Automated classification of convective downburst events in wind gust observations.

Nicholas J Cook

Wind Engineer, Highcliffe-on-Sea, Dorset, UK.

Orcid ID: 0000-0002-6390-6884

Email: wind@njcook.uk

Abstract

Wind observations near the ground are critical in assessing the impact of wind on structures. All wind climates comprise a mixture of several disjoint meteorological mechanisms that require separation before assessment. In this paper previous studies distinguishing between convective and non-convective gust events are reviewed. Classification by visual inspection of the gust speed timeseries is generally agreed to be easy and accurate, but it becomes impractical for very large datasets. Recent automated approaches, using statistics, pattern recognition and neural networks, are calibrated against 4000 visually classified gust events from 20 locations across the USA over 22 years. The most promising method is developed to use only gust speed statistics to distinguish five classes of gust event: synoptic scale storms, deep convection, the forward flank and the rear flank of gust fronts, and downbursts from isolated thunderstorms. A 6th class collects non-meteorological artefacts in the data. The ensemble-averaged timeseries of each class form a distinctive hierarchy. The misclassification error rate against the visual classification is 7.8%, with most errors between adjacent classes. When applied to $>10^7$ gust events ≥20kn from 450 locations across the USA, the class hierarchy remains stable. The method is implemented by open-source R scripts.

Keywords: Thunderstorms, frontal downbursts, mesoscale gusts, synoptic-scale gusts, ASOS, kernel density estimation, k-means, neural network, shapelet transform, fuzzy membership.

Acronyms

ABL	Atmospheric boundary layer
ASOS	Automated Surface Observing System of the US National Weather Service
CM	Confusion matrix
CONUS	Contiguous United States
FN	False negative
FP	False positive
HKD	Highest Kernel Density method

KDE	Kernel Density Estimation
METAR	Meteorological Aerodrome Report of current weather
NCEI	US National Center for Environmental Information
NN	Artificial Neural Network
NOAA	US National Oceanic and Atmospheric Administration
QC	Quality control
T/NT	Binary thunderstorm – non-thunderstorm classification
UTC	Coordinated Universal Time ("Zulu")
WMO	World Meteorological Organisation

1. Introduction

Wind observations in the atmospheric boundary layer (ABL) near the ground surface form the critical first link in the Davenport Chain [1] for assessing the impact of wind on the built environment. A concept now generally accepted is that all wind climates are mixed in the sense that physically different wind mechanisms govern surface winds at different times in an exclusive, or disjoint, manner. Gomes and Vickery [2] were first to recognise the need to separate observations from the various mechanisms for analysis before assessing their joint contribution to extreme winds, especially from rare (e.g., hurricanes and tropical cyclones) or localised (e.g., thunderstorm, tornado) events. Now that analysis and prediction models for hurricanes and tropical cyclones are well established [3], the focus of attention has moved to thunderstorm and other convective downbursts. The transient, non-stationary, meso-scale nature of these convective gusts, mixed with the ABL gusts in synoptic-scale windstorms, makes their detection and study a particular challenge.

The importance of thunderstorms to Wind Engineering is discussed by Lombardo et. al. [4] who report they are the cause of most of the damage to structures across the continental United States (CONUS). In addition to the peak gust speed in a thunderstorm downburst, sudden off-axis changes in wind direction are a particular issue for horizontal-axis wind turbines [5][6][7]. Zhang et. al. [8] report thunderstorms as the cause of 20% of all turbine accidents. Downbursts are also a major hazard to aircraft on take-off and landing.

An impressive number of full-scale field studies have used a variety of approaches: single and groups of anemometers [10] [11] [12], vertical arrays on single towers [13][14] and on groups of towers [15], by Doppler RADAR, SODAR or LiDAR soundings [10][11][16][17], and in various combinations. Most employ the classical directional decomposition method for synoptic observations, with adjustments for the transient nature of downbursts, as addressed by [12]. Key consensus findings are that the direction of the peak gust speed is invariant with height [10] [11] [14], and that the depth of the gust front increases with time [16]. Downbursts are reported to be "arranged either randomly, in squall-lines, or in mesoscale convective systems" [11].

Identification and classification of gust events in a timeseries of observations are generally implemented either as two sequential processes in either order: a) identify all gust events, then classify each event; or b) pre-define classes, then identify events of that class. In chronological order of publication:

- 1992, Twisdale and Vickery [18] classified thunderstorm gusts as the maximum gust observed on "thunderdays" – days where thunder is seen or heard.
- 1999, Choi [19] identified thunderstorm gusts by visual search through 13 years of anemograph charts and other records.
- 2002, Choi and Hidayat [20] identified thunderstorms in the 20 highest gusts per year as coinciding with "thunder and rain" to calibrate the difference in gust factor between thunderstorm and monsoon winds in Singapore.
- 2002 Kasperski [21], for Germany, defined the three classes: depression, gust front, and thunderstorm, for gusts separated by 24h, identifying by the peak, mean and gust factor, but experienced difficulty in separating depressions from fronts.
- 2009 Lombardo et. al. [22] used the thunderstorm flag in METAR to identify hourly maximum gusts from thunderstorms in CONUS.
- 2014, De Gaetano et. al. [23] adopted a statistical approach, evaluating 12 parameters: peak 1s gust; 1-minute mean speed; 10-minute mean, and 1-hour mean: speed, direction, turbulence intensity, skewness, and kurtosis; of 2Hz-10Hz sonic anemometer data. They used various combinations of gust factor in a logic tree to identify the same three classes as Kasperski [21] but could not separate fronts and thunderstorms definitively – "*classifying an event not attributable to a depression (D) as a thunderstorm (T) or a gust front (F) is the ratio G10/G60: when it is less than 0.90, the event is usually a thunderstorm (T); when it is greater than 0.90, the event is usually a gust front (F).*"
- 2019, Huang et. al. [24] used daily outliers in gust, mean temp and mean humidity: Gust >15m/s, plus outliers of temperature and humidity within 20 minutes of the peak gust, to identify "thermally-developed" wind.
- 2019, Guerova et. al. [25] used instability indices and integrated water vapour to predict thunderstorm activity some minutes before observed lightning flashes. This is just one example of similar predictive methods using satellite data.
- 2020, Samanta et.al. [26]: Detected thunderstorm days for pre-monsoon winds at Kolkata from cloud-base height and potential temperature at the 850hPa level.

- 2020, Chen and Lombardo [27]: Built a convolutional Neural Network (NN) that differentiated between thunderstorm and synoptic gusts in 1-minute interval wind observations from the US Automated Surface Observing System (ASOS). This was trained using 76480 records of 91-minute duration, centred on a maximum gust ≥40kn, and mutually separated by ≥45 minutes. The thunderstorms in the training set were automatically identified from the thunderstorm flag in the corresponding METAR, as in [22]. Training the NN took 161 CPU hours. They noted an issue of false positive (FP) spikes in the ASOS data that were classified as thunderstorm gusts.
- 2022, Arul et al [9]: Used 240 non-stationary 1h-periods of wind speed to train a Stationary Shapelet Transform (SST) – a method originally developed [28] to find patterns in electro-cardiograms. Here a "shapelet" is a short timeseries that is characteristic of an event class. From 168 records of one-hour duration at 10Hz from sonic anemometers, and without presuming any class structure, an automatic process generated 35998 candidate shapelets which were winnowed down to 32 "mother shapelets" that best represented recurring shapes in the timeseries. On visually classifying these shapelets, 21 indicated thunderstorm gusts and 11 indicated synoptic gusts. Transforming the whole timeseries with each mother shapelet gave a set of coefficients containing peaks, each corresponding to a section of the timeseries matching a mother shapelet, so indicating a classified gust event. Although the classification is only binary, thunderstorm /non-thunderstorm (T/NT), multiple mother shapelets were required because of the high variability of thunderstorm shapes.

2. Motivation for this study

In each of the previous studies, summarised above, the methodology used one of three basic approaches:

a) Conventional statistical moments of the wind speed timeseries [21][23].
b) Pattern recognition applied to the wind speed timeseries [9][27].
c) Non-anemometric meteorological parameters [18][20][24][25][26].

Each approach has its strengths and weaknesses. Methods using satellite data, e.g., [25], are principally for short-term forecasting and nowcasting, and cannot be applied to long-term historical records. Dependence on the METAR thunder flag is appropriate in selecting gust events for calibration or training, but as a general classification criterion it risks producing numerous false positives (FP) and false negatives (FN), because not all thunderstorms are flagged, and many do not produce downbursts at the anemometer. A potential weakness of [23] is that it used high-frequency observations to evaluate the statistical moments and it is not immediately clear whether the method will operate successfully with sparser observations, e.g., the 1-minute interval in [27]. A strength of SST in [9] is as a "white box" method where all workings are transparent and monitorable by the user. The convolutional NN in [27] is a "black box" with all its workings hidden. A linear NN lies somewhere between these, since the output of each node is the simple weighted sum of its inputs, but the process of training these weights is generally hidden, so it could be said to be "grey".

Although some of the above studies share the same source observations, each stands on its own in the literature. These methods deserve a considered and fair intercomparison for their effectiveness. There is also the prospect of releasing some hidden synergy between the methods, to capitalise on their respective strengths. The binary T/NT classification in all the methods distributes downbursts in gust fronts between the two classes with, as reported, [21] [27] biased to NT and [9][23] biased to T. Thus, each class remains a mixture, rather than the intended exclusive class. De Gaetano et al [23] called for a method that definitively identifies gust front downbursts and deep-convection downdrafts. This curiosity-driven study examines whether this aim is possible using only anemometric data.

3. Data

This study uses the 1-minute interval weather observations from some of the almost 1000 Automated Surface Observation System (ASOS) stations across the contiguous USA (CONUS). These data are available in the NCEI TD6405 database for bulk download from ftp://ftp.ncdc.noaa.gov/pub/data/asos-onemin/ by year and station. Each file gives the 2-minute mean and 1-minute maximum 3s gust wind speed and direction at a resolution of 1kn and 1°. Initially from 2000 Belfort cup/vane anemometers with a 5s gust response were used, most at 10m above ground, but some at 26 feet (7.9m). Between 2005 and 2010 these were replaced by Vaisala sonic anemometers averaged to give the WMO standard 3s gust. The progress of the ASOS implementation and upgrade programme is documented in a series of NOAA reports currently available at https://weather,gov/asos/ASOSImplementation.

TD6405 is a most valuable resource for Wind Engineering because it allows study of mesoscale events over much longer observational periods than is possible with targeted measurement campaigns. Although presented [29] as a homogeneous set of data in fixed format text files, it is neither of these things owing to incremental changes in instrumentation, acquisition, and quality control (QC) procedures, as well as errors in transmission and archiving. In addition to the typical artefacts found in long-term wind records [30], which can be detected automatically [31], there are two recurring artefacts particular to TD6405:

1. Occasionally, data between 00:00 and 12:00 UTC on one day are archived as having occurred on the previous day in addition to the valid data for that day. This is correctable by exploiting serial continuity to determine which of each pair of observations at the same time belongs to which day.
2. Having no moving parts and heated in the winter to prevent icing, the sonic anemometers provide ideal perches for birds which block the acoustic path and generate spurious large "spikes" of gust speed on landing and take-off. These spikes are difficult to distinguish from instrumentation glitches and localised thermal events (e.g., dust devils) lasting less than one minute. Although not correctable, they can be detected and removed.

