1   **Development of the HadISDH marine humidity climate monitoring dataset**

2   Kate Willett[1], Robert Dunn[1], John Kennedy[1] and David Berry[2]

4   [1]Met Office Hadley Centre, Exeter, UK

5   [2]National Oceanography Centre, Southampton, UK

7   *Correspondence to:* Kate Willett kate.willett@metoffice.gov.uk

9   **Abstract**

11   Atmospheric humidity plays an important role in climate analyses. Here we describe the production and key

12   characteristics of a new quasi-global marine humidity product intended for climate monitoring,

13   HadISDH.marine. It is an in-situ based multi-variable marine humidity product, gridded monthly at a 5° by 5°

14   spatial resolution from January 1973 to December 2018 with annual updates planned. Currently, only reanalyses

15   provide up to date estimates of marine surface humidity but there are concerns over their long-term stability. As

16   a result, this new product makes a valuable addition to the climate record and will help address some of the

17   uncertainties around recent changes (e.g. contrasting land and sea trends, relative humidity drying). Efforts have

18   been made to quality control the data, ensure spatial and temporal homogeneity as far as possible, adjust for

19   known biases in non-aspirated instruments and ship heights, and also estimate uncertainty in the data.

20   Uncertainty estimates for whole-number reporting and for other measurement errors have not been quantified

21   before for marine humidity. This is a companion product to HadISDH.land, which, when combined will provide

22   methodologically consistent land and marine estimates of surface humidity.

24   The spatial coverage of HadISDH.marine is good over the Northern Hemisphere outside of the high latitudes but

25   poor over the Southern Hemisphere, especially south of 20° S. The trends and variability shown are in line with

26   overall signals of increasing moisture and warmth over oceans from theoretical expectations and other products.

27   Uncertainty in the global average is larger over periods where digital ship metadata are fewer or unavailable but

28   not large enough to cast doubt over trends in specific humidity or air temperature. Hence, we conclude that

29   HadISDH.marine is a useful contribution to our understanding of climate change. However, we note that our

30   ability to monitor surface humidity with any degree of confidence depends on the continued availability of ship

31   data and provision of digitised metadata.

33   HadISDH.marine data, derived diagnostics and plots are available at

34   www.metoffice.gov.uk/hadobs/hadisdh/indexMARINE.html (Willett et al., 2019).

36   **1 Introduction**

38   Water vapour plays a key role as a greenhouse gas, in the dynamical development of weather systems, and

39   impacts society through precipitation and heat stress. Over land, all these aspects are important and recent

40   changes have been assessed by Willett et al. (2014). Over the oceans, a major source of moisture over land, a

41   similar analysis is essential to enhance our understanding of the observed changes generally and as a basis for

42   worldwide evaluation of climate models. In recognition of its importance, the surface atmospheric humidity has

43   been recognised as one of the Global Climate Observing System (GCOS) Essential Climate Variables (ECVs).

45   Observational sources of humidity over the ocean are limited. The NOCSv2.0 (Berry and Kent, 2011) is the

46   only recently updated (January 1971 to December 2015) marine surface humidity monitoring product based on

47   in-situ observations, but it only includes specific humidity (q). Satellite based humidity products exist (e.g.

48   HOAPS, Fennig et al., 2012) but these rely on the in-situ observations for calibration. Whilst quasi-global, the

49   uncertainties in the NOCv2.0 product are large outside the northern, mid latitudes. In these regions the

50   NOCSv2.0 product shows a reasonably steadily rising trend over the period of record, similar to that seen over

51   land but with slightly different year-to-year variability. Most notably, 2010, a peak year over land in specific

52   humidity, does not stand out over ocean. Figure 1 and Willett et al. (2019) show global land and ocean specific

53   humidity and relative humidity (RH) series from available in-situ and reanalyses products. Older, static

54   products for the ocean (HadCRUH: Willett et al., 2008; Dai: Dai 2006) show increasing specific humidity to

55   2003 with similar variability to NOCSv2.0, and near-constant relative humidity. Both HadCRUH and Dai show

56   a positive relative humidity bias pre-1982 and slightly higher specific humidity over1978-1984 compared to

57   NOCSv2.0. There is broad similarity between the reanalysis products and the in-situ products but with notable

58   differences for specific humidity in the scale of the 1998 peak and the overall trend magnitude. Differences are

59   to be expected given that the reanalyses are spatially complete in coverage, albeit derived only from their

60   underlying dynamical models over data sparse regions. The reanalyses exhibit near-constant to decreasing

61   relative humidity over oceans but with poorer agreement between both the reanalyses themselves and compared

62   to the in-situ products over land. This is to be expected given the larger sources of bias and error over ocean

63   (Sect. 2) and sparse data coverage. Importantly, land and marine specific humidity appear broadly similar

64   whereas for relative humidity, the distinct drying since 2000 over land is not apparent over ocean in reanalyses

65   and the previously available in-situ products finish too early to be informative. Note that the HadISDH.marine

66   described herein is shown here for comparison and will be discussed below.

67

68   A positive bias in global marine average relative humidity pre-1982 is apparent in Dai and HadCRUH, and has

69   previously been attributed to high frequencies of whole numbers in the dew point temperature observations prior

70   to January 1982 (Willett et al., 2008). This is less clear in the global average specific humidity timeseries.

71   ICOADS (International Comprehensive Ocean-Atmosphere Dataset) documentation

72   (http://icoads.noaa.gov/corrections.html) notes issues with the pre-1982 data especially mixed-precision

73   observations, where the air temperature has been recorded to decimal precision but the dew point temperature is

74   only available as a whole number. Such reporting was in accordance with the WMO Ship Code before 1982.

75   The documentation notes a truncation error in the dew point depression which would lead to a positive bias in

76   relative humidity. Alternatively, Berry (2009) show that patterns in the North Atlantic Oscillation coincide with

77   this time period and could have played a role. The NOCSv2.0 product is based on reported wet bulb temperature

78   rather than dew point temperature, where decimal precision is usually present. Hence, the NOCSv2.0 product is

79   expected to be unaffected by these rounding issues. Our analysis shows that changes to the code in January 1982

80   did not eliminate whole number reporting and high frequencies of whole numbers can be found throughout the

81   record in both air temperature and dew point temperature (Sect. 2.4 and Sect. 3.4).

82

83   Clearly, there is a need for more and up to date in-situ monitoring of humidity over ocean, especially for RH.

84   The structural uncertainty in estimates can only be explored if there are multiple available estimates so a new

85   product that explores different methodological choices, and extends the record, is complementary to the existing

86   NOCSv2.0 product and reanalyses estimates. Here we report the development of a multi-variable marine

87   humidity analysis HadISDH.marine.1.0.0.2018f (Met Office Hadley Centre; National Oceanography Centre,

88   2019 FINALISED AFTER REVIEW). This forms a companion product to the HadISDH.land monitoring

89   product, enabling the production of a blended global land and ocean product. We use existing methods where

90   possible from the systems used for building the long running HadSST dataset (Kennedy et al., 2011a, 2011b,

91   2019), and also use some of the bias adjustment methods employed for NOCSv2.0 (Berry and Kent 2011). We

92   have explored the data to design new humidity specific processes where appropriate, particularly in terms of

93   quality control and gridding.

94

95   HadISDH.marine is a climate-quality 5° by 5° gridded monthly mean product from 1973 to present (December

96   2018 at time of writing) with annual updates envisaged. Fields will be presented for surface (~10 m) specific

97   humidity, relative humidity, vapour pressure, dew point temperature, wet bulb temperature and dew point

98   depression. Air temperature will also be made available as a by-product but less attention has been given to

99   addressing temperature specific biases. The product is intended for investigating long-term changes over large

100   scales and so efforts have been made to quality control the data, ensure spatial and temporal homogeneity, adjust

101   for known biases and also estimate remaining uncertainty in the data. In particular, we estimate uncertainties

102   from whole-number reporting and other measurement errors that have not been quantified before for marine

103   humidity.

104

105   Section 2 discusses known issues with marine humidity data. Section 3 describes the source data and all

106   processing steps. Section 4 presents the gridded product and explores the different methodological choices and

107   comparison with NOCSv2.0 specific humidity and ERA-Interim marine humidity. This section also includes a

108   first look at the blended land and marine HadISDH product for each variable. Section 5 covers data availability

109   and Section 6 concludes with a discussion of the strengths and weaknesses of the product.

110

111   **2 Known issues affecting the marine humidity data**

112

113   **2.1 Daytime solar-biases**

114

115   Marine air temperature measurements on board ships during the daytime are known to be affected by the heating

116   of the ship or platform by the sun. This results in a positive bias during daylight and early night time hours. The

117   bias varies with sunlight strength/cloudiness (and thus also latitude), relative wind speed, size and material of

118   the ship. This solar heating bias affects both the wet bulb and dry bulb temperature measurements but, as noted

119   by Kent and Taylor (1996), the ships do not act as a source of humidity or change the humidity content of the

120 air. As a result, biases in the specific humidity and dew point temperature due to the solar heating errors will be

121 negligible. However, care needs to be taken with relative humidity because estimates of the saturation vapour

122 pressure from the uncorrected dry bulb air temperature will be too high, leading to an underestimate in relative

123 humidity. Ideally, relative humidity should be estimated using the corrected dry-bulb temperature to calculate

124 the saturation vapour pressure and uncorrected wet and dry bulb temperature or dew point temperature to

125 calculate the vapour pressure.

126

127 Previously, efforts have been made to bias-adjust the air temperature observations for solar heating by

128 modelling the extra heating over the superstructure of the ship, taking account of the relative wind speed,

129 cloudiness, time of day, time of year and latitude (Kent et al, 1993; Berry et al., 2004; Berry and Kent, 2011).

130 These adjustments are complex and so we have decided not to attempt to implement them for our first version of

131 a marine humidity product given the wide variety of other issues we have accounted for. We have, however,

132 produced daytime, night time and combined products to investigate differences that may be caused by the solar

133 heating bias. Later versions of HadISDH.marine that apply bias corrections for solar heating may reduce the

134 amount of daytime data removed.

135

136 **2.2 Un-aspirated psychrometer bias**

137

138 Humidity measurements can be made in a variety of ways. Instruments can be housed in a screen with

139 ventilation slats, with or without additional artificial aspiration, or handheld in a sling or whirling psychrometer.

140 There is information on instrument ventilation provided up to 2014. Approximately 30 % of ship observations

141 have information in 1973, peaking at ~75 % by the mid-1990s, as summarised in Fig. 2. Initially, slings were

142 more common for the hygrometer and thermometer, but by 1982 a screen was more common. There is a

143 tendency for the screened instruments, in the absence of artificial aspiration, to give a wet bulb reading that is

144 higher relative to the slings/whirling instruments where airflow is ensured by the whirling motion. Bias

145 adjustments have been applied to un-aspirated humidity observations by Berry and Kent (2011), building on

146 previous bias adjustments of Josey et al. (1999) and Kent et al. (1993). They have also estimated the uncertainty

147 in the bias adjustments. We implement a modified version of their method of bias adjustment for the un-

148 aspirated observation types (Sect. 3.3.1) and uncertainty estimation. Uncertainties from instrument bias

149   adjustments will have some spatial and temporal correlation structure as the ships move around (Kennedy et al.,

150   2011a).

151

152   **2.3 Ship height inhomogeneity**

153

154   Over time there has been a general trend for ship heights to increase. Kent et al. (2007; 2013) quantified the

155   increase from an average of ~ 16m in 1973 to ~24m by the end of 2006. Instrument height information is

156   available for some ships between the period of 1973 and 2014, providing heights for the barometer (HOB),

157   thermometer (HOT), anemometer (HOA) and visual observing platform (HOP). Figure 3 shows the availability

158   of height information and the mean and standard deviation of heights per year in each category for the ship

159   observations selected here. Similar to the ventilation metadata, height information availability is low in 1973,

160   peaking mid-1990s to 2000 and then declining slightly. Prior to 1994 only the platform height was available

161   from WMO Publication 47. This was replaced in 1994 by the barometer height and augmented with the

162   thermometer and visual observing heights from 2002 onwards (Kent et al., 2007). Anemometer heights have

163   been available from WMO 47 since 1970. All four types of heights increase over time. We conclude that the

164   mean height based on HOP/HOB/HOT increases from 17 m in 1973 to 23 m by 2014, which differs slightly to

165   that in Kent et al., (2007). If uncorrected, this likely leads to a small artificial decreasing trend in air temperature

166   and specific humidity, as, in general, these variables decrease with height away from the surface. The effect on

167   relative humidity is less clear and depends on the relative effects on air temperature and specific humidity.

168

169   Prior studies (e.g. Berry and Kent, 2011; Berry 2009; Josey et al., 1999; Rayner et al., 2003; Kent et al., 2013)

170   have applied height adjustments to the air temperature, specific humidity and wind speed measurements to

171   adjust the measurements to a common reference height and minimise the impact of the changing observing

172   heights on the climate record. These have been based on boundary layer theory and the bulk formulae, using the

173   parameterisations of Smith (1980, 1988). In the absence of high-frequency observations of meteorological

174   parameters for each observation location, allowing direct estimation of the surface fluxes, parameterisations

175   have to be made and an iterative approach is necessary to estimate a height adjustment (Sect. 3.3.2). We have

176   followed these previous approaches and estimated height adjustments for all observations and variables of

177   interest. Where observing heights are unavailable we have made new estimates (Sect. 3.3.2). We have also

178   provided an estimate of uncertainty on these height adjustments, which are larger where we have also estimated

179 the height of the observation. The uncertainties from height adjustments will have some spatial and temporal

180 correlation structure.

181

182 **2.4 Whole-number reporting biases**

183

184 Recording and reporting formats and practices have changed many times over the 20th century, affecting the

185 climate record. Some formats required the wet bulb temperature to be reported, others the dew point temperature

186 and some allowed either or both (https://www.wmo.int/pages/prog/amp/mmop/documents/publications-

187 history/history/SHIP.html). Some earlier formats restricted space to reporting temperature to whole numbers

188 only and this practice has continued with some ships continuing to report the dew point (or wet bulb)

189 temperature and sometimes even the dry bulb temperature to whole numbers. A practice of truncation of the

190 dew point depression has been noted for the pre-1982 data (http://icoads.noaa.gov/corrections.html) which

191 would result in spuriously high humidity (both in relative and actual terms). It is clear from the

192 ICOADS3.0.0/3.0.1 data that there has been a practice of reporting values to whole numbers rather than decimal

193 places, both for air temperature and dew point temperature. Rounding dew point temperature and air

194 temperature could result in a +/- 0.5° C error individually or a just less than +/- 1° C error in dew point

195 depression for a worst-case scenario combination.

196

197 Whole-number reporting is an issue throughout the record for both variables – a breakdown of air and dew point

198 temperature by decimal place over time is shown in Fig. S1. Air temperature also shows a disproportionate

199 frequency of half degrees (5s). The prevalence of whole numbers (0s) declines over time, dramatically in the

200 mid- to late 1990s for air temperature and from 2008 for both air and dew point temperature. This decline in the

201 1990s, and in part also the general decline, appears to be linked to an increase in numbers of moored buoys (see

202 Fig. 5), a similar analysis without the moored buoys (not shown) shows greater consistency over time. The dew

203 point temperature has two distinct peaks in whole number frequency in the 1970s and mid-1990 to early 2010s.

204 The latter peak is more pronounced when moored buoys are not included. The early peak is somewhat

205 consistent with the restriction in transmission space prior to January 1982. This was previously thought to have

206 been a possible cause of higher relative humidity over the period 1973-1981 compared to the rest of the record

207 in the HadCRUH marine relative humidity product (Willett et al., 2008). The pre-1982 moist bias was also

208 apparent in the global marine relative humidity product of Dai (2006). The NOCSv2.0 product preferentially

209 utilises the wet bulb temperatures from ICOADS which are not affected by whole number reporting to the same

210 extent. This could be part of the reason why NOCSv2.0 has a lower estimate of specific humidity anomalies

211 over the 1973-1981 period than HadCRUH or Dai, which use the dew point temperatures (Fig. 1).

