828 THE QUALITY CONTROL ALGORITHMS USED IN THE PROCESS OF CREATING THE NASA KENNEDY SPACE CENTER LIGHTNING PROTECTION SYSTEM TOWERS METEOROLOGICAL DATABASE

John M. Orcutt
Jacobs ESSSA Group, Huntsville, AL

James C. Brenton
Jacobs ESSSA Group, Huntsville, AL

1. INTRODUCTION

The methodology and the results of the quality control (QC) process of the meteorological data from the Lightning Protection System (LPS) towers located at Kennedy Space Center (KSC) launch complex 39B (LC-39B) are documented in this paper. Meteorological data are used to design a launch vehicle, determine operational constraints, and to apply defined constraints on day-of-launch (DOL). In order to properly accomplish these tasks, a representative climatological database of meteorological records is needed because the database needs to represent the climate the vehicle will encounter.

Numerous meteorological measurement towers exist at KSC; however, the engineering tasks need measurements at specific heights, some of which can only be provided by a few towers. Other than the LPS towers, Tower 313 is the only tower that provides observations up to 150 m. This tower is located approximately 3.5 km from LC-39B.

In addition, data need to be QC'ed to remove erroneous reports that could pollute the results of an engineering analysis, mislead the development of operational constraints, or provide a false image of the atmosphere at the tower's location.

2. LPS TOWER DATA CHARACTERISTICS

The three towers that comprise the LPS network are located around LC-39B at KSC. Each tower directly measures temperature (T), relative humidity (RH), mean wind speed (MWS), and wind direction (WD) in one-second intervals at four levels (40, 78, 116, and 139 m). One-minute averages are generated from these one-second data, and retrieved by the Marshall Space Flight Center Natural Environments Branch (MSFC NE). In the creation of the one-minute file, the one-second maximum wind speed and corresponding direction are used for the one-minute peak wind speed (PWS) and direction (PWD). The one-minute dew point temperature (T_d) observation is derived from the one-minute T and RH observations. The one-minute database contains records of T and T_d, to the nearest tenth of a degree Celsius; mean and peak wind speed, to the nearest tenth of a knot; mean and peak wind direction to the nearest whole degree east from north; and RH, to the nearest whole percent. MSFC NE received the LPS data in one-minute, non-consecutive intervals for the 1 January 2011 – 30 April 2015 period of record (POR). Frequent gaps exist in this database because the towers were not operationally accepted until spring of 2015.

*Corresponding author address: John M. Orcutt., Jacobs ESSSA Group, 1500 Perimeter Pkwy., Huntsville, AL 35806; email: john.m.orcutt@nasa.gov.

3. THE QUALITY CONTROL PROCESS

The methodology used is similar to QC procedures implemented on other tower databases by both the Applied Meteorological Unit (AMU) (Lambert 2002) and MSFC NE (Barbré 2008, Barbré 2014, Decker 2008) but has been customized for this application. The first portion of the QC routine consists of independently examining each sensor. This portion includes an unrealistic data check, a check on tower obstruction of wind measurements, and a temporal consistency check. The next portion of the QC routine consists of sensor-to-sensor checks, where each sensor at the same height or tower is compared with the surrounding sensors at that height or tower. This portion includes a data hang-up check, a climatological check, a horizontal sensor-to-sensor check, and a vertical sensor-to-sensor check. Data are manually examined, after each check to determine if the QC procedure is effectively flagging erroneous data. After each check, T_d is removed if either T or RH are missing, RH is removed when T_d are missing, and all wind observations are removed if any wind observation at that tower and height are missing. Finally, an up-wind tower is selected to help remove outside influences on the wind observations.

