

Aaron Christenberry¹, Alex Lamers¹, Brian Mejia¹, Andrew Dean², and Steven Weiss²

¹School of Meteorology, University of Oklahoma, Norman, OK

²Storm Prediction Center, Norman, OK

1. Introduction

Tornado watches are among the most highly visible products issued by the Storm Prediction Center (SPC). They are disseminated to a variety of users, including broadcast media, law enforcement agencies, emergency managers, and the general public. As a result of this wide distribution and the significant threat tornadoes pose to life and property, tornado watches are the SPC product that the general public is most likely to see and react to. Doswell et al. (1999) attribute an increase in public awareness toward tornadoes as a major factor in the decline of tornado-related deaths during the last half-century. However, Donner (2007) suggests that people may tend to take the risk entailed by a tornado watch less seriously owing to a high false alarm rate caused primarily by the small geographic areas affected by tornadoes. As false alarm rate can only be decreased by minimizing the number of false alarms, also known as null cases, a thorough understanding of the patterns and conditions associated with null cases will enable the SPC to determine more accurately if issuing a watch under certain circumstances is, in fact, the best course of action.

According to SPC guidelines, a tornado watch is issued when the forecaster expects that either (1) one significant tornado (F2/EF2 or greater), or (2) two tornadoes of any intensity will occur within the spatial and temporal extent of the prospective watch (Dean and Schaefer 2006). If these expectations are proven correct by post-event storm reports, the watch is considered verified. All tornado watches also include a risk for non-tornadic severe thunderstorms (large hail and damaging wind gusts) and can be considered to be severe thunderstorm watches by default. A non-verified tornado watch may therefore partially verify as a severe thunderstorm watch. For a severe thunderstorm watch to be verified, there

* *Corresponding author address:* Andrew Dean, NOAA/NWS/NCEP Storm Prediction Center, 120 David L. Boren Blvd., Suite 2300, Norman, OK 73072; E-mail: andy.dean@noaa.gov

must be at least six reports of large hail [0.75 inches (19 mm)¹ in diameter or greater] or damaging thunderstorm wind [50 kt (25.7 m s⁻¹) or greater] (Dean and Schaefer 2006). The presence of a single weak (F0/F1) tornado does not necessarily imply the verification of a watch: if there are five or fewer severe wind and/or hail reports, neither a tornado watch nor a severe thunderstorm watch is verified in this case. Because the Enhanced Fujita (EF) scale was designed to be compatible with the original Fujita (F) scale (McDonald et al. 2004), any further references to a tornado damage rating level on one scale in this paper should be construed to include the other scale unless context dictates otherwise.

Watch verification has been an important area of study for the SPC and its predecessor, the National Severe Storms Forecast Center (NSSFC). For example, in the late 1970s, a journal article examining the verification of tornado watches from the 1967–77 period was published (Pearson and Weiss 1979). The NSSFC published a further review of forecast trends in 1992, including severe thunderstorm as well as tornado watches (Anthony and Leftwich 1992). Both of these verifications make use of the false alarm ratio (FAR) – the ratio of null cases to total watches (in other words, the percentage of null cases). The formula used in Pearson and Weiss (1979) and in our research is

$$\text{FAR} = z/(x+z) \quad (1)$$

where x is the number of correctly predicted (i.e., verified) events, and z the number of null cases. The FAR as formulated above functions on a strict "pass/fail" basis; a watch that almost verified is treated the same as a watch that lacked even a single convective event. Though this is a simple calculation useful for basic verification, it cannot

¹ In 2010, the criterion for severe hail was raised to 1.00 inches (25.4 mm) (http://www.spc.noaa.gov/misc/scn09-52_1inch_hail_oper.txt), but as our research utilized data up to 2008, this new definition does not apply to our research.

distinguish between different levels of forecast performance.

