Journal of Wind Engineering and Industrial Aerodynamics 97 (2009) 120–131

Contents lists available at ScienceDirect

Journal of Wind Engineering and Industrial Aerodynamics

journal homepage: www.elsevier.com/locate/jweia

Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis

Franklin T. Lombardo [a], Joseph A. Main [b,*], Emil Simiu [b]

[a] Wind Science and Engineering Research Center, Texas Tech University, MS 1165, Lubbock, TX 79414, USA
[b] Building and Fire Research Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899-8611, USA

ARTICLE INFO

Article history:
Received 24 July 2007
Received in revised form
12 December 2008
Accepted 12 March 2009
Available online 18 April 2009

Keywords:
Extreme wind speeds
Thunderstorms
Wind climatology
Meteorological data
Statistics
Wind loads
Structural design

ABSTRACT

Design wind loads are partly based on extreme value analyses of historical wind data, and limitations on the quantity and spatial resolution of wind data pose a significant challenge in such analyses. A promising source of recent wind speed and direction data is the automated surface observing system (ASOS), a network of about 1000 standardized US weather stations. To facilitate the use of ASOS data for structural engineering purposes, procedures and software are presented for (a) extraction of peak gust wind data and thunderstorm observations from archived ASOS reports, (b) classification of wind data as thunderstorm or non-thunderstorm to enable separate analyses, and (c) construction of data sets separated by specified minimum time intervals to ensure statistical independence. The procedures are illustrated using approximately 20-year datasets from three ASOS stations near New York City. It is shown that for these stations thunderstorm wind speeds dominate the extreme wind climate at long return periods. Also presented are estimates based on commingled data sets (i.e., sets containing, indiscriminately, both non-thunderstorm and thunderstorm wind speeds), which until now have been used almost exclusively for extreme wind speed estimates in the US. Analyses at additional stations will be needed to check whether these results are typical for locations with both thunderstorm and non-thunderstorm winds.

1. 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 (ASCE, 2006), for example, is based on extreme value analysis of wind speed data from 487 stations in the United States grouped into "superstations" (Peterka and Shahid, 1998). In responding to criticisms of this wind speed map (Simiu et al., 2003, 2005), Peterka and Esterday (2005) suggested that future analyses aimed at improving the wind speed map should focus on inclusion of additional wind data that have become available since 1994, 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 (NWS, 1998).

A good source for archived ASOS weather reports is Data Set 9956 from the National Climatic Data Center (NCDC, 2003), which contains routine hourly weather reports from about 10 000 stations worldwide, as well as special weather reports issued at shorter intervals during events of particular interest. A single record from Data Set 9956 is shown in Fig. 1, wrapped for display purposes. This record represents a routine hourly weather report from the ASOS station at LaGuardia Airport in New York. Observations of interest in the present study are indicated in Fig. 1 using bold font and underlining, 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 Fig. 1. Manual extraction of relevant data from such massive text files could be prohibitive. In order to facilitate more widespread use of ASOS wind data for structural engineering purposes, this paper describes procedures that have been developed for automated extraction of peak gust wind data from NCDC Data Set 9956. The procedures described in this paper could also be applied to data from other sources, such as NCDC Data Set 3505 (NCDC, 2006), which provides data on peak winds and thunderstorm occurrences, although without the level of detail of Data Set 9956. Records from Data Set 3505 are available for download through the NCDC website and can

* Corresponding author. Tel.: +1 301 975 5286; fax: +1 301 869 6275.
E-mail address: joseph.main@nist.gov (J.A. Main).

0167-6105/$ - see front matter
doi:10.1016/j.jweia.2009.03.001

```
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
```

Fig. 1. Example of a routine hourly weather report from NCDC Data Set 9956 for LaGuardia Airport.

thus be obtained more easily than records from Data Set 9956, which must be specially ordered at a substantial charge.

Owing to phenomenological and climatological differences between winds generated by thunderstorms and those generated by larger scale synoptic events, it is appropriate to perform separate statistical analyses of thunderstorm (T) and non-thunderstorm (NT) extreme wind speeds. Gomes and Vickery (1978) originally proposed the idea of separating by wind type in areas of mixed wind climates and to identify combined distributions of these wind speeds, given the errors that result from using a single distribution. Gomes and Vickery (1978), along with Holmes (2001), showed that T wind speeds dominated the wind climate of Australia when hurricanes were excluded, while Holmes (2001) stated that in Melbourne, T winds dominated the climate at return periods of 100 years and greater. Twisdale and Vickery (1992) performed a similar analysis at four sites in the US and found that in the central areas of the US, thunderstorms dominated the wind climate at return periods of about 50 years or larger. Letchford and Ghosalkar (2004) also found that thunderstorms dominated the wind climate in West Texas. Following these studies, Peterka and Esterday (2005) suggested 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.

In analyzing the mixed weather climate of Singapore, Choi and Tanurdjaja (2002) used careful inspection of continuous wind records to separate small-scale wind events, such as thunderstorms, from larger scale wind events. For the wind climate of Singapore, Choi and Tanurdjaja (2002) considered this approach more appropriate than separating the wind data into thunderstorm and non-thunderstorm, as was done in a previous study (Choi, 1999). However, the approach proposed by Choi and Tanurdjaja (2002) requires continuous wind records, which are not available for peak gust data from ASOS weather reports. Twisdale and Vickery (1992), Cook et al. (2003), and Letchford and Ghosalkar (2004) 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 on the same calendar day as a thunderstorm, thus resulting in misclassification. In this paper, a procedure for automated classification of thunderstorm and non-thunderstorm winds is proposed that makes use of weather observations and thunderstorm beginning and ending times reported by manual observers at ASOS stations. The availability of such information in archived ASOS weather reports allows for more precise classification of thunderstorm and non-thunderstorm wind data than using the "thunderday" approach, and the automation of this procedure facilitates application to large sets of data from many stations that use the same reporting format.

In hurricane-prone regions, it is important to separate hurricane and non-hurricane wind data in extreme value analysis. However, even in hurricane-prone regions it can still be useful to further classify the non-hurricane wind data as thunderstorm or non-thunderstorm using the procedures described in this paper. While ASOS weather reports provide no information regarding hurricane or tropical storm passage, information on historical hurricane tracks is available from other sources (e.g., Neumann et al., 1993; National Hurricane Center, 2006). Inspection of ASOS wind data from the three stations near New York City revealed several significant wind speeds associated with tropical systems. Because the main focus of this paper is separation of T and NT wind speeds, these wind data (along with sea breeze effects) are simply included with the NT data in the examples of separation presented in this paper. However, in Section 5, wind speeds with tropical influences were excluded from the extreme value analysis.

An important requirement in extreme value analysis is that the data are statistically independent. For this reason a procedure is also described for constructing sets of data separated by specified minimum time intervals to reduce their statistical dependence. According to Brabson and Palutikof (2000), an improper separation interval or "dead time" between the data points used in extreme value analysis can lead to an artificial increase in the wind speeds obtained for long return periods. The influence of the separation interval is investigated for both T and NT data, in an effort to determine appropriate values for these distinct types of winds. The procedures presented in this paper are illustrated using data from three ASOS stations in the New York City area. The procedures have also been implemented in a public-domain software package called *ASOS-WX* (Lombardo and Main, 2006), which was developed using Version 7 of MATLAB (MathWorks, 2006) and is available for download at www.nist.gov/wind. In addition to MATLAB files, a "stand-alone" version of *ASOS-WX* is also available, for users who do not have MATLAB and are running in the Windows 2000/XP environment.

Using data sets assembled through these procedures, extreme value analysis of T and NT wind speeds are carried out for the three NYC stations. The results of separate analyses of T winds and NT winds are presented, regardless of their direction. The results are then used to obtain distributions of NT and T speeds, mixed distributions of NT and T speeds, and distributions of commingled data sets in which no differentiation is made between NT and T speeds. Interestingly, it is shown that for these stations thunderstorm wind speeds dominate the extreme wind climate, especially at longer return periods, to such an extent that non-thunderstorm winds can be disregarded in the analysis. Such results could be of considerable relevance for the future development of an improved US wind map (see e.g., Simiu et al., 2003).

2. Extraction of peak wind data

As noted by Sparks (1999), 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 (e.g., 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 Fig. 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 specified in NOAA (2005), the wind direction and wind speed are indicated in a peak wind report by either five or six numerical characters preceding a slash, 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 weather report. For example, the characters "199911220506" shown in bold in the first line of Fig. 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. The station code (shown underlined in Fig. 1) is used to extract peak wind data for the station of interest.

