Impact of ASOS real-time quality control on convective gust extremes in the USA.

Nicholas J Cook, D.Sc.(Eng), Ph.D., F.R.Eng., C.Eng., F.R.Met.S., F.I.C.E, F.I.Struct.E.

Independent Researcher, Highcliffe-on-Sea, Dorset, UK.

Orcid ID: 0000-0002-6390-6884

Email: wind@njcook.uk

Abstract

Most damage, in terms of number and total cost, to buildings across the contiguous United States is caused by gusts in convective events associated with thunderstorms. Assessment of the risk posed by these events is reliant on the integrity of the source meteorological observations. This study examines the impact on risk due to the loss of the valid gust observations in convective downbursts which are erroneously culled by the real-time quality control algorithm of the US Automated Surface Observation System (ASOS) after 2013. ASOS data before 2014 is used to simulate the effect of this algorithm on the 50-year mean recurrence interval (MRI) gust speeds in convective gust events from isolated thunderstorms and active cold fronts at 450 well exposed stations distributed across the contiguous USA. The peak gust is culled in around 10% of these events, but the impact on 50-year MRI gust speeds is mitigated by the contributions of the unaffected events to the extremes. Nevertheless, significant underestimates occur when values are culled from the upper tail of the distribution of extreme gusts. The full ASOS record, 2000-2021, is used to estimate and map the 50-year MRI for all 450 stations by the XIMIS method. It is concluded that recovery of erroneously culled observations is not possible. Spatial smoothing would spread any underestimate over a wider area and would also reduce legitimately high values. The only practical option to eliminate the risk of underestimation is to ensure that the 50-year mean recurrence interval (MRI) gust speed at any given station is not less than the mean for nearby surrounding stations. This also affects stations with

legitimately lower values than their neighbours, which represents the price that must be paid to eliminate unacceptable risk.

Introduction

Most damage, in terms of number and total cost, to buildings across the contiguous United States (CONUS) is caused by gusts from thunderstorm downbursts (Lombardo et. al. 2014). Assessment of design gust values for codes of practice and standards is reliant on the integrity of the source meteorological observations. The principal purpose of this study was to examine how the US Automated Surface Observation System (ASOS) resolves extreme wind gusts in convective events and to assess the impact on safety due to loss of valid observations culled by the real-time quality control (QC) algorithm which is applied to the observations at the source station before onward transmission and archiving.

The ASOS network provides weather observations at 1-minute intervals at 860 stations distributed across CONUS that are archived in the TD6405 and TD6406 databases at the National Centers for Environmental Information (NCEI). This is a most valuable resource because it permits study of many more mesoscale events over much longer observational periods than is possible with targeted measurement campaigns. Previous studies found the precise location and World Meteorological Organisation (WMO) exposure classification of each anemometer mast (Cook 2021) and curated the observations (Cook 2022) to correct or remove common acquisition, transmission and archiving errors (Cook 2014). All gust events >20kn between 2000 and 2021 at 450 WMO class 1 and 2 stations were sorted into six characteristic Classes (Cook 2023) and assigned a presumed causal mechanism, including downbursts from active fronts and from isolated thunderstorms.

Initially, from 2000, ASOS stations used Belfort cup and vane anemometers with a nominal 5s gust response, most at 10m above ground, but some at 26 feet (7.9m). Between 2005 and 2010 they were gradually replaced by Vaisala sonic anemometers averaged to give the WMO standard 3s gust. Having no moving parts, and heated in the winter to prevent icing, the sonic anemometers provide ideal perches for birds (Avery and Genchi 2004; Schwartz and Kays 2001) which block the acoustic

path, causing data loss, and which generate spurious large “spikes” of gust speed on landing and take-off.

The principal role of ASOS is to transmit accurate current weather information in real time from the station to aircraft. Accordingly, QC procedures are applied continuously before transmission. These real-time data are also the primary source for the routine METAR weather reports and the derived daily, monthly, etc., summaries. The previous studies (Cook 2022 and 2023) highlighted serious concerns about excessive false positive QC alerts by Test 10: the “low-speed non-meteorological event” detection algorithm (NOAA 2013) that leads to loss of valid data in meso-scale convective events. The locations of the 450 ASOS stations with WMO class 1 or 2 exposures that were addressed by these studies are shown in Fig. 1.

The ASOS Real-time Quality Control

The ASOS system reports mean and gust wind speeds in integer values of knots (1 kn = 0.5144 ms^{-1}) and it is essential for understanding how the QC works, as well as to avoid unit bias, to retain these original units.

At the start of 2014 NOAA added a new QC test algorithm – Test 10 – to identify high gust speed events of non-meteorological origin occurring at low mean wind speeds. This was specifically to address the issue of bird-generated gusts but also to find the numerous “spike” artefacts caused by acquisition or transmission glitches. Test 10 works by monitoring the running 2-minute mean speed, $\bar{V}_{2\mathrm{min}}$, and running 3-s mean speed, $\bar{V}_{3s}$, at 5s intervals and raises an alert when:

$$\bar{V}_{2\mathrm{min}} <= 6 \text{ kn} \;\; \& \;\; \hat{V}_{3s} > 6 \text{ kn} \;\; \& \;\; \hat{V}_{3s} > 2.5 \times \bar{V}_{2\mathrm{min}} \qquad (1)$$

after which wind speed and direction observations are suppressed for 5 minutes. NOAA reported (NOAA 2013) a 97% drop in erroneous reports at the cost of 0.033% loss of valid data. In late 2017 the $\hat{V}_{3s}$ threshold in (1) was raised from >6kn to >13kn to reduce the rate of false positives. Although the lost valid data is a very small fraction of the whole, its action in suppressing sudden gusts is strongly biased towards thunderstorm downbursts.

Fig. 2 presents two examples of the ASOS 1-minute interval timeseries through the two strongest recorded thunderstorm downbursts at Jackson International Airport, MS, both occurring while Test 10 applied at the 13kn $\hat{V}_{3s}$ threshold. (Note that the chart sub-headings indicate ranks 19 and 23, respectively, because ranks 1:18 and 20:22 are all non-meteorological artefacts.) The peak gust speed, $\hat{V}_{3s}$, is shown by the (black) linked **o** symbols relative to the left-hand scale and the 2-minute mean direction by the (green) unlinked **o** symbols relative to the right-hand scale. Also shown are the contemporaneous atmospheric pressure (scale ±0.05in Hg) and temperature (scale ±10°F) anomalies and the amount of rain (none/light/medium/heavy).

There is no loss of gust speed data, $\hat{V}_{3s}$, in (a) because $\bar{V}_{2\text{min}} > 6$kn prior to the downburst, but data are culled from (b) because $\bar{V}_{2\text{min}} < 6$kn. Note that the maximum $\hat{V}_{3s}$ of both events are concurrent with the minimum temperature and maximum pressure anomaly, so it is likely that the maximum $\hat{V}_{3s}$ in (b) may be the true maximum for this event. However, in the more general case, where the true maximum was excised by the cull, its value is unknown, and this is the main concern in assessing the impact of Test 10 false positives.

The pioneering insight of Gomes and Vickery (1978) established the need to sort extreme events in mixed climates into classes by causal mechanism for separate analysis. Of the 5 event Classes defined for the USA in (Cook 2023) only Class 4: *Front-up* events, which correspond to active cold fronts, and Class 5: *Thunderstorm* events are potentially sensitive to Test 10 by virtue of low initial $\bar{V}_{2\text{min}}$ values. Fig. 3 presents the ensemble-averaged gust speed, temperature and pressure timeseries of these two Classes. For both Classes there is a rapid fall in temperature and rise in pressure concurrent with the peak gust. The gust speed timeseries differ in that the speed quickly recovers to the initial value for Class 5 but remains high much longer for Class 4. The individual gust events, e.g., Fig. 2, are much more variable than these ensemble-averages suggest. Those recovering slower are clearly Class 4, those recovering faster are clearly Class 5, but there is some uncertainty in the Class of intermediate events. The overall classification accuracy of the neural network in (Cook 2023), which interprets the gust speed, temperature and pressure timeseries, is claimed to be better than 95%.

On average, the initial $\bar{V}_{2\mathrm{min}} > 6$kn for both Classes, so only subsets of these events with low initial speeds will be affected by Test 10: potentially, 0.76% of Class 4 and 0.65% of Class 5 at the 6kn $\bar{V}_{2\mathrm{min}}$ threshold; 0.23% of Class 4 and 0.33% of Class 5 at the 13kn threshold. These percentages are small but are an order of magnitude greater than the 0.033% overall loss of valid data (NOAA 2013), highlighting the bias in Test 10 false positives towards convective gust events.

