probability of precipitation (PoP) forecasts. WR Scientific Services Division (SSD) hosted a GoToMeeting seminar in July to review PoP and QPF forecasts from May and June. The weather pattern and coverage of precipitation in the West during these two months could be described as "polar opposites." The Great Basin was dry and the Pacific Northwest was wet in May. June was more atypical with record precipitation across the interior states and abnormally dry conditions in the Pacific Northwest.

WR SSD has been conducting weather reviews since last December to highlight a pronounced dry bias in PoP forecasts across Western Region. To reduce the dry bias, WR offices implemented the ClimoPoP GFE smart tool, which incorporates climatology into the PoP forecast. Verification scores indicate using a climatological approach in PoP forecasting has improved PoP forecasts in Western Region.

While progress has been made to reduce the dry bias, there remains room for improvement. Verification statistics from May and June indicate forecast offices are not aggressive enough when forecasting higher PoPs (e.g., a PoP forecast

Continued on next page…

Assessing the Accuracy of Western Region's Precipitation Forecasts – Continued from Page 5

of 40 percent verifies 50-70 percent of the time (**Figure 1**). Regional PoP reliability scores suggest that forecasters should be more aggressive with their PoP forecasts, especially at longer lead times.

To support forecast operations, WR SSD has continued work on tools to populate precipitation-related elements in the Graphical Forecast Editor (GFE). Enhanced versions of the ClimoPoP and QPFHelper smart tools will be released before the wet season begins in September. WR SSD is also developing a snowfall analysis to verify snow amount forecasts this upcoming winter.

The SSD is working on providing better feedback to the field on the reliability of PoP forecasts. A product is under development, which will give forecasters an in-depth look at their performance on previous precipitation events. They will be able to review their PoP reliability progression over the entire seven-day forecast period, as well as over an aggregate period of 90 days. This will help forecasters recognize situations in which they can be more aggressive in their PoP forecasts, and more effectively highlight the onset of high impact weather. This new interface will be available on the WR SSD intranet page at the end of the summer.

Please visit the following URL:
http://ww2.wrh.noaa.gov/ssd/digital_services/

The goal of these efforts is to give WR forecasters better tools and feedback, so they can more easily assess the accuracy of their forecasts. In turn, this will result in improved forecasts and service for high-impact events, especially at longer lead times, allowing local officials to make better decisions as they prepare their communities for such events.

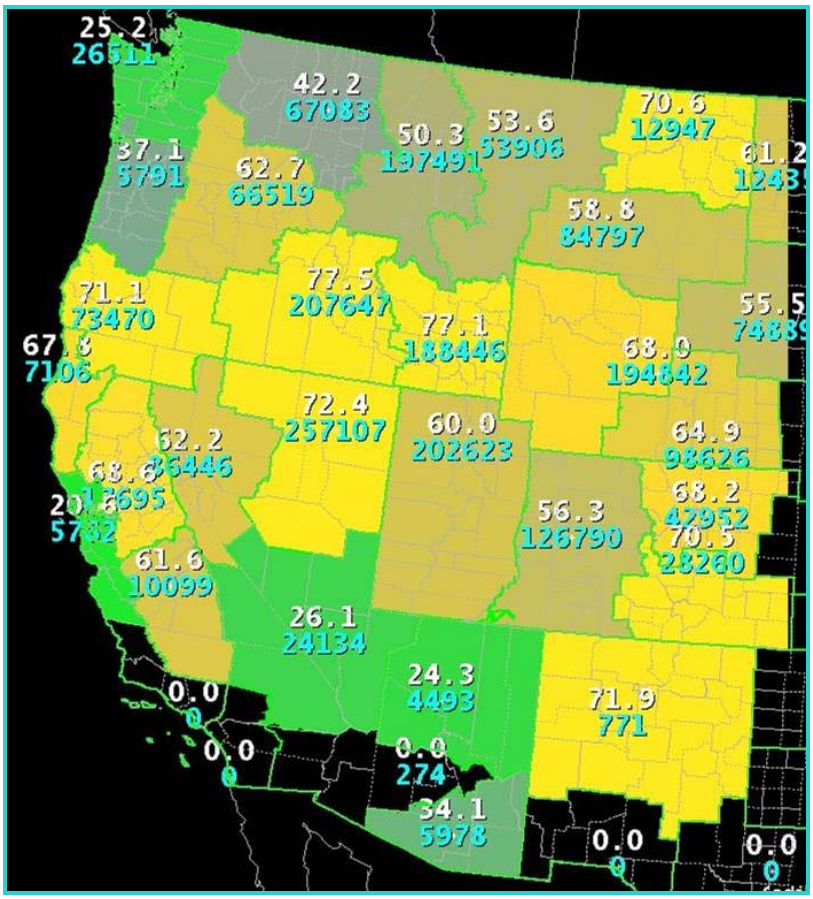
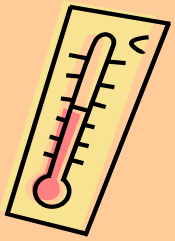


Figure 1: Reliability of 84-hour PoP forecasts containing a 40 percent PoP during June 2009 by county warning area (CWA) in the West. White numbers represent PoP reliability, and aqua numbers represent the number of grid points with a 40 percent PoP forecast in each CWA for June. Tan/yellow colors indicate CWA's with PoP reliabilities greater than 40 percent (dry bias), and grey/green colors show CWA's with PoP reliabilities less than 40 percent (wet bias).

Temperature Forecast Opportunities Across the NWS Southern Region



Bernard N. Meisner, NWS Southern Region Headquarters, and Néstor S. Flecha, University of Puerto Rico at Mayagüez

As the skill of the guidance from numerical weather prediction models continues to improve, the role of National Weather Service forecaster is evolving from primarily producing forecasts to interpreting the forecasts for our core partners. Nevertheless, there remains a fraction of days for which the forecaster can add substantial value to the numerical guidance. The purpose of this study was to identify the frequency of Forecast Opportunities, days for which the maximum and/or minimum temperature guidance is in error by 10 °F or more.

