Reviewer #1

Many thanks for your very helpful comments. We hope that we have made revisions in a satisfactory way. Please note that we did consider updating the HadISDH.marine dataset to include 2019. However, this would have involved substantial reprocessing of all of the figures in addition to bringing in ERA5 for comparison and uncertainty assessment instead of ERA-Interim because ERA-Interim does not continue to the end of 2019. This is not something that we felt we could achieve within the time frame. We envisage future papers where we compare HadISDH.marine with ERA5 in a comprehensive manner that would not be possible here.

Major issues:

1) The paper isn't sufficiently clear about whether data that is flagged for quality control issues like the buddy check or supersaturation are actually included in the final results or removed. On pages 12-13 some QC issues lead to "failures removed" but many issues lead to "failures flagged" rather than removal. It is later implied that the flagged values have been cleaned from the data (e.g. lines 718 and lines 966-968) but this doesn't seem to be explicitly written down as part of the process earlier in the paper and it is not clear that all flagged data is removed (e.g. the whole number flag). Please add a paragraph relatively early in the paper to explain in one place whether the flags and which ones are used to remove data (and at which iteration). Also some issues on page 13 are listed as "base qc" but the term "base qc" is never referred to again. In particular, state whether "raw (noQC)" includes the "base qc".

We agree that this was difficult to follow in the text. We have rewritten this section and added a new table (see below) that lists all of the quality control tests, whether they result in flags or removals and in which iterations. We have also added a final percentage data removal. We hope that this is now easier to follow.

Table 1. Description of quality control tests.

Test	**Description**	**1st and 2nd Iteration**	**3rd Iteration and Bias Adjusted**	**% of Observations Removed**
day / night	values likely to be affected by the solar heating of a ship where the sun was above the horizon an hour before the observation (based on the month, day, hour, latitude and longitude; Kent et al. (2013)) are flagged as 'day'	flagged	flagged	NA
climatology	T and T_d must be within a specified threshold of the nearest 1° by 1° pentad climatology	removed	removed	T = 2.39 and T_d = 5.14
supersaturation	T_d must not be greater than T (only T_d removed)	removed	removed	0.54
track	The distance and direction travelled by the ship must be plausible and consistent with the time between observations, normal ship speeds and observation locations before and after.	removed	removed	0.86
repeated value	A T or T_d value must not appear in more than 70 % of a ship track where there are at least 20 observations.	removed	removed	T = 0.04 and T_d = 0.06
repeated saturation	Saturation (T_d = T) must not persist for more than 48 hours within a ship track where there are at least 4 observations (only T_d removed).	removed	removed	0.54

buddy	T and T_d must be within a specified threshold of the average of nearest neighbours in space and time.	not applied	removed	T = 7.16 and T_d = 9.47
whole number	A T or T_d value must not appear as a whole number in more than 50 % of a ship track where there are at least 20 observations.	flagged	flagged	T = 11.73 and T_d = 8.20

2) A central feature of the dataset is that it involves three iterations. The iterations are mentioned throughout the paper but do not seem to be properly introduced (unless I've missed it). Please add a paragraph early on in the paper where you introduce the iterations, how they differ from each other, and why you use three iterations (rather than say 2 or 4).

This was also noted by the other reviewer. We have now swapped Figures 5 and 6 so that the flow chart can be introduced and discussed earlier – at the beginning of Section 3. We have also explained that we use three iterations to allow us to remove any artefacts from the ERA-Interim climatology by incrementally improving our own climatologies. Each iteration is computationally expensive and the 3rd iteration made only small changes compared to the 2nd iteration so we felt that three was sufficient. The new text is as below:

"The construction process, including the three iterations and all outputs, is visualised in Figure 5. Firstly, humidity variables are calculated. For the 1st iteration the hourly temperature and dew point temperature data are quality controlled (section 3.1) using an ERA-Interim based climatology. The data are then gridded, merged and a 1° by 1° pentad climatology produced for each variable (section 3.5). These 1st iteration climatologies are then used to quality control the original hourly data again; these data are then gridded, merged and a 2nd iteration climatology produced. The 2nd iteration climatology is then used to quality control the original hourly data for a third and final time. It is during this 3rd iteration that bias adjustments are applied and uncertainties estimated. The bias adjusted data and uncertainties are then gridded, merged and climatologies created. For future annual updates the 2nd iteration climatologies will be used to apply quality control. Having three iterations enables incremental improvements to the climatology used to quality control the data and therefore the skill of the quality control tests. It means that we can ensure that no artefacts remain from using ERA-Interim to quality control the data initially. Arguably more iterations could be done but each one is computationally expensive and the difference between the 2nd and 3rd iteration is already very small."

3) The paper seems to conclude that whole number rounding is not causing the pre 1982 positive bias and thus the negative trend in relative humidity, but I don't find this very convincing given that there is a large change in frequency of whole numbers in Td around 1980 (Fig S1b). Please address this issue in two ways: i) Calculate the trend in relative humidity for 1982 onwards to see if it is significant and include it in the paper. ii) Remove the whole-number flagged data and check if the trend in relative humidity remains negative.

We have done as you ask and agree that this was missing from the paper. The RH trends 1982 to 2018 are now shown in Figure 9 along with the full period trends. Although still negative these trends are now weaker (closer to 0) and generally not significant. Interestingly, ERA-Interim trends are only very slightly weaker and significant. We have added discussion on this in the text with some highlighted below:

Section 4

"Despite careful quality control and bias-adjustment the previously noted moist humidity bias pre-1982 is still apparent in the bias-adjusted (BA) data. The linear trend in relative humidity from 1982 to 2018 is -0.03 ± 0.13 %rh decade $^{-1}$, and therefore not significantly decreasing which is more consistent with expectation."

"Relative humidity is very sensitive to any differences in the data but even these differences are fairly small and do not change the overall conclusion of decreasing full-period trends and no significant trend over the 1982-2018 period."

"To explore whether the presence of whole numbers in the record has contributed to the pre-1982 bias we have processed a bias adjusted version with all whole number flagged data (Table 1) removed (BA_no_whole) which is shown against the noQC and BA versions in Fig. 9d. The resulting global average trend is largest in the BA_no_whole version, even over the 1982-2018 period, and the pre-1982 bias still clear. We conclude that the pre-1982 moist bias remains apparent in HadISDH.marine, and as yet not well understood, and quality control of the pre-1982 data is an area for more research in future versions."

Section 6

"The pre-1982 data have previously been noted as having a moist bias and our processing steps do not appear to have removed this feature. The trend excluding this earlier period (1982-2019) is no longer a significant decreasing trend and therefore more consistent with expectation. Removal of whole number flagged data appeared to exacerbate the pre-1982 bias and make the negative trends larger."

4) line 768-769: You don't mention correcting for serial correlation when calculating the uncertainty of the linear trends. Correcting for serial correlation could substantially increase the size of the 90th percentile confidence interval. Therefore, you should correct this estimate for serial correlation (or mention it if you are already doing so).

We have changed all trend fitting to OLS with AR(1) correction applied when fitting confidence intervals, following the Santer et al., (2008) paper which is now referenced. This has increased the confidence intervals but not changed the main conclusions.

5) Equation 3: Does the sqrt(9) result from the Gaussian distribution rather than assuming a uniform distribution (which would give sqrt(3))? Note 1 on page 14 of the cited BIPM document seems to suggest sqrt(9) would be correct for a 3 sigma range rather than a 1 sigma range as used here, so this could be an error.

Many thanks for pointing this out. We realise that we had assumed our estimated ranges covered 99.73% of possible values hence the sqrt(9) – using the methodology for a normal distribution. The range is based on 1sigma in the estimated height and so in fact covers 68.4% of possibilities so we realise that we should be using the 'two out of three' rule where u = (xHmax – xHmin)/2. This makes the height uncertainty larger. We have now redone the height uncertainty gridding, combined uncertainties and regional average uncertainties (Figure 12). This has made the height uncertainties larger and therefore expanded the total observation uncertainty and full uncertainty a little.

Minor issues:

line 49: "In these regions": Does this mean the region outside the northern mid latitudes or does it mean the northern mid latitudes?

We are referring to the northern mid-latitudes and so have changed this to 'In this region'.

lines 210-211: NOCS is not always lower in specific humidity over 1973-1981 - corre*ct?*

You are right that NOCSv2.0 is not always lower so we have removed this sentence.

line 225: It would be clearer to say "conversions between different units (e.g....) and between different variables" (currently it reads as a conversion between a unit and a variable which does not make sense)

We agree that this was not very clear and have changed as recommended.

line 270: Please clarify in the paper whether the absence of metadata from 2015 onwards is a temporary issue or something that is expected to persist.

We have added the following, based on the ICOADS website:

"It is likely that digitised metadata updates will be available periodically, depending on resource availability."

line 281: "pentad gridbox" is used without pentad being introduced. Please move the explanation for pentad from line 290 to here (i.e. that you mean pentad in time).

Done

Table S1: Introduce what Pmst is (not sure what mst stands for, Ps is used in the text). Also it is said in the table that e/es can be replaced by q/qs but these are clearly not equivalent. Clarify if you use e/es or q/qs.

We have changed this to Ps in Table S1 for consistency. This came about because we are using the climatological pentad mean surface pressure (from ERA-Interim nearest gridbox) but we agree that the notation is confusing. We use e/es to calculate RH and cannot recall why q/qs was listed in Table S1 and so have removed it.

line 356-357: Add a sentence to say how you determine if the track is 'plausible'

We have added the following to the new Table 1:

"The distance and direction travelled by the ship must be plausible and consistent with the time between observations, normal ship speeds and observation locations before and after."

line 481-482: I assume 'f' is being used here as a symbol for a generic function. Please instead explain (in words if necessary) what the function is.

We hope that the following is clearer:

"HOHest μ = 16 m + the linear trend in mean HOP/HOB/HOT height to the date of observation, σ = 4.6 m + the linear trend in standard deviation HOP/HOB/HOT height to the date of observation"

line 501-502: Is 'f' being used as a generic function? If so, writing 'a function of f(10/L)' doesn't make sense and should be 'a function of 10/L'.

The f()s have now been removed as we agree that they do not make sense.

line 508-509: Why does using T for SST mean that T is not adjusted?

When the SST is missing and T is used as a substitute there is no difference between the SST and T so the resulting adjustment to T will be zero.

lines 516-524: Multiple units are missing for temperatures and lengths in this section of the text (0.2, 50 etc. should all have units)

Now added.

line 532: 0.001 should have a unit

Done

line 538: Introduce that 'x' could be 'T, q, etc.'

Done

line 565: "and uncertainty"->"an uncertainty"

Done

line 587: Why is Nobs=10 the worst case scenario?

The climatology calculation requires there to be a minimum of 10 years of data present over the 30 year climatology period so Nobs=10 is the lowest number of observations possible. We have added that to the text.

line 616: Say how the gridding is done. Is it just a simple average of all data inside the grid box for those 3 hours?

It is just a simple mean. We have changed 'means from' to 'means of' in the text to try and make this clearer.

line 721-: I don't understand why you are showing results for the 2nd iteration rather than the 3rd iteration.

It is the 2nd iteration climatologies that are used to create anomalies and quality control the 3rd iteration data so we use that version to understand the difference between using the observation based climatology instead of ERA-Interim. We have added some text to explain this:

"To compare the use of ERA-Interim versus the observation based climatology to calculate anomalies and quality control the data we show difference maps of the 2nd iteration minus ERA-Interim pentad 1° by 1° grid climatologies and climatological standard deviations in Figs. S9 to S14 for a selection of pentads and variables."

line 772-773: Explain the abbreviations noQC, NBC, BClocal. I can guess the first two. I don’t know why BClocal is "local".

