Automated Extraction of Wind Data from Archived ASOS (Automated Surface Observing System) Weather Reports

Authors:

Franklin T. Lombardo, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8611, Gaithersburg, MD 20899-8611, franklin.lombardo@nist.gov
Joseph A. Main, National Institute of Standards and Technology, joseph.main@nist.gov

ABSTRACT

Wind loads for use in structural design are based in part on extreme value analysis of historical wind speed data for the location of interest. In some cases wind direction data are also required to characterize the directionality of the local wind climate. The Automated Surface Observing System (ASOS), a network of about 1000 standardized weather stations throughout the United States, is a good source of such local wind speed and direction data. In order to facilitate more widespread use of ASOS wind data for structural engineering purposes, this paper presents procedures and software for (a) extraction of peak gust wind data from archived ASOS reports, (b) extraction of thunderstorm observations from archived ASOS reports, (c) classification of wind data as thunderstorm or non-thunderstorm to enable separate statistical analysis of these distinct types of winds, and (d) construction of data sets separated by specified minimum time intervals to ensure statistical independence. These procedures are illustrated using ASOS data from three stations near New York City over a period of about 20 years.

INTRODUCTION

Wind loads for use in structural design are typically computed using design wind speeds obtained through extreme value analysis of historical wind speed data at the location of interest. The wind speed map in the American Society of Civil Engineers (ASCE) Standard 7-05 [1], for example, is based on extreme value analysis of wind speed data from 487 stations in the United States, grouped into "superstations" [2]. In responding to criticisms of this wind speed map, Peterka and Shahid [3] suggested that future analyses aimed at improving the wind speed map should focus on inclusion of additional wind data that have become available over the past 12 years, including data from new stations in locations without prior coverage. A promising source of such additional wind data is the Automated Surface Observing System (ASOS), a network of about 1000 weather stations throughout the United States that were largely automated in the 1990s [4].

Peterka and Shahid [3] also suggest that future efforts to improve the ASCE 7 wind speed map should consider separation of data into thunderstorm and non-thunderstorm winds to enable more accurate assessment of extreme wind speeds with long return periods. As pointed out by Twisdale and Vickery [5], thunderstorm winds have different statistical characteristics than non-thunderstorm winds, and errors can result from

analyzing both types of data together using a single distribution. In addition to wind data, ASOS weather reports also include thunderstorm observations, which allow for classification of winds as thunderstorm or non-thunderstorm.

In the design of some structures, notably high-rise buildings, wind direction data are required in addition to wind speed data to characterize the directionality of the local wind climate, because the assumption that extreme winds are equally likely from all directions leads to overly conservative designs. The Database-Assisted Design (DAD) methodology, applicable to both low-rise and high-rise buildings (e.g., [6]), provides a unified framework for using directional wind speed data in conjunction with aerodynamic data from wind tunnel tests in structural design. Because ASOS weather reports include wind directions in addition to wind speeds, ASOS wind data also have potential for use within the DAD framework.

A good source for archived ASOS weather reports is Data Set 9956 from the National Climatic Data Center (NCDC) [7], which contains data from about 10 000 stations worldwide. Data for specific stations can be purchased from NCDC as described in [7], and data are provided in ASCII format. These ASCII files contain routine hourly weather reports as well as special weather reports that were issued at shorter intervals between the hourly reports during events of particular interest. A single line from Data Set 9956 is shown in Figure 1, wrapped for display purposes. This line represents a routine hourly weather report from the ASOS station at LaGuardia Airport in New York. Observations of interest in the present study are shown in bold font and are further discussed subsequently. Because wind speeds with long recurrence intervals are of interest in structural design, archived weather reports are required over periods of decades, entailing hundreds of thousands of lines of text like that shown in Figure 1. Manual extraction of relevant data from such massive text files would clearly impose a considerable burden.