The data used were the extended range Global Forecast System Model Output Statistics (GFS MOS) Guidance (MEX) for 32 sites in the NWS Southern Region, one in each WFO's County Warning Area. We compared the MOS forecast daily maximum and minimum temperatures to those observed during the period January 2007–May 2009. We also noted the days when the observed temperatures departed substantially from the climatological normals for the month, and when there were large day-to-day changes in the observed temperatures, typically the result of frontal passages **(Figure 1)**.

We found that Forecast Opportunities occurred most often for calendar days, rather than for a particular lead time for successive model runs. Forecast Opportunities were most frequent for the extended lead times (Days Three through Six), indicating forecasters should not focus their efforts on the short-term only.

Forecast Opportunities typically occurred as frequently as 10–20 percent of the time during the cold half of the year, but were much less frequent, or even absent, during the summer **(Figure 2)**. To put those values in context, a frequency of 15 percent suggests one day out of a seven day forecast would be a Forecast Opportunity. For most sites, the number of Forecast Opportunities increased from 2007–2009; research to date has not provided a clear explanation for that increase.

Future research will focus on identifying indicators a forecaster might use to determine which day (if any) might be a Forecast Opportunity. Some preliminary work using the relative spread of the GFS ensemble as an indicator of the confidence in the guidance has not been promising.

Please review **Figures 1 and 2** on page 8.

Temperature Forecast Opportunities Across the NWS Southern Region – Continued from Page 7

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	KDFW December			Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	2007		Max/Min	Observed	Day 1	Day 2	Day 2	Day 3	Day 3	Day 4	Day 4	Day 5	Day 5	Day 6	Day 6	Day 7	Day 7	Day 8	Day 8	Day 9
3		Day			0000	1200	0000	1200	0000	1200	0000	1200	0000	1200	0000	1200	0000	1200	1200	0000
4		1	Max	72	70	71	73	73	76	74	71	72	59	58	70	67	61	63	65	
5			Min	62		57	58	58	60	60	58	55	42	45	51	55	49	47	46	40
6		2	Max	81	75	74	73	76	77	75	74	73	68	69	69	74	71	56	68	
7			Min	45		64	64	63	64	63	62	64	61	59	53	52	58	47	48	42
8		3	Max	59	55	60	61	63	70	63	59	62	67	63	62	63	61	60	64	
9			Min	35		41	39	40	43	41	42	39	42	44	42	39	44	46	47	37
10		4	Max	71	67	66	72	71	75	74	75	70	71	71	69	69	65	69	57	
11			Min	35		35	37	38	39	37	40	40	40	41	37	38	39	40	38	40
12		5	Max	76	71	72	74	73	69	71	69	66	71	71	71	67	69	69	67	
13			Min	41		40	45	44	49	46	46	44	49	46	46	43	46	40	41	44
14		6	Max	67	62	63	63	66	71	66	66	67	66	62	62	65	67	60	64	
15			Min	40		44	42	47	45	46	49	46	42	45	46	46	46	45	46	46
16		7	Max	84	78	80	79	81	81	76	78	75	70	68	67	74	68	69	68	
17			Min	57		55	57	56	58	57	59	56	50	50	55	49	45	50	48	41
18		8	Max	79	77	82	78	83	80	80	76	75	76	76	69	65	64	73	70	
19			Min	56		60	59	64	64	63	64	57	60	61	49	46	47	57	49	55
20		9	Max	56	55	58	63	62	67	62	53	54	66	76	73	77	68	70	71	
21			Min	37		58	63	62	67	62	53	54	61	64	61	60	55	51	49	52
22		10	Max	47	59	64	61	61	55	49	60	48	44	45	57	70	67	71	65	
23			Min	37		43	44	46	44	42	52	48	36	39	46	54	56	60	52	58
24		11	Max	68	71	66	64	65	66	65	57	70	55	47	53	54	57	66	62	
25			Min	44		54	59	63	61	61	54	49	51	35	33	41	46	58	51	56
26		12	Max	43	40	44	43	43	47	48	54	52	51	54	48	49	51	53	65	
27			Min	38		41	43	43	47	45	50	46	40	48	37	32	37	43	45	43
28		13	Max	52	56	60	56	59	57	59	63	58	67	61	55	56	54	53	56	
29			Min	35		36	34	35	34	38	40	39	42	40	36	37	37	35	36	33
30		14	Max	50	46	50	46	49	48	45	49	48	56	54	55	53	53	57	55	
31			Min	38		38	39	40	41	37	40	39	39	39	40	39	36	38	35	34
32		15	Max	47	47	48	49	49	48	46	47	45	42	50	60	58	54	57	55	
33			Min	34		38	40	38	37	37	36	36	33	35	36	37	37	38	35	40
34		16	Max	51	50	54	51	54	52	55	51	53	53	52	55	58	61	56	57	
35			Min	28		31	30	29	30	29	30	32	29	31	31	32	43	38	35	39
36		17	Max	56	57	58	58	60	63	58	56	57	57	53	55	58	62	63	60	
37			Min	29		29	30	30	32	30	29	35	32	32	30	32	39	41	41	36
38		18	Max	73	64	65	67	66	64	65	64	67	61	59	61	58	57	60	64	
39			Min	40		42	44	44	45	46	46	45	43	42	39	36	36	41	43	49
40		19	Max	69	72	70	69	68	67	67	68	65	69	70	65	69	62	61	59	
41			Min	39		40	38	41	42	46	45	47	48	49	46	43	45	45	42	46
42		20	Max	67	68	68	71	69	68	66	67	64	68	67	67	68	63	60	60	
43			Min	45		44	47	47	49	47	47	44	52	48	48	47	49	51	46	44
44		21	Max	73	75	73	75	72	73	73	68	66	70	69	66	62	63	63	59	
45			Min	38		40	40	40	42	43	42	43	44	44	43	43	41	44	40	44
46		22	Max	60	57	56	54	56	60	53	52	52	57	62	61	58	64	57	56	
47			Min	32		49	51	49	50	52	51	46	49	49	44	42	51	43	46	41
48		23	Max	55	52	54	54	56	52	52	52	51	57	51	51	56	56	62	51	
49			Min	24		34	30	30	32	30	34	38	32	35	31	35	35	36	40	41
50		24	Max	58	57	57	54	55	57	53	44	43	54	46	55	58	55	59	54	
51			Min	31		29	34	33	33	27	31	34	31	36	32	30	33	36	36	40
52		25	Max	62	61	60	61	61	62	60	58	54	48	49	55	45	56	56	58	
53			Min	33		33	33	33	33	35	35	33	34	33	32	34	34	38	35	40
54		26	Max	49	49	51	53	46	50	54	56	58	53	55	54	49	51	48	49	
55			Min	38		40	41	39	36	38	37	42	43	43	35	35	39	34	36	40
56		27	Max	45	52	55	50	52	52	55	52	50	58	60	55	51	50	50	51	
57			Min	39		34	31	32	31	31	34	37	36	32	33	34	35	36	34	31
58		28	Max	51	52	55	54	56	52	55	56	56	59	59	56	56	54	61	53	
59			Min	34		36	36	36	40	36	35	37	41	39	41	37	36	35	36	31
60		29	Max	55	56	57	56	57	57	59	51	56	62	62	64	63	58	56	55	
61			Min	29		32	32	30	31	32	31	34	32	38	37	39	40	38	35	33
62		30	Max	64	60	62	63	62	64	61	64	64	62	66	60	71	60	62	58	
63			Min	36		34	34	33	36	36	38	41	36	36	36	37	43	45	39	39
64		31	Max	60	62	61	62	59	63	59	59	57	59	59	57	61	61	62	57	
65			Min	37		37	37	39	43	40	41	42	42	36	38	41	42	45	37	39

