

This is a repository copy of *Evaluating biases in sea surface temperature records using coastal weather stations*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/125514/>

Version: Submitted Version

Article:

Cowtan, Kevin Douglas orcid.org/0000-0002-0189-1437, Rohde, Robert and Hausfather, Zeke (2018) Evaluating biases in sea surface temperature records using coastal weather stations. *Quarterly Journal of the Royal Meteorological Society*. pp. 670-681. ISSN 0035-9009

<https://doi.org/10.1002/qj.3235>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Estimating biases in Sea Surface Temperature records using coastal weather stations

Kevin Cowtan^{a*}, Robert Rohde^b, Zeke Hausfather^{b,c}

^a*Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom.*

^b*Berkeley Earth, Berkeley CA 94705.*

^c*University of California Berkeley, Berkeley CA 94720.*

*Correspondence to: Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom. E-mail:

kevin.cowtan@york.ac.uk

Sea surface temperatures form a vital part of global mean surface temperature records, however historical observation methods have changed substantially over time from buckets to engine room intake sensors, hull sensors and drifting buoys, rendering their use for climatological studies problematic. There are substantial uncertainties in the relative biases of different observations which may impact the global temperature record.

Island and coastal weather stations can be compared to coastal sea surface temperature observations to obtain an assessment of changes in bias over time. The process is made more challenging by differences in the rate of warming between air temperatures and sea surface temperatures, and differences across coastal boundaries. A preliminary sea surface temperature reconstruction homogenized using coastal weather station data suggests significant changes to the sea surface temperature record prior to 1980, with substantial uncertainties of which only some can be quantified. The differences to existing records are sufficient in magnitude to have implications for the estimates of climate sensitivity from the historical temperature record, and for the evaluation of internal variability from the difference between the observational record and an ensemble of climate model simulations.

Key Words: sea surface temperature; global mean surface temperature; bucket correction; climate change

Received ...

1 **1. Introduction**

2 Historical estimates of global mean surface temperature are
 3 generally constructed from a blend of land surface air temperature
 4 from weather stations and sea surface temperature (SST) estimates
 5 from ships and buoys. Changes to weather station equipment
 6 have had only a modest effect over the past one and a half
 7 centuries, which can be largely corrected by use of metadata
 8 and interstation comparisons (Menne and Williams Jr. 2009;
 9 Hausfather *et al.* 2016). By contrast sea surface temperatures have
 10 been measured using both canvas and insulated buckets, engine
 11 room intake sensors, ship hull sensors and free floating buoys,
 12 with the different systems measuring temperatures at different
 13 depths (Kent *et al.* 2010). The changing measurement methods
 14 require substantial corrections, the largest of which being the
 15 'bucket correction' of about 0.4°C around the start of the Second
 16 World War.

17 Different approaches have been used to homogenize sea
 18 surface temperature observations. The HadSST3 record from
 19 the UK Hadley Centre makes use of metadata to determine
 20 the most likely method used for a given observation, along
 21 with field replication of measurement methods and reconciliation
 22 of different observation types to correct for the heterogenous
 23 observation systems (Folland and Parker 1995; Rayner *et al.*
 24 2006; Kennedy *et al.* 2011a,b). The COBE-SST2 record (Hirahara
 25 *et al.* 2014) also uses metadata but adopts a different approach
 26 to dealing with observations where metadata is unavailable, with
 27 similar results. By contrast the NOAA Extended Reconstructed
 28 Sea Surface Temperature version 4 (ERSSTv4) product (Huang
 29 *et al.* 2015) makes use of nighttime marine air temperature
 30 (NMAT) observations (Kent *et al.* 2013) as a reference against
 31 which to correct the sea surface temperature observations from
 32 ships.

33 Both methods have limitations: the metadata approach depends
 34 on inference of the observational method for each observation
 35 and the correct determination of the resulting bias. The NMAT
 36 approach depends on the assumption that the NMATs themselves
 37 are unbiased, or at least less biased than the sea surface
 38 temperature observations. Nighttime marine air temperatures are
 39 used because they are less influenced by daytime heating of the

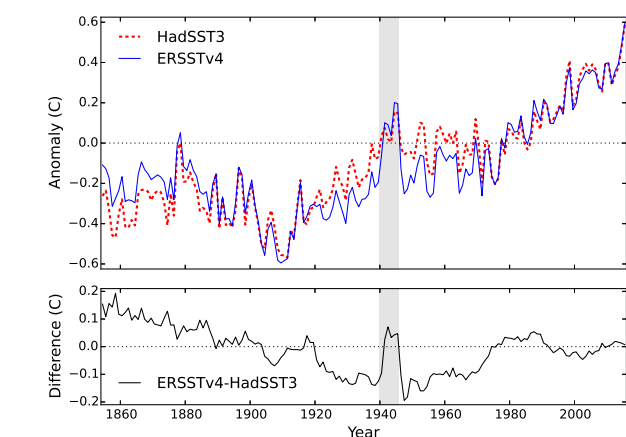


Figure 1. HadSST3 sea surface temperature anomalies with respect to the period 1961-1990, compared to ERSSTv4 aligned to HadSST3 on the period 1981-2010 (top panel), and differences (bottom panel), masked for common spatial coverage.

ship superstructure, however other factors such as the height of the
 deck above sea level also influence nighttime observations. The
 metadata and NMAT methods are largely independent, although
 NMATs have been used indirectly in estimating the prevalence of
 bucket types (Folland and Parker 1995). If both methods produced
 similar results this would increase our confidence in them,
 however in practice there are substantial differences between the
 reconstructions prior to 1980.

The substantial differences between HadSST3 and ERSSTv4
 can be seen in a common coverage comparison of the two records,
 shown in Figure 1, along with the difference between them. The
 records show fairly good agreement from the 1970s to the present.
 However, ERSSTv4 is significantly cooler than HadSST3 over the
 period 1920-1970, except for the World War 2 period (shown in
 the shaded area of Figure 1). ERSSTv4 is warmer than HadSST3
 prior to 1890 and shows further divergence earlier in the 19th
 century.

The differences around World War 2 are particularly striking,
 with ERSSTv4 showing a large spike in temperatures while
 HadSST3 shows only a modest peak. A drop in the number
 of observations coupled with changing data sources makes this
 period particularly problematic (Kennedy *et al.* 2011b). While
 ship-based measurements were greatly impacted by the war, land-
 based observations were less disrupted. Previous research has
 taken advantage of the more homogeneous land record during this
 period; for example Folland (2005) uses land temperatures and
 climate models to estimate the bias in bucket observations, while
 (Thompson *et al.* 2008) detected an inhomogeneity in the sea

68 surface temperature record arising from a change in the shipping
 69 fleet at the end of World War 2 by comparison of sea surface
 70 temperatures to temperatures from coastal weather stations and
 71 from climate models.

72 Similarly, Parker *et al.* (1995) and Rayner *et al.* (2003) used
 73 data from weather stations located on islands to assess the
 74 homogeneity of the sea surface temperature observations from
 75 ships passing close to those islands. Since ships are mobile
 76 platforms which can move between open ocean and coastal
 77 waters, a bias in the observations close to shore will generally also
 78 correspond to a bias in open ocean observations.

79 This paper will provide a preliminary evaluation of the
 80 use of island and coastal weather stations for the automatic
 81 homogenization of sea surface temperatures across the whole
 82 period of the sea surface temperature record. The existing
 83 HadSST3 sea surface temperature record (Kennedy *et al.* 2011a,b)
 84 will be compared to quality controlled coastal and island weather
 85 station data from version 4(beta) of the Global Historical
 86 Climatology Network-Monthly (GHCN-M v4) (Lawrimore *et al.*
 87 2011), and the differences used to correct the sea surface
 88 temperature record. The process is complicated by the presence
 89 of a climate signal in the difference in temperature between the
 90 sea surface and marine air temperatures (Cowtan *et al.* 2015),
 91 and differences in temperature on crossing the coastal boundary,
 92 which must be taken into account.