We have now changed these in the text and figures to noQC, noBA and BA and hopefully the amended text below makes this easier to understand:

“For all variables, there are only small differences in the global average timeseries between the various processing steps – from the raw data (noQC) to the 3rd iteration quality-controlled (noBA [no bias adjustment]) and then the bias-adjusted data (BA).”

line 783: A little more care is needed to discuss and cite expectations from theory and models. The first cited paper Byrne and O’Gorman 2013 indeed does shows results for weak positive changes in marine relative humidity. However, it doesn’t seem to give a theory for changes in marine relative humidity; it instead cites for theory the papers by Held and Soden 2010 and Schneider et al 2013 which could be cited here. The cited Byrne and O’Gorman 2018 paper does seem relevant in that it shows that the land changes in temperature and humidity are broadly consistent with simple theory and no changes in marine relative humidity.

We agree this was a little weak and in fact we had missed the Byrne and O’Gorman 2016 reference which shows modelled future changes in ocean relative humidity explicitly. This has now been added. We have also added the Held and Soden (2006) and Schneider et al., (2013) references. The text around this is now as follows:

Section 4.2

“This differs from theoretical expectation where changes in relative humidity over ocean are strongly energetically constrained to be small, of the order of 1% K^{-1} or less, and generally positive (Held and Soden, 2006; Schneider et al., 2010). Model-based expectations also suggest small positive changes (Byrne and O’Gorman, 2013, 2016, 2018).”

Section 6

“Theoretical and model-based analysis of changes in relative humidity over ocean under a warming climate suggest negligible or small positive changes (Held and Soden, 2006; Schneider et al., 2010; Byrne and O’Gorman, 2013, 2016, 2018).”

line 801-804: BClocal etc. include the quality-control step and the bias adjustment so they should be compared to the quality-controlled data but not the raw data when seeking to determine the effect of the bias adjustment.

This is a good point and we now only compare the BA (was BClocal etc) versions to the noBA (was NBC) versions.

line 818-819: I support the authors wise choice to focus on the ship data for the final product.

Thank you.

line 845: "compares well" Be more specific here about what aspect compares well. For example, the trends are quite different in magnitude.

We had already mentioned that this was in reference to the trend direction but agree that it was a little vague so have made it more explicit:

"In terms of linear trend direction, HadISDH.marine compares well with other monitoring estimates from NOCSv2.0 and ERA-Interim and to other reanalyses and older products (Fig. 1)."

line 868: The decreasing trend in relative humidity over ocean is said to be consistent with the decreasing trend over land. I don't see why this is "consistent" rather than just "similar". The papers cited earlier on models and theory suggest that land relative humidity can decrease even if marine relative humidity stays constant or increases slightly. Also, it would be helpful to give a value for the trend over land to compare with the trend over ocean to see how similar they are in magnitude.

We agree that this was misleading and have amended this to 'are similar to' and added the land trends in the text.

"; land linear trends are 0.03 %rh more negative at -0.12 (-027 to -0.03) %rh 10 yr^{-1} over the same 1973 to 2018 period"

fig 1: Why does JRA have values before 1980 for land but not marine relative humidity? How is missing data dealt with in this figure?

This was an error in plotting and has now been corrected.

fig 5: Might be less confusing if you use the same y axis range for both panels

Done.

fig 6: The blue path goes through the Quality Control box but then it later is labelled "no QC" which seems to be contradictory. Also "noQC" and "bias adjusted" are labelled for blue and yellow but not red.

We agree that this was confusing and have spotted a couple of errors in this figure which we have now corrected. We have also added text in the figure caption to help explain it. The 'no QC' boxes are now coloured gray to identify them as not being part of the 1st iteration.

fig 7: Annual mean climatologies are deemed acceptable if 9 months of the year are present. Couldn't this lead to a very large bias if for example November and December were missing given the large seasonal cycle?

We agree that this is true. These annual climatologies are produced just for this figure and not made available as part of the HadISDH.marine product. We chose 9 months to balance data coverage over data accuracy. We would expect users to want monthly or seasonal climatologies and compute their own seasonal and annual climatologies if required.

fig 8: It is probably less confusing to keep the same vertical order for the legend and the trends (currently they seem to be reversed).

Agreed, and now changed accordingly.

Reveiwer #2

Many thanks for your very helpful comments. We hope that we have made revisions in a satisfactory way. Please note that we did consider updating the HadISDH.marine dataset to include 2019. However, this would have involved substantial reprocessing of all of the figures in addition to bringing in ERA5 for comparison plots and uncertainty estimation instead of ERA-Interim because ERA-Interim does not continue to the end of 2019. This is not something that we felt we could achieve within the timeframe. We envisage future papers where we compare HadISDH.marine with ERA5 in a comprehensive manner that would not be possible here.

"Specific comments"

P1, L13: Please provide definition for “SDH” of “HadISDH”.

This is a little tricky to do because the name HadISDH was chosen originally for the land dataset many years ago and so is essentially a slightly random legacy. It utilised NOAA NCEI’s Integrated Surface Dataset, hence the ISD. The H stands for humidity. The marine data uses ICOADS rather than ISD but I felt that it was important to keep the HadISDH name because the land and marine products are related and intended to be used together. So, I’m not quite sure how to logically explain all of that in the paper. I could say: It is a Met Office ***Had****ley Centre led* ***I****ntegrated* ***S****urface* ***D****ataset of* ***H****umidity as this does make sense but doesn’t explicitly refer to the NCEI ISD dataset. I have added that sentence to the Introduction (4th paragraph) rather than the Abstract.*

P2, L43: Please provide citation for GCOS Essential Climate Variables (ECVs).

We have added the following: (Bojinski et al., 2014; https://gcos.wmo.int/en/essential-climate-variables)

P2, L54: Please provide definition of “CRUH” of “HadICRUH”.

Done – Met Office ***Had****ley Centre and* ***C****limatic* ***R****esearch* ***U****nit* ***H****umidity dataset.*

P9, L265: “Processing the hourly data into a gridded product”... Is this title appropriate for this section? (A possible alternative might be “Construction of the gridded dataset and uncertainty estimates”, for example.)

That is a much better title, thank you.

P11, L302: It turned out later (section 4.1) that buoy data were eventually excluded from the current version. I would suggest that this treatment (exclusion of buoy data) should be mentioned in the early part of this section. For example, the overall strategy might be summarized first using Figure 6.

We have made it much clearer that the buoys are used for the climatologies and the final version is ship only with the following statement:

“We include moored buoys to produce climatologies because spatial coverage is of high importance. Our final version recommended to users is a ship-only (SHIP) product but we have produced a combined (ALL) product for comparison.”

We have also now added a paragraph explaining the overall process – see below.

P13, L362: “3rd iteration” is referred without prior explanation. I think it would be helpful for the reader if the idea of the entire processing is presented first using Figure 6.

We agree that this is not clearly explained. We have swapped Figures 5 and 6 around and added a paragraph describing the flow of the dataset build in the beginning of section 3:

“The construction process, including the three iterations and all outputs, is visualised in Figure 5. Firstly, humidity variables are calculated. For the 1^{st} iteration the hourly temperature and dew point temperature data are quality controlled (section 3.1) using an ERA-Interim based climatology. The data are then gridded, merged and a 1° by 1° pentad climatology produced for each variable (section 3.5). These 1^{st} iteration climatologies are then used to quality control the original hourly data again; these data are then gridded, merged and a 2^{nd} iteration climatology produced. The 2^{nd} iteration climatology is then used to quality control the original hourly data for a third and final time. It is during this 3^{rd} iteration that bias adjustments are applied and uncertainties estimated. The bias adjusted data and uncertainties are then gridded, merged and climatologies created. For future annual updates the 2^{nd} iteration climatologies will be used to apply quality control. Having three iterations enables incremental improvements to the climatology used to quality control the data and therefore the skill of the quality control tests. It means that we can ensure that no artefacts remain from using ERA-Interim to quality control the data initially. Arguably more iterations could be done but each one is computationally expensive and the difference between the 2^{nd} and 3^{rd} iteration is already very small.”

P14, L398 (Fig. S7): Looking at Fig. S7 and its inset legend, “repeated saturation check” (pink, solid line) seems to be making only minor contributions.

Thank you for pointing this out. We realise we had mistaken the repeated saturation check for the track check which is the pink dotted line. We have corrected this in the text.

While looking at the figures in more detail we noticed that the has been an increase in failures for repeated saturation and supersaturation towards the end of the record which is also when many more electronic and capacitance sensors are in use instead of psychrometers. We have now pointed this out in the text too:

“There is an increase in removals from repeated saturation and supersaturation events over time, particularly the late 2000s. This may be related to the decrease in psychrometer deployment over time and increase in electric and capacitance sensors as shown in Fig. 4. The latter have increased significantly since the mid-2000s.”

P16, L456-458: It would be helpful if formulae are used to describe the procedures explained here.

We would prefer not to add formulae here as we think that this is covered by Table 1. Table 1 hadn’t yet been referenced at this point but we have now pointed the reader to it at the beginning of section 3.3. We have also tried to improve the text of the paragraph you refer to so that it is easier to understand:

“To carry these adjustments and uncertainties to all other humidity variables we start with q and then propagate the adjusted quantity and adjusted quantity plus uncertainty using the equations in Table S1. Using the original T (which does not need to be adjusted for poor ventilation) and ERA-Interim climatological surface pressure, e can be calculated from q. T_d and RH can be calculated from e and T. From these, the T_w and DPD can be calculated. The uncertainty is then obtained by subtracting the adjusted quantity from the adjusted quantity plus uncertainty for each variable.”

P22, L650 (Eq.7) and L655 (Eq.8): Where and how was Ugb woven into the uncertainty estimate?

We think that the text was not clear that Ugb is just a generic variable name when in fact the equation given is used for each of the five uncertainty sources (Ui, Um, Uw, Uc and Uh). We have modified the text to make this clearer. It was particularly misleading that we stated that all five quantities are combined to produce the total observation uncertainty for the gridbox before equations 7 and 8 which deal with the individual uncertainty sources. This sentence has now been removed. It is largely repeated later anyway.

P24, L700: Buoy products are excluded from the current version. I think this should be described earlier, for example, in section 3. Or the overall strategy along with the procedures (visualized in Fig, 6) could be presented earlier.

We think that we have now addressed this as described in our responses to your revisions listed above.

P27, L776-777, L798-799: How will the decadal trend for relative humidity look like when the pre-1982 period is excluded from the analysis?

We have now added trends for the 1982-2018 period to the annual time series comparison plots in Figure 9 and added some text in several places discussing this.

Section 4
"Despite careful quality control and bias-adjustment the previously noted moist humidity bias pre-1982 is still apparent in the bias-adjusted (BA) data. The linear trend in relative humidity from 1982 to 2018 is -0.03 ± 0.13 %rh decade $^{-1}$*, and therefore not significantly decreasing which is more consistent with expectation."*

"Relative humidity is very sensitive to any differences in the data but even these differences are fairly small and do not change the overall conclusion of decreasing full-period trends and no significant trend over the 1982-2018 period."

Section 6
"The pre-1982 data have previously been noted as having a moist bias and our processing steps do not appear to have removed this feature. The trend excluding this earlier period (1982-2019) is no longer a significant decreasing trend and therefore more consistent with expectation."

P30, L890: It would be worth briefly mentioning again here what comprise the total observation uncertainty.