U	V	W	X	Y	Z
Criteria			Percent Exceeded		
10	15		16%	4%	3%
Day	Max	Min			
1	72	62		Climate Normals	
2	81	45	Max	Min	Criterion
3	59	35	56.5	36.8	15
4	71	35			
5	76	41			
6	67	40			
7	84	57			
8	79	56			
9	56	37			
10	47	37			
11	68	44			
12	43	38			
13	52	35			
14	50	38			
15	47	34			
16	51	28			
17	56	29			
18	73	40			
19	69	39			
20	67	45			
21	73	38			
22	60	32			
23	55	24			
24	58	31			
25	62	33			
26	49	38			
27	45	39			
28	51	34			
29	55	29			
30	64	36			
31	60	37			

Figure 1: Sample Forecast Opportunities spreadsheet for KDFW December, 2007. Blue cells are the Forecast Opportunities, times when the MOS guidance error was 10 °F or greater; pink cells are the observed maximum and minimum temperatures that departed from the monthly normals by 15 °F or more; brown cells indicate when the day to day change in observed maximum or mimimum temperature was 15 °F or greater.

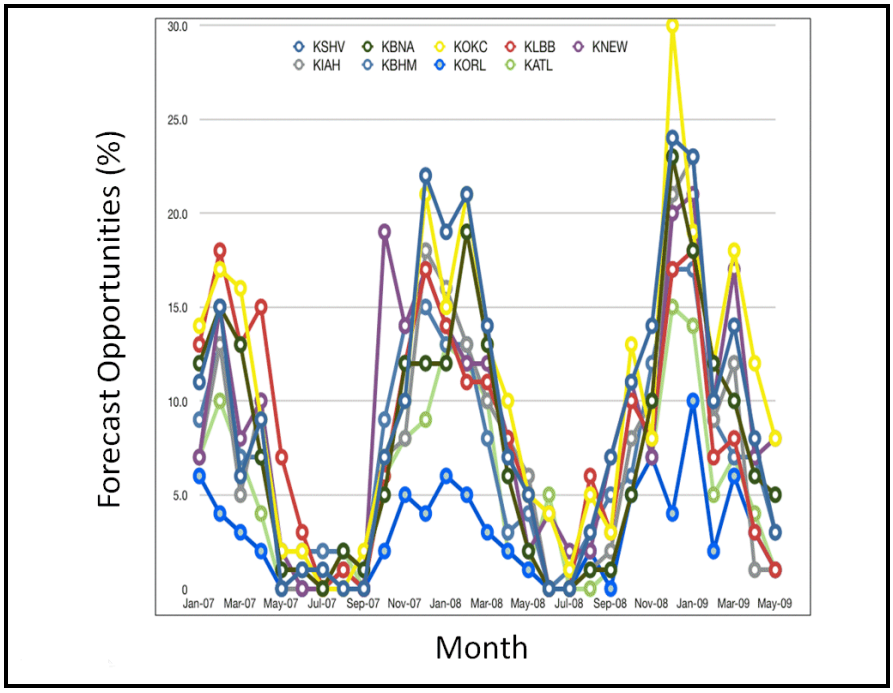
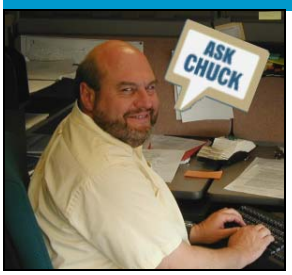


Figure 2: Secular variation in the frequency of Forecast Opportunities (January 2007–May 2009) for selected locations in the NWS Southern Region. The greater number of Opportunities during the cold half of the year is apparent, as is an increase in the number of Opportunities from 2007 to 2009. ●



ASK CHUCK!

By Chuck Kluepfel, NWS Headquarters

Question 1: Is there a reason the format of the Aviation page keeps changing? Also, is there a reason the Marginal Visual Flight Rules (MVFR) category was split up so that flight category statistics are much harder to compute?

