This document is part of the Data Supplement to which the IGSC 2019 paper "A benchmark suite for control algorithms of retractable wind-energy harvesters" refers.

1.0 HOW THE WINDSPEED DATA WAS DERIVED

1.1 FILTERING

We are using files in the Automated Surface Observing System (ASOS) data set DSI 6405 [2]. Some lines in those files contain non-windspeed-data where we expect windspeeds. To ignore those lines, we filtered out each line that

1. Do'es not have a number in its windspeed field,
2. Has a number or other character that is not a space in the column that immediately follows its windspeed field,
3. Has an identical local timestamp to its previous line,
4. Does not have a number in the least significant place of its wind heading field, or
5. Does not have a space that precedes its windspeed field.

The results of our filtering are shown in Table 2. The table shows that in all cases less than 3% of lines were removed by our filters.

1.2 SHIFTING FIELDS

We found that the windspeed and wind heading fields changed locations between years 2004 and 2014 for most weather stations. These are the field locations we encountered for windspeed and for wind heading respectively:

Speed1: Windspeed in columns 74, 75, and 76.

Speed2: Windspeed in columns 76, 77, and 78.

Heading1: Wind heading in columns 68, 69, and 70.

Heading2: Wind heading in columns 71, 72, and 73.

Heading3: Wind heading in columns 67, 68, and 69.

We extracted the windspeeds from the shifting field locations in ASOS data to create comma separated files that would eventually become this benchmark's wind-data .csv files. The shifting field locations in the ASOS data are shown in Table 1 as a function of timestamp ranges. For example, station KPHX's data has windspeed values in the location we call "Speed2" above for all records having timestamps in the range [PHX2013053007231423, end], which is from 2013-05-30 07:23, 14:23 UTC, to the end of KPHX's 2014 data, inclusive. All of KPHX's records outside that range (i.e., the balance of the records) have windspeed values in the location we call "Speed1" above.

1.3 REMOVING ANACHRONISTIC RECORDS

We omitted anachronistic records (i.e., in the sequence of the ASOS data, records having timestamps earlier than the timestamps of records occurring earlier in the sequence). Anachronistic records were found in only six of the thirty weather stations, KBWI, KEUG, KIAH, KLGA, KORD, and KPIT, and comprised a minuscule percentage (approximately 0.0007%, which equals the approximately 1000 anachronistic records shown in Table 3 divided by the approximately 150,000,000 samples, which comprise the eleven years of minute-by-minute data for 30 stations) of the ASOS data after we had filtered it according to the method described immediately above. As shown in Table 3, of six stations having anachronistic records, KIAH had the highest percentage of anachronistic records, which was 0.006%.

1.4 HANDLING IDENTICALLY TIMESTAMPED RECORDS

We discarded all records where timestamps were repeated in the ASOS data. The percentage of runs of identically timestamped samples compared to the number of uniquely timestamped

Table 1: Ranges of timestamps where fields are in certain columns in original ASOS data