In order to facilitate more widespread use of ASOS wind data, both for improvement of the ASCE 7 wind speed map and for implementation in DAD, this paper describes procedures that have been developed for automated extraction of wind data and thunderstorm observations from archived ASOS weather reports and for classification of wind data as thunderstorm or non-thunderstorm. An important requirement in extreme value analysis is that data are statistically independent, and for this reason a procedure is also described for constructing sets of data separated by specified minimum time intervals to ensure statistical independence. These procedures are illustrated using data from three ASOS stations in the New York City area. The procedures described in this paper have been implemented in a software package called ASOS-WX that is being made publicly available [8] (follow the link for "Extraction of wind data from ASOS records"). Further details on these procedures are provided in the documentation at this website.

```
0171725030199911220506+40770-073900FM-16+0009KLGA V0823201N01401000151CN
0004021N1+01201+01201999999ADDMW1171MW2451MW3001GF10919999990002519999
99GD15+000159MA1102511999999REMMET072KLGA 220506Z 32014KT 1/4SM FG VV0
01 12/12 PK WND 33043/24 A3027 TSB26E02 RMK AO2;EQDQ01+000003PRSWM1
```

FIGURE 1
EXAMPLE OF A ROUTINE HOURLY WEATHER REPORT FROM [7] FOR LAGUARDIA AIRPORT

EXTRACTION OF PEAK WIND DATA

As noted by Sparks [9], the most meaningful wind data available in ASOS reports are the peak wind observations, which report the highest 5 s averaged wind speed in knots since the last hourly routine weather observation. Only peak wind speeds in excess of 13 m/s (25 knots) are reported. It is noted that this reporting threshold has the effect of censoring the resulting peak wind data, and this effect must be properly accounted for in statistical analysis. The wind direction corresponding to each peak wind speed is also reported in increments of 10° from true north (i.e., 90° corresponds to winds from the east), and the time of the peak wind observation is reported in Coordinated Universal Time (UTC), which is 5 h ahead of Eastern Standard Time.

An example of an ASOS peak wind observation is shown in bold on the fourth line of Figure 1, where the characters "PK WND 33043/24" indicate a peak wind speed of 22 m/s (43 knots) with a wind direction of 330°. As explained in [4], the wind direction and wind speed are indicated by either five or six numerical characters preceding the slash in the peak wind report, with the first three digits indicating the wind direction and the remaining two or three digits indicating the wind speed in knots. Three digits are required only for wind speeds exceeding 51 m/s (99 knots). The time of the peak wind observation is indicated by either two or four numerical characters following the slash, which represent the minute of the observation as MM or the hour and minute as HHMM. If the hour is not reported, it can be inferred by noting that the observation must have occurred within the hour previous to the time of the current routine weather report. For example, the characters "199911220506" shown in bold in the first line of Figure 1 indicate that this routine weather report occurred at 05:06 UTC on November 22, 1999. It can then be inferred that the reported time of "24" for the peak wind observation represents a time of 04:24 UTC on the same day.

In order to maximize the amount of data available for use in extreme value analysis, it is generally desirable to include pre-ASOS data in addition to ASOS data from a particular station. In so doing it is important to note that prior to the implementation of ASOS, "instantaneous" peak gusts were reported rather than 5 s averages. The influence of averaging time on peak wind speeds can be accounted for as discussed in [10], and assuming an effective averaging time of 1 s for "instantaneous" gust speeds, an average ratio of about 1.05 is obtained between peak "instantaneous" speeds and peak 5 s speeds at 10 m (33 ft) elevation over open terrain. The ASCE 7 wind speed map uses 3 s averaged gust speeds, which can be obtained from 1 s and 5 s gust speeds through multiplication by factors of about 0.968 and 1.02, respectively.

It is also important to note that even after the implementation of ASOS, not all stations have the standard 10 m (33 ft) anemometer elevation, as Sparks points out [9]. Information on anemometer elevation changes can be obtained in Data Set 6421 from NCDC [11], and wind speeds can be scaled to account for anemometer height as discussed in [10]. For the three stations considered in this study (Newark, LaGuardia, and Kennedy airports), the anemometer height was changed from 6.1 m (20 ft) to 10 m (33 ft) at the same time that ASOS was implemented. The average ratio between wind speeds at 10 m (33 ft) and 6.1 m (20 ft) elevation over open terrain is about 1.08, which nearly cancels the effect of changing from "instantaneous" to 5 s gust speeds.