At the end of 2013, NOAA implemented a new QC test – "Test 10" – to the ASOS data to detect and remove spurious bird-generated gust values [32]. A curation of the ASOS data [33] reveals that Test 10 produces four times as many false positives (FP) than true positives (TP), and each FP unnecessarily culls the following 5 minutes of valid data. Although the proportion of lost data is small, less than 0.03% [32] and insignificant for the overall dataset, these false positives are strongly biased towards thunderstorm downbursts and gusts in otherwise calm periods [33]. The deleterious effects of this test became apparent in [33] and affect this study.

The automated procedures used here to identify and correct artefacts in the TD6405 data are fully described in [33]. They are principally intended to produce corrected synoptic-scale homogeneous datasets. Applied to this study, the automatic thresholds for a valid change in wind direction were found to cull some thunderstorm downbursts in light winds where direction changes are very large and can reverse. The wind direction artefact detection was disabled by setting an unreachably high threshold.

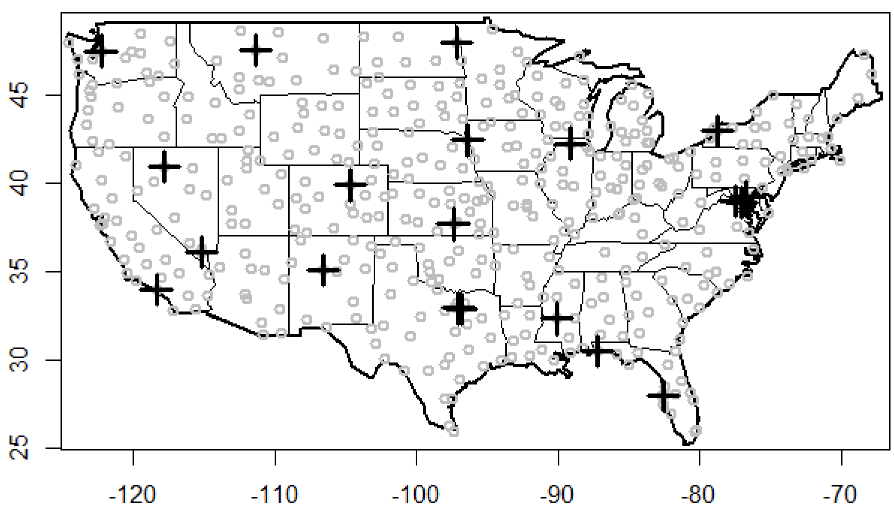


Figure 1. Locations of ASOS stations in the development set. Crosses: Development set of 20 stations. Circles: Analysis set of 450 stations.

A set of twenty ASOS stations, each with a nominal observational record from 1st January 2000 to 31st December 2021, were chosen for this study. Indicated by the crosses in Figure 1, they are distributed across CONUS, except for two stations serving Dallas, TX, and three stations serving Washington, DC, intended to expose any non-geographical disparities. Also shown by grey circles in Figure 1 are the locations of 450 ASOS stations with WMO Class 1 or 2 exposures [34], selected for fuller geographic cover. Table 1 lists relevant metadata for the 20 development stations.

Table 1. Parameters of the ASOS stations used in this study.

ICAO code	UTC (hours)	Elevation (feet)	Sonic, date installed	Latitude (degrees)	Longitude (degrees)	Station name
KABQ	-7	1618.5	22/05/2007	35.04191	-106.61545	AlbuquerqueNM
KBUF	-5	218.2	04/06/2009	42.94001	-78.73608	BuffaloNiagaraNY

KBWI	-5	47.5	20/09/2006	39.17332	-76.68414	BaltimoreWashingtonMD
KDAL	-6	134.1	28/05/2009	32.83838	-96.83584	DallasLoveFieldTX
KDCA	-5	3	26/09/2006	38.8472	-77.03454	WashingtonReaganDC
KDEN	-7	1650.2	12/09/2005	39.84661	-104.65624	DenverIntCO
KDFW	-6	170.7	27/05/2009	32.89744	-97.02196	DallasFtWorthTX
KGFK	-6	256.6	17/10/2002	47.94272	-97.18293	GrandForksND
KGTF	-7	1116.8	26/03/2007	47.4733	-111.3828	GreatFallsMT
KIAD	-5	88.4	03/10/2006	38.93487	-77.44728	WashingtonDullesVA
KICT	-6	402.6	06/10/2005	37.64754	-97.43	WichitaEisenhowerKS
KJAN	-6	100.6	22/05/2007	32.31986	-90.07778	JacksonIntlMS
KLAS	-8	664.5	25/04/2007	36.0719	-115.16344	LasVegasMcCarranNV
KLAX	-8	29.6	27/10/2006	33.9382	-118.3866	LosAngelesIntlCA
KPNS	-6	34.1	27/03/2007	30.478	-87.18686	PensacolaFL
KRFD	-6	222.5	22/05/2007	42.19325	-89.09335	RockfordIL
KSEA	-8	112.8	17/05/2007	47.44468	-122.31441	SeattleTacomaWA
KSUX	-6	333.8	30/04/2009	42.39171	-96.37949	SiouxCityIA
KTPA	-5	5.8	27/01/2009	27.96334	-82.54001	TampaFL
KWMC	-8	1309.4	17/11/2005	40.90179	-117.80811	WinnemuccaNV

4. Independent gust events

For this study, an “independent gust event” is defined as a one-hour period centred on the maximum gust and separated from other gust events by at least 30 minutes. The consensus from the earlier studies is that thunderstorm downbursts generally last less than 10 minutes, so this dead-time between events complies with the common Wind Engineering rule-of-thumb of 3×timescale for effective statistical independence. This rule is not sufficient for independence of non-convective gusts in synoptic-scale windstorms, which require a longer separation.

Following the approach of Lombardo and Zickar [35] and earlier studies, any observations that might have come from hurricanes were removed over the three-day period centred on the arrival of the hurricane eye into the relevant State. Owing to their rarity, hurricanes are assessed differently from the frequent synoptic and convective wind mechanisms [3]. This was relevant only to the coastal US States along the Gulf and Eastern Seaboard.

The 200 highest gust events at each development station were extracted to give the “Development set” of 4000 gust events – a number that was not too onerous to classify by visual inspection. The minimum

(200th) peak gust, averaged across all stations, was 38kn which is comparable to the 40kn threshold of Chen and Lombardo [27]. The initial method was a recursive search of the whole record from each station for the next highest gust event, which was simple to implement and validate, but very slow to execute. The execution time was shortened by a factor of 60 by first extracting the much smaller set of local maxima and their times of occurrence and searching this. By excluding the period of each gust event from future searches, the recursion cycle became progressively shorter. A “Demonstration set” of all gust events above ≥20kn was extracted by this optimised search, yielding >51 000 gust events. Finally, an “Analysis set” of all gust events above ≥20kn were extracted from each of the 450 stations in Figure 1, comprising $>10^7$ events.

5. Datum classification by visual inspection

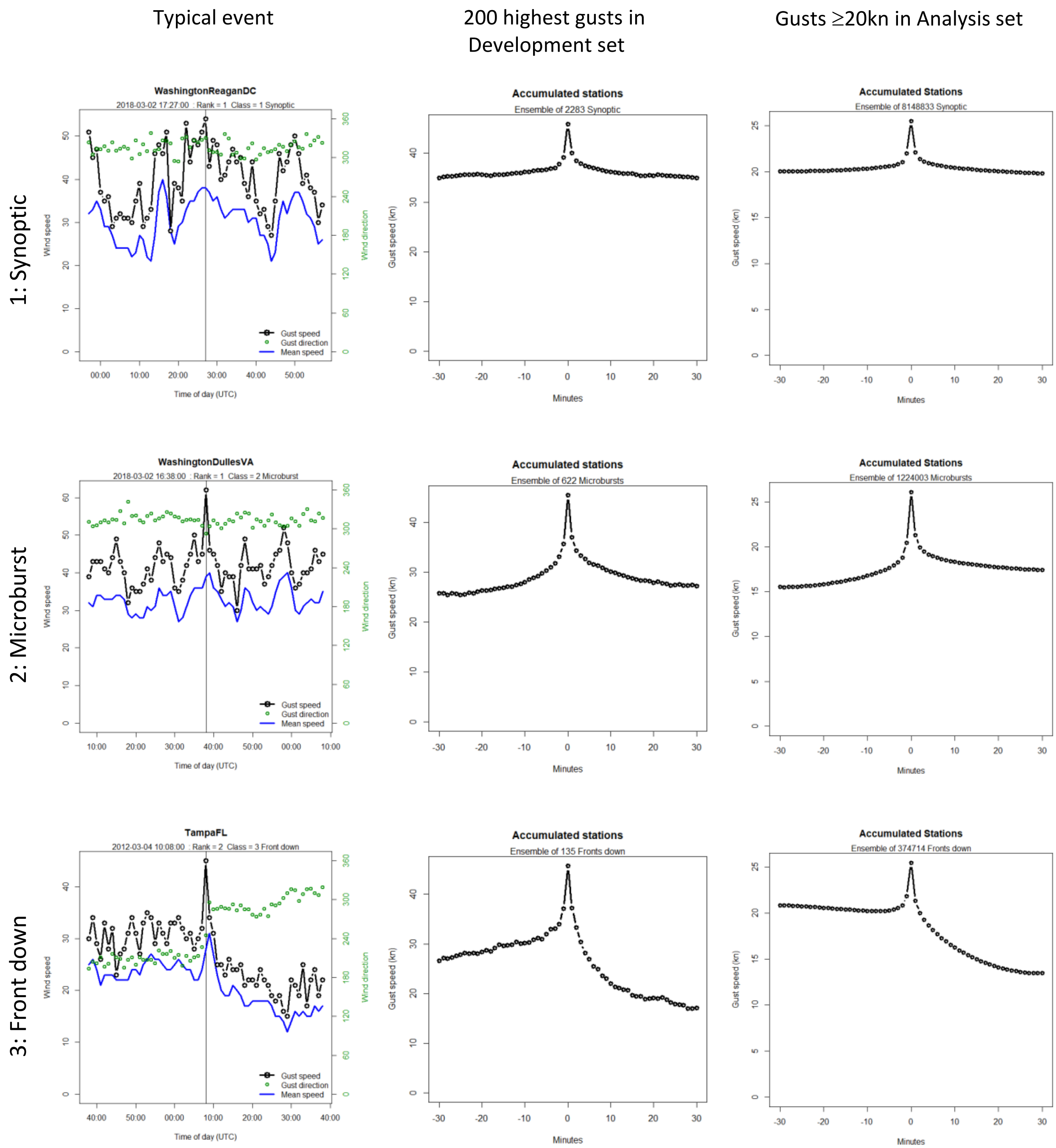
5.1 Gust event Classes

There is consensus in earlier studies, e.g., [10][19], that downbursts from thunderstorms and gust fronts are easily identifiable by visual inspection of the anemograph, and this also applies to charted digital timeseries with a fast-enough acquisition rate. The 2Hz – 10Hz rates of from sonic anemometers, as used in the current European “Wind and Ports” project [9][10][11][12][23], are more than sufficient and require application of a running-mean filter. The current state-of-the-art in windowed filtering is reported by Tubino and Solari [36]. The ASOS 1-minute interval 3s gust data permits resolution of gust events persisting for ~5 minutes, or longer. This includes downbursts from isolated thunderstorms, from downbursts on the forward and rear flanks of gust fronts, and microbursts able to penetrate through the ABL to the surface from deep convection in moderately strong and steady winds. The aim here is to identify classes corresponding to these physical mechanisms.

