

1 **Curating the TD6405 database of 1-minute interval wind observations across the USA for**
2 **use in Wind Engineering studies**

3 Nicholas J Cook

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

5 Orcid ID: 0000-0002-6390-6884

6 Email: wind@njcook.uk

7 **Abstract**

8 The TD6405 online database of wind observations at 1-minute intervals from 2000 onwards at almost 1000
9 stations across the USA is a highly valuable on-demand resource for Wind Engineering studies, but it
10 contains numerous defects. Many defects are common to other observational databases, for which detection
11 and repair methods already exist. Others are unique to TD6405 for various operational reasons, including
12 the requirement for real time quality control and the adoption of sonic anemometers. Having no moving
13 parts, sonic anemometers provide ideal perches for birds which interrupt the observations and create
14 spurious large gusts on landing and take-off. The real time quality control test added in 2013 to excise bird-
15 generated gusts is shown to miss many of them and to generate four false positives for each one it does
16 find. Each false positive also removes the following 5-minute period of valid observations. Consequently,
17 from 2013 onwards, the incidence of calms is suppressed, and thunderstorm downbursts are
18 disproportionally censored in light winds. An efficient retrospective test is proposed and validated using
19 observations prior to 2013, but the data missing after 2013 is permanently lost and cannot be recovered.
20 Urgent action is required to correct this ongoing situation.

21 **Keywords:** Wind observations, Validation, Artefacts, Quality control, Bird-generated gusts.

22 **Acronyms**

23	ACU	ASOS Acquisition Control Unit
24	ASCII	American Standard Code for Information Interchange
25	ASOS	Automated Surface Observing System
26	BGG	Bird-generated gust
27	CONUS	Contiguous United States
28	DCP	ASOS Data Control Panel
29	FAA	US Federal Aviation Administration
30	HOMR	Historical Observing Metadata Repository
31	ICAO	International Civil Aviation Organization

32	NCEI	US National Centers for Environmental Information
33	NOAA	US National Oceanic and Atmospheric Administration
34	NWS	US National Weather Service
35	QC	Quality control
36	RVR	Runway visible range
37	TD6405	Database of ASOS visibility and wind speed at 1-minute intervals
38	TD6406	Database of ASOS precipitation, pressure, and temperature at 1-minute intervals
39	UTC	Universal Coordinated Time
40	WBAN	Weather Bureau Army Navy
41	WMO	World Meteorological Organisation

42 **1. Introduction**

43 The NCEI TD6405 database holds 2-minute mean and 3s gust wind speeds at a resolution of 1kn and 1° at
 44 1-minute intervals from almost 1000 ASOS stations across the USA. This is a most valuable resource for
 45 Wind Engineering because its scope extends from the macrometeorological peak of the van der Hoven
 46 spectrum, through into the body of the micrometeorological peak. This allows study of mesoscale events,
 47 such as thunderstorm downbursts [1], over much longer observational periods than is possible with targeted
 48 measurement campaigns. TD6405 is presented [2] as a homogeneous set of data in fixed ASCII format files
 49 by station and year. In fact, it is neither of these things due to incremental changes in instrumentation,
 50 acquisition, and QC procedures, and errors in transmission and archiving. The companion TD6406
 51 database, which gives precipitation, atmospheric pressure, and wet/dry bulb temperatures, also at 1-minute
 52 intervals, is not addressed by this paper, but many of the issues in this study also apply to that.

53 TD6405 data begin at the start of 2000, when Belfort cup/vane anemometers with a 5s gust response were
 54 in operation at all the NWS stations, most at 10m above ground, some at 26 feet (7.9m), and a few at other
 55 heights. These were susceptible to icing in the winter. Later, two countries: USA, and Germany, each
 56 began a programme to replace all their cup/vane anemometers with sonic anemometers, all at 10m above
 57 ground. The perceived advantages being that sonics have a high frequency response, no moving parts that
 58 degrade, and can be heated to prevent icing. In the USA, this was one key element in a wider programme of
 59 improvements deemed “*necessary to keep pace with the expanding demands placed on the ASOS for*
 60 *timely, accurate delivery of more observation parameters to more external users in real time*” [3]. After
 61 installation, at a different date for each station, the gust duration changes to 3s to comply with the WMO
 62 standard gust. The progress of the ASOS implementation and upgrade programme is documented in a
 63 series of NOAA reports at <https://weather.gov/asos/ASOSImplementation>.

64 The principal, unforeseen [4] disadvantage of “no moving parts” sonic anemometers on isolated poles
 65 surrounded by grassland is that they provide ideal perches for birds which sit and block the acoustic path
 66 between sensors. However, in Germany, not content with just perching, the birds (principally corvids) peck
 67 and damage the sensors, requiring replacement of 1/3 of all the anemometers in every year [5]. The
 68 problem also seriously affected the Low-Level Windshear Alert System that had been installed at ~100
 69 susceptible US airports in response to the 1975 accident at JFK airport in New York. Installation of sonics
 70 at ASOS stations was temporarily halted in 2007 while a solution to this problem was sought. Various
 71 methods of bird deterrence were devised and trialled [5,6,7] but no single method was found to be effective
 72 for all bird species. ASOS installation was resumed on the conclusion that and that the issue was best
 73 addressed in the acquisition software, supplemented by deterrent spikes or other physical methods at the
 74 most susceptible sites. The final version of the QC algorithm [8] was implemented at the end of 2013,
 75 although it was tweaked in 2017 to reduce false positives. In addition to artefact types specific to ASOS,
 76 there are additional transmission and archiving artefacts in common with other databases, but, as ASOS is
 77 fully automated, none that are associated with manual input of data [9,10]. Of the few papers in this Journal
 78 [1,11,12] that have used TD6405 data so far, only Chen and Lombardo [1] take any account of issues
 79 affecting the reliability of the data, simply by discarding gusts >75kn.

80 This paper determines the types of artefacts present in the TD6405 database and demonstrates a targeted
 81 methodology implemented in the statistical language R to detect, correct them where possible, or otherwise
 82 remove them. Each type is addressed in the order it is encountered when accessing the database for
 83 analysis, which is approximately the reverse order in which it is imposed on the data.

84 **2. Methodology for finding and addressing artefacts**

85 *2.1 Formatting errors when reading the data*

86 The TD6405 database comprises fixed-format records. Examples of a fully populated record for KIAD
 87 Washington Dulles and KDCA Washington Reagan are:

```
88 93738KIAD IAD2011032218102310 0.079 D 0.101 D 0.083 D 37 6 42 6 01R60+
89 13743KDCA DCA2014010219230023 3.811 N 5.101 N 5 5 7 6 01 50
```

90 where:

93738KIAD 13743KDCA	The concatenated 5-digit WBAN and the 4-character ICAO codes for each station.
IAD2011032218102310 DCA2014010219230023	The concatenated 3-character FAA code for each station; the year, month, hour and minute in local standard time; and the hour and minute in UTC.

0.079 D	0.101 D	0.083 D	The visibility coefficients and day/night codes for the visibility photometers: three at						
3.811 N		5.101 N	KIAD and two at KDCA.						
37	6	42	6	The mean direction, mean speed, gust direction and gust speed in integer knots and					
5	5	7	6	degrees.					
01R60+	The runway bearing and visible range (RVR) codes for the principal runway: KIAD								
01 50	runway 01R, better than 6000 feet; KDCA runway 01, 5000 feet.								

91 The ASOS system transmits from the data control panel (DCP) of each instrument package to a central
92 Acquisition Control Unit (ACU) which assembles the record for onward transmission, prefixing each
93 record with the station identification and the local standard date-time. This prefix is indicated in the
94 examples by bold type to distinguish it from the data stream from the DCP. Transmission glitches between
95 DCUs and the ACU may insert stray characters but, more often, transmission is interrupted, and data are
96 lost. The value immediately prior to a data gap is often garbled. The simplest and fastest, approach is to
97 discard all such records, but this leads to a significant loss of recoverable data.

98 In these consecutive records from Washington Reagan:

```

99 13743KDCA DCA2013120418152315 0.167 N 0.189 N 66 6 70 6 01 60+
100 13743KDCA DCA201312041816231669 2 69 2 01 60+
101 13743KDCA DCA2013120421270227 0.209 0.228 N 69 2 63 2 01 60+
102 13743KDCA DCA2013120421280228 0.212 N 0.230 N 67 2 71 2 01 60+
103 13743KDCA DCA2013120421290229 0.213 N 0.233 N 3 01 60+
104 13743KDCA DCA2013120421300230 0.216 N 0.235 N 68 2 72 3 01 60+

```

105 data are missing from the middle of the 2nd, 3rd and 5th record. The four wind data integers that precede the
106 runway bearing are intact in the 2nd and 3rd, so are recoverable. Where the lost data includes the end-of-line
107 ASCII character, the next data may be appended to the record, as in this example where the data in the
108 record for 10:51 UTC belongs to 10:53 UTC:

```

109 13743KDCA DCA2014022205501050 0.088 N 0.153 N 213 5 212 5 01 60+
110 13743KDCA DCA2014022205511051 1053 0.085 N 0.144 N 214 5 212 5
111 01 60+
112 13743KDCA DCA2014022205541054 0.082 N 0.141 N 208 4 197 5 01 60+

```

113 This is deducible because a second UTC code (underlined) in the record corresponds to the minute before
114 the next valid record.

115 Rescuing the recoverable data requires each record to be parsed. The required wind data values, the RVR
116 runway bearing (at airports without parallel runways) and any second UTC code in the record are all
117 integers separated by spaces, and these are the target characteristics for parsing the data.

118 2.2 Duplicated date-times

119 Most duplicated date-times present in pairs which are either identical, when either can be deleted, or where
120 the first is erroneous and the second is a correction, but not always in that order. In this example:

121	13743KDCA DCA201003031034 1534	0.073 D	0.103 D	352	13	354	18	01	60+
122	13743KDCA DCA201003031034 1534	0.073 D	0.103 D	352	13	354	1	60+	

123 the corrected record precedes the erroneous one in the database. It is quite normal for a database to retain
124 erroneous records and their corrections, the UK Met Office MIDAS database does this too, but it is also
125 logical for the correction to appear after the error. This implies that the TD6405 files have been sorted by
126 date to bring error and correction together using a method that does not preserve the original order, i.e., by
127 a lexicographic sorting of the whole record. Sometimes there are two, and rarely three, such consecutive
128 error/correction pairs.

129 There are also many long runs of duplicated date-time pairs in which the wind data in each pair differ but
130 appear valid. Here, a minimum run length (10 pairs) of identical wind values was specified to distinguish
131 these from the error/correction pairs. Almost half of these runs start at 00:00 UTC and are exactly 12 hours
132 in length, indicating a systematic issue that affects all ASOS stations. A typical example is shown in Table
133 1, in which the left-hand panel lists the wind data for the first of each pair, the middle panel for the second
134 of each pair and the right-hand panel for the same time on the next day. The ↓ arrows indicate where data
135 has been omitted for brevity. At the head of the table, the data from 00:00 to 01:16 (in bold) for the first of
136 each pair are identical to that on the next day. At the tail of the table, the data from 06:54 to 11:59 (in bold)
137 for the second of each pair are identical to that on the next day. Between 01:17 and 08:11 all data on the
138 next day are missing. Over the middle period, it is seen that the trend in values continue to match the head
139 until 06:53 when it swaps over to match the tail. So, the data in bold type is the correct data for the
140 corresponding date-times on the next day and provides the missing data.

141 In a trial of 50 stations distributed across the contiguous USA, the heads and tails of all duplicated date-
142 time records show the same match to the same times where there is data on the next day, or when data is
143 missing for the whole period. In the latter case, to be sure, the R script searches 7 days ahead for any
144 matching data on those days. No exceptions were found, confirming that this is an issue only affecting
145 consecutive days, i.e., a date assignment problem when the local date/time is prefixed to the ASOS data
146 record.