212

213 Rounding of temperature alone should not affect the mean dew point temperature, specific humidity or vapour

214 pressure. However, as with the solar bias issue, it is sensitive to at what point the reported dew point

215 temperature was derived from the measured wet bulb temperature or relative humidity. Most likely, this would

216 be done prior to any rounding or truncating for reporting but during later conversion of various sources into

217 digital archives, or corrections, the dew point temperature may have been reconstructed

218 (https://icoads.noaa.gov/e-doc/other/dupelim_1980). The effect of rounding on a monthly mean gridbox average

219 should be small as these errors are random and should reduce with averaging. However, there is a risk of

220 removing very high humidity observations when a rounded dew point temperature then exceeds a non-rounded

221 air temperature. Such values are removed by our supersaturation check (Sect. 3.2). We do not feel able to

222 correct for this issue but instead include an uncertainty estimate for it. Overly frequent whole numbers are

223 identified both during quality control track analysis and deck analysis. This will be discussed in more detail in

224 Sect. 3.4. Clearly, there are various issues that can arise linked to the precision of measured and reported data in

225 addition to conversion between unit (e.g., Fahrenheit, Celsius and Kelvin, Fig. S1) and variable.

226

227 **2.5 Measurement errors**

228

229 All observations are subject to some level of measurement error and, outside of precision laboratory

230 experiments, the errors can be significant. The BIPM Guide to the Expression of Uncertainty in Measurement

231 (BIPM, 2008) describes two categories of measurement uncertainty evaluation. A Type A evaluation estimates

232 the uncertainty from repeated observations. A Type B evaluation of the uncertainty is based on prior knowledge

233 of the instrument and observing conditions. Within this study we use a Type B evaluation, adjusting for

234 systematic errors and inhomogeneities due to inadequate ventilation and changing observing heights (screen and

235 height adjustments) and estimate the residual uncertainty For the random components, we make the conservative

236 assumption that all measurements were taken using a psychrometer (wet bulb and dry bulb thermometers),

237 which allows us to follow the HadISDH.land methodology of Willett et al. (2013, 2014) as described in Sect.

238 3.4.. An assessment of the frequency of hygrometer types (TOH) within our selected ICOADS3.0.0/3.0.1 data

239   shows this to be a fair assumption as the vast majority of ships (where metadata is available: ~30 % increasing

240   to ~70 % 1973 to 1995 then decreasing to 60 % by 2014) are listed as being from a psychrometer (Fig. 4).

241   Electric sensors are becoming more common and made up ~30 % of observations by 2014 (the end of the

242   metadata information). There are no instrument type metadata for ocean platforms or moored buoys. As it is

243   likely that most buoy observations are made using RH sensors, we plan to develop an RH sensor specific

244   measurement uncertainty in future versions.

245

246   **2.6 Other sources of error**

247

248   There are other issues specific to humidity measurements that may be further sources of error. Hygrometers that

249   require a wetted wick (i.e., psychrometers), and thus a source of water, are vulnerable to the wick drying out or

250   contamination, especially by salt in the marine environment. The wick drying results in erroneous readings of

251   100 %rh where the wet bulb essentially behaves identically to the dry bulb thermometer. There can also be

252   issues when the air temperature is close to freezing depending on whether the wet bulb has become an ice bulb

253   or not and whether wet bulb or ice bulb calculations are used in any conversions. Humidity observing in low

254   temperature can be generally problematic. For radiosondes, there has previously been a practice of recording a

255   set low value when the humidity observation falls below a certain value (Wade 1994, Elliott et al. 1998). It is

256   debateable how likely such low humidity values are over oceans and this practice has not been documented for

257   ship observations. However, the set value issue is something to look out for. Wet bulb thermometers (and other

258   instruments) can experience some hysteresis at high humidity where it takes some time to return to a lower

259   reading. The wet bulb also requires adequate ventilation which has been discussed above.

260

261   These can be accounted for to a large extent through quality control but some error will inevitably remain. We

262   can increase our confidence in the data by comparison with other available products and general expectation

263   from theory.

264

265   **3 Processing the hourly data into a gridded product**

266

267   ICOADS Release 3.0 (Freeman et al., 2017) forms the base dataset for the HadISDH.marine humidity products.

268   From January 1973 to December 2014 we use ICOADS.3.0.0 from http://rda.ucar.edu/datasets/ds540.0/. These

269   data include a unique identifier (UID) for each observation, a station identifier/ship callsign (ID), metadata on

270   instrument type, exposure and height in many cases. From January 2015 onwards we use ICOADS.3.0.1 from

271   the same source. These data include an ID and UID but no instrument metadata. Each observation is associated

272   with a deck number. These are identifiers for ICOADS national and trans-national sub-sets of data relating to

273   source e.g., deck 926 is the International Maritime Meteorological (IMM) data

274   (https://icoads.noaa.gov/translation.html). We utilise the reported air temperature (T) and reported dew point

275   temperature (T_d) as the source for our humidity products. Sea surface temperature (SST) and wind speed (u) are

276   used for estimating height adjustments.

277

278   We calculate the specific humidity (q), relative humidity (RH), vapour pressure (e), wet bulb temperature (T_w,

279   not the thermodynamic wet bulb but a close approximation to it) and dew point depression (DPD) for each point

280   observation. All humidity variables are derived from reported air and dew point temperature and ERA-Interim

281   climatological (from the nearest 1° by 1° pentad gridbox) surface pressure P_s, using the set of equations from

282   Willett et al., (2014) which can be found in Table S1. This provides consistency with HadISDH.land for later

283   merging. Note that for consistency we also use a fixed psychrometric coefficient that is identical for all

284   observations, minimising the impact of changing instrument types (e.g. whirling sling / ventilated measurement

285   vs screen) on the wet bulb temperature record. This is also consistent with what is done for HadISDH.land.

286

287   Additionally, we use ERA-Interim (Dee et al., 2011) reanalysis data to provide initial marine climatologies and

288   climatological standard deviations for all variables to complete a first iteration climatological outlier test. We

289   extract 1° by 1° gridded 6 hourly 2 m air and dew point temperature) and surface pressure to create 6 hourly

290   humidity variables and then pentad (5 day mean) climatologies and standard deviations over the 1981-2010

291   period. Note that several iterations are passed before finalising the product. Only the initial iteration uses ERA-

292   Interim climatologies, later iterations use climatologies built from the previous iteration's quality-controlled

293   observations (Sects. 3.2, 3.5, 4.1).

294

295   **3.1 Data selection**

296

297   We screen all ICOADS data to sub-select only those observations passing the following criteria:

298   -   there must be a non-missing T and T_d value;

299   -   the platform type (PT) must be in one of the following categories: a ship (a US Navy or unknown

300   vessel, a merchant ship or foreign military ship, an ocean station vessel off station /at an unknown

301   location, an ocean station vessel on station, a lightship, an unspecified ship - PT = 0, 1, 2, 3, 4, 5);

302   or a stationary buoy (moored or ice buoy - PT = 6, 8);

303   -   the observation must have a climatology and standard deviation available for its closest 1° by 1°

304   pentad;

305   -   the observation must pass the gross error checks: calculated RH must be between 0 and 150 %rh

306   (supersaturated values are flagged during quality control); both T and T_d must be between -80 and

307   65 °C; and calculated q must be greater than 0.0 g kg^{-1}.

308

309   Other marine products (e.g., NOCSv2.0; Berry and Kent, 2011) solely use ship observations due to the lack of

310   buoy metadata available. In contrast, we include moored buoys for this version and to produce climatologies

311   because spatial coverage is of high importance. However, we provide ship-only and combined products. This

312   will be reassessed for future versions. Figure 5a shows the number of observations included in the initial

313   selection per year, broken down by platform type. The breakdown for day and night time observations

314   individually is near identical (not shown). Ship (PT = 5) observations make up almost the entire dataset until the

315   1990s. After this the number of moored buoys grows significantly to make up around ~50 % of observations

316   from 2000 onwards. The ship-only product (removal of moored buoys) significantly reduces the number of

317   observations in the recent period but gives a more consistent number of observations throughout the record. Our

318   use of climate anomalies should mitigate biasing due to uneven sampling to some extent. Note that the number

319   of gridboxes containing data may be a more relevant measure and that the vast increase in the number of buoys

320   has not actually resulted in the same level of increase in spatial coverage in terms of gridboxes (compare 2018

321   annual average maps for ship-only and combined HadISDH.marine in Fig. S2).

322

323   **3.2 Quality control processing**

324

325   We have not used any of the pre-set flags from ICOADS processing to ensure methodological independence of

326   HadISDH and a process that allows for exploration and analysis of different methodological choices. The

327   quality control processing employed here largely follows the methodology for HadSST4 (Kennedy et al., 2019)

328   with some changes to the climatology check and buddy check thresholds to increase regional sensitivity and

329   additional humidity specific checks. A flag for whole number prevalence has also been added but this is used for

330   uncertainty estimation and not to fail an observation. All observations have their nearest 1° by 1° pentad mean

331   climatology (source depends on iteration – Sect. 3.5) subtracted to create a climate anomaly.

332

333   Each observation is passed through a suite of basic quality control tests (base qc) which consist of:

334   -   position check (failures removed): latitudes must be between -90° and 90° and longitudes must be

335   between -180° and 360° (later converted to -180° to 180°);

336   -   date check (failures removed): hour, day, month and year must be valid quantities;

337   -   blacklist check (failures removed): any observation from Deck 732 from a specified year and

338   region is blacklisted (Rayner et al., 2006, Kennedy et al, 2011a, Table S2);

339   -   day check (daytimes are flagged): any value likely to be affected by the solar heating of a ship

340   where the sun was above the horizon an hour before the observation (based on the month, day,

341   hour, latitude and longitude; Kent et al. (2013)) is flagged;

342   -   climatology check (failures flagged): T and T_d must be within a specified threshold of the nearest

343   1° by 1° pentad climatology;

344   -   supersaturation check (failures flagged – T_d only): T_d must not be greater than T.

345

346   The climatology check differs from the static HadSST3 threshold of climatology +/- 10° C. We have allowed for

347   a variable threshold depending on the nearest 1° by 1° pentad climatology standard deviation σ. This is set at 5.5

348   σ. It accounts for the lower variability in the tropics and greater variability in the mid-latitudes. We have set

349   minimum and maximum σ values of 1° C and 4° C respectively resulting in a minimum range of +/- 5.5° C and

350   a maximum range of +/- 22° C. Several thresholds were tested with the selected threshold balancing avoiding

351   acute cut-offs in the data distribution while still removing obviously bad data (Figs. S3 to S6). Given that

352   outliers are assessed by comparing a point observation with a 1° by 1° pentad mean the thresholds have to be

353   relatively large.

354

355   Five additional checks are then applied at the ship track level where possible:

356   -   track check (failures flagged): the distance and direction travelled by the ship must be plausible

357   and consistent;

358   -   repeated value check (failures flagged): a T or T_d value must not appear in more than 70 % of a

359   ship track where there are at least 20 observations;

360   -   repeated saturation check (failures flagged – T_d only): saturation ($T_d = T$) must not persist for more

361   than 48 hours where there are at least 4 observations;

362   -   buddy check (failures flagged – 3rd iteration only): T and T_d must be within a specified threshold of

363   the average of its nearest neighbours in space and time;

364   -   whole number flag (whole numbers flagged): a T or T_d value must not appear as a whole number in

365   more than 50 % of a ship track where there are at least 20 observations.

366

367   The buddy check compares each observation's climate anomaly with the average of the climate anomalies of its

368   nearest neighbours in space and time, expanding the search area in space and time as necessary until at least one

369   neighbour observation is found. The permitted difference is set by the climatological standard deviation of the

370   candidate 1° by 1° pentad gridbox multiplied by an amount dependent on the number of neighbours present.

371   There are five levels of searches:

372   1.   ±1° latitude and longitude and ± 2 pentads: the climatological standard deviation is multiplied by

373   5.5, 5.0, 4.5 and 4.0 for 1-5, 6-15, 16-100 and >100 neighbouring observations respectively;

374   2.   ±2° latitude and longitude and ± 2 pentads: the climatological standard deviation is multiplied by

375   5.5 for >1 neighbouring observation;

376   3.   ±1° latitude and longitude and ± 4 pentads: the climatological standard deviation is multiplied by

377   5.5, 5.0, 4.5 and 4.0 for 1-5, 6-15, 16-100 and >100 neighbouring observations respectively;

378   4.   ±2° latitude and longitude and ± 4 pentads: the climatological standard deviation is multiplied by

379   5.5 for >1 neighbouring observation;

380   5.   no neighbour ±2° latitude and longitude and ± 4 pentads: the threshold is set at 500.

381   The thresholds used for the buddy check are wider than those previously used in HadSST3. This is to account

382   for the greater variability of air and dew point temperature, and sparser observation coverage. It is only applied

383   in the 3rd iteration of the quality control (Sect. 3.5).

384

385   Figure 5 shows the final number of observations passing through initial selection and then 3rd iteration quality

386   control by platform (PT) type. The quality control does not significantly affect one platform over another. The

387   performance of these tests is demonstrated for 4 example months in Figs. S3 to S6. These reveal a slight positive

388   bias in the removed air temperature observations and negative bias in removed dew point temperature.

389   Removals in terms of relative humidity and specific humidity similarly tend to have a negative bias. It is clear

390   that the majority of grossly erroneous observations are removed. The change in climatology between iterations

391   of the quality control process (Sect. 3.5) also makes a difference to removals. This is both because the

392   observation driven climatologies do not provide complete spatial coverage and because the ERA-Interim

393   climatologies are cooler and drier than the observations (Sect. 4.1). Removals are dense in the Northern

394   Hemisphere and especially sparse around the tropics. The addition of the buddy check in the 3rd iteration

395   considerably increases the removal rate, noticeably over the Southern Hemisphere and Tropics.

396

397   The quality-control flagging rate for the 3rd iteration reduces over time from ~25 % to ~18 %, as shown in Fig.

398   S7 This is driven by the buddy check and repeated saturation check. Proportionally more observations are

399   flagged during the daytime than night time but the interannual behaviour is very similar. The daytime increase is

400   driven by the larger number of air temperature buddy and climatology check failures. This could be due to the

401   issue of solar heating of the ship structure during the daytime. The main source of test fails by a large margin is

402   the buddy check, followed by the climatology check and repeated saturation check. There doesn't appear to be a

403   strong difference in the distribution of removals from each test between the 1973-1981 and 1982-1990 periods

404   that might explain the pre-1982 moist bias (Fig. S8, Sect. 4.2).

405

406   The whole number flags show very different behaviour to the other checks and to each other over time in Fig.

407   S7. These depend on the ability to assign each observation to a track/voyage and the frequency of whole number

408   observations on that voyage, hence, these flags are not a true reflection of the whole number frequency.

409   Compared to the actual proportion of whole numbers shown in Fig. S1, these tend to exaggerate the annual

410   patterns but the shape is broadly similar. This method of identifying problematic whole numbers appears to

411   under-sample the true distribution, especially for air temperature pre-1982. An additional deck-based check is

412   applied later for estimating uncertainty from whole numbers (Sect. 3.4).

413

414   Note that the NOCSv2.0 dataset, with which we compare our specific humidity data, includes an outlier check

415   that removes data greater than 4.5 standard deviations from the climatological mean. This test has already been

416   applied within the ICOADS format and so the NOCSv2.0 excludes any data with ICOADS trimming flags set

417   (Wolter 1997). They also apply a track check based on Kent and Challenor (2006).

418

3.3 Bias adjustments and associated uncertainties

419

420

421 Given the issues raised in Sect. 2, it is desirable to attempt to adjust the observations to improve the spatial and

422 temporal homogeneity and accuracy of the data. As discussed in Sect. 2.1, we have not attempted to adjust for

423 solar biases in this first version product. We have made adjustments for instrument and height biases and

424 estimated uncertainties in these adjustments.