93 A distinction is generally made between sea surface
 94 temperature (SST) of the surface ocean waters, marine air
 95 temperature (MAT) of the air at the ocean surface, and land
 96 surface air temperature (LSAT) as observed by weather stations.
 97 These will be assumed to refer to non-coastal regions, and the
 98 new terms coastal SST (CSST), coastal marine air temperature
 99 (CMAT) and coastal land surface air temperature (CLSAT) will
 100 be used for coastal regions. The differences between MAT and
 101 SST will be referred to as air-water difference. The difference
 102 between SST and CSST will be referred to as inshore difference.
 103 The differences between CMAT and CLSAT will be referred to as
 104 coastal difference. The difference between CLSAT and LSAT will
 105 be referred to as inland difference. Not all of these are resolvable
 106 in either models or observations due to the limited resolution of
 107 climate models and limited spatial coverage of the observations.

Temperatures will all be expressed in terms of anomalies
 with respect to the 1961-1990 baseline of HadSST3. As a result
 absolute temperature differences are ignored and only differences
 in temperature change between different types of observations will
 be discussed.

2. Change in coastal land surface air temperature as an indicator of sea surface temperature

The use of weather stations to assess inhomogeneities in SST
 assumes that change in land surface air temperature measured
 by coastal weather stations is a good indication of change in
 sea surface temperature, and this assumption must be evaluated.
 Globally, land warms faster than oceans, and so it is possible
 that coastal air temperatures might overestimate sea surface
 temperature change. Coastal air temperatures are less variable
 than temperatures in continental interiors, so land based weather
 stations will be most useful if they are sufficiently close to the
 coast. Island weather stations may be particularly useful in this
 regard.

To evaluate the utility of coastal land-based weather stations to
 estimate coastal sea surface temperatures, surface air temperatures
 were examined for the high resolution GFDL-HiRAM C360
 model runs, which are reported on a fine ~ 30 km grid (Harris
et al. 2016). Atmospheric Model Intercomparison Project-style
 historical experiments are available for the period 1979-2008,
 which is characterized by rapid greenhouse warming. Sea
 surface temperatures ('tos' in CMIP nomenclature), surface air
 temperatures ('tas'), and the land mask ('sftlf') are all available
 on the same grid (Taylor *et al.* 2012). Two runs of this model are
 available.

In order to determine whether land-based weather stations
 can give an indication of marine air temperature, the trend in
 the difference between surface air temperature and sea surface
 temperature (i.e. tas-tos) was examined while crossing coastal
 boundaries. No sea surface temperatures are available for pure
 land cells, however the variation in temperature difference can be
 examined as a function of increasing land fraction in cells with up
 to 99% land.

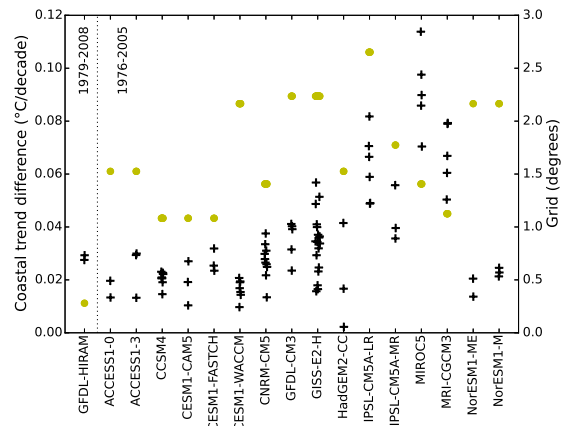
A map of the trend in the difference between tas and tos was
 calculated over the period 1979-2008 for every cell for which both

147 values were present. Every pair of adjacent cells between 60S and
 148 60N in the trend map were compared. For every pair of adjacent
 149 cells where both trend values were present and the land fraction
 150 in the two cells was different, the difference in trend and the
 151 difference in land fraction were calculated. Ordinary least squares
 152 regression was used to determine the contribution of increasing
 153 land fraction difference to increasing trend difference.

154 The data show an increase in tas-tos trend when moving
 155 from the cell with 0% land to a cell with 100% land (Figure
 156 S1). The coefficient of determination in the regression is small
 157 ($R^2 \sim 0.03$), suggesting that geographical variability is large
 158 compared to the coastal effect. The t-value of the prediction
 159 is large ($t \sim 35$); however it is likely to be overestimated due
 160 to spatial autocorrelation. The best indication of uncertainty in
 161 the regression coefficient therefore comes from repeating the
 162 experiment with different runs of the same climate model. The
 163 values of the coefficient of the land fraction difference in the
 164 regression are $0.028^\circ\text{C}/\text{decade}$ and $0.029^\circ\text{C}/\text{decade}$ for the two
 165 runs of the HiRAM model.

166 These values are about 20% of the sea surface temperature trend
 167 for the study period. However, the 30 km cells used in the HiRAM
 168 model are large compared to typical distances between a coastal
 169 weather station and the sea. In practice a coastal weather station is
 170 likely to be characterized by a grid cell which is part ocean, so the
 171 actual land effect on the air temperature trend may be less than
 172 this. If the ratio of land air to sea surface temperature change is
 173 roughly constant over time the land surface air temperatures can
 174 simply be scaled to address the impact of the coastal effect.

175 The same calculation was repeated for a selection of CMIP5
 176 historical simulations (described in Table S1) for which the
 177 appropriate fields were available. CMIP5 model runs typically
 178 use different grids for the land and ocean data, and so the sea
 179 surface temperatures were first transferred onto the surface air
 180 temperature grid using inverse distance weighting. Historical runs
 181 typically end in 2005, so the period 1986-2005 was used. The
 182 CMIP5 model grids are generally much coarser than the HiRAM
 183 grid (typically 100-200km), and so the air temperatures of high
 184 land fraction coastal cells will sample regions further inland than
 185 for the HiRAM model.



186 **Figure 2.** Coastal 30 year temperature trend differences for different climate
 187 models. Black crosses indicate the regression coefficient between the trend
 188 difference and the sea fraction between neighbouring cells with different land
 189 fractions for individual runs of a given model. Spots indicate the average of the
 190 latitude and longitude dimensions of a grid cell for that model at the equator for that
 191 model, with the scale on the right hand axis.

186 The trend and regression calculations were repeated for each
 187 model, with the results shown in Figure 2. There is significant
 188 variation between models, with the GISS-E2-H model showing
 189 a rather higher coastal effect than the HiRAM runs. Given that
 190 the coastal difference in air temperature trend moving from sea
 191 to land is non-negligible, the coastal weather stations will require
 192 adjustment before they are used to homogenize the sea surface
 193 temperature data. The coastal trend difference appears to increase
 194 roughly linearly with cell land fraction, and so a scaling should be
 195 applied to the weather station data which is linearly dependent on
 196 the land fraction around the weather station.

197 3. Coastal weather station record

198 A coastal weather station record was constructed using the
 199 GHCN-M v4 temperature data (Lawrimore *et al.* 2011), which
 200 uses data from the International Surface Temperature Initiative
 201 (Rennie *et al.* 2014) and includes data from 26,182 weather
 202 stations. The raw data were used in preference to the homogenized
 203 data, because (a) homogenization is expected to be of limited
 204 use for isolated island stations, and (b) homogenization may
 205 potentially increase coastal trends and reduce inland trends in
 206 order to bring them into agreement.

207 Information on station environment is not currently included
 208 in the GHCN-M version 4 data, and so coastal and island stations
 209 were identified using using a quarter degree global land mask from
 210 (Jet Propulsion Laboratory 2013). Stations north of 60N or south
 211 of 60S were omitted to avoid the effects of sea ice, and stations in

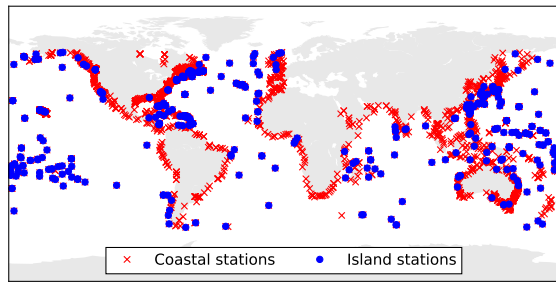


Figure 3. Map of coastal and island weather stations from GHCNv4 used in the construction of the coastal weather station record. Crosses show coastal stations, while dots show the subset of stations which are included on the island list on the basis of the land fraction in the surrounding cells.