147 The swap-over from head to tail in Table 1 is relatively obvious because the wind conditions on the two
148 days are quite different, but it is common for wind conditions on consecutive days to be very similar. The R
149 script tracks the trend in the wind data using the absolute difference in each mean and gust speed and

150 direction value between the current and previous minute, weighting the mean and gust speeds by a factor of
151 10 to equalise ranges, and summing to produce an overall difference “score”. In each comparison the
152 greater score indicates a possible swap, but a swap is authorised only when both tests agree, i.e., 1st→2nd
153 **and** 2nd→1st. There may be many such swaps in a duplicated period – for example, 40 swaps in one 12-
154 hour period at KTUS Tucson, AZ. Where the head and tail are swapped, as in the above example, the
155 number of swaps must always be odd, otherwise even where they are not swapped. In a few rare cases that
156 fail this rule, usually because conditions on the two days are very similar, the swap with the lowest
157 indicated score is discounted to comply with this rule.

158 *2.3 Summary of TD6405 record-parsing artefacts for KTPA Tampa International*

159 Figure 1 shows the number per month of each of type of artefact found when parsing the TD6405 record
160 for KTPA Tampa International, 2000-2020. As a reference, the upper envelope of $\sim 10^3$ per month
161 corresponds to $\sim 23\%$ of the observations. Apart from the change from cup to sonic anemometer at the date
162 marked by the thick vertical line, the process of acquisition through to archiving of the ASOS data
163 (described in more detail later) does not change fundamentally, but there are incremental upgrades of
164 hardware and software throughout the period, as described in the implementation reports. Although
165 “release dates” for these changes are given in the implementation reports, the dates they are implemented at
166 any station are not given there, nor in the Historical Observing Metadata Repository (HOMR) at
167 <https://www.ncei.noaa.gov/homr/>. However, these changes are reflected as clear patterns in Figure 1 that
168 indicate sudden changes in characteristics. For example: missing, bad (unrecoverable), minute-shifted, and
169 out-of-range records are consistent in the cup data until the start of 2004; in 2005 the rate of missing
170 increases, of bad remains consistent, and the rates for other two types drop; then from the start of 2006, the
171 characteristics suddenly become more random. A second large change in character occurs in the sonic data
172 from the start of 2014 with a higher, consistent rate of bad records that may be an unintended side-effect of
173 the major change to the ASOS real-time QC algorithm described next.

174 *2.4 Bird perching events and the ASOS QC algorithm.*

175 The concerns expressed by NOAA [13] about bird perching events at ASOS stations in the summer of 2007
176 appear to be more about spurious gusts caused when birds flapped their wings on landing or take-off than
177 about the period of data loss while the birds blocked the acoustic path. It was noted that 7 sites were
178 responsible for more than 50% of the error reports. This report [13] also notes an unrelated error that
179 whenever a visibility sensor checksum occurred it overwrote the DCU memory containing the wind data.

180 The vertical bars in Figure 2 show the gust wind speed values at Tampa International, FL, (KTPA) for four
181 occasions that match the reported characteristics of a bird perching event. The gust values immediately

182 before and after a loss of data are marked by the black circles and the corresponding 2-minute mean values
183 by the open circles. The two horizontal dashed lines are the threshold criteria for the ASOS QC algorithm.
184 The continuous line is the 10-minute centred running mean, discounting the missing gust values, which is
185 explained later. Whether a bird generates a gust (BGG) “spike” on landing or on take-off depends on
186 whether it flaps its wings to do so: (a) for landing and take-off, (b) and (c) only for take-off, (d) only for
187 landing, or not at all. The QC algorithm introduced at the start of 2014 is supposed just to remove the bird-
188 generated gusts as nothing can be done to recover the lost data. Note that a bird landing and taking off in
189 the same minute or in the next minute, could cause one or two adjacent BGG spikes with no loss of data,
190 and this will be indistinguishable from a very short duration meteorological event (e.g., a dust devil) or a
191 data transmission error. In (c) the take-off spike is followed by a larger value, so that either the bird took
192 off in the three running 3s periods spanning the change of minute (~1:20 chance), or another bird arrived
193 and left in the second minute.

194 Before reviewing the ASOS QC algorithm it is first necessary to understand the data acquisition software
195 logic [8,13]. The sonic anemometer generates speed and direction values 1s intervals, computes a running
196 3s average and a 5s block average, and every 5s sends the 5s average and maximum 3s average during that
197 5s to the DCP. The DCP forwards these 5s-interval data to the ACU unchanged. Before July 2012, the
198 ACU truncated the speeds (e.g., 22.9kn became 22kn), but this was corrected to rounding to the nearest
199 integer in the July 2012 software release [14]. **All sonic speeds prior to this date are biased by –0.5kn.**
200 Every 5s the ACU computes the average 2-minute mean from a minimum of 75% (18) of the previous 24
201 values, with any value failing the QC test treated as missing. When a perching bird causes a data gap, BGG
202 spikes must occur in the last/first 30s of the 2 minutes to avoid being automatically censored.

203 The QC algorithm works in real time at 5 s intervals, and this is its *Achilles heel* because it cannot
204 anticipate a future loss of data when assessing a landing BGG spike, nor can it anticipate future low values
205 when assessing a take-off BGG spike. It performs the 10 tests listed in Table 2 on the current 3s gust, 5s
206 mean and 2-minute mean values and the values are discounted if any test fails. Tests 1 through 8 assess the
207 validity of the values. Test 9, the “high wind” test, and Test 10, the “low wind” test, are intended to find
208 non-meteorological events. The internal heaters may not be sufficient to prevent sensors icing up in winter
209 freezing-rain storms, generating spurious high gusts [15], and this is the main target of Test 9. Birds tend to
210 perch at low wind speeds, and this is the main target of Test 10 – note that the gust threshold was raised
211 from 6kn to 13kn in late 2017 to reduce the rate of false positives.

212 The NOAA evaluation of the performance of the QC algorithm [8] includes the following comments:

213 a) *“When the QC Algorithm removes six or less 5-second samples in a 2-minute period, the removal*
214 *of those samples will be transparent to the end user.”*

- 215 b) *“When the QC Algorithm removes more than six 5-second samples, the wind report will go*
216 *missing and will take a minimum of 4 ½ minutes to return.”*
- 217 c) *“In order to remove 97% of the corrupted 5-second samples, the QC Algorithm must remove some*
218 *good data as well.”*
- 219 d) *“The median amount of good data removed by the algorithm is 0.33%, which will be*
220 *representative of the vast majority of the ASOS Network.”*
- 221 e) *“There will be some sites that will experience good data loss of 1% or more, and other sites where*
222 *good data loss will be near 0. Sites with unorthodox siting constraints, sites in mountain passes*
223 *and in desert areas tend to have above median good data rejection rates.”*
- 224 f) *“The majority of good data rejected are “ordinary” samples. However, the algorithm can*
225 *wrongfully reject a piece of good data that is of operational significance.”*

226 The 0.33% loss of valid data reduces to 0.17% after the gust threshold is increased from 6kn to 13kn in late
227 2017 – but this is a small percentage of a very large population of values, corresponding to ~1750 per year
228 and ~900 per year respectively. These false positives are biased towards sudden gusts in light winds, i.e.,
229 towards thunderstorm microbursts and other small thermally driven surface events, such as dust devils,
230 which are of great interest to Wind Engineering.

231 As Test 10 uses only the 2-minute mean and 3s gust data, its performance in finding bird-generated gusts
232 can be assessed retrospectively from these values in the TD6405 database – the only difference being a
233 minor mismatch in timing: in TD6405 the 2-minute averaging period is relative to the UTC minute
234 containing the 3s gust, whereas in the real-time QC it is the 2 minutes immediately preceding the gust.

235 For Test 10 applied to the “raw” uncorrected TD6405 data for KTPA, Figure 3(a) shows that the
236 distribution in time and value of gust speeds that would fail the test. These divide into three distinct time
237 periods:

- 238 1. Before 27 January 2009: Belfort cup anemometer – 28.4 per year. The cup anemometer was not
239 susceptible to bird perching, so this indicates a background level for QC failures from other causes.
240 Inspection indicates that only 8.9% are false positives: either sudden gusts in light winds as in
241 Figure 4(a), or after a lull in strong winds 4(b). The remaining 91% are spikes at the head or tail of
242 gaps 4(c) which mimic landing spikes, or isolated spikes 4(d), both types being very typical of
243 transmission errors [9,10].
- 244 2. After 27 January 2009 and before 2014: Sonic anemometer before Test 10 – 81.4 per year.
245 Inspection indicates that only 2.5% match the types of bird event in Figure 2, and 3.3% are isolated
246 spikes in consistently low wind speeds as in Figure 4(d). 83.3% are false positives of the same
247 types as in Figure 4(a) and (b). The remaining 11% are spikes in moderate winds, where it is

248 unclear if they are valid or artefacts. Figure 5 shows examples of false negatives, i.e., a gust that
249 should have been flagged: in (a) an isolated spike of 24kn is flagged, whereas an earlier isolated
250 spike >100kn and a later value at the head of a gap are not flagged because the corresponding mean
251 values exceed the 6kn threshold; similarly, in (b) the spike at the head of the data gap is flagged,
252 but not at the tail.

253 3. From 2014: Sonic anemometer after Test 10 – 2.25 per year. There are very few retrospective fails
254 after Test 10 was added to the ASOS QC algorithm. Most of the discounted events are false
255 positives and, for each one removed, the following 4 or 5 minutes of good data is lost due to the
256 QC deadtime (comment b). Some surviving events retrospectively caught by Test 10 are shown in
257 Figure 6.

258 (a) Here, a sudden gust is flagged because the mean in TD6405 has been rounded down to the
259 6kn maximum threshold. The length of the following gap matches the QC deadtime,
260 suggesting Test 10 was triggered 2 minutes later. This is likely to be a valid thunderstorm
261 downburst that has been wrongly censored by the QC algorithm. So, this is a false positive.

262 (b) Here the gust immediately after a brief lull is flagged. This is another false positive.

263 (c) The spike at the gap tail should have been removed by the QC algorithm as it meets both
264 threshold conditions. This is an unexpected and unexplained error.

265 The 4- or 5-minute deadtime is operational through the whole ASOS period but is a particular problem
266 after 2013 when triggered by Test 10. Figure 7 shows that the rate of 4- or 5-minute gaps increases by a
267 factor of ~20 after Test 10 is introduced and >80% of these are expected to be false positives. Raising the
268 gust threshold in 2017 only seems to make matters worse by increasing the number of gaps in the 6-minute
269 to 12-minute range. A major Wind Engineering concern is that many thunderstorm downbursts in relatively
270 calm winds will be censored in the record from 2014 onwards.

271 *2.5 A proposed replacement bird-perching algorithm.*

272 The period between installation of the sonic anemometer and adoption of Test 10 remains uncorrected for
273 bird-perching events and may therefore be used to test alternative algorithms that work retrospectively. The
274 benefit of hindsight allows a replacement algorithm to be devised and tested that targets the form of the
275 artefacts shown in Figure 2, specifically the BGG spikes in the minute immediately before and/or after a
276 gap in the data, without generating any false positives. The fixed 6kn and 13kn thresholds in Test 10, which
277 generate most of the false positives, are a weakness that can be avoided by requiring the presence of a data
278 gap immediately before or after a gust factor threshold is exceeded. This also allows for bird events where
279 the mean is above 6kn or the gust is below 13kn to be detected. The current QC gust factor is based on the
280 2-minute mean so it becomes infinite in periods of calm, when all gusts >13kn will be rejected. Using a

281 modified gust factor which is based on the running 10-minute average gust speed centred on the gust, the
282 (blue) line in Figure 2, avoids this issue.