2. Details and Methods

In order to conduct a more robust analysis of null cases, we used a top-down decision tree approach to verify tornado watches from 2003 to 2008. This six-year period provides a large data sample of watches ($n=1901$) and is bounded by the earliest availability of archived SPC hourly environmental parameters and the latest availability of final Storm Data. Watches were classified into nine different levels of performance, or classes:

- *Class 1* – verifies according to SPC guidelines (Dean and Schaefer 2006) as tornado watch
- *Class 2* – contains one weak (F0 or F1) tornado but verifies as severe thunderstorm watch
- *Class 3* – contains no tornadoes but verifies as severe thunderstorm watch
- *Class 4* – contains one weak tornado and 1–5 hail and/or wind reports (verifies as neither a tornado watch nor a severe thunderstorm watch but contains a tornado and small number of non-tornadic reports)
- *Class 5* – contains no tornadoes but does contain 1–5 hail and/or wind reports
- *Class 6* – contains no severe reports but at least one NWS tornado warning occurred within the watch
- *Class 7* – contains no severe reports but at least one NWS severe thunderstorm warning occurred within the watch
- *Class 8* – contains no reports or warnings, but cloud-to-ground (CG) lightning occurred in the watch according to National Lightning Detection Network (NLDN) lightning data
- *Class 9* – contains no reports, warnings, or CG lightning

In addition, we also grouped these nine classes into four superclasses: Verified Tornado Watch (Class 1 only), Severe Storm Verification (Classes 2 and 3), Near Miss (4 and 5), and Full Miss (6, 7, 8, and 9). The FAR for tornado watches as defined by Pearson and Weiss (1979) is therefore the sum of the severe storm verification (SSV) percentage, near miss (NM) percentage, and full miss (FM) percentage, or more simply 100% minus the verified tornado watch (VTW) percentage. The tools and functionalities of

Microsoft Excel 2007 were used to determine which class each watch belonged to as well as to calculate pertinent verification statistics. We recognize that issues may exist with this classification scheme, and emphasize that the F-scale is employed as a damage scale, rather than an intensity scale (Doswell and Burgess 1988). In certain rare cases, this may prove to be the difference between a single F1 (Classes 2 or 4) and a single F2 (Class 1); however, data are insufficient to account for this distinction.

3. Results and Discussion

a. General

The class-by-class breakdown of the 1901 Tornado Watches issued during the 2003–08 period is shown in Table 1.

Table 1; Breakdown of tornado watches by class and superclass.

Class	Superclass	Number of watches	% of watches – class	% of watches – superclass
1	VTW	830	43.66%	43.66%
2	SSV	196	10.31%	31.04%
3	SSV	394	20.73%	
4	NM	64	3.37%	14.73%
5	NM	216	11.36%	
6	FM	89	4.68%	10.57%
7	FM	38	2.00%	
8	FM	65	3.42%	
9	FM	9	0.47%	

Nearly half of all tornado watches verified, and of those that did not, the majority would have verified as severe thunderstorm watches. The most common false alarms were classes 3 and 5, or watches that contained only non-tornadic severe weather reports. The FM rate of 10.57% is somewhat surprising, as it indicates that approximately one out of every ten tornado watches has no severe storm reports associated with it, but these data above are for the entire country, entire year, and entire diurnal period.

Only nine tornado watches qualified as a Class 9, and many of these were clustered geographically and temporally. Five occurred in either Georgia or North Carolina between October and January inclusive, primarily reflecting cool-season situations when tornado forecasting is typically more challenging, and eight of the nine were

issued between 1700 UTC and 2300 UTC, during more diurnally favorable periods but between radiosonde observation times. Of these nine watches, there were two predominant failure modes, as evaluated by their respective SPC convective outlooks and forecast discussions and the radiosonde launches immediately preceding and following the watch issuance: (1) the capping inversion was stronger than expected, suppressing development of thunderstorms, or (2) the thermodynamic setup was very marginal, with low mid-level lapse rates (generally less than 6 K km^{-1}), and high shear but low CAPE.