Visual inspection was therefore expected to distinguish between five Classes of valid gust event, plus a 6th to collect spike artefacts, designated here as:

1) Synoptic – comprising non-convective gusts generated by the ABL in synoptic-scale weather systems and near-neutral atmospheric stability.
2) Microburst – comprising downdrafts from deep convection in steady winds.
3) Front-down – comprising convective downdrafts in the rear flank of gust fronts, where the mean wind speed is decreasing from a higher steady value.
4) Front-up – comprising convective downdrafts in the forward flank of gust fronts, where the mean wind speed is increasing to a higher steady value. Often, but not always, associated with a change in mean wind direction.
5) Thunderstorm – comprising downbursts from isolated thunderstorms in relatively light winds, where the initial wind speed and direction are restored after the event.
6) Spike – comprising isolated instrumentation/transmission spikes, also including some very short-duration surface-generated thermal events like dust-devils. Owing to the ASOS acquisition and QC

protocol [32], a single instrumentation spike may affect two consecutive values if it occurs in the three seconds spanning a change in UTC minute (~1:20 chance). Bird-generated gusts will also be included if the acoustic path of the sonics is blocked for less than 30s [32], which register as single spikes with no loss of data.



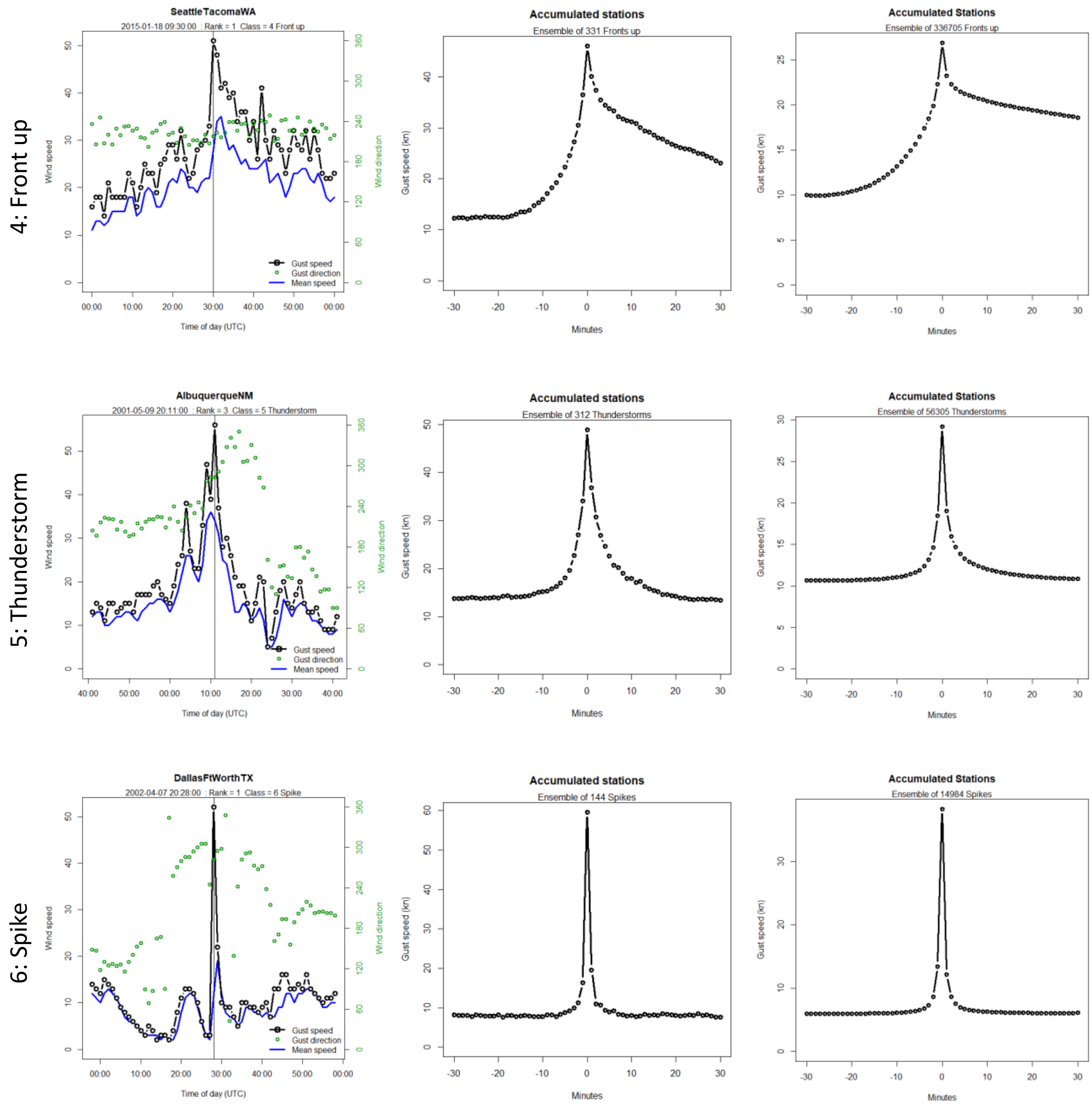


Figure 2. Typical (left) and averaged (centre) visually classified gust events of the development set, and (right) of all gust events >20kn at 450 stations classified by the Highest Kernel Density method.

5.2 Timeseries of classified gust events

The 4000 gust events of the Development set were classified by visual inspection into sub-sets, comprising:

Class 1: 2278 (57%); 2: 622 (16%); 3: 134 (3.4%); 4: 331 (8.3%); 5: 311 (7.8%); 6: 144 (3.6%).

with 180 (4.5%) events unclassifiable due to missing observations.

In Figure 2, the left-hand column shows examples of typical gust events in each Class. Each example in Class 1-5 is the largest valid gust recorded at that station, illustrating that any one of the 5 valid classes may dominate the extremes. The middle column presents the ensemble average of each Class in the

Development set and indicates a consistent trend of increasing sharpness with Class index. The frontal Class 3 & 4 are distinctly skewed in opposing sense, as expected. The right-hand column presents the ensemble average of each Class of all >10^7 gusts in the Analysis set, as identified by the "Highest Kernel Density" (HKD) method described later. Comparison of the two sets of averaged traces confirms that the distinctive ensemble-averaged shape of each Class persists when the population increases from 4000 to >10^7, although the value at the peak is reduced by the inclusion of many lower values – an encouraging result for pattern-recognition methods. Note that the peak gust of Synoptic will always emerge above the steady incident wind speed because the averaged variation in the other values is always less.

The shapes of each averaged Class, normalised to unity peak value, are shown together in Figure 3 for direct comparison and from a clear hierarchy. The principal difference from previous studies is the inclusion of the frontal and downburst classes between the binary T/NT classes. The distinctive characteristics of these new classes are:

- Microburst emerges from a steady wind speed with a sharpness intermediate between Synoptic and Thunderstorm.
- The ramp-up of Front-up is like Thunderstorm and the ramp-down is like Microburst.
- The ramp-up of Front-down is like Microburst and the ramp-down is like Thunderstorm.

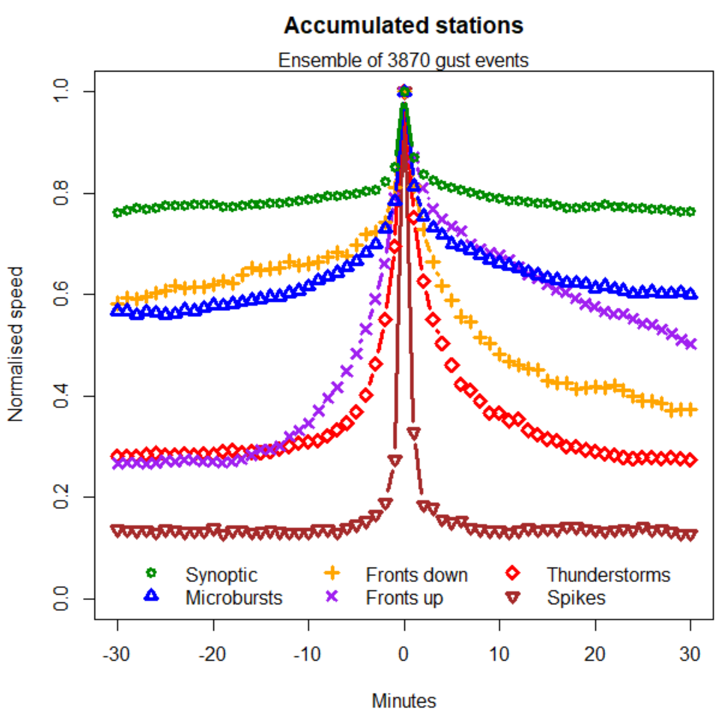


Figure 3. Normalised averaged gust events of the Development set for each visually identified class.

5.3 Statistical metrics of visually classified gust events

5.3.1 Conventional candidates

In addition to the mean, statistical metrics for timeseries include standard deviation, and the higher moments, normally expressed in non-dimensional form as intensity, skewness, and kurtosis. Peak values are expressed by gust factor = peak/mean, or peak factor = (peak – mean)/standard deviation. With only 4000 events in the Development set, unevenly divided between the six classes, it is impractical to evaluate their probability distributions by conventional binning methods, so Kernel Density Estimation (KDE) was used.