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, where the date of ASOS commissioning can be found in Data Set 6421 from NCDC (2002). Note that, prior to ASOS commissioning, "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 Simiu and Scanlan (1996), 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 0.97 and 1.02, respectively. Prior to ASOS commissioning, peak winds were reported in a slightly different format, and the *ASOS-WX* software can handle both ASOS and pre-ASOS reporting formats.

As Sparks (1999) points out, it is also important to note that even after ASOS commissioning, not all stations have the standard 10 m (33 ft) anemometer elevation. Information on anemometer elevation changes can be obtained in Data Set 6421 from NCDC (2002), and wind speeds can be scaled to account for anemometer height as discussed in Simiu and Scanlan (1996). For the three stations considered in this study (Newark, LaGuardia, and Kennedy airports near New York City), the anemometer height was changed from 6.1 m (20 ft) to 10 m (33 ft) at the same time that ASOS was commissioned. 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.

Fig. 2 shows a 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 commissioned at the Newark station on July 1, 1996. Because data before and after this change in threshold are censored at different levels, care must be taken in combining these data for statistical analysis. For example, in a "peaks over threshold (POT)" analysis (e.g., Simiu and Heckert, 1996), a threshold less than 18 m/s (35 knots) must not be used for the combined data set.

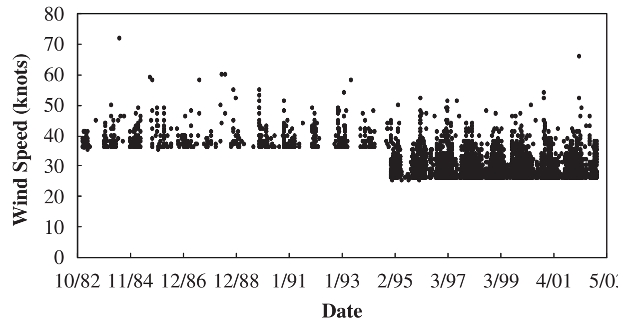

Fig. 2. Time history of peak wind speeds extracted from NCDC Data Set 9956 for Newark Airport.

3. Classification of thunderstorm and non-thunderstorm winds

Once the date and time of peak wind reports have been extracted, these can be compared with intervals of thunderstorm occurrence to classify the wind data as thunderstorm or non-thunderstorm. Archived ASOS weather reports contain two types of observations of thunderstorm occurrence: (1) thunderstorm beginning and ending times and (2) manual weather observations. Both of these types of observations are manually reported by human observers to augment the automated observations, and continuous staffing by human observers is required in order to reliably use such manual reports to classify winds as thunderstorm or non-thunderstorm. Therefore, the procedure described in this section is applicable only to ASOS stations designated Service Level A or Service Level B, which provide continuous manual reporting (AOPA, 1999). The three ASOS stations considered in this paper have been established as Service Level A.

3.1. Thunderstorm beginning and ending reports

According to the *Federal Meteorological Handbook No.* 1 (NOAA, 2005), "The beginning of a thunderstorm is to be reported as the earliest time: (1) thunder is heard; (2) lightning is observed at the station when the local noise level is sufficient to prevent hearing thunder; or (3) lightning is detected by an automated sensor". Conversely, "the ending of a thunderstorm shall be reported as 15 minutes after the last occurrence of any of the above criteria". The *Federal Meteorological Handbook No.* 1 also specifies that thunderstorm beginning and ending times are to be reported using the coding format "TSB(hh)mmE(hh)mm", where "TS" indicates thunderstorm, "B" indicates beginning, "E" indicates ending, and "(hh)mm" denotes the time of occurrence. If the hour of occurrence "hh" is not reported, it can be inferred from the time of the routine weather report, as discussed previously for peak winds. An example is shown in bold on the fourth line of Fig. 1, where the characters "TSB26E02" shown in bold on the fourth line of Fig. 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 number of alternative reporting formats for thunderstorm beginning and ending times have also been encountered in archived weather reports. For example, beginning and ending times are commonly reported in isolation as "TSB(hh)mm" or "TSE(hh)mm". The *ASOS-WX* software can extract beginning and ending times reported in any of the various formats that were encountered.

3.2. Manual weather observations

Manual weather observations, which indicate the atmospheric conditions at the time of each weather report, provide a second

Table 1
Descriptions of ASOS thunderstorm codes.

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

source of information on thunderstorm occurrence. Manual weather observations are indicated in NCDC Data Set 9956 by the characters "MW" followed by four digits. The second and third digits form a two-digit code that represents the atmospheric conditions at the time of the current report. One hundred different codes are available to denote different atmospheric conditions, and descriptions for all codes are provided in NCDC (2003). Seven different codes are available for indicating a thunderstorm in progress, and these codes are listed with their descriptions in Table 1. In any given weather report, as many as seven different codes can be used to represent the present weather, and the first of the four digits following "MW" is simply a counter for the number of codes used in the current weather report (i.e., MW1 denotes the first code, MW2 denotes the second, and so on). The fourth digit following "MW" indicates the quality status of the present weather observation: a quality code of 0 denotes no quality check, 1 denotes "good" quality, and higher values denote suspect, erroneous, or missing reports (NCDC, 2003). The characters "MW1171" shown in bold in the second line of Fig. 1 thus indicate a first manual weather observation (in this report there are three) with a code of 17 and a "good" quality check. The date and time of each manual weather observation can be determined from the date and time of the routine weather report.

3.3. Matching of thunderstorm beginning and ending times

Intervals of thunderstorm occurrence can be defined more precisely by using reported thunderstorm beginning and ending times than by using manual weather observations, because the precise hour and minute of each thunderstorm beginning and ending is reported, while only the time of the current weather report is available for manual weather observations. However, a challenge in making use of reported thunderstorm beginning and ending times is that coding errors sometimes result in beginning times with no matching ending time or vice versa. In other cases, the time between a thunderstorm beginning report and the next thunderstorm ending report may be unrealistically long, suggesting that intermediate beginning and ending reports may be missing. Let $\mathbf{b} = [b_1\ \ b_2\ \ \ldots]$ denote the vector of thunderstorm beginning times extracted from the archived weather reports, and let $\mathbf{e} = [e_1\ \ e_2 \ldots]$ denote the vector of extracted ending times. As a consequence of reporting errors, the vectors $\mathbf{b}$ and $\mathbf{e}$ generally do not have the same number of elements, and for any given index k, the ending time e_k may not be associated with the same thunderstorm as the beginning time b_k. To address this problem, a procedure has been developed that makes use of reported thunderstorm beginning and ending times in conjunction with manual weather observations to assemble lists of matching thunderstorm beginning and ending times.

The first step in this procedure is to estimate thunderstorm beginning and ending times from the manual weather observations. In so doing, thunderstorm observations occurring in consecutive weather reports are assumed to represent a single thunderstorm. A thunderstorm beginning is then estimated as the date/time associated with the first in a set of consecutive manual thunderstorm observations, while the corresponding thunderstorm ending is estimated as the date/time of the first subsequent weather report that does not contain a manual thunderstorm observation. If no weather report is found within an interval of two hours following a thunderstorm observation, then the thunderstorm is assumed to have ended one hour after the last thunderstorm observation. Conversely, if no weather report is found within an interval of two hours preceding a thunderstorm observation, then the thunderstorm observation is assumed to represent the beginning of a new thunderstorm. Using this procedure, vectors of thunderstorm beginning times and corresponding ending times can be assembled from the manual weather observations, denoted $\mathbf{b}_m = [b_{m1}\ \ b_{m2} \ldots b_{mM}]$ and $\mathbf{e}_m = [e_{m1}\ \ e_{m2} \ldots e_{mM}]$, respectively, where M is the total number of thunderstorms. This procedure ensures that each beginning time has a corresponding ending time, so the vectors $\mathbf{b}_m$ and $\mathbf{e}_m$ have the same length, and their corresponding elements are associated with the same thunderstorm.

