



ISTITUTO NAZIONALE DI RICERCA METROLOGICA  
Repository Istituzionale

Reassessing changes in diurnal temperature range: A new data set and characterization of data biases

This is the author's submitted version of the contribution published as:

*Original*

Reassessing changes in diurnal temperature range: A new data set and characterization of data biases / Thorne, P. W.; Menne, M. J.; Williams, C. N.; Rennie, J. J.; Lawrimore, J. H.; Vose, R. S.; Peterson, T. C.; Durre, I.; Davy, R.; Esau, I.; Klein Tank, A. M. G.; Merlone, Andrea. - In: JOURNAL OF GEOPHYSICAL RESEARCH. ATMOSPHERES. - ISSN 2169-897X. - 121:10(2016), pp. 5115-5137. [10.1002/2015JD024583]

*Availability:*

This version is available at: 11696/54561 since: 2017-02-24T15:13:48Z

*Publisher:*

Wiley

*Published*

DOI:10.1002/2015JD024583

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

1 Reassessing changes in Diurnal Temperature Range: A new dataset and  
2 characterization of data biases.

3

4 P. W. Thorne<sup>1,2</sup>, M. J. Menne<sup>3</sup>, C. N. Williams<sup>3</sup>, J. J. Rennie<sup>4</sup>, J. H. Lawrimore<sup>3</sup>, R. S.  
5 Vose<sup>3</sup>, T. C. Peterson<sup>5</sup>, I. Durre<sup>3</sup>, R. Davy<sup>2</sup>, I. Ezau<sup>2</sup>, A. M. G. Klein-Tank<sup>6</sup>, A. Merlone<sup>7</sup>

6

7 <sup>1</sup> National University of Ireland Maynooth, Maynooth, Ireland

8 <sup>2</sup> Nansen Environmental and Remote Sensing Center, Bergen, Norway

9 <sup>3</sup> NOAA's National Center for Environmental Information - Asheville, Asheville, NC,  
10 USA

11 <sup>4</sup> Cooperative Institute for Climate and Satellites, Asheville, NC

12 <sup>5</sup> Asheville, NC

13 <sup>6</sup> KNMI, De Bilt, Netherlands

14 <sup>7</sup> Istituto Nazionale di Ricerca Metrologica (INRiM), Torino, Italy

15

16

17 Corresponding author:

18 Peter Thorne

19 National University of Ireland Maynooth

20 Maynooth

21 Ireland

22 [peter@peter-thorne.net](mailto:peter@peter-thorne.net)

23



25 **Key points**

- 26 • Breakpoints are found to be more prevalent in DTR than other elements
- 27 • DTR has decreased since the early 20<sup>th</sup> Century but decrease is not linear
- 28 • Effects of homogenization change many details of global and regional DTR

29 **Abstract**

30 It is almost a decade since changes in Diurnal Temperature Range (DTR) globally  
31 have been explicitly assessed in a stand-alone data analysis. The present study takes  
32 advantage of substantively improved basic data holdings arising from the  
33 International Surface Temperature Initiative's databank effort and applies the  
34 National Climatic Data Center's automated pairwise homogeneity assessment  
35 algorithm to reassess global and regional DTR records. It is found that breakpoints  
36 are more prevalent in DTR series than other temperature elements and that the  
37 resulting adjustments have a broader distribution. This strongly implies that there  
38 is an over-arching tendency across the global meteorological networks for non-  
39 climatic artifacts to impart either random or anti-correlated rather than correlated  
40 biases in maximum and minimum temperature series. Future homogenization  
41 efforts would likely benefit from a consideration of DTR, maximum and minimum, in  
42 addition to average temperatures. Estimates of change in DTR are relatively  
43 insensitive to whether adjustments are calculated directly or inferred from  
44 adjustments returned for the maximum and minimum temperature series. The  
45 homogenized series exhibit a reduction in DTR since the early 20<sup>th</sup> Century globally.  
46 Adjustments serve to roughly half the magnitude of the long-term global reduction  
47 in DTR in the basic 'raw' data. Most of the estimated reduction in globally-averaged  
48 DTR occurred over 1960-1980. In several regions DTR has apparently increased  
49 since the 1990s, whilst globally it has exhibited very little change. Estimated  
50 changes in DTR are an order of magnitude smaller than in maximum and minimum  
51 temperatures, which have both been increasing rapidly on multi-decadal timescales.

52        **1. Introduction**

53

54 Diurnal Temperature Range (DTR) is defined as the daily maximum ( $T_x$ ) minus the  
55 daily minimum ( $T_n$ ) temperature. Herein consideration of DTR is restricted to land  
56 regions where DTR is far more dynamic than over the oceans. Over land areas DTR  
57 varies enormously both seasonally and geographically [Wang and Dillon, 2014]. The  
58 nature of DTR variability is important from a theoretical perspective for myriad  
59 reasons including in understanding microclimate impacts and the nature of changes  
60 within the deeper boundary layer [e.g. Christy et al., 2009, Pielke and Matsui, 2005,  
61 Zhou and Ren, 2011, Parker, 2006, Steenveld et al., 2011, McNider et al., 2012], and  
62 potentially as a determinant between forcings that have different Short Wave and  
63 Long Wave radiative fingerprints but may otherwise be similar [e.g. Jackson and  
64 Forster, 2013; Wang and Dickinson, 2013]. Trends and variability in DTR also have  
65 important practical implications for human health [Paaijmans et al., 2010], ecology  
66 [Peng et al., 2013, Vasseur et al., 2014], and agriculture [Battisti et al., 2009]  
67 amongst others.

68

69 Meteorological records have been undertaken at observing stations that extend back  
70 to the late 18<sup>th</sup> Century regionally and to the late 19<sup>th</sup> Century quasi-globally  
71 [Rennie et al., 2014]. Efforts have been made for at least three quarters of a Century  
72 [Callendar, 1938, Hawkins and Jones, 2013] to collate these data, apply homogeneity  
73 assessments and ascertain the nature of changes in Land Surface Air Temperatures  
74 (LSAT) over the globe. Today, there exist several such datasets globally [Lawrimore

75 et al., 2011 (see also Williams et al., 2012a,b, 2013), Jones et al., 2012, Rohde et al.,  
76 2013] and regionally [e.g. Bohm et al., 2010; Tietavainen et al., 2010, Li et al., 2010,  
77 Jain and Kumar, 2012, Trewin, 2012, Vincent et al., 2012, Falvey and Garreaud,  
78 2009, Christy et al., 2009; Van der Schrier et al., 2013]. Many of these analyses have  
79 been limited to a consideration of changes in average temperatures ( $T_m$ ), in part  
80 because records for average temperatures are more complete (Figure 1).  
81 Proportionately the effect becomes substantial prior to about 1950 and critical prior  
82 to 1895 (Figure 1 lower panel). Most US series post-1895 have been digitized to  
83 include  $T_x$  and  $T_n$  elements as part of the Climate Database Modernization Program.  
84 Elsewhere the situation is substantially more mixed and depends upon the data  
85 source.

86

87 Although DTR has been discussed as part of more general analyses globally [Rohde  
88 et al., 2012, Donat et al., 2013] and regionally [e.g. Makowski et al., 2008, Sen Roy  
89 and Balling, 2005, Christy et al., 2009, Zhou and Ren, 2011], it is almost a decade  
90 since the last stand-alone comprehensive analysis of global DTR data and its  
91 homogeneity was produced [Vose et al., 2005] and over twenty years since the first  
92 such assessment [Karl et al., 1993]. The IPCC in the most recent working group 1  
93 assessment [Hartmann et al., 2013] noted that there was only '*medium confidence*'  
94 (see Mastandrea et al. [2010] for an interpretation of the specific meaning of this  
95 term in an IPCC context) in available records of observed changes in DTR due to the  
96 presence of a number of unresolved issues raised in the literature [Fall et al., 2011,

97 Williams et al., 2012c, Christy et al., 2009] and the lack of recent studies and  
98 analyses.

99

100 In the last decade substantial progress has been made in:

- 101 • Creating better more complete records of daily data holdings of  $T_x$  and  $T_n$   
102 with better provenance and quality control [Menne et al., 2012];
- 103 • In combining disparate global holdings of monthly records with the  
104 improved daily holdings to provide a more robust data basis from which to  
105 undertake analyses of long-term LSAT changes [Rennie et al., 2014]; and
- 106 • The creation of automated monthly climatic timeseries homogeneity  
107 assessment methods and their performance benchmarking and assessment  
108 [Venema et al., 2012, Williams et al., 2012c, Menne and Williams, 2009].