icao	Spd.1	Speed2	Hdg.1	Heading2	Heading3[a]
KPHX	bal.[b]	[PHX2013053007231423, end[c]]	bal.	[PHX2013053007231423, end]	N/A
KLAX	all	none	bal.	none	[LAX2004010117230123, LAX2006013007391539)
KSAC	all	none	bal.	none	[SAC2004010118370237, SAC2004050500070807)
KSAN	all	none	bal.	none	[SAN2004010118060206, SAN2004031809381738)
KSFO	all	none	all	none	none
KSMX	bal.	[SMX2014082706071407, end]	bal.	[SMX2014082706071407, end]	[SMX2004010100000800, SMX2004030905111311)
KDEN	bal.	[DEN2013121113272027, end]	bal.	[DEN2013121113272027, end]	[DEN2004010100000700, DEN2004082506391339)
KMCO	bal.	[MCO2014022410051505, end]	bal.	[MCO2014022410051505, end]	[MCO2004010100000500, MCO2006092000330533)
KATL	all	none	bal.	none	[ATL2004010122110311, ATL2004021802330733)
KORD	bal.	[ORD2013121213101910, end]	bal.	[ORD2013121213101910, end]	[ORD2004010100350635, ORD2004060706491249)
KCVG	all	none	bal.	none	[CVG2004010119390039, CVG2004010709381438)
KBOS	all	none	all	none	none
KBWI	all	none	bal.	none	[BWI2004010120190119, BWI2004010601170617)
KDTW	all	none	bal.	none	[DTW2004010119490049, DTW2004071405401040)
KMSP	all	none	all	none	none
KMCI	bal.	[MCI2013121208581458, end]	bal.	[MCI2013121208581458, end]	[MCI2004010100000600, MCI2004070707581358)
KSTL	all	none	bal.	none	[STL2004010100000600, STL2005042502190819)
KCLT	all	none	bal.	none	[CLT2004010119360036, CLT2004012201480648)
KLAS	bal.	[LAS2014040909481748, end]	bal.	[LAS2014040909481748, end]	[LAS2004010117140114, LAS2004031002341034)
KLGA	all	none	bal.	none	[LGA2004010100200520, LGA2004030113571857)
KCLE	bal.	[CLE2013011510351535, end]	bal.	[CLE2013011510351535, end]	none
KEUG	bal.	[EUG2013121910281828, end]	bal.	[EUG2013121910281828, end]	[EUG2004010100000800, EUG2004030904571257)
KPHL	bal.	[PHL2014022011001600, end]	bal.	[PHL2014022011001600, end]	[PHL2004010123030403, PHL2004020204580958)
KPIT	bal.	[PIT2014030710371537, end]	bal.	[PIT2014030710371537, end]	[PIT2004010100000500, PIT2005012809041404)
KDFW	bal.	[DFW2013012411311731, end]	bal.	[DFW2013012411311731, end]	none
KIAH	all	none	bal.	none	[IAH2004010100000600, IAH2004031823150515)
KSAT	bal.	[SAT2014022709261526, end]	bal.	[SAT2014022709261526, end]	[SAT2004010120240224, SAT2005082308021402)
KDCA	all	none	all	none	none
KSEA	all	none	bal.	none	[SEA2004010100000800, SEA2006031402341034)

[a] The range could include the last line in each Heading3 range (e.g., LAX2006013007391539), but our filter does not include the last line in each range. We seem to have enough data that the last line in each Heading3 range is not needed for most simulations. Thus, we did not re-filter the data to include the last line in each Heading3 range.

[b] balance of records (i.e., all records having timestamps outside the range(s) given in this row for this field

[c] i.e., end of 2014 data

Table 2: Percentage of lines in ASOS data that the filters listed in Section 1.1 discarded) ordered descendingly

icao	lines discarded (%)	icao	lines discarded (%)	icao	lines discarded (%)
KORD	2.45%	KIAH	0.37%	KMCI	0.23%
KTPA	1.17%	KMCO	0.36%	KSEA	0.22%
KSMX	0.93%	KLAX	0.33%	KSFO	0.2%
KDEN	0.69%	KPHL	0.33%	KBWI	0.17%
KPIT	0.58%	KSAC	0.31%	KCLT	0.14%
KDCA	0.52%	KPHX	0.3%	KCVG	0.13%
KEUG	0.48%	KSAT	0.28%	KBOS	0.11%
KCLE	0.43%	KATL	0.26%	KSAN	0.11%
KSTL	0.41%	KDFW	0.26%	KLAS	0.08%
KLGA	0.4%	KDTW	0.23%	KMSP	0.07%
continues above right		continues above right			

Table 3: Statistics on anachronistic records

station	No. of anachronistic records	Max. run size	No. of samples	Anachronistic records as percentage of no. of samples
KBWI	190	190	5419390	0.004%
KEUG	199	110	5205945	0.004%
KIAH	337	337	5198385	0.006%
KLGA	21	21	5185102	0.000%
KORD	246	218	5147727	0.005%
KPIT	23	23	5306440	0.000%

samples is less than one tenth of a percent (i.e., 0.09%) and, on a station-by-station basis, a maximum of less than three tenths of a percent (0.29% for KORD).