More precise vectors of matching beginning and ending times, denoted $\mathbf{B} = [B_1\ \ B_2 \cdots B_M]$ and $\mathbf{E} = [E_1\ \ E_2 \cdots E_M]$, respectively, can then be assembled by using the matching beginning and ending times in $\mathbf{b}_m$ and $\mathbf{e}_m$ in conjunction with the more precise (but not necessarily matching) reported beginning and ending times in $\mathbf{b}$ and $\mathbf{e}$. The procedure for assembly of $\mathbf{B}$ and $\mathbf{E}$ involves looping through the elements of the vectors $\mathbf{b}_m$ and $\mathbf{e}_m$ and searching for corresponding elements in the vectors $\mathbf{b}$ and $\mathbf{e}$. Beginning with the index $k = 1$, the procedure is as follows:

(1) Search for reported thunderstorm beginning times in $\mathbf{b}$ that fall within the smaller of the following intervals: (a) the hour preceding b_{mk} or (b) the interval between E_{k-1} and b_{mk} (for $k>1$). (The time of a reported thunderstorm beginning must precede the time of the weather report in which it was indicated, but a beginning time should not precede the previous ending time.) If one or more reported beginning times are found in this interval, then set B_k equal to the earliest of these reported beginning times. If no reported beginning times are found in this interval, then set B_k equal to b_{mk}.
(2) Search for reported ending times in $\mathbf{e}$ that fall in the interval between B_k and e_{mk}. (The time of a reported thunderstorm ending must precede the time of the weather report in which it was indicated, but an ending time should not precede its corresponding beginning time.) If one or more reported ending times are found in this interval, then set E_k equal to the latest of these reported ending times. If no reported ending times are found in this interval, then set E_k equal to e_{mk}.
(3) If $k = M$, then terminate the procedure; otherwise increment the index k and repeat from 1.

The resulting vectors $\mathbf{B}$ and $\mathbf{E}$ are considered the best available estimates of the thunderstorm beginning and ending times. Both of the resulting vectors will have M elements and the indices of the beginning and ending times in $\mathbf{B}$ and $\mathbf{E}$ will correspond, so that a vector of thunderstorm durations can be computed as $\mathbf{D} = \mathbf{E} - \mathbf{B}$. Fig. 3 shows a histogram of the thunderstorm durations computed from the lists of matching beginning and ending times obtained from Newark Airport over the same period as the peak wind data in Fig. 2. All of the thunderstorms have durations of less than 5 h, and no thunderstorms have durations of less than 15 min, as a

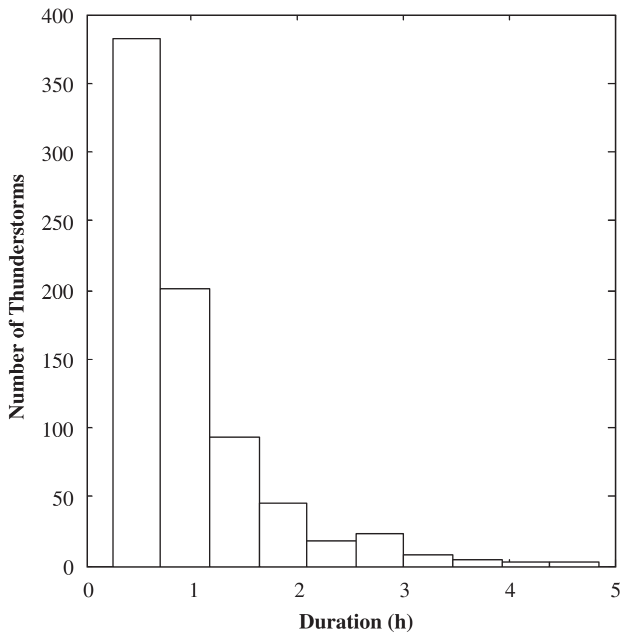

Fig. 3. Histogram of thunderstorm durations (Newark Airport).

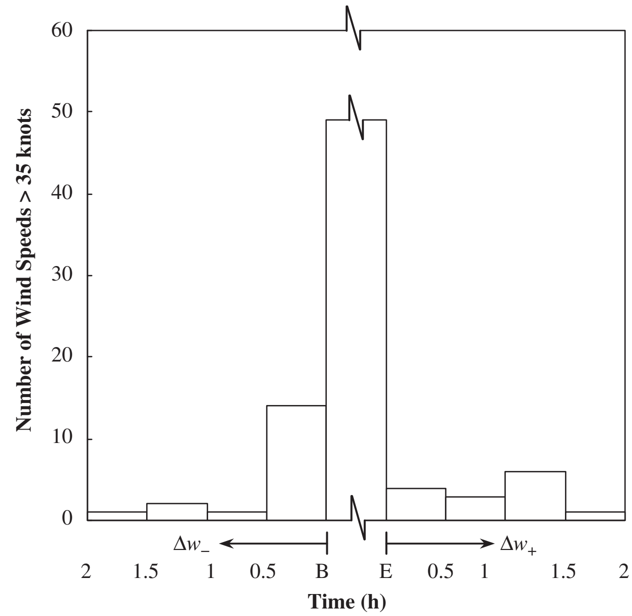

Fig. 4. Influence of extending the thunderstorm window (Newark Airport).

consequence of the thunderstorm ending definition in the *Federal Meteorological Handbook No.* 1 (NOAA, 2005) noted above.

3.4. Time windows for classification of thunderstorm winds

Once the vectors **B** and **E** of matching thunderstorm beginning and ending 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 ending times. These thunderstorm windows can be extended by specified intervals of time before the reported thunderstorm beginning times and after the reported ending 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 ending (i.e., winds associated with a thunderstorm outflow boundary or gust front). Let Δw_- denote the interval by which thunderstorm windows are to be extended before the reported thunderstorm beginning, and let Δw_+ denote the interval by which thunderstorm windows are to be extended after the reported ending. For the kth thunderstorm, the extended window to be searched for peak winds is then given by the time interval between $B_k - \Delta w_-$ and $E_k + \Delta w_+$, and the corresponding window is checked for each thunderstorm. 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.

The histogram in Fig. 4 shows the effect of extending the thunderstorm windows for Newark Airport. Only wind speeds greater than 35 knots are included, because of the change in reporting threshold discussed previously. The markers labeled "B" and "E" along the abscissa in Fig. 4 denote the reported thunderstorm beginning and ending times, respectively, and the broken horizontal line between these markers represents the variability of thunderstorm durations. The vertical bar plotted between the "B" and "E" markers in Fig. 4 represents the number of peak winds found in the windows between reported thunderstorm beginnings and endings (i.e., with $\Delta w_- = 0$ and $\Delta w_+ = 0$). The vertical bars to the left of the "B" marker indicate the number of additional peak winds found by extending the thunderstorm windows by an amount Δw_- before the reported beginnings, with $\Delta w_+ = 0$. Conversely, the vertical bars to the right of the "E" marker indicate the number of additional peak winds found by extending the thunderstorm windows by an amount Δw_+ after the reported endings, with $\Delta w_- = 0$.

Fig. 4 shows that a significant number of high wind speeds was found within an interval of $\Delta w_- = 0.5$ h before the reported thunderstorm beginnings, most likely associated with thunderstorm gust fronts. Further increasing this interval to $\Delta w_- = 1$ h before the reported thunderstorm beginnings resulted in a much smaller number of additional wind speeds. A significant number of additional wind speeds was found by extending the thunderstorm windows from $\Delta w_+ = 1$ h to 1.5 h after the reported endings. However, closer inspection of the surrounding weather reports—including such factors as wind direction, atmospheric pressure, and temperature—suggested that these wind speeds were associated with a larger-scale storm system, rather than with the preceding thunderstorm. Similar observations were made for data from Kennedy Airport and from LaGuardia Airport, and therefore, intervals of 1 h both before the reported beginning times and after the reported ending times were deemed appropriate for extending thunderstorm windows in the New York City area. However, in regions with different climates, different values of Δw_- and Δw_+ may be appropriate. In the *ASOS-WX* software, the intervals for extending thunderstorm windows in each direction can be independently specified by the user for classification of thunderstorm and non-thunderstorm winds.

4. Construction of data sets with reduced statistical dependence

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 peak wind reports are generally available from each storm. The resulting data set thus contains "clusters" of wind speeds from each storm, and 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 Harris (1999), 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 Fig. 2). Simiu and Heckert (1996) 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 value 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 Simiu and Heckert (1996). This procedure ensures that no two wind speeds are separated by less than a specified minimum separation interval, denoted Δt_{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 Δt_{min} for these distinct types of winds. Let $\mathbf{t} = [t_1 \;\; t_2 \cdots t_N]$ denote the vector of date/time values for peak wind speeds extracted from ASOS weather reports, sorted in ascending order, and let $\mathbf{s} = [s_1 \;\; s_2 \cdots s_N]$ denote the corresponding vector of peak wind speeds, where N is the total number of wind speeds. 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 $t_2 - t_1 \geqslant \Delta t_{min}$, then the difference between the next pair of time values $t_3 - t_2$ is checked, 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 from $\mathbf{s}$ 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 from $\mathbf{s}$ and s_k is retained, because s_k has a larger separation interval from subsequent wind speeds. When a wind speed value is deleted from $\mathbf{s}$, the corresponding time value is also deleted from $\mathbf{t}$, 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 Δt_{min}. In the implementation of this procedure in the *ASOS-WX* software, 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 Fig. 5 for both thunderstorm and non-thunderstorm wind speed data from Newark Airport over a period from May to June of 2000. A separation interval of $\Delta t_{min} = 4$ d was used for non-thunderstorm winds in Fig. 5(a), while a separation interval of $\Delta t_{min} = 6$ h was used for thunderstorm winds in Fig. 5(b). In both cases, the surviving wind speeds are indicated with circles. The values of the separation intervals used in Fig. 5 should not be taken as definitive recommendations. Rather, the influence of the separation interval should be investigated using data from specific stations, as different values of the separation intervals may be appropriate for different climates.