The unrealistic data check implements thresholds to remove data that either physically cannot exist or data that exist outside of KSC's climatology. Selecting criteria for the T and T_d ranges, as well as the maximum wind speed comprised of first examining the non-QC'ed data and then examining previous QC algorithms (Barbré 2008, Barbré 2014, Decker 2008, Lambert 2002). Data are removed if any of the following criteria were not met:

- $-26.0\ ^\circ C < T < 40.5\ ^\circ C$
- $-18.0\ ^\circ C < T_d < 35.0\ ^\circ C$
- $T_d < T$
- $0.0\ \% < RH \leq 100.0\ \%$
- $0\ m/s < MWS \leq 60.0\ m/s$
- $0\ m/s < PWS \leq 69.5\ m/s$
- $0^\circ <$ Wind Direction $< 360^\circ$
- MWS < PWS

Next, each wind sensor is checked for obstruction by the tower. This check removes instances where the wind blows through the tower. The instruments are situated on booms that point out from the tower. For Tower 1, the boom is located at 79º from north; Tower 2, 219º; and Tower 3, 323º. The boom locations are illustrated in Figure 1. Wind observations are removed if the WD is found to be in the following range of angles:

- Tower 1: 229˚ - 289˚
- Tower 2: 9˚ - 69˚
- Tower 3: 113˚ - 173˚

Finally, each T, T_d, and wind sensor is checked for temporal consistency. The difference from the mean of the surrounding hour (ΔT, ΔT_d) is calculated for T and T_d. For the wind data, the temporal consistency is evaluated by examining the vector wind difference (ΔV) immediately before and after the observation that is being tested. MWS are used to calculate ΔV. This check is designed to remove any spurious spikes and dips in the data; however, after a manual QC of the data, the temporal consistency check was removed from the QC process. Manual QC found the temporal consistency check to consistently remove valid peaks and spikes (especially from the PWS and MWS observations) while not removing obviously erroneous data. In addition, all invalid data that the temporal consistency check removes is removed by subsequent checks.

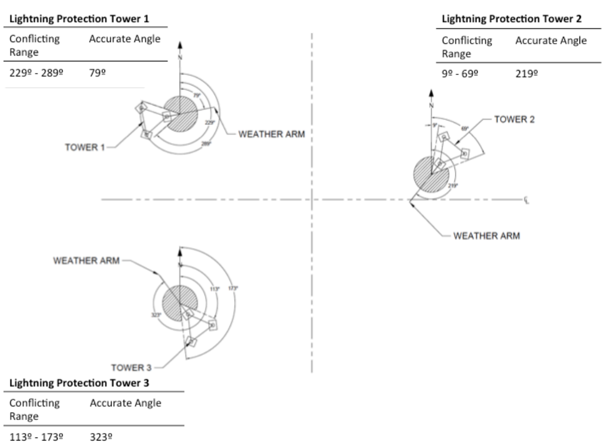

Figure 1: The angles of the instrument booms and the conflicting ranges for each LPS tower. The conflicting range is the set of angles where the wind would have to blow through the tower to reach the anemometer.

After the previous individual sensor checks are completed, sensor-to-sensor checks are performed. The first sensor-to-sensor check, the data hang-up check, finds periods of time where a sensor reports a constant measurement for more than 30 consecutive observations that do not coincide with observations from other sensors. First, consecutive data lasting for longer than 30 minutes are flagged, and are compared against the two other sensors at the given altitude. Then, T and T_d are removed if the magnitude of the difference is greater than 0.3 °C. Wind reports are removed if the magnitude of the difference of the MWS exceeds 0.6 m/s. The threshold for T and T_d is determined by examining cases where the data hang-up check removes data, and the threshold for mean wind speeds is a heritage value from previous work (Barbré 2008, Barbré 2014).

Next, each observation is checked against the mean and standard deviation of the given parameter for the given month and hour. The mean and standard deviation of T, T_d, MWS, and PWS for each month and each hour of each month is calculated. T and T_d observations are removed if the observation is outside of the mean ± five standard deviations for the variable during a given month and hour. MWS and PWS observations are removed if the observation is outside of the mean ± 10 standard deviations during a given month and hour. The thresholds used are heritage values (Barbré 2008 Barbré 2014 Lambert 2002). Figure 2 shows an example of the average MWS and PWS with ± one standard deviation for the month of April at the 139 m level.