We have done this by specifying the five observation uncertainty components in the paragraph before (old line 876).

"This includes the total observation uncertainty, which covers uncertainty components for instrument adjustment, height adjustment, measurement, climatology and whole number uncertainty (Table 1)."

"Technical corrections"

P10, L289: Remove right parenthesis ")" after "temperature".

Done.

P14, L398: Put periods "." after "Fig. S7".

Done.

P22, L644: "has" should perhaps read "as".

Yes – thanks. Done.

P33, L971: Put periods "." after "averaging".

Done.

P35, L1037: "over estimate" should read "overestimate".

Additional changes:

In response to reviewer 1 several other changes have been made to the paper. I hope you are able to see our response to reviewer 1 but we have summarised the major changes below.

We have changed all trend fits from median of pairwise slopes to ordinary least squares. The confidence intervals shown are now 90% confidence intervals corrected for AR(1) correlation following the Santer et al., 2008 paper.

We have now created a gridded dataset where the whole number flagged data are removed to check the trends. This data is shown in Figure 9 as an additional panel comparing the raw data, the bias adjusted data and the bias adjusted data where all whole number flagged data are removed. This does not remove the pre-1982 issue and in fact appears to exacerbate it. There are several additions to the text to note this.

Development of the HadISDH marine humidity climate monitoring dataset

Kate M. Willett[1], Robert J. H. Dunn[1], John J. Kennedy[1] and David I. Berry[2]

[1]Met Office Hadley Centre, Exeter, UK

[2]National Oceanography Centre, Southampton, UK

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

Abstract

Atmospheric humidity plays an important role in climate analyses. Here we describe the production and key characteristics of a new quasi-global marine humidity product intended for climate monitoring, HadISDH.marine. It is an in-situ based multi-variable marine humidity product, gridded monthly at a 5° by 5° spatial resolution from January 1973 to December 2018 with annual updates planned. Currently, only reanalyses provide up to date estimates of marine surface humidity but there are concerns over their long-term stability. As a result, this new product makes a valuable addition to the climate record and will help address some of the uncertainties around recent changes (e.g. contrasting land and sea trends, relative humidity drying). Efforts have been made to quality control the data, ensure spatial and temporal homogeneity as far as possible, adjust for known biases in non-aspirated instruments and ship heights, and also estimate uncertainty in the data. Uncertainty estimates for whole-number reporting and for other measurement errors have not been quantified before for marine humidity. This is a companion product to HadISDH.land, which, when combined will provide methodologically consistent land and marine estimates of surface humidity.

The spatial coverage of HadISDH.marine is good over the Northern Hemisphere outside of the high latitudes but poor over the Southern Hemisphere, especially south of 20° S. The trends and variability shown are in line with overall signals of increasing moisture and warmth over oceans from theoretical expectations and other products. Uncertainty in the global average is larger over periods where digital ship metadata are fewer or unavailable but not large enough to cast doubt over trends in specific humidity or air temperature. Hence, we conclude that HadISDH.marine is a useful contribution to our understanding of climate change. However, we note that our

ability to monitor surface humidity with any degree of confidence depends on the continued availability of ship data and provision of digitised metadata.

HadISDH.marine data, derived diagnostics and plots are available at www.metoffice.gov.uk/hadobs/hadisdh/indexMARINE.html (Willett et al., 20202019).

1 Introduction

Water vapour plays a key role as a greenhouse gas, in the dynamical development of weather systems, and impacts society through precipitation and heat stress. Over land, all these aspects are important and recent changes have been assessed by Willett et al. (2014). Over the oceans, a major source of moisture over land, a similar analysis is essential to enhance our understanding of the observed changes generally and as a basis for worldwide evaluation of climate models. In recognition of its importance, the surface atmospheric humidity has been recognised as one of the Global Climate Observing System (GCOS) Essential Climate Variables (ECVs) (Bojinski et al., 2014; https://gcos.wmo.int/en/essential-climate-variables).

Observational sources of humidity over the ocean are limited. The NOCSv2.0 (Berry and Kent, 2011) is the only recently updated (January 1971 to December 2015) marine surface humidity monitoring product based on in-situ observations, but it only includes specific humidity (*q*). Satellite based humidity products exist (e.g. HOAPS, Fennig et al., 2012) but these rely on the in-situ observations for calibration. Whilst quasi-global, the uncertainties in the NOCv2.0 product are large outside the northern, mid--latitudes. In thisese regionss the NOCSv2.0 product shows a reasonably steadily rising trend over the period of record, similar to that seen over land but with slightly different year-to-year variability. Most notably, 2010, a peak year over land in specific humidity, does not stand out over ocean. Figure 1 and Willett et al. (2019) show global land and ocean specific humidity and relative humidity (RH) series from available in-situ and reanalyses products. Older, static products for the ocean (HadCRUH – Met Office **Had**ley Centre and **C**limatic **R**esearch **U**nit **H**umidity dataset: Willett et al., 2008; Dai: Dai 2006) show increasing specific humidity to 2003 with similar variability to NOCSv2.0, and near-constant relative humidity. Both HadCRUH and Dai show a positive relative humidity bias pre-1982 and slightly higher specific humidity over 1978-1984 compared to NOCSv2.0. There is broad similarity between the reanalysis products and the in-situ products but with notable differences for specific

Commented [WK1]: TO be changed to the CEDA archive link once I have it – prior to publication.

humidity in the scale of the 1998 peak and the overall trend magnitude. Differences are to be expected given that the reanalyses are spatially complete in coverage, albeit derived only from their underlying dynamical models over data sparse regions. The reanalyses exhibit near-constant to decreasing relative humidity over oceans but with poorer agreement between both the reanalyses themselves and compared to the in-situ products over land. This is to be expected given the larger sources of bias and error over ocean (Sect. 2) and sparse data coverage. Importantly, land and marine specific humidity appear broadly similar whereas for relative humidity, the distinct drying since 2000 over land is not apparent over ocean in reanalyses and the previously available in-situ products finish too early to be informative. Note that the HadISDH.marine described herein is shown here for comparison and will be discussed below.

A positive bias in global marine average relative humidity pre-1982 is apparent in Dai and HadCRUH, and has previously been attributed to high frequencies of whole numbers in the dew point temperature observations prior to January 1982 (Willett et al., 2008). This is less clear in the global average specific humidity timeseries. ICOADS (International Comprehensive Ocean-Atmosphere Dataset) documentation (http://icoads.noaa.gov/corrections.html) notes issues with the pre-1982 data especially mixed-precision observations, where the air temperature has been recorded to decimal precision but the dew point temperature is only available as a whole number. Such reporting was in accordance with the WMO Ship Code before 1982. The documentation notes a truncation error in the dew point depression which would lead to a positive bias in relative humidity. Alternatively, Berry (2009) show that patterns in the North Atlantic Oscillation coincide with this time period and could have played a role. The NOCSv2.0 product is based on reported wet bulb temperature rather than dew point temperature, where decimal precision is usually present. Hence, the NOCSv2.0 product is expected to be unaffected by these rounding issues. Our analysis shows that changes to the code in January 1982 did not eliminate whole number reporting and high frequencies of whole numbers can be found throughout the record in both air temperature and dew point temperature (Sect. 2.4 and Sect. 3.4).

Clearly, there is a need for more and up to date in-situ monitoring of humidity over ocean, especially for RH. The structural uncertainty in estimates can only be explored if there are multiple available estimates so a new product that explores different methodological choices, and extends the record, is complementary to the existing NOCSv2.0 product and reanalyses estimates. Here we report the development of a multi-variable marine humidity analysis HadISDH.marine.1.0.0.2018f (Willett et al., 2020~~Met Office Hadley Centre; National~~

Oceanography Centre, 2019 FINALISED AFTER REVIEW). HadISDH.marine is a Met Office **Had**ley Centre led **I**ntegrated **S**urface **D**ataset of **H**umidity,This formings a companion product to the HadISDH.land monitoring product, and enabling the production of a blended global land and ocean product. We use existing methods where possible from the systems used for building the long running HadSST dataset (Kennedy et al., 2011a, 2011b, 2019), and also use some of the bias adjustment methods employed for NOCSv2.0 (Berry and Kent 2011). We have explored the data to design new humidity specific processes where appropriate, particularly in terms of quality control and gridding.

HadISDH.marine is a climate-quality 5° by 5° gridded monthly mean product from 1973 to present (December 2018 at time of writing) with annual updates envisaged. Fields will be presented for surface (~10 m) specific humidity, relative humidity, vapour pressure, dew point temperature, wet bulb temperature and dew point depression. Air temperature will also be made available as a by-product but less attention has been given to addressing temperature specific biases. The product is intended for investigating long-term changes over large scales and so efforts have been made to quality control the data, ensure spatial and temporal homogeneity, adjust for known biases and also estimate remaining uncertainty in the data. In particular, we estimate uncertainties from whole-number reporting and other measurement errors that have not been quantified before for marine humidity.

Section 2 discusses known issues with marine humidity data. Section 3 describes the source data and all processing steps. Section 4 presents the gridded product and explores the different methodological choices and comparison with NOCSv2.0 specific humidity and ERA-Interim marine humidity. This section also includes a first look at the blended land and marine HadISDH product for each variable. Section 5 covers data availability and Section 6 concludes with a discussion of the strengths and weaknesses of the product.

2 Known issues affecting the marine humidity data

2.1 Daytime solar-biases

Marine air temperature measurements on board ships during the daytime are known to be affected by the heating of the ship or platform by the sun. This results in a positive bias during daylight and early night time hours. The bias varies with sunlight strength/cloudiness (and thus also latitude), relative wind speed, size and material of the ship. This solar heating bias affects both the wet bulb and dry bulb temperature measurements but, as noted by Kent and Taylor (1996), the ships do not act as a source of humidity or change the humidity content of the air. As a result, biases in the specific humidity and dew point temperature due to the solar heating errors will be negligible. However, care needs to be taken with relative humidity because estimates of the saturation vapour pressure from the uncorrected dry bulb air temperature will be too high, leading to an underestimate in relative humidity. Ideally, relative humidity should be estimated using the corrected dry-bulb temperature to calculate the saturation vapour pressure and uncorrected wet and dry bulb temperature or dew point temperature to calculate the vapour pressure.

Previously, efforts have been made to bias-adjust the air temperature observations for solar heating by modelling the extra heating over the superstructure of the ship, taking account of the relative wind speed, cloudiness, time of day, time of year and latitude (Kent et al, 1993; Berry et al., 2004; Berry and Kent, 2011). These adjustments are complex and so we have decided not to attempt to implement them for our first version of a marine humidity product given the wide variety of other issues we have accounted for. We have, however, produced daytime, night time and combined products to investigate differences that may be caused by the solar heating bias. Later versions of HadISDH.marine that apply bias corrections for solar heating may reduce the amount of daytime data removed.

2.2 Un-aspirated psychrometer bias

Humidity measurements can be made in a variety of ways. Instruments can be housed in a screen with ventilation slats, with or without additional artificial aspiration, or handheld in a sling or whirling psychrometer. There is information on instrument ventilation provided up to 2014. Approximately 30 % of ship observations have information in 1973, peaking at ~75 % by the mid-1990s, as summarised in Fig. 2. Initially, slings were more common for the hygrometer and thermometer, but by 1982 a screen was more common. There is a tendency for the screened instruments, in the absence of artificial aspiration, to give a wet bulb reading that is higher relative to the slings/whirling instruments where airflow is ensured by the whirling motion. Bias adjustments have been applied to un-aspirated humidity observations by Berry and Kent (2011), building on previous bias adjustments of Josey et al. (1999) and Kent et al. (1993). They have also estimated the uncertainty in the bias adjustments. We implement a modified version of their method of bias adjustment for the un-aspirated observation types (Sect. 3.3.1) and uncertainty estimation. Uncertainties from instrument bias adjustments will have some spatial and temporal correlation structure as the ships move around (Kennedy et al., 2011a).