Fig. 6 shows the influence of the separation interval Δt_{min} on the number of data points per year that survive the separation procedure for the three stations in the New York City area. Results for both thunderstorm and non-thunderstorm winds are presented using data from a period of about 20 years, and only wind speeds greater than 35 knots are included, because of the change in reporting threshold shown in Fig. 2. Fig. 6 shows that the number of surviving data points stabilizes as the separation interval Δt_{min} exceeds the duration of most storm systems. Based

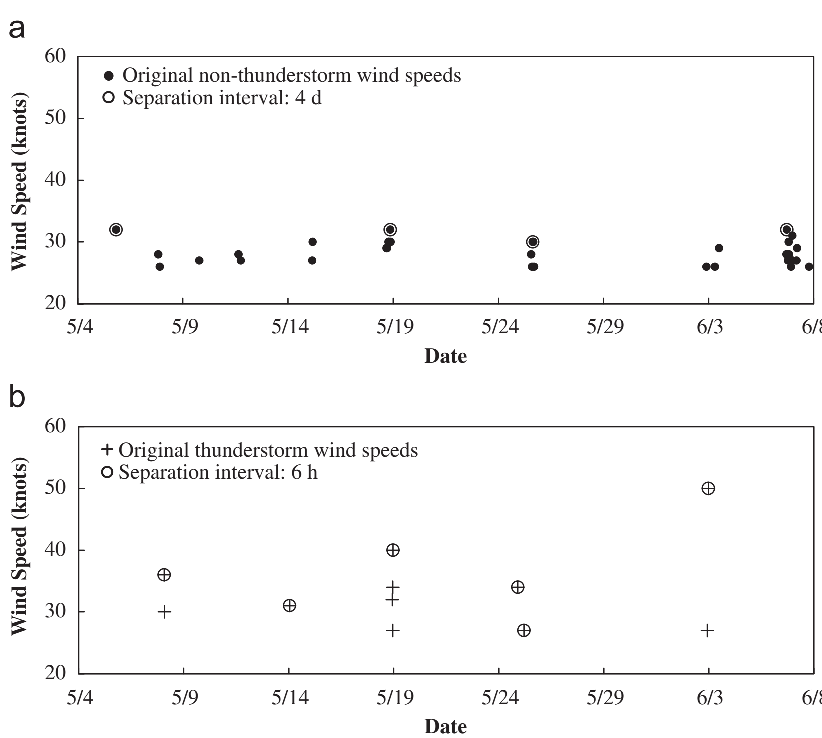

Fig. 5. Construction of data sets with reduced statistical dependence (raw data from NCDC Data Set 9956 for Newark Airport, year 2000). (a) Non-thunderstorm; (b) thunderstorm.

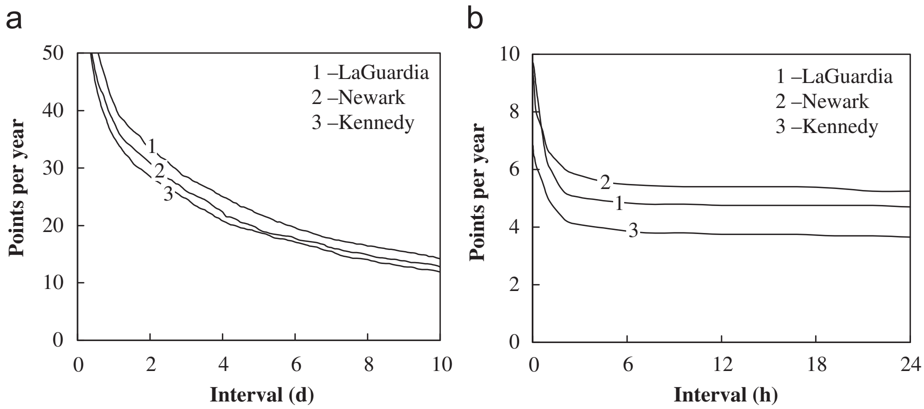

Fig. 6. Influence of separation interval on number of surviving data points. (a) Non-thunderstorm; (b) thunderstorm.

on statistical tests reported by Thom (1964), Simiu and Heckert (1996) indicate that durations of four to eight days are typical for non-thunderstorm systems, and Fig. 6(a) shows that the number of surviving non-thunderstorm wind speeds plateaus over this range. Brabson and Palutikof (2000), in a study of wind speeds in Scotland, found a separation interval or "dead time" of 1 d to be appropriate for 3 s gust speeds greater than 30 m/s (58 knots) and about 2 d for speeds greater than 24 m/s (47 knots). Although they did not study wind speeds lower than 47 knots, it can be inferred that as the wind speed approaches the ASOS threshold of 18 m/s (35 knots), the Δt_{min} interval chosen for this paper is reasonable. Robinson and Easterling (1988) report that most thunderstorms have durations less than 6 h, which is consistent with Fig. 3, and accordingly, Fig. 6(b) shows that the number of surviving thunderstorm wind speeds drops only slightly as the separation interval increases from 6 h to 12 h. This plateauing in the number of data points is an indication that "clustering" in the data has been eliminated and that the resulting data set includes no more than one data point from a single storm system or from a cluster of closely spaced storms. However, further study is needed before definitive recommendations can be made as to appropriate values of Δt_{min} for thunderstorm and non-thunderstorm winds, because the plots in Fig. 6 are based on relatively small quantities of data—especially for thunderstorm winds. It is suggested that the influence of Δt_{min} on the resulting extreme value statistics be investigated as data from larger numbers of stations will be subjected to extreme value analyses.

Fig. 7 shows a polar plot of thunderstorm and non-thunderstorm wind speed and wind direction data from Newark Airport that were obtained by applying the separation procedure described above to thunderstorm and non-thunderstorm wind data with the same separation intervals as in Fig. 5. These data were obtained from the original data shown in Fig. 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. The directionality of the wind climate is clearly evident in Fig. 7, with most of the strong wind speeds directed from the northwest for both thunderstorm and non-thunderstorm winds. Database-assisted design (e.g., Main and Fritz, 2006) provides a unified framework for using such directional wind speed data in structural design.

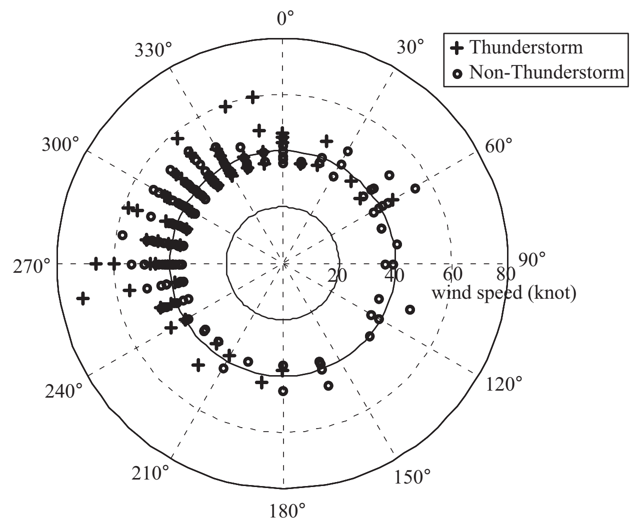

Fig. 7. Wind speed and direction data with reduced statistical dependence (raw data from NCDC Data Set 9956 for Newark Airport).