425

426 **3.3.1 Application of adjustments for biases from un-aspirated instruments**

427

428 We have shown that the majority of humidity observations have been made with a psychrometer (Fig. 4) and

429 that 30-70 % of instruments with metadata available have been housed within a non-aspirated screen (Fig. 2).

430 Berry and Kent (2011) found that applying a 3.4 % reduction to specific humidity observations from non-

431 aspirated screens was a reasonable adjustment to remove the bias relative to aspirated/well ventilated

432 observations (e.g., slings, whirled hygrometers or artificially aspirated instruments). Some uncertainty remains

433 after adjustment which they estimated to be ~0.2 g kg^{-1}. We have used the hygrometer exposure metadata

434 (EOH) or the thermometer exposure (EOT) if EOH does not exist. We assume good ventilation for any

435 instruments that are aspirated (A), from a sling (SL) or ship's sling (SG) or from a whirling instrument (W). We

436 assume poorer ventilation for instruments that are from a screen (S), ship's screen (SN) or are unscreened (US)

437 and apply a bias adjustment. The reported exposure type of Ventilated Screens (VS) does not appear to mean

438 that the screen is artificially ventilated and so bias adjustments are also applied to these. We do not apply

439 adjustments to buoys and other non-ship data based on the assumption that these generally measure relative

440 humidity directly. For any ship observations with no exposure information we apply 55 % of the 3.4 %

441 adjustment based on the mean percentage of observations with EOH metadata that require an adjustment over

442 the 1973-2014 (metadata) period. This partial adjustment factor follows the method of Berry and Kent (2011)

443 and Josey et al. (1999) but differs in quantity. They assessed this over a shorter time period and found then that

444 ~30 % of observations were from poorly ventilated instruments.

445

446 To estimate the uncertainty in the non-aspirated instrument adjustment applied U_i, we use the Berry and Kent

447 (2011) and Josey et al. (1999) uncertainty estimate of 0.2 g kg^{-1} and apply this in all cases where an adjustment

448   or partial adjustment has been applied. This is treated as a standard uncertainty (1 σ). In the case of partial

449   adjustments for the ship observations with no metadata there is large uncertainty in both the adjustment and

450   adjusted value. To account for this we use the amount of what would have been a full 3.4 % adjustment in

451   addition to the 0.2 g kg^{-1} as the 1 σ uncertainty.

452

453   To carry these adjustments to all other humidity variables we start with q and then propagate the adjustment and

454   uncertainty amounts using the equations in Table S1. Using the original T (which does not need to be adjusted

455   for poor ventilation) and ERA-Interim climatological surface pressure, e can be calculated from q. T_d and RH

456   can be calculated from e and T. From these, the T_w and DPD can be calculated. The uncertainty is obtained by

457   subtracting the new adjusted quantities from an adjusted quantity plus uncertainty, beginning again from the

458   adjusted q plus the 0.2 g kg^{-1} uncertainty and full adjustment magnitude in the case of ships without metadata.

459

460   **3.3.2 Application of adjustments for biases from ship heights**

461

462   After bias adjustment for poor ventilation, all variables are adjusted to approximately 10 m elevation. This

463   serves to account for the inhomogeneity from the systematic increase in ship height over time and for spatial

464   inhomogeneity between observations made at different heights. In the absence of height adjustments, increasing

465   ship heights likely lead to a small decrease in air temperature and specific humidity over time (Berry and Kent,

466   2011) because these quantities generally decrease with height. As Fig. 3 shows, the standard deviations in ships'

467   instrument heights exceed 5 m in most cases. Also, we have included buoys in the processing so far and these

468   can be very low (~4 m, e.g. Gilhousen, 1987) relative to ship observing heights.

469

470   The height of the hygrometer (HOH) must be estimated (HOHest) as no metadata is available. In the case of

471   psychrometers, which are the most common instruments listed in the ship metadata, the wet and dry bulb

472   thermometers are co-located. Figure 3 shows that the visual observation height (HOP) is the most commonly

473   available information, followed by the barometer height (HOB) and then thermometer height (HOT). It also

474   shows the mean and standard deviation of all observing heights including the anemometer (HOA). Hence,

475   HOHest is obtained using the following methods in preference order:

476

477   1.   HOP present and >2 m: HOHest μ = HOP, σ = 1 m

478   2.   HOB present and >2 m: HOHest μ = HOB, σ = 1 m

479   3.   HOT present and >2 m: HOHest μ = HOT, σ = 1 m

480   4.   HOA present and >12 m: HOHest μ = HOA – 10, σ = 9 m

481   5.   No height metadata: HOHest μ = f(linear trend in mean HOP/HOB/HOT height, date), σ = f(linear

482   trend in standard deviation HOP/HOB/HOT height, date)

483

484   The μ and σ of the combined HOP, HOB and HOT increases from 16 m and 4.6 m respectively in January 1973

485   to 23 m and 11 m respectively in December 2014. Kent et al. (2007) and Berry and Kent (2011) used 16 m to 24

486   m between 1971 and 2007 so our estimate is very similar. The anemometer height is also required for the

487   adjustments. We either use the provided HOA as long as it is greater than 2 m or set it to 10m above the

488   HOHest. All buoys are assumed to be observing at 4 m, with anemometers at 5 m

489   (http://www.ndbc.noaa.gov/bht.shtml).

490

491   Once HOHest has been obtained for each observation, the air temperature and specific humidity are adjusted to

492   10 m using bulk flux formulae. The methodology, assumptions and parameterisations largely follow that of

493   Berry and Kent (2011), Berry (2009), Smith (1980, 1988) and Stull (1988). Essentially, the quantity of interest x

494   can be adjusted to a reference height of 10 m as follows:

495

496         $$x_{10} = x - \frac{x_*}{\kappa}\left(\ln\left(\frac{z_x}{10}\right) - \psi_x + \psi_{x10}\right)$$ (1)

497

498   where x_* is the scaling parameter specific to that variable (e.g., friction velocity in the case of u, characteristic

499   temperature or specific humidity in the case of T or q respectively), κ is the von Karman constant (0.41 used

500   here), z_x is the observation height of the variable of interest, ψ_x is the stability correction for the variable of

501   interest and is a function of $f(z_x/L)$, ψ_{x10} is the stability correction for the variable of interest at a reference height

502   of 10m and is a function of $f(10/L)$ and L is the Monin-Obukov Length.

503

504   An iterative approach (as done for Berry and Kent 2011) is required to resolve Eq. (1) because we only have

505   basic meteorological variables available at a single height for each observation. We start from T, q, u, sea

506   surface temperature (SST), the co-located 1° by 1° gridbox pentad climatological surface pressure from ERA-

507   Interim (climP), HOHest which becomes both z_q and z_t and our estimated anemometer height which becomes z_u.

508   For some observations the SST or u is missing. If SST is missing it is given the same value as T so in effect, no

509   adjustment to T is applied. Either way, the SST is set to a minimum of -2° C and a maximum of 40° C. If u is <

510   0.5 m s^{-1} it is given a light wind speed of 0.5 m s^{-1}. If u is missing or >100 m s^{-1} it is assumed to be erroneous but

511   given a moderate wind speed of 6 m s^{-1}. We also approximate surface values T_0, q_0 and u_0 where T_0 = SST, q_0 =

512   q_{sat}(SST)x0.98 and $u_0 = 0$. Clearly, with so many necessary approximations there are many different plausible

513   methodological choices, hence the need for multiple independent analyses that explore these different choices in

514   order to quantify the structural uncertainty.

515

516   We begin the iteration by assuming a value for L depending on assumed stability:

517   -   if (SST - T) > 0.2: L = -50, unstable conditions are assumed;

518   -   if (SST - T) < -0.2: L = 50, stable conditions assumed;

519   -   if (SST = T) +/- 0.2: L = 5000, neutral conditions assumed where L tends to ∞.

520   We also start with an assumption that the 10 m wind speed in neutral conditions $u_{10n} = u$. The iteration is

521   continued until L converges to within 0.1, which it generally does. If after 100 iterations there is no convergence

522   we either apply no adjustment or if absolute L is large (> 500) we assume neutral conditions and take L (and all

523   other parameters) as they are. In cases where u_* is very large (it should be < 0.5 [Stull, 1988]) we also apply no

524   adjustment. The iteration involves 21 steps as described in the Supplementary Material.

525

526   For most observations we arrive at a plausible L, friction velocity u_*, ψ_x and ψ_{x10}. We then calculate the scaling

527   parameters T_* and q_*:

528

529        $$T_* = \kappa \left(\ln \left(\frac{z_t}{z_{t0}} \right) - \psi_t \right)^{-1} (T - T_0) \qquad \text{(2a)}$$

530        $$q_* = \kappa \left(\ln \left(\frac{z_q}{z_{q0}} \right) - \psi_q \right)^{-1} (q - q_0) \qquad \text{(2b)}$$

531

532   where the neutral stability heat transfer coefficient $z_{t0} = 0.001$ and the neutral stability moisture transfer

533   coefficient $z_{q0} = 0.0012$ (Smith 1988). The adjusted values for T_{10} and q_{10} can then be calculated from Eq. (1).

534   From these we recalculate the other humidity variables using the equations in Table S1.

535

536   There is uncertainty in the obtained HOHest. Given that this is a best estimate we assume that the uncertainty in

537   the height is normally distributed and use the standard deviation in the height estimate used to calculate an

538   uncertainty range in the height adjusted value x of xH_{min} to xH_{max}. Following the BIPM Guide to the Expression

539   of Uncertainty in Measurement (BIPM, 2008), the standard uncertainty (1 σ) for the height adjusted value (U_h)

540   is then given by:

541

542   $$U_h = \frac{\left(\frac{xH_{max}-xH_{min}}{2}\right)}{\sqrt{9}}$$ (3)

543

544   The range xH_{min} to xH_{max} depends on the source of HOHest and associated σ, as listed above. There are several

545   scenarios where estimating the uncertainty in this way is not possible or calculation of an adjustment is not

546   possible. Also, U_h for buoys is highly uncertain given the lack of height information available. These alternative

547   scenarios are documented in Table 1.

548

549   **3.4 Estimating residual uncertainty at the observation level**

550

551   Three other sources of uncertainty affect the marine humidity data at the observation level. These are

552   measurement uncertainty U_m, climatology uncertainty U_c and whole number uncertainty U_w. These are all

553   assessed as 1 σ standard uncertainties.

554

555   We have estimated U_m for each observation following the method used for HadISDH.land (Willett et al., 2013,

556   2014). This assumes that humidity was measured using a pyschrometer which is a reasonable assumption for the

557   marine ship data (Fig. 4). The HadISDH.land measurement uncertainty is based on an estimated standard (1 σ)

558   uncertainty in the wet bulb and dry bulb instruments of 0.15° C and 0.2° C respectively. As shown in Table 3,

559   the equivalent uncertainty for the other variables depends on the temperature. The uncertainty is applied as a

560   standard uncertainty in RH depending on which bin the air temperature falls in. This is then propagated through

561   the other variables starting with vapour pressure, using the equations in Table S1.

562

563   Whole numbers of air and/or dew point temperature that have either been flagged as such during quality control

564   (Sect. 3.2), or that belong to a source deck/year where whole numbers make up more than two times the

565   frequency of other decimal places (Table S4), are given and uncertainty U_w. These decks and years where whole

566   numbers are very common differ for air and/or dew point temperature. Clearly with so many decks affected, the

567   removal of entire decks to remove any whole number biasing could easily reduce sampling to critically low

568   levels. We cannot distinguish between observations that have been rounded versus those that have been

569   truncated so we assume that all offending whole numbers have been rounded. This means that the value could

570   be anywhere between $\pm\, 0.5°$ C, with a uniform distribution. Hence, where only air or dew point temperature is

571   an offending whole number the standard 1 σ uncertainty expressed in air or dew point temperature (° C) is:

572

573         $$U_w = \frac{0.5}{\sqrt{3}}$$ (4)

574

575   Where both air and dew point temperature are offending whole numbers the standard 1 σ uncertainty expressed

576   in air or dew point temperature (° C) for dew point depression, relative humidity and wet bulb temperature is:

577

578         $$U_w = \frac{1}{\sqrt{3}}$$ (5)

579

580   There is uncertainty U_c in the climatological values used to calculate climate anomalies because of missing data

581   over time, uneven and sparse sampling in space and also the inevitable mismatch between a point observation

582   and a 1° by 1° gridded pentad climatology. This uncertainty reduces with the number of observations

583   contributing to the climatology N_{obs} and with the variability of the region σ_{clim}. The climatologies used to create

584   the anomalies have undergone spatial and temporal interpolation to move from 5° by 5° gridded monthly

585   climatologies and climatological standard deviations σ_{clim} to maximise coverage and so it is not straightforward

586   to assess the number of observations contributing to each 1° by 1° gridded pentad climatology and the true σ_{clim}

587   is likely greater. Therefore, we assume a worst case scenario of $N_{obs} = 10$. Hence, for a standard 1σ uncertainty:

588

589         $$U_c = \frac{\sigma_{clim}}{\sqrt{N_{obs}}}$$ (6)

590

591   **3.5 Gridding of actual and anomaly values and uncertainty**

592

593   To create a quasi-global monitoring product the raw observations need to be gridded. The spatial density is too

594   low for high resolution grids and the intended purpose is for this marine product to be blended with the

595   HadISDH.land humidity product which is on a 5° by 5° grid at monthly resolution. Hence, the raw hourly

596   observations must be averaged to monthly mean gridded values.

597

598   The sparsity of the data means that there is a risk of bias due to poor sampling. A 5° by 5° gridbox covers an

599   area greater than 500 km^2 by 500 km^2 which, despite the large correlation decay distances of both temperature

600   and humidity, can include considerable variability. Furthermore, a monthly mean can be made up of a strong

601   diurnal cycle and considerable synoptic variability. This is minimised by the use of climate anomalies but

602   regardless, care should be taken to ensure sufficient sampling density while maximising coverage where

603   possible.

604

605   Several data-density criteria were trialled to balance spatial coverage and poor representativeness (high

606   variance) of the gridbox averages. Climate anomalies are created at the raw observation level by subtracting the

607   nearest 1° by 1° pentad climatology (1981-2010) and so we can grid both the actual values and the anomalies.

608   Gridding of the anomalies is safer than gridding actual values in terms of biasing through poor sampling density

609   because the correlation length scales of anomalies are higher than for actual temperatures. Initially, ERA-

610   Interim is used to provide a climatology. This then requires an iterative approach to produce an initial

611   observation-based climatology and improve the climatology through quality control. To reduce biasing further

612   we grid the data in six stages to create an average at each stage. The entire process including quality control,

613   bias adjustment, gridding and three iterations, is shown diagrammatically in Fig. 6 and each gridding stage

614   described below.

615

616   1. Create 1° by 1° 3-hourly gridded means from the hourly observations of actuals and anomalies;

617   there must be at least one observation.

618   2. Create separate 1° by 1° daytime and nighttime gridded means from the 1° by 1° 3-hourly gridded

619   mean actuals and anomalies; there must be at least one 1° by 1° 3-hourly grid.

620   3. Create 5° by 5° monthly daytime and nighttime gridded means from the 1° by 1° daytime and

621   nighttime gridded mean actuals and anomalies; there must be at least 0.3*days in the month of 1°

622   by 1° daily grids.

623  4. Create combined 5° by 5° monthly gridded means from the 5° by 5° monthly daytime and

624   nighttime gridded mean actuals and anomalies; there must be at least 1 5° by 5° monthly daytime

625   or nighttime gridded mean.

626  5. Create 1981-2010 5° by 5° monthly mean climatologies and standard deviations from the 5° by 5°

627   monthly gridded means of actuals and anomalies; there must be at least 10 5° by 5° monthly

628   gridded means.