283 The proposed replacement algorithm is simple: a) calculate the centred running 10-minute average of the
284 gust speeds; b) locate all gaps in the data; c) calculate the gust factor for the value immediately before and
285 after each gap; d) remove any which exceed a suitable gust factor (typically 1.5). Repeat this to catch any
286 double spikes as in Figure 2(c). Figure 3(b) shows the effect at KTPA of applying Test 10 after first
287 applying this targeted algorithm:

- 288 1. The rate of QC failures in the cup anemometer period is halved by the removal of just those spikes
289 adjacent to gaps that mimic bird-perching events. This leaves only the isolated spike events that are
290 not targeted.
- 291 2. The rate of failures in the sonic period before Test 10 is only marginally reduced because they are
292 mostly false positives. The 1.5% drop corresponds to bird-perching events that also match the Test
293 10 threshold criteria, about 60% of all the targeted events removed.
- 294 3. A very few failures in the sonic period after Test 10 are caught by again applying Test 10. All are
295 spikes that only just meet the threshold criteria, so might reasonably be deemed to be valid. This is
296 illustrated by Figure 8 where there are two 14kn spikes, 2 minutes apart, where the first passes and
297 the second fails. Both these spikes are passed by the proposed replacement algorithm.

298 *2.6 Identifying and correcting remaining artefacts.*

299 So far, only the artefacts unique to the ASOS data have been addressed. The typical artefacts generated by
300 acquisition, transmission and archiving errors remain in the data and these have been studied before in
301 some detail [9]. Those characteristic artefacts due to manual data entry, e.g., transpositions, double
302 keystrokes, etc., which are usually correctable [9,10], are not relevant to the fully automated ASOS data.
303 The artefacts that remain are mostly isolated spikes in value within the valid ranges which are caused by
304 transmission and archiving glitches. Procedures devised to identify these using thresholds on the **change** of
305 wind speed and direction (expressed as an arc) are fully described in [10]. To illustrate this approach,
306 Figure 9 shows the probability density distribution of the first differences of gust speed: viz., of the change
307 in gust speed between successive minutes, for each of the three time periods at KTPA. The dashed line
308 corresponds to the expectation of only one event in the whole record, so that any event that lies outside
309 where the trend in the body of the distribution meets this line is taken to be an artefact spike.

310 In Figure 9, it is obvious that there are many more out-of-trend events in the Belfort cup anemometer
311 period (a) than in the sonic anemometer periods (b) and (c). The implication is that these are caused by the
312 Belfort anemometer and/or its data feed to the DCP. Figure 9(b) shows a much-reduced incidence, but

313 those that do occur span the full $\pm 125\text{kn}$ range of valid values. Figure 9(c) shows that a few very large
314 spikes remain that are either large enough to raise the running 2-minute mean above 6kn, so Test 10 is not
315 triggered, or are subsequent transmission errors. In all three cases, the optimal threshold on first difference
316 is $\sim 30\text{kn}$. First difference thresholds for the mean speed and the mean and gust directions are obtained
317 similarly [10].

318 The methodology advocated in [10] is that a spike comprising one (singleton), two (pair), or three (triplet)
319 consecutive observations is an artefact only if both rising and falling first differences exceed the threshold.
320 As this requires valid before and after values to calculate the change, the method will not work next to a
321 data gap, which is why BGG spikes must be addressed separately. By this criterion the passage of fronts,
322 where wind speed rises rapidly then remains high, or thunderstorm downbursts, where wind speed rises
323 rapidly then decays more slowly [1], are not artefacts and are not censored. Note that the method in [10] is
324 applied separately to speed and direction because these are conventionally measured by two different
325 instruments. The sonic anemometer provides both, so would allow a combined test based on the change in
326 vector but, for compatibility, this study retains the [10] methodology.

327 **3. Results of the proposed methodology.**

328 Ten of the 50 ASOS sites, listed in the Appendix, were studied to demonstrate the performance of the
329 proposed methodology. Among these, KIAD Dulles and KDCA Reagan are the international and regional
330 hubs, respectively, for Washington DC so might reasonably be expected to be similar, while KMYV
331 Marysville Airport, CA, is the site with the highest reported occurrence of bird perching during August
332 2007 [13].

333 Table 3 lists the individual curation stages and the performances statistics for each step in the data
334 extraction and artefact cleaning processes at the ten sites. Counts of artefact types that are much more
335 frequent at certain sites are highlighted in the table and were examined by year to reveal the effects of the
336 incremental changes in instrumentation, acquisition, QC procedures and archiving.

337 Figure 10 shows the annual distributions of the highlighted artefacts found in the extraction phase, when
338 records are read and parsed:

339 **(a)-(b) Undecipherable records.** KTPA shows a consistent high rate in the first years of ASOS which
340 appears to be corrected in 2005. After sonic installation the rate is high in some years, building to a
341 peak after the 2017 QC tweak. KTUS follows a similar rising trend in the sonic period (note change in
342 scale) with 2018 considerably worse. This issue appears to remain current.

343 (c)-(d) **Single minute shifts.** These only occur in the cup anemometer period, as demonstrated at
344 KSEA and KTPA. They are resolvable, but this only assigns the observation to a different time and
345 there is no net gain.

346 (e)-(f) **Duplicate date/times.** At KSNA there is no coherent pattern, but at KTPA they correspond to
347 the introduction of Test 10. They are common across all stations, but this is a retrospectively resolvable
348 issue, as demonstrated in §2.2.

349 (g) **Missing observations.** These create multiple large gaps in the data record at many stations which
350 place constraints on a station's usefulness in Wind Engineering studies. Here at KDCA the incidence
351 rises after sonic installation in late 2006 to a peak in 2016. The fall to the current low incidence
352 coincides with the 2017 QC tweak.

353 (h) **Ambiguous wind data integers.** This defect occurs when the four wind values cannot be uniquely
354 found in the group of consecutive integers immediately before the RVR code, as shown in the §2.1
355 examples. At multiple-runway stations, or where no RVR is reported, there should be 4 integers. At
356 single-runway stations there should be 5 or 6 integers if the range is at or below maximum, respectively
357 because the code for maximum visibility (60+) is not interpreted as an integer. So, the wind values
358 should always be the first 4 integers of the group, and any fifth should always be the same runway
359 bearing. Any integer group that fails this rule is ambiguous: less than 4 indicates some wind values are
360 missing; and a fifth that is not the runway bearing indicates a spurious extra value somewhere in the
361 group. This occurs frequently only at KDCA in the year 2014, indicating a temporary problem that has
362 been fixed. It does not occur in any of the other 9 example stations, so may be a unique quirk.

363 Figure 11 shows annual distributions of the highlighted artefacts found in the cleaning phase, when
364 duplicated date/times are resolved, bird gusts are removed by two passes of the proposed algorithm of §2.5,
365 and the isolated spike artefacts are removed by the [10] method as in §2.6:

366 (a)-(d) **Mean and gust calms.** KDCA and KSNA show the highest rate of calms of the ten example
367 stations. The rate of mean and gust calms at both stations drops suddenly after the 2017 change to Test
368 10. This was expected because the gust factor becomes infinite, but why this drop did not occur in 2013,
369 when Test 10 was first introduced and was more stringent, is a mystery.

370 (e)-(f) **Identical observations.** Identical records were removed in the extraction phase. Here, identical
371 wind observations occur when a faulty record is corrected but the fault was not in the wind integer
372 group. They mostly occur in the early sonic period, before Test 10, as at KSNA (e); but KTUS also
373 shows a peak in 2018. This issue is resolved by removing one of the duplicates.

374 **(g)-(h) Duplicate date/times.** This is the issue, common across all 50 ASOS stations studied, that is
375 resolved by the method in §2.2. There is no clear pattern to when it occurs: this appears random at
376 KSNA (g) but is weighted to 2015 onwards at KTPA (h). That this issue consistently occurs at the UTC
377 change of day suggests that it is due to an incompatibility between local/UTC date/time conversions of
378 the various operating systems in the transmission-to-archiving data path that is ongoing.

379 **(i)-(l) Bird gusts.** Here, bird-generated gusts are defined by the characteristic spikes in sonic speeds
380 immediately before or after a data gap as in Figure 2. However, they appear much more frequently in
381 the cup anemometer period (Figure 9(a)), indicating that other gaps due to instrumentation faults or
382 icing share these characteristics. No significant rate drop coincides with the introduction of Test 10 in
383 2013 – instead, at KMYV and KTUS the rate increases to around 200 per year due to the high rate of
384 false positives.

385 **(m)-(n) Short runs.** The many short gaps in the cup anemometer period fragment the data record and
386 create many runs of data that are too short to be useful. This study is a precursor to a current
387 investigation that requires 10-minute means (the METAR standard mean) so, here a short run is
388 defined as a run shorter than 10 consecutive values. The examples for KSEA (m) and KTPA (n) show
389 that this issue was resolved after 2004, probably by the software upgrade [16] to allow cup and sonic
390 anemometers to operate together temporarily. Note that removal of isolated spikes by the §2.6 method
391 [10] in the next cleaning step can create additional short runs, i.e., the “Residual short runs” in Table 3.

392 **(o)-(p) Single spikes.** These are the singleton spikes addressed by the [10] methodology in §2.6. At this
393 point, the spike artefacts next to gaps which mimic BGGs have already been located and removed, so
394 these are the remaining isolated spikes within runs of good data. KIAD (o), KTPA (p) and KMYV
395 show the highest rates are centred around 2008, but the other 7 stations show no trend. The pair and
396 triplet spikes of §2.6 are far less numerous.

397 A spike at head or tail of a gap may not be a BGG because such spikes are also common due to equipment
398 malfunction, accounting for half occurring in the cup anemometer period, or is at the head/tail of a gap after
399 a Test 10 false positive, e.g., a censored thunderstorm downburst.

400 The results for KDCA Regan and KIAD Dulles were expected to be similar due to their closeness, but the
401 maximum validated 3s gust speed at KIAD (115kn) is almost twice that at KDCA (64kn) and close to the
402 maximum rated speed for the ASOS system of 125kn. The wind speeds for both these events are shown for
403 comparison in Figure 12. Inspection of the KIAD timeline shows that the gust speeds are characteristic of a
404 thunderstorm downburst [1]: a sudden increase followed by a gradual return over several minutes, after
405 which the original wind direction is restored. The gust speeds at KDCA are characteristic of the passage of

406 a front: a steady increase over ~5 minutes followed by a faster decrease, with the wind direction veering
407 from ~180° to settle at ~270° behind the front.

408 Gust speeds in the cup anemometer period at each station are nominally 5s gusts but, as cup anemometers
409 have a constant length scale, the gust duration varies with wind speed. The gust speeds from the sonic
410 anemometers are true 3s averages of 0.1s-interval samples. The factor by which the cup anemometer period
411 can be adjusted to homogenise the whole period may be derived from the sonic and cup gust factors at each
412 station as the ratio 3s gust factor/5s gust factor. The ensemble average for the 36 stations in this study
413 where the cup anemometer was at 10m above ground gives the value 1.063 ± 0.011 for gust speeds >10kn.
414 The corresponding value for the 9 stations where the cup anemometer was at 26 feet (7.9m) gives the value
415 1.070 ± 0.014 . Three caveats come with these estimates: 1) the wind climate is assumed to be identical in
416 the cup and sonic periods; 2) the effect of the -0.5kn bias before 2012 is not included; and 3) the factor is
417 assumed to depend solely on the anemometer response characteristics and not be affected by the wind
418 climate. A corollary of 3) is that the value should not exhibit any geographic variation but an unexpected
419 pattern emerges when the values for the 36 stations where both cup and sonic anemometer were at 10m
420 above ground are plotted on the contour map in Figure 13. Validating this pattern and determining its cause
421 requires analysis of more ASOS stations.