b. Regional differences

1) IN GENERAL

To describe more effectively differences in tornado watch performance based on geographic region of the United States, we employed regional boundaries based on those used by Thompson et al (2008). Their subdivisions were based on geographic terrain features, variations in seasonal flow regimes, and climatological variations in significant tornado environments. We found, however, that a distinct difference existed between watch performance in the Lower Mississippi River Valley (LMRV, comprising Arkansas, Mississippi, Alabama, Louisiana, and Tennessee) and those in the southeastern Atlantic Coast (SEAC, comprising Florida, Georgia, and the Carolinas), and we therefore separated Thompson et al.'s "Southeast" into these two regions. The total number of tornado watches issued in each region (determined by the centroid of each watch) is given in Figure 1. As one might expect, the Southern and Northern Plains had a large number

of tornado watches, as did the two regions of the Southeast and the Midwest. Far fewer tornado watches were issued over the Northeast and Rockies regions during the period. The West had only one watch issued over the entire six-year period (a Class 5) and was excluded from statistical calculations and comparisons owing to the small sample size.

Most VTW rates were in the 40–50% range except for the Northeast, which had a VTW rate of 31.94% (Table A1). This low success rate can perhaps be partially attributed to the comparatively low frequency of tornado occurrence and relative scarcity of tornado watches in this area ($n=72$ or an average of 12 yr^{-1}). Of greater concern is the relatively high FM rate for the Southeast Atlantic Coast, where more than 25% of watches did not contain a severe report. Five of the nine Class 9 events occurred within the region, and it had the highest percentages of Class 4, 6, and 8 watches (Table A1). In particular, Class 6 events (no severe storm reports but an NWS tornado warning issued for the area) were more frequent in the Southeast. This is consistent with the prevalence of marginally favorable tornadic environments [low CAPE and high shear (Schneider and Dean 2008)] over this region, which creates additional challenges for forecasting and warning of tornadoes. Finally, it is seen that the SSV is higher than the VTW rate over the Rockies, Midwest, and Northeast. The reasons for this result are not known at this time.

2) BY SEASON

Tornado watches from regions with more than 100 total tornado watches [the LMRV, SEAC, Midwest (MW), Northern Plains (NP), and Southern Plains (SP)] were stratified by meteorological season. Most tornado watches were issued during the spring (March–May), the traditional season for tornadoes, except over the NP when most occurred during the summer. No tornado watches at all were issued for the NP during winter throughout this six-year period (Table A2). Interestingly, the distribution of watches by season in the SP, MW, and LMRV regions are dominated by a spring peak, although a secondary peak is evident in the LMRV in the fall. In the SEAC, however, there is a more uniform distribution of tornado watches throughout the year. Spring also appears to be the season in which the verification statistics are best, as FM rates were typically lowest and VTW rates generally the highest (Fig. 2). The exception is over the LMRV and the

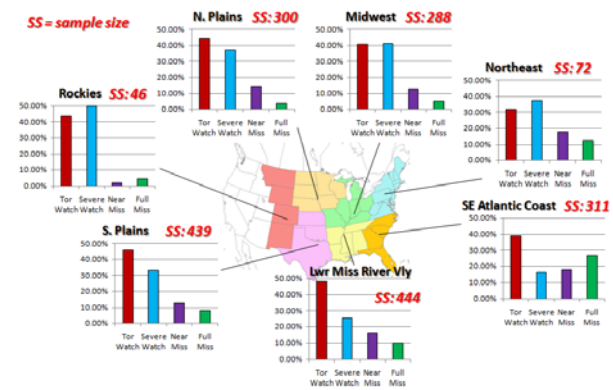


Figure 1: Verification of tornado watches by region, with the sample size of each included. VTWs are in red; SSVs, blue; NMs, purple; and FMs, green

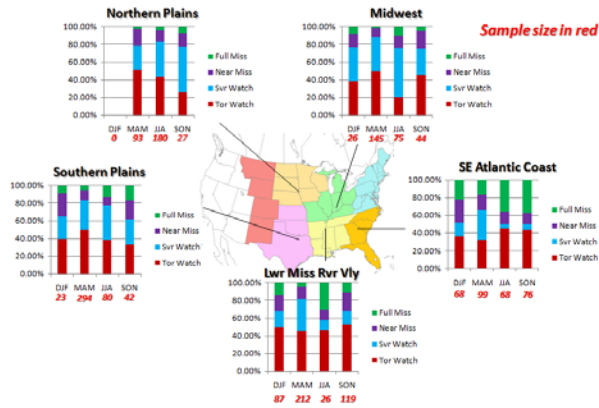


Figure 2: Verification breakdown by meteorological season for the five regions with more than 100 watches over the period. DJF means December-January-February, MAM March-April-May, JJA June-July-August, and SON September-October-November.