Table 4: Statistics about windspeeds between uniquely timestamped records and between uniquely timestamped records and the first records of each run of identically timestamped records

station	see legend*	mean	std. dev.	kur-tosis	skew-ness	n	mean(B)/ mean(A)
*Legend: A = statistics about normalized difference of windspeeds** in samples having unique timestamps. B = statistics about normalized difference of windspeeds** between first sample of run of samples sharing timestamps and the most recent sample having a unique timestamp B/A = quotient of group B's mean divided group's A mean **The normalized difference of windspeeds $\delta_{\text{knotsPerMinute}}$ equals the difference of windspeeds δ_{knots} over the difference between timestamps δ_{minute} (to account for timestamps that differ by more than one minute); i.e., $\delta_{\text{knotsPerMinute}} = \delta_{\text{knots}}/\delta_{\text{minute}}$.							
KATL	A	1.37E-04	0.87	3.95	0.17	5436757	
	B	7.27E-01	3.66	1.45	-0.39	11	
	B/A	5.29E+03					5.29E+03
KBOS	A	2.92E-05	0.90	2.45	0.09	5.46E+06	
	B	5.00E-01	3.87	4.57	1.62	10	
	B/A	1.71E+04					1.71E+04
KBWI	A	-2.55E-05	0.86	4.57	0.17	5394105	
	B	2.89E-01	3.49	9.10	2.40	38	
	B/A	-1.14E+04					-1.14E+04
KCLE	A	-1.16E-05	0.88	3.62	0.15	5393670	
	B	3.04E-01	1.58	8.56	1.94	121	

Table 4: (continued)

	B/A	-2.62E+04					-2.62E+04
KCLT	A	-4.16E-06	0.79	2.64	0.14	5487847	
	B	-8.33E-02	1.38	-1.00	0.18	12	
	B/A	2.00E+04					2.00E+04
KCVG	A	4.06E-06	0.84	14.43	0.13	5347487	
	B	2.50E-01	1.99	18.96	3.50	43	
	B/A	6.16E+04					6.16E+04
KDCA	A	3.87E-05	0.87	9.86	0.15	4736355	
	B	1.18E-01	4.38	267.99	14.65	397	
	B/A	3.06E+03					3.06E+03
KDEN	A	-2.12E-04	0.88	10.86	0.24	5338217	
	B	-2.25E-02	1.75	16.98	-2.74	89	
	B/A	1.06E+02					1.06E+02
KDFW	A	-3.74E-04	0.90	2.28	0.17	5300401	
	B	-1.52E-01	2.71	17.76	-2.98	79	
	B/A	4.06E+02					4.06E+02
KDTW	A	9.32E-05	0.85	2.90	0.15	5424015	
	B	1.54E+00	3.13	4.01	1.92	13	
	B/A	1.65E+04					1.65E+04
KEUG	A	-4.56E-05	0.70	12.06	0.19	5195953	
	B	-5.00E-02	1.39	8.43	-2.36	20	
	B/A	1.10E+03					1.10E+03
KIAH	A	-1.06E-05	0.83	2.49	0.20	5193286	
	B	6.09E-02	1.10	9.69	-0.56	115	
	B/A	-5.76E+03					-5.76E+03
KLAS	A	-2.33E-04	0.90	2.88	0.16	5433173	
	B	-2.50E-01	0.71	-0.23	0.40	8	

Table 4: (continued)

	B/A	1.07E+03					1.07E+03
KLAX	A	4.89E-05	0.71	6.46	0.08	5332533	
	B	-3.41E-01	3.08	35.96	-5.82	41	
	B/A	-6.99E+03					-6.99E+03
KLGA	A	1.28E-04	0.95	1.91	0.11	5174896	
	B	-5.56E-02	0.86	0.21	0.13	30	
	B/A	-4.33E+02					-4.33E+02
KMCI	A	-2.92E-05	0.85	3.15	0.15	5494553	
	B	-4.73E-01	1.85	12.01	-2.97	55	
	B/A	1.62E+04					1.62E+04
KMCO	A	-1.11E-04	0.80	216.34	1.62	5383438	
	B	1.39E-01	1.55	11.66	1.99	36	
	B/A	-1.25E+03					-1.25E+03
KMSP	A	-8.57E-05	0.88	3.84	0.11	5497700	
	B	-1.80E+00	4.29	1.21	-1.33	10	
	B/A	2.10E+04					2.10E+04
KORD	A	1.99E-05	0.91	2.16	0.13	5117591	
	B	7.93E-02	1.47	12.35	-2.16	74	
	B/A	3.98E+03					3.98E+03
KPHL	A	7.14E-05	0.86	2.48	0.16	5223332	
	B	-5.56E-01	1.81	5.46	-2.16	9	
	B/A	-7.78E+03					-7.78E+03
KPHX	A	-9.70E-04	0.84	4.68	0.22	5275447	
	B	2.73E-01	1.79	2.93	-1.13	11	
	B/A	-2.81E+02					-2.81E+02
KPIT	A	-1.50E-04	0.84	3.22	0.19	5298517	
	B	-1.79E-01	1.95	3.16	-0.65	13	