212 the Baltic and Mediterranean region were omitted since these may
 213 not reflect the global oceans. Stations were also omitted which
 214 lie more than 10km from the nearest coast according to metadata
 215 from Mosher (2017). Two station selections were used:

- 216 1. An island station list, consisting of 428 stations for which
 217 the average land fraction for the 8 cells surrounding the
 218 cell containing the weather station was less than 10%.
 219 By chance all 428 stations fall in cells for which the
 220 land fraction is recorded as zero, however the station list
 221 provides no coverage prior to the 1920s.
- 222 2. A coastal station list, consisting of 2386 stations for which
 223 the land fraction in the station cell was less than 50% or the
 224 land fraction in one of the four orthogonally adjacent cells
 225 was 0%. Some stations are available back to the start of the
 226 HadSST3 data in 1850. The coastal station list is a superset
 227 of the island station list.

228 The two station selections are shown in Figure 3.

229 To address the different warming rates of coastal air and
 230 sea surface temperatures, the temperature observations for each
 231 station were scaled according to equation 1, in accordance with
 232 the climate model results.

$$T_{scaled} = T_{anom}(a - bl(\phi, \lambda)) \quad (1)$$

233 T_{anom} is the original temperature anomaly, T_{scaled} is the scaled
 234 anomaly, $l(\phi, \lambda)$ is the land fraction in the given grid cell and a
 235 and b are coefficients whose determination will be described later.

236 The station records for the selected stations are first aligned
 237 using the Climatic Anomaly Method (Jones 1994), using a
 238 baseline period of 1961-1990 for consistency with HadSST3.

For stations with at least 25 months of data present in the 30
 year baseline period for a given month of the year, temperature
 anomalies were determined by subtracting a constant from all
 data for that month of the year to bring the mean on the baseline
 period to zero. If insufficient months of data were available, data
 were not used for that station for that month of the year. Data
 for 851 of the 2386 coastal stations were aligned in this way. A
 gridded temperature field was then calculated from the initial set
 of temperature anomalies, using a 5×5 degree grid.

A limitation of the climatic anomaly method is that stations
 or months cannot be used if insufficient data are available during
 the baseline period, reducing the number of available station
 records. Additional stations were therefore added iteratively by
 determining the offset required for each month of the year to fit
 the new station to the initial stations by the following method:
 The scaled station anomalies in each grid cell were averaged for
 each month of the record. The resulting sparse temperature field
 was extended to global coverage using kriging (Cressie 1990)
 following the method of (Cowtan and Way 2014). Anomalies
 were calculated for additional stations for which at least 15
 months of data were available during the baseline period by
 fitting them to the temperature record for the appropriate grid
 cell, yielding 1328 aligned stations. A second global temperature
 field was determined from the expanded station list. In a third
 step, anomalies were calculated for further additional stations for
 which 15 months of data were available at any time between
 1850 and the present by fitting them to the temperature record
 for the appropriate grid cell, yielding 2196 aligned stations. A
 spatially incomplete coastal temperature field was calculated from
 the resulting anomalies.

4. Coastal station homogenization of the sea surface temperature record

In addition to the corrected HadSST3 record, Kennedy *et al.* also
 distribute raw sea surface temperature fields with no adjustments
 for instrument type. The coastal weather station record was used to
 determine a time dependent (and optionally spatially dependent)
 correction to the raw sea surface temperature observations to
 bring them into agreement with the scaled coastal weather station
 record. The correction field is based on the difference field

278 between the (sparse) coastal weather station field and the raw sea
 279 surface temperature field. In order to ensure maximum coverage,
 280 the more complete sea surface temperature field was first infilled
 281 by kriging using the method of Cowtan and Way.

282 Both air-sea and coastal temperature differences can be
 283 influenced by weather (for example due to the greater heat
 284 capacity of the ocean), and so the differences between the coastal
 285 weather station and sea surface temperature anomaly fields show
 286 significant spatial and month-on-month variability. The correction
 287 to the raw sea surface temperatures must therefore be averaged
 288 both spatially, and over a moderate time window.

289 The HadSST3 corrections are spatially relatively uniform over
 290 most of the record, except for the periods where the sea surface
 291 temperatures come primarily from buckets, when there is a
 292 significant zonal variation in the bias arising from the varying air-
 293 sea temperature differential with latitude (Kent *et al.* 2016). The
 294 primary component of the zonal variation is a contrast between
 295 the tropics and higher latitudes, however during some periods
 296 (such as the late 1940s) hemispheric differences are also apparent
 297 due to differences in the shipping fleets in different regions.
 298 This suggests that the correction might be modelled by some
 299 combination of the zonally invariant spherical harmonics, Y_{00} , Y_{01}
 300 and Y_{02} :

- 301 1. Y_{00} is a constant field. Fitting Y_{00} is equivalent to fitting the
 302 global mean of the correction field.
- 303 2. Y_{01} changes sign between the hemispheres, and so captures
 304 hemispheric differences.
- 305 3. Y_{02} changes sign between the equator and the poles, and
 306 so captures differences between the tropics and the higher
 307 latitudes.

308 In the early record, the available weather stations are clustered
 309 in developed regions with varying concentrations, and so a
 310 naive fitting method would overweight the regions with more
 311 observations. To address this issue, the spherical harmonics
 312 were fitted to the coastal difference map using generalised
 313 least squares (GLS), which includes information about the
 314 expected covariances of the observations in order to weight each
 315 observation according to the amount of independent information it
 316 provides. The covariance matrix of observations was constructed

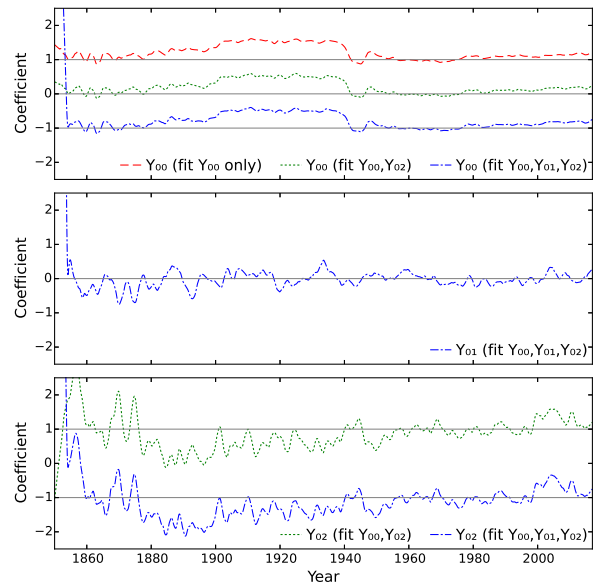


Figure 4. Smoothed coefficients of the spherical harmonics Y_{00} , Y_{01} and Y_{02} used in fitting the coastal temperature difference map for each month of the record, after application of a 36 month lowess smooth. Three different models are fitted, the first using just Y_{00} ; the second using Y_{00} and Y_{02} , and the third using Y_{00} , Y_{01} and Y_{02} . Each panel shows a single coefficient, with the model indicated in the key and the coefficients offset to allow comparison of the lines.

317 as an exponentially-declining function of distance in the same way
 318 as the variogram in Cowtan and Way, with an e-folding range
 319 of 800km determined empirically from the data over the period
 320 1981-2010 when the coastal stations have good geographical
 321 coverage.

322 Three different models are fitted, the first using just Y_{00} ; the
 323 second using Y_{00} and Y_{02} , and the third using Y_{00} , Y_{01} and Y_{02} .
 324 The coefficients for each spherical harmonic in each model are
 325 shown as a function of time in Figure S2. The Y_{00} (global mean)
 326 coefficient suggests a cool bias in the raw sea surface temperatures
 327 relative to the coastal air temperatures in the decades prior to
 328 World War 2, and to a lesser extent in the decade following the
 329 end of World War 2, consistent with previous analyses. This bias is
 330 apparent even without temporal smoothing of the coefficient. The
 331 Y_{01} and Y_{02} coefficients show rather greater monthly variability
 332 which is of a similar or greater amplitude to any persistent signal,
 333 and show very large excursions in the earliest decade of the record.