5. Extreme value analysis of separated wind speeds

Using the procedures outlined above, NT and T wind speed data sets spanning a period of approximately 20 years were obtained from each of the three ASOS stations near New York City, and extreme value analysis was performed on these data sets. Wind speeds associated with tropical storms were excluded from the data sets, so that the NT wind data correspond to extratropical synoptic storm wind speeds. Tropical storm wind speeds in the original data sets were due to storms whose energy remained mostly offshore or which were otherwise remnants of once stronger tropical systems. A notable example is Hurricane Gloria, a weak category 2 hurricane that made landfall on Long Island in September 1985 and produced wind speeds, according to the NYC stations, of over 50 knots at both Kennedy and LaGuardia airports. Several of these wind speeds, if not excluded, were found to slightly influence the results of the extreme value analysis for NT winds. For locations with strong and frequent hurricane wind speeds, a separate extreme value analysis should be performed for hurricane winds, in addition to T winds and NT (extratropical synoptic storm) winds.

The minimum separation interval between successive NT wind speeds used in extreme value analysis was $\Delta t_{min} = 4$ d. It was found that differences between extreme wind speed estimates based on data sets with at least 4 d separation on the one hand and at least 8 d separation on the other hand were negligible in practice. It was also found that estimates of T speeds from data sets with $\Delta t_{min} = 6$ h between successive peaks differ negligibly from estimates based on sets with 4 d separations. In the extreme value analysis of T speeds, $\Delta t_{\min} = 12\,\text{h}$ was chosen to be conservative. Note that the appropriate thunderstorm and non-thunderstorm separation intervals may vary depending upon geographical location and the climate at these particular locations. In the extreme value analysis of commingled T and NT wind speeds a minimum separation interval of $\Delta t_{\min} = 4\,\text{d}$ was used, the same as for NT wind speeds.

Two types of extreme value analysis were performed: peaks over threshold analyses and epochal analyses (see e.g., Simiu and Miyata, 2006, p. 29). POT estimates are deemed to be approximately correct for ranges of thresholds over which the wind speed estimates are reasonably constant. The POT analyses yield the parameters—including the tail length parameter c—of the generalized Pareto distribution (GPD) assumed to best fit the differences between a sample's wind speeds and the threshold being considered (see Appendix A). Wind speeds are recorded in integer units, and to avoid large numbers of identical speeds, the recorded speeds are transformed through the addition of a fractional part drawn randomly from a uniformly distributed set contained in the interval $[-0.5,+0.5]$.

In the US it has been commonly assumed that extreme wind distribution tails are of the Gumbel type, which corresponds asymptotically to a generalized Pareto distribution with tail length parameter $c = 0$. However, most data sets analyzed in this paper and elsewhere have estimated tail length parameter $c<0$ (see e.g., Simiu and Heckert, 1996), i.e., their tails correspond to reverse Weibull distributions. The Australian/New Zealand Standard AS/NZS 1170.2, Supplement 1:2002 (2002) also uses estimates based on the assumption $c<0$ (see Standard's Commentary Section C3.2, p. 14).

The Gumbel distribution has been used for the estimation of extreme non-hurricane speeds in the ASCE 7 Standard and is widely accepted among structural engineers. To avoid large differences between currently accepted estimates and estimates corresponding to best fitting distributions for which the tail is of the reverse Weibull type, it is appropriate for practical wind speed estimation purposes to use the value $c = -0.1$ whenever the estimated tail length parameter is $c<-0.1$ (tails corresponding to $c = -0.1$ being typically longer than tails corresponding to $c<-0.1$). The procedure for estimating extreme wind speeds for which the parameter c is prescribed is described in the Appendix A. Calculations based on simulated data sets indicated that probability distributions obtained by this procedure fit the data reasonably well. It was also verified that the prescribed value $c = -0.005$ yields a distribution with tail essentially equivalent to a Gumbel distribution tail, which was also used in Brabson and Palutikof (2000). (Strictly speaking, a Gumbel distribution tail corresponds to $c = 0$.) Unless otherwise indicated, the results presented in this section were obtained by using the POT approach.

Figs. 8–10 show the results of extreme value analysis of wind speeds from the Newark, Kennedy, and LaGuardia ASOS stations, respectively. In each figure, the lowest row of plots shows the estimated tail length parameter c as a function of threshold, and the upper three rows show estimates of the 50, 500, and 2000 year wind speeds, also as functions of threshold. In addition to estimates of speeds corresponding to the estimated tail length parameter c (shown as solid lines), Figs. 8–10 show estimates based on the assumption that $c = -0.1$ and on the assumption that $c = -0.005$ (i.e., that the best fitting distribution's tail is approximately Gumbel).

5.1. Non-thunderstorm wind speeds

Column (a) of Fig. 8 shows the results of the extreme value analysis of NT wind speeds at Newark Airport. As expected, the estimates based on the approximately Gumbel distribution ($c = -0.005$) are larger than those based on the best fitting distribution with parameter c and on the distribution with the prescribed parameter $c = -0.1$. The estimated c values are all less than the prescribed value of $c = -0.1$, but tend to vary as a function of threshold. Brabson and Palutikof (2000) found similar results for the extratropical wind climate of Scotland. Letchford and Ghosalkar (2004) also found the c parameter less than zero for both NT and T winds in West Texas. The results for Kennedy and LaGuardia airports (Figs. 9 and 10) are qualitatively similar to those for Newark. Note that at LaGuardia, the estimated $c \approx -0.1$ for higher thresholds, so there is little difference between estimates based on the assumed value $c = -0.1$ and those based on the estimated value of c (column (a) of Fig. 10). As noted above, these results were based on NT data sets from which hurricane

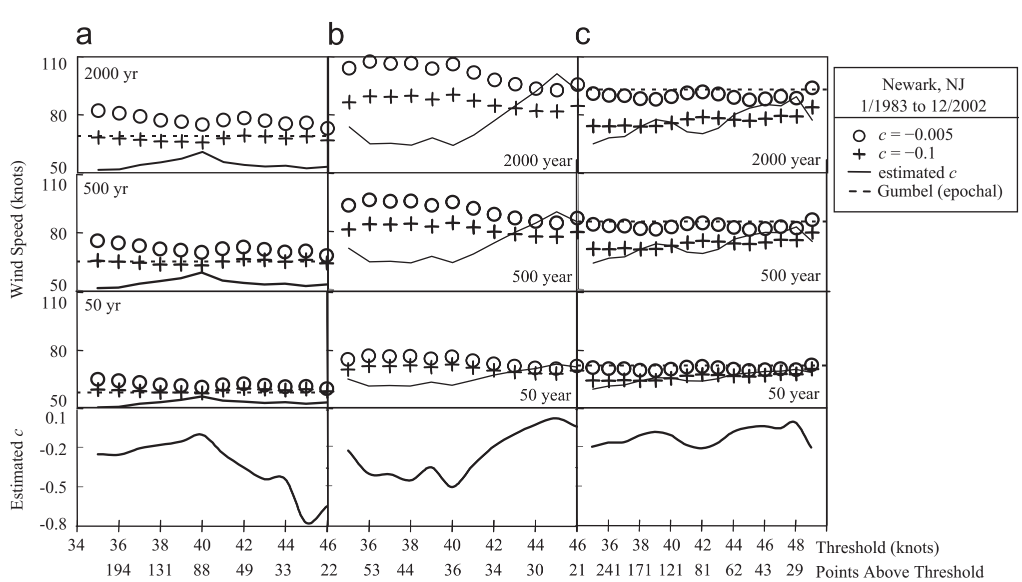

Fig. 8. Estimates of extreme wind speeds and tail length parameters for Newark Airport obtained using (a) non-thunderstorm, (b) thunderstorm, and (c) commingled wind data. Dashed lines represent estimates based on Gumbel distributions fitted to sets of 20 largest yearly speeds (epochal approach).