422 **4. Discussion.**

423 There are significant major deficiencies in the TD6405 database:

- 424 1. Spurious wind values occurred before 2014 whenever visibility sensor checksum errors overwrote
425 the wind data.
- 426 2. Wind speeds before 2012 have a -0.5kn bias due to truncation instead of rounding the 0.1kn
427 resolution values. Overall statistics are correctable, but not individual values.
- 428 3. Gust values in the cup anemometer period at each station are nominally 5s gusts (but this varies
429 with wind speed), while gust values after sonic installation are true 3s averages. The 5s gusts can
430 be adjusted by the estimated 3s/5s gust ratios.
- 431 4. Mean and gust calms are suppressed after 2013 by the action of QC Test 10. Underestimating the
432 proportion of calms distorts assessments of wind energy potential and pollution dispersal.
- 433 5. The high rate of false positives (>80%) after 2013 and the 5-minute gaps that follow fragment the
434 data timeline. This is detrimental to studies that require long continuous runs of data.
- 435 6. These false positives are biased towards sudden gusts in light winds. This disproportionately censors
436 thunderstorm downbursts and other short-duration thermal events of interest.
- 437 7. The false positives introduce gaps into continuous data that mimic bird-perching events. The
438 apparent increase in BGGs after 2013 in Figure 11(i)-(l) is an anomaly caused by the proposed

439 replacement algorithm of §2.5 interpreting the gusts at head and tail of the gaps as BGGs. Before
440 2014, where Test 10 has not been applied, this does not happen.

441 Not all these deficiencies can be adequately addressed, limiting the usefulness of the TD6405 database in
442 Wind Engineering studies. The extent of the deficit in calms after 2013 is clear, but the extent of
443 thunderstorm downburst censoring requires assessing by further studies, using the methodologies of [1]
444 and/or [17].

445 The principal rationale for ASOS data at 1-minute intervals is for immediate transmission of current
446 weather conditions to aircraft, and therefore the QC algorithm must operate in real time. However, this
447 study shows that Test 10, in use since 2013, is not fit for purpose in the Wind Engineering context. Test 10
448 is indiscriminate and over-aggressive in targeting BGGs, missing many and generating around 4 false
449 positives for each genuine one that is found. NOAA seem to acknowledge the QC algorithm is deficient
450 with the comment [8] “*Show me a perfect algorithm, and I will show you an algorithm that does nothing*”
451 which has echoes of the US idiom “*Good enough for government work*”. Perfection can never be achieved,
452 but better always can – and in this case, substantially better is needed. The fixed thresholds are not suitable
453 to a wind climate that varies geographically, and they can be avoided. The sonic anemometers raise an
454 error flag when the acoustic path is blocked and drops it when the path is restored. The replacement
455 algorithm of §2.6 works 5 minutes behind real time but, by using this flag, the delay could be reduced to
456 under 1 minute.

457 While the real-time QC may be justified for the primary use of the ASOS data, this constraint should not be
458 applied to the archived data. The raw ASOS data is retained for 14 hours [8], so that retrospective QC
459 could be applied to that. Also, the raw data could be passed to the TD6405 archive unchanged, ensuring
460 that no information is lost, so new and improved QC retrospective procedures can be applied as they are
461 developed. The 1-minute interval data in TD6405 occupy 60 times the space per station than the hourly
462 SYNOP/METAR data in the NCEI TD3505 global database. Archiving the raw 5s-interval data would
463 require only 12 times more space than TD6405 – indeed, why stop there? Archiving the raw 1s data at
464 0.1kn resolution requires only the same 60 factor on space as the change from TD3505 to TD6405 and
465 would extend studies to encompass virtually the whole of the micrometeorological range.

466 **Conclusions**

- 467 • The TD6405 database of 1-minute interval wind data remains a most valuable resource for Wind
468 Engineering studies but is compromised by significant defects.
- 469 • Many of the defects occurring before 2013 are correctable by the methods described in this paper.

- 470 • The defects from 2013 onwards caused by implementation of QC Test 10 are not correctable
471 because valid data has been permanently lost by censoring.
472 • From 2013 onwards, the incidence of valid calms is suppressed by Test 10.
473 • From 2013 onwards, thunderstorm downbursts and other transient thermal events are censored in
474 light winds.
475 • QC Test 10 should be replaced by an algorithm that properly targets bird-generated gusts without
476 generating false positives. Until this is done, valuable data continues to be distorted or lost.

477 **Supplementary information**

478 The R scripts that will allow this study to be replicated for any ASOS station in the TD6405 database are
479 available from the Mendeley database at URL: *to be supplied*.

480 **Funding**

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

483 **References**

- 484 [1] Chen G, Lombardo FT. An automated classification method of thunderstorm and non-thunderstorm
485 wind data based on a convolutional neural network. J. Wind Eng. Ind. Aerodyn. 2020, 207, 104407.
486 <https://doi.org/10.1016/j.jweia.2020.104407>.
- 487 [2] NCDC. Data documentation for Data Set 6405: ASOS surface 1-minute, page 1 data. National Climatic
488 Data Center, Asheville NC, USA, July 2006, pp 5. Available at:
489 <https://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td6405.pdf>.
- 490 [3] NOAA. ASOS Product Improvement: Master implementation plan, NOAA, January 31, 2003.
491 Available at: <https://www.weather.gov/media/asos/ASOS%20Implementation/masterplan0131.pdf>.
- 492 [4] NOAA. ASOS Product Improvement Implementation Plan for Ice Free Wind, NOAA, May 15, 2003.
493 Available at: <https://www.weather.gov/media/asos/ASOS%20Implementation/IFW030515A.pdf>.
- 494 [5] Schubotz K, Dorschel H, Lanzinger E, Theel M. Examination of different methods of bird repelling
495 techniques for ultrasonic anemometers, German Meteorological Service, 2014. Available at:
496 [https://library.wmo.int/pmb_ged/iom_116_en/Session%203/P3_21_Schubotz_Intercomp_birdrepellingtech
497 niques_ultrasonic_anemometers.pdf](https://library.wmo.int/pmb_ged/iom_116_en/Session%203/P3_21_Schubotz_Intercomp_birdrepellingtechniques_ultrasonic_anemometers.pdf).

498 [6] Avery ML, Genchi AC. Avian perching deterrents on ultrasonic at airport windshear alert systems,
499 United States Department of Agriculture, Wildlife Services, September 2004. Available at:
500 https://digitalcommons.unl.edu/icwdm_usdanwrc/74.

501 [7] Schwartz J, Kays T. Bird Deterrence at Low Level Windshear Alert System (LLWAS) Poles, 2001.
502 2001 Bird Strike Committee-USA/Canada, Third Joint Annual Meeting, Calgary, AB. 20. Available at:
503 <https://digitalcommons.unl.edu/birdstrike2001/20>.

504 [8] NOAA. Primer for the ASOS Software Version 3.10 Ice Free Wind Sensor quality control algorithm.
505 NOAA, July 24, 2013. Available at:
506 https://www.weather.gov/media/asos/ASOS%20Implementation/IFWS%20QC%20Algorithm_primer.pdf.

507 [9] N J Cook. Review of errors in archived wind data. *Weather*, 69 (2014), 3, 72-81.
508 <https://doi.org/10.1002/wea.2148>.

509 [10] N J Cook. Detecting artefacts in analyses of extreme wind speeds. *Wind and Structures*, 19 (2014),
510 271-294. <https://doi.org/10.12989/was.2014.19.3.271>.

511 [11] Wang J, Quan Y, Gu M. Estimation of directional design wind speeds via multiple population genetic
512 algorithm. *J. Wind Eng. Ind. Aerodyn.* 2021, 210, 104534. <https://doi.org/10.1016/j.jweia.2021.104534>.

513 [12] Zang X, Chen X. Assessing probabilistic wind load effects via a multivariate extreme wind speed
514 model: A unified framework to consider directionality and uncertainty. *J. Wind Eng. Ind. Aerodyn.* 2015,
515 147, 30-42. <https://doi.org/10.1016/j.jweia.2015.09.002>.

516 [13] NOAA. Automated Surface Observing System: Ice Free Wind Problem Resolution Project Plan.
517 NOAA, January 2008. Available at:
518 [https://www.weather.gov/media/asos/ASOS%20Implementation/ASOS%20IFW%20Problem%20Resoluti
519 on%20Project%20Plan.pdf](https://www.weather.gov/media/asos/ASOS%20Implementation/ASOS%20IFW%20Problem%20Resolution%20Project%20Plan.pdf).

520 [14] NOAA. Automated Surface Observing System: Release Note Software version 3.07. NOAA, 7
521 November 2012. Available at:
522 https://www.weather.gov/media/asos/ASOS%20Implementation/release_notes_307_draft.pdf.

523 [15] NOAA. Investigation of Ice Free Wind Sensor Problems during December 2007 Ice Storm: Interim
524 Report: January 18, 2008. Available at:
525 https://www.weather.gov/media/asos/ASOS%20Implementation/IFWS_iceup_intrim_rpt_011808.pdf.

526 [16] NOAA. Automated Surface Observing System: Release Note Software version 2.79. NOAA, 13
527 February 2004. Available at:
528 https://www.weather.gov/media/asos/ASOS%20Implementation/release_notes.279_final.pdf.

529 [17] De Gaetano P, Repetto MP, Repetto T, Solari G. Separation and classification of extreme wind events
530 from anemometric records. J. Wind Eng. Ind. Aerodyn. 2014, 126, 132-143.
531 <http://dx.doi.org/10.1016/j.jweia.2014.01.006>.

532

533 **Appendix. ICAO Airport codes and locations of cited stations**

534 KDEN Denver International Airport, CO, at 39.8328°N, 104.6575°W
535 KDCA Washington Ronald Reagan National Airport, VA, at 38.84720°N, 77.03454°W
536 KIAD Washington Dulles International Airport, VA, at 38.93486°N, 77.44728°W
537 KLAS Las Vegas International Airport, NV, at 36.071899°N, 115.163435°W
538 KLAX Los Angeles International Airport, CA, at 33.938201°N, 118.386603°W
539 KMYV Marysville, Yuba County Airport, CA, at 39.102071°N, 121.568833°W
540 KSEA Seattle-Tacoma International Airport, WA, at 47.444680°N, 122.314415°W
541 KSNA Santa Ana, Orange County Airport, CA, at 33.679809°N, 117.867405°W
542 KTPA Tampa International Airport, FL, at 27.963336°N, 82.540006°W
543 KTUS Tucson International Airport, AZ, at 32.131542°N, 110.956371°W

544 Latitudes and longitudes are to the centre of the anemometer mast. All stations used in this study are listed
545 in the Supplementary information file GustRatio.csv.

546

547

548

549

Table 1. Duplicated date-times at KDCA Washington Reagan Airport on 20 February 2012.

1st duplicated date/time					2nd duplicated date/time					Next day at same time				
UTC	Drn	kts	gDrn	gkts	UTC	Drn	kts	gDrn	gkts	UTC	Drn	kts	gDrn	gkts
00:00:00	333	8	333	9	00:00:00	52	9	36	11	00:00:00	333	8	333	9
00:01:00	332	8	327	8	00:01:00	49	9	48	11	00:01:00	332	8	327	8
00:02:00	329	7	332	8	00:02:00	49	9	59	9	00:02:00	329	7	332	8
00:03:00	328	7	326	8	00:03:00	55	9	56	11	00:03:00	328	7	326	8
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
01:13:00	46	1	24	1	01:13:00	79	6	95	8	01:13:00	46	1	24	1
01:14:00	47	1	37	2	01:14:00	85	6	59	8	01:14:00	47	1	37	2
01:15:00	37	2	32	2	01:15:00	65	5	72	7	01:15:00	37	2	32	2
01:16:00	44	2	68	2	01:16:00	61	4	89	6	01:16:00	44	2	68	2
01:17:00	66	2	65	2	01:17:00	79	6	100	11
01:18:00	72	2	54	3	01:18:00	97	9	105	10
01:19:00	64	2	59	3	01:19:00	107	10	109	12
01:20:00	61	2	65	3	01:20:00	110	9	99	12
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
06:50:00	257	2	259	2	06:50:00	10	4	349	6
06:51:00	255	2	254	2	06:51:00	10	5	11	9
06:52:00	256	2	263	2	06:52:00	5	7	358	10
06:53:00	258	2	256	2	06:53:00	2	9	357	13
06:54:00	6	12	3	18	06:54:00	256	2	254	2
06:55:00	9	13	5	14	06:55:00	253	2	247	2
06:56:00	6	13	355	16	06:56:00	248	2	241	2
06:57:00	359	14	354	17	06:57:00	241	2	240	2
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
08:08:00	5	19	10	23	08:08:00	142	1	135	1
08:09:00	3	19	10	22	08:09:00	134	1	132	1
08:10:00	5	19	9	22	08:10:00	126	1	119	1
08:11:00	4	18	4	19	08:11:00	118	1	104	1
08:12:00	2	16	8	17	08:12:00	111	1	104	1	08:12:00	111	1	104	1
08:13:00	6	16	17	20	08:13:00	103	1	92	1	08:13:00	103	1	92	1
08:14:00	8	17	11	20	08:14:00	92	1	75	2	08:14:00	92	1	75	2
08:15:00	6	15	11	17	08:15:00	82	1	80	2	08:15:00	82	1	80	2
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
11:55:00	355	11	352	14	11:55:00	214	3	209	3	11:55:00	214	3	209	3
11:56:00	350	13	352	17	11:56:00	211	3	220	3	11:56:00	211	3	220	3
11:57:00	349	15	351	19	11:57:00	215	3	216	3	11:57:00	215	3	216	3
11:58:00	355	16	5	18	11:58:00	215	2	212	3	11:58:00	215	2	212	3
11:59:00	352	16	342	19	11:59:00	214	2	219	3	11:59:00	214	2	219	3