109

110 This paper aims to take advantage of these methodological and data innovations to  
111 create a new estimate of long-term changes in DTR globally and regionally. A  
112 subsequent companion paper compares these results to a broad range of other  
113 observationally based estimates [Thorne et al. submitted]. These subsequent  
114 analyses permit an assessment of sensitivity to both structural and parametric  
115 uncertainties [Thorne et al., 2005] in DTR estimation. A holistic assessment of DTR  
116 and its changes is stayed to the companion piece. This paper focuses instead upon  
117 the effects of the Pairwise Homogenization Algorithm (PHA) technique upon the  
118 data and a characterization of the resulting series and a consideration of  
119 implications for trends in  $T_x$  and  $T_n$ .

120

121 The remainder of the paper is structured as follows. In section 2 the data and  
122 homogenization methods employed in this study are briefly introduced. Section 3  
123 summarizes the impacts of running the PHA algorithm on the data and discusses  
124 potential implications for the nature of non-climatic artifacts in the record. Section 4  
125 describes the spatial and temporal evolution of the homogenized series for the  
126 spatially incomplete global mean and a subset of regions for which data are  
127 complete enough to analyze back to 1901 (Europe, N. America and Australia).  
128 Section 5 provides a brief discussion. Section 6 contains details on the dataset  
129 availability and Section 7 concludes.

130

## 131 **2. Data and homogeneity assessment method**

132

### 133 2.1 Source data

134

135 The present analysis is exclusively based upon the version 1 ‘recommended merge’  
136 release of the Global Land Surface Databank [Rennie et al., 2014] at monthly data  
137 resolution. This databank release is a result of efforts by many international  
138 collaborators under the auspices of the International Surface Temperature Initiative  
139 [Thorne et al., 2011]. It has combined holdings from over 50 constituent sources  
140 ranging from single stations to holdings of many thousand stations. These sources  
141 have been merged hierarchically with merge decisions based upon both metadata  
142 and data similarity metrics. Sources with  $T_x$  and  $T_n$  and better provenance and

143 believed to be closer to the original recorded 'raw' basic data have been prioritized.  
144 The merge creates a single unique version per station that is as long as possible  
145 while minimizing potential discontinuities through false imputation of short period  
146 data. In total this version consists of just over 32,000 stations, most of which have  
147  $T_x$  and  $T_n$  series for at least part of their records and many of which extend over at  
148 least 100 years (although not necessarily continuously).

149

150 The processing of the databank series merged  $T_x$  and  $T_n$  series stations first and  
151 only then went back to look for record segments for which solely  $T_m$  records exist.  
152 Despite this deliberate effort to maximize the amount of  $T_x$  and  $T_n$  data pull-  
153 through, availability for these elements is always lower than for  $T_m$  (Figure 1). It is  
154 all but certain that  $T_x$  and  $T_n$  data, or at least observations at intervals over the day,  
155 were associated with the original records for which in the digital archives now only  
156  $T_m$  data exist in most cases. These data have either been lost or more likely were  
157 never digitized. This attests to the real importance of data rescue efforts, even for  
158 those stations which nominally already have records but for which the records are  
159 incomplete in important aspects such as availability of daily summaries which serve  
160 to inhibit understanding [e.g. Allan et al., 2011].

161

162 To facilitate the analysis herein a fourth field –  $T_{dtr}$  – the difference between  $T_x$  and  
163  $T_n$  has also been calculated and analyzed. In addition, for those analyzes of  
164 homogenization performance (Section 3) which include recourse to results for  $T_m$   
165 these consider solely  $T_m$  values derived directly from the  $T_x$  and  $T_n$  elements as

166 their average. This avoids conflation of data completeness and data characteristics  
167 in the analysis, which would otherwise ensue from use of the more temporally  
168 complete merged  $T_m$  series (Figure 1). In many cases for remaining  $T_m$  reports the  
169 archived  $T_m$  may not result simply from averaging  $T_x$  and  $T_n$ . For example in at  
170 least Australia (and perhaps many other regions) in recent years the monthly  
171 average reported in CLIMAT messages is the average of hourly reports. Regardless,  
172 given that PHA is a neighbor-based procedure it is important to have the same  
173 networks for each element to perform a fair comparison and evaluation.

174

175 Both the DTR and the  $T_m$  fields result from direct calculation from the monthly  
176 mean  $T_x$  and  $T_n$  series. So, the basic data used herein are internally consistent in  
177 that in the data presented to the homogenization algorithm DTR will always be the  
178 difference between  $T_x$  and  $T_n$ ,  $T_m$  will always be their average, and these elements  
179 are only ever calculated when both  $T_x$  and  $T_n$  are present. However, for months  
180 where either  $T_x$  and / or  $T_n$  have missing daily values this is not going to be  
181 equivalent to the average of the calculable daily DTRs (or  $T_m$ 's) within the month.  
182 While a more restrictive criteria of calculation of these values from the dailies could  
183 be applied to the subset of the databank arising from daily sources [Rennie et al.,  
184 2014] it would result in considerably fewer candidate station records, particularly  
185 prior to the 1950s. This comes at a potential cost regarding the monthly statistical  
186 mean and / or standard deviation characteristics for those stations where data is  
187 patchy on an intra-month basis due either to frequent missing days or frequent  
188 quality control flagging on the daily reports.

189

## 190 2.2 Pairwise Homogeneity Assessment

191

192 The data are presented to the exact same processing suite as those for Global  
193 Historical Climatology Network Monthly (GHCN, currently GHCN-Mv3.2.0)  
194 [Lawrimore et al., 2011, Williams et al. 2012a,b, 2013]. This consists of a set of  
195 quality control checks followed by application of a Pairwise Homogeneity  
196 Assessment (PHA) breakpoint identification and adjustment procedure [Menne and  
197 Williams, 2009]. The interested reader is directed to these papers for a fuller  
198 exposition of the methodology than is possible here if technical details are required.

199

200 The data are submitted separately for each of the four data streams considered ( $T_x$ ,  
201  $T_n$ ,  $T_m$  and DTR). No attempt is made herein to consider these data jointly to ensure  
202 consistency in returned adjustments across the elements when assessing  
203 homogeneity of the series, although such an approach is being actively developed  
204 for new versions of GHCNM. This is likely to yield inconsistencies at the station level  
205 between elements herein, which may occasionally be substantial (Section 3). The  
206 PHA algorithm analyzes timeseries of pairwise differences between nearby stations.  
207 It uses a Standardized Normal Homogeneity Test (SNHT) test statistic  
208 [Alexandersson, 1986] which is a t-test class of test, to identify potential  
209 discontinuities in each station pair. After doing so for all identified neighbor  
210 combinations the very large matrix of potential breakpoints is decomposed such  
211 that breakpoints are assigned iteratively to those stations in which they arise

212 concurrently across multiple inter-comparisons with the resulting counts reduced  
213 accordingly until no further plausible breakpoint candidates exist. Then  
214 adjustments are inferred for the resulting population of identified candidate real  
215 breakpoints through comparisons to apparently homogeneous neighbor segments  
216 and applied if the distribution of returned adjustment estimates is substantively  
217 non-zero. The process is run solely once and the resulting set of applied and rejected  
218 adjustments are returned. The stations have been adjusted based upon the  
219 adjustment estimates and quality control decisions returned by the PHA in its  
220 operational version settings. The ensemble analysis of Williams et al. [2012c]  
221 highlights potential impacts from giving different, plausible, parameter settings to a  
222 number of the uncertain parameters within the PHA algorithm. For the present  
223 analysis consideration of such ensembles is deemed beyond scope.