SEAC where VTW is not maximized in the spring. Conversely, the autumn months exhibit lower verification skill, especially over the NP and SP where VTW rates were less than 35%. The small tornado watch sample size probably influences this result, however, as fewer than 10% (69 out of 739) of all tornado watches in the plains were issued in the fall. In addition, nine of the 65 total class-8 events as well as three of the nine class-9 events happened in the SEAC during autumn. These were likely associated primarily with land falling tropical cyclones. Tornado Watches during the summer months also had a rather high FM rate (>10%) outside of the NP, peaking at 35.3% in the SEAC (Table A2), suggesting that environmental conditions extant during the summer and fall months contribute to greater frequency of false alarm watches.

c. Temporal differences

The time of day when a tornado watch was issued did appear to make a difference in the verification rate. Watches with onsets in the morning or afternoon (roughly 1200 UTC through 2359 UTC) had higher VTW rates and lower FM rates than watches with onsets in the evening or overnight hours (0000–1159 UTC) (Fig. 3, see also Table A3). There is a marked decrease in VTW from the 2100 UTC hour to the 2200 UTC hour, a time period that only correlates with sunset in the NE in late autumn and early winter, but which does match 1600 Central Standard Time and 1700 Central Daylight Time. This reduction in VTW may be related to the issuance time during a forecast shift, which is discussed below. An unexpected peak occurs during the 1000 UTC hour, but this is

likely an artifact of the small sample size, as this hour had the lowest number of watches ($n=24$), and is therefore the most susceptible to statistical abnormalities. These watches were not, however, for only a few weather systems, because if they were, any statistical aberrations would likely be more pronounced.

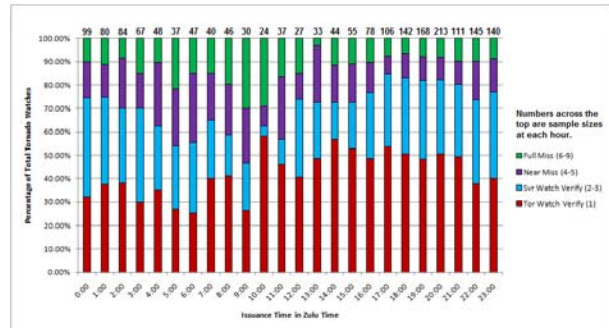


Figure 3: Verification breakdown by issuance time of Tornado Watch, hours in UTC. A watch issued at, e.g., 1550 UTC would be in the "15:00" column.

The SPC's watch product responsibilities are conducted in three work shifts, determined by local (i.e., Central) time and with adjustments made for Daylight Saving Time (DST): the day shift (0800–1600 LT), the swing shift (1600–0000 LT), and the midnight shift (0000–0800 LT). The day shift had a VTW rate more than 10% higher than the swing and midnight shifts, and the midnight shift had a higher FM rate (Fig. 4). Despite the lower VTW rate and higher NM and FM rates, however, there are meteorological and non-meteorological factors that may contribute to this difference. The midnight shift operates during a time of day when