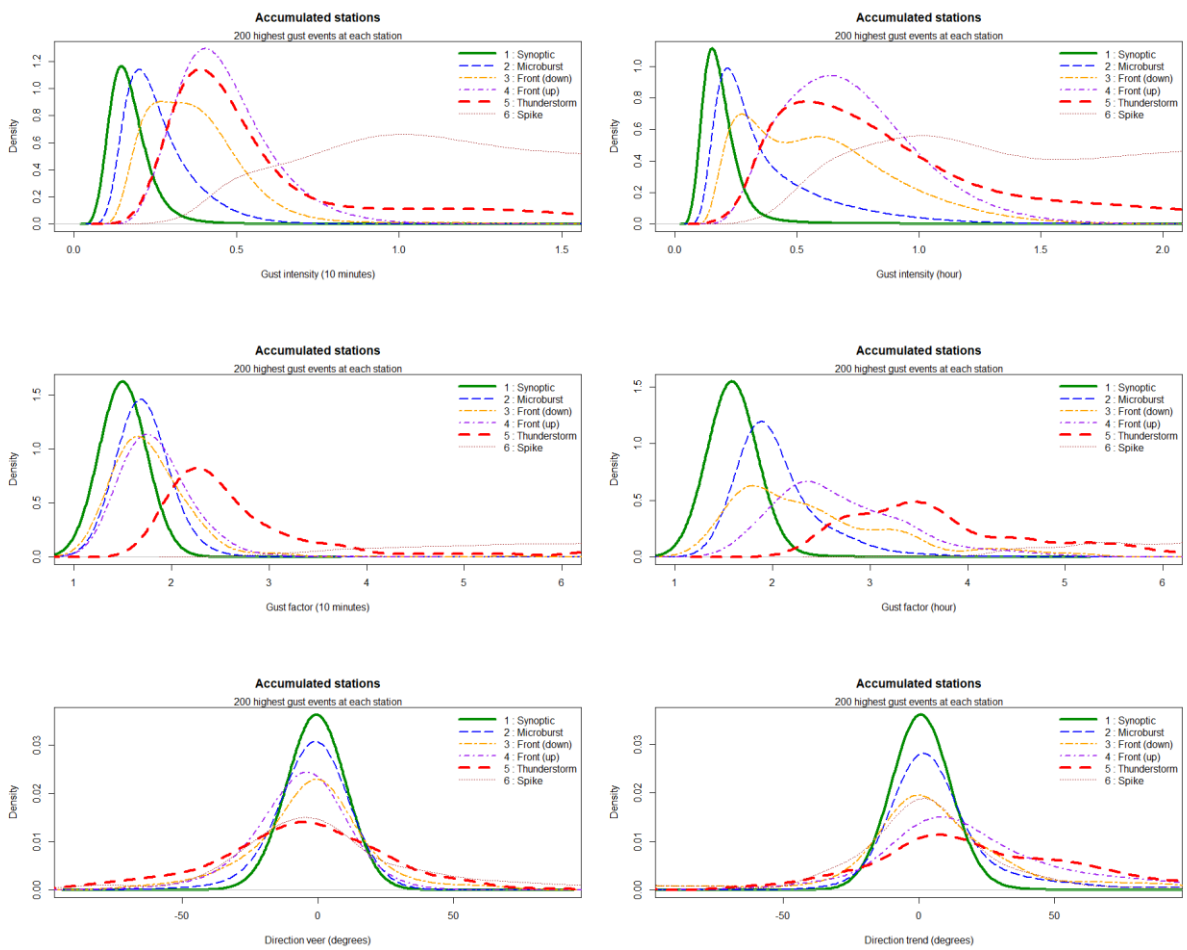
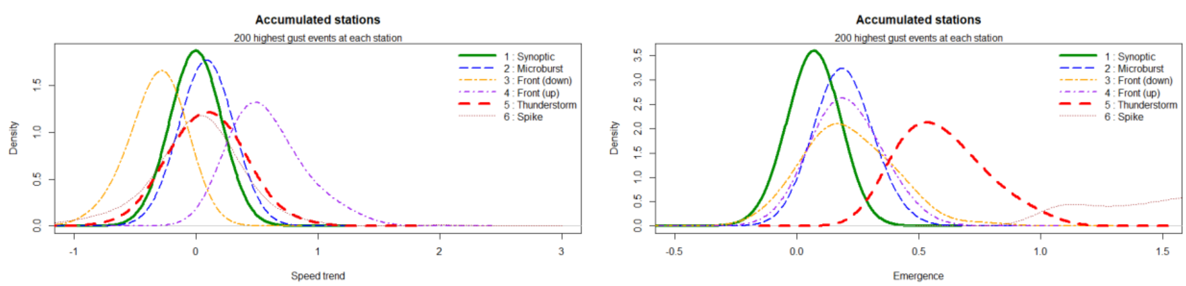


Figure 4. Kernel Density Estimates of conventional statistical metrics.

Figure 4 presents the KDEs of some of the candidate metrics assessed for their ability to discriminate between classes.

- In the top row “gust intensity” (the standard deviation of the 3s gust divided by its mean) is presented for 10-minute and 1h datum means. While there is a left-right trend in the mode with Class, Classes 4 and 5 coincide and there is considerable overlap between distributions.

- Gust factor is presented in the middle row for the same datum means. Now Class 1, 5 and 6 are reasonably separated, but Class 2, 3 and 4 are not. This is the reason [23] can classify T/NT, but not identify the frontal classes.
- The bottom row shows two metrics for the gust wind direction: "veer" – the difference between the peak and the incident mean; and "trend" – the change in 30-minute mean from before to after the event. These KDEs are clustered around the origin and provide no useful discrimination. Gomes and Vickery [2] commented in 1997/8 that the average peak thunderstorm gust direction remains consistent with the approach mean direction, so this negative result was expected. Both Class 4: Front-up and Class 5: Thunderstorm direction trends show a slight bias to veer which is too small to be useful.



(a) 3s gust speed trend. (b) 3s peak gust emergence.

Figure 5. Kernel density estimates of proposed statistical metrics: gust "trend" and "emergence".

5.3.2 Proposed candidates

Other metrics for the 3s gust timeseries were examined to find the best discriminator. The two new metrics proposed as "best" are:

1. Speed trend – This is the change from the mean over the 30 minutes before the peak to the 30 minutes after the peak, divided by the mean for the hour centred on the peak. The KDE, Figure 5(a), shows a good separation between Class 3: Front-down and Class 4: Front-up, with the other classes clustered together in between.
2. Peak gust emergence – This is defined as the peak gust divided by the mean of the 10 next-highest local peaks in the event. Like gust factor, it is a measure of sharpness of the central peak. In gust factor, the datum mean for Front-down and Front-up lies somewhere between the high/low value before the central peak and the low/high value afterwards, so is representative of neither. Gust factor tends to suppress Front-down events and to exaggerate Front-up events. The proposed new emergence metric indicates how far the peak emerges above the envelope of its peers, so treats all the event classes fairly. The KDE, Figure 5(b), shows a good separation between 1: Synoptic, 5: Thunderstorm and 6: Spike, with the frontal classes clustered together.

When used jointly, these two proposed metrics can discriminate between the six classes, as demonstrated by the next section.

6. Comparison of automated classification methods

Following the motivation for this study, this section compares the performance of the more successful current methodologies that use only anemographic data, then develops and assesses improvements. Here, each method is compared with the visual classification by a confusion matrix (CM) [9][27] in which correctly classified events lie on the diagonal and misclassified events lie off-diagonal. The overall error is indicated by the proportion of events that are misclassified.

6.1 Statistical methods

6.1.1 De Gaetano et. al. method

The principal parameters of the De Gaetano et. al. [23] method are the gust factors normalised by the 1-minute, 10-minute and the 60-minute mean. They are compared against datum threshold values sequentially in a logic tree to achieve a binary T/NT choice. The method was applied to the Development set exactly as specified in [23] to produce the CMs of Table 2.

Table 2. Confusion matrices for the De Gaetano method.

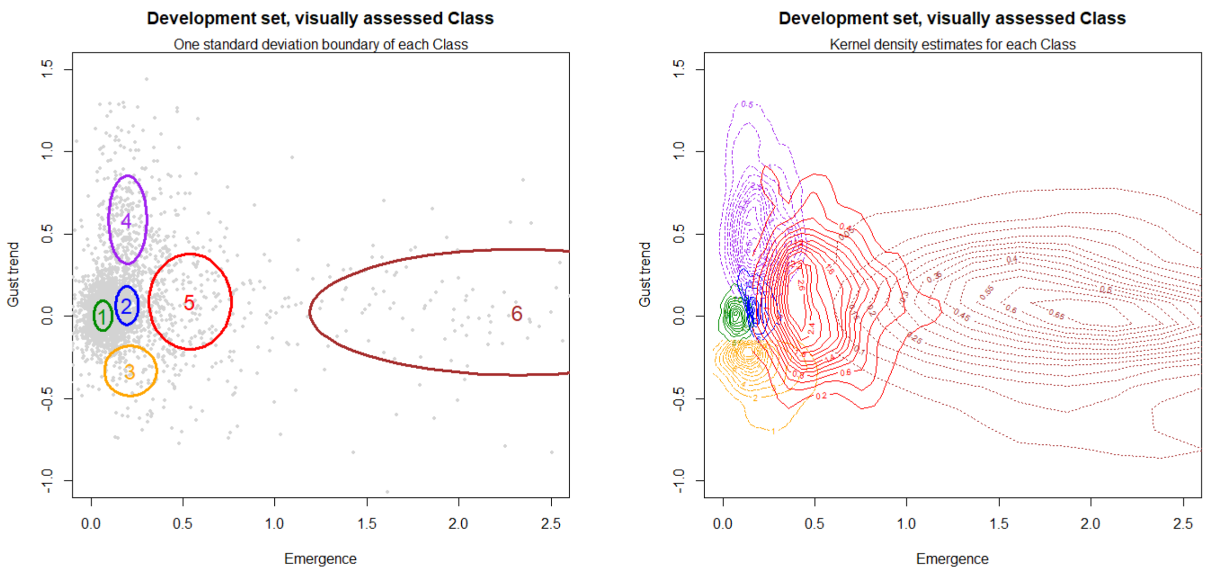
(a)	**De Gaetano**	
Class	NT	T
1 & 2	2774	136
3 - 5	282	494
Error 11.3%		

(b)	**De Gaetano**		
Class	NT	T	Error
1	2264	14	0.6%
2	510	122	19%
3	89	45	66%
4	172	159	52%
5	21	290	6.7%

Table 2(a) represents the binary T/NT classification with Classes 3-5 combined into T and Class 1 and 2 into NT as described in [23] and gives a classification error of 11%. Around 1/3 of Class 3 to 5 that should have been assigned to T are in NT and around ¼ of T are Class 1 and 2. Table 2(b) reveals the misclassifications of the individual classes. The classifications of Class 1 as NT and Class 5 as T are reasonably good, with only ~7% of Thunderstorm missed. As expected, the method fails to discriminate the intermediate classes – but the errors are biased towards NT and not as predicted by [23].

6.1.2 Bivariate k-means method

The classical *k*-means method is a way of clustering data into *k* similar sets and assigning membership of any observed value to the set with the closest mean value. Here, k = 6 and are already sorted by visual inspection, so only the assignment of membership is required. The grey dots in Figure 6(a) indicate the distribution of the Development set across the two-dimensional emergence-trend field. The ellipses represent the one standard deviation boundary around the mean of each Class and are reasonably well separated. Assignment of each event to the closest mean of each Class produces the CM in Table 3(a) and a classification error of 12.4%.



(a) *k*-means cluster ellipses (b) KDE contours

Figure 6. Bivariate *k*-means clusters and kernel density contours of gust "trend" and "emergence" metrics.

Table 3. Confusion matrices for emergence-trend methods

(a)	**Bivariate *k*-means class**					
Class	1	2	3	4	5	6
1	2123	189	7	1	0	0
2	27	563	22	5	6	0
3	14	0	123	0	0	0
4	23	44	0	263	2	0
5	0	38	37	35	200	2
6	0	0	0	2	25	117

Error 12.4%

(b)	**Highest Kernel Density class**					
Class	1	2	3	4	5	6
1	2202	75	31	12	0	0
2	28	565	15	15	0	0
3	2	1	134	0	0	0
4	0	17	0	319	0	0
5	0	40	25	29	209	9
6	0	0	0	0	4	140

Error 7.8%

6.1.3 Highest Kernel Density method

The two-dimensional KDEs, evaluated[1] for each Class, are shown in Figure 6(b) as contours on the emergence-trend plane. For each event the HKD method selects the Class with the highest KDE. The selection was tuned by optimising the kernel bandwidths to give the lowest overall error, 7.8%, in the CM,

[1] By kde2d() in the "MASS" package for R, using axis-aligned bivariate Normal kernels. Based on Venables WN and Ripley BD (2002) *Modern Applied Statistics with S*, Springer, pp510.

Table 3(b). Implementation of the method was simplified by evaluating a look-up table in small increments (0.01) of emergence and trend, shown as a chart in Figure 7. Events in the tail of a KDE will be misclassified and leak into a neighbouring Class when its value falls below that of the neighbour. The method can be viewed as making a sharp cut along the boundary curves of equal KDE value in Figure 7.