629  6. Renormalise the gridded anomalies by subtracting the monthly anomaly 1981-2010 climatology to

630   remove biases from use of the previous iteration climatology (Sect. 4.1).

631

632 At each iteration the gridded observation based climatologies are infilled linearly over small gaps in space and

633 time and then interpolated down to 1° by 1° pentad resolution. The observations are too sparse to create such

634 high-resolution grids directly.

635

636 The observation uncertainties also need to be gridded and the total observation uncertainty U_o calculated. Ships

637 move around, and so their uncertainties also track around the globe. This means that the uncertainty in any one

638 point / gridbox bears some relationship to nearby points / gridboxes over time and space and cannot be treated

639 independently. Correlation needs to be accounted for both in gridding and subsequently creating regional

640 averages from gridboxes to avoid underestimation. The five sources of observation uncertainty are summarised

641 in Table 1. The non-aspirated instrument adjustment uncertainty U_i, height adjustment uncertainty U_h and

642 climatology uncertainty U_c persist over time and space as ships move around. These are accordingly treated as

643 correlating completely within one gridbox month. The measurement uncertainty U_m, and whole number

644 uncertainty U_w are likely to differ observation to observation and so treated has having no correlation within one

645 gridbox month. Hence, observation uncertainty sources are first gridded individually, following the first four

646 steps outlined above and taking into account correlation where necessary. The gridded uncertainty sources are

647 then combined to give a total observation uncertainty for each gridbox. For those that do not correlate (U_m and

648 U_w) the gridbox mean uncertainties U_{gb} are combined over N points in time and space as follows:

649

650   $$U_{gb} = \frac{\sqrt{a^2 + b^2 \ldots + n^2}}{N} \qquad (7)$$

651

652   For those sources that do correlate (U_c, U_i and U_h), assuming $r = 1$, the gridbox mean uncertainties U_{gb} are

653   combined over N points in time and space as follows:

654

655        $$U_{gb} = \frac{a+b\ldots+n}{N} \qquad\qquad (8)$$

656

657   To create the total observational uncertainty the gridbox quantities of the five uncertainty sources can then be

658   combined in quadrature:

659

660        $$U_o = \sqrt{U_c{}^2 + U_m{}^2 + U_w{}^2 + U_h{}^2 + U_i{}^2} \qquad\qquad (9)$$

661

662   Given the general sparsity of observations across each gridbox month and the uneven distribution of

663   observations across each gridbox and over time there is also a gridbox sampling uncertainty component, U_s.

664   This is estimated directly at the 5° by 5° monthly gridbox level and follows the methodology applied for

665   HadISDH.land (Willett et al., 2013, 2014), denoted SE^2, which is based on station-based observations from

666   Jones et al (1997):

667

668        $$U_s = \frac{\left(\bar{s}_i^2 \, \bar{r}(1-\bar{r})\right)}{(1+(N_s-1)\bar{r})} \qquad\qquad (10)$$

669   .

670   where $\bar{s}_i^2$ is the mean variance of individual stations within gridbox, $\bar{r}$ is the mean inter-site correlation and N_s is

671   the number of stations contributing to the gridbox mean in each month. The mean variance of individual stations

672   within the gridbox is estimated as:

673

674        $$\bar{s}_i^2 = \frac{(\hat{S}^2 N_{SC})}{(1+(N_{SC}-1)\bar{r})} \qquad\qquad (11)$$

675

676   where $\hat{S}^2$ is the variance of the gridbox monthly anomalies over the 1982-2010 climatology period and N_{SC} is

677   the mean number of stations contributing to the gridbox over the climatology period. The mean inter-site

678   correlation is estimated by:

679

680        $$\bar{r} = \frac{x_0}{X}\left(1 - exp\left(-\frac{x_0}{X}\right)\right) \qquad\qquad (12)$$

681

682   where X is the diagonal distance across the gridbox and x_0 is the correlation decay length between gridbox

683   means. We calculate x_0 as the distance (gridbox midpoint to midpoint) at which correlation reduces to 1/e. To

684   account for the fact that marine observations generally move around at each time point we use the concept of

685   pseudo-stations to modify this methodology. For any one day there could be 25 1° by 1° gridboxes and so we

686   assume that the maximum number of pseudo-stations per gridbox is 25 which is broadly consistent with the

687   number of stations per gridbox in HadISDH.land. Over a month then, there could be a maximum of 775 1° by 1°

688   daily gridboxes contributing to each 5° by 5° monthly gridbox. Given ubiquitous missing data and sparse

689   sampling the maximum in practice is closer to 600. Using these values we then scale the actual number of 1° by

690   1° daily gridboxes contributing to each 5° by 5° monthly gridbox to provide a pseudo-station number between 1

691   and 25 for each month (N_s) and then the average over the climatology period (N_{SC}).

692

693   The gridbox U_o and U_s uncertainties are then combined in quadrature, assuming no correlation between the two

694   sources. This gives the full gridbox uncertainty U_f. Calculation of regional average uncertainty and spatial

695   coverage uncertainty is covered in Sect. 4.

696

697   **4 Analysis and validity of the gridded product**

698

699   The final gridded marine humidity monitoring product presented as HadISDH.marine.1.0.0.2018f is the result of

700   the 3rd iteration quality-control and bias-adjustment of ship-only observations average into 5° by 5° gridded

701   monthly means (Fig. 6). There are four reasons for only using the ship observations. Firstly, the increase in

702   spatial coverage in the combined ship and buoy product is actually fairly small (Fig. S2) and only during the

703   latter part of the record. Secondly, a dataset intended for detecting long-term changes in climate should have

704   reasonably consistent input data and coverage over time. Thirdly, we believe that the buoy data are less reliable

705   given their proximity to the sea surface and exposure to sea spray contamination in addition to the lower

706   maintenance frequency compared to ship data. Fourthly, there are no metadata available for buoy observations

707   which makes it difficult to apply necessary bias adjustments or estimate uncertainties. Actual monthly means,

708   anomalies from the 1981-2010 climatology (not standardised by division with the standard deviation), the

709   climatological means and standard deviation of the climatologies, uncertainty components and number of

710   observations for both products are all made available as netCDF from www.metoffice.gov.uk/hadobs/hadisdh/.

711

4.1 Comparison of climatologies between HadISDH.marine and ERA-Interim

713

714   At the end of each iteration (Fig. 6), observation-based climatology fields are created at both the monthly 5° by

715   5° grid and, by interpolation, pentad 1° by 1° grid (Sect. 3.5). These are then used to quality control and create

716   anomaly values for the next iteration. Hence, the 2nd iteration quality-controlled data are used to build the final

717   3rd iteration and therefore, there should be no lasting effect from having used the ERA-Interim fields initially.

718   The quality-controlled, buddy-checked and bias-adjusted 3rd iteration is used to create the final climatology

719   provided to users.

720

721   Specific humidity, relative humidity and air temperature difference maps of the 2nd iteration minus ERA-Interim

722   pentad 1° by 1° grid climatologies and climatological standard deviations are shown in Figs. S9 to S14 for a

723   selection of pentads. Note that ERA-Interim fields are for 2 m above the ocean surface whereas the raw

724   observations range between approximately 10 m to 30 m above the surface. In normal conditions we may

725   therefore expect ERA-Interim to provide climatologies that are warmer and moister than the observations.

726   However, overall, ERA-Interim appears drier (both in absolute and relative terms) and cooler than the

727   observation based climatologies. For humidity this is consistent with the results of Kent et al. (2014). For the

728   majority of gridboxes these differences are within ± 2 g kg^{-1}, %rh and °C. However, differences are especially

729   strong around coastlines with magnitudes exceeding ± 10 g kg^{-1}, %rh and °C. This is to be expected given that

730   ERA-Interim coastal gridboxes will include effects from land, especially at the relatively coarse 1° by 1° grid

731   resolution. For relative humidity there are more regions where ERA-Interim is more saturated and there is more

732   seasonality in the differences. Relative humidity is less stable spatially and on synoptic time scales and also

733   more susceptible to biases and errors than specific humidity and air temperature, largely because it is affected by

734   errors in both air temperature and dew point temperature. For temperature, the coastal difference can be positive

735   or negative depending on the season.

736

737   The climatological standard deviations are generally lower in the 2nd iteration observations compared to ERA-

738   Interim. Differences are generally between ± 2 g kg^{-1}, %rh and °C but for relative humidity there are expansive

739   regions in the extratropics to mid-latitudes, especially in the Northern Hemisphere where climatological

740   standard deviations are up to 5 %rh lower in the observations. The generally lower variability in the

741   observation-based climatology is to be expected given the interpolation from monthly mean resolution and

742   interpolation over neighbouring gridboxes where data coverage is limited. However, much of the tropics,

743   particularly in the Southern Hemisphere tends to show more variability in the observations. Similarly, many of

744   the peripheral gridboxes (those at the edge of the spatial coverage and therefore more likely to be interpolated

745   from nearby gridboxes rather than based on actual data) show higher variability for specific and relative

746   humidity and lower variability for air temperature. All of these gridboxes are in data sparse regions which likely

747   contributes to the higher variability. Ideally, observation based climatologies would be created directly at the

748   pentad 1° by 1° grid but this severely reduces spatial coverage of the climatology fields and any product based

749   on them. A balance has to be made between coverage and quality.

750

751   Annual mean 5° by 5° climatologies (no interpolation) from the 3rd iteration quality-controlled, bias-adjusted

752   ship-only product are shown in Fig. 7 for specific humidity, relative humidity, air temperature and dew point

753   temperature. These have a minimum data presence threshold of 10 years for each month over the climatology

754   period and at least 9 climatological months present for the annual climatology. Data coverage is virtually non-

755   existent in the Southern Hemisphere below 40° S and Northern Hemisphere coverage diminishes drastically

756   above 60° N. These climatologies are as expected for these variables and compare well in terms of broad spatial

757   patterns with ERA-Interim (not shown). There is good spatial consistency considering that no interpolation has

758   been conducted meaning that any erroneous gridboxes should stand out. We conclude that as a first version

759   product, these climatologies look reasonable.

760

761   **4.2 Analyses of global average for various processing stages and with other products**

762

763   Global average quantities are key measures of climate change and so we focus here on the differences arising

764   from the various processing steps of HadISDH.marine along with the NOCSv2.0 specific humidity and ERA-

765   Interim reanalysis products. Global averages (70° S to 70° N) have been created by weighting each gridbox by

766   the cosine of its latitude at gridbox centre. All timeseries shown are the renormalised anomalies with a zero-

767   mean over the 1981-2010 period. Figs. 8 to 11 show timeseries for specific humidity, relative humidity, dew

768   point temperature and air temperature respectively. Decadal linear trends (shown) are computed using the

769   median of pairwise slopes with ranges representing the 90th percentile confidence interval.

770

771   For all variables, there are only small differences in the global average timeseries between the various

772   processing steps – from the raw data (noQC) to the 3rd iteration quality-controlled (NBC) and then the bias-

773   adjusted data (BClocal). They are smallest for air temperature and largest for relative humidity. Both the

774   interannual variability and long-term linear trends are very similar, and the trends in the global average are

775   positive over the 1973-2018 period for specific humidity, dew point temperature and air temperature, and

776   negative for relative humidity. We consider these trends to be significant because the 90th percentile confidence

777   intervals around the trend are not large enough to bring the direction of the trends into question. The linear

778   trends for the final HadISDH.marine.1.0.0.2018f version are 0.07 ± 0.01 g kg^{-1} decade^{-1}, -0.09 ± 0.02 %rh

779   decade^{-1}, $0.08 \pm 0.01°$ C decade^{-1} and $0.11 \pm 0.01°$ C decade^{-1} for specific humidity, relative humidity, dew

780   point temperature and air temperature respectively. Hence, we conclude that HadISDH.marine shows

781   moistening and warming since the 1970s globally in actual terms but that the air above the oceans has become

782   less saturated and drier in relative terms. This differs from theoretical and model-based expectations of a small

783   positive or no change in relative humidity over ocean (Byrne and O'Gorman, 2013, 2018).

784

785   Since there are considerable known issues affecting the marine humidity data, and because there are large

786   outliers (Figs. S3 to S6), the effect of quality (noQC compared to NBC), might be expected to be large.

787   Furthermore, approximately 25 %, dropping steadily over time to 18 % of the initial selection of data have been

788   removed by the quality control (Fig. 6), so there is a considerable difference in the amount of data contributing

789   to the quality-controlled version compared to the raw version. Despite all of this, differences are relatively

790   small. Overall, the quality control makes the positive trends smaller (specific humidity, dew point temperature

791   and air temperature) and negative trends larger (relative humidity). The effect of quality control, including

792   buddy checking, is largest in the 1970s to early 1980s, when the largest amount of data was removed by quality

793   control. This is especially noticeable for relative humidity and dew point temperature, and the same period as

794   the previously noted moist relative humidity bias, suggesting that the pre-1982 bias could be an artefact of the

795   quality control. This could be due to erroneous removal of good data but investigation (Figs. S3 to S8) suggests

796   that much of the data removal was appropriate – many very low relative humidity values were removed. It could

797   also be an artefact of the reduced number of observations after quality control, reducing the chance of averaging

798   out random error. Either way, the pre-1982 moist bias is apparent in HadISDH.marine and quality control of the

799   pre-1982 data is an area for more research in future versions.

800

801   The bias adjustment (BClocal, BClocalHGT, BClocalINST) reduces the negative trends in relative humidity

802   both compared to the raw (noQC) and quality-controlled (NBC) data. It increases the positive trends in specific

803   humidity and dew point temperature relative to the quality-controlled data but reduces the trends compared to

804   the raw data. The effect of bias adjustment is negligible for air temperature, which only has adjustment for ship

805   height applied. For the humidity variables the height adjustment has a far larger effect than the non-aspirated

806   instrument adjustment. The non-aspirated instrument adjustment makes the positive trends in specific humidity

807   and dew point temperature slightly smaller and the negative trends in relative humidity slightly larger. The

808   height adjustment has the opposite effect. For relative humidity, the bias adjustments appear to have introduced

809   greater intra-decadal scale variability but retained the interannual patterns, again highlighting the sensitivity of

810   relative humidity compared to the other variables. Given that these biases exist we do have to try and mitigate

811   their impact. However, this is a focus area for investigation and improvements in future versions of

812   HadISDH.marine.

813

814   The timeseries that include data from moored buoys compared to those from ships only ('all' versus 'ship')

815   show smaller positive trends for specific humidity, dew point temperature and air temperature and larger

816   negative trends for relative humidity. Moored buoys begin to play a role from the late 1980s, increasing in

817   number dramatically to make up over 50 % of the observations by 2015. The 'all' timeseries can be seen to

818   diverge slightly from the 'ship' timeseries in the latter part of the record. Therefore, it is more consistent to

819   produce the final HadISDH.marine version without inclusion of moored buoy data.