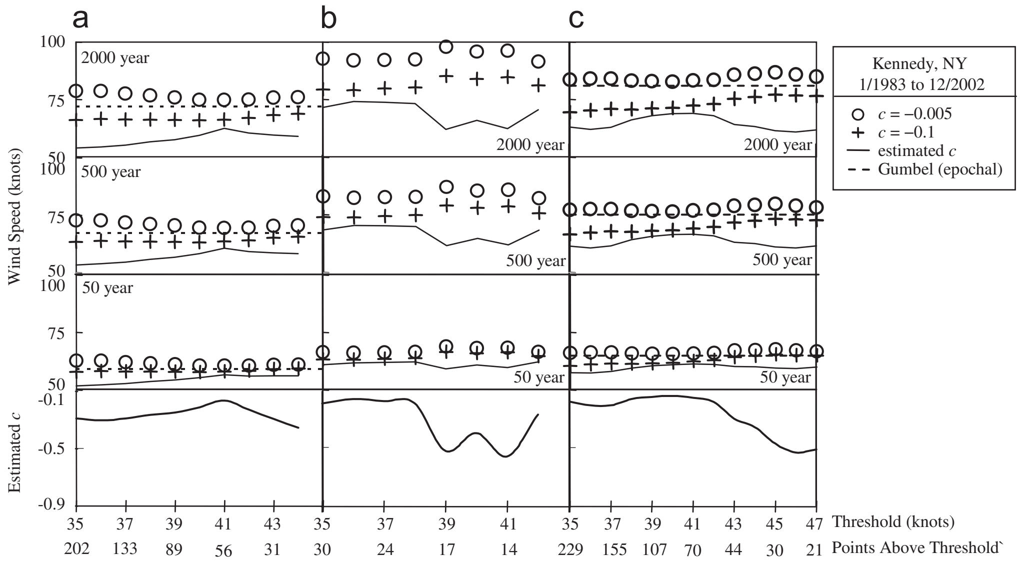

Fig. 9. Estimates of extreme wind speeds and tail length parameters for Kennedy Airport obtained using (a) non-thunderstorm, (b) thunderstorm, and (c) commingled wind data. Dashed lines represent estimates based on Gumbel distributions fitted to sets of 20 largest yearly speeds (epochal approach).

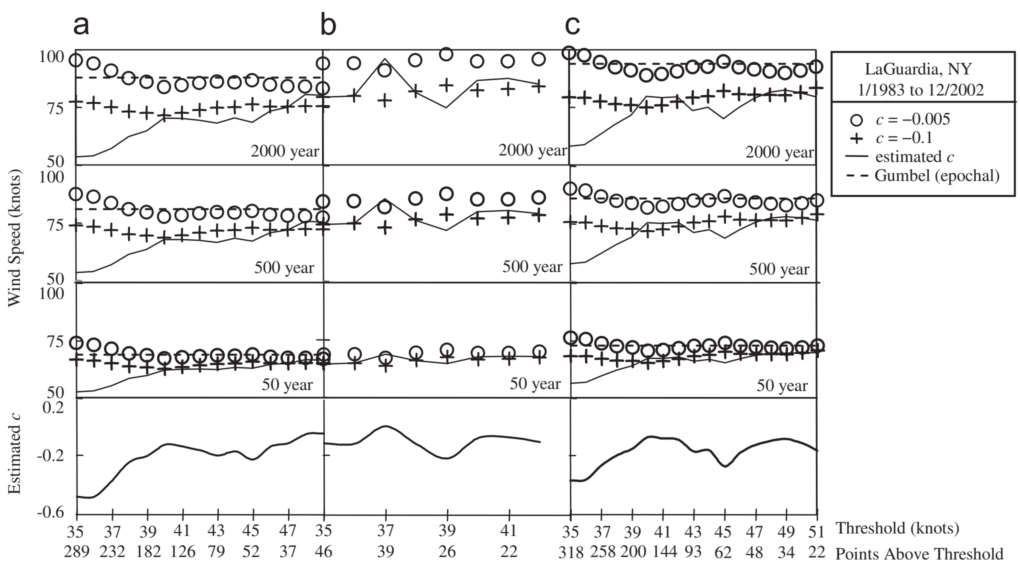

Fig. 10. Estimates of extreme wind speeds and tail length parameters for LaGuardia Airport obtained using (a) non-thunderstorm, (b) thunderstorm, and (c) commingled wind data. Dashed lines represent estimates based on Gumbel distributions fitted to sets of 20 largest yearly speeds (epochal approach).

wind speeds, which occurred on rare occasions during the period of record, had been eliminated. The results shown were found to slightly differ from those obtained by leaving the hurricane data points in the data sets. Larger differences might be expected if very strong hurricane wind speeds had occurred during the period of record.

5.2. *Thunderstorm wind speeds*

Column (b) of Fig. 8 shows estimates for T wind speeds at Newark Airport. It appears that estimates corresponding to thresholds between 18.5 m/s (36 knots) and 20 m/s (39 knots), say, are appropriate. Note also that, for each of the three mean recurrence intervals, the estimated T wind speeds are higher than their NT counterparts. This is also true for Kennedy and LaGuardia (column (b) of Figs. 9 and 10). Twisdale and Vickery (1992) also examined LaGuardia Airport data, and found that NT wind speeds dominated the wind climate. However, in the present study a number of high T wind speeds were identified that occurred after the period analyzed by Twisdale and Vickery (1992), including peak gust speeds of 70, 62, and 60 knots, which accounted for three of the top 10 wind speeds over the 20 year period, including the highest. Also, a high NT wind speed recorded during the period of Twisdale and Vickery's study was eliminated from the present analysis because it was associated with a tropical storm system. These factors, perhaps more importantly than the different approaches for classifying T and NT winds, are believed to contribute to the different conclusions regarding the importance of T winds at LaGuardia.

5.3. *Commingled data sets*

Commingled data sets are sets that include all wind speeds exceeding the threshold being considered, regardless of whether those speeds are associated with T or NT winds. Results of analyses based on commingled sets are shown in column (c) of Figs. 8–10.

In the US extreme wind speeds analyses have routinely been estimated on the basis of commingled sets. However, estimates that are appropriate from both a physical and a statistical point of view should be based on mixed distributions (Eq. (1)). Those based on commingled data sets will at least in principle yield incorrect results. Commingled distributions, containing wind speed data from two or more distributions have sometimes been confused with a type II, or Frechet distribution (Gomes and Vickery, 1978). Comparisons between results in Figs. 8–10 show that, for all three stations, the analyses based on commingled data sets yield unconservative estimates of the extreme wind speeds. However, if it is postulated that the probability distribution of commingled sets has Gumbel type tail, then the underestimation of the extreme wind speeds due to the use of commingled sets and the overestimation of the extreme wind speeds inherent in the use of the Gumbel distribution balance each other to a large extent, and yield estimates of extreme speeds that are close to those obtained by accounting for the distinct probabilities of NT and T wind speeds.

5.4. Estimates based on epochal approach

The results discussed so far were based on the POT approach. Estimates based on largest annual wind speeds and the assumption that all distributions are Gumbel are also shown in Figs. 8–10 by horizontal dotted lines. Gumbel wind speed estimates have been shown to be less sensitive to the length of the data set (i.e., the duration in years) than estimates obtained using the GPD (Brabson and Palutikof, 2000). It has also been argued that all wind speeds in the upper tail correspond to Gumbel distribution if convergence is accounted for (Cook et al., 2003). The fact that the epochal estimates are represented, for convenience, in plots whose abscissas are threshold values should not be construed as meaning that the estimates are functions of threshold; in fact, the estimates have nothing whatsoever to do with any threshold considerations. Note that the ASOS data are censored below at 18 m/s (35 knots). In conducting the epochal analyses it was verified for the NT and commingled data that no largest annual wind speed was lower than 18 m/s (35 knots), and the annual maxima data sets are therefore unaffected by this censoring. For the T data, however, some years existed in which no wind speed exceeded 18 m/s (35 knots), and therefore, epochal estimates are not presented for the T wind speed data.

5.5. Mixed distributions

For structural engineering purposes it is of interest to estimate extreme wind speeds with specified mean recurrence intervals regardless of the type of storm with which the extreme winds are associated. For T and NT speeds, the following expression holds:

$$P[\max(v_T, v_{NT}) \leqslant V] = P(v_T \leqslant V)P(v_{NT} \leqslant V) \tag{1}$$

where the left-hand side denotes the probability that T and NT wind speeds are less than V; $P(v_T \leqslant V)$ and $P(v_{NT} \leqslant V)$ denote, respectively, the probability that T winds are less than V and the probability that NT winds are less than V. Eq. (1) is a consequence of the mutual independence of T and NT winds (see Simiu and Miyata, 2006).

The nature of the mixed distribution is illustrated in Fig. 11 using data from LaGuardia Airport. Curves of wind speed versus return period are shown for the mixed distribution based on Eq. (1) (labeled M), for a GPD fit to the T wind speeds alone (labeled T), for a GPD fit to the NT wind speeds alone (labeled NT), and for a GPD fit to commingled T and NT wind speeds (labeled C). The GPD fits used in Fig. 11 are based on the assumption that $c = -0.005$ and on a threshold of $u = 40$ knots. The mixed distribution yields the highest wind speeds over the entire range of return periods, converging to the commingled distribution for short return periods and converging to the T distribution for long return periods. Comparing the T and NT distributions shows that the NT winds are dominant for very short return periods, while the T winds are dominant for long return periods. While the commingled distribution matches the mixed distribution at very short return periods, it yields substantially smaller wind speeds at long return periods. For return periods of 500 years and greater, the mixed distribution is virtually indistinguishable from the T distribution, meaning that the estimated extreme speeds are determined solely by the T wind speeds.