224

### 225 2.3 Station gridding

226

227 For subsequent analysis only stations and months with sufficient data to create a  
228 1971-2000 climatology under a Climate Anomaly Method have been retained in the  
229 gridded fields. As is discussed in the accompanying paper [Thorne et al., submitted]  
230 this is one of several possible approaches to gridding. For each station and calendar  
231 month, the minimum data requirement for calculating a climatology is  $2/3$  of data in  
232 the 30 year period taken as a whole and at least  $1/2$  in each decade (1971-1980,  
233 1981-1990 and 1991-2000). This implies that a climatology may have been  
234 computed for some, but not all, calendar months at a particular station. For example,

235 if the station's operator always took a vacation in July, then an insufficient amount  
236 of data may have been available for July while data for the other months of the year  
237 were sufficiently complete. In practice stations tend to be either substantively  
238 complete over the climatology period or have a marked data paucity that precludes  
239 their inclusion, meaning this affect is relatively minor in the retained station set.  
240 Stations for which a climatology can be calculated for any month tend to have  
241 climatologies for all twelve calendar months.

242

243 The climatology value has only been calculated with a trimmed mean based upon  
244 solely months within 3 standard deviations ( $\sigma$ ) of the climatology period data mean  
245 for the given calendar month. An additional simple  $5\sigma$  anomaly QC check has then  
246 been applied to the resulting anomaly series on a calendar month basis to remove  
247 gross outliers. Data between 3 and 5 standard deviations are retained but do not  
248 inform the climatological estimate. In stations with a strong secular trend this  
249 quality control step may remove real points far away in time from the climatology  
250 period. A high critical threshold of  $5\sigma$  was chosen to mitigate this risk while still  
251 ensuring grossly questionable data did not get gridded. The check removes solely a  
252 handful of grossly questionable data points.

253

254 Resulting anomalies have simply been gridded, without any further weighting, into  
255 bins of 5 degrees latitude by 5 degrees longitude. Data have been gridded for all  $T_x$ ,  
256  $T_n$  and DTR for both the raw and adjusted series. Gridded  $T_m$  series are not  
257 considered herein but will be documented in forthcoming GHCNM analyses instead.

258

259 For DTR it is possible to estimate the adjustments and resulting gridded series both  
260 directly from applying PHA to the timeseries and indirectly, through applying the  
261 net effect of the returned adjustments to  $T_x$  and  $T_n$ . The latter approach will yield a  
262 set of physically consistent estimates by construction, but at a potential cost if it  
263 misses breaks more amenable to identification and / or adjustment in DTR.  
264 Regardless, differences arising between ‘directly adjusted’ and ‘indirectly adjusted’  
265 series provide some indication of likely uncertainties / sensitivities of the resulting  
266 analyses using the PHA method. However, these are very much an incomplete  
267 indicator of the likely true uncertainties. Comparisons to other estimates,  
268 constructed using distinct methods for all processing choices including quality  
269 control, adjustment, climatology calculation and gridding, will likely give a more  
270 realistic assessment of the true magnitude of the uncertainties in DTR estimates and  
271 are discussed further in the accompanying paper [Thorne et al., submitted].

272

### 273 **3. Analysis of homogeneity adjustments**

274

#### 275 3.1 A consideration of the potential structure and magnitude of breakpoints

276

277 The four sets of series submitted to PHA consist of the two primary elements ( $T_x$   
278 and  $T_n$ ), their average ( $T_m$ ), and their difference (DTR). To ascertain the possible  
279 effects of the different data artifact characteristics on breakpoint magnitudes and  
280 distributions all possible combinations of  $T_x$  and  $T_n$  breakpoints between -5 and 5 K

281 have been considered in Figure 2. By construction breakpoints in  $T_m$  are always  
282 smaller than the break in either  $T_x$  or  $T_n$  except in the special case where the breaks  
283 in both elements are identical in sign and magnitude (perfectly correlated). Because  
284 DTR is the difference between the two elements there is no such cancellation in  
285 breakpoints of DTR and absolute breakpoint magnitudes reach 10K at  $[-5K, 5K]$  and  
286  $[5K, -5K]$ . Hence DTR has twice as large a potential breakpoint magnitude for  
287 combinations explored as any of the other elements. By construction breakpoint  
288 magnitudes in DTR and  $T_m$  are orthogonal. In the limit of perfectly correlated  
289 breakpoints in  $T_x$  and  $T_n$  ( $T_x$  break =  $T_n$  break) there will be no breakpoints in DTR.  
290 Similarly for perfectly anti-correlated breakpoints ( $T_x$  break =  $-T_n$  break) there will  
291 be no breakpoints in  $T_m$ .

292

293 In cases where the breakpoints in  $T_x$  and  $T_n$  are correlated (both of the same sign)  
294 one or other of the breakpoints in  $T_x$  and  $T_n$  will always be the largest breakpoint.  
295 Where the breakpoints in  $T_x$  and  $T_n$  are anti-correlated (one positive, one negative)  
296 the largest breakpoint will always be in DTR. Restricting to a consideration of solely  
297  $T_m$  and DTR, the breakpoint in DTR will be largest both when the breakpoints in  $T_x$   
298 and  $T_n$  are anti-correlated, and when they are only weakly correlated (same sign  
299 but substantially distinct magnitude whereby the difference is greater than their  
300 mean).

301

302 Assuming that the inter-station noise arising from random effects and real physical  
303 effects is similar across the elements such that Signal-to-Noise Ratios (SNRs) are

304 similar in all resulting pairwise comparisons for breakpoint detection (Section 2.2)  
305 there is therefore a set of *a priori* expectations that can be made:

- 306 1. If the breakpoints in  $T_x$  and  $T_n$  are entirely randomly distributed and not  
307 conditionally dependent such that the break in  $T_x$  has no *a priori*  
308 distributional basis given a break in  $T_n$ , then it would be expected that there  
309 would be more and larger breaks in DTR than in  $T_x$  or  $T_n$  and fewest in  $T_m$ .
- 310 2. If the breaks in  $T_x$  and  $T_n$  are conditionally dependent such that if the break  
311 in  $T_n$  is positive it is more likely that  $T_x$  is also positive and vice-versa then  
312 most and larger breakpoints would be expected to be found in  $T_x$  and  $T_n$   
313 with fewest in DTR or  $T_m$  (depending upon whether the conditioning was  
314 weak ( $T_m$ ) or strong (DTR))
- 315 3. If the breaks in  $T_x$  and  $T_n$  are conditionally independent such that a negative  
316 break in  $T_n$  has a tendency to lead to a positive break in  $T_x$  and vice-versa  
317 then it would be expected that most breaks would be found in DTR and they  
318 would be substantially larger than in  $T_x$  and  $T_n$  with fewer, much smaller  
319 breaks in  $T_m$

320

### 321 3.2. Analysis of returned breakpoint adjustments from the PHA algorithm

322

323 The PHA algorithm (Section 2.2) was run on the subset of stations which had  
324 sufficiently long records and for which sufficient neighbor estimates existed. The  
325 data masks are exactly equivalent for  $T_m$  and DTR as they require  $T_x$  and  $T_n$  to both  
326 be available (Section 2.1). For  $T_x$  and  $T_n$  some additional data exists for some

327 stations. However, to a first approximation the number of stations and record length  
328 are equivalent for all four elements presented to PHA. Despite this similarity in  
329 input data availability there exist marked differences in the estimated frequency,  
330 magnitude and distribution of adjustments returned across the 4 elements (Figure  
331 3). There are more adjustments returned for DTR (66,572) than for  $T_n$  (62,013), for  
332 which there are more again than for both  $T_x$  (51,777) and  $T_m$  (50,378). The  
333 standard deviation of the returned adjustment estimates is largest for DTR (1.24K),  
334 roughly equivalent for  $T_x$  (0.98K) and  $T_n$  (1.00K), and smallest for  $T_m$  (0.75K).  
335 There is no obvious substantial departure for any element from Gaussian  
336 distributional assumptions. In all cases there is a 'missing middle' of undetectable /  
337 unadjustable real-world breakpoints that must in reality exist.