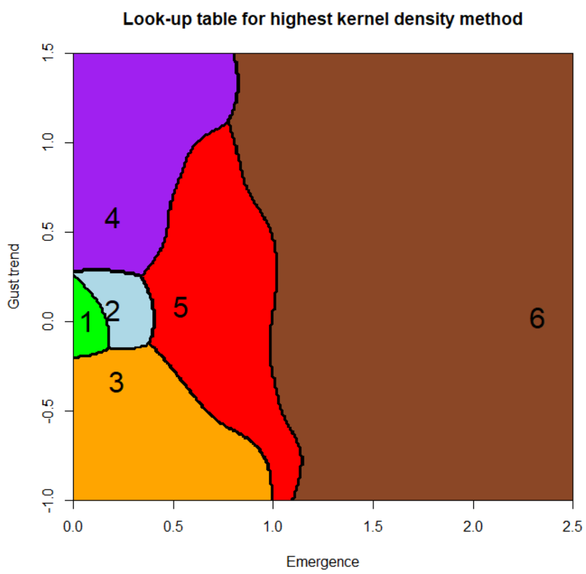


Figure 7. Look-up table for Highest Kernel Density method

As an incidental by-product of the method, the membership probability, F_i, of Class *i* for each event, *e*, may be obtained from the KDE, *p*, by:

Equation 1. $$F_i\{e\} = p_i\{e\} / \sum_{j=1}^{6} p_j\{e\}$$

Membership probability may be used as a weighting factor in any statistic that implements fuzzy logic, e.g., weighted moments, allowing events that fall outside the sharp boundaries to contribute partially to their parent Class.

6.1.4 Neural Network method

The NN in the "neuralnet" package for R was trained to predict the Class from the emergence and trend, using resilient backpropagation with weight backtracking, and disjoint training/testing data. The Development set was split in half by latitude/longitude in two ways: north/south halves, and east/west halves. Training used one half, then testing used the disjoint other half. The critical quality metric was the average prediction accuracy all disjoint combinations of the training/testing sets.

It can be assumed that each set consists of the coherent Class components, which are to be classified, plus an incoherent random "noise" component that differs between sets. Optimisation of the NN configuration started by increasing the number of neurons in a single-layer network, analogous to increasing the number of free parameters in a parametric fit, which eventually leads to overfitting by including the random component. Testing accuracy increases with increasing numbers of neurons up to the point where overfitting begins, after which it starts to fall. The overall error was minimised with 50 neurons at 11.4%. Extending to two- and three-layer networks produced only a marginal reduction in error, but with significantly shorter training times. The optimal NN was found to be three layers of 4 neurons (64 d.f.). Applied to the full development set, this NN produced the CM of Table 4 and an error of 11.1%.

Table 4. Confusion matrix for emergence-trend neural network

	Neural network class					
Class	1	2	3	4	5	6
1	2236	40	6	1	0	0
2	14	538	60	10	0	0
3	1	13	121	0	0	0
4	2	12	18	297	2	0
5	0	17	35	21	231	8
6	0	0	0	0	9	135

Error 11.1%

6.2 Pattern recognition methods

6.2.1 Nearest Normalised Trace method

This is essentially the shapelet method of Arul et al [9], but it uses the normalised averaged gust events in Figure 3 for the whole 1-hour period as the mother shapelets. Training was much faster, however, as the shapelet search was not required, and the comparison was made with the events of the Development set and not the whole record. The timeseries of each event, normalised to unity at the peak, was compared with each of the timeseries in Figure 3 and the Class with the minimum squared Euclidean distance was selected. The CM is in Table 5(a) and the error is 17.9%.

Table 5. Confusion matrices for pattern recognition methods

(a)	**Nearest normalised trace class**					
Class	1	2	3	4	5	6
1	2120	173	17	4	6	0
2	109	395	45	41	326	1
3	25	16	70	0	24	2
4	8	42	0	241	40	1
5	0	6	17	18	218	53
6	0	0	0	2	14	130

Error 17.9%

(b)	**Optimised neural network class**					
Class	1	2	3	4	5	6
1	2193	89	1	0	0	0
2	84	517	16	2	3	0
3	7	30	98	0	0	0
4	0	3	0	322	6	0
5	0	1	6	42	258	0
6	0	0	0	0	8	136

Error 11.9%

6.2.2 Optimised Neural Network method

This method inputs all 61 gust speed values of a gust event directly into the “neuralnet” NN, whittling them down to the single output class through a series of neuron layers. As before, the configuration was optimised using disjoint training/testing sets to minimise the testing error. The optimal network configuration was found to be two layers: the first with 5 neurons and the second with 3 neurons; and produced the CM in Table 5(b) and an error of 11.9%.

6.2 Optimal method

The five methods tested in this study are compared in Table 6 for effectiveness based on minimum overall classification error. The worst performer is the Nearest Normalised Trace method which is the shapelet transform method of Arul et. al. [9] operating with only one mother shapelet for each class. As each individual event trace differs in shape from the ensemble average, Arul et. al. [9] required 32 mother shapelets to make the binary T/TN choice, so this result comes as no surprise.

Table 6. Comparison of method overall error

Input data	Method	Error
Emergence and trend	Bivariate *k*-means method	12.4%
	Highest Kernel Density	**7.8%**
	Optimised neural network	11.1%
Directly from traces	Nearest normalised trace	17.9%
	Optimised neural network	11.9%

The Highest Kernel Density method of §*6.1.3* stands out as the clear winner, and its application is further simplified by means of a look-up table, Figure 7. The HKD method has the additional benefit of providing the class membership probabilities to implement fuzzy logic.

7. Applying the optimal Highest Kernel Density classification method

R scripts were coded to automate the download of the TD6405 observations from any ASOS site, the detection and correction/removal of artefacts, the extraction of all independent gust events above a specified threshold speed, and classification by each of the above methods. These scripts follow four key principles:

1. A central database keeps track of the ASOS stations to be processed and their status.
2. Processing runs fully automatically, in stages, under the control of a master script.
3. Processing may be interrupted and re-started at any time without loss of completed stages.
4. Multiple instances of R can be run on the same PC, or on multiple PCs with network access to the central database.

Principle 3 will allow incremental updating as future observations are added to the TD6405 database.

The scripts were validated by extracting and classifying the Demonstration set of gust events and comparing the top 200 events for each station with the earlier Development set. The scripts were then run to extract and classify the Analysis set for the 450 ASOS stations indicated in Figure 1. The ensemble averaged traces of these were presented earlier in Figure 2.

Brabson and Palutikof [37] demonstrate that wind observations may be represented as a mixture of disjoint Poisson processes when $P\{t\}$, the inter-arrival time t between events of each process, is exponentially distributed. The standard test for this is to plot $-\ln(1-P\{t\})$ against t, when a mixed distribution forms straight-line segments of different slopes, each slope indicating the corresponding Poisson rate parameter $\lambda = 1/T$, where T is the timescale, but a single Poisson process forms a single straight line. Independence between successive events in a correlated timeseries may be obtained by declustering [38]. The most common approaches are:

a) Removing all events below a suitable value threshold (peak-over-threshold).
b) Selecting the maximum value event within equal periods of time (epoch maxima).
c) Taking the maximum value between successive up-crossings of the mean (Davenport-Rice) [39][40].
d) Selecting the larger of any pair of events within a suitable period (Simiu-Heckert) [41].

Approach b) does not result in a Poisson point process, because of the fixed time intervals, but the other three approaches converge asymptotically. The issue with a) is that the threshold may need to be so high that insufficient data remain to be analysed. Approach c) retains the maximum population of independent events if the time constant of the running mean is optimised [38], but it is not compatible with the event extraction/classification processes of this and the earlier studies. Approach d), introduced by Simiu and Heckert [41], is completely compatible with this study. A Poisson process is desirable because it forms the basis of extreme-value theory and it assists in the analysis of extremes by ensuring compatibility with the Harris [43] XIMIS plotting positions and fitting weights.

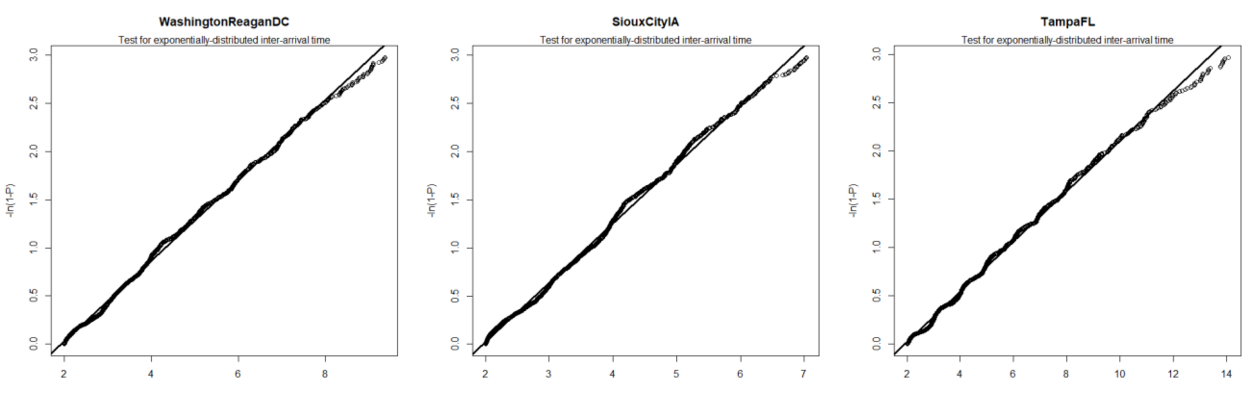
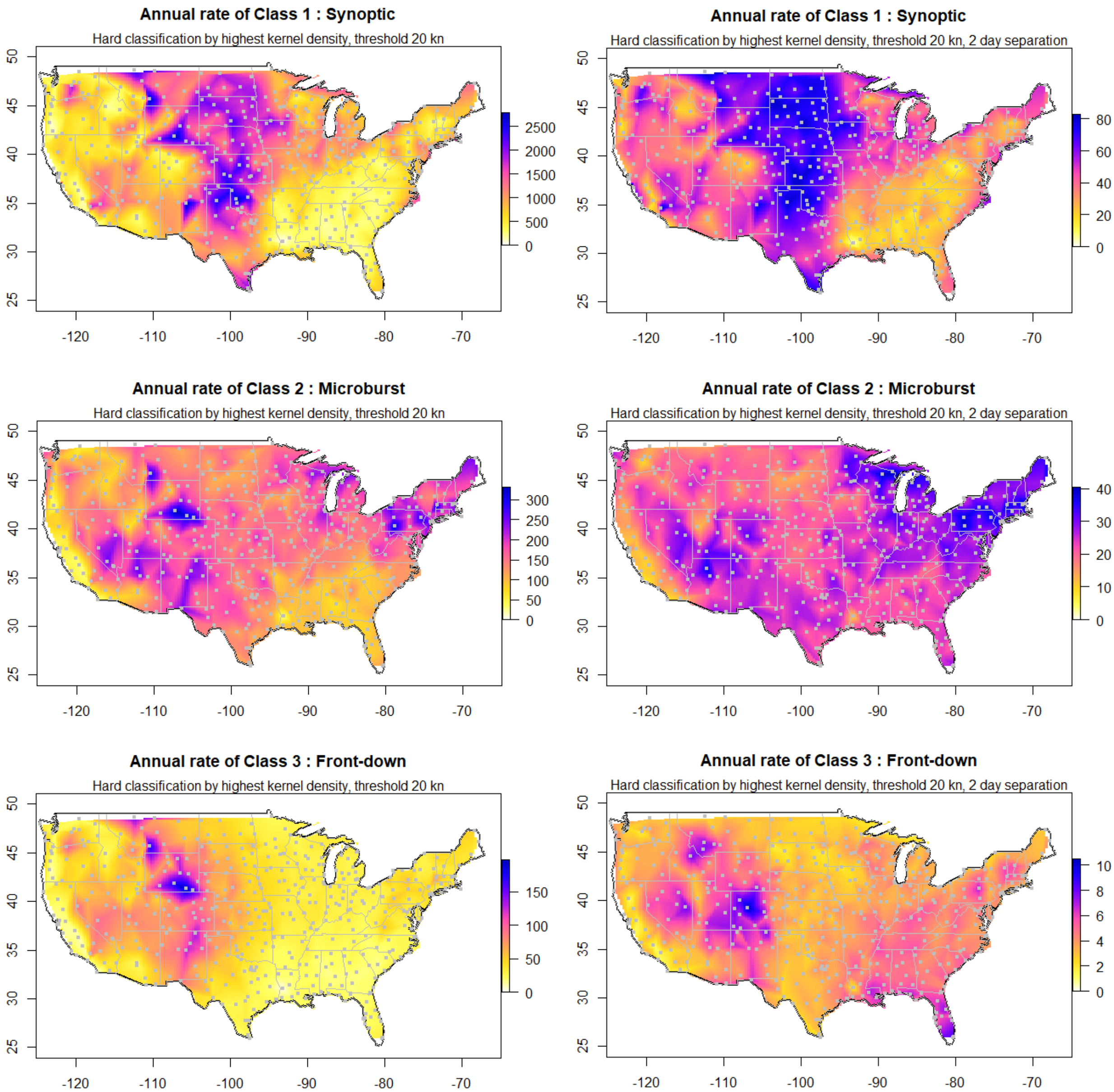