2.3 Ship height inhomogeneity

Over time there has been a general trend for ship heights to increase. Kent et al. (2007; 2013) quantified the increase from an average of ~ 16m in 1973 to ~24m by the end of 2006. Instrument height information is available for some ships between the period of 1973 and 2014, providing heights for the barometer (HOB), thermometer (HOT), anemometer (HOA) and visual observing platform (HOP). Figure 3 shows the availability of height information and the mean and standard deviation of heights per year in each category for the ship observations selected here. Similar to the ventilation metadata, height information availability is low in 1973, peaking mid-1990s to 2000 and then declining slightly. Prior to 1994 only the platform height was available from WMO Publication 47. This was replaced in 1994 by the barometer height and augmented with the thermometer and visual observing heights from 2002 onwards (Kent et al., 2007). Anemometer heights have been available from WMO 47 since 1970. All four types of heights increase over time. We conclude that the mean height based on HOP/HOB/HOT increases from 17 m in 1973 to 23 m by 2014, which differs slightly to that in Kent et al., (2007). If uncorrected, this likely leads to a small artificial decreasing trend in air temperature and specific humidity, as, in general, these variables decrease with height away from the surface. The effect on relative humidity is less clear and depends on the relative effects on air temperature and specific humidity.

Prior studies (e.g. Berry and Kent, 2011; Berry 2009; Josey et al., 1999; Rayner et al., 2003; Kent et al., 2013) have applied height adjustments to the air temperature, specific humidity and wind speed measurements to adjust the measurements to a common reference height and minimise the impact of the changing observing heights on the climate record. These have been based on boundary layer theory and the bulk formulae, using the parameterisations of Smith (1980, 1988). In the absence of high-frequency observations of meteorological parameters for each observation location, allowing direct estimation of the surface fluxes, parameterisations have to be made and an iterative approach is necessary to estimate a height adjustment (Sect. 3.3.2). We have followed these previous approaches and estimated height adjustments for all observations and variables of interest. Where observing heights are unavailable we have made new estimates (Sect. 3.3.2). We have also provided an estimate of uncertainty on these height adjustments, which are larger where we have also estimated the height of the observation. The uncertainties from height adjustments will have some spatial and temporal correlation structure.

2.4 Whole-number reporting biases

Recording and reporting formats and practices have changed many times over the 20th century, affecting the climate record. Some formats required the wet bulb temperature to be reported, others the dew point temperature and some allowed either or both (https://www.wmo.int/pages/prog/amp/mmop/documents/publications-history/history/SHIP.html). Some earlier formats restricted space to reporting temperature to whole numbers only and this practice has continued with some ships continuing to report the dew point (or wet bulb) temperature and sometimes even the dry bulb temperature to whole numbers. A practice of truncation of the dew point depression has been noted for the pre-1982 data (http://icoads.noaa.gov/corrections.html) which would result in spuriously high humidity (both in relative and actual terms). It is clear from the ICOADS3.0.0/3.0.1 data that there has been a practice of reporting values to whole numbers rather than decimal places, both for air temperature and dew point temperature. Rounding dew point temperature and air temperature could result in a +/- 0.5° C error individually or a just less than +/- 1° C error in dew point depression for a worst-case scenario combination.

Whole-number reporting is an issue throughout the record for both variables – a breakdown of air and dew point temperature by decimal place over time is shown in Fig. S1. Air temperature also shows a disproportionate frequency of half degrees (5s). The percentage~~revalence~~ of whole numbers (0s) declines over time, dramatically in the mid- to late 1990s for air temperature and from 2008 for both air and dew point temperature. This decline in the 1990s, and in part also the general decline, appears to be linked to an increase in numbers of moored buoys (see Fig. 5), a similar analysis without the moored buoys (not shown) shows greater consistency over time. The dew point temperature has two distinct peaks in whole number frequency in the 1970s and mid-1990 to early 2010s. The latter peak is more pronounced when moored buoys are not included. The early peak is somewhat consistent with the restriction in transmission space prior to January 1982. This was previously thought to have been a possible cause of higher relative humidity over the period 1973-1981 compared to the rest of the record in the HadCRUH marine relative humidity product (Willett et al., 2008). The pre-1982 moist bias was also apparent in the global marine relative humidity product of Dai (2006), which like HadCRUH used dew point temperatures. The NOCSv2.0 product preferentially utilises the wet bulb temperatures from ICOADS which are not affected by whole number reporting to the same extent. This could be part of the reason why NOCSv2.0 has a lower estimate of specific humidity anomalies over the 1973-1981 period than HadCRUH or Dai, which use the dew point temperatures (Fig. 1).

Rounding of temperature alone should not affect the mean dew point temperature, specific humidity or vapour pressure. However, as with the solar bias issue, it is sensitive to at what point the reported dew point temperature was derived from the measured wet bulb temperature or relative humidity. Most likely, this would be done prior to any rounding or truncating for reporting but during later conversion of various sources into digital archives, or corrections, the dew point temperature may have been reconstructed (https://icoads.noaa.gov/e-doc/other/dupelim_1980). The effect of rounding on a monthly mean gridbox average should be small as these errors are random and should reduce with averaging. However, there is a risk of removing very high humidity observations when a rounded dew point temperature then exceeds a non-rounded air temperature. Such values are removed by our supersaturation check (Sect. 3.2). We do not feel able to correct for this issue but instead include an uncertainty estimate for it. Overly frequent whole numbers are identified both during quality control track analysis and deck analysis. This will be discussed in more detail in Sect. 3.4. Clearly, there are various issues that can arise linked to the precision of measured and reported data in addition to conversion between different units (e.g., Fahrenheit, Celsius and Kelvin, Fig. S1) and between different variables.

2.5 Measurement errors

All observations are subject to some level of measurement error and, outside of precision laboratory experiments, the errors can be significant. The BIPM Guide to the Expression of Uncertainty in Measurement (BIPM, 2008) describes two categories of measurement uncertainty evaluation. A Type A evaluation estimates the uncertainty from repeated observations. A Type B evaluation of the uncertainty is based on prior knowledge of the instrument and observing conditions. Within this study we use a Type B evaluation, adjusting for systematic errors and inhomogeneities due to inadequate ventilation and changing observing heights (screen and height adjustments) and estimate the residual uncertainty For the random components, we make the conservative assumption that all measurements were taken using a psychrometer (wet bulb and dry bulb thermometers), which allows us to follow the HadISDH.land methodology of Willett et al. (2013, 2014) as described in Sect. 3.4.. An assessment of the frequency of hygrometer types (TOH) within our selected ICOADS3.0.0/3.0.1 data shows this to be a fair assumption as the vast majority of ships (where metadata is available: ~30 % increasing to ~70 % 1973 to 1995 then decreasing to 60 % by 2014) are listed as being from a psychrometer (Fig. 4). Electric sensors are becoming more common and made up ~30 % of observations by 2014 (the end of the metadata information). There are no instrument type metadata for ocean platforms or moored buoys. As it is likely that most buoy observations are made using RH sensors, we plan to develop an RH sensor specific measurement uncertainty in future versions.

2.6 Other sources of error

There are other issues specific to humidity measurements that may be further sources of error. Hygrometers that require a wetted wick (i.e., psychrometers), and thus a source of water, are vulnerable to the wick drying out or contamination, especially by salt in the marine environment. The wick drying results in erroneous relative humidity readings of 100 %rh where the wet bulb essentially behaves identically to the dry bulb thermometer. There can also be issues when the air temperature is close to freezing depending on whether the wet bulb has become an ice bulb or not and whether wet bulb or ice bulb calculations are used in any conversions. Humidity observing in low temperature can be generally problematic. For radiosondes, there has previously been a practice of recording a set low value when the humidity observation falls below a certain value (Wade 1994, Elliott et al. 1998). It is debateable how likely such low humidity values are over oceans and this practice has not been documented for ship observations. However, the set value issue is something to look out for. Wet bulb thermometers (and other instruments) can experience some hysteresis at high humidity where it takes some time to return to a lower reading. The wet bulb also requires adequate ventilation which has been discussed above.

These can be accounted for to a large extent through quality control but some error will inevitably remain. We can increase our confidence in the data by comparison with other available products and general expectation from theory.

3 ~~Processing the hourly data into a gridded product~~Construction of the gridded dataset and uncertainty estimates

ICOADS Release 3.0 (Freeman et al., 2017) forms the base dataset for the HadISDH.marine humidity products. From January 1973 to December 2014 we use ICOADS.3.0.0 from http://rda.ucar.edu/datasets/ds540.0/. These data include a unique identifier (UID) for each observation, a station identifier/ship callsign (ID), metadata on instrument type, exposure and height in many cases. From January 2015 onwards we use ICOADS.3.0.1 from the same source. These data include an ID and UID but no instrument metadata. It is likely that digitised metadata updates will be available periodically, depending on resource availability. Each observation is associated with a deck number. These are identifiers for ICOADS national and trans-national sub-sets of data relating to source e.g., deck 926 is the International Maritime Meteorological (IMM) data (https://icoads.noaa.gov/translation.html). We utilise the reported air temperature (T) and reported dew point temperature (T_d) as the source for our humidity products. Sea surface temperature (SST) and wind speed (u) are used for estimating height adjustments.

We calculate the specific humidity (q), relative humidity (RH), vapour pressure (e), wet bulb temperature (T_w, not the thermodynamic wet bulb but a close approximation to it) and dew point depression (DPD) for each point observation. All humidity variables are derived from reported air and dew point temperature and ERA-Interim climatological (from the nearest 1° by 1° 5 day mean [pentad] gridbox) surface pressure P_s, using the set of equations from Willett et al., (2014) which can be found in Table S1. This provides consistency with HadISDH.land for later merging. F~~Note that f~~or consistency we ~~also~~ use a fixed psychrometric coefficient that is identical for all observations~~,~~ when estimating the approximate thermodynamic wet bulb temperature rather the observed value which depends on the type of psychrometer used. ~~minimising the impact of changing instrument types (e.g. whirling sling / ventilated measurement vs screen) on the wet bulb temperature record.~~ This is also consistent with what is done for HadISDH.land.

Additionally, we use ERA-Interim (Dee et al., 2011) reanalysis data to provide initial marine climatologies and climatological standard deviations for all variables to complete a 1st ~~first~~ iteration climatological outlier test. We extract 1° by 1° gridded 6 hourly 2 m air and dew point temperature~~)~~ and surface pressure to create 6 hourly humidity variables and then pentad ~~(5 day mean)~~ climatologies and standard deviations over the 1981-2010 period. Note that 3~~several~~ iterations are passed before finalising the product. Only the 1st ~~initial~~ iteration uses ERA-Interim climatologies, later iterations use climatologies built from the previous iteration's quality-controlled observations (Sects. 3.2, 3.5, 4.1).