338

339 Following from Section 3.1 if there is no difference in effective power of PHA to  
340 detect and adjust for breaks between elements then the implication is that the  
341 breakpoints in  $T_x$  and  $T_n$  are either entirely random or conditionally independent.  
342 However, there are also reasons why DTR may be expected to exhibit lower noise as  
343 it is the difference between two variables,  $T_x$  and  $T_n$ , which tend to co-vary on  
344 monthly timescales. If the noise in the pairwise station comparators, which form the  
345 basis for the breakpoint statistical assessment, was lower then it may simply be that  
346 PHA can more efficiently detect smaller breakpoints from the 'missing middle'  
347 clearly evident in all panels of Figure 3. It is obvious given the broader distribution  
348 of DTR adjustments from Figure 3 that the increased number of breakpoints found

349 and adjusted in DTR results from larger discontinuities rather than any difference in  
350 efficacy of breakpoint identification.

351

352 The breakpoint behavior can be further investigated by consideration of directly  
353 inferred and indirectly inferred adjustment estimates for DTR and  $T_m$  (Figures 4  
354 and 5). Breaks in the derived variables would be expected to be coincident in timing  
355 and resulting magnitude with those estimated from the  $T_x$  and  $T_n$  analyses.

356 Comparing direct and indirect adjustment estimates therefore provides a check on  
357 internal consistency of results. The direct and indirect adjustment estimates should  
358 be correlated and show no overall offset from one another. Scatter would be  
359 expected to arise due to variations in breakpoint date assignments and neighbor  
360 segments used to adjust. The degree of scatter provides some indication of the  
361 probable uncertainty in the resulting station series estimates.

362

363 For DTR these comparisons exhibit substantial scatter, even when a collocation  
364 error of 12 months in the breakpoint locations found is allowed for (Figure 4 left  
365 hand panel). There are many cases where either a DTR adjustment is made without  
366 a corresponding adjustment to either  $T_x$  or  $T_n$  and vice-versa (points along either  
367  $y=0$   $x \neq 0$  or  $x=0$   $y \neq 0$  respectively). In numerous cases the adjustments differ in sign  
368 (top left and lower right quadrants). Overall, however, there is a tendency to  
369 broadly agree with the cloud of points scattered around the 1:1 line rather than  
370 entirely randomly. The histogram of adjustment comparators (Figure 4 right hand  
371 panel) is zero mean and broadly Gaussian, albeit with a large sigma such that almost

372 23% of differences exceed 1K in magnitude. A similar analysis of  $T_m$  (Figure 5)  
373 exhibits far less scatter between directly and indirectly inferred adjustments (left  
374 hand panel, points lie much closer to the 1:1 line) with only just under 5% of  
375 differences exceeding 1K in magnitude (right hand panel).

376

377 Both direct and indirect adjustments to DTR act to reduce the apparent spread in  
378 individual station linear trend fit estimates over 1901-2012 and 1951-2012 (Figure  
379 6). This is consistent with what would be expected if reasonable adjustments were  
380 being applied to data containing inhomogeneities. Individual station series in the  
381 basic data contain systematic data errors. Such systematic effects are equivalent to  
382 adding units of red noise to the time-series, causing artificial dispersion in the  
383 distribution of long-term station series behavior. Figure 6 suggests that many such  
384 systematic biases are being effectively removed in a reasonable manner by the PHA  
385 algorithm.

386

### 387 3.3 Synthesis of adjustments analysis

388

389 Breakpoints are more easily discoverable using PHA in DTR than they are in  $T_x$  or  
390  $T_n$  which in turn are somewhat more discoverable than in  $T_m$ . Earlier analyses over  
391 the European domain [Wijngaard et al., 2003] and globally using HadISD [Dunn et  
392 al., 2014] found similarly that breakpoints in DTR were somewhat more amenable  
393 to detection. Not only were more breakpoints found in DTR but they were on  
394 average larger and had a broader standard deviation than other elements. When

395 calculated directly from DTR or indirectly from  $T_x$  and  $T_n$  adjustments, individual  
396 adjustment estimates show similar behavior but with substantial dispersion.  
397 Therefore care should be taken in interpretation of individual adjusted station DTR  
398 series. However, the overall distribution of station trend estimates is less dispersive  
399 following application of adjustments with many obviously questionably large  
400 station trends removed. Taken as a whole this analysis provides confidence in the  
401 efficacy of PHA when applied to DTR series at least at regional or global scales.

402

403 Overall, results from PHA strongly imply that breakpoints in  $T_x$  and  $T_n$  are either  
404 randomly distributed or conditionally independent. Strong conditional dependence  
405 whereby  $T_x$  and  $T_n$  breakpoints are almost always of the same sign and similar  
406 magnitude can be ruled out by the present analysis. Reasons and implications are  
407 returned to in the discussion (Section 5).

408

#### 409 **4. Analysis of gridded fields and regional averages**

410

##### 411 4.1 Data completeness

412

413 As with most preceding analyses of DTR [e.g. Vose et al., 2005] data is globally  
414 incomplete and the data density in those areas sampled varies over at least two  
415 orders of magnitude. Figure 7 shows gridbox DTR station data counts for the month  
416 when data density is globally maximal (October 1987). Sampling is dense over  
417 much of Australia, China and Japan, Europe and in particular North America.

418 Sampling is particularly poor (or even non-existent) over much of Africa, SE Asia,  
419 the Arabian Peninsula, the Amazon basin and the ice sheets of Antarctica and  
420 Greenland. Sampling varies substantively through time both globally and regionally  
421 in those regions with records that extend back to the early 20<sup>th</sup> Century (Figure 8).  
422 Outside North America there exists a step-change in availability in 1960 with far  
423 fewer stations prior to this. As a result trends and variability in DTR for analyses  
424 across 1960 may be an artifact of coverage changes rather than true changes. As  
425 discussed further in Section 2.1 there likely exist records which if rescued digitized  
426 and shared could mitigate this issue.

427

## 428 4.2 Diurnal Temperature Range

429

430 Herein analysis is made of changes in DTR from the original 'raw' data records and  
431 following adjustments calculated directly and indirectly from applying the  
432 adjustments returned to  $T_x$  and  $T_n$  and then calculating DTR from these series as  
433 outlined in Section 2.3. The analysis starts with spatial patterns of trends over  
434 increasingly shorter periods to present. Recourse is then made to regionally  
435 averaged timeseries behavior and linear trend estimates.

436

### 437 4.2.1 Spatial trends

438

439 Trends calculated since the beginning of the 20th Century greatly reduce coverage if  
440 a data completeness mask is applied to ensure early and late period data availability

441 in addition to total timeseries completeness (Figure 9 c.f. Figure 7). Data remain  
442 only for N. America, Europe, parts of Australia, E. China and Japan and a handful of  
443 dispersed additional locations. The spatial domains sampled in Figure 9 govern the  
444 designation of sub-domains considered in subsequent regional analyses and  
445 denoted henceforth by geographic shorthand as: N. America (45W-135W, 25-60N);  
446 Europe (10W-60E, 25-60N); and Australia (110E-155E, 10S-45S). The cluster over  
447 Japan and E. China is deemed too small to calculate a reasonable regional average.  
448

449 Century timescale trends in DTR (Figure 9) are of the order 0.1K/decade at most  
450 across the sampled gridboxes in the raw data and in the two adjusted products.  
451 Trends are significant at the gridbox level in many of the gridboxes sampled in the  
452 input data, but this decreases substantially following application of adjustments  
453 either using the direct or the indirect approach. In the input data most gridboxes  
454 exhibit a reduction in DTR over time. Although a majority of gridboxes still indicate  
455 a reduction in DTR following the application of adjustments, the magnitude of the  
456 DTR reduction is far less significant. Adjustments change the sign of the DTR trends  
457 in much of the South Western / Western United States from negative to positive and  
458 reduce the negative trends elsewhere in N. America. This change is more marked  
459 when adjustments are calculated indirectly than when they are calculated directly.  
460 There are less spatially consistent changes in remaining regions with many  
461 gridboxes experiencing large changes including changing the sign of the DTR trend.  
462

463 Starting in 1951 as expected from Figure 8, spatial sampling is much more complete  
464 although Africa, the Indian sub-continent and S. America remain substantively  
465 incompletely sampled in addition to Greenland and Antarctica (Figure 10). Over this  
466 62 year period in the input data records the vast majority of gridboxes exhibit  
467 substantial reductions in DTR that are particularly marked over much of Asia and N.  
468 America. Application of adjustments substantively changes the trend behavior over  
469 N. America where trends are reduced with a sign change in many gridboxes west of  
470 the Rockies to an increasing DTR and very few gridbox series remain significant. In  
471 Southern Europe adjustments indicate small increases in DTR. Overall, adjusted  
472 series are visually somewhat more spatially homogeneous than the input data  
473 trends lending some support to the findings detailed in Section 3 regarding the  
474 efficacy of the PHA when applied either directly or indirectly to DTR records.