Prior to the implementation of ASOS, peak winds were reported in a slightly different format, with wind directions given in tens of degrees (e.g., "9" indicated 90° and "18"

indicated 180°). The ASOS-WX software [8] can handle peak wind reports in both ASOS and pre-ASOS formats. By searching through archived ASOS records for the "PK WND" coding and extracting the relevant information as described above, the wind speed, wind direction, and date/time of each peak wind observation can be obtained. The station code (shown underlined and in bold in the first line of Figure 1) is also extracted for each peak wind observation, so that data from the station of interest can be selected.

Figure 2 shows a resulting plot of peak wind speeds versus date from the ASOS station at Newark Airport over a period of about 20 years. Raw wind speed values are presented in knots, and no scaling is applied to either ASOS or pre-ASOS data. This figure clearly shows a drop in the reporting threshold from 18 m/s (35 knots) to 13 m/s (25 knots) on or about January 1, 1995, more than a year before ASOS was implemented at the Newark station on July 1, 1996. Because data before and after this change in threshold are left-censored at different levels, care must be taken in combining these data for statistical analysis. For example, in a "peaks over threshold" analysis [12], a threshold less than 18 m/s (35 knots) must not be used for the combined data set.
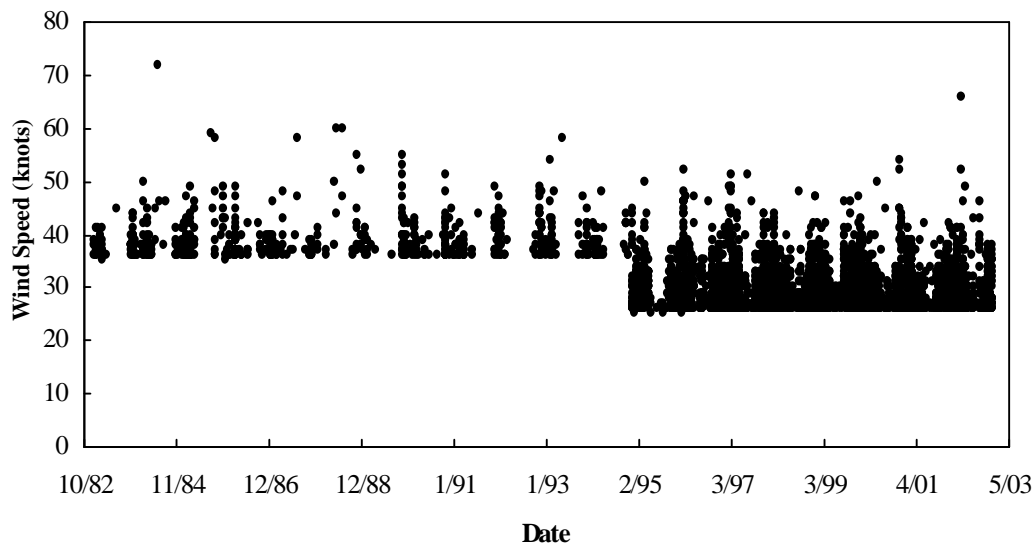

FIGURE 2
TIME HISTORY OF PEAK WIND SPEEDS EXTRACTED FROM [7] FOR NEWARK AIRPORT

IDENTIFICATION OF THUNDERSTORM WIND DATA

Once the date and time of peak wind reports have been extracted, these can be compared with the date and time of thunderstorm reports to classify the winds as thunderstorm or non-thunderstorm. Previous studies (e.g., [5]) used a "thunderday" approach, in which any wind that occurred on the same calendar day as a thunderstorm was considered a thunderstorm wind. However, because thunderstorms typically last only a few hours and are often associated with larger-scale weather systems, it is not unlikely that significant non-thunderstorm winds could be recorded within 24 h of a thunderstorm. The availability in ASOS reports of date/time information for thunderstorm beginnings and ends allows for more precise classification of thunderstorm and non-thunderstorm winds.