Answer: Here is a record of recent changes to the Terminal Aerodrome Forecast (TAF) system and the reasons for the changes. Not all changes were well received so we reversed one (the last bullet).

- We standardized the new format of the contingency tables among all of the programs we serve. The TAF contingency tables now appear similar to the tables on the marine verification page. Until now, we have had nothing but favorable comments about the new format. People seem to find them easier to read, and if you don't like the fancy colors, you can toggle them off. Unlike before, for entries in the contingency table, you can toggle back and forth between minutes, hours, total 5-minute intervals, and percentages of the entire contingency table. In the past, you had to re-run another report to change from percentages to hours. The toggle switches for these features are found on the "Report Format" line in the report header.
- The definition of MVFR has not changed. It is (and always has been) 1000 feet to 3000 feet ceilings, or visibilities 3 to 5 statute miles. The worst of the two elements (ceiling vs. visibility) always rules. If the ceiling remains in the 1000 to 3000 foot range, but the visibility drops below 3 miles; the visibility is the driver, and you have dropped below the MVFR category to Instrument Flight Rules (IFR) or lower. If the visibility rises to six miles or more, Visual Flight Rules (VFR), but 1000 to 3000 foot ceilings continue (MVFR), the ceiling becomes the driver, and the flight category is MVFR.

A few more upgrades are planned during the course of the next six months. A special graphing routine to be added will require some training for full understanding. Also, in the interest of reducing paper, we will add a feature that will allow users to *opt out* of requesting certain statistics now routinely included in every report (such as the 5x5 or 6x6 contingency table). We already have this option in the public forecast verification statistics. Similar to the public statistics, you will soon be able to run several related *Stats on Demand* requests at one time with output to a spreadsheet. You can then use the spreadsheet to display tabular and/or graphic statistics.

Question 2: When I run aviation statistics and click on problematic Aviation Routine Weather Reports (METARS) at the bottom of the report, I get a large list of METARs. Most of the remarks contain *Thunderstorm information not available* (TSNO), setting thunderstorm to missing. Under these circumstances, do you keep the rest of the METAR for verification, including the ceiling and visibility?

Answer: Yes. TSNO remarks have no affect on any aspect of TAF verification other than thunderstorms. Our system quality controls (QC) each observation and lists all observations with inconsistencies and noteworthy remarks, such as TSNO. The presence of this remark tells the algorithm the thunderstorm observation is missing, which is much different than the non-occurrence of a thunderstorm. The former doesn't count in the contingency table; the latter counts as a thunderstorm non-occurrence and is tallied in the "no" row. ●

Aviation Verification A Central Region Method

By Dan Bloom, Central Region Headquarters

Method Setup:

Central Region uses the Performance Branch's Aviation Stats on Demand to create three separate aviation verification spread sheets 1) A monthly spread sheet containing aviation verification data for each month since January 2003, 2) An annual spreadsheet which contains yearly statistics for each fiscal year since FY03, and 3) Another monthly spreadsheet, like the first monthly spreadsheet, except this one includes both scheduled and amended Terminal Aviation Forecasts (TAFs). The first two spreadsheets only include scheduled TAFs and can be found on the Central Region Intranet. With these preceding exceptions, all three spreadsheets use the same selected parameters.

For "Element Type," *Flight Category* has been chosen. For "Forecast Type," *Operational Impact* has been chosen. These choices reflect the conditions for which National Weather Service (NWS) customers (pilots, dispatchers, etc.) must plan.

For "Guidance Type," *None* is selected. Government Performance and Results Act (GPRA) performance measures are based on *no guidance*. NWS customers likely do not care about how the NWS performs against its guidance products, but rather how the NWS forecast compares to what occurs.

Stats on Demand gives the user an opportunity to get verification data for above or below a certain critical threshold. In this method, Central Region uses ceilings 2000 feet and visibility 3 miles. This selection was made based on advice from then Central Region Aviation and Fire Weather Focal Point, and now retired, Gary Schmeling.

For "TAF Begin Times," the default (all boxes checked) was selected. For "Forecast Projections," the defaults, ">0-3" and ">3-6" hours, were selected. These selections cover what TAF information NWS customers find most critical and use most often.

Data Selected to Monitor:

In this method, Central Region keeps track of the five, what are called "slices" (Visual Flight Rules (VFR), Marginal Visual Flight Rules (MVFR), Instrument Flight Rules (IFR), Low Instrument Flight Rules (LIFR) and Very Low Instrument Flight Rules (VLIFR)). The reasoning for this is based on customer wants and needs. When the NWS forecasts MVFR conditions, the customer expects the observation to verify MVFR conditions. When the NWS forecasts VLIFR conditions, the customer expects to see VLIFR conditions. Customers make their decisions to fly or not fly based on these forecasts. The "slice" Probability of Detection (POD), False Alarm Ratio (FAR) and the Critical Success Index (CSI) measure how well NWS aviation forecasters are meeting this customer need.

In this method, Central Region also tracks the POD, FAR and CSI for "IFR & Below." POD and FAR for "IFR & Below" are the GPRA measures. Central Region also keeps track of POD, FAR, and CSI for thunderstorms, since thunderstorms can be life-threatening to NWS aviation customers. The

Continued on next page…

Aviation Verification- A Central Region Method- Continued from Page 10

element type "Significant Weather" instead of the "Flight Category" must be selected to obtain the thunderstorm verification data.

The bias for each of the four lowest "slices" is monitored. For any given "slice," the bias is the number of forecasts divided by the number of observations. Ideally this is near one. With a bias significantly less than one, customers may be caught with lower conditions than expected and need to find an alternate airport to land. With a bias significantly greater than one, customers may, too often, refrain from flying because of expected lower conditions which do not develop. These customers may be inhibited from meeting their business obligations and other travel plans needlessly. All of these factors can have financial and other consequences for NWS aviation customers.