Figs. 12 and 13 show similar plots for Newark and Kennedy airports, respectively. At these stations, the mixed distribution converged to the T distribution at lower return periods than for LaGuardia, and for return periods of 50 years and greater, the wind

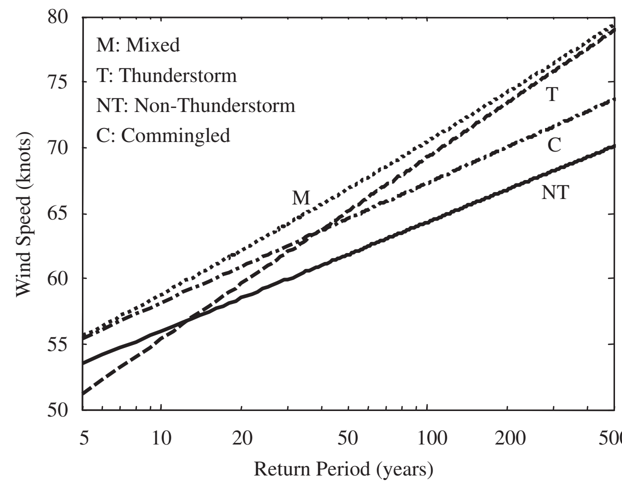

Fig. 11. Estimates of wind speed versus return period for LaGuardia Airport obtained using a mixed distribution (M), shown with estimates based on a GPD fit to T wind speeds alone (T), NT wind speeds alone (NT), and commingled T and NT wind speeds (C). ($u = 40$ knots, $c = -0.005$.)

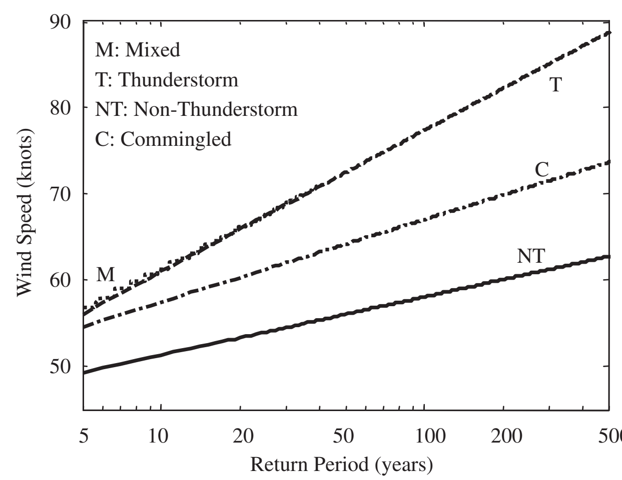

Fig. 12. Estimates of wind speed versus return period for Newark Airport obtained using a mixed distribution (M), shown with estimates based on a GPD fit to T wind speeds alone (T), NT wind speeds alone (NT), and commingled T and NT wind speeds (C). ($u = 40$ knots, $c = -0.005$.)

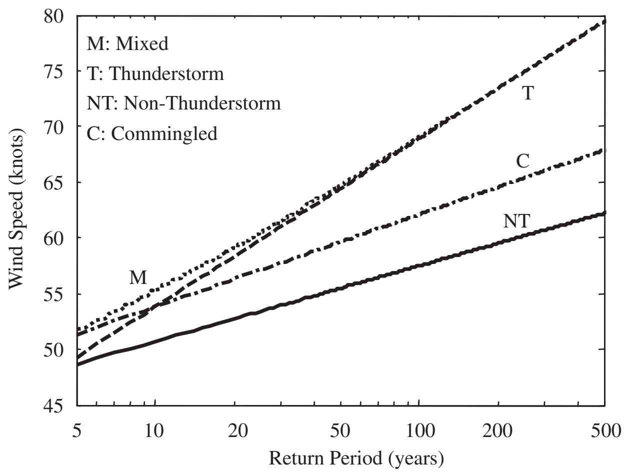

Fig. 13. Estimates of wind speed versus return period for Kennedy Airport obtained using a mixed distribution (M), shown with estimates based on a GPD fit to T wind speeds alone (T), NT wind speeds alone (NT), and commingled T and NT wind speeds (C). ($u = 40$ knots, $c = -0.005$.)

speeds are determined solely by the T wind speeds. It is thus observed that T wind speeds dominate the extreme wind climate at long return periods for all three NYC stations. Twisdale and Vickery (1992) found similar results for Dallas and Minneapolis, and Holmes (2001) found similar results for Melbourne.

6. Conclusions

In this paper, procedures have been described for extracting peak gust wind data and thunderstorm observations from archived ASOS weather reports in NCDC Data Set 9956, and for classifying the resulting wind data as thunderstorm or non-thunderstorm. The procedure for identification of thunderstorm wind data involves comparing the date and time of each peak wind report with reported intervals of thunderstorm occurrence. To handle errors encountered in thunderstorm reports, a procedure for defining time windows of thunderstorm occurrence has been described that makes use of manual weather observations in conjunction with reported thunderstorm beginning and ending times. Using data from three ASOS stations in the New York City area, a significant number of high wind speeds were found within the 30 min intervals preceding reported thunderstorm beginnings, most likely associated with thunderstorm outflow boundaries. For identification of thunderstorm winds, the thunderstorm windows can be extended by specified time intervals before the reported thunderstorm beginnings and after the reported endings, and intervals of 1 h in each direction were deemed appropriate for the three stations near New York City.

A modified procedure for constructing data sets with reduced statistical dependence has also been presented, which involves specifying a minimum separation interval Δt_{min} for the resulting data. This procedure can be applied separately to thunderstorm and non-thunderstorm wind data, to account for the difference in the typical durations of these weather systems. The influence of the separation interval Δt_{min} on the number of surviving data points has been investigated for both thunderstorm and non-thunderstorm winds, and in both cases the number of data points was observed to stabilize as Δt_{min} exceeded the duration of typical storm systems. It is noted that different values of the separation intervals may be appropriate for other climates. The procedures described in this paper have been implemented in a publicly available software package called *ASOS-WX*, which is freely available for download at www.nist.gov/wind. The software and procedures presented in this paper hold the potential to significantly expand the body of wind data available for structural engineering purposes. These data could be used for improving the current wind speed map in ASCE 7 and could also be used within the database-assisted design framework.

These capabilities make it possible to conduct extreme value analyses based on a realistic representation of the probability distribution of extreme wind speeds in climates in which both types of wind occur. Extreme wind speeds were estimated by using probability distributions that account in a physically and probabilistically rigorous manner for the individual probability distributions of thunderstorm wind speeds on the one hand and non-thunderstorm wind speeds on the other. Estimates obtained for Newark Airport, NJ, LaGuardia Airport, NY, and Kennedy Airport, NY, showed that, at those stations, thunderstorm wind speeds dominate the extreme wind climate at long return periods. For return periods greater than 50 years (greater than 500 years for LaGuardia) thunderstorm winds were dominant to such an extent that non-thunderstorm wind speeds can be disregarded in the analysis.

Results of the analyses at the three stations were found to have negligible dependence on whether the Δt_{min} between successive peak wind speeds was 4 d or 8 d for non-thunderstorm wind speeds, and 6 h or 12 h for non-thunderstorm wind speeds. This conclusion is tentative and its validity for other stations would need to be checked.

Commingled data sets (i.e., sets containing, indiscriminately, both non-thunderstorm and thunderstorm wind speeds) have been used almost exclusively for extreme wind speed estimates in the United States. Comparisons between results of analyses based on separate sets of non-thunderstorm and thunderstorm wind speeds on one hand and on commingled data sets on the other suggest that the latter may yield unconservative results. However, if for the sake of conservatism it was postulated that the tail of the distribution based on commingled data sets is of the Gumbel type (even though the best fitting distributions had shorter tails than the Gumbel distribution), then the estimated extreme wind speeds were found to be only marginally larger than those based on mixed distributions and the assumption that $c = -0.1$.

Analyses conducted at numerous additional stations throughout the US will be needed to check the validity of estimates similar to those presented in this paper. To improve the reliability of the estimates such analyses could be based on sets that would include converted fastest-mile wind speed data recorded before the institution of ASOS as well as the identification of additional thunderstorm wind speed data over a longer period of time. Improved and standardized meteorological observations and methods would also be helpful.