The construction process, including the three iterations and all outputs, is visualised in Figure 5. Firstly, humidity variables are calculated. For the 1st iteration the hourly temperature and dew point temperature data are quality controlled (section 3.1) using an ERA-Interim based climatology. The data are then gridded, merged and a 1° by 1° pentad climatology produced for each variable (section 3.5). These 1st iteration climatologies are then used to quality control the original hourly data again; these data are then gridded, merged and a 2nd iteration climatology produced. The 2nd iteration climatology is then used to quality control the original hourly data for a third and final time. It is during this 3rd iteration that bias adjustments are applied and uncertainties estimated. The bias adjusted data and uncertainties are then gridded, merged and climatologies created. For future annual updates the 2nd iteration climatologies will be used to apply quality control. Having three iterations enables incremental improvements to the climatology used to quality control the data and therefore the skill of the quality control tests. It means that we can ensure that no artefacts remain from using ERA-Interim to quality control the data initially. Arguably more iterations could be done but each one is computationally expensive and the difference between the 2nd and 3rd iteration is already very small.

3.1 Data selection

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

- there must be a non-missing T and T_d value;
- the platform type (PT) must be in one of the following categories: a ship (a US Navy or unknown vessel, a merchant ship or foreign military ship, an ocean station vessel off station /at an unknown location, an ocean station vessel on station, a lightship, an unspecified ship - PT = 0, 1, 2, 3, 4, 5); or a stationary buoy (moored or ice buoy - PT = 6, 8);
- the observation must have a climatology and standard deviation available for its closest 1° by 1° pentad;

- the observation must pass the gross error checks: calculated RH must be between 0 and 150 %rh (supersaturated values are flagged during quality control); both T and T_d must be between -80 and 65 °C; and calculated q must be greater than 0.0 g kg^{-1};
- position check (failures removed): latitudes must be between -90° and 90° and longitudes must be between -180° and 360° (later converted to -180° to 180°);
- the date check (failures removed): hour, day, month and year must be valid quantities;
- blacklist check (failures removed): any observation from Deck 732 from a specified year and region is blacklisted (Rayner et al., 2006, Kennedy et al, 2011a, Table S2).;

Other marine products (e.g., NOCSv2.0; Berry and Kent, 2011) solely use ship observations due to the lack of buoy metadata available. WIn contrast, we include moored buoys for this version and to produce climatologies because spatial coverage is of high importance. OHowever, ur final version recommended to users is a we provide ship-only (SHIP) and combined product but we have produced a combined (ALL) product for comparisons. This will be reassessed for future versions. Figure 65a shows the number of observations included in the initial selection per year, broken down by platform type. The breakdown for day and night time observations individually is near identical (not shown). Ship (PT = 5) observations make up almost the entire dataset until the 1990s. After this the number of moored buoys grows significantly to make up around ~50 % of observations from 2000 onwards. The ship-only product (removal of moored buoys) significantly reduces the number of observations in the recent period but gives a more consistent number of observations throughout the record. Our use of climate anomalies should mitigate biasing due to uneven sampling to some extent. Note that the number of gridboxes containing data may be a more relevant measure and that the vast increase in the number of buoys has not actually resulted in the same level of increase in spatial coverage in terms of gridboxes (compare 2018 annual average maps for ship-only and combined HadISDH.marine in Fig. S2).

3.2 Quality control processing

We have not used any of the pre-set flags from ICOADS processing to ensure methodological independence of HadISDH and a process that allows for exploration and analysis of different methodological choices. The quality control processing employed here largely follows the methodology for HadSST4 (Kennedy et al., 2019) with some changes to the climatology check and buddy check thresholds to increase regional sensitivity and additional humidity specific checks. A flag for whole number prevalence has also been added but this is used for uncertainty estimation and not to remove~~fail~~ an observation. All observations have their nearest 1° by 1° pentad mean climatology (source depends on iteration – Sect. 3.5) subtracted to create a climate anomaly.

Each observation is passed through a suite of ~~basic~~ quality control tests ~~(base qc)~~ which are summarised in Table 1 along with whether the quality control tests are used~~y~~ to ~~are~~ remove~~d~~ or just to flag~~ged~~ the observations, and the stage of processing at which they are applied. ~~consist of:~~

~~climatology check (failures flagged): T and T_d must be within a specified threshold of the nearest 1° by 1° pentad climatology;~~

~~supersaturation check (failures flagged – T_d only): T_d must not be greater than T.~~The climatology check differs from the static HadSST3 threshold of climatology for air temperature of +/- 8~~10~~° C. We have allowed for a variable threshold depending on the nearest 1° by 1° pentad climatology standard deviation σ. This is set at 5.5 σ. It accounts for the lower variability in the tropics and greater variability in the mid-latitudes. We have set minimum and maximum σ values of 1° C and 4° C respectively resulting in a minimum range of +/- 5.5° C and a maximum range of +/- 22° C. Several thresholds were tested with the selected threshold balancing avoiding acute cut-offs in the data distribution while still removing obviously bad data (Figs. S3 to S6). Given that outliers are assessed by comparing a point observation with a 1° by 1° pentad mean the thresholds have to be relatively large.

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

- ~~track check (failures flagged): the distance and direction travelled by the ship must be plausible and consistent;~~
- ~~repeated value check (failures flagged): a T or T_d value must not appear in more than 70 % of a ship track where there are at least 20 observations;~~
- ~~repeated saturation check (failures flagged – T_d only): saturation ($T_d = T$) must not persist for more than 48 hours where there are at least 4 observations;~~
- ~~buddy check (failures flagged – 3rd iteration only): T and T_d must be within a specified threshold of the average of its nearest neighbours in space and time;~~
- ~~whole number flag (whole numbers flagged): a T or T_d value must not appear as a whole number in more than 50 % of a ship track where there are at least 20 observations.~~

The buddy check compares each observation's climate anomaly with the average of the climate anomalies of its nearest neighbours in space and time, expanding the search area in space and time as necessary until at least one neighbour observation is found. The permitted difference is set by the climatological standard deviation of the candidate 1° by 1° pentad gridbox multiplied by an amount dependent on the number of neighbours present. There are five levels of searches:

1. ±1° latitude and longitude and ± 2 pentads: the climatological standard deviation is multiplied by 5.5, 5.0, 4.5 and 4.0 for 1-5, 6-15, 16-100 and >100 neighbouring observations respectively;
2. ±2° latitude and longitude and ± 2 pentads: the climatological standard deviation is multiplied by 5.5 for >1 neighbouring observation;
3. ±1° latitude and longitude and ± 4 pentads: the climatological standard deviation is multiplied by 5.5, 5.0, 4.5 and 4.0 for 1-5, 6-15, 16-100 and >100 neighbouring observations respectively;
4. ±2° latitude and longitude and ± 4 pentads: the climatological standard deviation is multiplied by 5.5 for >1 neighbouring observation;
5. no neighbour ±2° latitude and longitude and ± 4 pentads: the threshold is set at 500.

The thresholds used for the buddy check are wider than those previously used in HadSST3. This is to account for the greater variability of air and dew point temperature, and sparser observation coverage. It is only applied in the 3rd iteration of the quality control (Sect. 3.5).

Figure 65 shows the final number of observations passing through initial selection and then 3rd iteration quality control by platform (PT) type. The quality control does not significantly affect one platform over another. The performance of these tests is demonstrated for 4 example months in Figs. S3 to S6. These reveal a slight positive bias in the removed air temperature observations and negative bias in removed dew point temperature. Removals in terms of relative humidity and specific humidity similarly tend to have a negative bias. It is clear that the majority of grossly erroneous observations are removed. The change in climatology between iterations of the quality control process (Sect. 3.5) also makes a difference to removals. This is both because the observation driven climatologies do not provide complete spatial coverage and because the ERA-Interim climatologies are cooler and drier than the observations (Sect. 4.1). Removals are dense in the Northern Hemisphere and especially sparse around the tropics. The addition of the buddy check in the 3rd iteration considerably increases the removal rate, noticeably over the Southern Hemisphere and Tropics.

The quality-control flagging rate for the 3rd iteration reduces over time from ~25 % to ~18 %, as shown in Fig. S7. This is driven by the buddy check and track~~repeated saturation~~ check. Proportionally more observations are flagged during the daytime than night time but the interannual behaviour is very similar. The daytime increase is driven by the larger number of air temperature buddy and climatology check failures. This could be due to the issue of solar heating of the ship structure during the daytime. The main source of test fails by a large margin is the buddy check, followed by the climatology check and track~~repeated saturation~~ check. There doesn't appear to be a strong difference in the distribution of removals from each test between the 1973-1981 and 1982-1990 periods that might explain the pre-1982 moist bias (Fig. S8, Sect. 4.2). There is an increase in removals from repeated saturation and supersaturation events over time, particularly the late 2000s. This may be related to the decrease in psychrometer deployment over time and increase in electric and capacitance sensors as shown in Fig. 4. The latter have increased significantly since the mid-2000s.

The whole number flags show very different behaviour to the other checks and to each other over time in Fig. S7. These depend on the ability to assign each observation to a track/voyage and the frequency of whole number observations on that voyage, hence, these flags are not a true reflection of the whole number frequency. Compared to the actual proportion of whole numbers shown in Fig. S1, these tend to exaggerate the annual patterns but the shape is broadly similar. This method of identifying problematic whole numbers appears to under-sample the true distribution, especially for air temperature pre-1982. An additional deck-based check is applied later for estimating uncertainty from whole numbers (Sect. 3.4).

Note that the NOCSv2.0 dataset, with which we compare our specific humidity data, includes an outlier check that removes data greater than 4.5 standard deviations from the climatological mean. This test has already been applied within the ICOADS format and so the NOCSv2.0 excludes any data with ICOADS trimming flags set (Wolter 1997). We do not use the trimming flags to select data. They also apply a track check based on Kent and Challenor (2006).

3.3 Bias adjustments and associated uncertainties

Given the issues raised in Sect. 2, it is desirable to attempt to adjust the observations to improve the spatial and temporal homogeneity and accuracy of the data. As discussed in Sect. 2.1, we have not attempted to adjust for solar biases in this first version product. We have made adjustments for instrument and height biases and estimated uncertainties (summarised in Table 1) in these adjustments.

3.3.1 Application of adjustments for biases from un-aspirated instruments

We have shown that the majority of humidity observations have been made with a psychrometer (Fig. 4) and that 30-70 % of instruments with metadata available have been housed within a non-aspirated screen (Fig. 2). Berry and Kent (2011) found that applying a 3.4 % reduction to specific humidity observations from non-aspirated screens was a reasonable adjustment to remove the bias relative to aspirated/well ventilated observations (e.g., slings, whirled hygrometers or artificially aspirated instruments). Some uncertainty remains after adjustment which they estimated to be ~0.2 g kg^{-1}. We have used the hygrometer exposure metadata (EOH) or the thermometer exposure (EOT) if EOH does not exist. We assume good ventilation for any instruments that are aspirated (A), from a sling (SL) or ship's sling (SG) or from a whirling instrument (W). We assume poorer ventilation for instruments that are from a screen (S), ship's screen (SN) or are unscreened (US) and apply a bias adjustment. The reported exposure type of Ventilated Screens (VS) does not appear to mean that the screen is artificially ventilated and so bias adjustments are also applied to these. We do not apply adjustments to buoys and other non-ship data based on the assumption that these generally measure relative humidity directly. For any ship observations with no exposure information we apply 55 % of the 3.4 % adjustment based on the mean percentage of observations with EOH metadata that require an adjustment over the 1973-2014 (metadata) period). This partial adjustment factor follows the method of Berry and Kent (2011) and Josey et al. (1999) but differs in quantity. They assessed this over a shorter time period and found then that ~30 % of observations were from poorly ventilated instruments.