Central Region also keeps track of the "Percent Hits" and the "Percent >1 Category Errors." The "Percent Hits" is a measure how often the NWS gave its customers the correct forecast. The "Percent >1 Category Errors" is a measure of how often the NWS significantly missed the correct forecast which can lead to adverse consequences for our NWS aviation customers.

Conclusion:

This method provides a more comprehensive evaluation of how the NWS is serving our aviation customers. ●

DID YOU KNOW?

Performance Branch Meteorologist Brent MacAloney is now on NWS Chat to answer your questions and/or help out with Performance Management website support issues.

The NWS Chat username is: nws-brent.macaloney

Feel free to add Brent to your Buddy List in NWS Chat.

Administration Wants Performance Management Data Put to Better Use

Article courtesy of Gregg Carlstrom,
Staff Writer, Federal Times

The Obama administration's performance management agenda, to be released within the next few months, aims to improve the way agencies use their performance management data, according to Dustin Brown, the Office of Management and Budget's deputy assistant director for management.

"We collect a lot of information in the federal government. A ton of it," Brown said, speaking last week at the Excellence in Government conference. "But one of the things we don't do very well is organizing and using the information we're currently collecting."

The new performance management program will encourage agencies to use data to look for inefficiencies within their organization—why veterans' claims are processed more quickly in one field office than another, for example.

"Just weighing the pig doesn't make it fatter," Brown said. "I think we've done a lot of measuring and reporting up until now. And we'll continue to do that, and it's important. But just doing that in and of itself, and that process, is not going to change the dial. We need to take that next step going forward."

Several experts at the July 20 conference said agencies will face a number of challenges in implementing and meeting performance management goals. One is tight budgets: After big increases this year, many agencies will see flatter budgets in 2011 and beyond.

"The challenge in front of all of you is to think about how you're going to meet those increasing demands with the level of resources you have now," said Nancy Killefer, the director of McKinsey & Co.'s public sector practice.

Robert Shea, a former OMB associate director for management during the Bush administration, said OMB will also need to convince agencies that it isn't just imposing goals from above. Many federal managers have complained that the Program Assessment and Rating Tool, the current performance management system, is just a compliance exercise.

Shea also said agencies are struggling to develop metrics for programs with more abstract goals, such as grant programs and public awareness campaigns.

"You need to figure out exactly what it is you want to do, collect data go out and promote these strategies, implement these strategies," said Shea, now a director of Grant Thornton's public sector practice. "It might seem that what we do is remote to someone buckling their seatbelt, but you can, in a sophisticated way, identify the strategies that get people to buckle their seatbelt."

Brown said OMB has met with more than a dozen agencies to discuss their performance goals. ●

On the Road Again…

By Brent MacAloney, NWS Headquarters

Summer is normally a great time to travel away from the Washington DC metro area. I really enjoy escaping to the north and west to avoid the miserable heat and humidity that socks the region beginning in late May. This year is somewhat different though in that the weather has been rather "spring-like" with temperatures being 10-15 degrees below normal here in Maryland. Since the weather has been so nice, I think I have sub-consciously cut back on my summer traveling marathons in order to stay around and actually enjoy what Maryland summers have to offer.

Kansas City, MO

I was given a unique opportunity in April 2009, to attend my first residence course at the NWS Training Center in Kansas City, MO. This came about because there was a last minute opening in the "Field Operations Management" course and I was asked if I would like to go.

Initially, I didn't know if it was appropriate for me to be in a class named "Field Operations Management" since I have never worked in the field. I asked if I could have a day to think about it. After reading the course description, I really felt I was being given an excellent opportunity to not only get a better understanding of challenges that lead forecasters undergo while overseeing a shift, but also to improve my leadership skills and network with a diverse group of individuals. I must say, the class was absolutely awesome and I am so glad I decided to take it.

A few of the many highlights I had from this class started on the first day when a very dynamic speaker named George Ferguson of The Schuster

"If there's any way we can streamline data entry or the analysis process, we'll work on it."

Kane Alliance, Inc. really set the tone for the week. He led a session called "Leadership Development for Managing People" where we discussed ways to excel at leadership in the workplace. This session got me, and the rest of my classmates, thinking about how we can return to our jobs and play a more active roll in leading the organization in the right direction. He also showed some wonderfully empowering quotes from great leaders, which the class really enjoyed.

I also enjoyed the "Conflict Communications" session lead by Sky Young-Wick. Dealing with conflict makes many people feel uncomfortable. In many cases, people would rather avoid conflict rather than confront it in a professional manner. This leads to bigger problems such as lack of trust, inability to communicate, and lackluster performance as a team, if the conflict is not addressed properly. Sky did an awesome job in teaching the class techniques on dealing with conflict, giving everyone the confidence to go back to their workplace and effectively implement tools to resolve future situations where conflict may arise.

Overall, this was a great class, and I recommend it to anyone in the organization who wants to take more of a leadership role and help bring this organization to the next level. It serves as a reminder to step back and look at the whole picture before judging or making assumptions on decisions made around your workplace.

Continued on next page…

On the Road Again– Continued from Page 13

Boulder, CO

Nothing is more enjoyable than a trip to Colorado in May. On top of the amazing sights, the weather is top notch…warm and dry! The purpose of going to Boulder was to meet with Jeff Lazo and Emily Laidlaw of the National Center for Atmospheric Research's (NCAR) Societal Impacts Program (SIP) regarding the findings from the survey they conducted with Warning Coordination Meteorologists (WCMs) and Storm Data Focal Points in 2008; the entry of storm data and storm damage estimates.

The three main findings of this survey are as follows:

- A lot more training is needed on storm data.
- Assistance is needed in deriving quality estimates on damage amounts.
- Time allocated to collection and entry of storm data is lacking.