Observed effect of Test 10 from 2014-2021

Observations from 2014 to late 2017 were subject to Test 10 at the 6kn $\hat{V}_{3\mathrm{s}}$ threshold, and those from 2018 onwards to the 13kn $\hat{V}_{3\mathrm{s}}$ threshold. A Test 10 false positive culls 5 minutes of valid wind speed and direction data, sometimes longer if $\bar{V}_{2\mathrm{min}}$ falls below its trigger threshold again. This should not affect the timeseries of temperature and pressure. The Class 4 and 5 gust events over the whole period 2000-2021 were searched for gaps in speed and direction greater than 4 minutes. Those gaps immediately preceded by $\bar{V}_{2\mathrm{min}} \leq 6$kn and where temperature and pressure remained continuous were counted as a Test 10 false positive, while those >6kn were counted as due to other causes. The average annual rates of gaps per station for the Class 4 and 5 events are plotted in Fig. 4. This shows a sharp rise in the rate of gaps on the introduction of Test 10 which continues rising slowly thereafter, while the rate due to other causes remains reasonably constant. Some may have been caused by Test 10 after $\bar{V}_{2\mathrm{min}}$ fell below the 6kn threshold briefly in the 1-minute overlap between observations.

The annual rates of Test 10 culls of valid Class 4 and 5 events is mapped across CONUS in Fig. 5 for 2014-2017 and for 2018-2021, representing each threshold. The highest rates occur in the high-altitude stations of UT, CO, AZ and NM, with the maximum at KASE Aspen, CO. The annual rates of data gaps from other causes shown in Fig. 6 are an order of magnitude lower and more randomly distributed.

Simulating Test 10 using data before 2014.

The major issue with the Test 10 false positives after 2013 is that the culled observations are unrecoverable, so it is uncertain whether this includes the peak gust and, if it was, how much it was

underestimated by the surviving observations. The gust events which occurred before the introduction of Test 10 are suitable for retrospectively assessing its effect at both 6kn and 13kn thresholds. The only significant difference is that the reported observations are at 1-minute intervals, whereas Test 10 operates at 5s intervals. This has two potential counteracting consequences:

1. The 5s intervals give 12 opportunities in each minute for the running $\bar{V}_{2\mathrm{min}}$ to fall to 6kn or less and trigger Test 10, whereas the observations give only one which should lead to fewer simulated culls.
2. The values of $\hat{V}_{3s}$ in the gust events are the largest observed in the previous 1 minute, not the in just the previous 5s, and should lead to more simulated culls.

The simulated annual rates of Class 4 and 5 false positives at either Test 10 threshold are mapped across CONUS in Fig. 7. At the 6kn $\hat{V}_{3s}$ threshold (left), the highest rates again occur in the high-altitude stations of UT, CO, AZ and NM, with the maximum at KASE Aspen, CO. At the 13kn $\hat{V}_{3s}$ threshold (right), the rates are considerably reduced. While the geographical distributions are the same as observed in Fig. 5, but the trend in rate between thresholds is reversed.

The period culled by Test 10 does not always include the peak gust. The distribution of the simulated annual rate of Class 4 and 5 $\hat{V}_{3s}$ values that are underestimated by culling the peak is shown in Fig. 8 to have a similar geographical pattern as Fig. 5, but at annual rates around 10 times smaller. The simulated rate is lower for the 13kn $\hat{V}_{3s}$ threshold which was the intent of the threshold change. The underestimated values of $\hat{V}_{3s}$ caused by the ~10% of simulated Test 10 false positives that cull the peak are shown in the Q-Q plots of Fig. 9 for both Classes and thresholds. As these are in integer knots, each point may represent multiple events. The solid line shows the linear regression, and the chained line denotes 1:1 correspondence. The scatter is too large and random to allow any reasonable correction. It is therefore necessary to directly assess how these false-positive culls would affect the distribution of extremes.

The XIMIS method of extreme-value analysis

The XIMIS method (Harris, 2009) of extreme-value analysis (EVA), fitted by weighted least-mean-squares (wLMS) and displayed on Gumbel axes, was preferred over other methods for the following reasons:

1. Probability estimated from the order statistics are for independent events following a Poisson recurrence process model leading to a Fisher-Tippett Type 1 (FT1) asymptote.
2. The test for validity of the Poisson model is exponentially distributed inter-arrival times (Brabson and Palutikof 2009) which is simple to implement.
3. The order statistics are ranked downwards from the highest value, so XIMIS is the only method (apart from ACER (Karpa and Naess, 2013)) that works for left-censored data where the population, *N*, is unknown.
4. Gust event data from (Cook 2023) are left-censored at $\hat{V}_{3s} > 20$kn.
5. Asymptotic convergence is not a requirement, i.e., XIMIS is a penultimate method.
6. Confidence limit outliers are easily detected, and their contribution removed.
7. Any extreme model can be fitted using any fitting method, but wLMS was used here with the recommended fitting weights (Harris, 2009) to account for statistical variance.

XIMIS methodology

The variate, V, is normalised by the Gumbel "reduced variate" (Gumbel 1958), y_V:

$$y_V = (V - U)/b \tag{2}$$

where U is the mode and b is the dispersion. The annual probability of exceedance, Φ_m, for the m-th rank from the top is linearised in the FT1 model by the Gumbel "plotting position" (Gumbel 1958), y_m:

$$y_m = -\ln(-\ln(\Phi_m)) \tag{3}$$

so that a plot of y_V as abscissa against y_m as ordinate presents the FT1 asymptote as a straight line of slope 1 through the origin. In practice, V is usually plotted as abscissa, whereupon the slope is b and the intercept is U. Deviation from the asymptote due to non-convergence resolves as a curve which may be fitted by any method but, when necessary, XIMIS seeks to linearise this by the transformation:

$$y_V = (V^w - U^w)/b^w \tag{4}$$

which is exact when the upper tail of the parent distribution of V has Weibull equivalence with shape parameter *w*.

XIMIS uses the mean plotting position, $\bar{y}$, given (Gumbel 1958, Eq.4.2.1(11)) for each m by the recursive formula:

$$\bar{y}_{m+1} = \bar{y}_m - 1/m \qquad \text{with} \qquad \bar{y}_1 = \gamma + \ln(R) \tag{5}$$

where $\gamma = 0.5772\ldots$ (Euler's constant) and R is the observation period in years. Equation (5) is exact for an Exponential parent ($w = 1$) but, when $w \neq 1$, progressively accumulates error at large m which is avoided by left censoring. The corresponding variances, $\sigma_m^2(y)$, are given in a similar manner (Gumbel 1958, Eq.4.2.1(13)) by:

$$\sigma_{m+1}^2(y) = \sigma_m^2(y) - 1/m^2 \qquad \text{with} \qquad \sigma_1^2 = \pi^2/6 \tag{6}$$

The distributions of plotting position around the mean, $\bar{y}_m$, are also the distributions of the *m*-th highest values for repeated samples of the observation period, R. In practice, as there is only one observation period, the distributions of the *m*-th highest values are used to assess the confidence that can be placed in the one sample. The PDF of the *m*-th highest values is given (Gumbel, 1958), Eq.5.3.1(2)) by:

$$p_m(y_m) = \frac{m^m}{(m-1)!} \exp(-my_m - me^{-y_m}) \tag{7}$$

and the corresponding CDF of the *m*-th highest values is given (Gumbel 1958, Eq.5.3.2(3)) by:

197 $$P_m(y_m) = P_1{}^m(y_m) \sum_{v=0}^{m-1} \frac{m^v e^{-vy_m}}{v!}$$

198 (8)

199 Note that $p_m(y)$ and $P_m(y)$ are continuous functions for each *m*-th extreme but are evaluated by (7)
200 and (8) only for the discrete ranks. Confidence limits for an XIMIS model are evaluated by solving
201 (8) for y_m at each desired confidence level P_m and for the median, $\breve{y}_m = y_m(P_m = 0.5)$, then
202 applying their differences, $\delta y_m(P_m) = y_m(P_m) - \breve{y}_m$, either side of the model. Owing to the huge
203 powers of m encountered when solving (8), the practical range using 64-bit floating point arithmetic
204 is limited to $m \leq 100$.

205 ***XIMIS charts***

206 All the XIMIS charts which follow are presented on the conventional Gumbel axes (Gumbel 1958),
207 with maximum gust speed $\hat{V}_{3s}$ as abscissa and $y = -\ln(-\ln(\Phi))$ as ordinate. When all the integer
208 $\hat{V}_{3s}$ values are plotted, the many tied values create a staircase effect with steps wider than the
209 confidence limits. For clarity of presentation, only one point was plotted for each integer $\hat{V}_{3s}$ at the
210 ensemble mean plotting position of each set of ties, $\langle \bar{y}_m \rangle_{\hat{V}}$. All observations above the lowest
211 indicated speed were included in the FT1 model fit shown by the straight lines which was fitted by
212 wLMS using $1/\sigma_m^2$ as the fitting weights(Harris, 2009). Confidence limits of 5% and 95% on $\hat{V}_{3s}$ are
213 shown by the chained curves. Strictly, the confidence is in the value of y_m assigned to each rank
214 (Harris 2009), so indicates the uncertainty in the abscissa. But following the convention of Gumbel
215 (1958), here they are applied through (2) to indicate the uncertainty in the ordinate $\hat{V}_{3s}$ at the mean
216 plotting positions.