334 The coefficients were therefore smoothed using a 36 month
 335 linear lowess smooth with a cubic window (Cleveland 1979),
 336 chosen to provide the most smoothing possible without distorting
 337 the World War 2 feature in the Y_{00} coefficient (Figure 4). The
 338 smoothed Y_{01} (hemispheric) coefficient still does not display
 339 a persistent signal, however the Y_{02} (equator-pole) coefficient
 340 tends to be negative in the periods dominated by canvas bucket

341 observations (1880-1940 and 1945-1950) (Folland and Parker
 342 1995), and positive in the 21st century when buoy observations
 343 become dominant. Prior to 1880 the Y_{02} coefficient shows large
 344 excursions, arising from most of the available coastal temperature
 345 data being confined to the mid latitudes.

346 The detection of the uninsulated bucket signal both in the global
 347 mean coastal bias (i.e. Y_{00}), and in the zonal distribution (Y_{02})
 348 provides support for the use of coastal temperature differences
 349 in the detection of sea surface temperature biases. However the
 350 zonal distribution signal only becomes apparent with temporal
 351 smoothing, which suggests that the method is already approaching
 352 its limits in terms of the isolation of geographical components of
 353 the coastal temperature difference.

354 Once the fit to the difference field has been determined, the
 355 spherical harmonics are then scaled by the fitted coefficients to
 356 determine a global correction field, which is then added to the
 357 raw HadSST3 field to produce a corrected sea surface temperature
 358 record. The corrected record is dependent on the values of a
 359 and b which scale the coastal temperature anomalies to account
 360 for the differential warming rates across the air-sea and coastal
 361 boundaries. Values for a and b are determined by assuming that
 362 the trend in the coastal temperature difference over the period
 363 1981-2010 is dominated by the warming signal, justified by the
 364 rapid warming over this period and the comparatively limited
 365 metadata based corrections identified by Kennedy *et al.* (2011b).
 366 The HadSST3 trend is therefore assumed to be correct over this
 367 period, and the coefficients a and b determined such that the
 368 global mean of the temperatures in the co-located corrected field
 369 yields the same trend. The island stations have $l(\phi, \lambda) = 0$ for all
 370 stations, and so can be used to determine a value for a , giving
 371 $a = 0.99$. A value for b is then determined such that the trend
 372 in the corrected record using the coastal station list also matches
 373 the HadSST3 trend, giving $b = 0.58$. The coefficients a and b do
 374 not vary significantly with the introduction of additional spherical
 375 harmonics to the regression.

376 The temperature field resulting from adding the correction field
 377 to the raw HadSST3 temperature field for each month will be
 378 referred to as a coastal hybrid sea surface temperature. The mean
 379 sea surface temperature for each month was then calculated from
 380 the mean of the cells for which HadSST3 observations were

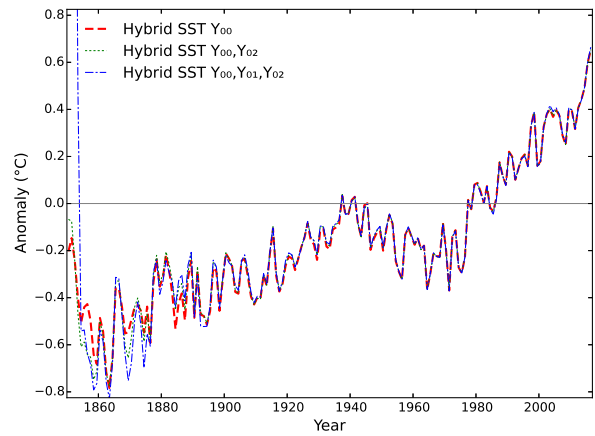


Figure 5. Coastal hybrid temperature reconstructions determined by fitting the coastal temperature difference map for each month of the record and using the resulting model to correct the sea surface temperature field. Three different models are fitted, the first using just Y_{00} ; the second using Y_{00} and Y_{02} , and the third using Y_{00} , Y_{01} and Y_{02} .

available, weighting each grid cell according to the area of the cell. 381
 The annual means using one, two or three spherical harmonics 382
 were then plotted for the whole period of the record (Figure 5). 383

The number of spherical harmonics makes essentially no 384
 difference to the resulting geographical means after 1900, and 385
 little difference between 1880 and 1900. However in the earliest 386
 decades, the inclusion of additional spherical harmonics increases 387
 the annual variability in the record. The remainder of this study 388
 will therefore focus primarily on the most parsimonious model 389
 where only the global mean of the coastal difference map (Y_{00}) is 390
 fitted; this will also allow the sensitivity of the results to different 391
 subsets of the coastal temperature record to be evaluated. 392

5. Results 393

Global marine temperature reconstructions were determined using 394
 the coastal hybrid method fitting a single global term to the 395
 coastal temperature difference field, and applying the 36 month 396
 lowess smooth to the resulting coefficients. Two temperature 397
 reconstructions were calculated as follows: 398

1. A reconstruction from HadSST3 using just the island 399
 stations. 400
2. A reconstruction from HadSST3 using the full list of 401
 coastal stations. 402

The resulting fields were masked to common coverage with 403
 the HadSST3 dataset before calculation of an area weighted 404
 monthly mean temperature series for each reconstruction. The 405
 island temperature series begins in 1920 due to limited island 406

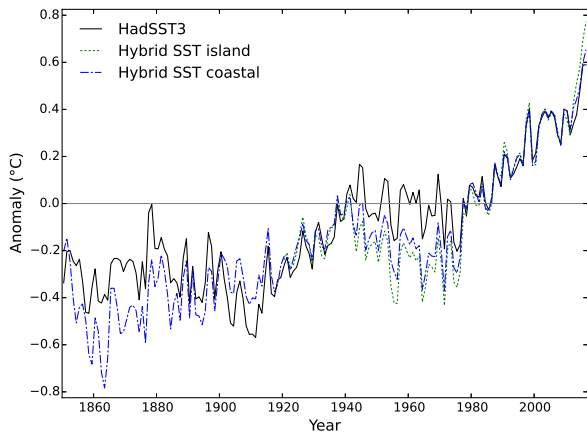


Figure 6. Comparison of two versions of the coastal hybrid temperature record to HadSST3. The two hybrid records use only island stations to correct HadSST3 over the period 1920-2016, or all coastal stations to correct HadSST3 over the period 1850-2016.

407 station coverage. Annual means were calculated from the monthly
 408 series, and compared to HadSST3 in Figure 6. Both of the coastal
 409 hybrid reconstructions show a cooler mid 20th century plateau
 410 than HadSST3. The coastal reconstruction rejects the coolness of
 411 the first two decades of the 20th century found in existing SST
 412 datasets and also suggests a cooler 19th century.

413 5.1. Sensitivity of the hybrid SSTs to the coastal temperature 414 record

415 If the corrections to the sea surface temperature arise from
 416 global biases in the observational platforms and procedures,
 417 they should be detectable across the globe rather than arising
 418 from just one region. To test this the calculation was repeated
 419 omitting a hemisphere of data from the coastal difference
 420 field. The generalized least squares calculation reconstructs the
 421 missing hemisphere with the optimal average of the remaining
 422 hemisphere. The calculation was performed ten times, omitting
 423 the northern hemisphere, the southern hemisphere and eight
 424 hemispheres centered on points on the equator separated by 45
 425 degrees of longitude. The resulting ensemble of 10 reconstructions
 426 is compared to HadSST3 in Figure 7. The ensemble members
 427 show cooler temperatures for most of the mid 20th century
 428 plateau, but are spread around HadSST3 in the 1930s. The
 429 ensemble members show warmer temperatures around 1910, and
 430 cooler temperatures in the mid 19th century. The ensemble is
 431 somewhat bimodal in the late 19th century, with some members
 432 much cooler than and others similar to HadSST3.