For those of you who actively work with storm data, these are probably "no-brainers," but I think from a NWS Headquarters end of things, until the survey results were presented, we did not realize how serious the situation really is. All three of these findings will be kept in mind with the next-generation storm data collection program. I will have more to say on that in the winter 2009-2010 issue of the Peak Performance Newsletter.

Cheyenne, WY

At the end of my trip to Boulder, I decided to head 2 hours north to Cheyenne, WY. WFO Cheyenne Meteorologist-in-Charge (MIC) John Eise, asked me to come up so that we could discuss verification matters with him and his staff. As an added bonus, we were fortunate enough to have the MIC from the Riverton WFO, Kevin Lynott, in attendance as well. This served as a perfect forum for us to discuss how the WFOs in Wyoming could use the tools on the Performance Management site to improve forecasts and warnings, as well as better serve the public.

After the great discussions we had at this meeting, I walked away with the feeling that these two offices now have the confidence to analyze their long-duration warning verification data (Winter Storm and High Wind) and pick out areas in which they can improve public service. In this part of the country, these products are absolutely crucial and for many, can mean the difference between life and death.

Fort Worth, TX

NWS Southern Region was holding a regional WCM workshop in Fort Worth, TX the week of June 22nd. Since I had not been to Southern Region Headquarters (SRH) to give a presentation on the Performance Branch in almost three years, I decided to head out to the workshop a day earlier so I could meet with the SRH employees. I typically like to give these Performance Branch overviews to the regional headquarters on a biennial basis since so many changes occur from year to year within the Performance Branch.

The meeting with the employees at SRH was productive. After this meeting, I realized the Performance Branch still has a long way to go in marketing the tools and information we have available on the Performance Management website. I feel this way because no matter where I go, whenever I demonstrate the various tools, people respond with the comment, "Wow, I wish I knew that was there when I was working on a project a month ago" or "How long have you had that tool available?" In most cases, these tools

Continued on next page…

On the Road Again– Continued from Page 14

have been available for quite some time. I am going to think about how we can better market our tools. If you have any ideas, I'm open for suggestions.

I'm sorry I was unable to stay longer at the WCM Workshop. It is rare that I get to interact with a large group of WCMs, and pick their brains on how to make our tools more usable. During the short time I was there, I was able to brief the WCMs on the next-generation of storm data collection, the future of service assessments, and answer generic performance management questions.

As always, the resounding response from the WCMs is that they are constantly being asked to do more with less. So, any improvements to our tools that can make the entry or analysis of data easier will be appreciated. One of my commitments to NWS employees is that with all future development, the impact on workload, and time constraints will be kept in mind. If there's any way we can streamline data entry or the analysis process, we'll work on it. We encourage all of you who are interested to send us ideas on how to streamline our processes.

Burlington, VT

The official NWS travel was over, so it was time for me to retreat to my family's summer house, deep in the woods of north central Vermont. We were fortunate to have National WCM Program Lead, Chris Maier, and his wife along for vacation. During our visit, we were lucky enough to capture a photo of a *double* rainbow, which appeared over Montpelier **(Figure 1)** on the 4th of July!

Figure 1: A Fourth of July rainbow over Montpelier, Vermont. (Photo by Brent MacAloney)

Continued on next page…

On the Road Again– Continued from Page 15

While in Vermont, I made it a point to visit Burlington. Since this trip was pleasure and no business, there was no forecast visit. Instead, we met with WFO Burlington WCM, Scott Whittier for a Vermont Lake Monsters baseball game. It was fun to catch up with Scott, talk a little shop, and enjoy a beautiful night at the ball park **(Figure 2)**.

Well, that is it from the road this time. In future Issues I'll have highlights from a trip with Doug Young (Performance Branch Chief), to Norman, OK for a "meeting of the minds" to discuss implementing more meaningful tornado warning performance metrics. I also hope to have some stories from a trip to Salt Lake City, UT for a Western Region Societal Impacts meeting. Until then, I hope all of your summer travels are fun and safe!