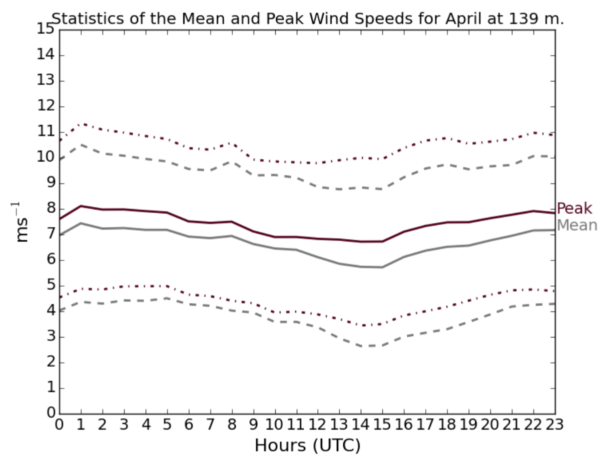

Figure 2: The average of the peak (maroon) and mean (grey) wind speed for each hour in April at the 139 m level. The solid lines represent the hourly and monthly average of the PWS and MWS. The maroon dot-dashed line is ± one standard deviation from the mean of the PWS, and the grey dot-dashed line is ± one standard deviation from the mean of the MWS.

After the climatological check, a horizontal sensor-to-sensor comparison is performed. In this check each sensor is compared to the other two sensors at the same level. This check is done at all elevations. Cumulative distributions of differences are examined. Thresholds have been derived from an examination of the distributions, an examination of the time series plots, and previous work (Barbré 2008, Barbré 2014). An example of a cumulative distribution examined to determine the thresholds is shown in Figure 3. Data are removed if the following criteria are not met:

- $|\Delta T|$, $|\Delta T_d| < 4.0$ ºC
- $|\Delta RH| < 10$ %
- $|\Delta MWS| < 5.0$ m/s

Observations are removed from the sensor reporting the erroneous data if all three towers are reporting at that time. Otherwise if one sensor is missing, both sensors being compared are removed, as it is not possible to determine which sensor is “correct”. If two sensors are missing, the difference cannot be determined, and no data are removed.

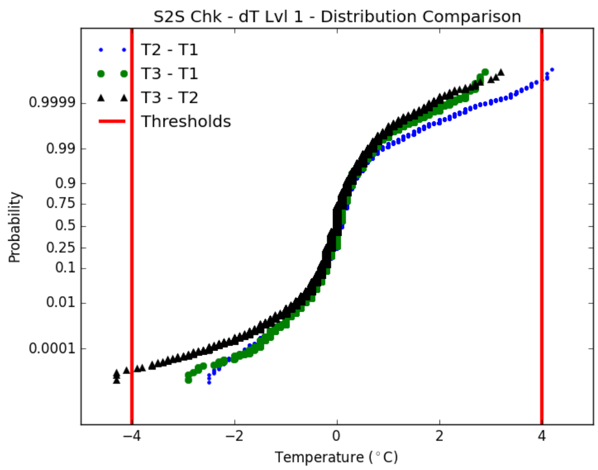

Figure 3: Cumulative distribution of the temperature differences of adjacent sensors at the 40 m level. Red bars represent the heritage thresholds used to remove data.

The final check done on the data is a vertical sensor-to-sensor check. In this check, data are compared against the vertically adjacent sensors. This check is only performed on the MWS and T data from the middle two sensors of each tower. First, the mean wind vector differences and T differences are computed from the average parameter of the vertically adjacent sensors. Cumulative distributions of the differences are plotted and examined to determine thresholds used to remove data.

- $|\Delta MWS| < 5.0$ m/s if the $|\Delta MWS|$ from one of the other towers exceeds 3.0 m/s
- -1.5 ºC $< \Delta T < 2.5$ ºC if the $|\Delta T|$ from one of the other towers exceeds 1.0 °C

An example of the cumulative distribution of the T difference is shown in Figure 4. Note that Figure 4 shows only the threshold, and not any other dependencies that are taken into account when determining the validity of data. This explains why the tails of Tower 1 and 3 appear to be removed by the threshold, but they are not removed unless the T condition is also violated. Tower 2 Sensor 3 had several periods of erratic measurements, which explains the large tail seen in the distribution in Figure 4. Thus this check implements thresholds different than previous work.