Thunderstorm beginning and end times are indicated in ASOS reports using the codes "TSB" and "TSE" respectively, followed by either two or four numerical characters

representing the minute of the observation as MM or the hour and minute as HHMM. If the hour is not reported, it can be inferred from the time of the routine weather report, as discussed previously for peak winds. Occasionally, a thunderstorm beginning and end will both occur within an hour, and in such cases, subsequent beginning and end times are reported immediately following the initial report using "B" and "E" to denote beginning and end, respectively. For example, the characters "TSB26E02" shown in bold on the fourth line of Figure 1 indicate (in conjunction with the date and time of the routine weather report) that a thunderstorm began at 04:26 UTC and ended at 05:02 UTC on November 22, 1999.

A challenge in making use of reported thunderstorm beginning and end times is that coding errors sometimes result in beginning times with no matching end time or vice versa (i.e., two successive "TSB" reports may be encountered without an intervening "TSE" report). In other cases, the time between a "TSB" report and the next "TSE" report may be unrealistically long (intervals of several days have been observed). In such ambiguous cases, *manual weather observations* can be used as an additional source of information to check the legitimacy of questionable thunderstorm beginning and end reports and to estimate corresponding end or beginning reports that may be missing.

Manual weather observations are indicated in ASOS reports by the characters "MW" followed by three numeric characters. The first of these digits is simply a counter to indicate the number of manual weather observations on the current data line, while the last two digits are a code that represents the type of weather in progress at the time of the current report. The characters "MW117" shown in bold in the second line of Figure 1 thus indicate a manual weather observation with code 17. Descriptions for all codes are provided in [7], and as shown Table 1, seven different codes are available for indicating a thunderstorm in progress. The time of each manual weather observation can be determined from the time of the routine weather report.

Code	Description
17	Thunderstorm, but no precipitation at time of observation
29	Thunderstorm (with or without precipitation)
95	Thunderstorm, slight or moderate, without hail, but with rain and/or snow at time of observation
96	Thunderstorm, slight or moderate, with hail at time of observation
97	Thunderstorm, heavy, without hail, but with rain and/or snow at time of observation
98	Thunderstorm, combined with duststorm or sandstorm at time of observation
99	Thunderstorm, heavy, with hail at time of observation

TABLE 1
DESCRIPTIONS OF ASOS THUNDERSTORM CODES

The ASOS-WX software [8] searches through archived ASOS records and extracts all thunderstorm beginning and end reports and all manual weather observations having one of the seven codes in Table 1. This information is then used to construct lists of matching thunderstorm beginning and end times for use in classification of peak winds. Further details about these procedures are provided in the online documentation [8], and it is noted that the software can also handle pre-ASOS reporting formats. Figure 3 shows a histogram of the thunderstorm durations computed from the lists of matching beginning and end times obtained for Newark Airport over the same time period as the peak wind data in Figure 2. Nearly all of the thunderstorms have a duration of less than 6 h, which is consistent with the observations in [13].
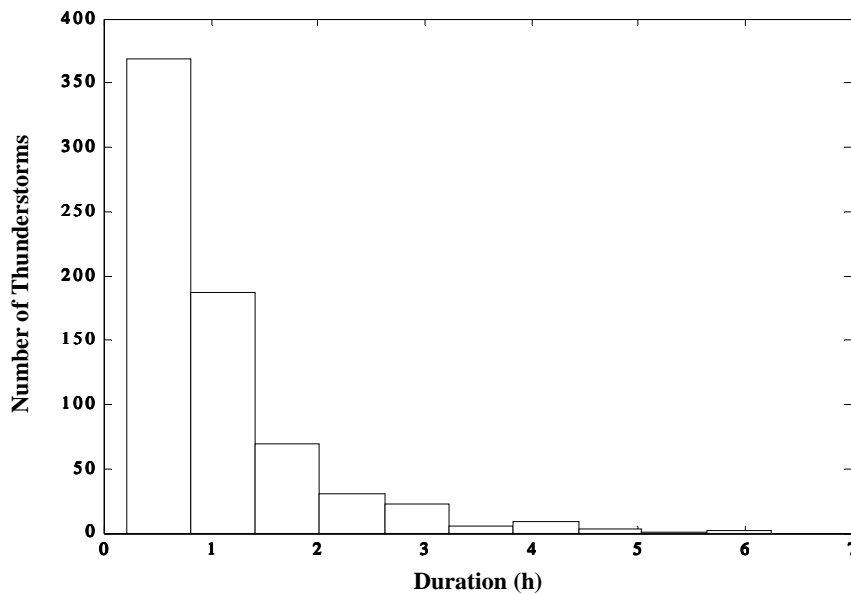
FIGURE 3
HISTOGRAM OF THUNDERSTORM DURATIONS (NEWARK AIRPORT)