Cheers!
Brent

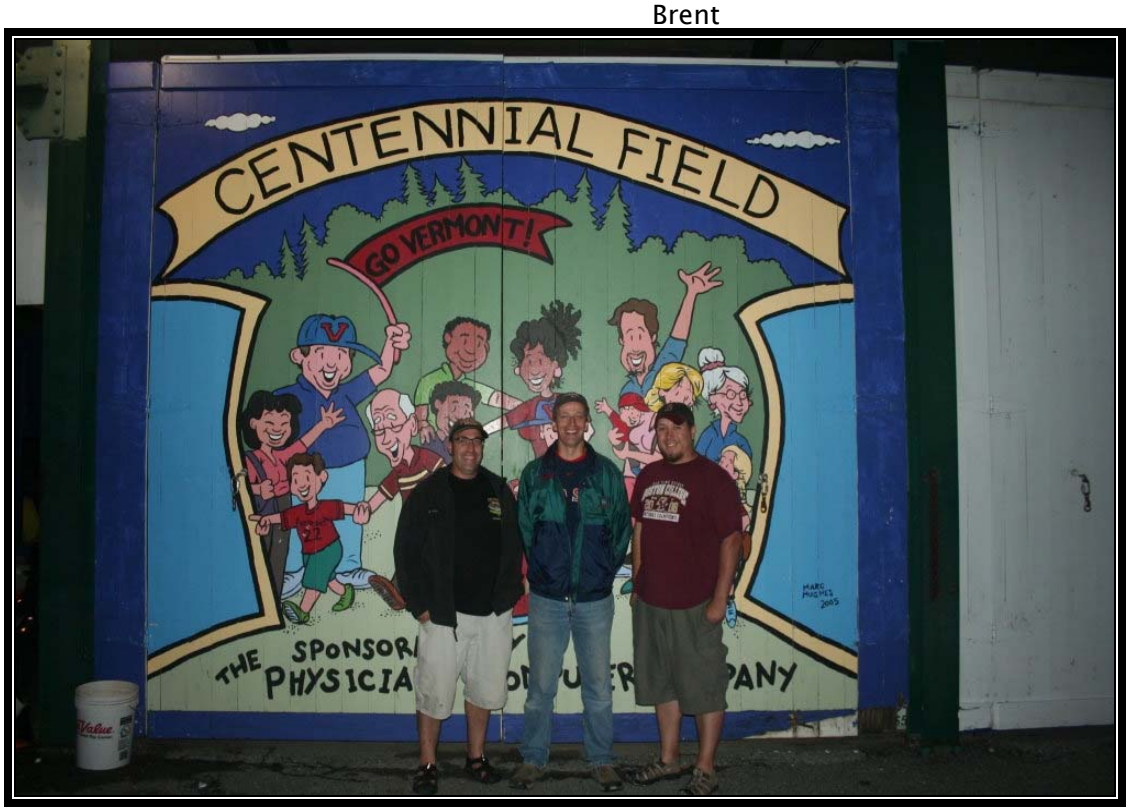


Figure 2: Two WCMs and a verification guy at the Vermont Lake Monsters baseball game. Pictured from Left to Right– Chris Maier, National WCM Program Manager; Scott Whittier, WCM Burlington; and Brent MacAloney, Performance Branch. (Photo by Manina MacAloney)

New Verification Tool
Sky Cover

By Chuck Kluepfel, NWS Headquarters

We will soon be adding a sky cover product to the *Stats on Demand* public forecast verification suite. Forecast data come from the Point Forecast Matrices (PFM) so this is a point-based verification system. Observations come from a combination of Aviation Routine Weather Reports (METARS) and satellite estimates. Total *opaque* sky cover is the observed quantity used for verification because cloud forecasts issued by the NWS are assumed to be opaque. All clouds reported from the Automated Surface Observing System (ASOS) and Automated Weather Observing System *(*AWOS*)* instruments with bases at or below 12,000 feet are used. Cloud estimates above this height threshold come from the satellite (GOES) cloud product (SCP). We had considered using the high cloud reports that get augmented by human observers to a small number of METAR points, but these terminals make up a small percentage of the terminals that are verified, and we have found these data to be inconsistent.

The SCP provides a convenient but imperfect estimate of total sky cover at or above 12,000 feet. The categorical value (clear, scattered, broken, and overcast) includes clouds from thin cirrus to thick altostratus. One of these categorical values is provided for middle and/or high clouds, and whenever both are provided, the value representing the greater sky coverage is used to represent total sky cover at or above 12,000 feet. The column in the SCP labeled ECA (effective cloud amount) provides an estimate of average opacity for all clouds at or above 12,000 feet. The ECA runs from zero (no clouds or very transparent clouds) to 100 (opaque clouds) and is used by the following algorithm to decrease the estimate of total opaque sky cover whenever the ECA is low. Consequently, transparent cirrus clouds play a minimal role in verification.

https://verification.nws.noaa.gov/content/pm/verif/public/pubSky.doc

The SCP has some known short-comings, which are more completely documented in the NWS Technical Procedures Bulletin 410 (Kluepfel et al., 1994):

- Data are not available for Alaska so Alaska is not included in sky cover verification at this time. Efforts are being made towards using some estimates from the polar orbiting satellites, but so far the data lack consistency.
- When thin clouds are present, the cloud top heights in the SCP should be used judiciously. This does not present a problem for verification because once the SCP has identified one or more cloud layers above 12,000 feet, the verification program is only interested in cloud opacity, not height.
- Sometimes very thin cirrus clouds are undetected so the total sky cover in these situations is CLR and the ECA is reported as zero. Since these clouds are very thin, they have a very minimal contribution to cloud opacity and don't affect verification.
- A strong surface-based temperature inversion within an Arctic air mass may result in a false report of clouds above 12,000 feet Mean Sea Level (MSL) This presents enough of a problem for sky

Continued on next page…

New Verification Tool- Sky Cover- Continued from Page 17

cover verification that we do not verify *any* point in space and time with a surface temperature below -20 °C.

We recently ran a Beta test of the contiguous states and Hawaii from September 2008 through May 2009, and the results are presented in the following contingency tables. Contingency table cell entries are given as percentages of the total number of entries in the contingency table. Forecasts come from the Point Forecast Matrix (PFM) and are issued twice a day. The 12-hour forecast periods 1 through 5 (hours 3 through 60) are pooled together, and forecasts are available for and verified every 3 hours. The following five sky cover categories are used for forecasts, observations, and verification: clear, few, scattered, broken, and overcast. Each forecast category for sky cover is represented by a separate column in the contingency table, and each observation category is represented by a separate row. A separate contingency table is given for NWS (PFM) forecasts **(Table 1)**, the Global Forecast System Model Output Statistics (GFS MOS) **(Table 2)**, and North American Mode (NAM MOS) **(Table 3)**. The NWS forecasts for this time period seem to have huge bias toward under-forecasting the clear and overcast categories, whereas GFS and North American Model Model Output Statistics (NAM MOS) do not exhibit this bias. For example, clear skies are only forecast 7.4 percent of the time, whereas clear skies are observed 28.7 percent of the time.

I will close by asking a simple question. Could the under-forecasting of these two categories be a psychological issue with the Graphical Forecast Editor (GFE)? Sky cover is forecast on a scale of zero to 100 percent, just like probability of precipitation (PoP). A couple years ago when I prepared forecasts on the GFE at WFO Sterling, I noticed that I tended to avoid forecasting values at both ends of the extremes, and my neighboring WFOs tended to do the same. If I did forecast zero or 100 percent sky cover, I then had to worry about coordinating with my neighboring offices. While this is often a good strategy for probabilistic forecasts with low forecaster confidence, sky cover forecasts are not probabilistic. Keep in mind I am only speculating about a possible cause and effect relationship. Any thoughts? Feel free to send me an e-mail.