Table 4: (continued)

	B/A	1.19E+03					1.19E+03
KSAC	A	1.63E-04	0.81	40.27	0.07	5402902	
	B	-1.10E+00	2.85	3.93	-2.05	371	
	B/A	-6.72E+03					-6.72E+03
KSAN	A	-4.12E-05	0.69	10.00	-0.04	5389450	
	B	-8.16E-02	0.81	0.30	-0.09	98	
	B/A	1.98E+03					1.98E+03
KSAT	A	-2.34E-04	0.93	3.13	0.20	5444969	
	B	1.61E-02	0.91	1.12	0.05	620	
	B/A	-6.90E+01					-6.90E+01
KSEA	A	9.68E-05	0.75	9.46	0.18	5375319	
	B	4.88E-01	3.47	16.76	3.36	41	
	B/A	5.04E+03					5.04E+03
KSFO	A	4.77E-05	0.79	2.40	0.09	5344491	
	B	2.52E-02	1.25	12.16	0.51	278	
	B/A	5.28E+02					5.28E+02
KSMX	A	5.15E-04	0.75	26.05	0.09	5371896	
	B	1.51E-01	2.13	33.83	4.56	93	
	B/A	2.92E+02					2.92E+02
KSTL	A	1.96E-05	0.87	2.54	0.16	5524684	
	B	1.17E-01	1.54	27.07	3.45	94	
	B/A	5.96E+03					5.96E+03
KTPA	A	2.76E-05	0.79	7.61	0.14	5148249	
	B	0.00E+00	0.80	0.30	-0.22	57	
	B/A	0.00E+00					0.00E+00
						Average:	3.85E+03

1.5 PROCESSED OUTLIERS

After the data cleaning described above, we identified outliers using two methods. The first method is to use the *adjacent windspeed difference per minute* defined as the quotient of the difference in windspeeds of two consecutive records divided by the number of minutes separating those two windspeeds:

$$\delta_{tb,ta} = \frac{s_{tb}^{speed} - s_{ta}^{speed}}{tb - ta}$$

where

- ta and tb are the timestamps of two consecutive samples,
- s_{ta}^{speed} is the windspeed in knots of sample s having timestamp ta,
- $tb - ta$ is the difference of timestamps tb and ta in minutes, and
- $tb > ta$ (i.e., timestamp tb is later than ta) so that $\delta_{tb,ta} < 0$ only when $s_{tb}^{speed} < s_{ta}^{speed}$.

For each weather station ws in our list of 30 stations, we examined the distribution of non-zero $\delta_{tb,ta}$'s by first finding the standard deviation of $\delta_{tb,ta}$ (to which we refer as $\sigma_{\delta,ws}$) listed in Table 5. Second, we created a frequency distribution table of the product $(\delta)(\sigma_{\delta,ws})$ for each weather station ws, as listed in Table 6. For each range of the 30 frequency distributions in Table 6, we summed the count to produce the combined frequency distribution table shown as Table 7, which is plotted as a histogram in Figure 1 on page 17.

Figure 1 shows that an extreme-value test [1] may be appropriately applied to find outlying windspeeds. The figure shows that approximately 99% of the non-negative δ's are less than 3 standard deviations from 0. However, we deemed extreme-values of δ to be those values that are greater than 20 times $\sigma_{\delta,ws}$. Thus, let's refer to 20 times $\sigma_{\delta,ws}$. as *thresh*, hereafter.

When a $\delta_{tb,ta}$ exceeded *thresh*, we discarded the sample s_{tb}, which is the sample having timestamp tb, and all samples s_t following s_{tb} until $delta_{t,ta} \leq thresh$. All samples that passed that filter were subject to another outlier filter: the strong-gale filter.