Once lists of matching thunderstorm beginning and end times have been assembled, the procedure for identification of thunderstorm winds involves searching for peak wind observations that occurred within windows of time defined by these pairs of beginning and end times. These thunderstorm windows can be extended by specified intervals of time before the reported thunderstorm beginning times and after the reported end times, so that winds can be classified as thunderstorm winds even if they arrive at a station somewhat before a reported thunderstorm beginning or after a reported end (i.e., winds associated with a thunderstorm outflow boundary or gust front). Any peak winds that fall within these extended thunderstorm windows are classified as thunderstorm winds, while the remaining peak winds are classified as non-thunderstorm.

For the three stations in the New York City area, intervals of 1 h both before the reported beginning times and after the reported end times were deemed appropriate for extending thunderstorm windows. In the ASOS-WX software [8], the intervals for extending thunderstorm windows in each direction can be independently specified by the user for classification of thunderstorm winds. By varying these intervals, it was observed that a significant number of high wind speeds occurred within the 30 min intervals preceding the reported thunderstorm beginnings, most likely associated with thunderstorm outflow boundaries.

CONSTRUCTION OF STATISTICALLY INDEPENDENT DATA SETS

An important requirement in extreme value analysis is that the data are statistically independent, and for this reason, only one wind speed from each storm system should be used. Because hourly peak wind speeds are available in ASOS records, while storm systems typically last for several days (several hours for thunderstorms), multiple wind speeds are generally available for each storm. Some method is therefore required to extract the maximum wind speed from each storm and to eliminate other wind speeds

associated with the same storm. The "method of independent storms," discussed in [14], is one possible approach. However, this method requires continuous wind speed records, which are unavailable from ASOS records because peak wind speeds are reported only if they exceed a certain threshold (see Figure 1). Simiu and Heckert [12] present an alternative approach, which involves partitioning the data into periods with duration greater than or equal to the duration of a typical storm system. The maximum value from each period is then selected, subject to the additional requirement that maxima of adjacent periods must be separated by at least half a period – otherwise, the smaller of the adjacent maxima is replaced by the next smaller value in the respective period, which itself must be separated by at least half a period from maxima of adjacent periods.

A new procedure is proposed in this paper that does not require continuous time histories and is more easily automated than the procedure in [12]. This procedure ensures that no two wind speeds are separated by less than a specified a minimum separation interval, denoted $\Delta t_{\min}$. This interval should be greater than or equal to the duration of a typical storm system. Because thunderstorms typically have shorter durations than larger-scale storm systems, the procedure can be applied separately to thunderstorm and non-thunderstorm winds, using different values of $\Delta t_{\min}$ for these distinct types of winds. Let t_1, t_2, ... t_N denote the times of peak wind speeds extracted from ASOS records, and let s_1, s_2, ... s_N denote the corresponding peak wind speeds, where N is the total number of wind speeds (either thunderstorm or non-thunderstorm). The time values are measured with respect to some fixed reference date and time, so that $t_N - t_1$ represents the total time span covered by the data, which is typically on the order of decades.

The procedure then works as follows. The time of the first peak wind speed, t_1, is checked against the time of the second peak wind speed, t_2. If $t_2 - t_1 \geq \Delta t_{\min}$, then the procedure moves forward to check $t_3 - t_2$, and so on through the data set. If $t_{k+1} - t_k < \Delta t_{\min}$, then the lesser of the two corresponding wind speeds (s_k or s_{k+1}) is deleted and the greater is retained. If the two wind speeds are equal ($t_{k+1} - t_k < \Delta t_{\min}$ and $s_k = s_{k+1}$), then s_{k+1} is deleted and s_k is retained, because s_k has a larger separation interval from subsequent wind speeds. When a wind speed value is deleted, the corresponding time value is also deleted, and the surviving time value is then compared with the next time value in the data set. This procedure continues through the entire time history, to ensure that all of the resulting data points are separated by at least $\Delta t_{\min}$. In the implementation of this procedure in the ASOS-WX software [8], the indices of the surviving data points in the original data set are saved, so that the wind directions corresponding to the surviving wind speed and time values can also be obtained.