Simulating the effect of Test 10 on XIMIS extremes 2000-2013

CONUS superstation

The ensemble of all independent events at a group of stations representative of an area into a "superstation" was used by Peterka (1992) to reduce the sampling error from short records of fastest-mile wind speeds across the US Midwest. Fig. 10 presents the XIMIS charts for Class 4 and 5 events amalgamated from all 450 stations, corresponding to 8360 station-years of observations. The charts show the fit for the original observations and for both Test 10 thresholds. Both charts show a good linear fit to the FT1 asymptote, requiring no correction for non-convergence offered by (4). All observations lie within the confidence limits. If the Class characteristics were uniform across CONUS – which, of course, they are not – the highest observed $\hat{V}_{3s}$ in each Class could be interpreted as having an 8360-year mean recurrence interval (MRI). The validity of the superstation concept depends on two contradictory conditions: 1) that the individual stations are well separated to ensure independence; and 2) that the area they represent is sufficiently small for uniformity.

The principal conclusions from Fig. 10 are that the FT1 asymptote is a good model for these strongly convective gust events and that the impact of Test 10 on the 50-year MRI gust speed $\hat{V}_{50}$ is minimal for the CONUS superstation. The provenance of these conclusions is tracked for the highest observed $\hat{V}_{3s}$, via the corresponding US state superstations, to the individual ASOS stations in the following sections.

The Kansas and Illinois state superstations

The highest observed $\hat{V}_{3s}$ occurred in Kansas for Class 4 and to Illinois for Class 5. Fig. 11 presents the corresponding XIMIS charts for KS (290 station-years) and IL (194 station-years) state superstations. Again, both charts show a good linear fit to the FT1 asymptote with all points within the confidence limits, no discernible influence of Test 10 on $\hat{V}_{50}$. The highest observed $\hat{V}$ now lie above the fitted asymptote and projection onto the fitted model gives an MRI of 667 years for Class 4 in KS and 686 years for Class 5 in IL, a large reduction from the CONUS values.

KGCK Garden City, KS, and KMLI Moline, IL

The highest observed $\hat{V}_{3s} =$ 82kn for Class 4 occurred at KGCK Garden City, KS, and $\hat{V}_{3s} =$ 93kn for Class 5 at KMLI Moline, IL. Fig. 12 presents the XIMIS charts for these two stations. Allowing for the higher statistical variance due to the smaller populations, both charts still show a good linear fit to the FT1 asymptote with all points just within the confidence limits and no discernible influence of Test 10 on $\hat{V}_{50}$. The corresponding MRI of the highest observed $\hat{V}_{3s}$ are now 36 years and 213 years, respectively, with the latter only just within the confidence limits. Fig. 13 presents the 1-minute interval timeseries in the same format as Fig. 2.

- The event at KGCK (left) shows typical the Class 4 characteristics of an active cold front:
  - The peak gust occurs at the forward flank of the front and recovers to a sustained value greater than the incident value.
  - A sudden sustained change in direction is contemporaneous with the peak gust.
  - The temperature falls through the event to a sustained minimum ~15 minutes after the peak gust, whereas in a typical Class 5 event, e.g., Fig. 2, the temperature minimum typically coincides with the peak gust.
  - The pressure rises to a sustained value.
  - A burst of heavy rain starts at the peak gust and lasts for 10-20 minutes.
- The 93kn gust at KMLI occurred during a severe outbreak of thunderstorms and tornados across MO, IL and AR. It is reported by the NOAA Storm Prediction Center (SPC 2006) wind report for the day as a non-tornadic event which destroyed the primary ASOS anemometer but was recorded by the back-up equipment. It differs from the typical Class 5 thunderstorm events in Fig. 2 in some important respects:
  - The $\hat{V}_{3s}$ timeseries shape is typical of a Class 5 thunderstorm event but the duration is untypically short.
  - The transient change in direction is untypically short.
  - There is a transient rise in temperature instead of the expected fall.

- The transient changes in pressure are untypically abrupt.
- The continuous rainfall is interrupted at the time of the peak gust.

This is cited as an example of a "warm" event by Cook (2023) which include local thermally induced events and represent 7.3% of Class 5. The characteristics of this event are sufficiently atypical of the training set for both neural networks of Cook (2023) for the event to have been classified as non-meteorological in origin (TNNDT = 6 and XTNNDT = 6 in the chart subheadings). Reclassification to Class 5 was made by visual inspection of the timeseries and confirmation by SPC (2006).

Influence of Test10 at all stations

Although simulating Test 10 appears to have no discernible influence on majority of stations, about 10% are affected. Fig. 14 shows the distribution of error in the values of $\hat{V}_{50}$ predicted by XIMIS for both Classes and thresholds. These distributions are skewed towards underestimation, with the largest underestimate of $\hat{V}_{50}$ for Class 4 at KBKB Fort Polk Fullerton, LA, and for Class 5 at KTQE Tekamah, NE. Raising the threshold to 13kn reduces the frequency of errors for both Classes, and the range of errors for Class 4.

Underestimation is a concern for the design of safe structures. The XIMIS charts for the largest underestimate of each Class are presented in Fig. 15.

- At KBKB the underestimate of 4.3kn (10.7%) for Class 4 at the 6kn threshold is principally caused by culling the $m = 1$ event, which removes the peak value and demotes the event rank to $m = 22$. The effect is amplified by the small population of Class 4 events at this station and the large statistical variance denoted by the wide confidence limits. Raising the threshold to 13kn restores the culled peak and $\hat{V}_{50}$ is then marginally (1%) overestimated.
- At KTQE the underestimate of 5.2kn (5.7%) for Class 5 at both thresholds is principally caused by culling the $m = 3$ and 11 events. Raising the threshold to 13kn does not restore these missing events.

The distribution of $\hat{V}_{50}$ underestimates for both Classes and thresholds are mapped across CONUS in Fig. 16. The locations of errors appear confined to individual randomly located stations, although the larger errors resolve as patches around each affected station due to the barycentric linear interpolation. There is no discernible correlation between the locations of Class 4 and Class 5 locations. The reduction in errors on raising the threshold from 6kn to 13kn is evident, more so for Class 4 than Class 5.

Results

The values of $\hat{V}_{50}$ predicted by XIMIS for all stations over the observation period 2000-2021 are mapped across CONUS in Fig. 17. These are designated as "raw" in that no corrections have been applied to compensate for the effect of Test 10 false positives. Again, underestimates resolve as patches around each affected station, although these do not correlate well with the locations in Fig. 16. Locations of underestimates that would have occurred in 2000-2013, had QC Test 10 operated then, do not predict where underestimates occur in 2014-2021. Neither do the locations where the observed rate of culls are high, Fig. 5, illustrating the inherent uncertainty of the impact of Test 10. There is a complex balance between militating effects, such as the rate of all culls, and mitigating effects, such as the small proportion that result in underestimates and the lower weighting given to the smaller ranks (higher values) at each station. Stations with the highest values are least influenced by Test 10, implying that the highest peaks start from initial mean values higher that the threshold. The effect on risk of exceedance by culling the highest value at a station is the same as if that value had, by chance, not occurred and therefore it has the smallest weighting in EVA. The larger underestimates occur when there are several culls from the upper tail and the rate of gust events is low.

For the safe design of structures, underestimation must be avoided at all costs. Overestimates are an economic issue, but Fig. 14 shows the effect on $\hat{V}_{50}$ is limited to about 1kn which is acceptable. Reduction of the underestimates by simple geographical smoothing is not an adequate solution because, while the underestimate at the affected station is reduced, it is redistributed over a larger surrounding area. Smoothing also reduces valid peak values. One solution is to clip the value at a

station by the mean of other surrounding stations within a given radius, i.e., by taking the larger of the mean at stations within a doughnut shaped area around the station or the original station value. The results of using a 3° radius are shown in Fig. 18, in which ~40% of stations have been clipped for each Class. The frequency of each level of underestimate correction is plotted in Fig. 19, together with the names of stations where the correction is greater than 10kn. Of these, only four are common to both Classes: Aspen, CO; Fort Polk Fullerton, LA; Harrison Boone, AR; and Hayward, WI. The proportion of stations corrected in this way is much higher than the 10% expected from the simulation of Test 10 because it includes stations where $\hat{V}_{50}$ is legitimately lower than its neighbours. Nevertheless, it indicates that reliance on a single station without verification by its neighbours is unsafe.

The estimates of $\hat{V}_{50}$ presented in Fig. 18 should be regarded as a preliminary "proof of concept", because they lack full consideration of statistical independence, asymptotic convergence and local exposure. These considerations will be addressed in a later study and these preliminary estimates serve to inform how this study should proceed.