Figure 8. Example tests for Poisson process at m = 2 days minimum separation.

Adopting d), the gust events were subjected to a further screening process that enforced a minimum m-day separation period. The same top-down search procedure as in §4, was applied with a minimum separation of m = 2 days, and this reduced the total population of events by a factor ~15. The value m = 2 days was expected to enclose the longest Synoptic timescale and all the shorter timescales of the other classes, but not the very long timescales associated with substantial gaps in the observational record. Figure 8 shows typical standard independence test plots for three of the Development stations after removing the data gaps in the top 5% of the distribution, confirming that m = 2 days provides convergence to a Poisson process and statistical independence at Synoptic scales.

30-minute minimum separation **2-day minimum separation**



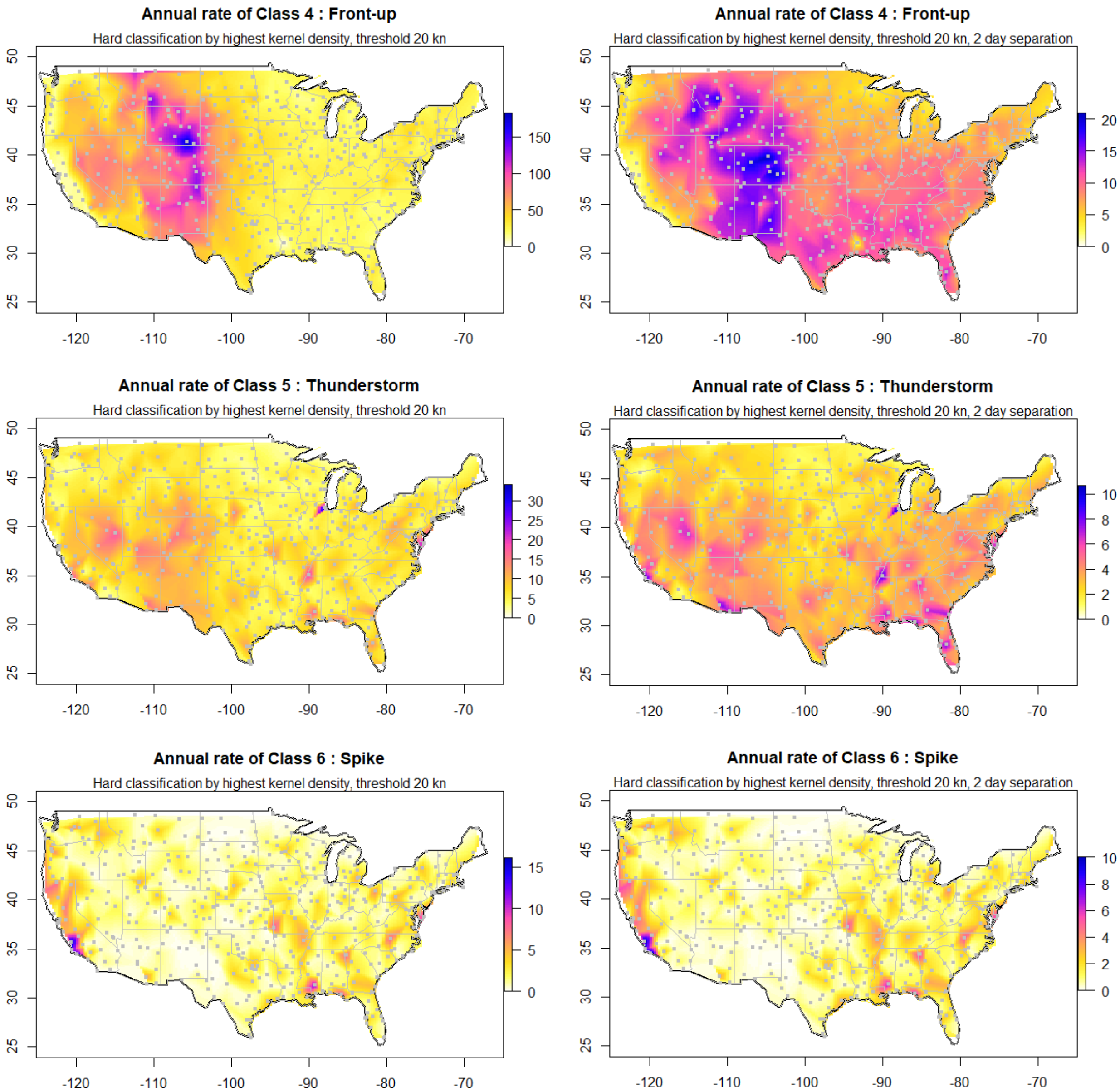


Figure 9. Annual rate of occurrence of gust events ≥20kn in each Class

In Figure 9 the geographical distribution of the annual rate of gust events ≥20kn in each Class is mapped for minimum separations of 30 minutes and 2 days. These maps show that:

- The highest rate of Synoptic gust events occurs in a north-south band through the Great Plains, and the lowest across the Southeast Region. The 2-day separation reduces the Class 1 rate by a factor of 30.
- A concentration of Microburst events in the valley of the North Platte in southern Wyoming corresponds to a westward-protruding arm in the Synoptic distribution. This does not appear in the 2-day events, indicating that these are localised events occurring in short, correlated bursts. The highest rates of 2-day events occur in the NV/UT/AZ tristate area in the Southwest, the Northeast region, and around the Great Lakes.

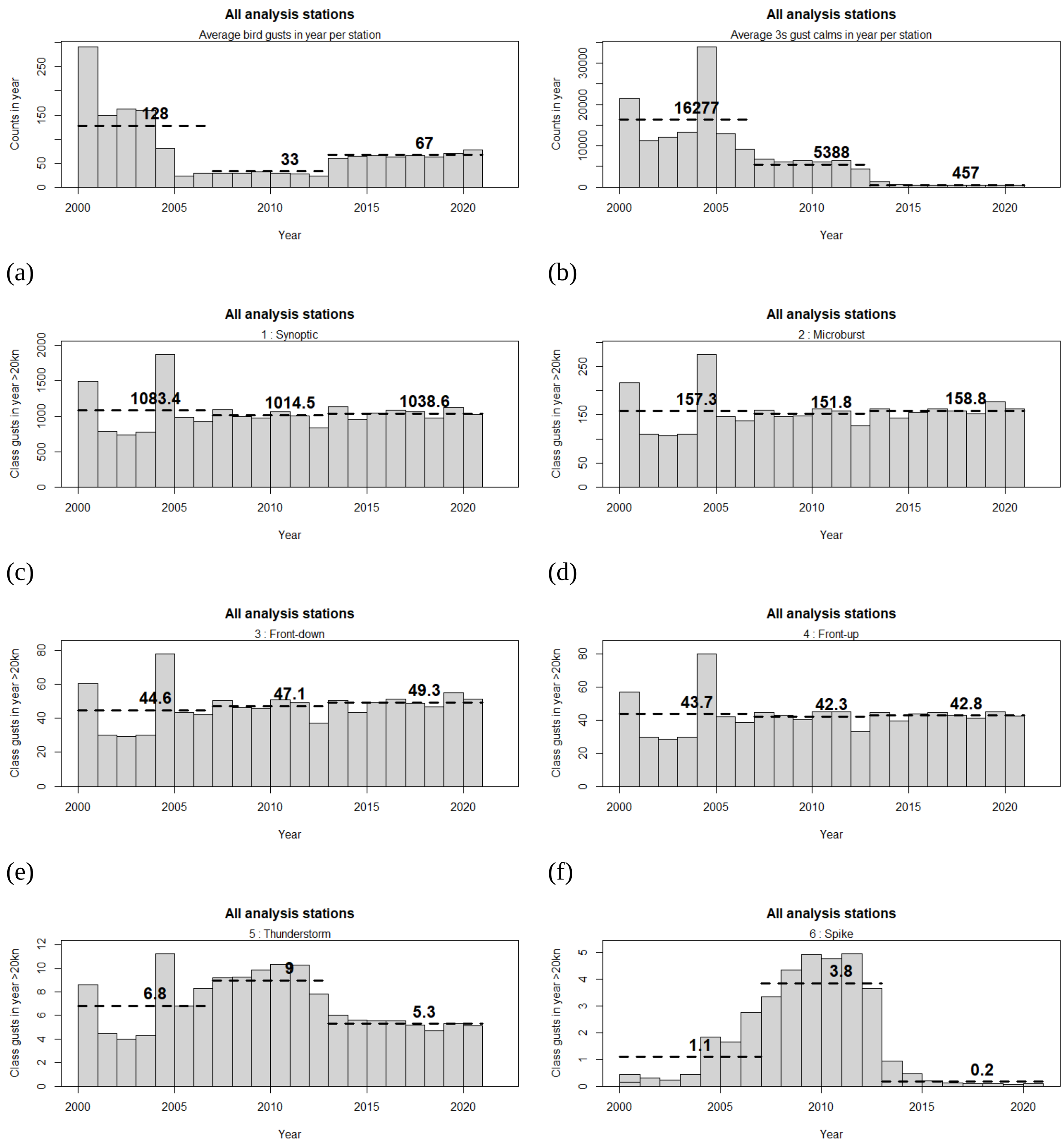
- The two frontal classes show similar distributions, with the maximum frequency in the north-south band in the High Plains along the eastern flank of the Rockies. The high concentration in southern Wyoming moves southwards into northern Colorado when the 2-day separation is imposed.
- If the single concentration of 814 Thunderstorm events in 22 years at KORD Chicago O'Hare, IL, were valid, Chicago would be the thunderstorm capital of the USA, but nearby KMDW Chicago Midway indicates only 102 Thunderstorm events. Visual inspection of the timeseries shows that the majority at KORD are probably low-valued instrumentation spikes that have avoided the Spike classification. Discounting KORD, the Thunderstorm rate is highest in an east-west band across Colorado, Utah, and Nevada, with isolated pockets across the Great Plains and the Southeast Region. Owing to their relative rarity, imposition of the 2-day separation reduces the Class 5 rate by only half.
- Spike is so rare that there is very little difference in the rate upon imposition of the 2-day separation. The single concentration centred on KPRB, Paso Robles, CA, shows that this ASOS station joins KORD in being particularly prone to instrumentation glitches.