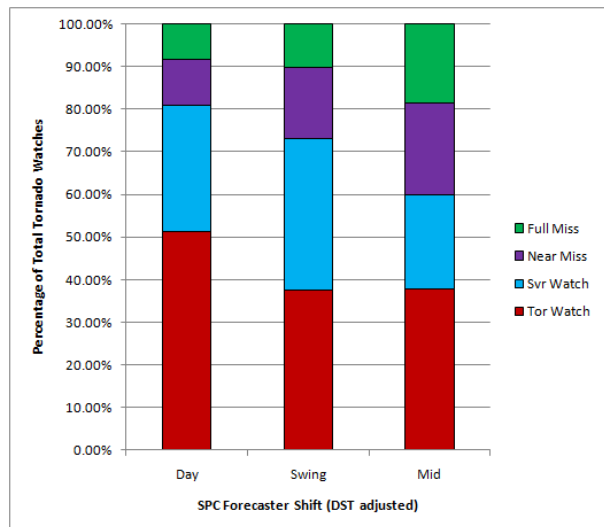


Figure 4: Verification breakdown by SPC forecaster shift, adjusted for Daylight Saving Time. "Mid" means the midnight shift.

convective activity is typically approaching or at its minimum, as implied by the small sample sizes displayed in Fig. 3 for these hours. Thus, the statistics may be skewed somewhat by the more limited number of severe weather opportunities on the midnight shift. Furthermore, because verification is predicated on accurate identification and reporting of storms, a human element is involved. Watches issued by the midnight shift are likely to be valid in the late-night or early morning hours when many people are likely to be asleep during much of this time period, and darkness makes it less likely that anyone awake will see tornadoes clearly. Finally, a minor but noticeable improvement was evident as each shift progressed. Watches issued during the final two hours of each shift, on average, had a 5% higher VTW rate than watches issued during the first two hours of each shift, and SSV, NM, and FM rates were all lower toward the end of the shift compared to the beginning (Fig. 5). This suggests that as forecasters monitor the evolution of the atmosphere consistently for several hours, the accuracy of their decision-making improves slightly but noticeably.

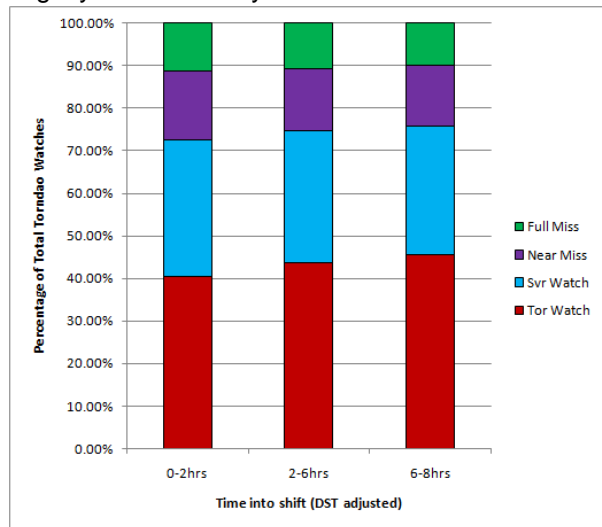


Figure 5: Verification breakdown by amount of time into the SPC forecaster shifts; i.e., "0-2 hrs" refers to the first two hours of each shift (0800-1000 LT for the day shift, 1600-1800 LT for the swing shift, and 0000-0200 LT for the midnight shift).

d. "Particularly dangerous situations"

The SPC distinguishes between "particularly dangerous situation" (PDS) tornado watches, characterized by a strong expectation of F2+ tornado occurrences, and non-PDS tornado watches. The PDS label is used sparingly; of the

1901 watches that were issued during the time period of study, only 151 (7.9%) were PDS watches. The failure rates of PDS watches were very low compared to tornado watches as a whole. Almost 80% verified as a tornado watch, and of those that did *not* verify, over 80% (27 of 33) were SSVs (Fig. 6). Only one PDS watch was a full miss, and it was a class 6; no PDS watches in the entire six-year period qualified as a class 7, 8, or 9. Since PDS watches are characterized by a higher probability of intense, highly destructive tornadoes, a much lower failure rate is expected, and verification results confirm this with only 4% of all PDS watches either near or full misses.