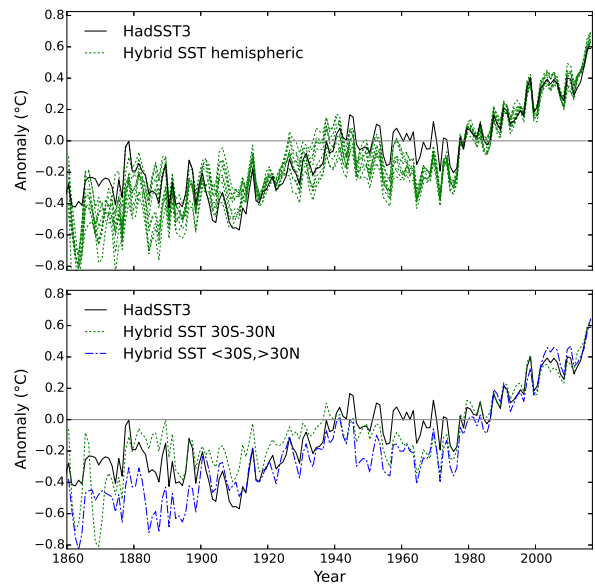


Figure 7. Coastal hybrid temperature reconstructions using different subsets of the coastal weather station record. The correction field is determined by fitting the Y_{00} coefficient to each of ten hemispheric subsets of the coastal difference field (top panel), or to just the equatorial or mid latitude cells of the coastal difference field (lower panel).

Global sea surface temperature reconstructions based on just 433
 the equatorial or mid latitude data show a somewhat greater 434
 contrast, with the mid-latitude data showing a cooler mid-century 435
 plateau than the equatorial data (which is still cooler than 436
 HadSST3). The bucket bias is greatest at the equator, and so 437
 correction using mid latitude data leads to a smaller correction 438
 and therefore cooler temperatures than HadSST3 in the 19th 439
 century, while the tropical data lead to a reconstruction which 440
 is similar to or slightly warmer than HadSST3 for most of the 441
 early period. Prior to 1880, the tropical data are very sparse so the 442
 coastal hybrid record is likely to be cool biased due to the lower 443
 corrections from the mid-latitude stations. 444

The coastal hybrid temperature reconstruction is strongly 445
 determined by the coastal weather station record, which is in turn 446
 dependent on both the station selection (which has already been 447
 explored through the island-only record and the hemispheric and 448
 zonal subsets), and the scale terms a and b which account for the 449
 difference in warming rate between sea surface temperatures and 450
 weather stations with different degrees of exposure to the sea. 451
 Since only an ad-hoc estimate of the values of these parameters 452
 is available, the sensitivity of the resulting record to those values 453
 must be explored. 454

Reducing the parameter a (which controls the scaling of all 455
 weather stations relative to coastal sea surface temperatures) while 456

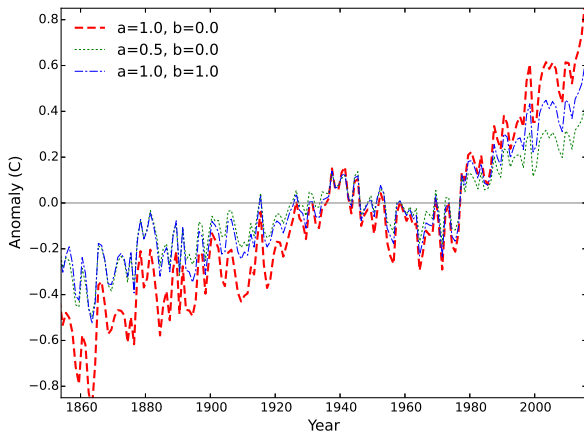


Figure 8. Comparison of hybrid temperature reconstructions using different values of the weather station scaling parameters a and b .

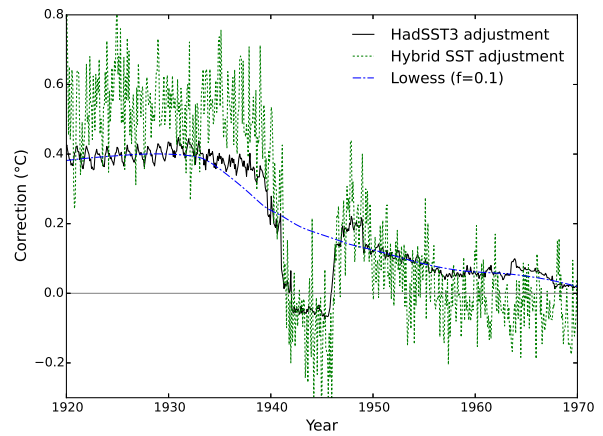


Figure 9. Comparison of the corrections applied to the raw sea surface temperature reconstruction by either the hybrid coastal method, or by the metadata-based HadSST3 method. The dashed line is a lowess smooth through the HadSST3 corrections, smoothed to emulate the smoothing used in the ERSSTv4 algorithm.

457 holding b at zero (or more generally, scaling a and b together),
 458 reduces the amount of warming fairly uniformly across the whole
 459 record (Figure 8). Thus a misestimation of a could lead to a
 460 misestimation of the total amount of warming since the 19th
 461 century, but the resulting record would maintain its shape, still
 462 showing a cooler mid century plateau and no dip around 1910.
 463 Increasing the b parameter leads to reduced warming prior to
 464 World War 2 but has a rather smaller effect on late 20th century
 465 warming. This behaviour arises from the sparsity of island stations
 466 in the early record, hence the b term which controls for the inland
 467 effect of less exposed stations plays a greater role.

468 The dependence of the coastal hybrid record on a novel
 469 temperature reconstruction using raw rather than homogenized
 470 temperature data must also be considered. Hybrid coastal
 471 temperature reconstructions were therefore determined using the
 472 existing CRUTEM version 4 and GHCN version 3 gridded
 473 temperature fields (Jones *et al.* 2012; Lawrimore *et al.* 2011),
 474 using a single scale factor in each case to preserve the trend in the
 475 resulting record on the period 1981-2010 (Figure S3). Using the
 476 CRUTEM data produces a coastal hybrid record which is broadly
 477 similar to that obtained using the custom coastal weather station
 478 record.

479 If the GHCN gridded data are used the resulting record shows
 480 significantly more warming prior to 1970. Part of this difference
 481 can be explained by the automated homogenization used in the
 482 GHCN record, because a hybrid reconstruction using the GHCN
 483 version 4 homogenized data also shows more early warming
 484 (Figure S4). The GHCN version 3 based record would imply an

implausible sign change in the bucket bias in the early period; this
 is more likely to arise from the GHCN homogenization algorithm
 not accounting for the different rates of warming of coastal and
 inland stations. The remaining differences probably arise from the
 smaller weather station inventory for GHCN version 3 compared
 to GHCN version 4, and changes in the mix of coastal and non-
 coastal stations in the large 5×5 degree cell used by the GHCN
 gridded data.

5.2. World War 2

The ERSSTv4 and HadSST3 records show a large discrepancy
 during World War 2, with ERSSTv4 showing substantial warmth
 over most of the conflict, while HadSST3 shows only a modest
 warm period spanning two to three years. To assess this period
 a coastal hybrid record was constructed with no temporal smoothing.
 The resulting adjustments to the raw record are compared to the
 corresponding metadata-based HadSST3 adjustments in Figure 9.

Without the temporal smoothing term the adjustments from
 the coastal hybrid method show greater inter-monthly variability,
 however the shape of the adjustment matches the metadata-based
 HadSST3 adjustments well. The size of the adjustment suggested
 by the coastal hybrid method is larger than that for HadSST3, and
 falls outside the range of the 100 member HadSST3 ensemble
 (Kennedy *et al.* 2011b). The similarity in shape provides a
 validation of both the metadata assignments of observation type in
 HadSST3, and the utility of the coastal hybrid method in detecting
 that bias.

511 The discrepancy in the size of the World War 2 bias between
 512 HadSST3 and the hybrid record could arise from non-uniformity
 513 in the zonal distribution of coastal observations, given the latitude
 514 dependence of the bucket bias. To test this possibility the World
 515 War 2 period was also examined in reconstructions based on
 516 hemispheric subsets of the coastal temperature data, or on the
 517 tropical or mid-latitude data alone (Figure 10). The use of a
 518 hemispheric or zonal subset of the coastal stations can lead to an
 519 estimate of the post-war bias which is larger or smaller than the
 520 HadSST3 estimate. As expected the equatorial data lead to a larger
 521 estimate of the pre-war bias than the mid-latitude data, however in
 522 both cases the estimated bias is larger than in HadSST3.