To estimate the uncertainty in the non-aspirated instrument adjustment applied U_i, we use the Berry and Kent (2011) and Josey et al. (1999) uncertainty estimate of 0.2 g kg^{-1} and apply this in all cases where an adjustment or partial adjustment has been applied. This is treated as a standard uncertainty (1 σ). In the case of partial adjustments for the ship observations with no metadata there is large uncertainty in both the adjustment and adjusted value. To account for this we use the amount of what would have been a full 3.4 % adjustment in addition to the 0.2 g kg^{-1} as the 1 σ uncertainty.

To carry these adjustments and uncertainties to all other humidity variables we start with q and then propagate the adjusted quantity~~ment~~ and adjusted quantity plus uncertainty~~ amounts~~ using the equations in Table S1. Using the original T (which does not need to be adjusted for poor ventilation) and ERA-Interim climatological surface pressure, e can be calculated from q. T_d and RH can be calculated from e and T. From these, the T_w and DPD can be calculated. The uncertainty is then obtained by subtracting the ~~new~~ adjusted quantit~~ies~~y from the ~~an~~ adjusted quantity plus uncertainty for each variable~~, beginning again from the adjusted q plus the 0.2 g kg^{-1} uncertainty and full adjustment magnitude in the case of ships without metadata~~.

3.3.2 Application of adjustments for biases from ship heights

After bias adjustment for poor ventilation, all variables are adjusted to approximately 10 m elevation. This serves to account for the inhomogeneity from the systematic increase in ship height over time and for spatial inhomogeneity between observations made at different heights. In the absence of height adjustments, increasing ship heights likely lead to a small decrease in air temperature and specific humidity over time (Berry and Kent, 2011) because these quantities generally decrease with height. As Fig. 3 shows, the standard deviations in ships' instrument heights exceed 5 m in most cases. Also, we have included buoys in the processing so far and these can be very low (~4 m, e.g. Gilhousen, 1987) relative to ship observing heights.

The height of the hygrometer (HOH) must be estimated (HOHest) as no metadata is available. In the case of psychrometers, which are the most common instruments listed in the ship metadata, the wet and dry bulb thermometers are co-located. Figure 3 shows that the visual observation height (HOP) is the most commonly available information, followed by the barometer height (HOB) and then thermometer height (HOT). It also shows the mean and standard deviation of all observing heights including the anemometer (HOA). Hence, HOHest is obtained using the following methods in preference order:

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

3. HOT present and >2 m: HOHest μ = HOT, σ = 1 m
4. HOA present and >12 m: HOHest μ = HOA – 10, σ = 9 m
5. No height metadata: HOHest μ = 16 m + ~~f(~~the linear trend in mean HOP/HOB/HOT height to the~~,~~ date of observation~~)~~, σ = 4.6 m +~~f(~~ the linear trend in standard deviation HOP/HOB/HOT height to the~~,~~ date of observation~~)~~

The μ and σ of the combined HOP, HOB and HOT increases from 16 m and 4.6 m respectively in January 1973 to 23 m and 11 m respectively in December 2014. Kent et al. (2007) and Berry and Kent (2011) used 16 m to 24 m between 1971 and 2007 so our estimate is very similar. The anemometer height is also required for the adjustments. We either use the provided HOA as long as it is greater than 2 m or set it to 10m above the HOHest. All buoys are assumed to be observing at 4 m, with anemometers at 5 m (http://www.ndbc.noaa.gov/bht.shtml).

Once HOHest has been obtained for each observation, the air temperature and specific humidity are adjusted to 10 m using bulk flux formulae. The methodology, assumptions and parameterisations largely follow that of Berry and Kent (2011), Berry (2009), Smith (1980, 1988) and Stull (1988). Essentially, the quantity of interest x can be adjusted to a reference height of 10 m as follows:

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

where x_* is the scaling parameter specific to that variable (e.g., friction velocity in the case of u, characteristic temperature or specific humidity in the case of T or q respectively), κ is the von Karman constant (0.41 used here), z_x is the observation height of the variable of interest, ψ_x is the stability correction for the variable of 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 of 10m and is a function of ~~f(~~$10/L$~~)~~ and L is the Monin-Obukov Length.

An iterative approach (as done for Berry and Kent 2011) is required to resolve Eq. (1) because we only have basic meteorological variables available at a single height for each observation. We start from T, q, u, sea surface temperature (SST), the co-located 1° by 1° gridbox pentad climatological surface pressure from ERA-Interim (climP), HOHest which becomes both z_q and z_t and our estimated anemometer height which becomes z_u.

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 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 < 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 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 = q_{sat}(SST)*~~x~~0.98 and $u_0 = 0$. Clearly, with so many necessary approximations there are many different plausible methodological choices, hence the need for multiple independent analyses that explore these different choices in order to quantify the structural uncertainty.

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

- if (SST - T) > 0.2 °C: L = -50 m, unstable conditions are assumed;
- if (SST - T) < -0.2 °C: L = 50 m, stable conditions assumed;
- if (SST = T) +/- 0.2 °C: L = 5000 m, neutral conditions assumed where L tends to ∞.

We also start with an assumption that the 10 m wind speed in neutral conditions $u_{10n} = u$. The iteration is continued until L converges to within 0.1 m, which it generally does. If after 100 iterations there is no convergence we either apply no adjustment or if absolute L is large (> 500 m) we assume neutral conditions and take L (and all other parameters) as they are. In cases where u_* is very large (it should be < 0.5 m s^{-1} [Stull, 1988]) we also apply no adjustment. The iteration involves 21 steps as described in the Supplementary Material.

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

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

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

where the neutral stability heat transfer coefficient z_{t0} = 0.001 m and the neutral stability moisture transfer coefficient z_{q0} = 0.0012 m (Smith 1988). The adjusted values for T_{10} and q_{10} can then be calculated from Eq. (1). From these we recalculate the other humidity variables using the equations in Table S1.

There is uncertainty in the obtained HOHest. Given that this is a best estimate we assume that the uncertainty in the height is normally distributed and use the standard deviation in the height estimate used HOHest to calculate an uncertainty range in the height adjusted value *x* (where *x* is any of *T*, *q* etc.) of xH_{min} to xH_{max}. Following the 'two out of three chance' rule in the BIPM Guide to the Expression of Uncertainty in Measurement (BIPM, 2008), the standard uncertainty (1 σ) for the height adjusted value (U_h) is then given by:

$$U_h = \frac{xH_{max} - xH_{min}}{2} \quad (3)$$

The range xH_{min} to xH_{max} depends on the source of HOHest and associated σ, as listed above. There are several scenarios where estimating the uncertainty in this way is not possible or calculation of an adjustment is not possible. Also, U_h for buoys is highly uncertain given the lack of height information available. These alternative scenarios are documented in Table 21.

3.4 Estimating residual uncertainty at the observation level

Three other sources of uncertainty affect the marine humidity data at the observation level. These are measurement uncertainty U_m, climatology uncertainty U_c and whole number uncertainty U_w. These are all assessed as 1 σ standard uncertainties.

We have estimated U_m for each observation following the method used for HadISDH.land (Willett et al., 2013, 2014). This assumes that humidity was measured using a pyschrometer which is a reasonable assumption for the marine ship data (Fig. 4). The HadISDH.land measurement uncertainty is based on an estimated standard (1 σ) uncertainty in the wet bulb and dry bulb instruments of 0.15° C and 0.2° C respectively. As shown in Table S3, the equivalent uncertainty for the other variables depends on the temperature. The uncertainty is applied as a standard uncertainty in RH depending on which bin the air temperature falls in. This is then propagated through the other variables starting with vapour pressure, using the equations in Table S1.

Whole numbers of air and/or dew point temperature that have either been flagged as such during quality control (Sect. 3.2), or that belong to a source deck/year where whole numbers make up more than two times the frequency of other decimal places (Table S4), are given and uncertainty U_w. These decks and years where whole numbers are very common differ for air and/or dew point temperature. Clearly with so many decks affected, the removal of entire decks to remove any whole number biasing could easily reduce sampling to critically low levels. We cannot distinguish between observations that have been rounded versus those that have been truncated so we assume that all offending whole numbers have been rounded. This means that the value could be anywhere between ± 0.5° C, with a uniform distribution. Hence, where only air or dew point temperature is an offending whole number the standard 1 σ uncertainty expressed in air or dew point temperature (° C) is:

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

Where both air and dew point temperature are offending whole numbers the standard 1 σ uncertainty expressed in air or dew point temperature (° C) for dew point depression, relative humidity and wet bulb temperature is:

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

There is uncertainty U_c in the climatological values used to calculate climate anomalies because of missing data over time, uneven and sparse sampling in space and also the inevitable mismatch between a point observation and a 1° by 1° gridded pentad climatology. This uncertainty reduces with the number of observations contributing to the climatology N_{obs} and with the variability of the region σ_{clim}. The climatologies used to create the anomalies have undergone spatial and temporal interpolation to move from 5° by 5° gridded monthly climatologies and climatological standard deviations σ_{clim} to maximise coverage and so it is not straightforward to assess the number of observations contributing to each 1° by 1° gridded pentad climatology and the true σ_{clim} is likely greater. The minimum number of years required to be present over the 30 year climatology period is 10. Therefore, we assume a worst case scenario of N_{obs} = 10. Hence, for a standard 1σ uncertainty:

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

3.5 Gridding of actual and anomaly values and uncertainty

To create a quasi-global monitoring product the raw observations need to be gridded. The spatial density is too low for high resolution grids and the intended purpose is for this marine product to be blended with the HadISDH.land humidity product which is on a 5° by 5° grid at monthly resolution. Hence, the point~~raw~~ hourly observations must be averaged to monthly mean gridded values.

The sparsity of the data means that there is a risk of bias due to poor sampling. A 5° by 5° gridbox covers an area greater than 500 km^2 by 500 km^2 which, despite the large correlation decay distances of both temperature and humidity, can include considerable variability. Furthermore, a monthly mean can be made up of a strong diurnal cycle and considerable synoptic variability. This is minimised by the use of climate anomalies but regardless, care should be taken to ensure sufficient sampling density while maximising coverage where possible.

Several data-density criteria were trialled to balance spatial coverage and poor representativeness (high variance) of the gridbox averages. Climate anomalies are created at the raw observation level by subtracting the nearest 1° by 1° pentad climatology (1981-2010) and so we can grid both the actual values and the anomalies. Gridding of the anomalies is safer than gridding actual values in terms of biasing through poor sampling density because the correlation length scales of anomalies are higher than for actual temperatures. Initially, ERA-Interim is used to provide a climatology. This then requires an iterative approach to produce an initial observation-based climatology and improve the climatology through quality control. To reduce biasing further we grid the data in six stages to create an average at each stage. The entire process including quality control, bias adjustment, gridding and three iterations, is shown diagrammatically in Fig. 5~~6~~ and each gridding stage described below.

1. Create 1° by 1° 3-hourly gridded means of~~from~~ the~~the~~ hourly observations of actuals and anomalies; there must be at least one observation.
2. Create separate 1° by 1° daytime and night time gridded means of the~~from the~~ 1° by 1° 3-hourly gridded mean actuals and anomalies; there must be at least one 1° by 1° 3-hourly grid.
3. Create 5° by 5° monthly daytime and night time gridded means of~~from~~ the 1° by 1° daytime and night time gridded mean actuals and anomalies; there must be at least 0.3*days in the month of 1° by 1° daily grids.