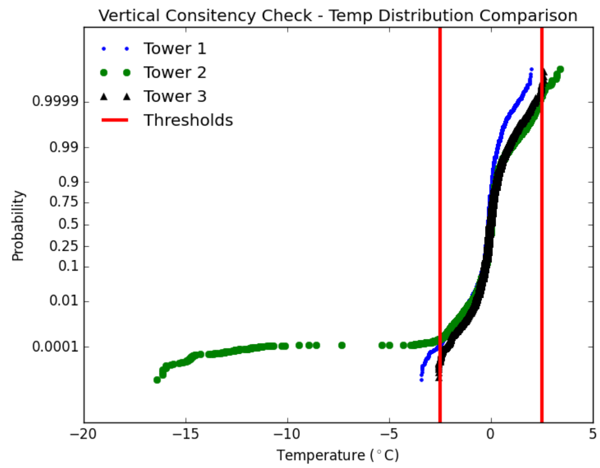

Figure 4: Distribution comparison of T from the vertical sensor-to-sensor check. The red bars represent the thresholds used to determine the validity of data.

After the previous QC checks are completed, an exclusive up-wind sensor is defined for each altitude. The overall methodology in determining the up-wind tower is nearly identical to the methodology used in Barbré 2008 and Barbré 2014. This procedure is implemented on the mean wind if the MWS is considered to exceed “light and variable” (i.e. wind speeds ≥ 3.0 m/s) from any tower at a given height and timestamp. The booms that hold the meteorological instruments away from the tower face different directions on each tower. Limits that dictate which tower can be used as the up-wind tower are determined by engineering judgment and examining the effect of turbulence from other towers and objects on the wind measurements (Lubitz 2008). The limits for each tower used as the up-wind tower are as follows:

- 0° ≤ Tower 1 < 140°
- 140° ≤ Tower 2 < 280°
- 244° ≤ Tower 3 < 360°

If the up-wind tower did not report, but the winds are within the overlapping up-wind sector from a down-wind tower, then the wind report from the down-wind tower is used. If two towers exist that could be used as the up-wind tower, the most recent exclusive up-wind tower is used. The most recent exclusive up-wind tower is also used if there is only one tower that reports, and if the maximum wind speed does not exceed 3.0 m/s. Finally, any wind speed observation that has a wind vector shear value greater than 0.2 s^{-1} is removed. The T, T_d, and RH of the up-wind tower consist of the mean values from all available towers at each timestamp and height.

4. ATTRIBUTES OF THE QUALITY CONTROLLED DATABASE

In the POR examined in this study, 2,276,280 timestamps exist at each of the 12 sensors, which results a total of 27,315,360 potential observations of each variable. The total number of potential observations, number of missing observations, percent of missing observations, the number of available observations, and the percent of available observations are shown in Table 1. Of the potential 27,315,360 observations, approximately 60 – 68% (8.8 M – 10.4 M) of the observations are missing, depending on the variable. T has the most available observations (10,440,488) and MWS has the least (8,836,399). The majority of the missing data is from 2013. More data exists in 2011, 2012, and 2014, but 2015 is the most temporally complete year. One should note that the LPS towers have only been operationally accepted since the spring of 2015. Before operational acceptance, the towers were undergoing testing and troubleshooting, which resulted in the poor data availability.