523 The wartime warmth in the ERSSTv4 reconstruction arises
 524 from a failure to correct for the sharp changes in bias during
 525 this period. ERSSTv4 applies a lowess smooth to the difference
 526 between the SST and NMAT data to determine the bias correction
 527 using a window of 10% (i.e. about 200 months) of the data. The
 528 same smoothing operation applied to the HadSST3 adjustments is
 529 shown in Figure 9: the smoothed correction does not capture the
 530 World War 2 bias. Both the metadata adjustments of HadSST3
 531 and the coastal hybrid method reject the World War 2 warmth
 532 in ERSSTv4, and the smoothing term provides a sufficient
 533 explanation for the bias. Removal of the smoothing step may
 534 therefore resolve the bias in ERSSTv4, contingent on there being
 535 no corresponding wartime bias in the NMAT data.

536 5.3. The post-1998 “hiatus” period

537 The ERSSTv4 and HadSST3 records also show a difference in
 538 trend over the period since 1997, which while smaller is relevant to
 539 discussions of a “hiatus” in warming. Karl *et al.* (2015) reject the
 540 existence of a hiatus on the basis of the larger trends in ERSSTv4.
 541 Hausfather *et al.* (2017) find independent support for the higher
 542 trend in ERSSTv4 in SST records constructed using homogeneous
 543 observation platforms to address the inhomogeneities in the
 544 observational record.

545 Temperature trends for co-located observations from HadSST3,
 546 ERSSTv4, and from the coastal and island hybrid records
 547 constructed from the raw HadSST3 data without smoothing are
 548 given in Table 1 for the period from 1997 to 2016. Hausfather
 549 *et al.* (2017) note that the uncertainty in the trends is dominated by

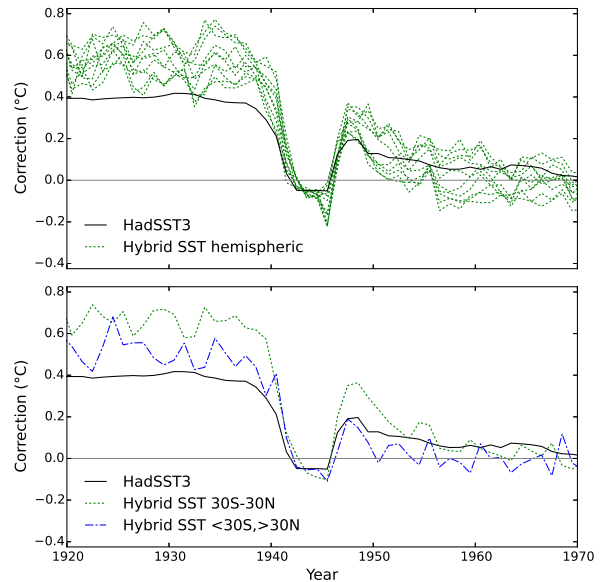


Figure 10. Coastal hybrid temperature corrections for the World War 2 period, using different subsets of the coastal weather station record. The correction field is determined by fitting the Y_{00} coefficient to each of ten hemispheric subsets of the coastal difference field (top panel), or to just the equatorial or mid latitude cells of the coastal difference field (lower panel).

Table 1. Temperature trends for the period 1997-2016 for the common coverage of the HadSST3, hybrid, and ERSSTv4 records. Trends of the difference series against HadSST3 are given with the corresponding uncertainties.

Dataset	Trend (1997-2016) (°C/decade)	Trend difference with HadSST3
HadSST3	0.081	
Hybrid/coastal	0.106	0.025 ± 0.016
Hybrid/island	0.129	0.048 ± 0.018
ERSSTv4	0.111	0.030 ± 0.010

the weather signal, and is therefore not a measure of the structural
 uncertainty in the trend, and that the uncertainties in the trends
 in the difference series should therefore be used to assess the
 trend significance. The coastal and island hybrid records both
 show higher trends which are closer to ERSSTv4 than HadSST3,
 consistent with the results of Hausfather *et al.* While the coastal
 hybrid record does not reject either the ERSSTv4 or HadSST3
 trend at the 95% confidence level, the island hybrid record does
 reject the HadSST3 trend at the 95% confidence level.

6. Discussion

The homogenization of the sea surface temperature record
 is challenging, owing to a constantly changing fleet of
 mobile observation platforms and variability in the observation
 protocols. Both metadata and external temperature data sources
 have been used to homogenize the data by HadSST3 and
 ERSSTv4 respectively, with differing results. Coastal weather

566 stations provide an alternative and independent check on those
 567 homogenizations, but are subject to uncertainties and biases due
 568 to the temperature differences across coastal boundaries as well as
 569 any uncorrected biases in the weather station observations.

570 This study presents a preliminary attempt at the use of
 571 coastal weather station records to correct inhomogeneities in
 572 the sea surface temperature record. The challenges of removing
 573 the climate signal from the coastal temperature differences are
 574 substantial, and so the results should be considered an indication
 575 of possible problems in existing series rather than a definitive
 576 temperature history. The new record suggests, in decreasing order
 577 of confidence, that:

- 578 1. The World War 2 warm spike in ERSSTv4 is spurious. The
 579 coastal temperature data support the shape of the meta-data
 580 based correction of HadSST3, providing evidence for the
 581 wartime corrections. The coastal temperature data suggest
 582 more tentatively that the size of the correction (due to a
 583 transition between bucket and engine room observations) is
 584 slightly underestimated in HadSST3.
- 585 2. The mid-century plateau spanning the 1940s to the 1970s
 586 is cooler in the coastal hybrid record than in HadSST3.
 587 This supports the cooler temperatures of ERSSTv4 over
 588 this period. The same result is obtained when using all of
 589 the coastal weather station data or spatially distinct subsets
 590 of the data.
- 591 3. The larger estimate of the size of the bucket correction
 592 in the coastal hybrid record leads to a greater upward
 593 correction of pre-World War 2 temperatures, leading to
 594 warmer temperatures since 1900 and an earlier start to the
 595 mid-century plateau. The large dip in temperatures around
 596 1910 in existing records is largely eliminated in the coastal
 597 hybrid record.
- 598 4. The rate of warming in HadSST3 since 1998 is likely to
 599 be underestimated, consistent with previous work showing
 600 less warming in ship observations over that period than in
 601 more reliable buoy measurements (Kennedy *et al.* 2011b;
 602 Karl *et al.* 2015; Hausfather *et al.* 2017).
- 603 5. The coastal hybrid record is also cooler than existing
 604 records between 1880 and 1900, however this result is

contingent on the station selection, with some subsets of 605
 the data yielding temperatures similar to HadSST3. 606

6. The coastal hybrid record shows cooler temperatures 607
 between 1850 and 1880 than the existing SST records. 608
 However coastal weather station coverage in the tropics is 609
 poor and island station coverage non-existent during this 610
 period. 611

The sparsity of data in the tropics in the earliest part of the 612
 record presents a problem in estimating the bias in the sea surface 613
 temperature observations due to the zonal dependence of the air- 614
 water temperature difference. When the Y_{02} coefficient is included 615
 in the model, the resulting temperature record only shows 616
 significantly different behaviour prior to about 1880 (Figure 5). 617

While the coastal hybrid method is likely to have a cool bias at 618
 the start of the record, the agreement of the different spherical 619
 harmonic models after 1880 point is consistent with the cool bias 620
 being confined to the period prior to that date. 621

The coastal hybrid record is compared to co-located data from 622
 both HadSST3 and ERSSTv4 in Figure 11, and shows significant 623
 differences with both. The existing records disagree over the 624
 warmth of the mid 20th century plateau with ERSSTv4 being 625
 cooler than HadSST3, however the hybrid record is cooler than 626
 either. The hybrid record rejects the warm spike in ERSSTv4 627
 during World War 2. The hybrid record is broadly consistent 628
 with HadSST3 between 1915 and 1935, however it rejects 629
 the unexplained cool period between 1900 and 1915 in the 630
 existing records. Prior to 1900 HadSST3 is generally cooler than 631
 ERSSTv4, however the hybrid record is cooler than either. 632