820

821   Before quality control there are more daytime ship observations than night time ship observations in the early

822   record (~1000000 compared to ~800000 per year) but this evens out by the end of the record to ~900000 per

823   year. However, the quality control removes more daytime observations than night time observations, especially

824   in the 1970s and 1980s such that both contribute ~700000 observations per year, dipping in the middle of the

825   record. There has been no bias adjustment for solar heating of ships applied in this version of HadISDH.marine

826   so the daytime data may contain some artefacts of solar heating. If this is a problem it should affect the air

827   temperature and relative humidity but not the dew point temperature or specific humidity (Sect. 2.1). While the

828   full dataset (both) combines both daytime and night time data, for various gridboxes and seasons there is only

829   either a daytime or night time value present. As such, the 'both' timeseries and its linear trend may not be a

830   straightforward average of the 'day' and 'night' timeseries and trends. In the case of specific humidity, the

831   daytime and night time global average timeseries have slightly larger positive trends than the combined

832   timeseries and for relative humidity they have smaller negative trends than the combined series. For specific

833   humidity, dew point temperature and air temperature the 'day' and 'night' trend differences are essentially

834   negligible, with linear trend differences within 0.01 g kg $^{-1}$ decade $^{-1}$ or 0.01° C decade $^{-1}$. Even for relative

835   humidity the differences are small. The 'day' timeseries gives the largest negative trend followed by 'both'

836   which is 0.01 %rh decade $^{-1}$ smaller and then 'night' which is 0.02 %rh decade $^{-1}$ smaller again. The negligible

837   differences in air temperature suggest that solar heating is not a significant concern at least at the global average

838   scale. Relative humidity is very sensitive to any differences in the data but even these differences are fairly

839   small and do not change the overall conclusion of decreasing long-term trends. 'Night' trends are often thought

840   to provide a better signal of change because they are generally free from convective and shortwave radiative

841   processes and more a measure of outgoing longwave radiation. The main conclusion here is that trends and

842   variability are very similar in the daytime, nighttime and combined timeseries which adds confidence in their

843   representativeness of real-world trends and variability.

844

845   Overall, at least in terms of linear trend direction, HadISDH.marine compares well with other monitoring

846   estimates from NOCSv2.0 and ERA-Interim and to other reanalyses and older products (Fig. 1). ERA-Interim

847   has been masked to ocean coverage using a 1° by 1° land-sea mask and also to HadISDH.marine coverage for

848   comparison. Note that the ERA-Interim timeseries shown in Figs. 8 to 11 are from analysis fields of 2 m air

849   temperature and dew point temperature, whereas the timeseries shown in Fig. 1 are from background forecast

850   values to avoid biases introduced from ship data and ocean-only points over open sea. They are very similar at

851   least in terms of the global average. Both NOCSv2.0 and HadISDH.marine are estimates of 10 m quantities and

852   the NOCSv2.0 coverage is similar to that of HadISDH.marine but it only extends to 2015. NOCSv2.0 shows the

853   largest trends in specific humidity over the 1979-2015 common period, 0.03 g kg $^{-1}$ decade $^{-1}$ greater than

854   HadISDH.marine. The interannual patterns are broadly similar but with some differences showing that

855   methodological choices do make a difference, given that the underlying observations are from the same source.

856   ERA-Interim shows very weak moistening compared to HadISDH.marine for specific humidity and dew point

857   temperature and slightly weaker warming for air temperature. Over the longer 1979-2018 period ERA-Interim

858   trends are slightly larger for specific humidity but still weaker than in HadISDH.marine. The decreasing

859   saturation in relative humidity is very strong in ERA-Interim at more than 3 times the HadISDH.marine trend

860   over the common period. The masking to HadISDH.marine coverage surprisingly makes very little difference in

861   the linear trends and only small year-to-year differences. Interannual behaviour does differ, especially for

862   relative humidity and especially in the period up to the early 1990s where ERA-Interim is warmer and wetter

863   generally, thus moderating the long-term trends in specific humidity, dew point temperature and air temperature.

864   Note that the ERA-Interim background field relative humidity shown in Fig, 1 also shows a decrease but to a

865   lesser extent than the analysis fields (Fig. 9) which include ship data. Agreement is closest for air temperature in

866   both trends and variability.

867

868   The decreasing relative humidity trends over ocean are consistent with the drying seen in HadISDH.land and

869   ERA-Interim land relative humidity (Fig. 1). The timeseries pattern is quite different though with marine

870   relative humidity decreasing throughout the period around large variability and land relative humidity clearly

871   decreasing from 2000. The greater sensitivity of relative humidity to observation errors, biases and sampling

872   issues makes the conclusion of long-term drying an uncertain one but agreement with ERA-Interim adds weight

873   to this conclusion.

874

875   For the final HadISDH.marine.1.0.0.2018f product the regional average uncertainty is also computed and shown

876   for the global average in Fig. 12. This includes the total observation uncertainty, sampling uncertainty and also a

877   spatial coverage uncertainty, following the method applied for HadISDH.land (Willett et al., 2014). The

878   coverage uncertainty essentially uses the variability between ERA-Interim full coverage compared to ERA-

879   Interim with HadISDH.marine coverage to estimate uncertainty. To obtain uncertainty in the global average

880   from the gridbox uncertainties correlation in time and space should be taken into account. It is not trivial to

881   assess the true spatial and temporal correlation of the various uncertainty sources. In reality, although ships

882   move around over space and time, implying some correlation, the contributing sources to each ~500 km^2

883   gridbox monthly mean differ widely. Therefore, for this first version product we assume no correlation between

884   gridboxes in time or space and take the simple approach of the quadrature combination of uncertainty sources,

885   noting that this is a lower limit on uncertainties.

886

887   The uncertainty in the global averages (Fig. 12) are larger than the equivalent time series for land (see Fig. 12 in

888   Willett et al., 2014). The coverage uncertainty (accounting for observation gaps in space and time) is generally

889   the largest source of uncertainty with the exception of relative humidity and dew point depression. For the latter

890   two, the total observation uncertainty makes up the greatest contribution. In all cases the total observation

891   uncertainty is larger at the beginning and especially the end of the records, where there are fewer/no metadata

892   with which to apply bias adjustments. The contribution from sampling uncertainty (gridbox spatial and temporal

893   coverage) is generally very small except for relative humidity. This is as expected given that the correlation

894   decay distance of humidity should generally be larger over ocean than over land given the homogeneous surface

895   altitude and composition. Overall, the magnitudes of the uncertainties are small relative to the magnitudes of

896   long-term trends and variability in all variables except for relative humidity and dew point depression. This

897   suggests that there is good confidence in changes in absolute measures of humidity over ocean (e.g., specific

898   humidity), and also air temperature, but lower confidence in changes in the relative humidity. The warming and

899   moistening are further corroborated by strong theoretical reasoning based on laws of physics governing the

900   expectation that specific humidity should have increased over the period of record given the warming of the

901   oceans and atmosphere that has occurred (Hartmann et al., 2013). The uncertainty model makes many

902   assumptions over correlation of uncertainty in space and time. It is likely that we have overestimated the

903   uncertainty at the gridbox scale by assuming complete correlation for height adjustment uncertainty, instrument

904   adjustment uncertainty and climatological uncertainty. Conversely, we have likely underestimated the

905   uncertainty at the regional average level by assuming no correlation. This is certainly an area for improvement

906   in future versions.

907   **4.3 Decadal trends across the globe presented by HadISDH.marine**

908

909   Figure 13 shows the decadal linear trends for specific humidity, relative humidity, dew point temperature and air

910   temperature for HadISDH.marine.1.0.0.2018f. The completeness criteria for trend fitting is 70 %, more strict

911   than for the climatologies (Fig. 7). This results in poorer spatial coverage especially in the Southern

912   Hemisphere. Clearly, there are no data points outside 70° S to 70° N, hence the restriction of the global average

913   timeseries to this region is sensible. The tropical and Southern Hemisphere Pacific Ocean, and Southern

914   Hemisphere Atlantic Ocean have virtually no data coverage. Overall, the appearance of the trends shows good

915   spatial consistency, with few gridboxes standing out as obviously erroneous. There has been no interpolation

916   across gridboxes that would have smoothed out any outliers, and so the lack of these outlying gridboxes

917   suggests that the data are of reasonable quality for this long-term analysis at least. Trends are as expected from

918   the global average timeseries – generally moistening and warming but becoming less saturated. The same is true

919   over land (Willett et al., 2014).

920

921   The moistening shown in specific humidity and dew point temperature (Fig. 13 panels a, b and e, f) is

922   widespread. The majority of gridboxes are considered to be statistically significant in that the 90th percentile

923   confidence interval around the trend magnitude is the same sign as the trend and does not encompass zero. The

924   largest increases in specific humidity are in the lower latitudes where as the largest increases in dew point

925   temperature are more spread out with a tendency towards the extratropics and mid-latitudes. There are a few

926   regions where there are clusters of gridboxes with drying trends. These are generally consistent between the

927   specific humidity and dew point temperature, especially in the few cases where these negative trends are

928   significant such as the central Pacific, the east coast of Brazil, the southern coast of Australia and around New

929   Zealand.

930

931   Marine air temperature shows widespread and significant warming, in agreement with HadNMAT (Kent et al.,

932   2013). Very few of the gridboxes with a negative trend are significant. In some cases they are in similar

933   locations to the drying trends seen in specific humidity and/or dew point temperature e.g., the coast south of

934   Australia around Tasmania, the east coast of Brazil. The warming is stronger in the northern mid latitudes with

935   the Baltic, Mediterranean and Red Seas showing particularly strong warming consistent with strongly increasing

936   dew point temperature and specific humidity.

937

938   Whilst relative humidity is more sensitive to methodological choices and observational errors, the broad

939   spatially coherent structures to the regions of increasing and decreasing saturation, with broadscale significance,

940   are very encouraging in terms of data quality. Furthermore, the drying trends tend to be around the mid-latitudes

941   while the increasing saturation trends are more around the tropics, as seen over land. We still urge caution in the

942   use of marine relative humidity but these results collectively suggest that decreasing saturation might be a real

943   feature.

944

945   **5 Code and data availability**

946

947   HadISDH.marine is available as 5° by 5° gridded fields of monthly means and anomalies along with a 1981-

948   2010 climatology and uncertainty estimates at the gridbox scale. The data begin in January 1973 and continue to

949   December 2018 (at time of print) and will be updated annually. HadISDH.marine is publicly available from

950   www.metoffice.gov.uk/hadobs/hadisdh/ under an Open Government license

951   (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/) as netCDF and text files.

952   Processing code (Python) can also be made available on request. HadISDH.marine data, derived diagnostics and

953   plots can be found at www.metoffice.gov.uk/hadobs/hadisdh/indexMARINE.html (Willett et al., 2019). It

954   should be cited using this paper and the following: Willett, K., Dunn, R., Kennedy, J. and Berry, D.:

955   HadISDH.marine: gridded global monthly marine surface humidity data (version 1.0.0.2018f) [Data set]. Met

956   Office Hadley Centre HadOBS Datasets, www.metoffice.gov.uk/hadobs/hadisdh/indexMARINE.html, 2019.

957

958   This product forms one of the HadOBS (www.metoffice.gov.uk/hadobs) climate monitoring products and will

959   be blended with the HadISDH.land product to create a global land and marine humidity monitoring product.

960   Updates and exploratory analyses are documented at http://hadisdh.blogspot.co.uk and through the Met Office

961   HadOBS twitter account @metofficeHadOBS.

962

963   **6 Discussion and conclusions**

964

965   Marine humidity data are susceptible to a considerable number of biases and sources of error that can be large in

966   magnitude. We have cleaned the data where possible by applying quality control for outliers, supersaturation,

967   repeated values and neighbour inconsistency which has removed up to 25 % of our initial selection in some

968   years. We have also applied adjustments to account for biases arising from un-aspirated instrument types and

969   differing observation heights over space and time. Care has also been taken to avoid diurnal and seasonal

970   sampling biases as far as possible when building the gridded fields and the use of gridbox mean climate

971   anomalies reduces remaining random error through averaging

972

973   Spatial coverage of HadISDH.marine differs year to year. The coverage is generally poorer than seen for

974   variables such as SST which benefit significantly from drifting buoy observations. Any further decline in

975   observation and transmission of humidity from ships is of concern to our ability to robustly monitor surface

976   humidity over oceans. Future versions may be able to make more use of humidity data from buoys but their

977   proximity to the sea surface and difficulty of regular maintenance can lead to poor quality observations. The

978   provision of digital metadata significantly improves our ability to quantify and account for biases in the data.

979   Hence, the continuity of this metadata beyond 2014, and ideally an increase in quantity, also strongly affects our

980   ability to robustly monitor ocean surface humidity. Given the current availability of ship data and metadata, and

981   necessarily strict selection criteria and quality control, the resulting spatial coverage is good over the Northern

982   Hemisphere outside of the high latitudes. There is very poor coverage over the Southern Hemisphere, especially

983   south of 20° S. This means that our 'global' analyses are biased to the Northern Hemisphere. Care should be

984   taken to account for different spatial coverage when comparing products. However, when comparing HadISDH

985   to masked and unmasked ERA-Interim fields differences were surprisingly small.

986

987   We have shown that the observations are warm and moist relative to ERA-Interim reanalysis for the majority of

988   the observed globe apart from the northwestern Pacific. This is despite ERA-Interim fields representing 2 m

989   above the surface compared to the general observation heights of 10-30 m above the surface. Differences are

990   largest around coastlines, particularly in the Red Sea and Persian Gulf. There is insufficient spatial coverage to

991   produce a high resolution climatology from the data themselves, hence our use of ERA-Interim initially and then

992   interpolated observation based fields. However, the lower resolution (5° by 5°) monthly mean climatologies

993   from the final HadISDH.marine.1.0.0.2018f version show expected spatial patterns and have good spatial

994   consistency, providing evidence that our data selection methods have resulted in reasonably high quality data.

995

996   The quality control and bias adjustment procedures have made small differences to the global average anomaly

997   timeseries for specific humidity, dew point temperature and air temperature. This overall agreement in the

998   global average timeseries between versions, and also between the daytime, night time and combined versions,

999   increases confidence in the overall signal of increased moisture and warmth over oceans. These features show

1000 widespread spatial consistency in the HadISDH.marine.1.0.0.2018f gridbox decadal trends which also adds

1001 confidence. Hence, we can conclude that the ICOADS data are a useful source of humidity data for climate

1002 monitoring. However, we expect differences to be larger on smaller spatial scale analyses. HadISDH.marine

1003 shows consistency with other products in terms of long-term linear trends in the global averages. There are some

1004 differences year to year, with ERA-Interim showing warmer and moister anomalies prior to the early 1990s, and

1005 hence, smaller trends overall. For relative humidity, differences between the versions can be large for any one

1006 year but the overall decreasing saturation trend appears to be robust. We conclude this because the trend is

1007 consistent across all processing steps, apparent in ERA-Interim fields and also has spatial consistency across the

1008 extratropics and mid-latitudes. This is a somewhat surprising result and one that should be treated cautiously.

1009 Model based analysis of changes in relative humidity over ocean under a warming climate suggest negligible or

1010 small positive changes (Byrne and O'Gorman, 2013, 2018). The temporal patterns in global average relative

1011   humidity are quite different to those over land whereas specific humidity shows similarity with the

1012   HadISDH.land timeseries, largely driven by the El Niño related peaks. Further work to assess the physical

1013   mechanisms that might lead to such trends is needed.

1014

1015   There are known issues with ERA-Interim in terms of its stability. For example, sea surface temperatures cooled

1016   around mid-2001 due to a change in the SST analysis product used (Simmons et al., 2014). This is very likely to

1017   affect humidity over the ocean surface in ERA-Interim. Similarly, changes in satellite streams over time can also

1018   affect the long-term stability of ERA-Interim, even in the surface fields. Also, the assimilated ship data are not

1019   adjusted for biases in the ERA-Interim assimilation. Clearly, there are various issues affecting both in-situ based

1020   monitoring products and reanalysis products such that neither one can be easily identified as the more accurate

1021   estimate. Analyses should take into account all available estimates and their strengths and weaknesses.

1022   Comparison of HadISDH.marine with satellite-based estimates of humidity over ocean will be an important next

1023   step.