The entire QC process removes between 0.9% to 11.9% of available data with T having the least amount of data removed, and PWD having the most removed. The unrealistic data check removes 0.03% of T, 0.01% of T_d, 7.2% of RH, and between 3.5% and 10.6% of the wind observations. The unrealistic check removes such a large amount of wind observations as it includes the initial removing of wind data when any other wind variable is initially missing. The obstructing tower check removes 0.02% of wind observations. No thermodynamic observations are removed by this check as it is only performed on wind observations. No data are removed by the temporal consistency check, as this check is not implemented due to it eliminating too much valid data regardless of thresholds used. RH and T_d have the same number of observations removed from the data hang-up check through the sensor-to-sensor checks because if T_d fails a QC check, then RH is removed too, and vice versa. The data hang-up check removes 0.8% of T, 1.7% of T_d, 1.5% of RH, and 0.1% of all wind observations. The climatological check removes 0.3% of T, 0.2% of T_d, 0.1% of RH, and a negligible amount from any wind observation. The horizontal sensor-to-sensor check removes a negligible amount of T, 0.3% of T_d and RH, and approximately 1.2% of wind observations. The last check, the vertical sensor-to-sensor check, removes a negligible amount of data from any variable. After the QC process is performed, the database contains 99.1% of the initial available T observations, 97.9% of available T_d observations, 90.8% of RH observations, 95.2% of WS, 94.1% of WD, 92.9% of PWS, and 88.1% of PWD relative to the original sample. The up-wind tower archive consists of 3,880,550 T observations, 3,634,361 T_d and RH observations, and 2,410,077 wind observations. The shear check of the up-wind tower removes 0.2% of the wind observations. The number and percent removed from each check, and the number and percent of data remaining for each variable is shown in Table 2.

While this bulk data retention information is useful, for engineering purposes it is more useful to examine the data retention at each sensor and month; as analyses performed by MSFC NE and others are generally done with respect to a particular month (or months) and height. For all variables, at most heights and towers, all months retain over 85 % of the data. Tower 3 Sensor 2 did not record any T or T_d observations in November, and had a long standing error with the wind speed sensor. As a reminder, the towers had not been operationally accepted until spring of 2015. Neither altitude nor month appears to be a factor in determining how much data are retained.

Table 1: The number of potential observations, number of missing observations, percent of missing observations, number of available observations, and the percent of available observations for each variable measured on the LPS network at all heights and towers.

	T	T_d	WS	WD	PWS	PWD	RH
Time Stamps	2276280	2276280	2276280	2276280	2276280	2276280	2276280
# of Sensors	12	12	12	12	12	12	12
# of Potential Observations	27315360	27315360	27315360	27315360	27315360	27315360	27315360
# Missing	16874872	17672268	18478961	18375500	18259066	17773728	16926322
% Missing	61.8	64.7	67.7	67.3	66.8	65.1	62.0
# Available	10440488	9643092	8836399	8939860	9056294	9541632	10389038
% Available	38.2	35.3	32.3	32.7	33.2	34.9	38.0

Table 2: The number and percentage of the number of available observations that were removed in each QC check for each variable on the LPS network at all heights and towers.

		T	T_d	WS	WD	PWS	PWD	RH
	# Available	10440488	9643092	8836399	8939860	9056294	9541632	10389038
Unrealistic Data Check	# Removed	2812	1178	305883	409344	525778	1011116	747124
	% Removed	0.03	0.01	3.5	4.6	5.8	10.6	7.2
Conflicting with Tower	# Removed	0	0	1496	1496	1496	1496	0
	% Removed	0.00	0.00	0.02	0.02	0.02	0.02	0.00
Temporal Consistency Check	# Removed	0	0	0	0	0	0	0
	% Removed	0	0	0	0	0	0	0
Data Hang-Up Check	# Removed	83776	159995	11247	11247	11247	11247	159995
	% Removed	0.8	1.7	0.1	0.1	0.1	0.1	1.5
Climatological Check	# Removed	3134	14697	2	2	2	2	14697
	% Removed	0.03	0.2	0.00	0.00	0.00	0.00	0.1
Horizontal Sensor-to-Sensor Check	# Removed	2752	30314	106826	106826	106826	106826	30314
	% Removed	0.0	0.3	1.2	1.2	1.2	1.1	0.3
Vertical Sensor-to-Sensor Check	# Removed	370	222	0	0	0	0	222
	% Removed	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total Removed	# Removed	92844	206406	425454	528915	645349	1130687	952352
	% Removed	0.9	2.1	4.8	5.9	7.1	11.9	9.2
Total Available	# Available	10347644	9436686	8410945	8410945	8410945	8410945	9436686
	% Available	99.1	97.9	95.2	94.1	92.9	88.1	90.8