The application of this separation procedure is illustrated in Figure 4 for both thunderstorm and non-thunderstorm wind speed data from Newark airport over a period from April to June of 2000. A separation interval of $\Delta t_{\min} = 4$ d was used for non-thunderstorm winds in Figure 4(a), while a separation interval of $\Delta t_{\min} = 6$ h was used for thunderstorm winds in Figure 4(b). In both cases, the surviving wind speeds are indicated with circles. Figure 5 shows the influence of the separation interval $\Delta t_{\min}$ on the number of data points per year that survive the separation procedure. Results for both thunderstorm and non-thunderstorm winds are presented for the three airports in the New York City area using data over a period of about 20 years. Figure 5 shows that the
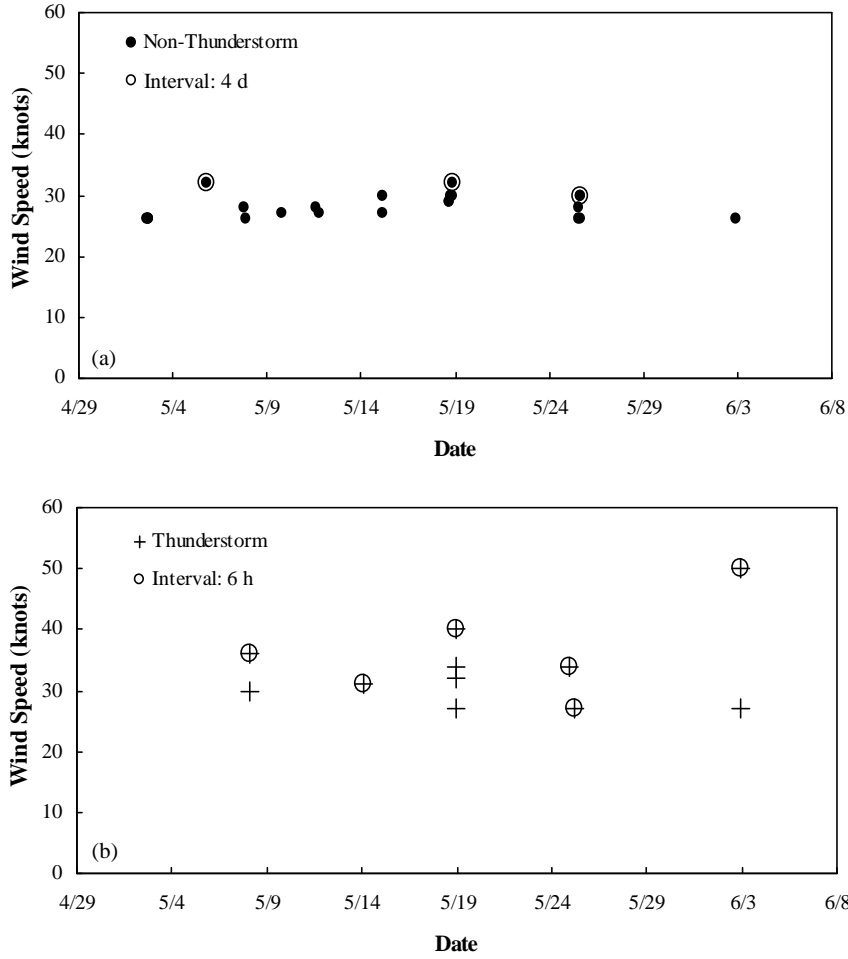
FIGURE 4
CONSTRUCTION OF STATISTICALLY INDEPENDENT WIND SPEED DATA SETS (DATA FROM [7] FOR NEWARK AIRPORT, YEAR 2000). (A) NON-THUNDERSTORM; (B) THUNDERSTORM.
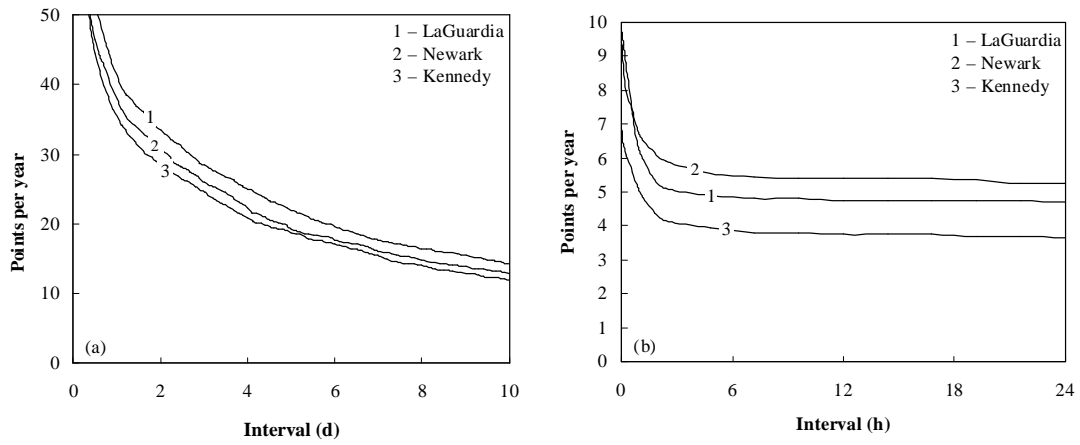