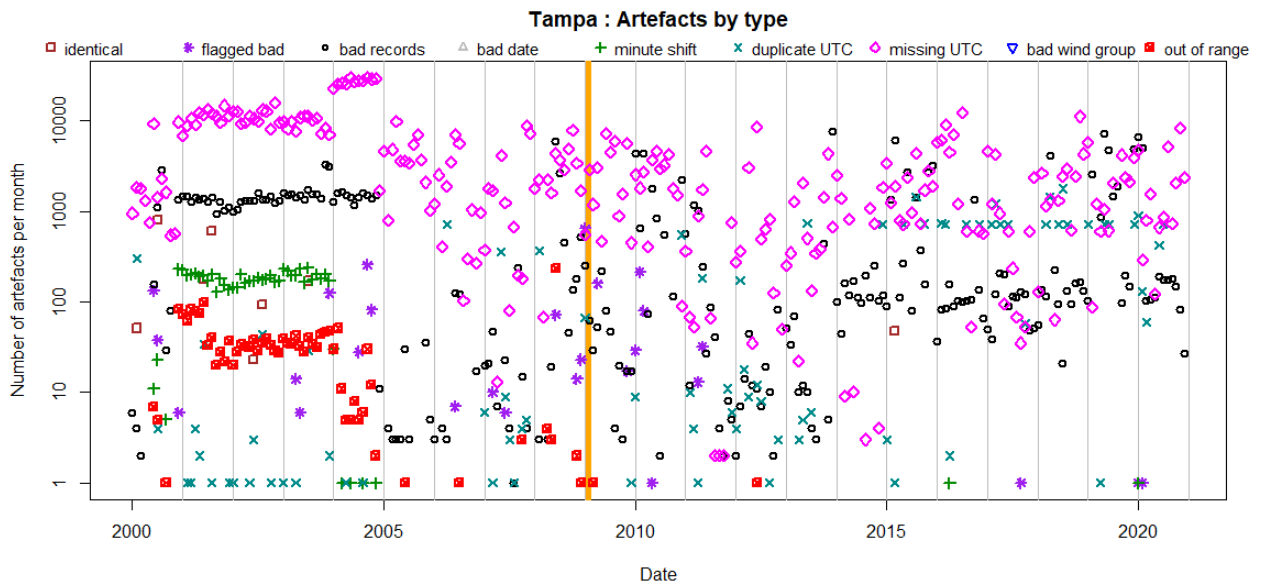
551 **Table 2.** Sonic anemometer real-time QC algorithm [8] test criteria.

Test	Error criteria
1	Sensor did not respond when polled by ASOS
2	Wind Speed out of range (less than 0 or more than 165)
3	Averaging Times wrong (Average = 5 seconds, Peak = 3 seconds)
4	Sonic Pass/Fail flag is set to "F" (Fail)
5	5-second average speed exceeds 3-second peak speed by 1 or more knots
6	5-sec average < 12kn AND 3-sec peak > 30kn
7	5-sec average speed >= 12 kn AND average direction – peak direction exceeds 30°
8	Signal Quality is less than 79
9	5-sec average >= 12kn AND 3-sec peak is greater than 2.5 times the 2 min average wind speed
10	2014→2017 : The 2 min average =< 6kn AND 3-sec peak >6kn AND 3-sec peak is greater than 2.5 times the 2 min average wind speed. 2018→ : The 2 min average =< 6kn AND 3-sec peak >13kn AND 3-sec peak is greater than 2.5 times the 2 min average wind speed.

553 **Table 3.** Performance statistics for 10 ASOS sites.

EXTRACTION	KDEN Denver		KDCA Reagan		KIAD Dulles		KLAS Las Vegas		KLAX Los Angeles	
Nominal minutes	11045861		11045991		11045921		11045844		11045804	
Processed records	10130670	91.71%	9201448	83.30%	10389912	94.06%	10099171	91.43%	10290976	93.17%
Identical records	2236	0.02%	7698	0.07%	2445	0.02%	7314	0.07%	1053	0.01%
Undecipherable records	65584	0.59%	77119	0.70%	30021	0.27%	26378	0.24%	1984	0.02%
Single minute shifts	6	0.00%	8	0.00%	0	0.00%	9	0.00%	242	0.00%
Duplicate UTC	9634	0.09%	6551	0.06%	4453	0.04%	6806	0.06%	15936	0.14%
Missing UTC	683979	6.19%	1710953	15.49%	552965	5.01%	763297	6.91%	668516	6.05%
Ambiguous wind values	0	0.00%	35159	0.32%	0	0.00%	0	0.00%	0	0.00%
Good records	10062849	91.10%	9116630	82.53%	10357445	93.77%	10065479	91.12%	10256537	92.85%
Out of range	22	0.00%	45	0.00%	22	0.00%	15	0.00%	48	0.00%
Valid records	10062655	91.10%	9116530	82.53%	10357403	93.77%	10065449	91.12%	10256439	92.85%
CLEANING	KDEN Denver		KDCA Reagan		KIAD Dulles		KLAS Las Vegas		KLAX Los Angeles	
Nominal minutes	11045861		11045991		11045921		11045844		11045804	
Raw observations	10062655	91.10%	9116530	82.53%	10357403	93.77%	10065449	91.12%	10256439	92.85%
Mean calms	34092	0.31%	90573	0.82%	278154	2.52%	70711	0.64%	117961	1.07%
Gust calms	22695	0.21%	64282	0.58%	179461	1.62%	34879	0.32%	65101	0.59%
Raw max 2min mean	111		94		113		49		89	
Raw max gust	119		125		125		67		88	
Identical observations	27	0.00%	81	0.00%	1103	0.01%	217	0.00%	846	0.01%
Unique observations	10062628	91.10%	9116449	82.53%	10356300	93.76%	10065232	91.12%	10255593	92.85%
Duplicate UTCs	9236	0.08%	4472	0.04%	3038	0.03%	6478	0.06%	14823	0.13%
All pair runs	87		726		18		20		119	
Runs > 10 pairs	25		10		11		14		41	
After pairs resolved	10059093	91.07%	9114544	82.51%	10354783	93.74%	10063972	91.11%	10251362	92.81%
Landing pass 1	377		500		422		159		256	
Take-off pass 1	1024		582		424		771		280	
Unique bird events pass 1	1384		1048		819		927		524	
Landing pass 2	135		282		178		51		104	
Take-off pass 2	500		361		214		348		118	
Unique bird events pass 2	628		586		381		339		213	
Bird gusts removed	2012	0.018%	1634	0.015%	1200	0.011%	1326	0.012%	737	0.007%
After birds	10057081	91.05%	9112910	82.50%	10353583	93.73%	10062646	91.10%	10250625	92.80%
Short runs	819		7742		540		537		3551	
Short run observations	1813		17685		2141		2527		13295	
After short runs removed	10055268	91.03%	9095225	82.34%	10351442	93.71%	10060119	91.08%	10237330	92.68%
Single spikes	28		12		77		4		3	
Pair spikes	3		1		31		0		0	
Triplet spikes	2		6		16		0		0	
Repeated gust pairs >=50kn	24		3		11		3		2	
Repeated gust triplets >=40kn	63		5		7		5		3	
Unique dirty observations	266	0.002%	39	0.000%	194	0.002%	24	0.000%	12	0.000%
Residual short runs	25		8		41		0		2	
Residual short run observations	52		29		142		0		9	
Clean observations	10054950	91.03%	9095157	82.34%	10351106	93.71%	10060095	91.08%	10237309	92.68%
Clean max 2min mean	76		45		51		49		40	
Clean max gust	85		64		115		61		59	

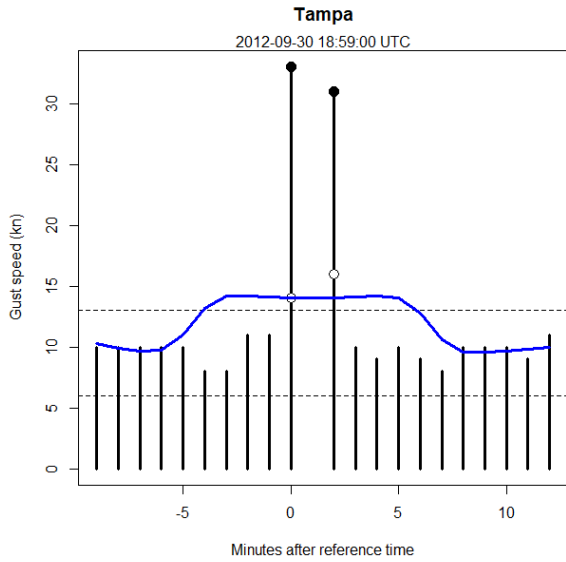
EXTRACTION	KMYV Marysville		KSEA Seattle		KSNA Santa Ana		KTPA Tampa		KTUS Tucson	
Nominal minutes	8318946		11045753		8326934		11046000		11045859	
Processed records	7728766	92.91%	9969606	90.26%	7535317	90.49%	9854870	89.22%	10076180	91.22%
Identical records	36	0.00%	2391	0.02%	3	0.00%	1971	0.02%	1594	0.01%
Undecipherable records	80091	0.96%	69901	0.63%	19579	0.24%	177554	1.61%	119515	1.08%
Single minute shifts	1	0.00%	7171	0.06%	42	0.00%	6928	0.06%	1062	0.01%
Duplicate UTC	8061	0.10%	5190	0.05%	17406	0.21%	24119	0.22%	6068	0.05%
Missing UTC	514714	6.19%	925276	8.38%	716624	8.61%	1115426	10.10%	759175	6.87%
Ambiguous wind values	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Good records	7648639	91.94%	9897314	89.60%	7515735	90.26%	9675345	87.59%	9955071	90.12%
Out of range	66	0.00%	2720	0.02%	21	0.00%	1988	0.02%	374	0.00%
Valid records	7648560	91.94%	9894569	89.58%	7515443	90.25%	9672856	87.57%	9954652	90.12%
CLEANING	KMYV Marysville		KSEA Seattle		KSNA Santa Ana		KTPA Tampa		KTUS Tucson	
Nominal minutes	8318946		11045753		8326934		11046000		11045859	
Raw observations	7648560	91.94%	9894569	89.58%	7515443	90.25%	9672856	87.57%	9954652	90.12%
Mean calms	143591	1.73%	129691	1.17%	288516	3.46%	120872	1.09%	101705	0.92%
Gust calms	78621	0.95%	94760	0.86%	162244	1.95%	86456	0.78%	73917	0.67%
Raw max 2min mean	124		95		120		124		120	
Raw max gust	125		125		120		125		123	
Identical observations	26	0.00%	52	0.00%	1633	0.02%	1398	0.01%	29	0.00%
Unique observations	7648534	91.94%	9894517	89.58%	7513807	90.23%	9671458	87.56%	9954623	90.12%
Duplicate UTCs	7848	0.09%	4896	0.04%	10352	0.12%	20987	0.19%	5790	0.05%
All pair runs	32		44		534		96		13	
Runs > 10 pairs	18		11		24		41		9	
After pairs resolved	7644582	91.89%	9890670	89.54%	7507881	90.16%	9667336	87.52%	9952680	90.10%
Landing pass 1	1342		1338		258		1899		911	
Take-off pass 1	896		1150		327		1908		1977	
Unique bird events pass 1	2083		2416		563		3736		2839	
Landing pass 2	571		647		122		851		284	
Take-off pass 2	407		776		173		1028		921	
Unique bird events pass 2	917		1255		288		1634		1159	
Bird gusts removed	3000	0.036%	3671	0.033%	851	0.010%	5370	0.049%	3998	0.036%
After birds	7641582	91.86%	9886999	89.51%	7507030	90.15%	9661966	87.47%	9948682	90.07%
Short runs	1609		74242		1171		104620		15688	
Short run observations	6073		276368		3583		360899		58177	
After short runs removed	7635509	91.78%	9610631	87.01%	7503447	90.11%	9301077	84.20%	9890505	89.54%
Single spikes	35		9		1		91		5	
Pair spikes	7		0		0		8		0	
Triplet spikes	3		0		0		4		0	
Repeated gust pairs >=50kn	2		0		2		2		0	
Repeated gust triplets >=40kn	6		1		2		0		2	
Unique dirty observations	72	0.001%	12	0.000%	8	0.000%	117	0.001%	11	0.000%
Residual short runs	16		1		0		61		3	
Residual short run observations	51		2		0		287		19	
Clean observations	7635386	91.78%	9610617	87.01%	7503439	90.11%	9300673	84.20%	9890475	89.54%
Clean max 2min mean	45		45		49		45		43	
Clean max gust	65		61		51		66		73	