4. Create combined 5° by 5° monthly gridded means of~~from~~ the 5° by 5° monthly daytime and night time gridded mean actuals and anomalies; there must be at least 1 5° by 5° monthly daytime or night time gridded mean.
5. Create 1981-2010 5° by 5° monthly mean climatologies and standard deviations from the 5° by 5° monthly gridded means of actuals and anomalies; there must be at least 10 5° by 5° monthly gridded means.
6. Renormalise the gridded anomalies by subtracting the monthly anomaly 1981-2010 climatology to remove biases from use of the previous iteration climatology (Sect. 4.1).

At each iteration the gridded observation based climatologies are infilled linearly over small gaps in space and time and then interpolated down to 1° by 1° pentad resolution. The observations are too sparse to create such high-resolution grids directly.

The observation uncertainties also need to be gridded and the total observation uncertainty U_o calculated. Ships move around, and so their uncertainties also track around the globe. This means that the uncertainty in any one point / gridbox bears some relationship to nearby points / gridboxes over time and space and cannot be treated independently. Correlation needs to be accounted for both in gridding and subsequently creating regional averages from gridboxes to avoid underestimation. The five sources of observation uncertainty are summarised in Table 2~~1~~. The non-aspirated instrument adjustment uncertainty U_i, height adjustment uncertainty U_h and climatology uncertainty U_c persist over time and space as ships move around. These are accordingly treated as correlating completely within one gridbox month. The measurement uncertainty U_m, and whole number uncertainty U_w are likely to differ observation to observation and so treated ~~h~~as having no correlation within one gridbox month. Hence, observation uncertainty sources are first gridded individually, following the first four steps outlined above and taking into account correlation where necessary. ~~The gridded uncertainty sources are then combined to give a total observation uncertainty for each gridbox.~~ For those that do not correlate (U_m and U_w) the gridbox mean uncertainties U_{gb} for each source are combined over N points in time and space as follows:

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

For those sources that do correlate (U_c, U_i and U_h), assuming r = 1, the gridbox mean uncertainties U_{gb} for each source are combined over N points in time and space as follows:

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

To create the total observational uncertainty for each gridbox the gridbox quantities of the five uncertainty sources can then be combined in quadrature:

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

Given the general sparsity of observations across each gridbox month and the uneven distribution of observations across each gridbox and over time there is also a gridbox sampling uncertainty component, U_s. This is estimated directly at the 5° by 5° monthly gridbox level and follows the methodology applied for HadISDH.land (Willett et al., 2013, 2014), denoted SE^2, which is based on station-based observations from Jones et al (1997):

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

.

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 the number of stations contributing to the gridbox mean in each month. The mean variance of individual stations within the gridbox is estimated as:

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

where $\hat{S}^2$ is the variance of the gridbox monthly anomalies over the 1982-2010 climatology period and N_{SC} is the mean number of stations contributing to the gridbox over the climatology period. The mean inter-site correlation is estimated by:

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

where X is the diagonal distance across the gridbox and x_0 is the correlation decay length between gridbox means. We calculate x_0 as the distance (gridbox midpoint to midpoint) at which correlation reduces to 1/e. To account for the fact that marine observations generally move around at each time point we use the concept of pseudo-stations to modify this methodology. For any one day there could be 25 1° by 1° gridboxes and so we assume that the maximum number of pseudo-stations per gridbox is 25 which is broadly consistent with the number of stations per gridbox in HadISDH.land. Over a month then, there could be a maximum of 775 1° by 1° daily gridboxes contributing to each 5° by 5° monthly gridbox. Given ubiquitous missing data and sparse sampling the maximum in practice is closer to 600. Using these values we then scale the actual number of 1° by 1° daily gridboxes contributing to each 5° by 5° monthly gridbox to provide a pseudo-station number between 1 and 25 for each month (N_s) and then the average over the climatology period (N_{SC}).

The gridbox U_o and U_s uncertainties are then combined in quadrature, assuming no correlation between the two sources. This gives the full gridbox uncertainty U_f. Calculation of regional average uncertainty and spatial coverage uncertainty is covered in Sect. 4.

4 Analysis and validity of the gridded product

The final gridded marine humidity monitoring product presented as HadISDH.marine.1.0.0.2018f is the result of the 3rd iteration quality-control and bias-adjustment of ship-only observations average into 5° by 5° gridded monthly means (Fig. 56). There are four reasons for only using the ship observations. Firstly, the increase in spatial coverage in the combined ship and buoy product is actually fairly small (Fig. S2) and only during the latter part of the record. Secondly, a dataset intended for detecting long-term changes in climate should have reasonably consistent input data and coverage over time. Thirdly, we believe that the buoy data are less reliable given their proximity to the sea surface and exposure to sea spray contamination in addition to the lower maintenance frequency compared to ship data. Fourthly, there are no metadata available for buoy observations which makes it difficult to apply necessary bias adjustments or estimate uncertainties. Actual monthly means, anomalies from the 1981-2010 climatology (not standardised by division with the standard deviation), the climatological means and standard deviation of the climatologies, uncertainty components and number of observations for both products are all made available as netCDF from www.metoffice.gov.uk/hadobs/hadisdh/.

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

At the end of each iteration (Fig. 5~~6~~), observation-based climatology fields are created at both the monthly 5° by 5° grid and, by interpolation, pentad 1° by 1° grid (Sect. 3.5). These are then used to quality control and create anomaly values for the next iteration. Hence, the 2nd iteration quality-controlled data are used to build the final 3rd iteration and therefore, there should be no lasting effect from having used the ERA-Interim fields initially. The quality-controlled, buddy-checked and bias-adjusted 3rd iteration is used to create the final climatology provided to users.

To compare the use of ERA-Interim versus the observation based climatology to calculate anomalies and quality control the data we show~~Specific humidity, relative humidity and air temperature~~ difference maps of the 2nd iteration minus ERA-Interim pentad 1° by 1° grid climatologies and climatological standard deviations ~~are shown~~ in Figs. S9 to S14 for a selection of pentads and variables. Note that ERA-Interim fields are for 2 m above the ocean surface whereas the raw observations range between approximately 10 m to 30 m above the surface. In normal conditions we may therefore expect ERA-Interim to provide climatologies that are warmer and moister than the observations. However, overall, ERA-Interim appears drier (both in absolute and relative terms) and cooler than the observation based climatologies. For humidity this is consistent with the results of Kent et al. (2014). For the majority of gridboxes these differences are within ± 2 g kg^{-1}, %rh and °C. However, differences are especially strong around coastlines with magnitudes exceeding ± 10 g kg^{-1}, %rh and °C. This is to be expected given that ERA-Interim coastal gridboxes will include effects from land, especially at the relatively coarse 1° by 1° grid resolution. For relative humidity there are more regions where ERA-Interim is more saturated and there is more seasonality in the differences. Relative humidity is less stable spatially and on synoptic time scales and also more susceptible to biases and errors than specific humidity and air temperature, largely because it is affected by errors in both air temperature and dew point temperature. For temperature, the coastal difference can be positive or negative depending on the season.

The climatological standard deviations are generally lower in the 2nd iteration observations compared to ERA-Interim. Differences are generally between ± 2 g kg^{-1}, %rh and °C but for relative humidity there are expansive regions in the extratropics to mid-latitudes, especially in the Northern Hemisphere where climatological standard deviations are up to 5 %rh lower in the observations. The generally lower variability in the observation-based climatology is to be expected given the interpolation from monthly mean resolution and interpolation over neighbouring gridboxes where data coverage is limited. However, much of the tropics, particularly in the Southern Hemisphere tends to show more variability in the observations. Similarly, many of the peripheral gridboxes (those at the edge of the spatial coverage and therefore more likely to be interpolated from nearby gridboxes rather than based on actual data) show higher variability for specific and relative humidity and lower variability for air temperature. All of these gridboxes are in data sparse regions which likely contributes to the higher variability. Ideally, observation based climatologies would be created directly at the pentad 1° by 1° grid but this severely reduces spatial coverage of the climatology fields and any product based on them. A balance has to be made between coverage and quality.

Annual mean 5° by 5° climatologies (no interpolation) from the 3rd iteration quality-controlled, bias-adjusted ship-only product are shown in Fig. 7 for specific humidity, relative humidity, air temperature and dew point temperature. These have a minimum data presence threshold of 10 years for each month over the climatology period and at least 9 climatological months present for the annual climatology. Data coverage is virtually non-existent in the Southern Hemisphere below 40° S and Northern Hemisphere coverage diminishes drastically above 60° N. These climatologies are as expected for these variables and compare well in terms of broad spatial patterns with ERA-Interim (not shown). There is good spatial consistency considering that no interpolation has been conducted meaning that any erroneous gridboxes should stand out. We conclude that as a first version product, these climatologies look reasonable.

4.2 Analyses of global averages for various processing stages and with other products

Global average quantities are key measures of climate change and so we focus here on the differences arising from the various processing steps of HadISDH.marine along with the NOCSv2.0 specific humidity and ERA-Interim reanalysis products. Global averages ~~(70° S to 70° N)~~ have been created by weighting each gridbox by the cosine of its latitude at gridbox centre. All timeseries shown are the renormalised anomalies with a zero-mean over the 1981-2010 period. Figs. 8 to 11 show timeseries for specific humidity, relative humidity, dew point temperature and air temperature respectively. Decadal linear trends (shown) are computed using ordinary least squares regression with ~~the median of pairwise slopes with~~ ranges representing the 90^{th} percentile confidence interval calculated using AR(1) correction (Santer et al., 2008).

For all variables, there are only small differences in the global average timeseries between the various processing steps – from the raw data (noQC) to the 3^{rd} iteration quality-controlled (no~~N~~BA~~C~~ [no bias adjustment]) and then the bias-adjusted data (~~BClocal~~BA). They are smallest for air temperature and largest for relative humidity but all steps result in global average trends that are significant and in the same direction, and have similar interannual variability. We consider these trends to be significant because the 90^{th} percentile confidence intervals around the trend are not large enough to bring the direction of the trends into question. ~~Both the interannual variability and long-term linear trends are very similar~~T~~, and t~~he trends in the global average are positive over the 1973-2018 period for specific humidity, dew point temperature and air temperature, and negative for relative humidity. ~~We consider these trends to be significant because the 90^{th} percentile confidence intervals around the trend are not large enough to bring the direction of the trends into question.~~ The linear trends for the final HadISDH.marine.1.0.0.2018f version are 0.07 ± 0.02~~1~~ g kg $^{-1}$ decade $^{-1}$, -0.09 ± 0.08~~2~~ %rh decade $^{-1}$, 0.09~~8~~ ± 0.02~~1~~° C decade $^{-1}$ and 0.11 ± 0.03~~1~~° C decade $^{-1}$ for specific humidity, relative humidity, dew point temperature and air temperature respectively. Hence, we conclude that HadISDH.marine shows moistening and warming since the 1970s globally in actual terms but that the air above the oceans appears to have~~has~~ become less saturated and drier in relative terms. This differs from theoretical expectation where changes in relative humidity over ocean are strongly energetically constrained to be small, of the order of 1% K^{-1} or less, and generally positive (Held and Soden, 2006; Schneider et al., 2010). ~~and m~~Model-based expectations also suggest~~of a~~ small positive changes~~or no change in relative humidity over ocean~~ (Byrne and O'Gorman, 2013, 2016, 2018). Despite careful quality control and bias-adjustment the previously noted moist humidity bias pre-1982 is still apparent in the bias-adjusted (BA) data. The linear trend in relative humidity from 1982 to 2018 is -0.03 ± 0.13 %rh decade $^{-1}$, and therefore not significantly decreasing which is more consistent with expectation.