FIGURE 5
INFLUENCE OF SEPARATION INTERVAL ON NUMBER OF SURVIVING DATA POINTS.
(A) NON-THUNDERSTORM; (B) THUNDERSTORM.

number of surviving data points stabilizes as the separation interval Δt_{min} exceeds the duration of most storm systems. Simiu and Heckert [12] indicate that a duration of four to eight days is typical for non-thunderstorm systems, and Figure 5(a) shows that the number of surviving non-thunderstorm wind speeds plateaus over this range. Figure 3 shows that most thunderstorms have a duration of less than $6\,h$, and Figure 5(b) shows that the number of surviving thunderstorm wind speeds drops only slightly as the separation interval increases from $6\,h$ to $12\,h$. Figure 6 shows a polar plot of thunderstorm and non-thunderstorm wind speed and wind direction data from Newark airport that were deemed statistically independent using the same separation intervals as in Figure 4. These data were obtained from the original data shown in Figure 2, but only surviving data with wind speeds greater than 18 m/s $(35\ knots)$ are presented, because data lower than this threshold are not properly represented, as noted above.
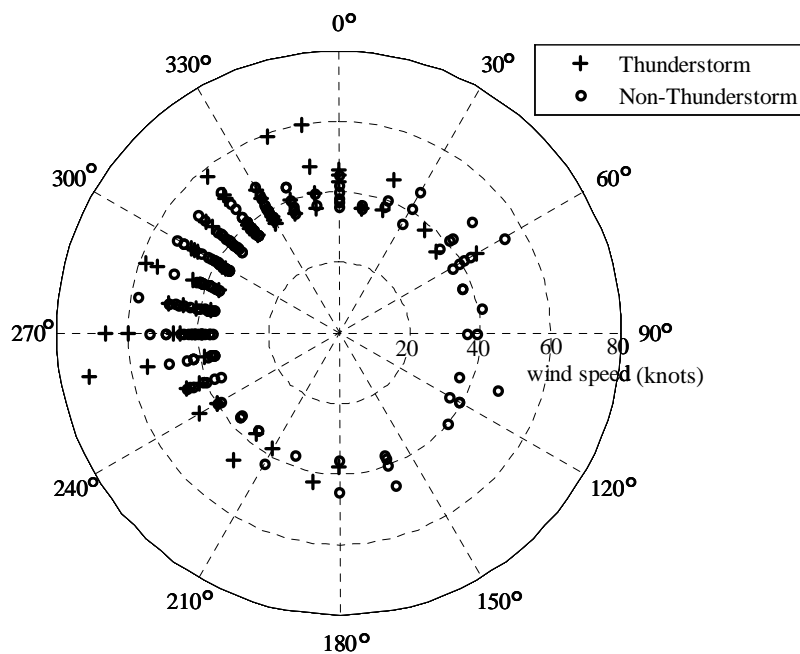