1024

1025   We have attempted to quantify uncertainty in HadISDH.marine. The uncertainty analysis comprises observation

1026   uncertainty at the point of measurement which is then propagated through to gridbox averages taking correlation

1027   in space and time into account where relevant. Sampling uncertainty at the gridbox level due to uneven

1028   sampling across the gridbox in space and time is assessed. We have also provided uncertainty estimates in

1029   regional and global averages including coverage uncertainty. The propagation of gridbox observation and

1030   sampling uncertainty to large scale averages does not explicitly take into account correlation in these uncertainty

1031   quantities in space and time. As this is a first version monitoring product this simple method is seen as an

1032   appropriate first attempt to assess uncertainty. The ranges presented should be seen as a lower limit on the

1033   uncertainty. Overall, uncertainty in the global average is dominated by the coverage uncertainty for all variables

1034   except relative humidity and dew point depression. The total observation uncertainty is larger at the beginning,

1035   and especially the end of the record, where digital metadata are fewer or non-existent (post-2014). Overall, the

1036   uncertainty is small relative to the magnitude of long-term trends with the exception of relative humidity. We

1037   suspect that this is an over estimate at the gridbox level owing to assumptions of complete correlation in the

1038   height adjustment, instrument adjustment and climatology uncertainty components, and an underestimate at the

1039   regional average level given assumptions of no correlation. This is a first attempt to comprehensively quantify

1040   marine humidity uncertainty and future methodological improvements are envisaged.

We conclude that this first version marine humidity monitoring product is a reasonable estimate of large-scale

trends and variability and contributes to our understanding of climate changes as a new and methodologically-

independent analysis. The trends and variability shown are mostly in concert with expectation; widespread

moistening and warming is observed over the oceans (excluding the mostly data-free Southern Hemisphere)

from 1973 to present. These are also large relative to the magnitude of our uncertainty estimates. Our key

finding is that the marine relative humidity appears to be decreasing (the air is becoming less saturated). We

have explored various processes for ensuring high quality data and shown that these do not make large

differences for large scale analyses of specific humidity, dew point temperature and air temperature but that

there is greater sensitivity to methodological choices for relative humidity.

The spatial coverage of surface humidity data is very low outside of the Northern Hemisphere. If only those data

with digitised metadata are included then this coverage deteriorates further. Although moored buoy numbers

have increased dramatically since the 1990s, their measurements are more prone to error through proximity to

the water, and hence, contamination, in addition to less frequent manual checking and maintenance. Hence, our

ability to monitor surface humidity with any degree of confidence depends on the continued availability of ship

data and provision of digitised metadata.

Author Contributions

Kate Willett undertook the majority of the methodology, coding, writing and plotting. John Kennedy designed
and coded the quality control methodology and software with some contribution from Kate Willett. Robert
Dunn designed and coded the gridding methodology and software with some contribution from Kate Willett.
David Berry designed and reviewed the height adjustment methodology and provided guidance on marine
humidity data biases, inhomogeneities and issues. All authors have contributed text and edits to the main paper.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

Kate Willett, Robert Dunn and John Kennedy were supported by the Met Office Hadley Centre Climate
Programme funded by BEIS and Defra. (GA01101).

References

Berry, D., 2009: Surface forcing of the North Atlantic: accuracy and variability, University of Southampton,
176pp.

Berry, D. I., Kent, E. C. and Taylor, P. K. : An analytical model of heating errors in marine air temperatures
from ships, J. Atmospheric and Oceanic Technology, 21(8), 1198–1215, 2004.

1083   Berry, D. I. and Kent, E. C. : A new air-sea interaction gridded dataset from ICOADS with uncertainty
1084   estimates, Bulletin of the American Meteorological Society, 90, 645-656, 2009.
1085

1086   Berry, D. I. and Kent, E. C.: Air–Sea fluxes from ICOADS: the construction of a new gridded dataset with
1087   uncertainty estimates, Int. J. Climatol., 31, 987–1001, 2011.
1088

1089   BIPM, 2008: Evaluation of measurement data – Guide to the expression of uncertainty in measurement. JCGM
1090   100:2008. https://www.bipm.org/en/publications/guides/gum.html
1091

1092   Bosilovich, M. G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A.,
1093   Norris, P., Wargan, K., Chao, W., Reichle, R., Takacs, L., Vikhliaev, Y., Bloom, S., Collow, A., Firth, S.,
1094   Labow, G., Partyka, G., Pawson, S., Reale, O., Schubert, S. D. and Suarez, M. : MERRA-2: Initial Evaluation of
1095   the Climate, Technical Report Series on Global Modeling and Data Assimilation, Volume 43, NASA/TM–2015-
1096   104606/Vol. 43, pp. 136. http://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/, 2015.
1097

1098   Buck, A. L. : New equations for computing vapor pressure and enhancement factor, J. Appl. Meteor., 20, 1527–
1099   1532, 1981.
1100

1101   Byrne, M. P. and P. A. O'Gorman, 2013: Link between land-ocean warming contrast and surface relative
1102   humidities in simulations with coupled climate models. Geophysical Ressearch Letters, 40 (19), 5223-5227,
1103   https://doi.org/10.1002/grl.50971.
1104

1105   Byrne, M. P. and O'Gorman, P. A.: Trends in continental temperature and humidity directly linked to ocean
1106   warming, Procedings of the National Academy of Sciences USA. 115(19), 4863-4868. doi:
1107   10.1073/pnas.1722312115, 2018.
1108

1109   Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses
1110   of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), February 2019.
1111   https://cds.climate.copernicus.eu/cdsapp#!/home
1112

1113   Dai, A. : Recent climatology, variability, and trends in global surface humidity, J. Climate., 19, 3589–3606,
1114   2006.
1115

1116   Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.
1117   A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L. J., Bidlot, L., Bormann, N., Delsol,
1118   C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L.,
1119   Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K.,
1120   Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis:
1121   configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597,
1122   doi:10.1002/qj.828, 2011.
1123

1124   Ebita, A., and co-authors, : The Japanese 55-year reanalysis "JRA-55": An interim report. SOLA, 7, 149-152,
1125   doi:10.2151/sola.2011-038, 2011.
1126

1127   Elliott, W. P., Ross, R. J. and Schwartz, B. : Effects on climate records of changes in National Weather Service
1128   humidity processing procedures, Journal of Climate, 11, 2424-2436, 1998.
1129

1130   Freeman, E.; Woodruff, S. D.; Worley, S. J.; Lubker, S. J.; Kent, E. C.; Angel, W. E.; Berry, D. I.; Brohan, P.;
1131   Eastman, R.; Gates, L.; Gloeden, W.; Ji, Zaihua; Lawrimor, J.; Rayner, N. A.; Rosenhagen, G.; Smith, S. R. :
1132   ICOADS Release 3.0: a major update to the historical marine climate record, International Journal of
1133   Climatology, 37 (5). 2211-2232.10.1002/joc.4775, 2017.
1134

1135   Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A.,
1136   Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty,
1137   A., da Silva, A. M., Gu, W., Kim, G., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G.,
1138   Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M. and Zhao, B. : The Modern-Era
1139   Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate, 30, 5419–
1140   5454,https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.
1141

Gilhousen, D., : A Field evaluation of NDBC Moored Buoy Winds, Journal of Atmospheric and Oceanic Technology, 4, 94 – 104, 1987.

Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M. and Zhai, P. M.: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 159 – 254, doi:10.1017/CBO9781107415324.008, 2013.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janisková, M., Laloyaux, P., Lopez, P., Muñoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J.-N., Vitart, F., Yang, X., Zsótér, E. and Zuo, H.: Operational global reanalysis: progress, future directions and synergies with NWP, ERA Report 27, 63pp. Available from www.ecmwf.int, 2018.

Jensen, M. E., Burman, R. D. and Allen, R. G.: Evapotranspiration and Irrigation Water Requirements: A Manual. American Society of Civil Engineers, 332 pp, 1990

Jones, P. D., Osborn, T. J., and Briffa, K. R. : Estimating sampling errors in large-scale temperature averages, Journal of Climate, 10, 2548-2568, 1997.

Josey, S. A., Kent, E. C. and Taylor, P. K.: New insights into the ocean heat budget closure problem from analysis of the SOC air–sea flux climatology, J. Climate, 12, 2685–2718, 1999.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Saunby, M. and Parker, D. E. : Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: measurement and sampling errors, J. Geophys. Res., 116, D14103, doi:10.1029/2010JD015218, 2011a.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Saunby, M. and Parker, D. E. : Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation, J. Geophys. Res., 116, D14104, doi:10.1029/2010JD015220, 2011b.

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., & Killick, R. E. : An ensemble data set of sea-surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set, Journal of Geophysical Research: Atmospheres, 124. https://doi.org/10.1029/2018JD029867, 2019.

Kent, E. C. and Challenor, P. G. : Towards estimating climatic trends in SST. Part II: random errors, Journal of Atmospheric and Oceanic Technology. 23, 476-486. DOI: 10.1175/JTECH1844.1, 2006.

Kent, E. C., and Taylor, P. K. : Accuracy of humidity measurement on ships: Consideration of solar radiation effects, J. Atmos. Oceanic Technol., 13, 1317–1321, 1996.

Kent, E. C., Tiddy, R. J. and Taylor, P. K.: Correction of marine air temperature observations for solar radiation effects, J. Atmos. Oceanic Technol., 10, 900–906, 1993.

Kent, E. C., Woodruff, S. D. and Berry D. I. : Metadata from WMO Publication No. 47 and an Assessment of Voluntary Observing Ship Observation Heights in ICOADS, J. Atmospheric and Oceanic Technology 2007 24:2, 214-234, doi: http://dx.doi.org/10.1175/JTECH1949.1, 2007.

Kent, E. C., Rayner, N. A., Berry, D. I., Saunby, M., Moat, B. I., Kennedy, J. J. and Parker, D. E. : Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set, J. Geophys. Res. Atmos., 118, doi:10.1002/jgrd.50152, 2013.

Kent, E. C., Berry, D. I., Prytherch, J., Roberts, J. B. : . International Journal of Climatology, 34 (2). 355-376.https://doi.org/10.1002/joc.3691, 2014.

Met Office Hadley Centre; National Oceanography Centre, 2019: HadISDH.marine: gridded global monthly marine surface humidity data version 1.0.0.2018f. Centre for Environmental Data Analysis, *date of citation*. doi:xx.xxxx/XXXXX. http://dx.doi.org/xx.xxxx/XXXXX (FINALISED AFTER REVIEW)

Peixoto, J. P., and Oort, A. H. : The climatology of relative humidity in the atmosphere, J. Climate, 9, 3443–3463, 1996.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., Kaplan, A. : Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research – Atmospheres, 108, No. D14, 4407. https://doi.org/10.1029/2002JD002670, 2003.

Rayner, N., Brohan, P., Parker, D., Folland, C., Kennedy, J., Vanicek, M., Ansell, T. and Tett, S. : Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 data set, J. Clim., 19(3), 446–469, doi:10.1175/JCLI3637.1, 2006.

Simmons, A., Willett, K. M., Jones, P. D., Thorne, P. W., and Dee, D.: Low-frequency variations in surface atmospheric humidity, temperature and precipitation: inferences from reanalyses and monthly gridded observational datasets, *J. Geophys. Res.*, 115, D01110, doi:10.1029/2009JD012442, 2010.

Simmons, A. J., Poli, P., Dee, D. P., Berrisford, P., Hersbach, H., Kobayashi S. and Peubey, C. : Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim, Q.J.R. Meteorol. Soc., 140: 329-353. doi:10.1002/qj.2317, 2014.

Smith, S. D,: Wind stress and heat flux over the ocean in gale force winds. J. Physical Oceanography, 10, 709-726, 1980.

Smith, S. D. : Coefficients for sea surface wind stress, heat flux and wind profiles as a function of wind speed and temperature, J. Geophys. Res., 93, 15467-15472, 1988.

Stull, R. B. : An Introduction to Boundary Layer Meteorology Klewer Academic Publishers, 666 pp, 1988.

Wade, C. G.,: An evaluation of problems affecting the measurement of low relative humidity on the United States radiosonde, Journal of Atmospheric and Oceanic Technology, 11, 687-700, 1994.

Willett, K. M., Jones, P. D., Gillett N. P. and Thorne, P. W. : Recent changes in surface humidity: development of the HadCRUH dataset, J. Climate, 21, 5364–5383, 2008.

Willett, K. M., Williams Jr., C. N., Dunn, R. J. H., Thorne, P. W., Bell, S., de Podesta, M., Jones, P. D. and Parker, D. E. : HadISDH: An updated land surface specific humidity product for climate monitoring. Climate of the Past, 9, 657-677, doi:10.5194/cp-9-657-2013, 2013.

Willett, K. M., Dunn, R. J. H., Thorne, P. W., Bell, S., de Podesta, M., Jones, P. D., Parker, D. E. and Williams Jr., C. N. : HadISDH land surface multi-variable humidity and temperature record for climate monitoring, Climate of the Past, 10, 1983-2006, doi:10.5194/cp-10-1983-2014, 2014.

Willett, K., Dunn, R., Kennedy, J. and Berry, D: HadISDH.marine: gridded global monthly marine surface humidity data (version 1.0.0.2018f) [Data set]. Met Office Hadley Centre HadOBS Datasets, www.metoffice.gov.uk/hadobs/hadisdh/indexMARINE.html, 2019.

Willett, K. M., Berry, D., Bosilovich, M. and Simmons, A. : [Global Climate] Surface Humidity [in "State of the Climate in 2018"], Bulletin of the American Meteorological Society, accepted, 2019.

Wolter, K., : Trimming problems and remedies in COADS, Journal of climate. 10. 1980-1997. DOI: 10.1175/1520-0442(1997)010<1980:TPARIC>2.0.CO;2, 1997.

Woodruff, S., : Archival of data other than in IMMT format: The International Maritime Meteorological Archive (IMMA) format. NOAA Earth System Research Laboratory (ESRL), Boulder, CO, USA. http://icoads.noaa.gov/e-doc/imma/R2.5-imma.pdf, 2010.