Detailed analysis of these data will be made in a separate study that builds on the foundations laid by Lombardo and Zickar [35]. The opportunity is taken here, in Figure 10, to examine the effects of the change in anemometer from cup to sonic and the introduction of QC Test 10. Bear in mind, from §3 and [33], that Test 10 creates four times as many FP than TP and imposes a 5-minute gap in data after each, and that these FP are biased towards calms and thunderstorms in light winds, i.e., to Class 5. The observational period in Figure 10 is sub-divided into three periods: from 2000-2006 for the cup anemometers[2]; 2007-2013 for the sonic anemometers without QC Test 10; and 2014-onwards for sonic with Test 10.

Figure 10(a) shows the total count of bird gusts in each year found by the new algorithm in [33] which registers a bird-generated gust only when a spike in value immediately precedes or succeeds a loss of data. This also includes spikes immediately preceding an instrumentation failure, mimicking the characteristics of bird gusts, and accounts for all counts in the initial cup period. The counts start to drop before the average installation date of the sonics, suggesting that the most unreliable stations were prioritised. All bird gusts after the start of 2014 should have been eliminated by Test10, but the count doubles. These are all FP created by Test 10 which the new algorithm [33] interprets as additional bird-perching events when preceded/succeeded by a spike, particularly after culling valid Thunderstorm events.

Figure 10(b) shows there were more gust calms recorded by the cup anemometers than by the sonics, presumably due to friction in the bearings, but that virtually no gust calms are recorded after introduction of Test 10. Values in years 2000 and 2004 appear to be outliers.

[2] After a slow start, the cup-sonic change was phased in quickly at all ASOS stations. 14 February 2017 represents the mean installation date. Not all stations contribute to earlier years and the results are weighted accordingly.



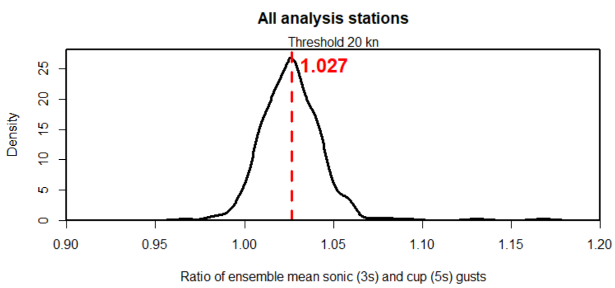
(a) (b) (c) (d) (e) (f) (g) (h)

Figure 10. Effect of instrumentation and quality control changes through the observation period.

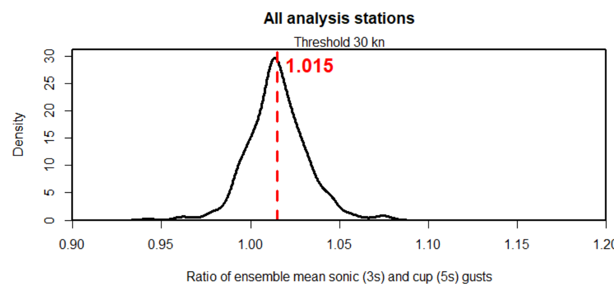
The ensemble-averaged annual rates for Classes 1 to 4 in Figure 10(c)-(f) all show very little variation between the three sub-periods, and between years, except that 2000 and 2004 again appear to be outliers. These gusts all occur in moderate to strong winds, so are insensitive to Test 10 which is triggered only in lighter winds [33].

Figure 10(g) shows the annual rate of Class 5 Thunderstorm rises slightly after installation of the sonics, but almost halves when QC Test 10 is imposed. Curation of the data [33] revealed a bias towards culling the strongest Class 5 events, due to their rapid ramp-up in speed. This limits the usefulness of the TD6405 data for assessing extreme thunderstorm events from 2014 onwards.

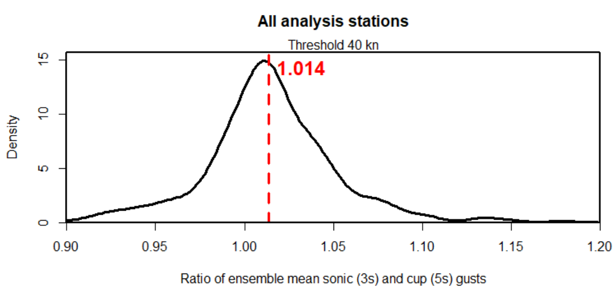
Figure 10(h) shows that Class 5 Spike is a rare occurrence. On average there was only one Spike per year through the cup period, because spikes typically occurring immediately before an instrument failure were culled by the new bird-gust procedure in (a). Spike rate quadruples after installation of the sonics, but it is impossible to be sure of the cause: instrumentation/transmission/bird/etc. Introduction of Test 10 removes nearly all the Spike events, and this is its only perceived virtue.



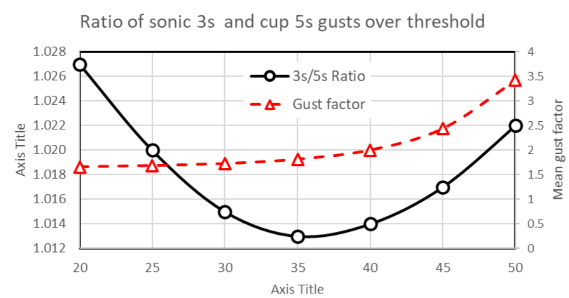
(a) Probability density, ≥20kn



(b) Probability density, ≥30kn



(c) Probability density, ≥40kn



(d) Mean above threshold

Figure 11. Ratio of the mean sonic and cup observations for peak gusts over thresholds, ensemble averaged for all analysis stations.

The Belfort cup anemometers were quoted as having an 5s response time, but cup anemometers have a fixed run-of-wind response length, so that the response time reduces with increasing mean wind speed [42]. The superior response of the Vaisala sonic anemometers was restricted by a running 3s-mean to give the standard WMO response, irrespective of wind speed. The sonic/cup gust ratio represents the factor that could be applied to correct the cup gust speeds to the 3s standard. This was estimated by ensemble averaging all peak gust speeds in the Analysis set exceeding various thresholds over the cup and the sonic periods and assuming that the wind climate remained constant. Figure 11(a)-(c) show the distribution of the 3s/5s ratio for three thresholds. Figure 11(d) shows that the mean ratio varies in a consistent manner with threshold: initially falling as expected, but reaching a minimum at a threshold of 35kn, before rising again at higher thresholds. Perturbation theory [42] predicts that the response to a gust depends not just on the current mean wind speed but also on the change in speed, i.e., also on the gust factor. Figure 11(d) shows that the mean gust factor rises rapidly for thresholds above 35kn, and this is the reason that the 3s/5s ratio rises again. The parabolic trend of the 3s/5s ratio is well fitted in the range 20kn to 50kn by the expression:

$$3\mathrm{s}/5\mathrm{s}\,\text{ratio} = 5\times10^{-5}G_*^2 - 0.0037G_* + 1.0806$$

where G_* is the gust factor based on the hourly-mean of the 1-minute maximum gust, as provided by the TD6405 database. Lombardo et.al [22] report using the factor 1.02 for extremes at the design risk, a value which lies in the middle of this observed range.

8. Conclusions

In response to the call by De Gaetano et al [23] for an automated method that definitively discriminates gust front downbursts and deep-convection downdrafts, this study examined whether this was possible using only anemometric data. A datum set of 4000 gust events from 20 stations were selected for training and assessing the various approaches to classification. Six characteristic classes proved easy to identify by visual inspection, as demonstrated by earlier studies, e.g., [10][19]. The distinct visual disparity in characteristic shape of the six Classes, and the subsequent persistence of these shapes after ensemble-averaging a huge population (>10^7) from 450 ASOS stations, support the view that they successfully discriminate between disjoint physical mechanisms. The disparity in geographic distribution of their annual rates adds further support to this view.

The ensemble-averaged shapes of each Class form a stepped hierarchy, but shapes of individual events fall between these steps. Similarly, there is overlap in the distributions of the statistical metrics of each Class. It follows that no zero-error classification method can ever exist. The study shows the typical error rate of current methods to be ~10%, with the errors lying close to the diagonal of their confusion matrices.

Manual classification by visual inspection relies on the remarkable pattern-recognition of the human brain. Replicating this in software is particularly difficult, but major advances have been made in recent years using artificial neural networks. These methods performed poorly in this study due to the presence of secondary transient events in the 30 minutes before and after the principal gust which introduce distortions from the characteristic datum patterns. The human brain has no difficulty in coping with these distortions.

The best performing “Highest Kernel Density” method, with an error rate of 7.8%, evolved directly from the statistical method of De Gaetano et. al. [23] based on gust factor, but the proposed new “emergence” and “trend” gust speed metrics proved to be more effective discriminators in classifying 1-minute interval observations. This enabled the complex logic tree of [23] to be replaced by a simple bivariate choice. The potential usefulness of the additional fuzzy membership probabilities that this method provides is in weighting the contributions of events that lie close to the Class boundaries, and their effectiveness will be assessed in a future study. The directional metrics provided no practical advantage and so the method relies solely on the gust speed timeseries. The HKD method also has the advantage of simplicity and transparency over the poorer performing and more complex pattern-recognition methods that were tested. This represents a practical example of the fine balance between Occam’s Razor: “entities should not be multiplied beyond necessity" (William of Occam, circa 1287–1347) and Menger's Law: "entities must not be reduced to the point of inadequacy" (Karl Menger, 1902-1985).

Application of the HKD method to the 450 Analysis stations confirmed the consistent flaws in the TD6405 data found by [33] caused by instrumentation and QC changes during the observational period 2000-2021. The introduction of QC Test 10, which affects all observations after the end of 2013, is of particular concern as it continues to cull valid thunderstorms in light winds. The strongest Class 5 3s gust recorded at KIAD: Washington Dulles, VA, 115kn at 23:02 UTC on 26th August 2007, would have been culled by Test 10 had it occurred after 2013. The effect on extreme values will be investigated. In reviewing the statistical metrics, the years 2000 and 2004 consistently appear as outliers and this warrants further investigation. That the maximum gust at KIAD: Washington Dulles, VA, is Class 5 Thunderstorm, while the maximum gust nearby at KDCA: Washington Reagan, DC, is Class 1: Synoptic, illustrates the "hit-and-miss" nature of Class 5. Further, the example of KORD Chicago O'Hare shows that the automated differentiation of Classes 5 & 6 is fallible.

Culling the classified gust events ≥20kn by selecting the highest local peaks separated by a minimum of m = 2 days produces a Poisson Point Process that meets the standard test of exponentially distributed inter-arrival times [37]. This validates using the Harris [43] XIMIS plotting positions and fitting weights in extreme value analysis of the individual Class events.