Charles.Kluepfel@noaa.gov

I will have more to say about skill scores through future training modules and future articles in this newsletter. For now, just think that the Peirce Skill Score (PSS) is very heavily tied to the percent correct score (the sum of the upper left to lower right diagonal entries of the contingency table), whereas the Gerrity Equitable Skill Score (ESS) is nonlinear and a little more complex. It is designed to reward the correct forecasting of rare events and play down the correct forecasting of common events. The Gerrity ESS also allows partial credit based on the size of the miss, whereas the PSS does not. The Gerrity ESS is the same score that we use in the marine *Stats on Demand* reports.

Please review the Tables on Page 19.

References:

Kluepfel, C.K., A.J. Schreiner and D.A. Unger (1994): The Satellite-Derived Cloud Cover Product (Sounder), U.S. Department of Commerce, NOAA/National Weather Service *Technical Procedures Bulletin 410*, 15 pp.

Continued on next page…

New Verification Tool– Sky Cover– Continued from Page 18

		Forecasts					
		CLR	FEW	SCT	BKN	OVC	Total
Obser-vations	CLR	5.9%	10.5%	8.3%	3.7%	0.3%	28.7%
	FEW	0.9%	3.0%	4.5%	3.4%	0.4%	12.2%
	SCT	0.3%	1.7%	3.7%	4.3%	0.6%	10.7%
	BKN	0.1%	0.7%	2.4%	4.8%	1.3%	9.3%
	OVC	0.2%	1.0%	4.2%	17.4%	16.3%	39.2%
	Total	7.4%	16.9%	23.2%	33.7%	18.9%	100.0%
	Bias	0.26	1.39	2.17	3.62	0.48	----
Percent hits = 33.7%			PSS = 0.227		Gerrity ESS = 0.390		

Table 1: Verification of forecasts vs. observations for NWS forecasts (from PFM)—contiguous states and Hawaii. 12-hour forecast periods 1 to 5; 0400 and 1600 LT forecast issuances. October 2008 thru May 2009. Data matched GFS MOS (Table 2).

		Forecasts					
		CLR	FEW	SCT	BKN	OVC	Total
Obser-vations	CLR	17.0%	2.8%	2.4%	3.0%	3.5%	28.7%
	FEW	4.6%	1.4%	1.4%	1.9%	2.9%	12.2%
	SCT	3.0%	1.0%	1.2%	1.9%	3.6%	10.7%
	BKN	2.2%	0.8%	0.9%	1.7%	3.7%	9.3%
	OVC	5.6%	2.1%	2.6%	5.4%	23.5%	39.2%
	Total	32.4%	8.0%	8.4%	13.9%	37.3%	100.0%
	Bias	1.13	0.66	0.79	1.49	0.95	----
Percent hits = 44.7%			PSS = 0.242		Gerrity ESS = 0.378		

Table 2: Verification of forecasts vs. observations for GFS MOS—contiguous states and Hawaii. 12-hour forecast periods 1 to 5; 0000 and 1200 UTC model cycles. October 2008 thru May 2009. Data matched to PFM forecasts (Table 1).

		Forecasts					
		CLR	FEW	SCT	BKN	OVC	Total
Obser-vations	CLR	13.8%	1.7%	4.7%	3.2%	5.2%	28.7%
	FEW	3.4%	0.7%	2.1%	1.9%	4.1%	12.1%
	SCT	2.1%	0.5%	1.6%	1.7%	4.7%	10.6%
	BKN	1.5%	0.4%	1.2%	1.3%	4.9%	9.2%
	OVC	3.5%	0.9%	2.9%	3.8%	28.2%	39.4%
	Total	24.4%	4.1%	12.5%	11.9%	47.1%	100.0%
	Bias	0.85	0.34	1.18	1.29	1.20	----
Percent hits = 45.7%			PSS = 0.237		Gerrity ESS = 0.364		

Table 3: Verification of forecasts vs. observations for NAM MOS—contiguous states and Hawaii. 12-hour forecast periods 1 to 5; 0000 and 1200 UTC model cycles. October 2008 thru May 2009. Data matched to PFM forecasts (this set of PFM forecasts not shown). ●

Contributors to this Summer – Fall Issue

Colin Baxter
Western Region Headquarters
Scientific Services Division
Colin.Baxter@noaa.gov

Dan Bloom
Central Region Headquarters
Regional Quality Assurance Meteorologist
Daniel.Bloom@noaa.gov

Gregg Carlstrom
Federal Times Newspaper
Staff Writer
gcarlstrom@federaltimes.com

Nestor S. Flecha
(University of Puerto Rico, at Mayaguez)
Southern Region Headquarters
Scientific Services Division
Nestor.Flecha@noaa.gov

Bob Glahn
NWS Headquarters
Meteorological Development Lab
Harry.Glahn@noaa.gov

Chuck Kluepfel
Performance Branch
Verification
Charles.Kluepfel@noaa.gov

Brent MacAloney
Performance Branch
Verification
Brent.Macaloney@noaa.gov

Bernard Meisner
Southern Region Headquarters
Scientific Services Division
Bernard.Meisner@noaa.gov

Jim Rawls
Performance Branch
Information Technology Specialist
James.Rawls@noaa.gov

Freda Walters, Co-Editor and Designer
Performance Branch
Service Assessment
Alfreda.Walters@noaa.gov

Doug Young, Editor
Performance Branch Chief
Douglas.Young@noaa.gov

Questions and comments on this publication should be directed to Freda Walters.

Web Links

Stats on Demand
https://verification.nws.noaa.gov

NDFD Verification:
https://bestpractices.nws.noaa.gov/contents/ndfd-stats/verification/
(National Verification)
https://bestpractices.nws.noaa.gov/contents/ndfd-stats/wfosummary/
(WFO Verification)

Real-Time Forecast System:
http://www-ad.fsl.noaa.gov/fvb/rtvs/

Please consider contributing to our next edition– Winter 2009 – 2010

Articles due Nov. 6th