475

476 The last period for which geographical trends are considered is from 1979, a start  
477 date typically used in climate studies because it is the advent of regular polar-  
478 orbiter satellite measurements. Although the current analysis is in-situ only it is still  
479 potentially informative to other studies to document changes over this period  
480 (Figure 11). Over this period sampling is more complete again, particularly so over  
481 South America although large areas remain data void. Since 1979 trends are  
482 substantively larger in magnitude and of more mixed sign. That trends over shorter  
483 periods are larger, more spatially heterogeneous, and of mixed sign is to be  
484 expected as shorter periods increasingly reflect decadal-scale regional variability  
485 [Santer et al., 2011]. Over this shorter period, the application of adjustments leads

486 to large changes in apparent sign and magnitude of DTR trends in many regions.  
487 This is particularly marked in the United States, in parts of Europe and over much of  
488 China and SE Asia.

489

490 Over the United States the adjustments in the post-1979 era lead to a change from a  
491 slight reduction in DTR to a larger increase in many gridboxes. The adjusted DTR  
492 increases are significant in several gridboxes in the South Western states. This  
493 adjustment is consistent with understanding of the transition from Cotton Region  
494 Shelters (CRS, termed Stevenson Screens elsewhere) to electronic Maximum  
495 Minimum Temperature Sensor (MMTS) starting in the 1980s and substantively  
496 completed by the late 1990s. In this change both the instrument and its shielding  
497 were changed substantively, often associated with a change in measurement  
498 location. This change affected roughly 70% of the COOP network, which is the  
499 backbone of the US records. Field based studies and statistical analyses have  
500 variously concluded that the CRS to MMTS transition led to a positive bias in  $T_n$  and  
501 a negative bias in  $T_x$  artificially reducing DTR in the raw data [Fall et al., 2011,  
502 Williams et al., 2012c and references therein]. Assuming that the PHA algorithm is  
503 adequate the effect of this change is larger than the underlying real-world DTR  
504 signal over much of the United States. The size of the effect found and adjusted for  
505 here is consistent in magnitude with understanding from various side-by-side  
506 comparisons under the assumption that c.70% of the network experienced the  
507 change.

508

509 In Europe adjustments lend support to the propensity for increased DTR in recent  
510 years [Vautard et al., 2009]. In China and SE Asia, although gridbox trends remain  
511 significant the reductions in DTR are generally less following adjustment than is  
512 implied by the raw data.

513

#### 514 4.2.2 Regional and global timeseries and trends

515

516 As is visually obvious from Figures 9-11 linear trend estimates do not describe all  
517 facets of the timeseries behavior globally or regionally. Timeseries for global (Figure  
518 12) and regional (Figure 13) DTR averages serve to highlight the presence of  
519 substantial interannual to multi-decadal variability in DTR even globally. In all cases  
520 these timeseries have been derived from averaging all available gridded data at each  
521 timestep using  $\cos(\text{lat})$  area weighting. As noted earlier, given the varying station  
522 count and gridbox availability care should be taken in interpretation in particular of  
523 pre-1960 data. The effects of different completeness inclusion criteria for this step  
524 are further discussed and analyzed in the accompanying paper [Thorne et al.,  
525 submitted].

526

527 Following adjustments it is estimated that globally averaged DTR was elevated  
528 relative to present day until the late 1950s, declined by of the order 0.2C by the  
529 early 1980s and has then been relatively steady since according to both adjusted  
530 series considered. There are substantial differences between directly and indirectly  
531 adjusted series estimates prior to around 1950. Overall the adjusted series are more

532 similar to each other than they are to the input data both in terms of the long-term  
533 trend and also decadal timescale variability. Globally adjustments have a substantial  
534 impact in the most recent period since 2000 when (semi-)automation has been  
535 prevalent across the global network as a whole (although some regions experienced  
536 this change 10-20 years earlier), and prior to the 1970s.

537

538 Global and regional average trends are substantively impacted by the PHA  
539 homogenization procedures. Adjusting either directly or indirectly the net effect is  
540 to reduce the magnitude of the apparent long-term trends in global DTR (Table 1).  
541 Nonetheless, trends towards globally reduced DTR are statistically significant over  
542 the period 1901 to 2012 and the shorter sub-period 1951 to 2012 for the 'raw'  
543 series and remain so for the adjusted series. Over the period 1979 to 2012 the  
544 global mean trend reverses from a significant reduction in the 'raw' data, to a slight  
545 increase in both of the adjusted series neither of which are statistically significant  
546 (c.f. Figure 11 and associated discussion).

547

548 In North America the adjustments reduce DTR prior to 1950 and increase DTR since  
549 the 1980s yielding a large reduction in the apparent narrowing of DTR implied by  
550 the basic 'raw' data (Figure 13, top panel). As discussed previously post-1980  
551 changes are consistent with understanding of the effects of transition from CRS to  
552 MMTS across roughly 70% of the US observing network. Earlier period adjustments  
553 may relate either to the effects of changes in time of observation [Karl et al., 1986]  
554 or a propensity to relocate from city to airport locations. Trends over 1901-2012 are

555 significantly negative in the basic 'raw' data and both adjusted series, but are halved  
556 in magnitude following adjustments. Over the two shorter periods considered  
557 neither adjusted series exhibits significant trend behavior. Estimates are slightly  
558 negative over 1951-2012 and slightly positive over 1979-2012 (Table 1). The two  
559 adjusted series are very similar to each other and very distinct from the basic 'raw'  
560 data behavior.

561

562 Over the European domain adjustments act to increase DTR both since the 1980s  
563 and prior to the 1950s (Figure 13, middle panel). This yields a marked change in  
564 multi-decadal variability in this region removing an apparent trend of increasing  
565 DTR in the first half of the twentieth Century in the basic 'raw' data. On the longest  
566 timescales this leads to an increased negative trend in DTR following adjustments,  
567 which is significant in both adjusted estimates but not the basic data (Table 1). Over  
568 1951-2012 again all estimates are significantly negative. Since 1979 both adjusted  
569 series imply positive trends in DTR over the European domain taken as a whole but  
570 these are not statistically significant. As is the case globally and over N. America the  
571 adjusted series are much more similar to each other than they are to the basic 'raw'  
572 data.

573

574 Australian DTR series exhibit far greater variability than those over Europe and  
575 America (Figure 13, lower panel). Variability appears to be highly correlated with  
576 continental scale aridity / rainfall (and by extension ENSO). For example the very  
577 wet year of 2010/11 is associated with a marked negative DTR anomaly, consistent

578 with basic theoretical understanding of partitioning of fluxes [Peterson et al., 2011].  
579 The effect of the adjustments is more muted for this region with slight increases in  
580 DTR in the mid-20<sup>th</sup> Century and reductions in the early 20<sup>th</sup> Century. Trends are  
581 generally not significant in the adjusted series with the exception of indirectly  
582 adjusted series for 1901-2012 (Table 1) and confidence intervals are larger than for  
583 other regions considered reflecting the much greater year to year variability in the  
584 series. Over this region there is less obvious concordance between the adjusted  
585 series.

586

#### 587 4.3 Maximum and minimum temperatures

588

589 For  $T_x$  and  $T_n$  only direct adjustments exist so analysis is limited to the raw and  
590 directly adjusted series. Trends over 1951-2012 for  $T_x$  (Figure 14) and  $T_n$  (Figure  
591 15) both exhibit strong warming in the vast majority of the gridboxes that are  
592 sampled. Adjustments remove an apparent cooling in  $T_x$  in the eastern United States  
593 consistent with the United States Historical Climatology Network (USHCN) [Menne  
594 et al., 2010] and our understanding of US biases arising from the CRS to MMTS  
595 transition. Cooling in  $T_x$  in Southern China is also reduced and several obviously  
596 erroneous gridbox series look more similar to surrounding series after  
597 homogenization. Adjustments to  $T_n$  adjust several obviously erroneous gridbox  
598 trends and increase slightly the apparent warming in eastern North America but  
599 otherwise have little obvious effect at the gridbox scale.