FIGURE 6
STATISTICALLY INDEPENDENT WIND SPEEDS AND DIRECTIONS (DATA FROM [7] FOR NEWARK AIRPORT)

CONCLUDING REMARKS

Procedures have been described for extracting wind speed and wind direction data from archived ASOS weather reports, and for extracting thunderstorm beginning and end times and using these to classify wind data as thunderstorm or non-thunderstorm. A new procedure for constructing statistically independent data sets has also been presented, which involves specifying a minimum separation interval for the resulting data. These procedures have been implemented in a publicly available software package called ASOS-WX. This software is designed to read archived weather reports from Data Set 9956 of the National Climatic Data Center, which contains hourly weather reports over periods of decades from about 10 000 stations worldwide. The software and procedures presented in this paper hold great potential to significantly expand the body of wind data

available for structural engineering purposes. This data could be used for improving the current wind speed map in ASCE 7 and could also be used in Database-Assisted Design.

REFERENCES

[1] American Society of Civil Engineers (ASCE). (2006). *Minimum Design Loads for Buildings and Other Structures, ASCE/SEI 7-05*, Reston, VA.

[2] Peterka, J.A. and Shahid, S. (1998). "Design gust wind speeds in the United States." *J. Struct. Eng.*, 124(2), 207-214.

[3] Peterka, J.A. and Esterday, W.S. (2005). "Discussion of 'Wind speeds in ASCE 7 standard peak-gust map: assessment' by Emil Simiu, Roseanne Wilcox, Fahim Sadek, and James J. Filliben." *J. Struct. Eng.*, 131(6), 994-996.

[4] National Weather Service (NWS). (1998). *Automated Surface Observing System (ASOS) User's Guide*, <www.weather.gov/asos/aum-toc.pdf>

[5] Twisdale, L.A. and Vickery, P.J. (1992). "Research on thunderstorm wind design parameters." *J. Wind Eng. Ind. Aerodyn.*, 41-44, 545-556.

[6] Main, J.A. and Fritz, W.P. (2006). "Database-assisted design for wind: concepts, software, and examples for rigid and flexible buildings." *Building Science Series 180*, National Institute of Standards and Technology, Gaithersburg, MD.

[7] National Climatic Data Center (NCDC). (2003). *Data Documentation for Data Set 9956 (DSI-9956)*, <www.ncdc.noaa.gov/oa/documentlibrary/pdf/tddoc/td9956.pdf>

[8] Heckert, N.A. (2006). *Extreme Winds and Wind Effects on Structures*, National Institute of Standards and Technology. <www.nist.gov/wind>

[9] Sparks, P.R. (1999). "Using wind data from the automated surface observing system." *J. of the Roof Consultants Institute*, 6, 24-28. <www.rci-mercury.com/files/bz080111.pdf>

[10] Simiu, E. and Scanlan, R.H. (1996). *Wind Effects on Structures*, 3rd Ed., Wiley, New York.

[11] National Climatic Data Center (NCDC). (2002). *Data Documentation for Data Set 6421 (DSI-6421)*, <www.ncdc.noaa.gov/oa/documentlibrary/pdf/tddoc/td6421.pdf >

[12] Simiu, E. and Heckert, N.A. (1996). "Extreme wind distribution tails: a 'peaks over threshold' approach", *J. Struct. Eng.*, 122(5), 539-547.

[13] Robinson, P.J. and Easterling, D.R. (1988). "The frequency distribution of thunderstorm durations." *J. of Applied Meteorology*, 27(1), 77-82.

[14] Harris, R.I. (1999). "Improvements to the 'method of independent storms.'" *J. Wind Eng. Ind. Aerodyn.*, 80, 1-30.