The late 19th century and early 20th century periods are 633
 of particular interest, with the coastal hybrid record showing a 634
 gradual warming which is more consistent with climate model 635
 simulations than the existing records. The bucket bias is estimated 636
 by Folland and Parker (1995) to increase linearly from 1850 to 637
 1920, however the coastal hybrid suggests a bias which remains 638
 small until around 1890 and then increases rapidly until 1910. 639

The differences between the coastal hybrid and existing sea 640
 surface temperature reconstructions are not necessarily indicative 641
 of problems in the existing records, although divergence between 642
 the existing records means that both cannot be correct. The coastal 643
 record may be more realistic if the coastal weather station record 644

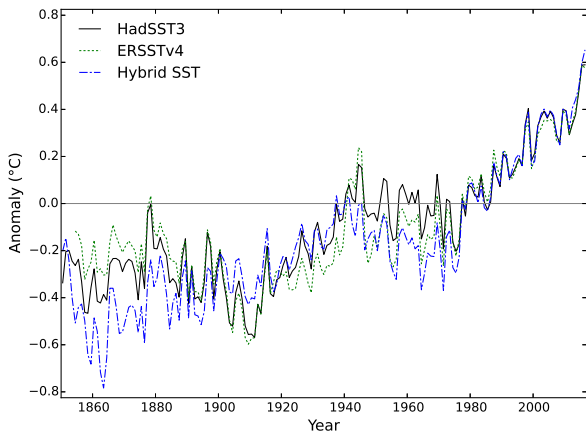


Figure 11. Comparison of the coastal hybrid temperature reconstruction (using all coastal stations and fitting the global mean of the coastal temperature differences only) to co-located data from HadSST3 and ERSSTv4 for the period 1850–2016.

645 is reliable and if the relationship between coastal air temperature
646 and offshore sea surface temperature is correctly modelled.

647 Possible problems with the coastal temperature record include
648 changing weather station coverage and the use of raw rather than
649 homogenized temperature data. For the period after 1920, the
650 similarity of the hybrid record when using the more strict island
651 station selection provides additional support for the results, but does
652 not address the earlier period. Use of homogenized data in the
653 preparation of the coastal hybrid record leads to *much* greater
654 warming in the 19th century, however this is unlikely to be correct
655 because it would require a change in the sign of the bucket bias.
656 It is more likely that homogenization exaggerates the trend for
657 coastal stations.

658 The differences between the coastal hybrid record and
659 HadSST3 could arise from changes in the air-sea temperature
660 difference, inshore temperature difference or coastal temperature
661 difference which are not accounted for by the simple scaling
662 scheme of equation 1. The inshore temperature difference may be
663 partially captured in the HadSST3 record due to the presence of
664 vessels traversing coastal waters, however the large 5×5 degree
665 grid cells may offset this. Uncertainties in the scaling of the coastal
666 weather station data relative to sea surface temperatures will affect
667 the evaluation of long term changes in sea surface temperature
668 bias, but not rapid changes like those around World War 2 or in
669 the 1970s.

670 It is notable that there are large changes in difference between
671 HadSST3 and the coastal hybrid reconstruction in the 1940s and
672 the late 1970s, corresponding roughly to changes in the sign of the

Pacific Decadal Oscillation (PDO). While the corresponding wind 673
changes may affect inshore or coastal temperature differences, the 674
coastal corrections are largely conserved between hemispheres 675
so cannot be driven by Pacific variability alone. Furthermore 676
the ERSSTv4 record also shows a somewhat cooler mid-century 677
plateau without the use of coastal temperatures, suggesting that the 678
PDO on its own cannot explain all of the differences between the 679
coastal hybrid and HadSST3 records. 680

681 Given the inherent uncertainties it would be premature to adopt
682 the coastal hybrid record as a historical record of sea surface
683 temperature. The limited spatial resolution of the correction limits
684 the utility of the record for estimating temperatures at a sub-
685 global scale, and the changing station coverage in the 19th century
686 certainly biases the record prior to 1880. Metadata-based analyses
687 like that of HadSST3 still provide the best tools for evaluating
688 regional sea surface temperature variation, however it is possible
689 that the approach presented here may provide a useful tool in
690 improving the parameterisation of the metadata-based corrections.

691 If the coastal hybrid record were correct, there would be
692 implications both for the estimation of climate sensitivity and
693 for the assessment of multidecadal internal variability from the
694 historical temperature record. Estimates of climate sensitivity
695 which rely on a 19th century temperature baseline (Otto *et al.*
696 2013; Richardson *et al.* 2016) would be too low due to the
697 warm bias in the early sea surface temperature record. Differences
698 between temperature observations and the mean of an ensemble
699 of climate model simulations are often attributed to internal
700 variability in the real climate system, because internal variability
701 is expected to cancel out when averaging multiple simulations.
702 Observation-model differences typically show a peak in the late
703 19th century, a dip in the early 20th century (Mann *et al.* 2016).
704 Both of these are reduced if the coastal hybrid record is used in
705 place of existing records, which might suggest a reduced role
706 for multidecadal internal variability in the observed temperature
707 record.

708 The consequences for the climate sensitivity and internal
709 variability highlight the importance of possible inhomogeneities
710 in the sea surface temperature record. The differences between
711 existing sea surface temperature reconstructions demonstrate that
712 there is a problem to be addressed. The coastal hybrid sea

713 surface temperature reconstruction cannot solve this problem
 714 outright because the results are contingent on correctly combining
 715 inhomogeneous observations across coastal boundaries; however
 716 the method does bring an additional source of observational data
 717 to help assess the biases in the sea surface temperature record.

718 Data and methods for this paper are available at [doi://](https://doi.org/10.1002/2015GL067640)
 719 TBA with updates at [http://www-users.york.ac.uk/](http://www-users.york.ac.uk/~kdc3/papers/estimating2017)
 720 [~kdc3/papers/estimating2017](http://www-users.york.ac.uk/~kdc3/papers/estimating2017).

721 Acknowledgements

722 We acknowledge the World Climate Research Programme's
 723 Working Group on Coupled Modelling, which is responsible for
 724 CMIP, and we thank the climate modeling groups (listed in Table
 725 S1) for producing and making available their model output. For
 726 CMIP the U.S. Department of Energy's Program for Climate
 727 Model Diagnosis and Intercomparison provides coordinating
 728 support and led development of software infrastructure in
 729 partnership with the Global Organization for Earth System
 730 Science Portals.

731 We are grateful to S. Mosher for providing metadata for
 732 the GHCNv4 station inventory, and K. Haustein for helpful
 733 discussion.

734 References

735 Cleveland W. 1979. Robust locally weighted regression and smoothing
 736 scatterplots. *Journal of the American Statistical Association* **74**(368): 829–
 737 836, doi:10.1080/01621459.1979.10481038.
 738 Cowtan K, Hausfather Z, Hawkins E, Jacobs P, Mann M, Miller S,
 739 Steinman B, Stolpe M, Way R. 2015. Robust comparison of climate
 740 models with observations using blended land air and ocean sea surface
 741 temperatures. *Geophysical Research Letters* **42**(15): 6526–6534, doi:10.
 742 1002/2015GL064888.
 743 Cowtan K, Way R. 2014. Coverage bias in the HadCRUT4 temperature series
 744 and its impact on recent temperature trends. *Quarterly Journal of the Royal*
 745 *Meteorological Society* **140**(683): 1935–1944, doi:10.1002/qj.2297.
 746 Cressie N. 1990. The origins of kriging. *Mathematical geology* **22**(3): 239–
 747 252, doi:10.1007/BF00889887.
 748 Folland C. 2005. Assessing bias corrections in historical sea surface
 749 temperature using a climate model. *International Journal of Climatology*
 750 **25**(7): 895–911, doi:10.1002/joc.1171.
 751 Folland C, Parker D. 1995. Correction of instrumental biases in historical sea
 752 surface temperature data. *Quarterly Journal of the Royal Meteorological*
 753 *Society* **121**(522): 319–367, doi:10.1002/qj.49712152206.