The strong-gale filter uses the fact that winds faster than strong-gales (i.e., winds above 47 knots) rarely occur on land [3]. We compared any winds faster than strong gales to

Weather Underground's archive, which includes ASOS data in addition to over 250,000 personal weather stations and over 26,000 Meteorological-Assimilation-Data-Ingest-System stations [5]. If our above-strong-gale windspeed for a certain day d exceeded the maximum windspeed for day d archived at Weather Underground, we discarded our windspeed; otherwise, we kept the windspeed in all cases except for one case. Weather Underground mistakenly records 1,000 mph as the maximum windspeed for 2010-10-19 at station KMCO. Since 1,000 mph is obviously false, we turned to another another source [4], which showed less than 5 mph for that date.

Table 5: Descriptive statistics of δ for each station

station	mean	standard deviation	kurtosis	skewness
KATL	0.00	0.87	3.95	0.17
KBOS	0.00	0.90	2.45	0.09
KBWI	0.00	0.86	4.57	0.17
KCLE	0.00	0.88	3.62	0.15
KCLT	0.00	0.79	2.64	0.14
KCVG	0.00	0.84	14.43	0.13
KDCA	0.00	0.87	9.86	0.15
KDEN	0.00	0.88	10.86	0.24
KDFW	0.00	0.90	2.28	0.17
KDTW	0.00	0.85	2.90	0.15
KEUG	0.00	0.70	12.06	0.19
KIAH	0.00	0.83	2.49	0.20
KLAS	0.00	0.90	2.88	0.16
KLAX	0.00	0.71	6.46	0.08
KLGA	0.00	0.95	1.91	0.11
KMCI	0.00	0.85	3.15	0.15
KMCO	0.00	0.80	216.34	1.62
KMSP	0.00	0.88	3.84	0.11
KORD	0.00	0.91	2.16	0.13
KPHL	0.00	0.86	2.48	0.16
KPHX	0.00	0.84	4.68	0.22
KPIT	0.00	0.84	3.22	0.19
KSAC	0.00	0.81	40.27	0.07
KSAN	0.00	0.69	10.06	-0.04
KSAT	0.00	0.93	3.13	0.20
KSEA	0.00	0.75	9.46	0.18
KSFO	0.00	0.79	2.40	0.09
KSMX	0.00	0.75	26.06	0.09
KSTL	0.00	0.87	2.54	0.16
KTPA	0.00	0.79	7.61	0.14

Table 6: For each weather station ws, the frequency distributions of $(\delta \geq 0)(\sigma_{\delta,ws})$