600

601 Global average timeseries of  $T_x$  and  $T_n$  are strongly positive (Figure 16), particularly  
602 since the early 1970s. Adjustments serve to narrow the difference in trends (which  
603 is consistent with a reduction in the estimated rate of decrease in DTR in the  
604 preceding subsection). The overall effect of PHA adjustments is to increase the long-  
605 term trend in both  $T_x$  and  $T_n$  with the effect being larger for  $T_x$  (although the  $T_x$   
606 trend is still smaller than that for  $T_n$ , Table 2). Trends in  $T_x$  and  $T_n$  are highly  
607 significant over all three periods considered in the present analysis and, in the  
608 adjusted series, roughly an order of magnitude larger than DTR trends. Trends in  $T_x$   
609 and  $T_n$  are consistent with GHCNv3.2.0 trends for  $T_m$  even though the station basis  
610 set differs substantially.

611

## 612 **5. Discussion**

613

614 The adjustments returned by the PHA algorithm strongly imply that breakpoints in  
615  $T_x$  and  $T_n$  are either random or conditionally independent. Random breaks would  
616 mean that the break size and magnitude in  $T_n$  on average had no influence upon the  
617 resulting break size and magnitude in  $T_x$ . Conditionally independent would imply an  
618 overall tendency for  $T_x$  and  $T_n$  breakpoints to be of opposite sign such that they  
619 partially or completely cancel in the mean. This raises two interesting questions:  
620 first whether there are more optimal approaches to homogenization than analyzing  
621  $T_m$  as is commonly the case for global centennial timescale LSAT reconstructions to  
622 date; and second why, metrologically, the over-arching tendency may be so.

623

624 5.1 Future homogenization efforts considerations

625

626 Homogenization of surface meteorological station records is inherently a signal-to-  
627 noise problem. Small, relative to meteorological and climatological variability,  
628 breakpoints arising for myriad reasons must be found and then accurately  
629 quantified. Therefore it is important to search in an optimal direction. State of the  
630 art algorithms like PHA perform pairwise comparisons that act to remove common  
631 real-world variations between candidate nearby stations and leave a difference  
632 series that in the absence of any biases in the two comparators should behave as iid  
633 white noise arising from random measurement errors and real inter-site variability.  
634 The white noise places a hard lower limit on signal detectability. No break will be  
635 discoverable that is of comparable magnitude to the standard deviation of the  
636 series. Yet, small breaks arguably matter substantively because they are systematic  
637 effects that do not cancel, so methods should try to optimize breakpoint  
638 detectability and adjustments whilst simultaneously minimizing false alarm rates.  
639 All breakpoint algorithms return bivariate distributions (cf. Figure 3) that in reality  
640 are the two wings of the true Gaussian distribution of real-world breaks with breaks  
641 around zero not being found and / or adjusted for.

642

643 If the breakpoints in  $T_x$  and  $T_n$  were strongly conditionally dependent (similar sign  
644 and magnitude) then searching for breakpoints in  $T_m$  would be quasi-optimal. The  
645 further towards conditional independence of  $T_x$  and  $T_n$  breakpoints the less optimal  
646 use of  $T_m$  series to locate and adjust for breakpoints will become as the dominant

647 direction of breaks becomes increasingly orthogonal to  $T_m$  (Figure 3). Section 3  
648 strongly implies breakpoints are at best random, if not conditionally independent. If  
649 the breakpoints are random then a search should be made in all four elements. If the  
650 breakpoints are mainly conditionally independent then consideration could be  
651 limited to DTR,  $T_x$  and  $T_n$ . Thus in future, homogenization procedures that search  
652 for breakpoints in  $T_m$ ,  $T_x$ ,  $T_n$  and DTR simultaneously will very likely yield a more  
653 accurate and optimal set of breakpoint locations.

654

655 Finding the breakpoints is just the first part of the problem. The resulting  
656 adjustment estimates then need to be reconciled. Here, no such effort has been  
657 made and instead the difference between direct DTR and indirect DTR adjustments  
658 has been used to illustrate potential sensitivities. In future, efforts could be made  
659 given a set of 4 adjustment estimates (or better still conditional density functions of  
660 the adjustments) and a closure condition that the adjustments to  $T_x$  and  $T_n$  must  
661 average to the adjustment of  $T_m$  and difference to the adjustment to DTR to form a  
662 combined set of adjustments. Such an approach is being pursued to develop future  
663 versions of GHCNM.

664

665 All of the above considerations are moot if the station series are only available as  
666  $T_m$ , as is the case for many of the stations in the current databank (Figure 1, lower  
667 panel). Therefore to optimize future analyses of surface temperature changes over  
668 land efforts should be made to recover  $T_x$  and  $T_n$  records for stations and periods of  
669 record for which currently only  $T_m$  records exist in addition to rescuing that data

670 for new stations to improve both coverage and station periods of record [Allan et al.,  
671 2011].

672

673 5.2 Why metrologically may breakpoints in  $T_x$  and  $T_n$  be random or conditionally  
674 independent?

675

676 All meteorological temperature measurements are undertaken by a proxy that is  
677 correlated with the target measurand be that the expansion of liquid, electrical  
678 resistance or some other means. Ideally, the calibration processes for thermometers  
679 would be defined by robust and well documented procedures, under highly  
680 controlled conditions, leading to a full evaluation and definition of calibration  
681 uncertainty components budgets and total values, according to the kind of sensors  
682 used and environments experienced.

683

684 Far from being in thermal adiabatic condition, a thermometer used to measure air  
685 temperature actually measures the mix of convective, radiative and contact heat  
686 transfers. All of these thermodynamic effects are difficult to be corrected with an  
687 uncertainty on the correction. Some devices permitting evaluation of the influence  
688 of such parameters on the sensors under calibration are being developed, but are  
689 still under experimental prototype status [Lopardo et al 2014, Merlone et al. 2014,  
690 Musacchio et al. 2014]. Moreover, since the calibration is performed in stable  
691 temperature conditions, while the measurement of daily air temperature  
692 fluctuations is anything but stable, sensor dynamics can introduce deviations due to

693 the response inertia and delay, not evaluated during calibration. For example, the  
694 behavior of two different thermometers calibrated both in a climatic chamber and in  
695 a liquid bath, was compared to their performance in a Stevenson Screen (CRS)  
696 (Grykalowska, 2014). While both the controlled calibration methods resulted in  
697 consistency within uncertainty, when placed in the Stevenson Screen, the readings  
698 of the two thermometers differed by substantially more than the sum of their  
699 calibration uncertainties, demonstrating that hitherto unaccounted for sensor  
700 dynamics effects remained.

701

702 In the atmosphere there are two critical aspects: the response to heat transfer  
703 effects; and dynamic behavior in capturing temperature fluctuations. Having long  
704 established and recognized the difficulties in estimating the errors induced by these  
705 quantities of influence on the sensors there have been the attempts to reduce the  
706 effects through e.g. screens protecting from direct radiation on the sensing element,  
707 reduced contact surface with the supporting structure, models to minimize the  
708 convective effects, and ventilation to reduce extra heating due to stagnant air. The  
709 range of measurement, shielding and mounting techniques likely yields differing  
710 error characteristics across the meteorological networks, which further are likely to  
711 be climatically dependent.

712

713 In principle, three physical co-variates shall influence the temperature  
714 measurements: radiation, wind speed and humidity. In days with wind blowing and  
715 limited sun radiation these effects are expected to be of low amplitude regardless of

716 instrument configuration whereas in days with sun, absence of wind and larger  
717 night-day temperature fluctuations the effects would be maximal. Such conditions  
718 amplify the possible differences in DTR recording arising from changes in  
719 instrumentation and practices through time.

720

721 There are two broad classes of instrumentation: artificially aspirated and non-  
722 aspirated. Artificially aspirated measurements exhibit substantially lower  
723 sensitivity to prevailing meteorological conditions so long as adequately screened  
724 from direct and indirect radiative effects. They may tend to read slightly high during  
725 daytime due to imperfect shielding from radiation or thermal contact and slightly  
726 low during nighttime due to cooling effects from condensation of the drawn air.