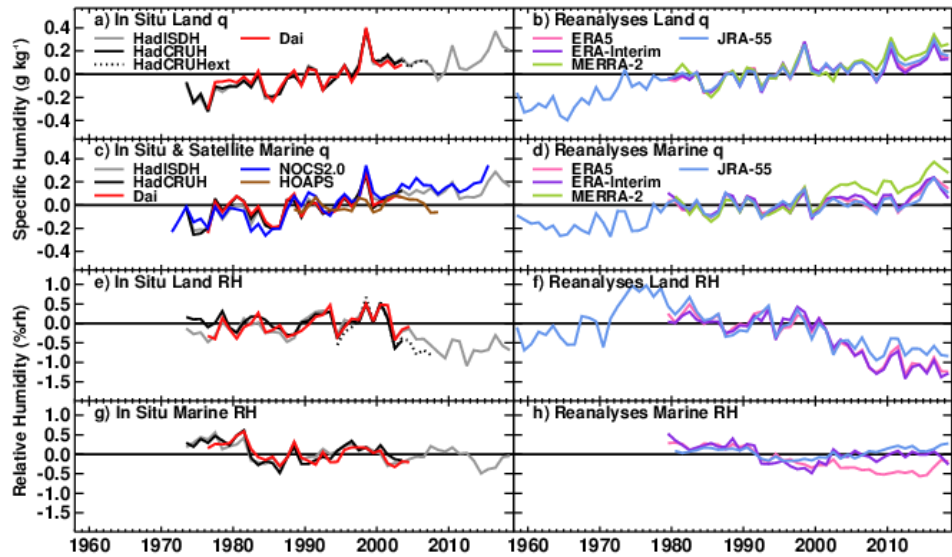
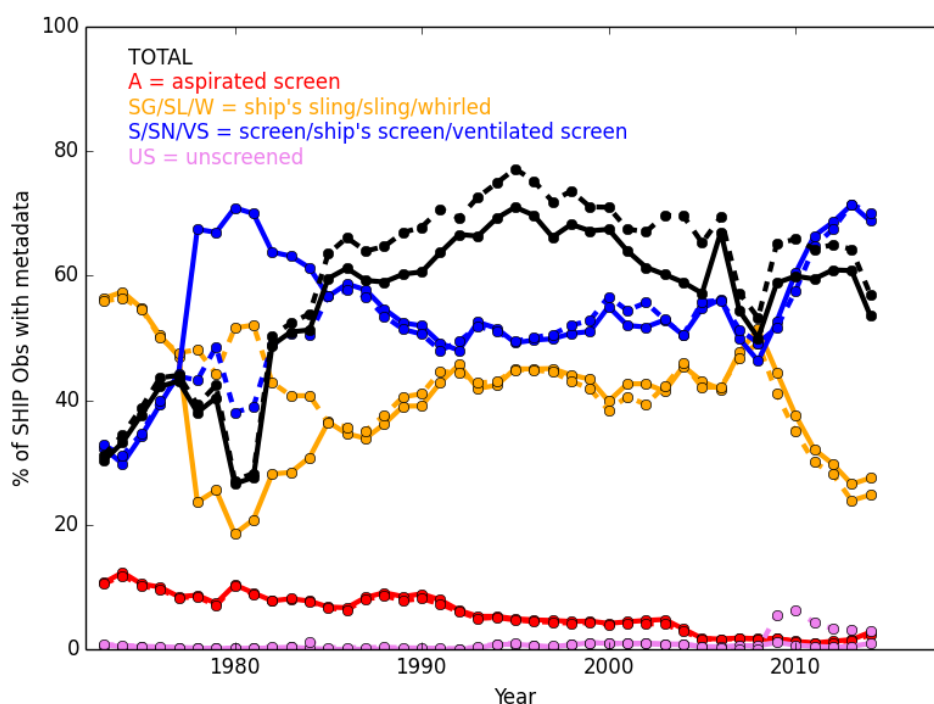
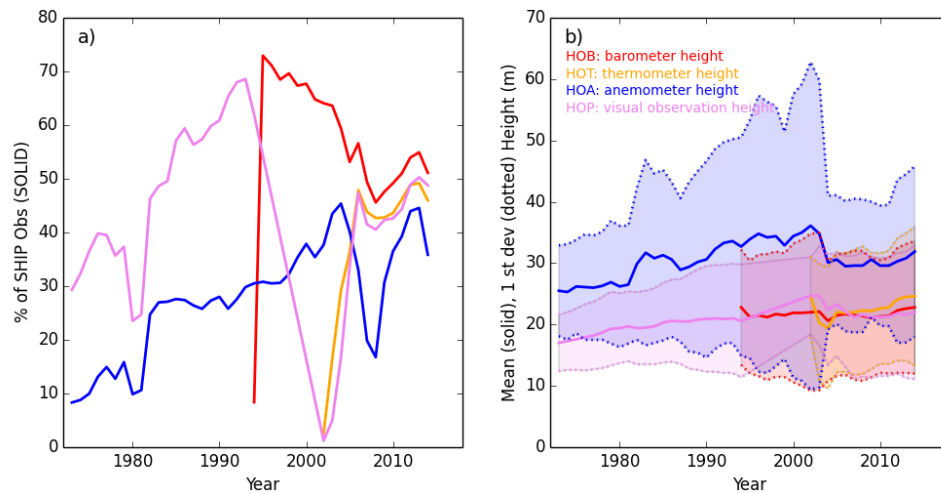
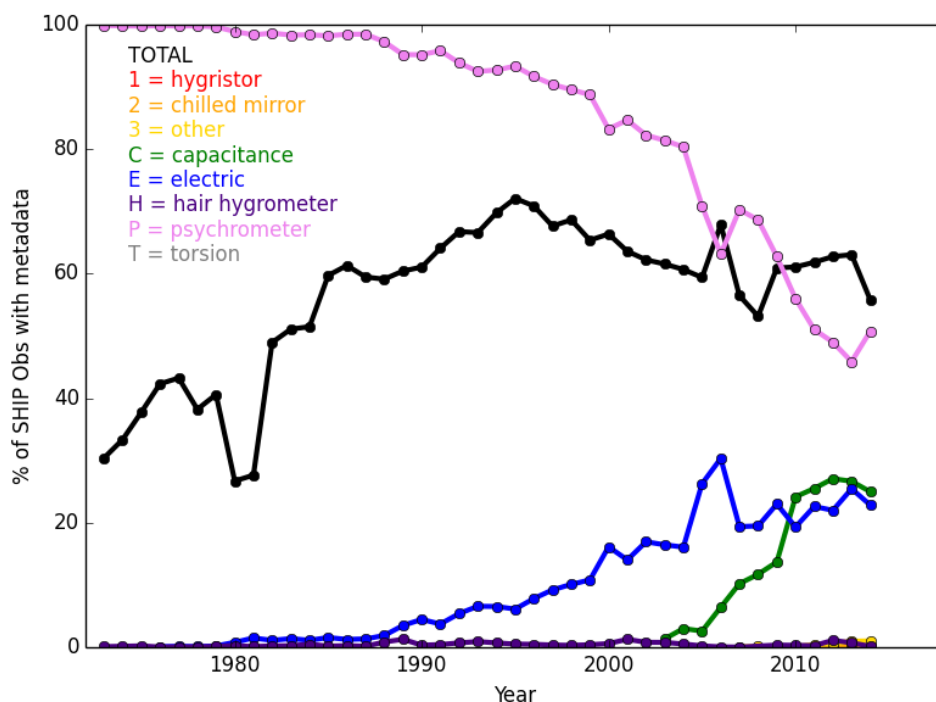
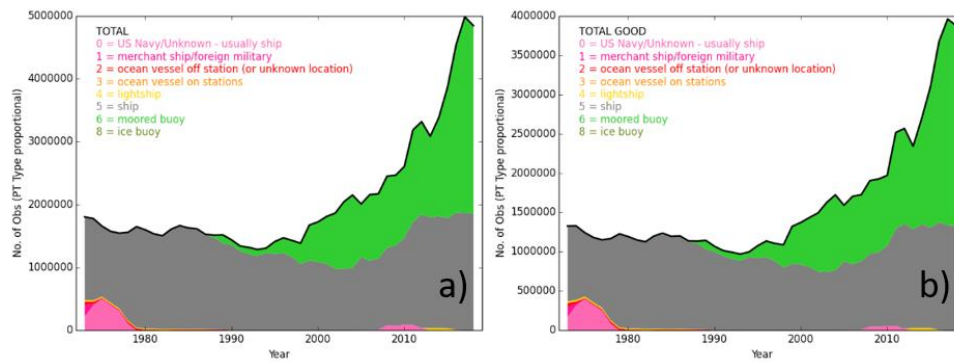
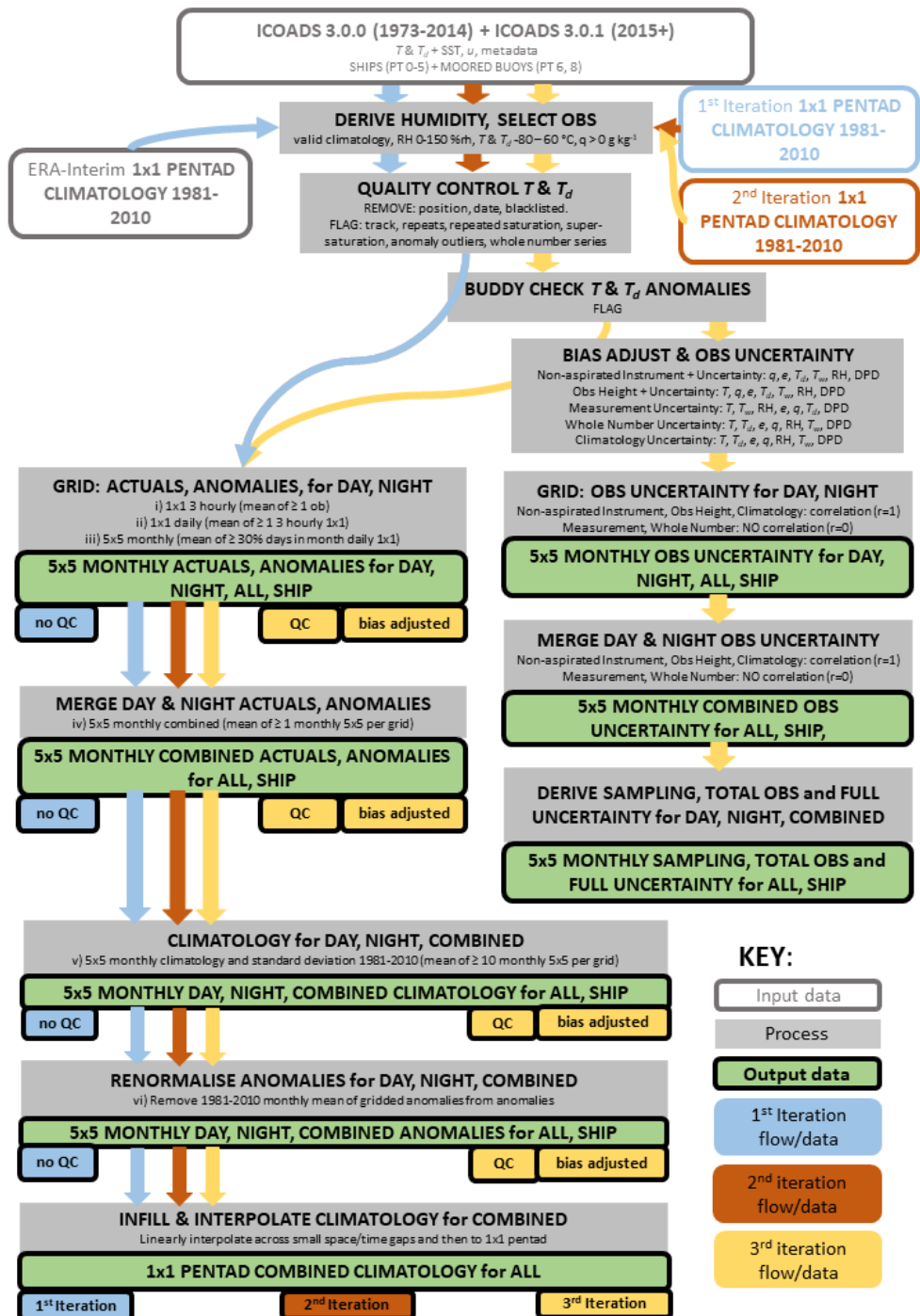
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313 **Tables**
1314
1315 Table 1. Description of the uncertainty elements affecting marine humidity. All uncertainties are assessed as 1σ
1316 uncertainty.

Uncertainty Source	Description	Type	Formula	Correlation

U_i	Non-aspirated instrument adjustment uncertainty. Expressed as q (g kg^{-1}) and then propagated to other humidity variables	Adjusted poorly aspirated instrument: 0.2 g kg^{-1} in terms of q (following Berry and Kent, 2011 standard uncertainty assessment).	Standard	0.2	Space and time, r = 1
		Partially adjusted unknown instrument: 0.2 g kg^{-1} + the full adjustment amount in terms of q.		$0.2 + 100\left(\dfrac{q_{adj} - q}{55}\right)$	
U_h	Observation height adjustment uncertainty. Expressed as T (°C) and q (g kg^{-1}) and then propagated to other humidity variables	Height adjusted ship and valid SST: assessed using the range of adjustments from a 1σ uncertainty in the height estimate.	Normally distributed	$\dfrac{\left(\dfrac{xH_{max} - xH_{min}}{2}\right)}{\sqrt{9}}$	Space and time, r = 1
		Height adjusted ship and invalid SST or height adjusted buoy: the larger of the adjustment value or 0.1 °C in terms of T and 0.007q.		$\dfrac{x_{adj}}{\sqrt{9}}$ Or 0.1 °C in terms of T 0.007q_{adj}	
		Height adjustment or uncertainty range not resolved, valid SST: half of the difference between the observation value and the surface value (SST or q$_{sf}$).		$\dfrac{T_{(adj)} - SST}{2}$ $\dfrac{q_{(adj)} - q_{sf}}{2}$ $q_{sf} = 0.98q_{sat}f(SST)$	
		Height adjustment or uncertainty range not resolved, no valid SST: 0.1 °C in terms of T and 0.007q.		0.1 °C in terms of T 0.007q_{adj}	

U_m	Measurement uncertainty. Expressed as T (°C), T_w (°C) and RH (%rh) and then propagated to other humidity variables.	Standard uncertainty in the thermometer (T) and psychrometer (T_w) is 0.2 °C and 0.15 °C respectively. This equates in an uncertainty in RH dependent on T.	Standard	0.2 °C in terms of T 0.15 °C in terms of T_w x %rh depending on the temperature and RH bins in Table S3	None, r = 0
U_w	Whole number uncertainty. Expressed as T (°C) and T_d (°C) and then propagated to other humidity variables.	Observation either has the Whole Number flag set or is a whole number and from a red listed source deck in Table S4.	Uniformly distributed	$\dfrac{0.5}{\sqrt{3}}$	None, r = 0
		If both T and T_d are offending whole numbers then RH, T_w and DPD have a combined uncertainty.		$\dfrac{1}{\sqrt{3}}$	
U_c	Climatology uncertainty. Assessed for each variable independently.	The 1° by 1° pentad gridbox climatological standard deviation for the variable is divided by the square root of the number of observations used to create it.	Standard	$\dfrac{\sigma_{clim}}{\sqrt{N_{obs}}}$	Space and time, r = 1
U_{og}	Total observation uncertainty of the gridbox	All gridbox observation uncertainty sources are combined, assuming no correlation between sources.	Standard	$\sqrt{U_i{}^2 + U_h{}^2 + U_m{}^2 + U_w{}^2 + U_c{}^2}$	Space and time to some extent, decreasing with space and time
U_{sg}	Temporal and spatial sampling uncertainty of the gridbox	Sampling uncertainty follows Jones et al., (1997) depending on the mean	Standard	$\dfrac{\left(\bar{s}_i^2 \bar{r}(1 - \bar{r})\right)}{(1 + (N_s - 1)\bar{r})}$	Space and time to some extent, decreasing

		'station' variance, the mean inter-site correlation and the number of 'stations' contributing to the gridbox.			with space and time
U_{fg}	Full uncertainty of the gridbox	All gridbox uncertainty sources are combined, assuming no correlation between sources.	Standard	$\sqrt{U_{og}{}^2 + U_{sg}{}^2}$	Space and time to some extent, decreasing with space and time

1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355 **Figures**
1356

1357
1358 Figure 1 Global average surface humidity annual anomalies (base period: 1979–2003). For in-situ datasets, 2-m
1359 surface humidity is used over land and ~10-m over the oceans. For the reanalysis, 2-m humidity is used across
1360 the globe. For ERA-Interim and ERA5, ocean-only points over open sea are selected and background forecast
1361 values are used as opposed to analysis values to avoid incorporating biases from unadjusted ship data. All data
1362 have been given a mean of zero over the common period 1979–2003 to allow direct comparison, with HOAPS
1363 given a zero mean over the 1988–2003 period. [Sources: HadISDH (Willett et al., 2013, 2014); HadCRUH
1364 (Willett et al., 2008); Dai (Dai 2006); HadCRUHext (Simmons et al. 2010); NOCSv2.0 (Berry and Kent, 2009,
1365 2011); HOAPS (Fennig et al. 2012), ERA-Interim (Dee et al., 2011), ERA5 (C3S 2017, Hersbach et al., 2018),
1366 MERRA-2 (Gelaro et al. 2017; Bosilovich et al. 2015) and JRA-55 (Ebita et al. 2011). Adapted from Willett et
1367 al., 2019.
1368
1369

1370
1371 Figure 2 Availability of instrument exposure information (black) for ships (platform (PT) = 0, 1, 2, 3, 4, 5) for
1372 the hygrometer (EOH, SOLID) and thermometer (EOT, DASHED) for each year. All ICOADS 3.0.0/3.0.1
1373 observations passing 3rd iteration quality control are included. The percentage of EOHs/EOTs in each exposure
1374 category is also shown. Aspirated (A) screens are shown in red. Handheld instruments (ship's sling [SG], sling
1375 [SL], whirling [W]) are shown in orange. Unaspirated/unventilated screens (S) and ship's screens (SN) are
1376 shown in blue. Additionally, ventilated screens (VS) are also shown in blue as these are generally not artificially
1377 aspirated. Unscreened (US observations are shown in violet.
1378
1379
1380
1381
1382