station	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)	[10,11)	[11,12)	[12,13)	[13,14)	[14,15)	[15,16)	[16,17)	[17,18)	[18,19)	[19,20)
KATL	3997537	169853	33031	5926	1079	277	148	32	8	10	4	1	4	3	0	2	0	1	2	0
KBOS	3937380	180312	34554	6558	1166	245	51	11	11	5	1	0	0	1						
KBWI	4024500	164549	34700	7099	1495	351	118	6	6	2	2	0	3	0	0	0	0	0	0	0
KCLE	3947912	172187	32377	6105	1217	356	114	77	19	4	3	5	1	1	1	2	2	0	0	2
KCLT	4196470	137101	22047	3947	90	31	21	7	3	1										
KCVG	4009064	153373	29400	5448	1102	368	57	25	16	6	9	2	1	0	0	0	0	0	0	0
KDCA	3525021	134720	29315	6721	1582	404	156	29	16	7	6	4	4	0	0	2	0	1	0	0
KDEN	3952564	154902	36839	8972	2315	688	222	157	30	13	4	6	4	4	4	0	3	0	1	1
KDFW	3817708	193430	38617	6833	1105	231	86	33	26	14	4	0	2	1	1					
KDTW	4004092	162864	29205	5052	953	334	49	23	26	14	8	6	1	0	0	1				
KEUG	4169716	88396	15575	315	99	8	6	8	2	1	2	0	1	1	0	0	1	0	0	0
KIAH	3870370	153100	27766	4548	790	247	38	21	9	4	3	0	1							
KLAS	3981235	175039	41368	9271	1851	380	81	20	22	5	1									
KLAX	4171954	78622	7055	157	44	2	2	0	0	0	0	0	1	1	0	0	0	0	0	1
KLGA	3579399	208577	40568	7598	1327	307	77	24	13	6	4	2	1	2	0	1	0	0	0	1
KMCI	4102080	167817	31459	5445	1012	302	28	21	12	13	10	2	1	2	0	0	0	1	0	0
KMCO	4136849	134191	23207	4564	244	102	45	33	6	6	5	6	1	1	4	0	0	0	0	0
KMSP	3985184	181355	33295	5818	995	191	63	37	8	4	4	1	2	0	1	0	0	0	0	0
KORD	3652652	185014	34176	6043	1113	245	81	38	13	17	3	1	1	2						
KPHL	3836098	154173	28853	5480	1050	272	97	18	3	9	2	1	4							
KPHX	3977796	143135	33588	7720	1620	438	40	20	9	10	5	2	0	0	1	1				
KPIT	3979226	151656	30269	6090	1231	451	49	36	13	4	4	2	2							
KSAC	4098961	144062	21219	2784	527	38	37	19	29	20	9	6	15	4	4	3	2	5	8	1
KSAN	4273561	76449	6127	77	23	5	0	1	5	1	0	1	0	2	0	1	0	0	1	
KSAT	3842823	210265	42621	7844	1398	263	84	32	27	13	9	3	2	0	0	0	0	0	0	0
KSEA	4165960	105754	13541	2265	112	27	31	8	3	5	1	0	0	0	0	0	0	0	0	0
KSFO	4080242	126901	18026	3288	137	38	18	16	3	2										
KSMX	4226634	102326	13901	2810	310	203	212	47	50	61	18	10	12	3	1	14	3	2	5	0
KSTL	4036294	176657	31860	5542	1008	241	161	27	21	15	7	3	1	2						
KTPA	3889128	116711	14283	2349	139	47	37	66	38	21	11	11	2	2	2	3	2	1	0	1

station	[20,21)	[21,22)	[22,23)	[23,24)	[24,25)	[25,26)	[26,27)	[27,28)	[28,29)	[29,30)	[30,31)	[31,32)	[32,33)	[33,34)	[34,35)	[35,36)	[36,37)	[37,38)	[38,39)	[39,40)
KATL	0	0	0	0	1															
KBOS																				
KBWI	1	1	0	0	1	1														
KCLE	0	0	1																	
KCLT																				
KCVG	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
KDCA	2	0	2	0	0	1	0	0	0	1	1	0	1							
KDEN	0	1	1	0	2	0	0	0	0	1	1	0	0	1	0	0	1			
KDFW																				
KDTW																				
KEUG	0	0	0	0	0	1	0	0	0	0	0	1								
KIAH																				
KLAS																				
KLAX	0	1	1																	
KLGA																				
KMCI	0	1																		

Table 6: (continued)

KMCO	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
KMSP	1																			
KORD																				
KPHL																				
KPHX																				
KPIT																				
KSAC	2	3	0	0	2	3	0	0	1	1	3	1	0	2	2	0	0	1		
KSAN																				
KSAT	0	1	0	1	1	0	0	1												
KSEA	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1				
KSFO																				
KSMX	2	2	2	1	1															
KSTL																				
KTPA	0	1	0	0	0	0	0	0	0	0	0	0	0	1						

station	[40,41)	[41,42)	[42,43)	[43,44)	[44,45)	[45,46)	[46,47)	[47,48)	[48,49)	[49,50)	[50,51)	[51,52)	[52,53)	[53,54)	[54,55)	[55,56)	[56,57)	[57,58)	[58,59)	[59,60)
KATL																				
KBOS																				
KBWI																				
KCLE																				
KCLT																				
KCVG																				
KDCA																				
KDEN																				
KDFW																				
KDTW																				
KEUG																				
KIAH																				
KLAS																				
KLAX																				
KLGA																				
KMCI																				
KMCO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KMSP																				
KORD																				
KPHL																				
KPHX																				
KPIT																				
KSAC																				
KSAN																				
KSAT																				
KSEA																				
KSFO																				
KSMX																				
KSTL																				
KTPA																				

station	[60,61)	[61,62)	[62,63)	[63,64)	[64,65)	[65,66)	[66,67)	[67,68)	[68,69)	[69,70)	[70,71)	[71,72)	[72,73)	[73,74)	[74,75)	[75,76)	[76,77)	[77,78)	[78,79)	[79,80)
KATL																				