Conclusions

As gusts from thunderstorm downbursts are the principal cause of damage, in terms of numbers and overall cost, to structures in the contiguous USA, the accurate assessment of risk posted by these events is reliant on the suitability, availability and integrity of the relevant meteorological observations. The ASOS network provides observations at 1-minute intervals which are sufficient to resolve these mesoscale convective events, but observations after 2013 are compromised by false-positive alerts by the introduction of a real-time quality control algorithm that culls valid data at the source stations before it is transmitted and archived. Although the number of false alerts is very small, they are biased towards these important mesoscale events. Simulation of the algorithm on observations before 2014 suggests that the peak gust is culled in about 10% of these alerts and its value cannot be estimated from the surviving values. The missing data are therefore unrecoverable, but their impact on the assessment of extremes is diluted by the contributions from the unaffected

events. As there is no reasonable prospect of retrospectively correcting the action of the algorithm on past observations, attention must be focussed in mitigating its impact on extreme value analysis, represented here by $\hat{V}_{50}$, and the consequential risk to structures. The QC simulation predicted that the locations of underestimating stations are randomly distributed across CONUS and not predictable from the observed rates of culls. The mitigation proposed here is remove underestimates by ensuring that $\hat{V}_{50}$ predicted at any given station is not less than the average value of other stations within a radius (here, 3°) around the station. This also impacts on stations where $\hat{V}_{50}$ is legitimately lower than the neighbours, which represents the price that must be paid to eliminate unacceptable risk.

Acknowledgements

The work reported in this paper is the result of unfunded, curiosity-driven research undertaken at the author's pleasure.

Notation

The following symbols are used in this paper:

b	Dispersion of Gumbel (FT1) distribution
m	Rank of observation in descending value
MRI	Mean Recurrence Interval (Return period)
p	Probability density (PDF)
P	Cumulative probability (CDF)
U	Mode of Gumbel (FT1) distribution
V	Wind speed
$\bar{V}_{2\mathrm{min}}$	2-minute mean wind speed
$\hat{V}_{3\mathrm{s}}$	Peak 3-second mean wind speed in previous minute

367	$\hat{V}_{50}$	Peak $\hat{V}_{3s}$ with predicted MRI = 50 years
368	w	Weibull shape parameter (in Eqn(4))
369	y_m	Gumbel plotting position (linearised Φ, Eq (3))
370	y_V	Gumbel reduced variate for V (non-dimensional V, Eq (2))
371	Φ	Annual probability of exceedance
372	σ	Standard deviation
373	$\langle\cdot\rangle_x$	Ensemble average of values corresponding to x, e.g., $\langle\bar{y}_m\rangle_{\hat{V}}$

374 Data Availability Statement

375 The meteorological observations used in this study, TD6405 and TD6406 files from 2000 to
376 December 2021, may be downloaded from NCEI by FTP or HTTP. NCEI has lately transitioned to
377 HTTP only, with observations from January 2022 onwards available here, updated monthly. The R
378 scripts and instructions to extract gust events from any ASOS station and classify by the TNNDT
379 method in Cook (2023) are available from Mendeley at URL: https://doi.org/10.17632/88jp3swkn6.1.
380 R scripts to replicate the analyses in this paper are available from the author, on application by email.

381 References

382 Avery, M. L., and A. C. Genchi. 2004. "Avian perching deterrents on ultrasonic at airport windshear
383 alert systems." United States Department of Agriculture, Wildlife Services, September 2004.
384 Available at: https://digitalcommons.unl.edu/icwdm_usdanwrc/74.

385 Brabson, B. B., and J. P. Palutikof. 2009. "Tests of the Generalised Pareto Distribution for predicting
386 extreme wind speeds." *J. Appl. Met.*, **39**: 1627-1640. https://doi.org/10.1175/1520-
387 0450(2000)039%3C1627:TOTGPD%3E2.0.CO;2.

388 Cook, N. J. 2014. "Detecting artefacts in analyses of extreme wind speeds." *Wind and Structures*, **19**:
389 271-294. https://doi.org/10.12989/was.2014.19.3.271.

Cook, N. J. 2021. “Locating the anemometers of the US ASOS network and classifying their local shelter.” *Weather*, wea.4131. https://doi.org/10.1002/wea.4131.

Cook, N. J. 2022. “Curating the TD6405 database of 1-minute interval wind observations across the USA for use in Wind Engineering studies.” *J. Wind Eng. Ind. Aero.* **224**: 104961. https://doi.org/10.1016/j.jweia.2022.104961.

Cook, N. J. 2023. “Automated classification of gust events in the contiguous USA.” *J. Wind Eng. Ind. Aero.*

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

Gumbel, E. J. 1958. *Statistics of Extremes*. Columbia University Press, New York. ISBN 0-231-02190-9.

Harris, R. I. 2009. “XIMIS, a penultimate extreme value method suitable for all types of wind climate.” *J. Wind Eng. Ind. Aero*. **97**: 271-286. https://doi.org/10.1016/j.jweia.2009.06.011.

Karpa, O., and A. Naess. 2013. “Extreme value statistics of wind speed data by the ACER method.” *J. Wind Eng. Ind. Aero.* **112**: 1-10. https://doi.org/10.1016/j.jweia.2015.01.011.

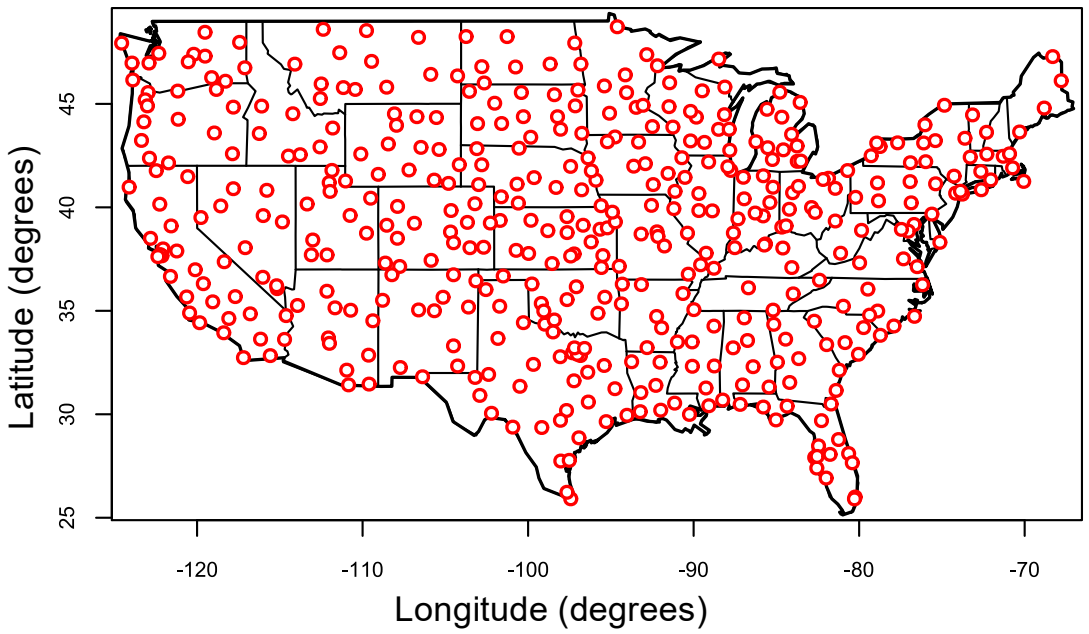
Lombardo, F. T., D. A. Smith, D. L. Schroeder, and K. C. Mehta. 2014. “Thunderstorm characteristics of importance to wind engineering.” *J. Wind Eng. Ind. Aero.* **125**: 121-132. https://doi.org/10.1016/j.jweia.2013.12.004.

NOAA. 2013. “Primer for the ASOS Software Version 3.10 Ice Free Wind Sensor quality control algorithm”. NOAA, July 24, 2013. Last accessed 13 November 2022: https://www.weather.gov/media/asos/ASOS%20Implementation/IFWS%20QC%20Algorithm_primer.pdf.

Peterka, J. A. 1992. “Improved extreme wind prediction for the United States.” *J. Wind Eng. Ind. Aero.* **41**: 533-541. https://doi.org/10.1016/0167-6105(92)90459-N.

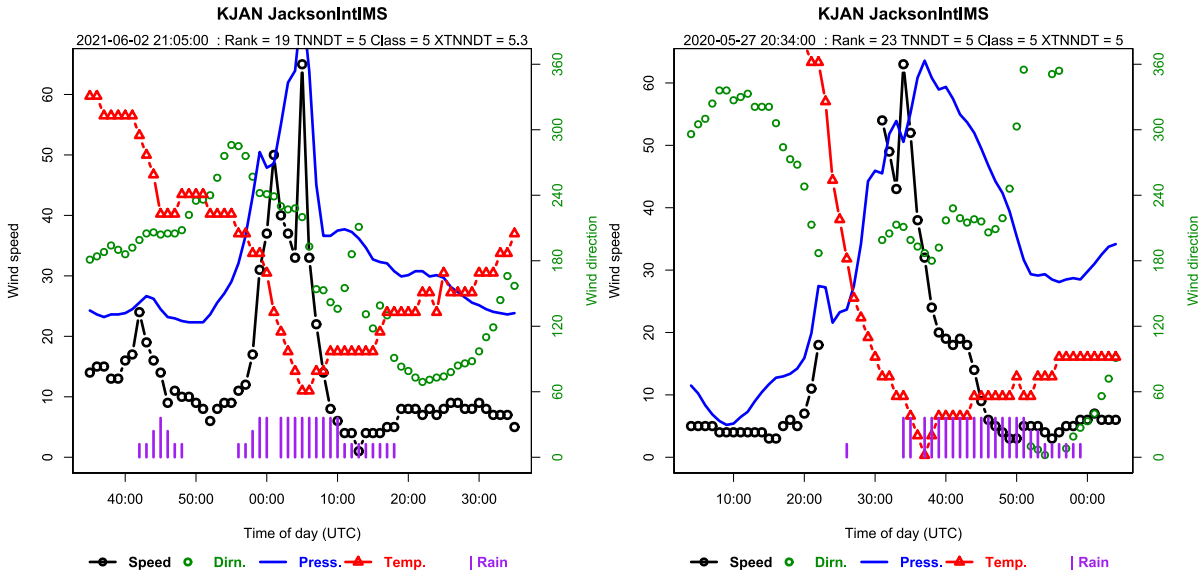
Schwartz, J., and T. Kays. 2001. “Bird Deterrence at Low Level Windshear Alert System (LLWAS) Poles.” 2001 Bird Strike Committee-USA/Canada, Third Joint Annual Meeting, Calgary, AB. Available at: https://digitalcommons.unl.edu/birdstrike2001/20.