727 Non-aspirated measures will exhibit substantially greater sensitivity to prevailing  
728 meteorological conditions. On average the measures may be warm biased for both  
729  $T_x$  and  $T_n$  due to a mix of radiative and ventilation effects. The biases will be highly  
730 dependent upon configuration and site micro-environment. The change from CRS to  
731 MMTS (both non-aspirated but very distinct) had differential effects on  $T_x$  and  $T_n$   
732 with  $T_x$  decreasing and  $T_n$  increasing. Changing from non-aspirated to aspirated  
733 measurements will tend to yield an apparent and spurious increase in DTR that is  
734 larger than any concurrent change in  $T_m$ .

735

736 5.3 Caveats pertaining to use of current data products

737

738 For analyses of DTR using the dataset constructed herein, the effects of the changing  
739 station availability through time are potentially an insidious effect. The primary  
740 effects are two-fold. Firstly changing the neighbor constraint substantively through  
741 time will affect the efficacy of any homogenization algorithm and PHA is not  
742 immune to this. Secondly, the changing data mask may confound a clean  
743 interpretation of global and regional trends even if the data were perfect (which  
744 they are not). Care should be taken in interpreting pre-1960 records when the  
745 station mix changes substantively both globally and regionally.

746

## 747 **6. Dataset availability**

748

749 The dataset is made available through [website to be appended here once decided,  
750 can we host through NUIM?]. The following series shall be made available:

- 751 • Adjusted station series as CF-compliant netcdf files (one per station)  
752 containing several timeseries fields.
- 753 • Gridded raw and adjusted series for  $T_x$ ,  $T_n$  and DTR (including indirectly  
754 adjusted) as CF-compliant netcdf files (a total of 7 files)

755 At this time there are no plans to update the series beyond 2012. Dataset users  
756 should cite this paper.

757

## 758 **7. Conclusions**

759

760 The present analysis has re-examined changes in DTR globally and regionally using  
761 improved holdings and NCDC's PHA algorithm. Adjustments to the basic 'raw' data  
762 have a non-negligible impact upon the resulting series behavior on multi-decadal  
763 timescales and are comparable in magnitude to the apparent trend in the basic 'raw'  
764 data globally and regionally. DTR is estimated to have decreased globally since the  
765 mid-twentieth Century but the adjustments reduce by half the trend compared to  
766 that in the basic 'raw' data. Both maximum and minimum temperatures have  
767 increased rapidly and changes in these elements are an order of magnitude greater  
768 than in DTR globally. Adjustments are more prevalent in DTR than in  $T_x$  or  $T_n$ ,  
769 which in turn are more common than in  $T_m$ . This implies that overall the biases in  
770  $T_x$  and  $T_n$  are either random or conditionally independent and has potentially  
771 important implications for future homogenization strategies. It implies that  
772 searching for and adjusting breaks in average temperatures is likely to be sub-  
773 optimal as the signal to noise ratio will tend to be a minimum in average  
774 temperatures. Instead efforts that search in addition for breakpoints in DTR,  $T_x$ , and  
775  $T_n$  would likely be more efficient at finding and adjusting for non-climatic artifacts  
776 in the records.

777

778

779 **Acknowledgements**  
780

781 We thank 2 NOAA NCEI internal reviewers for their insights. Fabio Bertiglia,  
782 provided useful input to Section 5.2. Details as to how to ascertain data and  
783 materials are given in Section 6 and can also be attained from the lead author.

784

785

786 **References**

787 Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *Journal of*  
788 *Climatology*, 6, 661-675

789

790 Allan, R. J., et al., 2011: The International Atmospheric Circulation Reconstructions  
791 over Earth (ACRE) initiative. *Bull. Amer. Meteor. Soc.*, **92**, 1421–1425.

792

793 Battisti, D. S., and R. L. Naylor, 2009, Historical warnings of future food security with  
794 unprecedented seasonal heat. *Science*, 323, 240-244

795

796 Bohm, R., P. D. Jones, J. Hiebl, D. Frank, M. Brunetti, and M. Maugeri, 2010: The early  
797 instrumental warm-bias: A solution for long central European temperature series  
798 1760–2007. *Clim. Change*, **101**, 41–67.

799

800 Callendar, G. S. (1938), The artificial production of carbon dioxide and its influence  
801 on temperature. *Q.J.R. Meteorol. Soc.*, 64: 223–240. doi: 10.1002/qj.49706427503

802

803 Christy, J. R., W. B. Norris, and R. T. McNider, 2009: Surface temperature variations  
804 in East Africa and possible causes. *J. Clim.*, **22**, 3342–3356.

805

806 Donat, M. G., L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013a: Global  
807 land-based datasets for monitoring climatic extremes. *Bull. Am. Meteor. Soc.*, **94**,

808 997-1006.

809

810 Falvey, M., and R. D. Garreaud, 2009: Regional cooling in a warming world: Recent  
811 temperature trends in the southeast Pacific and along the west coast of subtropical  
812 South America (1979–2006). *J. Geophys. Res. Atmos.*, **114**, D04102.

813

814 Grykalowska A. et al., The basics for definition of calibration procedure of  
815 temperature sensors for weather station Submitted to Meteorological Application as  
816 MMC2014 Proceedings

817

818 Hawkins, E., and P. D. Jones, 2013: On increasing global temperatures: 75 years after  
819 Callendar *Quarterly Journal of the Royal Meteorological Society*

820

821 Jackson, L. S. and P. M. Forster, 2013: Modeled rapid adjustments in diurnal  
822 temperature range response to CO<sub>2</sub> and solar forcings. *J. Geophys. Res.*, **118**, 2229-  
823 2240, doi: 10.1002/jgrd.50243

824

825 Jain, S. K., and V. Kumar, 2012: Trend analysis of rainfall and temperature data for  
826 India. *Curr. Sci.*, **102**, 37–49.

827

828 Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012:  
829 Hemispheric and large-scale land-surface air temperature variations: An extensive  
830 revision and an update to 2010. *J. Geophys. Res. Atmos.*, **117**, D05127.

831

832 Karl, Thomas R., Philip D. Jones, Richard W. Knight, George Kukla, Neil Plummer,  
833 Vyacheslav Razuvayev, Kevin P. Gallo, Janette Lindsey, and Thomas C. Peterson,  
834 1993: A new perspective on recent global warming: Asymmetric trends of daily  
835 maximum and minimum temperatures. *Bulletin of the American Meteorological*  
836 *Society*, **14**, 1007-1023.

837

838 Karl, T.R., C.N. Williams, Jr., P.J. Young, and W.M. Wendland, 1986: A model to  
839 estimate the time of observation bias associated with monthly mean maximum,  
840 minimum, and mean temperature for the United States, *J. Climate Appl. Meteor.*, **25**,  
841 145-160.

842

843 Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose,  
844 and J. Rennie, 2011: An overview of the Global Historical Climatology Network  
845 monthly mean temperature data set, version 3. *J. Geophys. Res. Atmos.*, **116**, D19121.

846

847 Li, Q., W. Dong, W. Li, X. Gao, P. Jones, J. Kennedy, and D. Parker, 2010: Assessment of  
848 the uncertainties in temperature change in China during the last century. *Chin. Sci.*  
849 *Bull.*, **55**, 1974–1982.

850

851 Lopardo G. et al. 2014. Traceability of Ground-Based Air-Temperature  
852 Measurements: A Case Study on the Meteorological Observatory of Moncalieri  
853 (Italy); International Journal of Thermophysics; available as 'Online First' on  
854 SpringerLink: <http://link.springer.com/article/10.1007/s10765-014-1806-y>

855

856

857 Makowski, K., M. Wild, and A. Ohmura, 2008: Diurnal temperature range over

858 Europe between 1950 and 2005. *Atmos. Chem. Phys.*, **8**, 6483–6498.

859

860 Mastrandrea, M.D., C.B. Field, T.F. Stocker, O. Edenhofer, K.L. Ebi, D.J. Frame, H. Held,

861 E. Kriegler, K.J. Mach, P.R. Matschoss, G.-K. Plattner, G.W. Yohe, and F.W. Zwiers,

862 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on

863 Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change

864 (IPCC). Available at <<http://www.ipcc.ch>>

865

866 McNider, R. T., et al., 2012: Response and sensitivity of the nocturnal boundary layer

867 over land to added longwave radiative forcing. *J. Geophys. Res.*, **117**, D14106.