Table 6: (continued)

KBOS																				
KBWI																				
KCLE																				
KCLT																				
KCVG																				
KDCA																				
KDEN																				
KDFW																				
KDTW																				
KEUG																				
KIAH																				
KLAS																				
KLAX																				
KLGA																				
KMCI																				
KMCO	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0
KMSP																				
KORD																				
KPHL																				
KPHX																				
KPIT																				
KSAC																				
KSAN																				
KSAT																				
KSEA																				
KSFO																				
KSMX																				
KSTL																				
KTPA																				

station	[80,81)	[81,82)	[82,83)	[83,84)	[84,85)	[85,86)	[86,87)
KATL							
KBOS							
KBWI							
KCLE							
KCLT							
KCVG							
KDCA							
KDEN							
KDFW							
KDTW							
KEUG							
KIAH							
KLAS							
KLAX							
KLGA							
KMCI							
KMCO	0	0	0	0	0	0	1
KMSP							

Table 6: (continued)

KORD								
KPHL								
KPHX								
KPIT								
KSAC								
KSAN								
KSAT								
KSEA								
KSFO								
KSMX								
KSTL								
KTPA								

1.6 INTERPOLATING

Gaps in the training and testing files were linearly interpolated. A gap is one or more missing minute-by-minute records. For example, the following sequence of windspeed records (where each record has a windspeed field preceded by a timestamp field having the format `YYYYMMDDHHmmhhmm` where `YYYY` is the year, `MM` is the month, `DD` is the day, `HH` is the local standard-time hour, `mm` is the minute, and `hh` is the UTC hour) has two gaps (a one-record or two-minute gap between UTC 05:02 and UTC 05:04 and a three-record or four-minute gap between UTC 05:05 and UTC 05:09):

```
2004010100010501,5
2004010100020502,5
2004010100040504,5
2004010100050505,5
2004010100090509,6
```

After interpolation, the sequence is

```
timestamp,wind_knots,actual vs. interpolated
2004010100010501,5,a
2004010100020502,5,a
2004010100030503,5,i
2004010100040504,5,a
2004010100050505,5,a
2004010100060506,5,i
2004010100070507,5,i
2004010100080508,5,i
2004010100090509,6,a
```

The interpolation routine (a portion of which is shown immediately below in Java code) truncates fractional values instead of rounding to the nearest integer:

```java
double dWindSpeedKnotsInterval = (iWindSpeedKnotsNext −
    iWindSpeedKnots)/ (double) diffInMinutes;
double dWindSpeedInterpolated = (double) iWindSpeedKnots;
for (int j=1; j<diffInMinutes; j++) {
  calLocal.add(Calendar.MINUTE, 1);
  calUtc.add(Calendar.MINUTE, 1);

  dWindSpeedInterpolated += dWindSpeedKnotsInterval;

  // build field
```