417 SPC. 2006. “SPC Storm Reports for 03/12/06.” NOAA Storm Prediction Center. (Accessed 14
418 January 2023.) https://www.spc.noaa.gov/exper/archive/event.php?date=20080615.
419



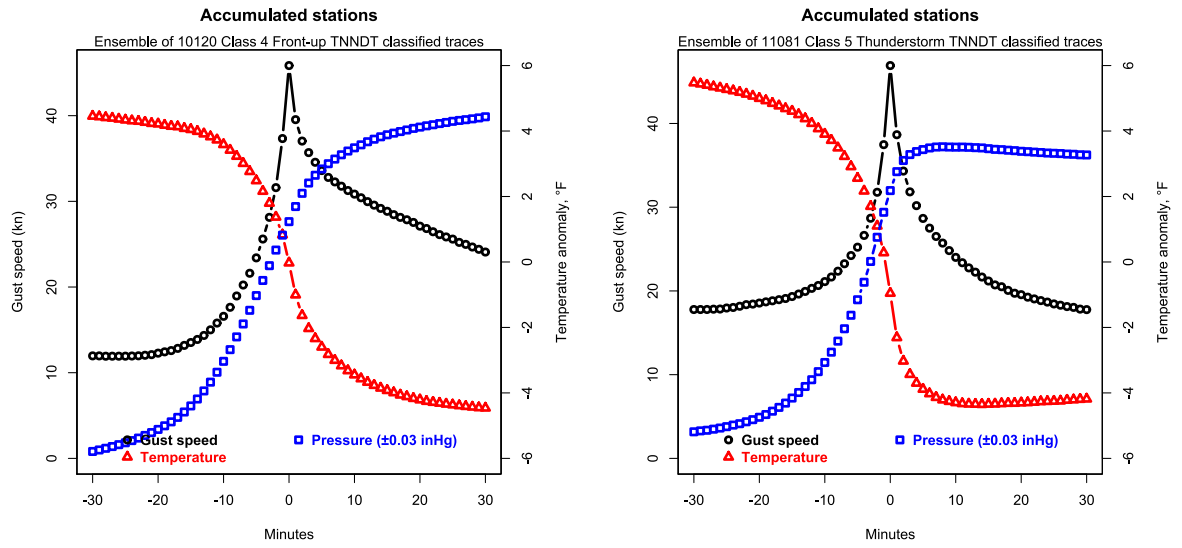
420

421 Fig. 1. Locations of the 450 WMO exposure class 1 & 2 stations.



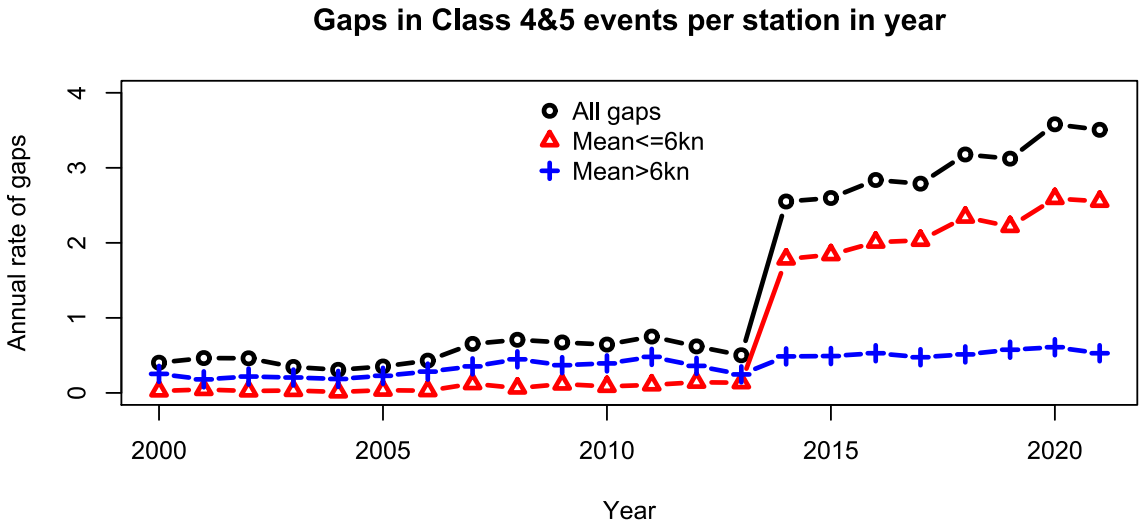
(a) Highest valid *Thunderstorm* gust (b) Second-highest valid *Thunderstorm* gust

422 Fig. 2. Extreme Class 5 *Thunderstorm* gust events at Jackson International AP, MS.



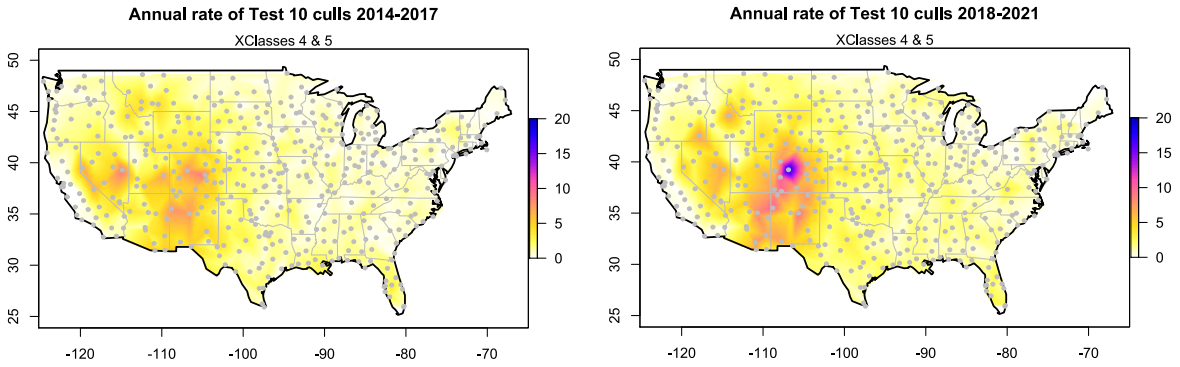
423 Fig. 3. Ensemble-averaged gust speed, temperature and pressure timeseries for all Class 4 and 5 gust

424 events at the 450 stations in Fig. 1.