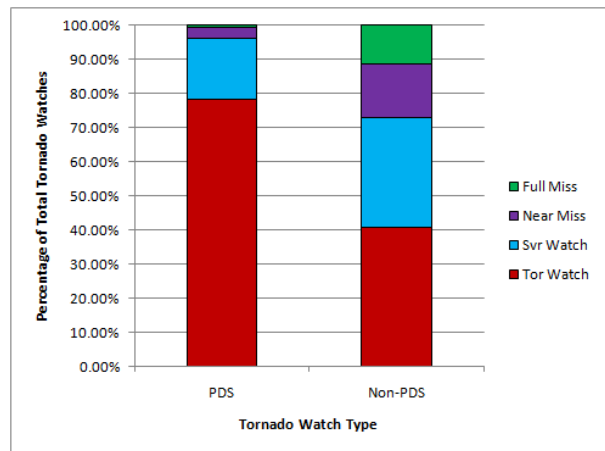


Figure 6: Verification breakdown by status of watch as a PDS.

4. Conclusions

Tornado watches officially verified by the occurrence of 2 or more tornadoes or at least 1 F2+ tornado comprised approximately 44% of all Tornado Watches, with full misses (no severe reports) at about 10%. Verification rates were generally best for watches issued in the spring and in the daylight hours, improving somewhat towards the end of each work shift. Not surprisingly, PDS tornado watches had a much higher success rate than non-PDS tornado watches. Failure rates were higher in the east-coast states, in particular the southeast Atlantic coast states, partly because of the lower frequency of tornado occurrence and higher frequency of marginal but still supportive environments for tornadogenesis. We recommend that additional research be undertaken with particular emphasis on these aspects of tornado forecasting. Research of the effects of various environmental parameters using these watch verification categories may prove useful as well.

5. References

- Anthony, R. W., and P. W. Leftwich, 1992: Trends in severe local storm watch verification at the National Severe Storms Forecast Center. *Wea. Forecasting*, **7**, 613–622.
- Dean, A. R. and J. T. Schaefer, 2006: The skill of forecasting relatively isolated severe thunderstorm events. Preprints, *18th Conf. Probability and Statistics in the Atmos. Sciences*. Atlanta, GA, Amer. Meteor. Soc.
- Donner, W. R., 2007: The political ecology of disaster: an analysis of factors influencing US tornado fatalities and injuries. *Demography*, **44**, 669–685.
- Doswell, C. A., and D. W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.*, **116**, 495–501.
- Doswell, C. A., A. R. Moller, and H. E. Brooks, 1999: Storm spotting and public awareness since the first tornado forecasts of 1948. *Wea. Forecasting*, **14**, 544–557.
- Edwards, R., 2008: Tropical cyclone tornadoes -- a research and forecasting overview. Part 1: Climatologies, distribution and forecast concepts. Preprints, *24th Conf. Severe Local Storms*, Savannah, GA, 7A.1. Amer. Meteor. Soc.
- McDonald, J. R., G. S. Forbes, and T. P. Marshall, 2004: The enhanced Fujita (EF) scale. Preprints, *22nd Conf. Severe Local Storms*, Hyannis, MA. Amer. Meteor. Soc.
- Pearson, A., and S. J. Weiss, 1979: Some trends in forecast skill at the National Severe Storms Forecast Center. *Bull. Amer. Meteor. Soc.*, **60**, 319–326.
- Thompson, R. L., J. S. Grams, and J. A. Prentice, 2008: Synoptic environments and convective modes associated with significant tornadoes in the contiguous United States. Preprints, *24th Conf. Severe Local Storms*. Savannah, GA. Amer. Meteor. Soc.