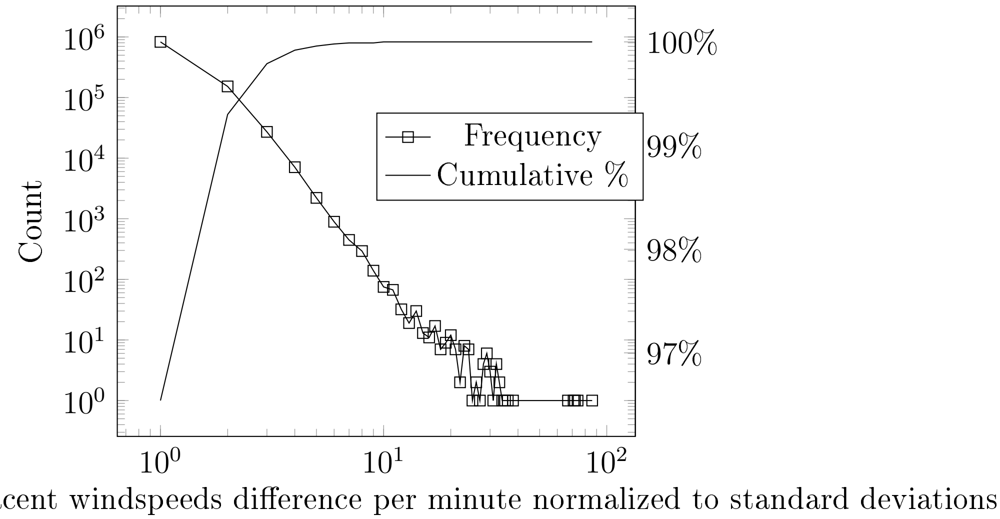

Figure 1: Histogram of Table 7 sans frequencies of 0: Frequency distribution table summarizing the 30 stations' frequency distributions of $\sigma_{\delta,ws}$ for all $\delta > 0$

```
        String formattedLocalTimestamp = sdf.format(calLocal.getTime());
        String formattedUtcHourAndMinutes = sdfUtc.format(calUtc.getTime
            ());
        String interpolatedLine = formattedLocalTimestamp +
          formattedUtcHourAndMinutes +
          "," + (int) dWindSpeedInterpolated;
          writeLine(interpolatedLine + ",i\n", writer);
    }
```

Table 7: Frequency distribution table summarizing the 30 stations' frequency distributions of $\sigma_{\delta,ws}$ for all $\delta \geq 0$

$\sigma_{\delta,ws} \forall \delta \geq 0$	Frequency	$\sigma_{\delta,ws} \forall \delta \geq 0$	Frequency	$\sigma_{\delta,ws} \forall \delta \geq 0$	Frequency
[0,1)	4503491	[29,30)	6	[58,59)	0
[1,2)	828842	[30,31)	3	[59,60)	0
[2,3)	152669	[31,32)	1	[60,61)	0
[3,4)	27134	[32,33)	4	[61,62)	0
[4,5)	7092	[33,34)	2	[62,63)	0
[5,6)	2209	[34,35)	1	[63,64)	0
[6,7)	892	[35,36)	1	[64,65)	0
[7,8)	447	[36,37)	1	[65,66)	0
[8,9)	293	[37,38)	0	[66,67)	0
[9,10)	139	[38,39)	1	[67,68)	1
[10,11)	75	[39,40)	0	[68,69)	0
[11,12)	67	[40,41)	0	[69,70)	0
[12,13)	32	[41,42)	0	[70,71)	0
[13,14)	19	[42,43)	0	[71,72)	1
[14,15)	30	[43,44)	0	[72,73)	1
[15,16)	13	[44,45)	0	[73,74)	0
[16,17)	11	[45,46)	0	[74,75)	1
[17,18)	17	[46,47)	0	[75,76)	0
[18,19)	7	[47,48)	0	[76,77)	0
[19,20)	9	[48,49)	0	[77,78)	0
[20,21)	12	[49,50)	0	[78,79)	0
[21,22)	7	[50,51)	0	[79,80)	0
[22,23)	2	[51,52)	0	[80,81)	0
[23,24)	8	[52,53)	0	[81,82)	0
[24,25)	7	[53,54)	0	[82,83)	0
[25,26)	1	[54,55)	0	[83,84)	0
[26,27)	2	[55,56)	0	[84,85)	0
[27,28)	1	[56,57)	0	[85,86)	0
[28,29)	4	[57,58)	0	[86,87)	1
cont. above right		cont. above right		[87,∞]	0

2.0 BIBLIOGRAPHY

[1] C.C. Aggarwal. *Outlier Analysis*. Springer International Publishing, 2016. ISBN: 9783319475783. URL: https://books.google.com/books?id=KyG1DQAAQBAJ.

[2] National Oceanic and Atmospheric Administration et al. *Automated Surface Observing System (ASOS) User's Guide*. Tech. rep. Mar. 1998. URL: https://www.weather.gov/media/asos/aum-toc.pdf.

[3] NOAA. *Beaufort Wind Scale*. URL: https://www.spc.noaa.gov/faq/tornado/beaufort.html.

[4] Time and Date AS. *Past Weather in Orlando, Florida, USA — October 2010*. URL: https://www.timeanddate.com/weather/usa/orlando/historic?month=10&year=2010.

[5] Weather Underground. *About Our Data*. URL: https://www.wunderground.com/about/data.