5. CONCLUSION

MSFC NE constructed a database from the meteorological instrumentation on the KSC LPS tower network for the POR of January 2011 – April 2015. This database contains one-minute records of T, T_d, RH, and mean and peak WS and WD. Extensive QC algorithms are applied to the database to ensure that invalid and incorrect observations are removed. Manual QC analysis results indicate that the temporal consistency check consistently removes the peaks of real wind events, and often misses erroneous spikes in the data, regardless of the thresholds used. Thus the temporal consistency check is removed from the QC routine for the LPS tower database. After the QC algorithms are completed, an up-wind tower is selected. In total, 99.1% of T observations, 97.9% of T_d observations, between 88.1% and 95.2% of wind measurements, and 90.8% of RH measurements remain after this QC process is implemented. However, these percentages vary per month, sensor, and tower. Most sensors have at least 85% availability during all months. Data files exist for all QC'ed data from each of the LPS towers as well as the up-wind tower and all data files have been archived.

Performing this QC routine reiterated several important considerations that must be taken into account when performing any QC process. Most automated QC checks implement subjectively derived thresholds as objective criteria. However, as shown in this paper, this can lead to problems. Examining the data during and after the QC process to aid in the determination of thresholds and which algorithms are even necessary is extremely important. In addition, one must accept that it is very difficult if not impossible to set a threshold that ensures removal of all erroneous data while retaining all valid data. As such, one of two general philosophies must be followed during the entire QC process: keep as much valid data as possible, but allow some erroneous data to exist, or remove as much erroneous data as possible, but also risk removing valid data. In the creation of this database, the philosophy of keeping as much valid data as possible, but allowing some erroneous data to exist was followed. One of the two philosophies must be used each time a new database is created, and it must be followed throughout its creation. A final important note exists in that while the methodology of this QC process is similar to QC processes performed on other meteorological towers at KSC, algorithms have been tailored to the LPS tower database. As is, this QC process is not appropriate to be used with another database, and it should not be expected to produce similar results if used on another database.

6. Acknowledgements

Many colleagues provided invaluable insight and contributions to this report. James Brenton (Jacobs ESSSA/MSFC NE) collected and processed the one-minute LPS tower data files used in this report from MIDDS. Ryan Decker (MSFC NE), Kevin Decker (KSC), Tatiana Bonilla (KSC), the AMU, and the 45th Weather Squadron of the United States Air Force found or provided data that was missing from the database MSFC NE received. BJ Barbré (Jacobs ESSSA/MSFC NE) compiled data received by other sources into MSFC NE's format. Much appreciation goes to these and the many others for their invaluable contributions to this report. This work was performed under NASA contract MSFC-NNM12AA41C.

7. References

Barbré, R.E., 2008: Quality Control Algorithms Used for the KSC Tower 313 Database. Jacobs ESTS Group Analysis Report. ESTSG-FY08-1481.

Barbré, R.E., 2014: Readme File for MSFC NE's Lightning Protection Tower Quality Controlled Database. Jacobs ESSSA Group Analysis Report. ESSSA-FY15-309.

Decker, R. K., 2008: Kennedy Space Center Launch Complex 39 Meteorological Databases. NASA/MSFC/MSFC NE. Presentation to the Space Shuttle Program Natural Environments Panel.

Lambert, W. C., 2002: Statistical Short-Range Guidance for Peak Wind Speed Forecasts on Kennedy Space Center / Cape Canaveral Air Force Station: Phase 1 Results. NASA / Applied Meteorological Unit. NASA Contractor Report NASA/CR-2002-21180.

Lubitz, W. D., 2008: Effects of Tower Shadowing on Anemometer Data. In proceedings of 11th Americas conference on Wind Engineering.