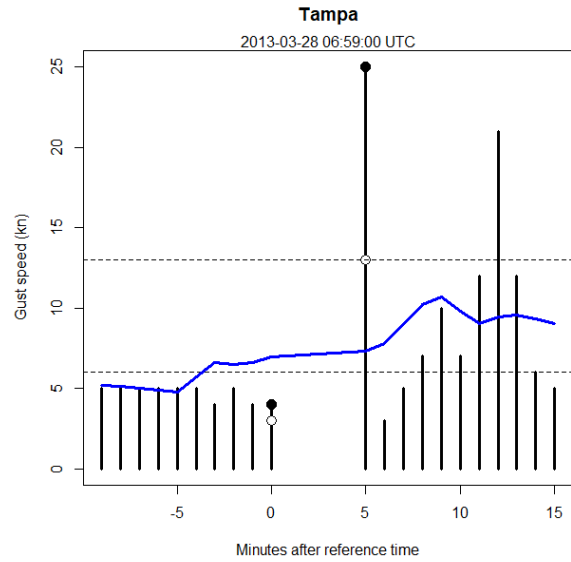
556

557 **Figure 1.** Number per month of each artefact type found by reading and parsing the TD6405 record for
 558 KTPA Tampa International, 2000-2020. The thick vertical line indicates the date of change from cup to sonic
 559 anemometer.

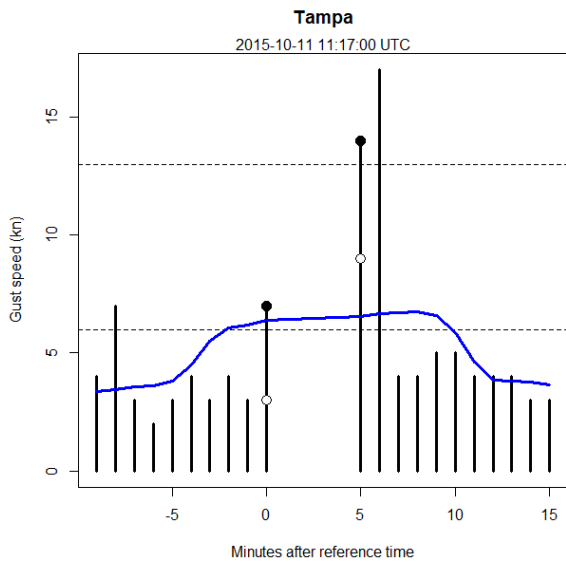
560



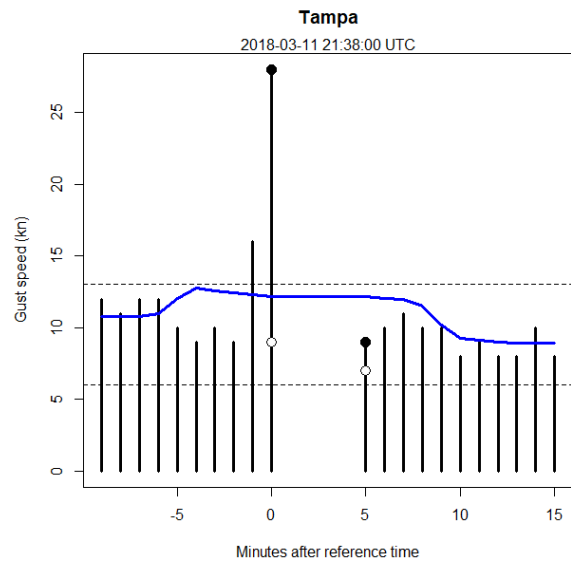
(a)



(b)

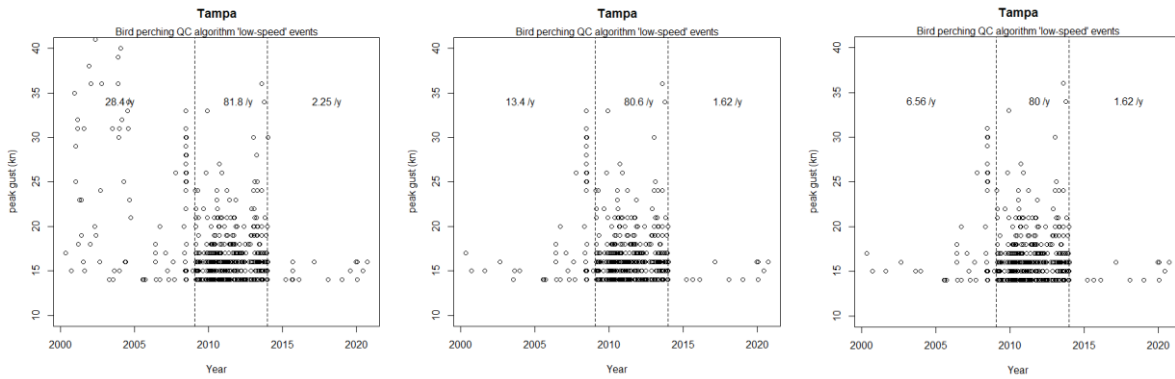


(c)



(d)

561 **Figure 2.** Presumed bird perching events at KTPA. Bars – 3s gust values; open and black circles – 2-minute
 562 mean and 3s gust, respectively, immediately before and after gap in data; continuous line – running 10-minute mean
 563 of 3s gusts, discounting gap; dashed horizontal lines – 6kn and 13kn thresholds for the ASOS QC algorithm. Note (c)
 564 and (d) occur after the 2014 upgrade of the QC algorithm.
 565

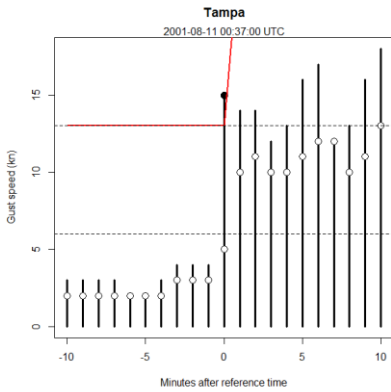


(a) Raw data

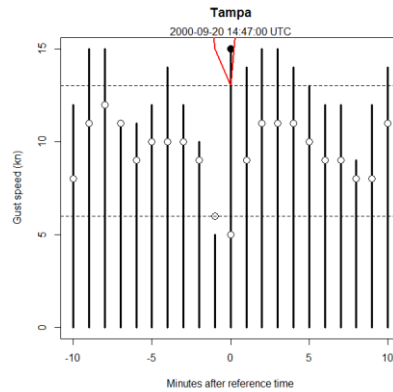
(b) De-birded data

(c) Cleaned data

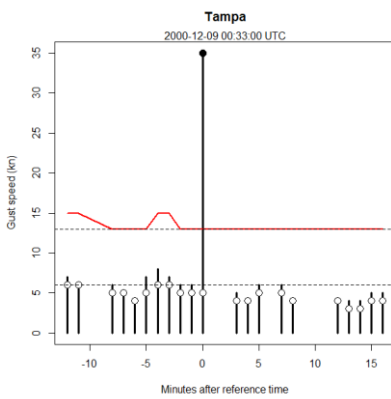
566 **Figure 3.** Performance of QC Test 10 applied to TD6405 data at KTPA. (a) “raw” data corrected for
 567 formatting errors and duplicated date-times only; (b) after “de-birding” by the proposed replacement algorithm
 568 in §2.5; and (c) fully “cleaned” data using the [10] methodology in §2.6.
 569



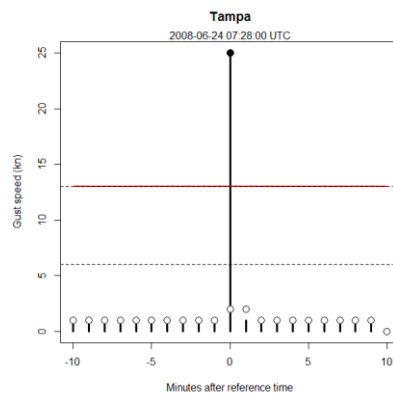
(a) Sudden gust in light winds



(b) Lull to 6kn in strong winds



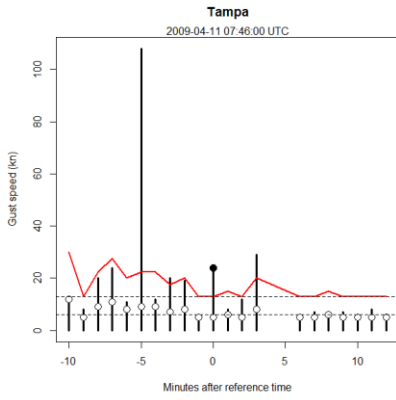
(c) Spike at head of gap in data



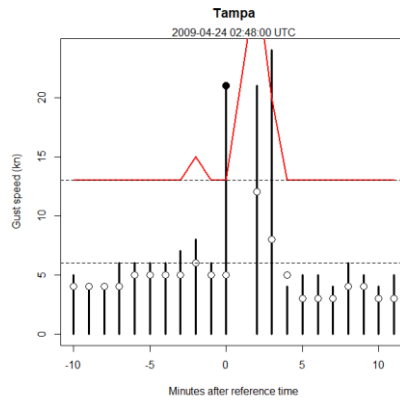
(d) Spike at low wind speeds

570 **Figure 4.** Examples of gusts failing QC Test 10 in the Belfort cup anemometer period at KTPA. Bars –gust
 571 values; open circles – 2-minute mean; dashed horizontal lines – 6kn maximum mean and 13kn minimum gust
 572 thresholds; thin (red) line – effective minimum gust threshold, where running 2.5 gust factor >13kn. The gusts failing
 573 Test 10 are shown by the black circles.