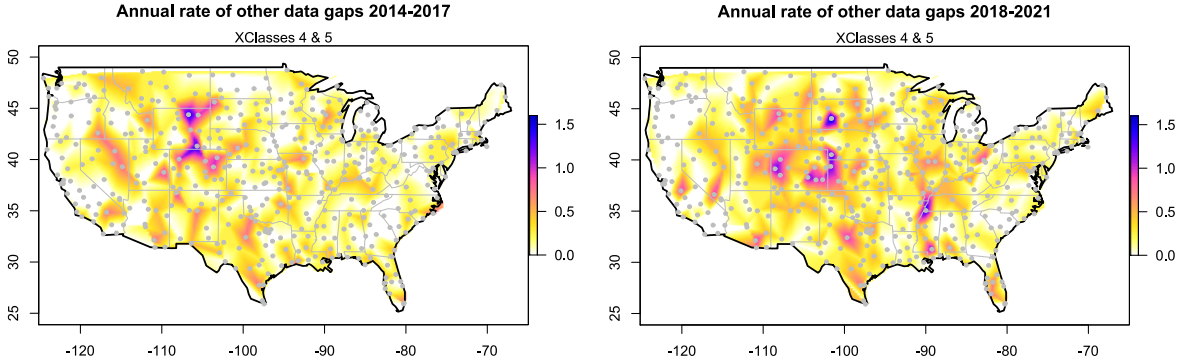
425

426 Fig. 4. Average annual rate of data gaps found in Class 4 and 5 events, 2000-2021.



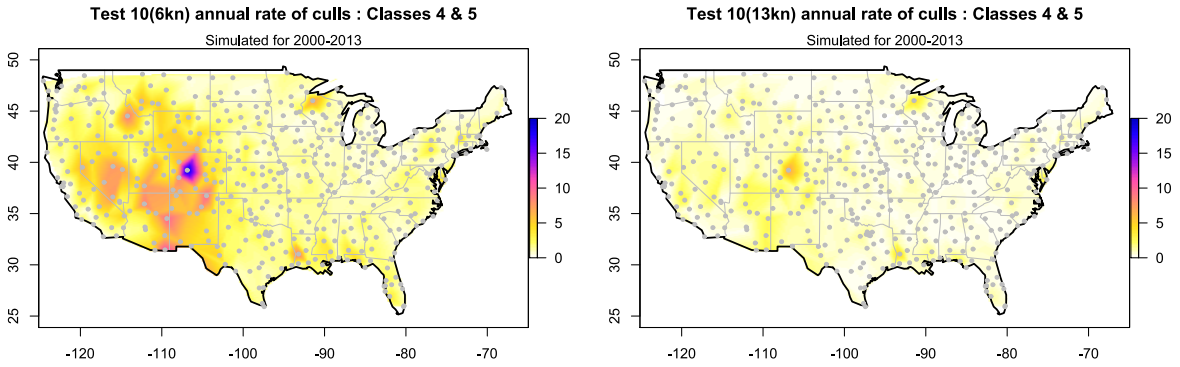
427 Fig. 5. Observed annual rates of Classes 4 & 5 events culled by Test 10.

428

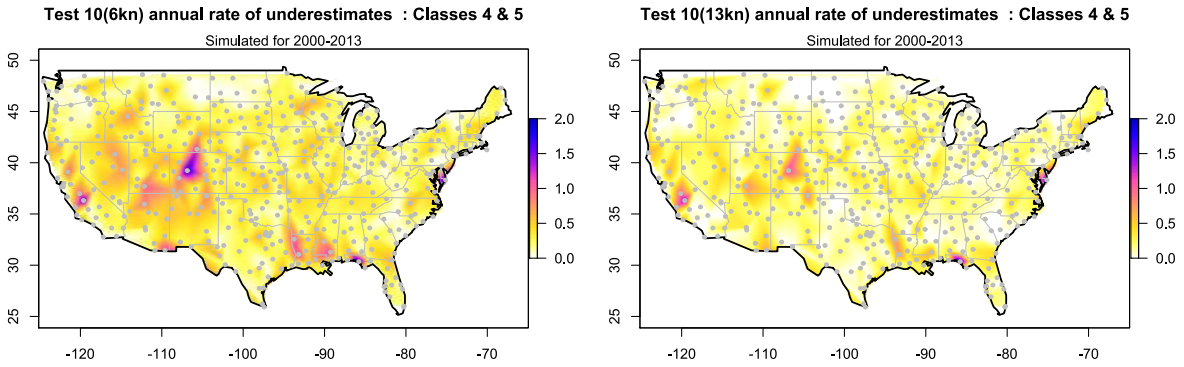


429 Fig. 6. Actual annual rates of gaps from other causes in Class 4 and 5 events.

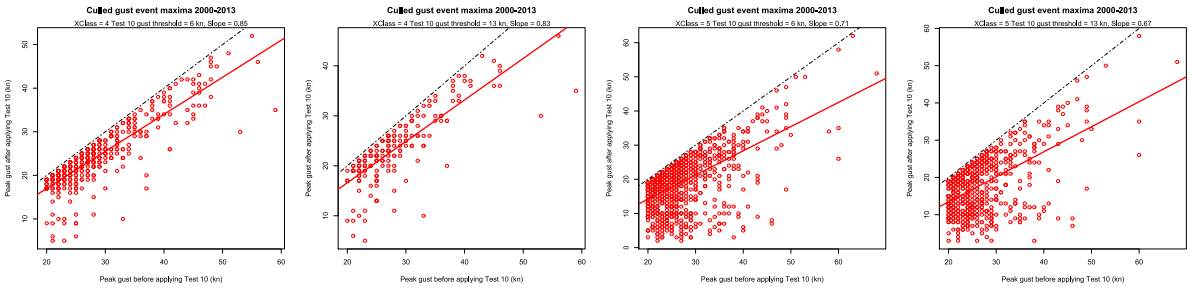
430



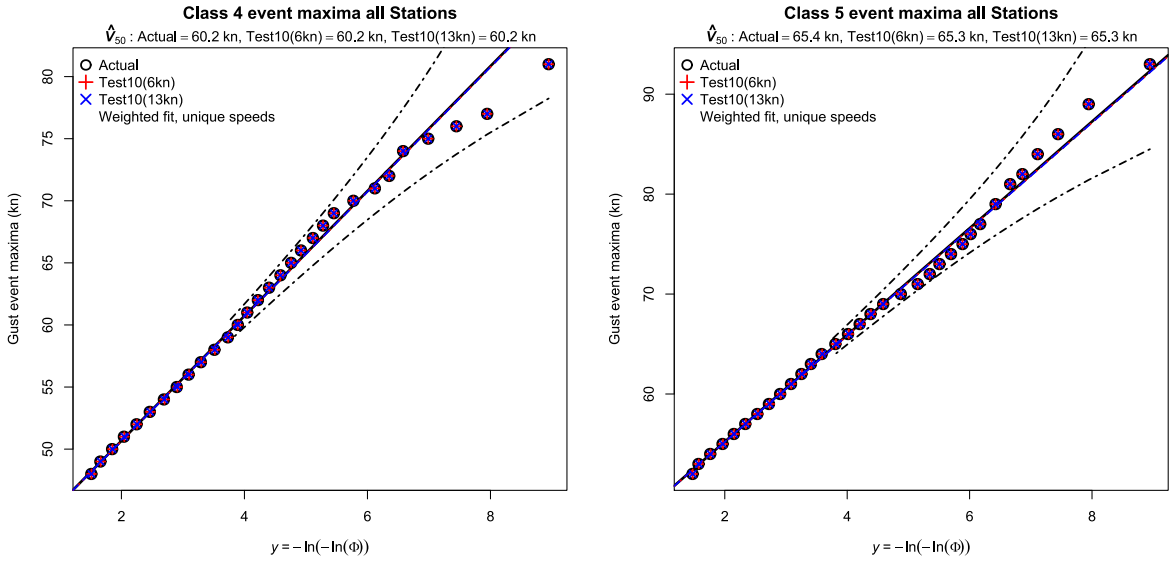
431 Fig. 7. Simulated annual rates of Class 4 and 5 Test 10 false positives from 2000-2013.



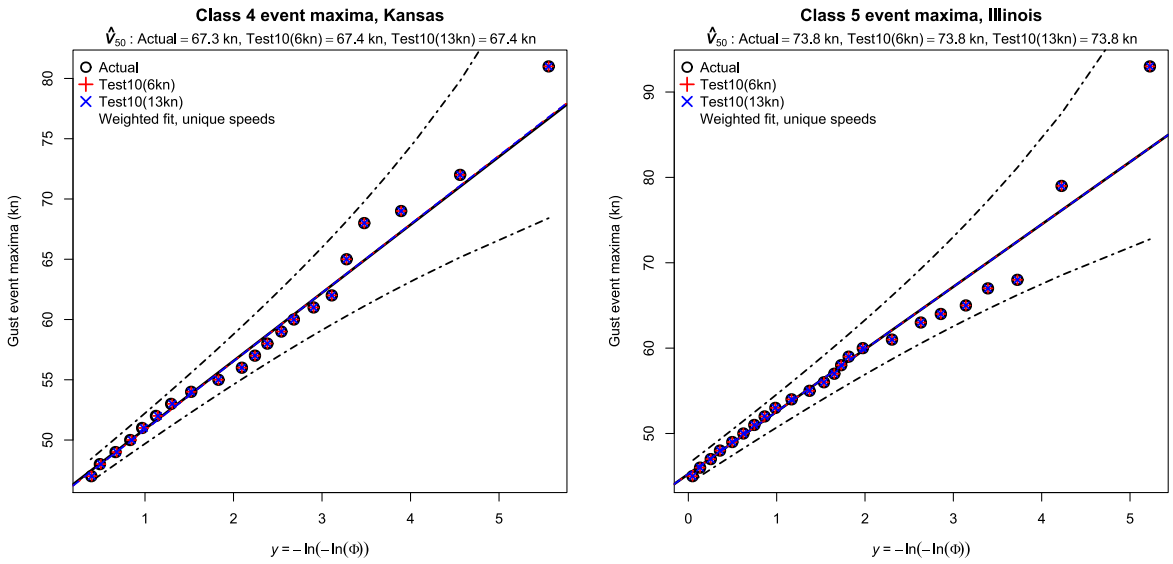
432 Fig. 8. Simulated annual rates of Class 4 and 5 $\hat{V}_{3s}$ underestimated by Test 10 from 2000-2013.