868

869 Menne, M. J., and C. N. Williams, 2009: Homogenization of temperature series via

870 pairwise comparisons. *J. Clim.*, **22**, 1700–1717.

871

872 Menne, M.J., C.N. Willaims, Jr., and M.A. Palecki, 2010: On the reliability of the U.S.

873 surface temperature record. *J. Geophys. Res.* , doi:10.1029/2009JD013094

874

875 Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An overview

876 of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric*

877 *and Oceanic Technology*, 29, 897-910, doi:10.1175/JTECH-D-11-00103.1.

878

879 Merlone A. et al. 2014, In situ calibration of meteorological sensor in Himalayan high  
880 mountain environment. Submitted to Meteorological Application as MMC2014  
881 Prioceedings

882

883 Musacchio C. et al. METROLOGY ACTIVITIES IN NY-ÅLESUND (SVALBARD),  
884 Submitted to Meteorological Application as MMC2014 Prioceedings

885

886 Parker, D. E., 2006: A demonstration that large-scale warming is not urban. *J. Clim.*,  
887 **19**, 2882–2895.

888

889 Paaijmans, K. P., et al., 2009: Influence of climate on malaria transmission depends  
890 on daily temperature variation. *PNAS*, 107, 15135-15139

891

892 Peng, S. et al., 2013: Asymmetric effects of daytime and night-time warming on  
893 Northern Hemisphere Vegetation. *Nature*, **501**, 88-94, doi:10.1038/nature12434

894

895 Peterson, T. C., K. M. Willett, and P. W. Thorne, 2011: Observed changes in surface  
896 atmospheric energy over land. *Geophys. Res. Lett.*, **38**, L16707.

897

898 Pielke, R. A., and T. Matsui, 2005: Should light wind and windy nights have the same  
899 temperature trends at individual levels even if the boundary layer averaged heat

900 content change is the same? *Geophys. Res. Lett.*, **32**, L21813.

901

902 Rennie, J. J. et al., 2014: The International Surface Temperature Initiative global land  
903 surface databank: monthly temperature data release description and  
904 methods. *Geoscience Data Journal*, doi: 10.1002/gdj3.8

905

906 Rohde, R., et al., 2012: A new estimate of the average Earth surface land temperature  
907 spanning 1753 to 2011. *Geoinfor. Geostat.: An Overview*, **1**.

908

909 Rohde, R., et al., 2013: Berkeley Earth temperature averaging process. *Geoinfor*  
910 *Geostat: An Overview*, **1**.

911

912 Santer, B. D., et al., 2011: Separating signal and noise in atmospheric temperature  
913 changes: The importance of timescale. *J. Geophys. Res. Atmos.*, **116**, D22105.

914

915 Sen Roy, S., and R. C. Balling, 2005: Analysis of trends in maximum and minimum  
916 temperature, diurnal temperature range, and cloud cover over India. *Geophys. Res.*  
917 *Lett.*, **32**, L12702.

918

919 Steeneveld, G. J., A. A. M. Holtslag, R. T. McNider, and R. A. Pielke, 2011: Screen level  
920 temperature increase due to higher atmospheric carbon dioxide in calm and windy  
921 nights revisited. *J. Geophys. Res. Atmos.*, **116**.

922

923 Thorne, P. W., et al 2005: Uncertainties in climate trends - Lessons from upper-air  
924 temperature records. *BAMS* **86**(10): 1437+  
925

926 Thorne, P. W. et al., 2011: "Guiding the Creation of a Comprehensive Surface  
927 Temperature Resource for 21st Century Climate Science.", *Bulletin of the American*  
928 *Meteorological Society*, doi: 10.1175/2011BAMS3124.1  
929

930 Thorne et al. submitted  
931

932 Tietavainen, H., H. Tuomenvirta, and A. Venalainen, 2010: Annual and seasonal  
933 mean temperatures in Finland during the last 160 years based on gridded  
934 temperature data. *Int. J. Climatol.*, **30**, 2247–2256.  
935

936 Trewin, B., 2012: A daily homogenized temperature data set for Australia. *Int. J.*  
937 *Climatol.*, **33**, 1510-1529.  
938

939 van der Schrier, G., E. J. M. van den Besselaar, A. M. G. Klein Tank, and G.  
940 Verver (2013), Monitoring European average temperature based on the E-OBS  
941 gridded data set, *J. Geophys. Res. Atmos.*, **118**, 5120–5135, doi:[10.1002/jgrd.50444](https://doi.org/10.1002/jgrd.50444).  
942

943 Vasseur, D. A., et al ., 2014, Increased temperature variation poses a greater risk to  
944 species than climate warming. *Proc. Roy. Soc. B.*, **281**, 20132612  
945

946 Vautard, R., P. Yiou, and G.J. Van Oldenborgh, 2009: Decline of fog, mist and haze in  
947 Europe over the past 30 years. *Nature Geoscience*, **2**, 115-119, doi:10.1038/ngeo414  
948

949 Venema, V. K. C., et al., 2012: Benchmarking homogenization algorithms for monthly  
950 data. *Clim. Past*, **8**, 89–115.  
951

952 Vincent, L. A., X. L. L. Wang, E. J. Milewska, H. Wan, F. Yang, and V. Swail, 2012: A  
953 second generation of homogenized Canadian monthly surface air temperature for  
954 climate trend analysis. *J. Geophys. Res. Atmos.*, **117**, D18110.  
955

956 Vose, R. S., D. R. Easterling, and B. Gleason, 2005: Maximum and minimum  
957 temperature trends for the globe: An update through 2004. *Geophys. Res. Lett.*, **32**,  
958 L23822.  
959

960 Wang, K. and R. E. Dickinson, 2013: Contribution of solar radiation to decadal  
961 temperature variability over land. *PNAS*, doi:10.1073/pnas.1311433110  
962

963 Wang, G. and M. E. Dillon, 2014, Recent Geographic convergence in diurnal and  
964 annual temperature cycling flattens global thermal profiles. *Nature Climate Change*,  
965 doi: 10.1038/NCLIMATE2378  
966

967 Wijngaard, J. B, A. M. G. Klein-Tank and G. P. Konnen, 2003: Homogeneity of 20<sup>th</sup>  
968 Century European daily temperature and precipitation series. *Int. J. Clim*, 23: 679-  
969 692 doi:10.1002/joc.906  
970

971 Williams, C. N., M. J. Menne, J. H. Lawrimore, 2012a: Modifications to Pairwise  
972 Homogeneity Adjustment software to improve run-time efficiency. NCDC Technical  
973 Report. NCDC No. GHCNM-12-01R  
974 [[http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20Report](http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20Report%20NCDC%20No12-01R-27Jul12.pdf)  
975 [t%20NCDC%20No12-01R-27Jul12.pdf](http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20Report%20NCDC%20No12-01R-27Jul12.pdf)]. Accessed 11/13/14  
976

977 Williams, C. N., M. J. Menne, and J. H. Lawrimore, 2012b: Modifications to Pairwise  
978 Homogeneity Adjustment software to address coding errors and improve run-time  
979 efficiency. NCDC Technical Report. NCDC No. GHCNM-12-02  
980 [[http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20Report](http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20Report%20NCDC%20No12-02-3.2.0-29Aug12.pdf)  
981 [t%20NCDC%20No12-02-3.2.0-29Aug12.pdf](http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20Report%20NCDC%20No12-02-3.2.0-29Aug12.pdf)]. Accessed 11/13/14  
982

983 Williams, C. N., M. J. Menne, and P. W. Thorne, 2012c: Benchmarking the  
984 performance of pairwise homogenization of surface temperatures in the United  
985 States. *J. Geophys. Res. Atmos.*, **117**.  
986

987 Zhou, Y. Q., and G. Y. Ren, 2011: Change in extreme temperature event frequency  
988 over mainland China, 1961–2008. *Clim. Res.*, **50**, 125–139.  
989