Appendix: Extended Data Tables

Table A1: Regional Breakdown by Class

Class	Midwest	N. Plains	Northeast	Rockies	S. Plains	LMRV	SEAC
1 (VTW)	117 (40.63%)	133 (44.33%)	23 (31.94%)	20 (43.48%)	201 (45.79%)	215 (48.42%)	121 (38.91%)
2 (SSV)	44 (15.28%)	37 (12.33%)	8 (11.11%)	10 (21.74%)	50 (11.39%)	33 (7.43%)	14 (4.50%)
3 (SSV)	75 (26.04%)	74 (24.67%)	19 (26.39%)	13 (28.26%)	96 (21.87%)	80 (18.02%)	37 (11.90%)
4 (NM)	5 (1.74%)	8 (2.67%)	2 (2.78%)	0 (0.00%)	11 (2.51%)	20 (4.50%)	18 (5.79%)
5 (NM)	32 (11.11%)	36 (12.00%)	11 (15.28%)	1 (2.17%)	45 (10.25%)	52 (11.71%)	38 (12.22%)
6 (FM)	5 (1.74%)	1 (0.33%)	3 (4.17%)	0 (0.00%)	11 (2.51%)	25 (5.63%)	44 (14.15%)
7 (FM)	0 (0.00%)	3 (1.00%)	5 (6.94%)	1 (2.17%)	10 (2.28%)	10 (2.25%)	9 (2.89%)
8 (FM)	8 (2.78%)	7 (2.33%)	1 (1.39%)	1 (2.17%)	14 (3.19%)	9 (2.03%)	25 (8.04%)
9 (FM)	2 (0.69%)	1 (0.33%)	0 (0.00%)	0 (0.00%)	1 (0.23%)	0 (0.00%)	5 (1.61%)
Total	288	300	72	46	439	444	311

Table A2: Seasonal Breakdown by Class

Class	Total	1	2	3	4	5	6	7	8	9
NP DJF	0	0	0	0	0	0	0	0	0	0
NP MAM	93	48	12	13	3	15	0	0	2	0
NP JJA	180	78	20	52	4	18	0	3	4	1
NP SON	27	7	5	9	1	3	1	0	1	0
SP DJF	23	9	2	4	1	5	0	1	1	0
SP MAM	294	147	33	64	8	26	2	4	9	1
SP JJA	80	31	11	20	2	5	4	3	4	0
SP SON	42	14	4	8	0	9	5	2	0	0
MW DJF	26	10	3	7	1	3	2	0	0	0
MW MAM	145	72	19	37	2	12	0	0	3	0
MW JJA	75	15	19	23	1	9	2	0	4	2
MW SON	44	20	5	8	1	8	1	0	1	0
LMRV DJF	87	43	4	12	4	12	6	3	3	0
LMRV MAM	212	97	26	50	6	22	5	3	3	0
LMRV JJA	26	12	0	3	2	1	7	0	1	0
LMRV SON	119	63	3	15	8	17	7	4	2	0
SEAC DJF	68	25	2	8	7	11	6	3	4	2
SEAC MAM	99	32	9	24	4	14	6	4	6	0
SEAC JJA	68	31	2	1	4	6	17	1	6	0
SEAC SON	76	33	1	4	3	7	15	1	9	3

Table A3: Time of Day Breakdown by Class

UTC hour	Total	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
00	99	32	10	32	4	11	3	3	4	0
01	80	30	5	25	2	9	3	3	3	0
02	84	32	8	19	3	15	5	1	1	0
03	67	20	5	22	1	9	5	1	4	0
04	48	17	3	10	2	11	2	1	2	0
05	37	10	2	8	2	7	5	1	2	0
06	47	12	5	9	4	10	5	0	2	0
07	40	16	5	5	2	6	3	1	1	1
08	46	19	2	6	0	10	4	1	4	0
09	30	8	0	6	0	7	4	3	2	0
10	24	14	0	1	0	2	3	3	1	0
11	37	17	1	3	1	9	5	0	1	0
12	27	11	3	6	1	2	2	0	2	0
13	33	16	1	7	4	4	0	1	0	0
14	44	25	1	6	4	3	3	1	1	0
15	55	29	3	8	3	6	3	0	3	0
16	78	38	8	14	1	9	4	0	4	0
17	106	57	16	17	0	8	4	0	2	2
18	142	72	20	26	4	11	2	2	3	2
19	168	81	20	37	5	12	6	2	5	0
20	213	108	29	38	8	13	7	3	5	2
21	111	55	17	17	2	9	4	4	2	1
22	145	55	11	41	7	17	5	3	5	1
23	140	56	21	31	4	16	2	4	6	0