433 Fig. 9. Actual $\hat{V}_{3s}$ and simulated Test 10 underestimates for Class 4 and 5 from 2000-2013.

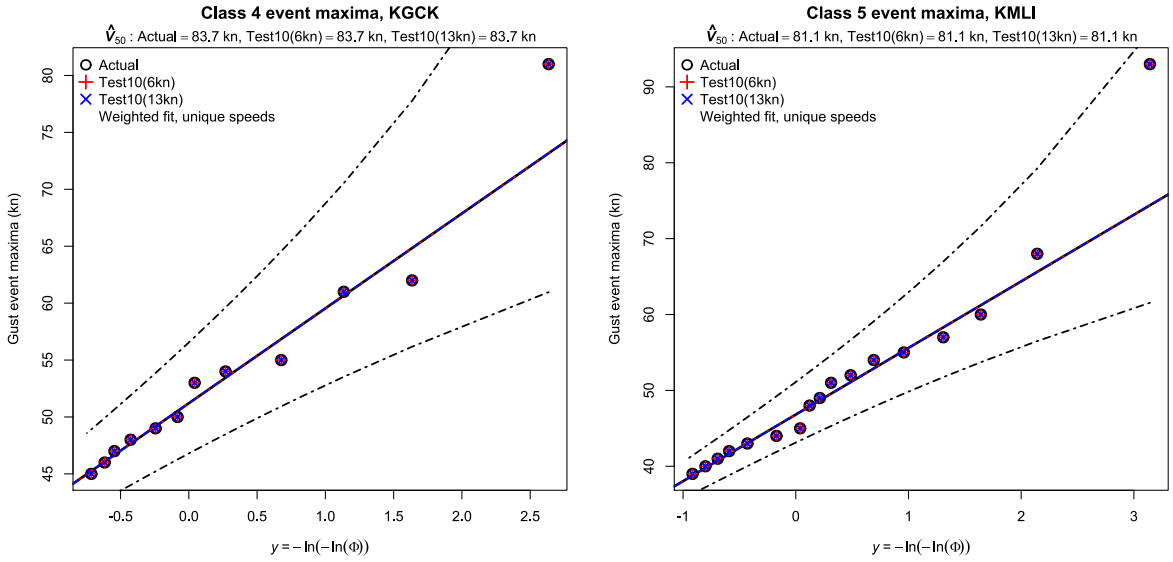


434 Fig. 10. XIMIS analysis of pre-2014 Class 4 and Class 5 gust events for the CONUS superstation.



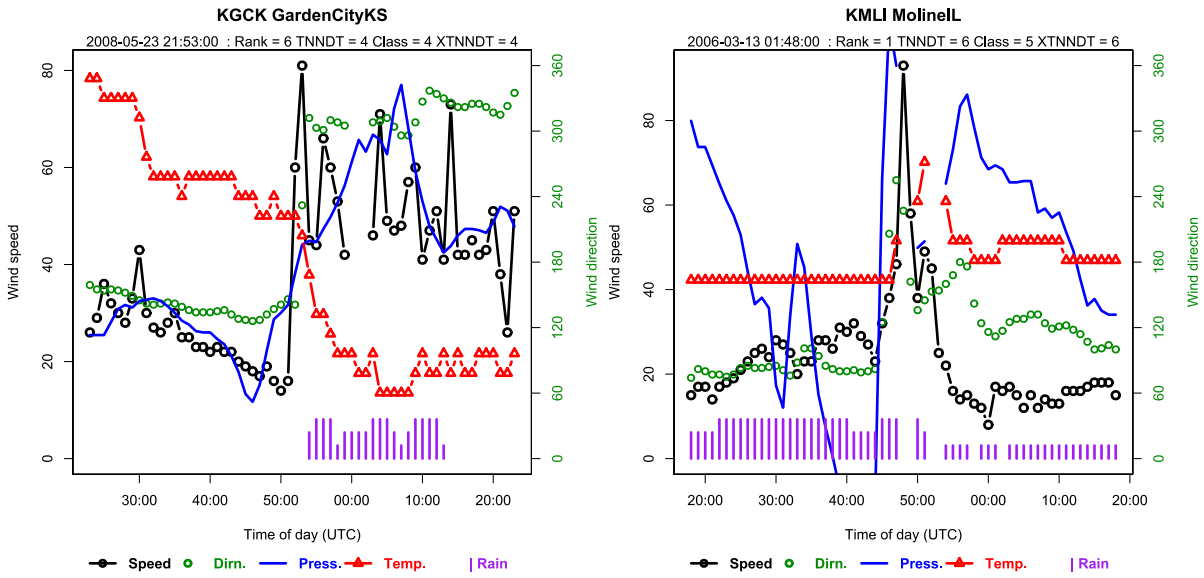
435 Fig. 11. XIMIS analysis of pre-2014 Class 4 for Kansas and Class 5 gust events for Illinois

436 superstations.



437 Fig. 12. XIMIS analysis of pre-2014 gust events, Class 4 at KGCK and Class 5 at KMLI.

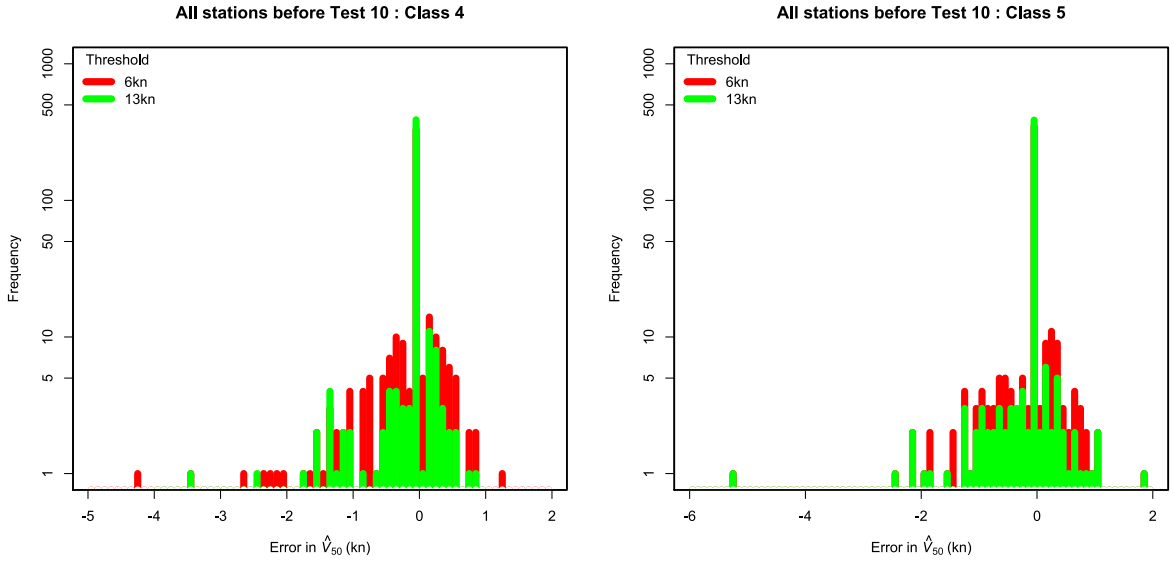
438



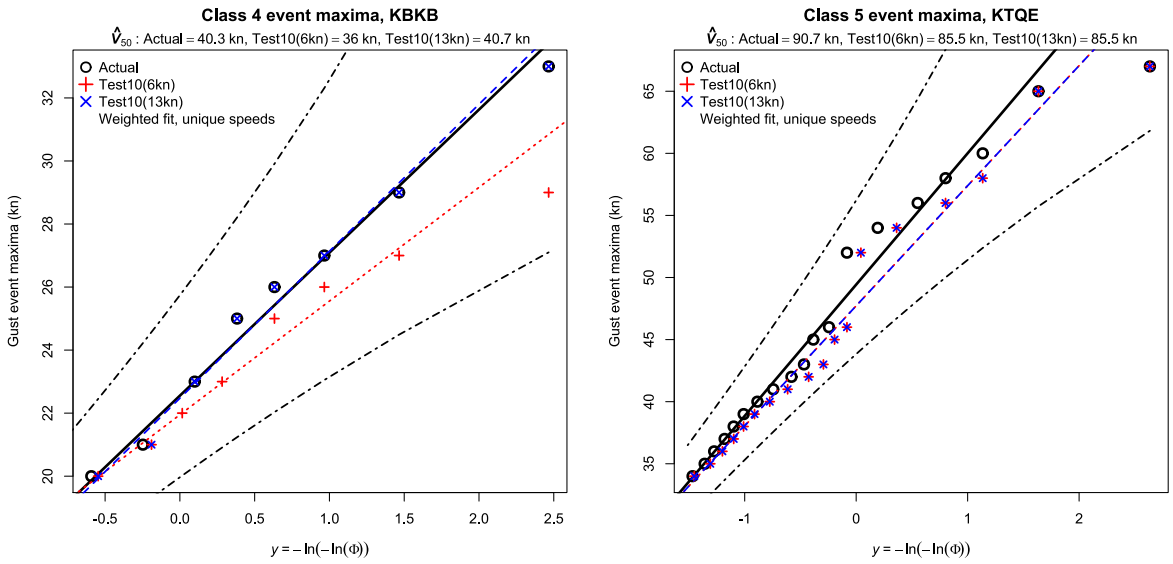
439 Fig. 13. Gust event timeseries for highest observed $\hat{V}_{3s}$: Class 4 at Garden City, KS, and Class 5 at