574

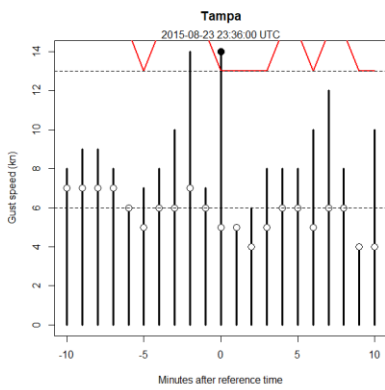


(a) Spike & head of gap

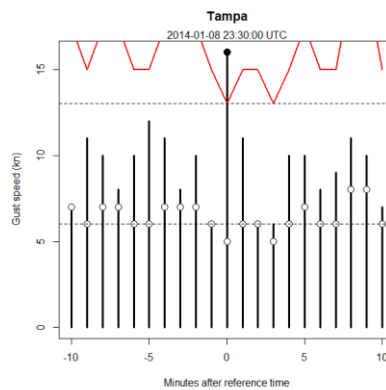


(b) Tail of gap

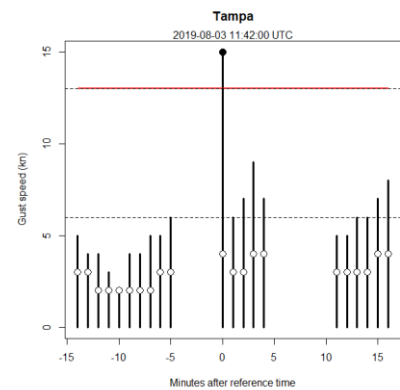
575 **Figure 5.** False negatives for QC Test 10 in the sonic anemometer period at KTPA **before** adoption of Test
 576 10. Key same as Figure 4.



(a) Sudden gust in light winds

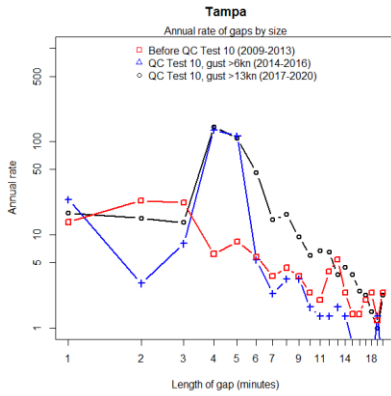


(b) Lull to 6kn in strong winds



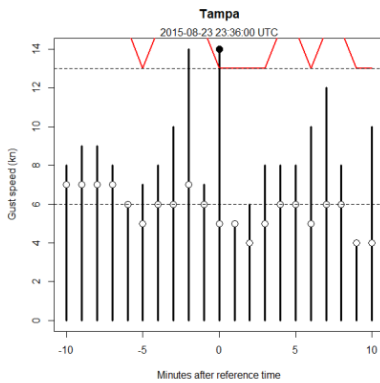
(c) Spike at tail of gap in data

577 **Figure 6.** Examples of gusts failing QC Test 10 in the sonic anemometer period at KTPA **after** adoption of
 578 Test 10. Key same as Figure 4.
 579



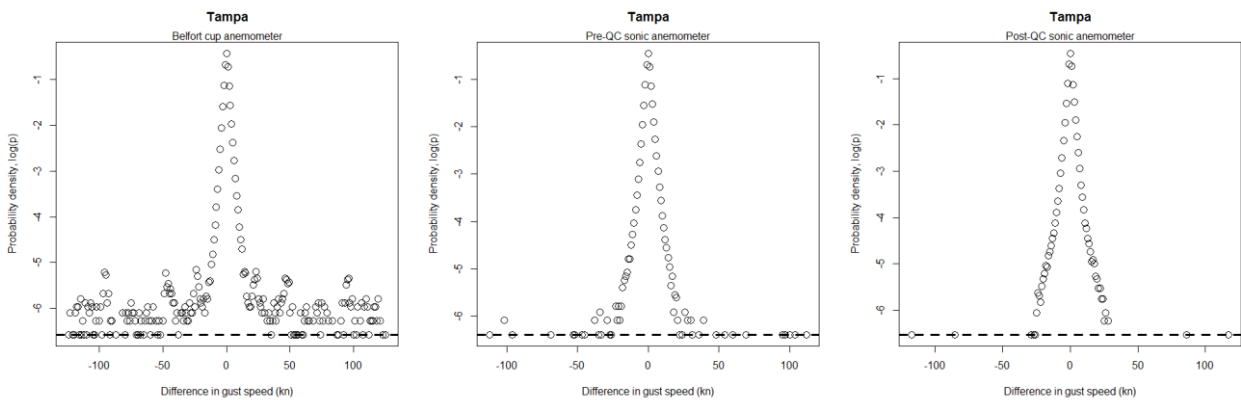
580

581 **Figure 7.** Annual rate of data gaps between 1s and 20s duration at KTPA for the sonic anemometer
 582 periods: before Test 10; Test 10, gusts >6kn; Test 10, gusts >13kn.



583

584 **Figure 8.** Example of Test 10 failure after applying the replacement bird-perching algorithm at KTPA.

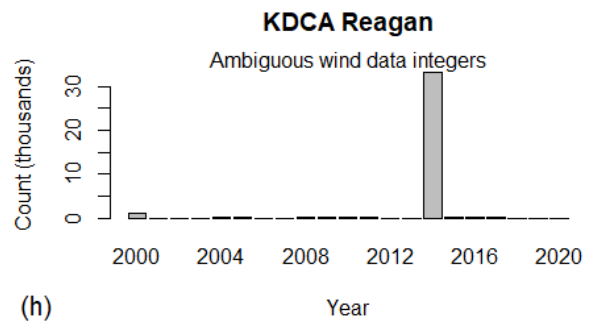
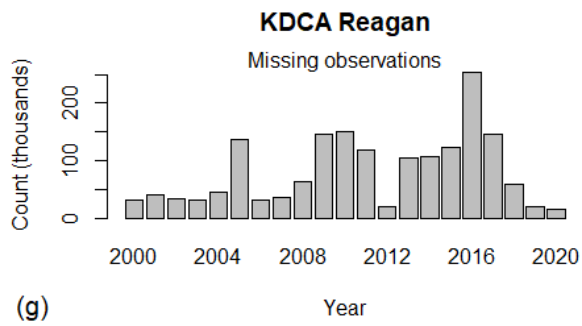
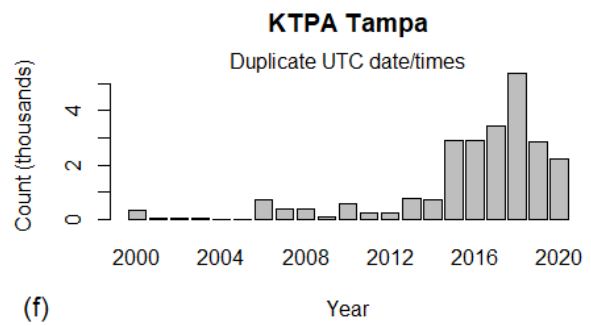
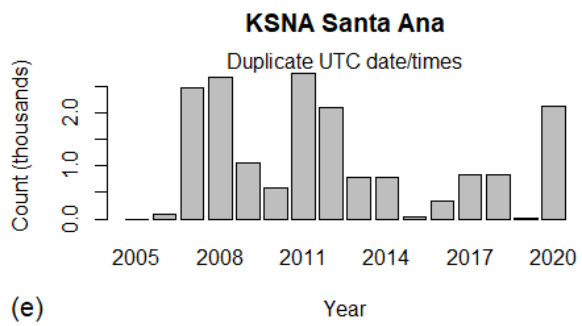
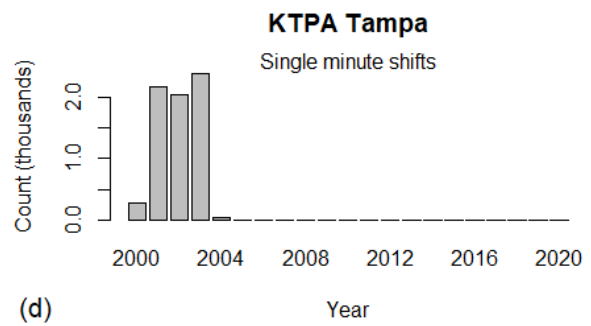
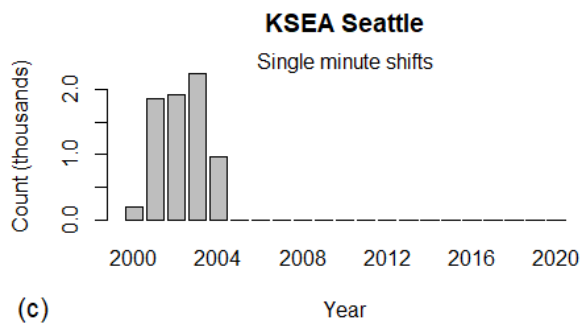
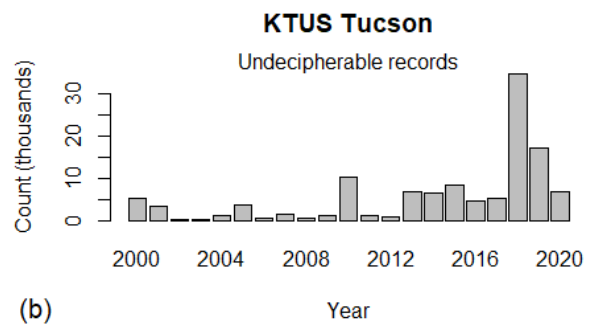
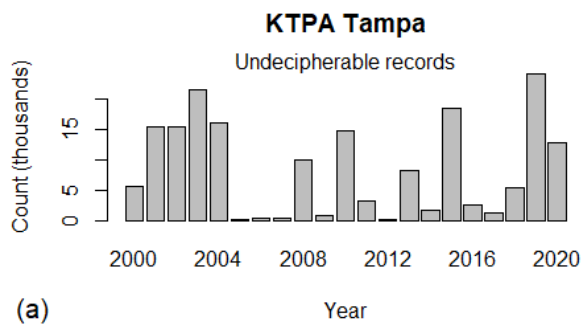