Harris L, Lin SJ, Tu C. 2016. High-resolution climate simulations using GFDL
 HiRAM with a stretched global grid. *Journal of Climate* **29**(11): 4293–
 4314, doi:10.1175/JCLI-D-15-0389.1. 754 755 756
 Hausfather Z, Cowtan K, Clarke DC, Jacobs P, Richardson M, Rohde R. 757
 2017. Assessing recent warming using instrumentally homogeneous sea 758
 surface temperature records. *Science Advances* **3**(1): e1601207, doi:10. 759
 1126/sciadv.1601207. 760
 Hausfather Z, Cowtan K, Menne M, Williams CN J. 2016. Evaluating the 761
 impact of U.S. historical climatology network homogenization using the 762
 U.S. climate reference network. *Geophysical Research Letters* **43**(4): 1695– 763
 1701, doi:10.1002/2015GL067640. 764
 Hirahara S, Ishii M, Fukuda Y. 2014. Centennial-scale sea surface temperature 765
 analysis and its uncertainty. *Journal of Climate* **27**(1): 57–75, doi:10.1175/ 766
 JCLI-D-12-00837.1. 767
 Huang B, Banzon V, Freeman E, Lawrimore J, Liu W, Peterson T, Smith 768
 T, Thorne P, Woodruff S, Zhang HM. 2015. Extended reconstructed 769
 sea surface temperature version 4 (ERSST.v4). part i: Upgrades and 770
 intercomparisons. *Journal of Climate* **28**(3): 911–930, doi:10.1175/JCLI- 771
 D-14-00006.1. 772
 Jet Propulsion Laboratory. 2013. ISLSCP II land and water masks with 773
 ancillary data. <http://daac.ornl.gov/>, doi:10.3334/ORNLDAA/ 774
 1200. Oak Ridge National Laboratory Distributed Active Archive Center, 775
 Oak Ridge, Tennessee, USA. 776
 Jones P. 1994. Hemispheric surface air temperature variations: a reanalysis 777
 and an update to 1993. *Journal of Climate* **7**(11): 1794–1802, doi:10.1175/ 778
 1520-0442(1994)007<1794:HSATVA>2.0.CO;2. 779
 Jones P, Lister D, Osborn T, Harpham C, Salmon M, Morice C. 2012. 780
 Hemispheric and large-scale land-surface air temperature variations: An 781
 extensive revision and an update to 2010. *Journal of Geophysical Research* 782
Atmospheres **117**(5), doi:10.1029/2011JD017139. 783
 Karl T, Arguez A, Huang B, Lawrimore J, McMahon J, Menne M, Peterson 784
 T, Vose R, Zhang HM. 2015. Possible artifacts of data biases in the 785
 recent global surface warming hiatus. *Science* **348**(6242): 1469–1472, doi: 786
 10.1126/science.aaa5632. 787
 Kennedy J, Rayner N, Smith R, Parker D, Saunby M. 2011a. Reassessing 788
 biases and other uncertainties in sea surface temperature observations 789
 measured in situ since 1850: 1. measurement and sampling uncertainties. 790
Journal of Geophysical Research Atmospheres **116**(14), doi:10.1029/ 791
 2010JD015218. 792
 Kennedy J, Rayner N, Smith R, Parker D, Saunby M. 2011b. Reassessing 793
 biases and other uncertainties in sea surface temperature observations 794
 measured in situ since 1850: 2. measurement and sampling uncertainties. 795
Journal of Geophysical Research Atmospheres **116**(14), doi:10.1029/ 796
 2010JD015220. 797
 Kent E, Kennedy J, Berry D, Smith R. 2010. Effects of instrumentation 798
 changes on sea surface temperature measured in situ. *Wiley Interdisci-* 799
plinary Reviews: Climate Change **1**(5): 718–728, doi:10.1002/wcc.55. 800

- 801 Kent E, Rayner N, Berry D, Saunby M, Moat B, Kennedy J, Parker D. 485–498, doi:10.1175/BAMS-D-11-00094.1. 848
- 802 2013. Global analysis of night marine air temperature and its uncertainty Thompson D, Kennedy J, Wallace J, Jones P. 2008. A large discontinuity in the 849
- 803 since 1880: The HadNMAT2 data set. *Journal of Geophysical Research* mid-twentieth century in observed global-mean surface temperature. *Nature* 850
- 804 *Atmospheres* **118**(3): 1281–1298, doi:10.1002/jgrd.50152. **453**(7195): 646–649, doi:10.1038/nature06982. 851
- 805 Kent EC, Berry DI, Carella G, Kennedy JJ, Parker DE, Atkinson CP, Rayner
- 806 NA, Smith TM, Hirahara S, Huang B, *et al.* 2016. A call for new approaches
- 807 to quantifying biases in observations of sea-surface temperature. *Bulletin*
- 808 *of the American Meteorological Society* **0**(0): null, doi:10.1175/BAMS-D-
- 809 15-00251.1, URL [http://dx.doi.org/10.1175/BAMS-D-15-](http://dx.doi.org/10.1175/BAMS-D-15-00251.1)
- 810 00251.1.
- 811 Lawrimore J, Menne M, Gleason B, Williams C, Wuertz D, Vose R, Rennie J.
- 812 2011. An overview of the global historical climatology network monthly
- 813 mean temperature data set, version 3. *Journal of Geophysical Research*
- 814 *Atmospheres* **116**(19), doi:10.1029/2011JD016187.
- 815 Mann M, Rahmstorf S, Steinman B, Tingley M, Miller S. 2016. The likelihood
- 816 of recent record warmth. *Scientific Reports* **6**, doi:10.1038/srep19831.
- 817 Menne M, Williams Jr C. 2009. Homogenization of temperature series via
- 818 pairwise comparisons. *Journal of Climate* **22**(7): 1700–1717, doi:10.1175/
- 819 2008JCLI2263.1.
- 820 Mosher S. 2017. Station metadata for GHCN version 4. Personal
- 821 communication.
- 822 Otto A, Otto F, Boucher O, Church J, Hegerl G, Forster P, Gillett N, Gregory
- 823 J, Johnson G, Knutti R, Lewis N, Lohmann U, Marotzke J, Myhre G,
- 824 Shindell D, Stevens B, Allen M. 2013. Energy budget constraints on climate
- 825 response. *Nature Geoscience* **6**(6): 415–416, doi:10.1038/ngeo1836.
- 826 Parker D, Folland C, Jackson M. 1995. Marine surface temperature: Observed
- 827 variations and data requirements. *Climatic Change* **31**(2-4): 559–600, doi:
- 828 10.1007/BF01095162.
- 829 Rayner N, Brohan P, Parker D, Folland C, Kennedy J, Vanicek M, Ansell
- 830 T, Tett S. 2006. Improved analyses of changes and uncertainties in sea
- 831 surface temperature measured in situ since the mid-nineteenth century:
- 832 The HadSST2 dataset. *Journal of Climate* **19**(3): 446–469, doi:10.1175/
- 833 JCLI3637.1.
- 834 Rayner N, Parker D, Horton E, Folland C, Alexander L, Rowell D, Kent E,
- 835 Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and
- 836 night marine air temperature since the late nineteenth century. *Journal of*
- 837 *Geophysical Research D: Atmospheres* **108**(14): ACL 2–1 – ACL 2–29.
- 838 Rennie J, Lawrimore J, Gleason B, Thorne P, Morice C, Menne M, Williams C,
- 839 Almeida WG, Christy J, Flannery M, *et al.* 2014. The international surface
- 840 temperature initiative global land surface databank: Monthly temperature
- 841 data release description and methods. *Geoscience Data Journal* **1**(2): 75–
- 842 102, doi:10.1002/gdj3.8.
- 843 Richardson M, Cowtan K, Hawkins E, Stolpe M. 2016. Reconciled climate
- 844 response estimates from climate models and the energy budget of earth.
- 845 *Nature Climate Change* **6**(10): 931–935, doi:10.1038/nclimate3066.
- 846 Taylor K, Stouffer R, Meehl G. 2012. An overview of CMIP5 and the
- 847 experiment design. *Bulletin of the American Meteorological Society* **93**(4):