Acknowledgments

William Brown of the National Climatic Data Center is gratefully acknowledged for providing the data sets analyzed in this report and for assistance in interpreting these data. Al Wissman and Melody Magnus of the National Weather Service are also gratefully acknowledged for providing additional site-specific information about ASOS stations. J.A. Main performed this work during his tenure as NIST/NRC Postdoctoral Research Associate. Texas Tech University and the National Science Foundation (Grant no. 0221688) provided part of the funding for this research through support of F.T. Lombardo's tenure as a student intern at NIST during phases of this work.

Disclaimer

The policy of the National Institute of Standards and Technology is to use the International System of Units (metric units) in all its publications. In this document, however, works of other authors outside NIST are cited which describe measurements in certain non-SI units. Specifically, figures present wind speed data in knots. Certain trade names or company products are mentioned in the text to specify adequately the procedure used. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the product is the best available for the purpose.

Appendix A

Let the number of data above the threshold u be denoted by k. The speed u and the k speeds larger than u form a set of size k+1. The rate of arrival of wind speeds larger than u is $\lambda = k/n_{\mathrm{yrs}}$, where n_{yrs} denotes the length of the record in years. The highest, second highest, …, kth, (k+1)th highest speeds are denoted by $X_{n,n}$, $X_{n-1,n},\ldots$, $X_{n-(k-1),n}$, $X_{n-k,n}\equiv u$, respectively, where n is the total number of data points. Compute the quantities:

$$M_n^{(r)} = \frac{1}{k}\sum_{i=0}^{k-1}[\log(X_{n-i,n}) - \log(X_{n-k,n})]^r, \quad r = 1,2 \tag{A.1}$$

The estimators of the tail length parameter c and location parameter a of the generalized Pareto distribution (see Eq. A1.36a, p. 605, Simiu and Scanlan, 1996) are from DeHaan (1994):

$$\hat{c} = M_n^{(1)} + 1 - \frac{1}{2\{1-[M_n^{(1)}]^2/[M_n^{(2)}]\}} \tag{A.2a}$$

$$\hat{a} = uM_n^{(1)}/\rho_1 \tag{A.2b}$$

$$\rho_1 = 1, \quad \hat{c}\geqslant 0; \quad \rho_1 = 1/(1-\hat{c}), \quad \hat{c}\geqslant 0 \tag{A.3}$$

If c is specified, an estimated value of the location parameter a can be obtained by substituting c for $\hat{c}$ in (A.3), and using (A.2b). It follows from the expression for the Generalized Pareto distribution that

$$\hat{y} = -\hat{a}[1-(\lambda R)^c]/c \tag{A.4}$$

where R is the mean recurrence interval in years. The wind speed with mean recurrence interval, R, x_n, is then

$$\hat{x}_r = \hat{y} + u \tag{A.5}$$

Software for the estimation of extreme wind speeds based on DeHaan expressions (A.2), (A.3) and (A.5), is provided at the site www.nist.gov/wind.

References

Aircraft Owners and Pilots Association (AOPA), 1999. ASOS: automated surface observing system. Air Safety Foundation Safety Advisor, Technology no. 2. 〈www.aopa.org/asf/publications/sa09.pdf〉 (accessed 30.03.06).

American Society of Civil Engineers (ASCE), 2006. Minimum design loads for buildings and other structures. ASCE/SEI 7-05, Reston, VA.

AS/NZS 1170.2 Supplement 1:2002, 2002. Structural design actions—wind actions—commentary.

Brabson, B.B., Palutikof, J.P., 2000. Tests of the generalized Pareto distribution for predicting extreme wind speeds. J. Appl. Meteorol. 39, 1627–1640.

Choi, E.C.C., 1999. Extreme wind characteristics over Singapore—an area in the equatorial belt. J. Wind Eng. Ind. Aerodyn. 83, 61–69.

Choi, E.C.C., Tanurdjaja, A., 2002. Extreme wind studies in Singapore. An area with mixed weather system. J. Wind Eng. Ind. Aerodyn. 90, 1611–1630.

Cook, N.J., Harris, R.I., Whiting, R., 2003. Extreme wind speeds in mixed climates revisited. J. Wind Eng. Ind. Aerodyn. 91, 403–422.

DeHaan, L., 1994. In: Galambos, J., Lechner, J.A., Simiu, E. (Eds.), Extreme Value Theory and Applications, vol. 1. Kluwer Academic Press, Dordrecht.

Gomes, L., Vickery, B.J., 1978. Extreme wind speeds in mixed climates. J. Ind. Aerodyn. 2, 331–344.

Harris, R.I., 1999. Improvements to the method of independent storms. J. Wind Eng. Ind. Aerodyn. 80, 1–30.

Holmes, J.D., 2001. Wind Loading of Structures. Spon Press, London.

Letchford, C., Ghosalkar, M., 2004. Extreme wind speed climatology in the United States Mid-West. In: The Sixth UK Conference on Wind Engineering, Cranfield University, UK.

Lombardo, F.T., Main, J.A., 2006. Documentation for ASOS-WX software. 〈www.itl.nist.gov/div898/winds/asos-wx/asos-wx-doc.pdf〉 (accessed 15.03.06).

Main, J.A., Fritz, W.P., 2006. Database-assisted design for wind: concepts, software, and examples for rigid and flexible buildings. In: Building Science Series, vol. 1806. National Institute of Standards and Technology, Gaithersburg, MD, USA.

The MathWorks, Inc., 2006. MATLAB and Simulink for technical computing. 〈www.mathworks.com〉 (accessed 15.03.06).

National Climatic Data Center (NCDC), 2002. Data documentation for Data Set 6421 (DSI-6421). 〈www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td6421.pdf〉 (accessed 11.07.07).

National Climatic Data Center (NCDC), 2003. Data documentation for Data Set 9956 (DSI-9956). 〈www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td9956.pdf〉 (accessed 11.07.07).

National Climatic Data Center (NCDC), 2006. Data documentation for Data Set 3505 (DSI-3505). 〈www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td3505.pdf〉 (accessed 11.07.07).

National Hurricane Center (NHC), 2006. NHC/TPC archive of past hurricane seasons. 〈www.nhc.noaa.gov/pastall.shtml〉 (accessed 17.03.06).

National Oceanic and Atmospheric Administration (NOAA), 2005. Federal Meteorological Handbook No. 1: Surface Weather Observations and Reports (FCM-H1-2005). 〈www.ofcm.gov/fmh-1/pdf/FMH1.pdf〉 (accessed 17.03.06).

National Weather Service (NWS), 1998. Automated Surface Observing System (ASOS) User's Guide. 〈www.weather.gov/asos/aum-toc.pdf〉 (accessed 30.03.06).

Neumann, C.J., Jarvinen, B.R., McAdie, C.J., Elms, J.D., 1993. Tropical Cyclones of the North Atlantic Ocean, 1871–1992. In: Historical Climatology Series 6-2. National Climatic Data Center, Asheville, NC, USA.

Peterka, J.A., 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.

Peterka, J.A., Shahid, S., 1998. Design gust wind speeds in the United States. J. Struct. Eng. 124 (2), 207–214.

Robinson, P.J., Easterling, D.R., 1988. The frequency distribution of thunderstorm durations. J. Appl. Meteorol. 27 (1), 77–82.

Simiu, E., Heckert, N.A., 1996. Extreme wind distribution tails: a 'peaks over threshold' approach. J. Struct. Eng. 122 (5), 539–547.

Simiu, E., Scanlan, R.H., 1996. Wind Effects on Structures, third ed. Wiley, New York.

Simiu, E., Wilcox, R., Sadek, F., Filiben, J.J., 2003. Wind speeds in ASCE 7 standard peak gust map: assessment. J. Struct. Eng. 129 (4), 427–439.

Simiu, E., Wilcox, R., Sadek, F., Filiben, J.J., 2005. Closure to '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), 997–998.

Simiu, E., Miyata, T., 2006. Design of Buildings and Bridges for Wind. Wiley, Hoboken.

Sparks, P.R., 1999. Using wind data from the automated surface observing system. J. Roof Consultants Inst. 6, 24–28.

Thom, H.C.S., 1964. Prediction of design and operating velocities for large steerable radio antennas. Large steerable radio antennas—climatological and aerodynamic considerations. Ann. N. Y. Acad. Sci. 116, 90–100.

Twisdale, L.A., Vickery, P.J., 1992. Research on thunderstorm wind design parameters. J. Wind Eng. Ind. Aerodyn. 41–44, 545–556.