440 Moline, IL.

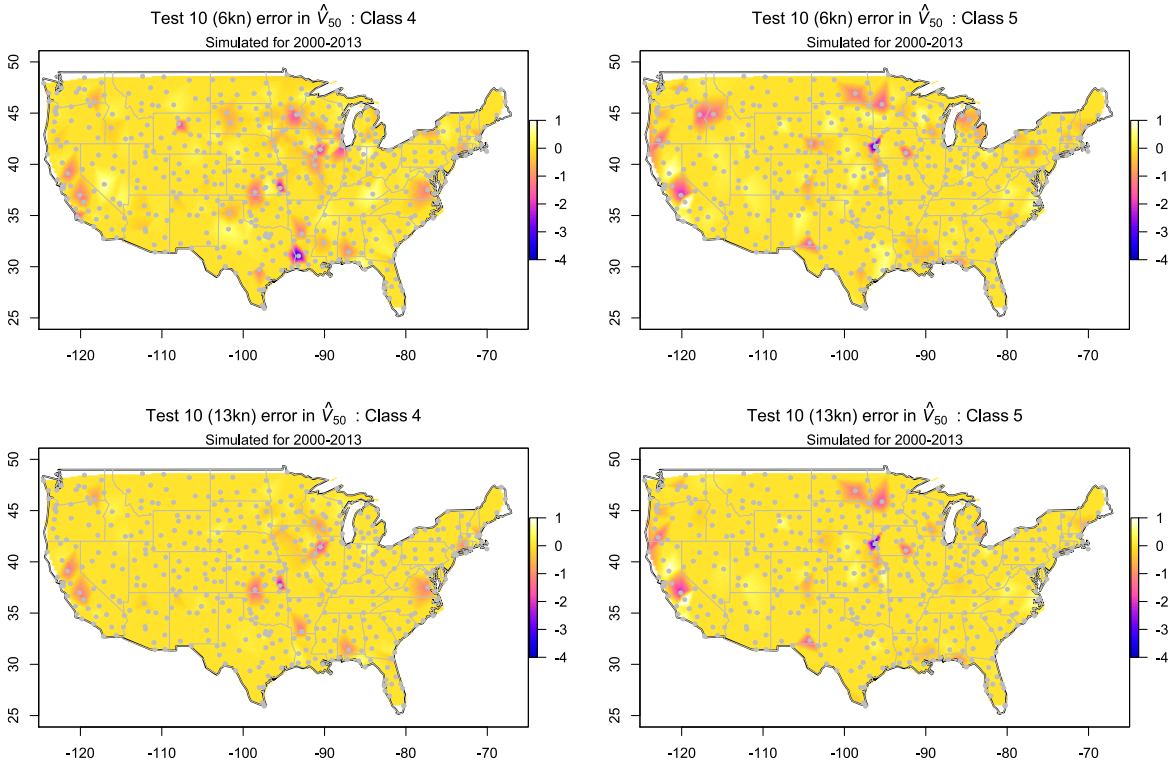
441



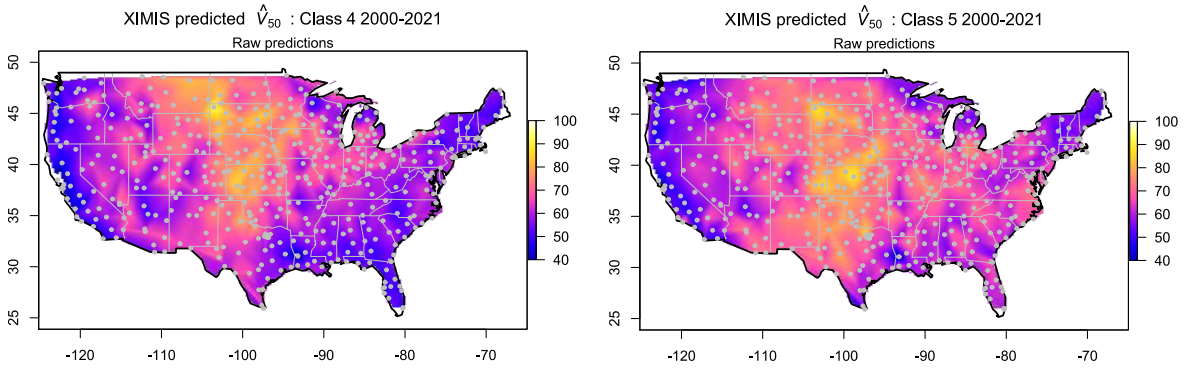
442 Fig. 14. Effect of simulated Test 10 threshold value on error distribution of $\hat{V}_{50}$ for Classes 4 and 5.



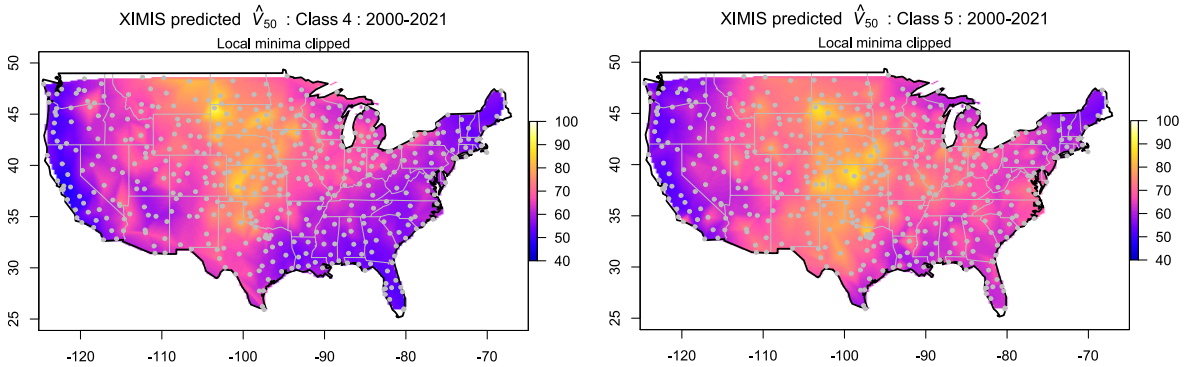
443 Fig. 15. XIMIS analysis of pre-2014 gust events, Class 4 at KBKB and Class 5 at KTQE.



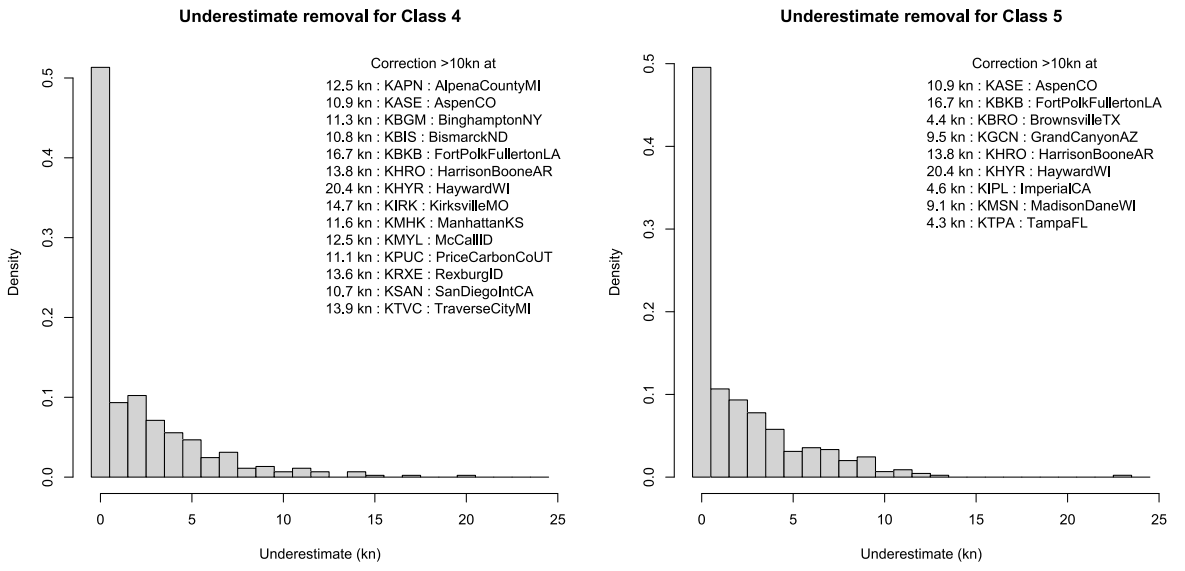
444 Fig. 16. Distribution of error of V_{50} from Test 10 applied to pre-2014 Class 4 and 5 gust events.



445 Fig. 17. Raw XIMIS predictions of $\hat{V}_{50}$ for Class 4 and Class 5 events.



446 Fig. 18. XIMIS predictions of $\hat{V}_{50}$ for Class 4 and Class 5 events with underestimates removed.



447 Fig. 19. Distributions of underestimate correction of $\hat{V}_{50}$ for Class 4 and Class 5 events.