(a) Cup anemometer

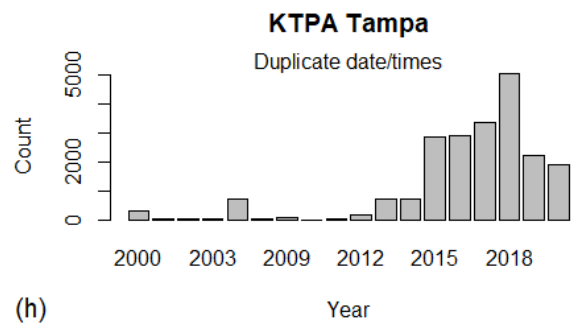
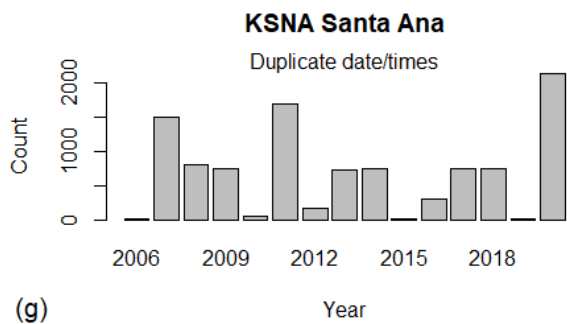
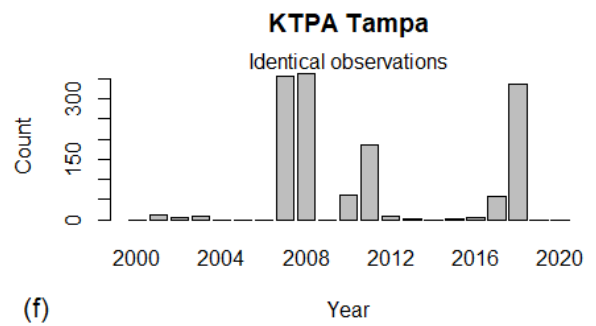
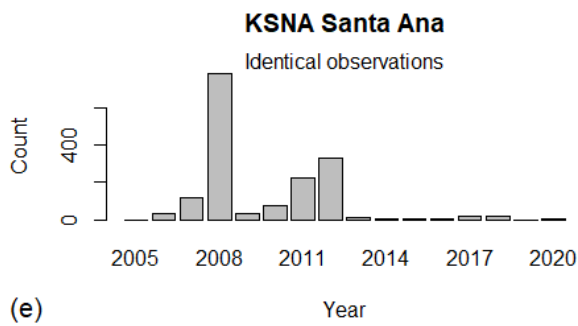
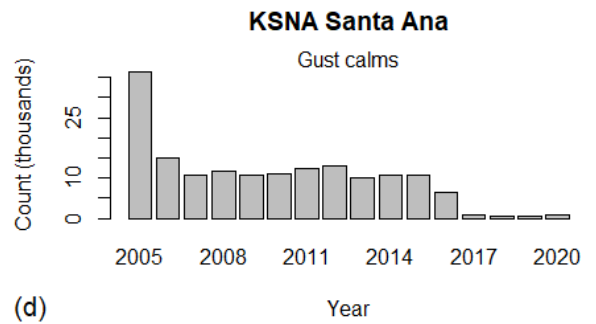
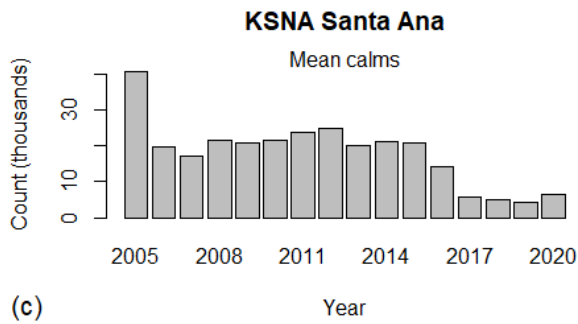
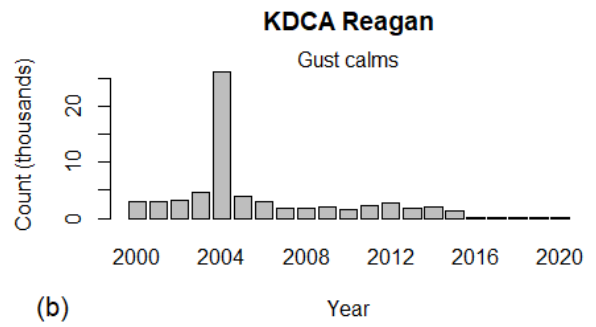
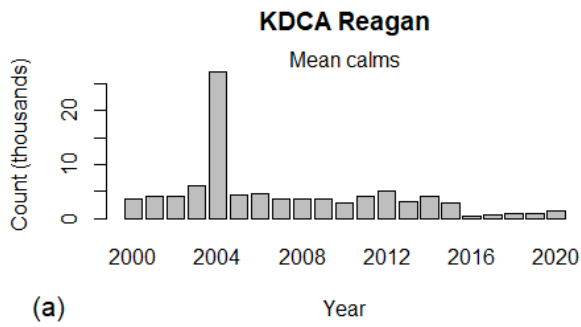
(b) Sonic anemometer before Test 10

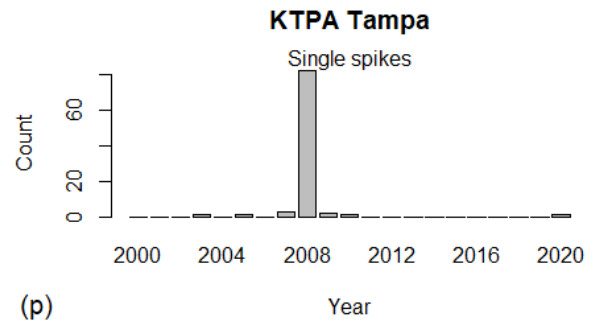
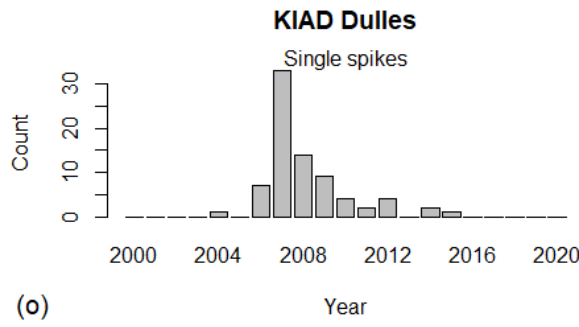
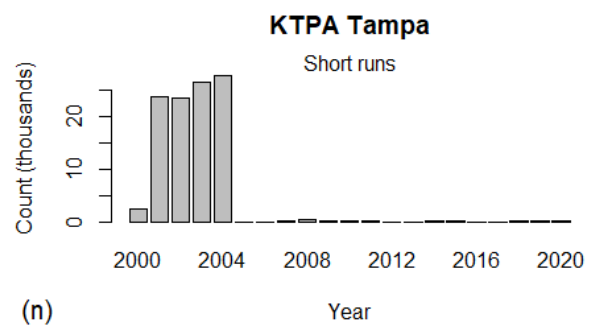
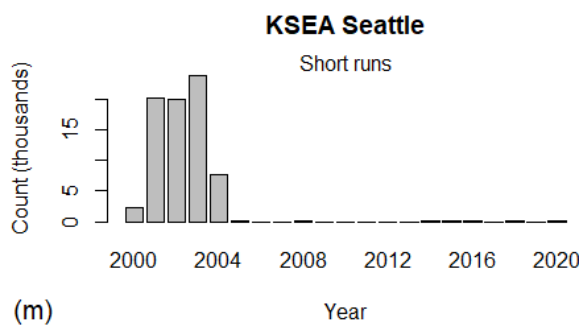
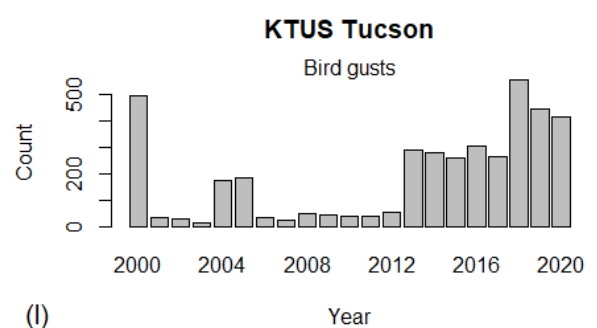
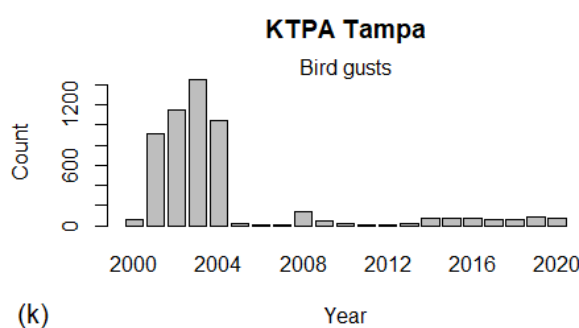
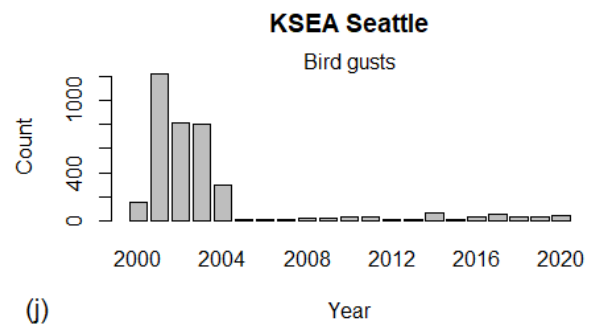
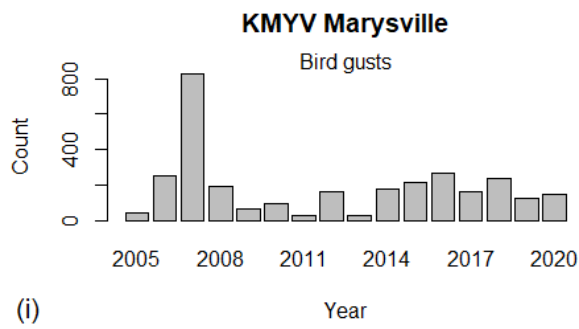
(c) Sonic anemometer after Test 10

585 **Figure 9.** Probability density of first differences in gust speed (change of gust speed between successive
 586 minutes) at KTPA for the three characteristic time periods. The dashed line corresponds to the expectation of a
 587 single event in the whole record.

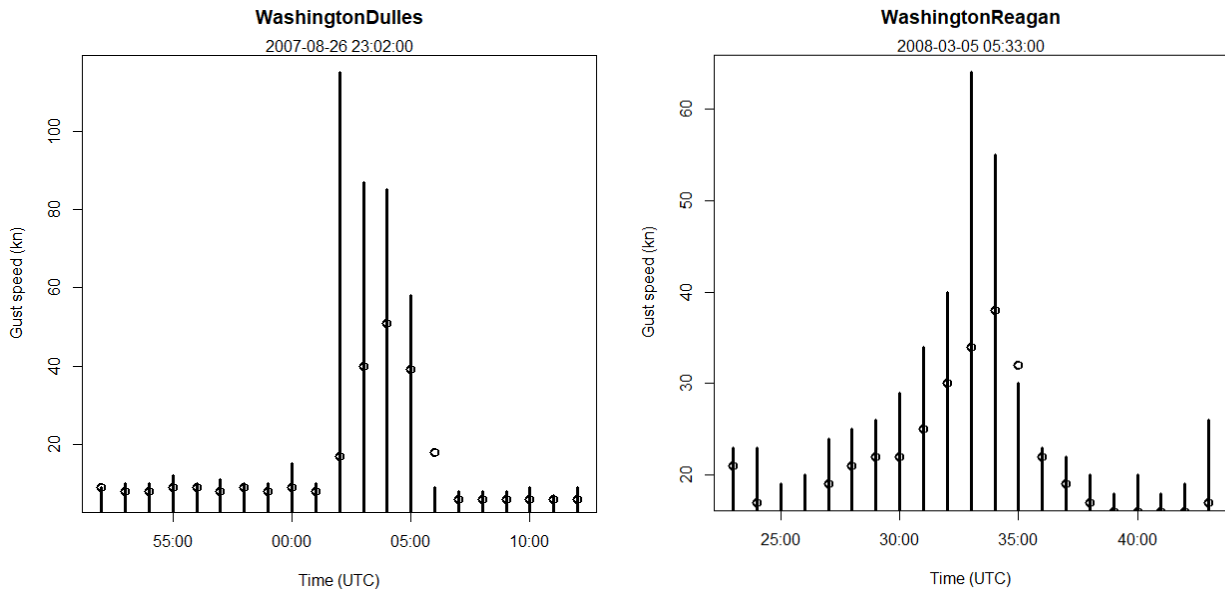


588 **Figure 10.** Histograms by year of dominant TD6405 artefacts found in extraction phase.



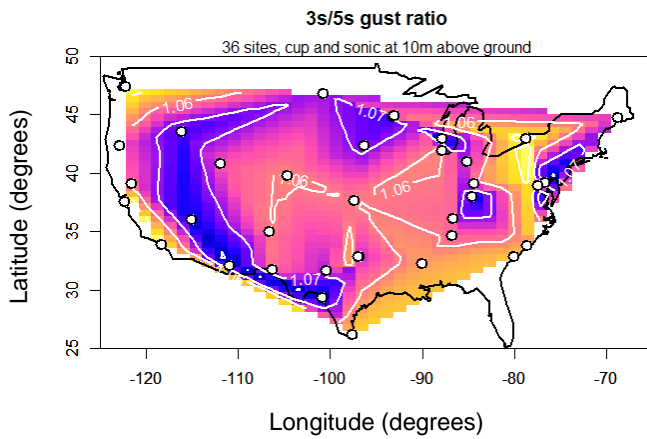


590 **Figure 11.** Histograms by year of dominant TD6405 artefacts found in cleaning phase.



591 **Figure 12.** Maximum gust event recorded at KIAD Dulles (left) and KDCA Reagan (right) Airports.

592



593 **Figure 13.** Distribution of the 3s/5s gust ratio across CONUS. The open circles show the locations of the 36
 594 stations where cup anemometers were at 10m above ground, with linear interpolation in between.

595