NOAA have been informed that the ongoing negative impact of QC Test 10 greatly outweighs any marginal benefits. Its poor performance stems from its real-time operation: it is unable to anticipate a few minutes into the future, so it cannot distinguish the fast ramp-up caused by a bird (followed by data loss) from that caused by a thunderstorm (followed by ramp-down). The ASOS observations should be archived in their original raw state so that better QC procedures can be applied, retrospectively, as they are developed.

Supplementary information

The R scripts and instructions to extract gust events classified by the three emergence-trend classification methods from any ASOS station in the TD6405 database are available from Mendeley at URL: https://doi.org/10.17632/88jp3swkn6.1. An Rdata file of gust events ≥20kn at the 450 ASOS stations of the Analysis set, classified by the three methods, is also provided (826Mb).

Note to Editor and Reviewers: The above URL is reserved and will not be activated until publication. In the meantime, the supplementary information is available for download at: https://data.mendeley.com/datasets/88jp3swkn6/draft?a=4fdfc188-d390-4765-bf63-655360c54811.

CRediT authorship contribution statement

Nicholas J Cook did everything.

Funding statement

This curiosity-driven research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] IAWE. Announcement of the Alan G. Davenport Wind Loading Chain. J. Wind Eng. Ind. Aero. 99 (2011) 998-999. https://doi.org/10.1016/j.jweia.2011.09.005.

[2] Gomes L, Vickery BJ. Extreme wind speeds in mixed wind climates. J. Wind Eng. Ind. Aero. 2 (1978) 331-344. https://doi.org/10.1016/0167-6105(78)90018-1.

[3] Vickery PJ, Masters FJ, Powell MD, Wadhera D. Hurricane hazard modeling: The past, present, and future. J. Wind Eng. Ind. Aero. 97 (2009), 392-405. https://doi.org/10.1016/j.jweia.2009.05.005.

[4] Lombardo FT, Smith DA, Schroeder DL, Mehta KC. Thunderstorm characteristics of importance to wind engineering. J. Wind Eng. Ind. Aero. 125 (2014) 121-132. https://doi.org/10.1016/j.jweia.2013.12.004.

[5] Kwon DK, Kareem A, Butler K. Gust-front loading effects on wind turbine tower systems. J. Wind Eng. Ind. Aero. 104–106 (2012) 109-115. https://doi.org/10.1016/j.jweia.2012.03.030.

[6] Ahmed MR, El Damatty AA, Dai K, Ibrahim A, Lu W. Parametric study of the quasi-static response of wind turbines in downburst conditions using a numerical model. Eng. Struct. 250 (2022) 113440. https://doi.org/10.1016/j.engstruct.2021.113440.

[7] Lu N Y, Hawbecker P, Basu S, Manuel L. On wind turbine loads during thunderstorm downbursts in contrasting atmospheric stability regimes. Energies, 12 (2019), 2773. https://doi.org/10.3390/en12142773.

[8] Zhang Y, Sarkar PP, Hu H. An experimental investigation on the characteristics of fluid–structure interactions of a wind turbine model sited in microburst-like winds. J. Fluids Struct. 57 (2015) 206-218. https://doi.org/10.1016/j.jfluidstructs.2015.06.016.

[9] Arul M, Kareem A, Burlando M, Solari G. Machine learning based automated identification of thunderstorms from anemometric records using shapelet transform. J. Wind Eng. Ind Aero. 220 (2022) 104856. https://doi.org/10.1016/j.jweia.2021.104856.

[10] Canepa F, Burlando M, Solari G. Vertical profile characteristics of thunderstorm outflows. J. Wind Eng. Ind. Aero. 206 (2020) 104332. https://doi.org/10.1016/j.jweia.2020.104332.

[11] Xhelaj A, Burlando M, Solari G. A general-purpose analytical model for reconstructing the thunderstorm outflows of travelling downbursts immersed in ABL flows, J. Wind Eng. Ind. Aero. 207 (2020) 104373. https://doi.org/10.1016/j.jweia.2020.104373.

[12] Zhang S, Solari G, Burlando M, Yang Q. Directional decomposition and properties of thunderstorm outflows. J. Wind Eng. Ind. Aero. 189 (2019) 71-90. https://doi.org/10.1016/j.jweia.2019.03.014.

[13] Zhang S, Yang Q, Solari G, Li B, Huang G. Characteristics of thunderstorm outflows in Beijing urban area. J. Wind Eng. Ind. Aero. 195 (2019) 104011. https://doi.org/10.1016/j.jweia.2019.104011.

[14] Choi ECC. Field measurement and experimental study of wind speed profile during thunderstorms. J. Wind Eng. Ind. Aero. 92 (2004) 275-290. https://doi.org/10.1016/j.jweia.2003.12.001.

[15] Orwig KD, Schroeder JD. Near-surface wind characteristics of extreme thunderstorm outflows. J. Wind Eng. Ind. Aero. 95 (2007) 565-584. https://doi.org/10.1016/j.jweia.2006.12.002.

[16] Gunter WS, Schroeder JL. High-resolution full-scale measurements of thunderstorm outflow winds. J. Wind Eng. Ind. Aero. 138 (2015) 13-26. https://doi.org/10.1016/j.jweia.2014.12.005.

[17] Shu ZR, Li QS, He YC, Chan PW. Vertical wind profiles for typhoon, monsoon and thunderstorm winds. J. Wind Eng. Ind. Aero. 168 (2017) 190-199. https://doi.org/10.1016/j.jweia.2017.06.004.

[18] Twisdale LA, Vickery PJ. Research on thunderstorm wind design parameters. J. Wind Eng. Ind. Aero. 41 (1992) 545-556. https://doi.org/10.1016/0167-6105(92)90461-I.

[19] Choi ECC. Extreme wind characteristics over Singapore – an area in the equatorial belt. J. Wind Eng. Ind. Aero. 83 (1999) 61-69. https://doi.org/10.1016/S0167-6105(99)00061-6.

[20] Choi ECC, Hidayat FA. Gust factors for thunderstorm and non-thunderstorm winds. J. Wind Eng. Ind. Aero. 90 (2002) 1683-1696. https://doi.org/10.1016/S0167-6105(02)00279-9.

[21] Kasperski M. A new wind zone map of Germany. J. Wind Eng. Ind. Aero. 90 (2002) 1271-1287. https://doi.org/10.1016/S0167-6105(02)00257-X.

[22] Lombardo FT, Main JA, Simiu E. Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis. J. Wind Eng. Ind. Aero. 97 (2009) 120-131. https://doi.org/10.1016/j.jweia.2009.03.001.

[23] De Gaetano P, Repetto MP, Repetto T, Solari G. Separation and classification of extreme wind events from anemometric records. J. Wind Eng. Ind. Aero. 126 (2014), 132-143. https://doi.org/10.1016/j.jweia.2014.01.006.

[24] Huang G, Jiang Y, Peng L, Solari L, Liao H, Li M. Characteristics of intense winds in mountain area based on field measurement: Focusing on thunderstorm winds. J. Wind Eng. Ind. Aero. 190 (2019) 166-182. https://doi.org/10.1016/j.jweia.2019.04.020.

[25] Guerova G, Dimitrova T, Georgiev S. Thunderstorm Classification Functions Based on Instability Indices and GNSS IWV for the Sofia Plain. Remote Sensing 11 (2019) 2988. http://dx.doi.org/10.3390/rs11242988.

[26] Samanta S, Tyagi B, Vissa NK, Sahu RK. A new thermodynamic index for thunderstorm detection based on cloud base height and equivalent potential temperature. J. Atmos. Sol.-Terr. Phys. 207 (2020) 105367. https://doi.org/10.1016/j.jastp.2020.105367.

[27] Chen G, Lombardo FT. An automated classification method of thunderstorm and non-thunderstorm wind data based on a convolutional neural network. J. Wind Eng. Ind. Aerodyn. 207 (2020) 104407. https://doi.org/10.1016/j.jweia.2020.104407.

[28] Ye L, Keogh E. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. Data Min Knowl Disc 22 (2011), 149–182. https://doi.org/10.1007/s10618-010-0179-5.

[29] NCDC. Data documentation for Data Set 6405: ASOS surface 1-minute, page 1 data. National Climatic Data Center, Asheville NC, USA, July 2006, pp 5. Available at: https://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td6405.pdf.

[30] N J Cook. Review of errors in archived wind data. Weather, 69 (2014), 3, 72-81. https://doi.org/10.1002/wea.2148.

[31] N J Cook. Detecting artefacts in analyses of extreme wind speeds. Wind and Structures, 19 (2014), 271-294. https://doi.org/10.12989/was.2014.19.3.271.

[32] NOAA. Primer for the ASOS Software Version 3.10 Ice Free Wind Sensor quality control algorithm. NOAA, July 24, 2013. Last accessed 3 February 2022: https://www.weather.gov/media/asos/ASOS%20Implementation/IFWS%20QC%20Algorithm_primer.pdf.

[33] Cook NJ. Curating the TD6405 database of 1-minute interval wind observations across the USA for use in Wind Engineering studies. *Submitted to JWEIA : INDAER-S-21-01039*.

[34] Cook NJ. Locating the anemometers of the US ASOS network and classifying their local shelter. Weather, (2021) wea.4131. https://doi.org/10.1002/wea.4131.

[35] Lombardo FT, Zickar AS. Characteristics of measured extreme thunderstorm near-surface wind gusts in the United States. J. Wind Eng. Ind. Aerodyn. 193 (2019) 103961. https://doi.org/10.1016/j.jweia.2019.103961.

[36] Tubino F, Solari G. Time varying mean extraction for stationary and nonstationary winds. J. Wind Eng. Ind. Aerodyn. 203 (2020) 104187. https://doi.org/10.1016/j.jweia.2020.104187.

[37] Brabson BB, Palutikof JP. Tests of the Generalised Pareto Distribution for predicting extreme wind speeds. J. Appl. Met., 39 (2000), 1627-1640. https://doi.org/10.1175/1520-0450(2000)039%3C1627:TOTGPD%3E2.0.CO;2.

[38] Cook NJ. On extracting independent peak values from correlated time series for assessing extremes. J. Wind Eng. Ind. Aerodyn. 170 (2017), 274-282. https://doi.org/10.1016/j.jweia.2017.09.003.

[39] Rice SO. Mathematical analysis of random noise. Part III. Bell Syst. Tech. J. 19 (1945), 46–156.

[40] Davenport AG. Note on the distribution of the largest value of a random function with application to gust loading. Proc. Inst. Civ. Eng. 28 (1964), 187–196.

[41] Simiu E, Heckert NA. Extreme wind distribution tails: A “Peaks over Threshold” approach, J. Struct. Eng. ASCE, 122 (1996) 539-547. https://doi.org/10.1061/(ASCE)0733-9445(1996)122:5(539).

[42] Kristensen L, Frost Hansen O. Distance constant of the Risø cup anemometer. Forskingcenter Risoe. Riso-R No. 1320(EN), 2002, pp23. https://orbit.dtu.dk/en/publications/distance-constant-of-the-ris%C3%B8-cup-anemometer.

[43] Harris RI. XIMIS, a penultimate extreme value method suitable for all types of wind climate. J. Wind Eng. Ind. Aerodyn. 97 (2009) 271-286. https://doi.org/10.1016/j.jweia.2009.06.011.