1383
1384 Figure 3 a) Availability of instrument height information for ships (platform (PT) = 0, 1, 2, 3, 4, 5) for the
1385 barometer (HOB), thermometer (HOT), anemometer (HOA) and visual observing platform (HOP) with b) mean
1386 heights (solid lines) and standard deviations (dotted lines) for each year. All ICOADS 3.0.0/3.0.1 observations
1387 passing 3rd iteration quality control are included.
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418

1419
1420   Figure 4 Availability of instrument type information (black) for ships (platform (PT) = 0, 1, 2, 3, 4, 5) for the
1421   hygrometer (TOH) for each year. All ICOADS 3.0.0/3.0.1 observations passing 3rd iteration quality control are
1422   included. The percentage of TOHs in each type category is also shown.
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437

1438
1439   Figure 5 Annual observation count for the initial selection (a) and only those observations passing the final 3rd
1440   iteration quality control (b), broken down by platform (PT) type.
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474

1475
1476   Figure 6 Flow chart of the build process from raw hourly observations to gridded fields.
1477
1478

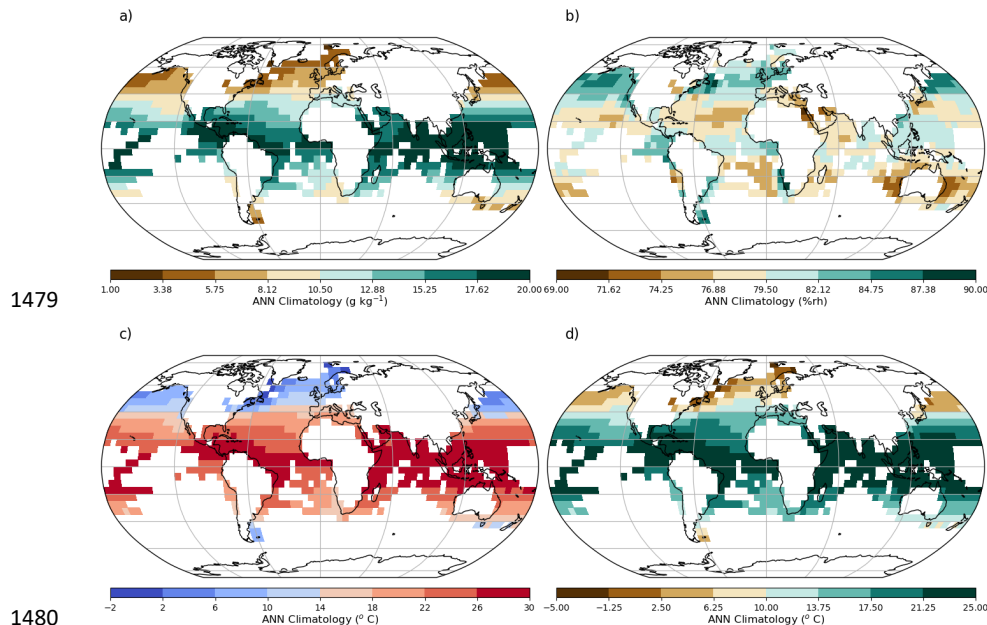
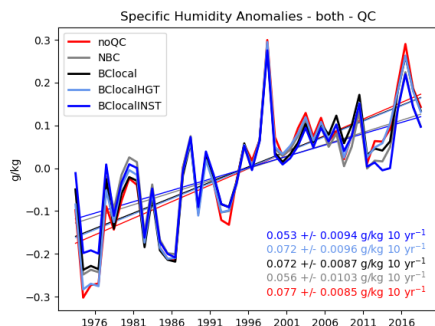
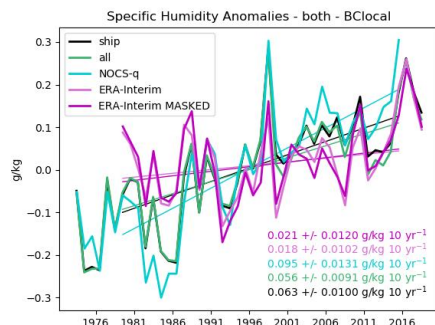
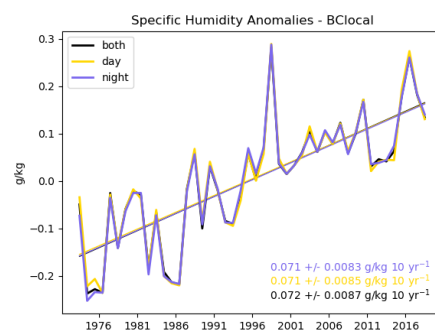
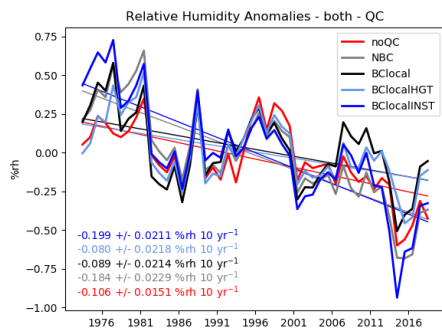
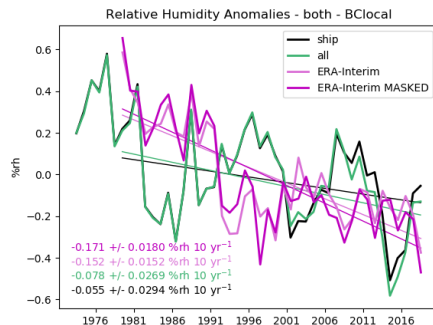
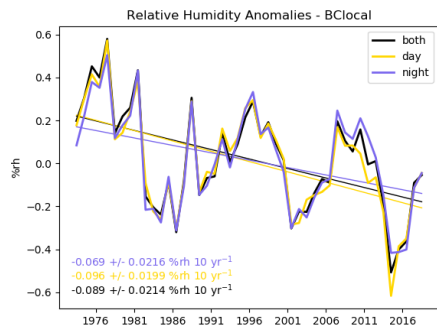
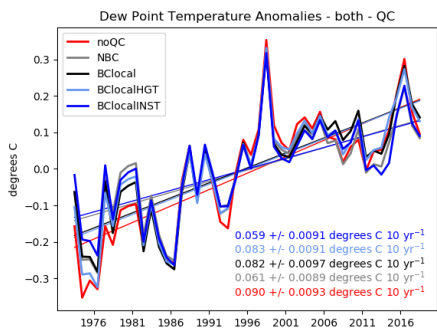
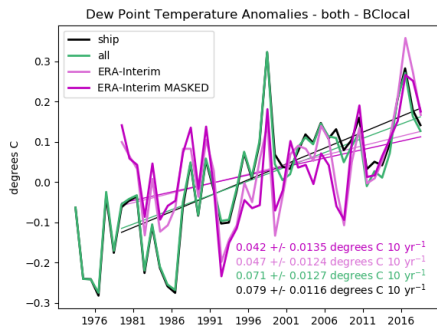
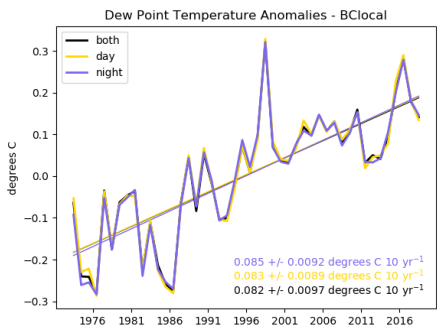
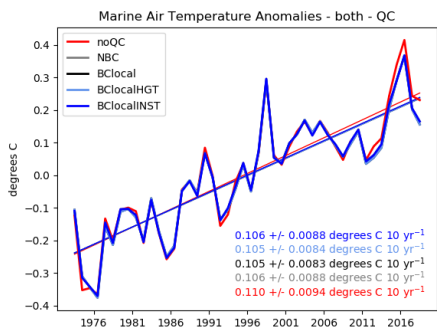
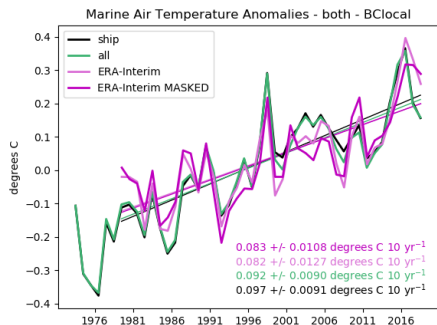
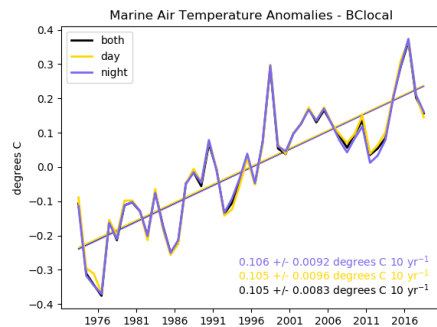



Figure 7 Annual mean climatologies relative to 1981-2010 for a) specific humidity (g kg⁻¹), b) relative humidity (%rh), c) air temperature (° C) and d) dew point temperature (° C) for 3rd iteration quality-controlled and bias-adjusted ship version. Climatological means are calculated for gridboxes and months with at least 10 years present over the climatology period. Annual mean climatologies require at least 9 months of the year to be represented climatologically.

1503   **a)**

1504   **b)**

1505   **c)**
1506   Figure 8 Global (70 °S to 70 °N) annual average anomaly timeseries and decadal trends (+/- 90 % confidence
1507   interval) for specific humidity. a) Processing comparison for ships only: raw data (noQC), 3rd iteration quality-
1508   controlled with no bias adjustment (NBC), 3rd iteration quality-controlled and bias-adjusted (BClocal), 3rd
1509   iteration quality-controlled and bias-adjusted for ship height only (BClocalHGT), 3rd iteration quality-controlled
1510   and bias-adjusted for instrument ventilation only (BClocalINST). b) Platform and alternative product
1511   comparison: 3rd iteration quality-controlled and bias-adjusted ship-only (ship), 3rd iteration quality-controlled
1512   and bias-adjusted for ships and moored buoys (all), NOCSv2.0 in-situ quality-controlled and bias-adjusted
1513   product based on ships only (NOCS-q), ERA-Interim reanalysis 2m fields using complete ocean coverage at the
1514   1° by 1° scale (ERA-Interim), ERA-Interim reanalyses 2m fields using complete ocean coverage at the 1° by 1°
1515   scale and masked to HadISDH.marine spatio-temporal coverage (ERA-Interim MASKED). Trends cover the
1516   common 1979 to 2015 period. 1979 to 2018 trends for ERA-Interim are 0.034 ± 0.098 and 0.029 ± 0.0098 for
1517   the full and masked versions respectively. c) Time of observation comparison for 3rd iteration quality-controlled
1518   and bias-adjusted ship-only: all times (both), daytime hours only (day), night time hours only (night). Linear
1519   trends were fitted using the median of pairwise slopes.

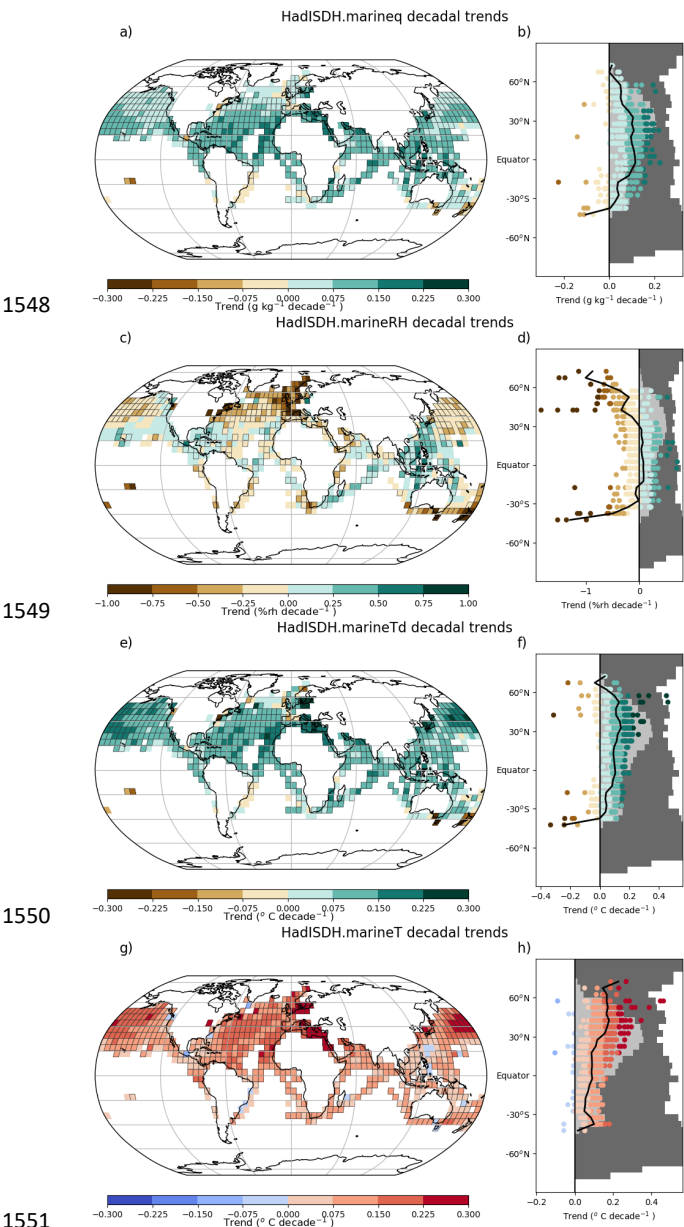
1520   **a)**

1521   **b)**

1522   **c)**
1523   Figure 9 Global (70 °S to 70 °N) annual average anomaly timeseries and decadal trends (+/- 90 % confidence
1524   interval) for relative humidity. See Figure 8 caption for details. Trends cover the common 1979 to 2018 period.
1525

1526   **a)**

1527   **b)**

1528   **c)**
1529   Figure 10 Global (70 °S to 70 °N) annual average anomaly timeseries and decadal trends (+/- 90 % confidence
1530   interval) for dew point temperature. See Figure 8 caption for details. Trends cover the common 1979 to 2018
1531   period.

1532 **a)**

1533 **b)**

1534 **c)**
1535 Figure 11 Global (70 °S to 70 °N) annual average anomaly timeseries and decadal trends (+/- 90 % confidence
1536 interval) for marine air temperature. See Figure 8 caption for details. Trends cover the common 1979 to 2018
1537 period.
1538

1539
1540 Figure 12 Global average timeseries of annual mean climate anomalies for all variables. The 2 sigma uncertainty
1541 ranges for total observation (blue), sampling (red) and coverage (gold) uncertainty contributions combined are
1542 shown. All series have been given a zero mean over the common 1981-2010 period. Linear trends were fitted
1543 using the median of pairwise slopes with the range representing the 90 % confidence interval in the trend.
1544
1545
1546
1547

1548

1549

1550

1551
1552 Figure 13 Linear decadal trends from 1973 to 2018 for a, b) specific humidity (g kg⁻¹), c, d) relative humidity
1553 (%rh), e, f) dew point temperature (° C) and g, h) air temperature (° C) for the 3ʳᵈ iteration quality-controlled
1554 and bias-adjusted ships only. Trends have been fitted using the median of pairwise slopes when there are at least
1555 70 % percent of months present over the trend period. Gridboxes with boundaries show significant trends in that
1556 the 90 % confidence interval around the trend magnitude is the same sign as the trend and does not encompass
1557 zero. The right-hand panels (b, d, f, h) show the distribution of gridbox trends by latitude with the mean shown
1558 as a solid black line. The dark grey shading shows the proportion of the globe at that latitude which is ocean.
1559 The light grey shading shows the proportion of the globe that contains observations.

1560
1561