Since there are considerable known issues affecting the marine humidity data, and because there are large outliers (Figs. S3 to S6), the effect of quality (noQC compared to noBA~~NBC~~), might be expected to be large. Furthermore, approximately 25 %, dropping steadily over time to 18 % of the initial selection of data have been removed by the quality control (Fig. 5~~6~~), so there is a considerable difference in the amount of data contributing to the quality-controlled version compared to the raw version. Despite all of this, differences are relatively small. Overall, the quality control makes the positive trends smaller (specific humidity, dew point temperature and air temperature) and negative trends larger (relative humidity). The effect of quality control, including buddy checking, is largest in the 1970s to early 1980s, when the largest amount of data was removed by quality control. This is especially noticeable for relative humidity and dew point temperature~~, and the same period as the previously noted moist relative humidity bias~~, suggesting that the pre-1982 bias, although present to some extent in the raw (noQC) data, could be ~~an artefact of~~exacerbated by the quality control. This could be due to erroneous removal of good data but investigation (Figs. S3 to S8) suggests that much of the data removal was appropriate – many very low relative humidity values were removed. It could also be an artefact of the reduced number of observations after quality control, reducing the chance of averaging out random error. To explore whether the presence of whole numbers in the record has contributed to the pre-1982 bias we have processed a bias adjusted version with all whole number flagged data (Table 1) removed (BA_no_whole) which is shown against the noQC and BA versions in Fig. 9d. The resulting global average trend is largest in the BA_no_whole version, even over the 1982-2018 period, and the pre-1982 bias is still clear. ~~Either way,~~We conclude that the pre-1982 moist bias remains~~is~~ apparent in HadISDH.marine, and is ~~as yet~~ not yet well understood, and quality control of the pre-1982 data is an area for more research in future versions.

The bias adjustment (BA~~Clocal~~, ~~BClocalHGT~~BA_HGT, ~~BClocalINST~~BA_INST) reduces the negative trends in relative humidity ~~both~~ compared to the ~~raw (noQC) and~~ quality-controlled (no~~N~~BA~~C~~) data,~~. It~~ and increases the positive trends in specific humidity and dew point temperature relative to the quality-controlled data~~ but reduces the trends compared to the raw data~~. The effect of bias adjustment is negligible for air temperature, which only has adjustment for ship height applied. For the humidity variables the height adjustment has a far larger effect than the non-aspirated instrument adjustment. The non-aspirated instrument adjustment makes the positive trends in specific humidity and dew point temperature slightly smaller and the negative trends in relative humidity slightly larger. The height adjustment has the opposite effect. For relative humidity, the bias adjustments appear to have introduced greater intra-decadal scale variability but retained the interannual patterns, again highlighting the sensitivity of relative humidity compared to the other variables. Given that these biases exist we do have to try and mitigate their impact. However, this is a focus area for investigation and improvements in future versions of HadISDH.marine.

The timeseries that include data from moored buoys compared to those from ships only (‘all’ versus ‘ship’) show smaller positive trends for specific humidity~~, dew point temperature~~ and air temperature and larger negative trends for relative humidity. Moored buoys begin to play a role from the late 1980s, increasing in number dramatically to make up over 50 % of the observations by 2015. The ‘all’ timeseries can be seen to diverge slightly from the ‘ship’ timeseries in the latter part of the record. Therefore, it is more consistent to produce the final HadISDH.marine version without inclusion of moored buoy data.

Before quality control there are more daytime ship observations than night time ship observations in the early record (~1 000 000 compared to ~800 000 per year) but this evens out by the end of the record to ~900 000 per year. However, the quality control removes more daytime observations than night time observations, especially in the 1970s and 1980s such that both contribute ~700 000 observations per year, dipping in the middle of the record. There has been no bias adjustment for solar heating of ships applied in this version of HadISDH.marine so the daytime data may contain some artefacts of solar heating. If this is a problem it should affect the air temperature and relative humidity but not the dew point temperature or specific humidity (Sect. 2.1). While the full dataset (both) combines both daytime and night time data, for various gridboxes and seasons there is only either a daytime or night time value present. As such, the ‘both’ timeseries and its linear trend may not be a straightforward average of the ‘day’ and ‘night’ timeseries and trends. ~~In the case of specific humidity, the daytime and night time global average timeseries have slightly larger positive trends than the combined timeseries and for relative humidity they have smaller negative trends than the combined series.~~ For specific humidity, dew point temperature and air temperature the ‘day’ and ‘night’ trend differences are essentially negligible, with linear trends identical or ~~differences~~ within 0.01 g kg^{-1} decade^{-1} ~~or 0.01° C decade^{-1}~~. Even for relative humidity the differences are small. The ‘day’ timeseries gives the largest negative trend followed by ‘both’ which is 0.01 %rh decade^{-1} smaller and then ‘night’ which is 0.02 %rh decade^{-1} smaller again. The negligible differences in air temperature suggest that solar heating is not a significant concern at least at the global average scale. Relative humidity is very sensitive to any differences in the data but even these differences are fairly small and do not change the overall conclusion of decreasing full-period~~long-term~~ trends and no significant trend over the 1982-2018 period. ‘Night’ trends are often thought to provide a better signal of change because they are generally free from convective and shortwave radiative processes and more a measure of outgoing longwave radiation. The main conclusion here is that trends and variability are very similar in the daytime, night time and combined timeseries which adds confidence in their representativeness of real-world trends and variability.

IOverall, at least in terms of linear trend direction, HadISDH.marine compares well with other monitoring estimates from NOCSv2.0 and ERA-Interim and to other reanalyses and older products (Fig. 1). ERA-Interim in Figs. 8 to 11 is from analysis fields of 2 m air temperature and dew point temperature and has been masked to ocean coverage using a 1° by 1° land-sea mask and also to HadISDH.marine coverage for comparison. Note that the ERA-Interim timeseries shown in Figs. 8 to 11 are from analysis fields of 2 m air temperature and dew point temperature, whereas the timeseries shown in Fig. 1 are from background forecast values to avoid biases introduced from ship data and ocean-only points over open sea. They are very similar at least in terms of the global average. Both NOCSv2.0 and HadISDH.marine are estimates of 10 m quantities and the NOCSv2.0 coverage is similar to that of HadISDH.marine but it only extends to 2015. NOCSv2.0 shows the largest trends in specific humidity over the 1979-2015 common period, 0.043 g kg $^{-1}$ decade $^{-1}$ greater than HadISDH.marine. The interannual patterns are broadly similar but with some differences showing that methodological choices do make a difference, given that the underlying observations are from the same source. ERA-Interim shows very weak moistening compared to HadISDH.marine for specific humidity and dew point temperature and slightly weaker warming for air temperature. Over the longer 1979-2018 period ERA-Interim trends are slightly larger for specific humidity but still weaker than in HadISDH.marine. The decreasing saturation in relative humidity is very strong in ERA-Interim at more than 23 times the HadISDH.marine trend over the common period. The masking to HadISDH.marine coverage surprisingly makes very little difference in the linear trends, they are slightly more negative, and only small year-to-year differences. Interannual behaviour does differ, especially for relative humidity and especially in the period up to the early 1990s where ERA-Interim is warmer and wetter generally, thus moderating the long-term trends in specific humidity, dew point temperature and air temperature. Note that the ERA-Interim background field relative humidity shown in Fig, 1 also shows a decrease but to a lesser extent than the analysis fields (Fig. 9) which include ship data. Agreement is closest for air temperature in both trends and variability.

The decreasing relative humidity trends over ocean are similar toconsistent with the drying seen in HadISDH.land and ERA-Interim land relative humidity (Fig. 1); land linear trends are 0.03 %rh more negative at -0.12 (-027 to -0.03) %rh 10 yr^{-1} over the same 1973 to 2018 period. The timeseries pattern is quite different though with marine relative humidity decreasing throughout the period around large variability and land relative humidity clearly decreasing from 2000. The greater sensitivity of relative humidity to observation errors, biases and sampling issues makes the conclusion of long-term drying an uncertain one but agreement with ERA-Interim adds some weight to this conclusion.

For the final HadISDH.marine.1.0.0.2018f product the regional average uncertainty is also computed and shown for the global average (70° S to 70° N) in Fig. 12. This includes the total observation uncertainty, which covers uncertainty components for instrument adjustment, height adjustment, measurement, climatology and whole number uncertainty (Table 2). In addition, the regional average uncertainty includes the gridbox sampling uncertainty and also a spatial coverage uncertainty, following the method applied for HadISDH.land (Willett et al., 2014). The coverage uncertainty essentially uses the variability between ERA-Interim full coverage compared to ERA-Interim with HadISDH.marine coverage to estimate uncertainty. To obtain uncertainty in the global average from the gridbox uncertainties correlation in time and space should be taken into account. It is not trivial to assess the true spatial and temporal correlation of the various uncertainty sources. In reality, although ships move around over space and time, implying some correlation, the contributing sources to each ~500 km^2 gridbox monthly mean differ widely. Therefore, for this first version product we assume no correlation between gridboxes in time or space and take the simple approach of the quadrature combination of uncertainty sources, noting that this is a lower limit on uncertainties.

The uncertainty in the global averages (Fig. 12) isare larger than the equivalent time-series for land (see Fig. 12 in Willett et al., 2014). The coverage uncertainty (accounting for observation gaps in space and time) is generally the largest source of uncertainty with the exception of relative humidity and dew point depression. For the latter two, the total observation uncertainty makes up the greatest contribution. In all cases the total observation uncertainty is larger at the beginning and especially the end of the records, where there are fewer/no metadata with which to apply bias adjustments. The contribution from sampling uncertainty (gridbox spatial and temporal coverage) is generally very small except for relative humidity. This is as expected given that the correlation decay distance of humidity should generally be larger over ocean than over land given the homogeneous surface altitude and composition. Overall, the magnitudes of the uncertainties are small relative to the magnitudes of long-term trends and variability in all variables except for relative humidity and dew point depression. This suggests that there is good confidence in changes in absolute measures of humidity over ocean (e.g., specific humidity), and also air temperature, but lower confidence in changes in the relative humidity. The warming and moistening are further corroborated by strong theoretical reasoning based on laws of physics governing the expectation that specific humidity should have increased over the period of record given the warming of the oceans and atmosphere that has occurred (Hartmann et al., 2013). The uncertainty model makes many assumptions over correlation of uncertainty in space and time. It is likely that we have overestimated the uncertainty at the gridbox scale by assuming complete correlation for height adjustment uncertainty, instrument adjustment uncertainty and climatological uncertainty. Conversely, we have likely underestimated the uncertainty at the regional average level by assuming no correlation. This is certainly an area for improvement in future versions.

4.3 Decadal trends across the globe presented by HadISDH.marine

Figure 13 shows the decadal linear trends for specific humidity, relative humidity, dew point temperature and air temperature for HadISDH.marine.1.0.0.2018f. The completeness criteria for trend fitting is 70 %, more strict than for the climatologies (Fig. 7). This results in poorer spatial coverage especially in the Southern Hemisphere. Clearly, there are no data points outside 70° S to 70° N, hence the restriction of the global average timeseries to this region is sensible. The tropical and Southern Hemisphere Pacific Ocean, and Southern Hemisphere Atlantic Ocean have virtually no data coverage. Overall, the appearance of the trends shows good spatial consistency, with few gridboxes standing out as obviously erroneous. There has been no interpolation across gridboxes that would have smoothed out any outliers, and so the lack of these outlying gridboxes suggests that the data are of reasonable quality for this long-term analysis at least. Trends are as expected from the global average timeseries – generally moistening and warming but becoming less saturated. The same is true over land (Willett et al., 2014).

The moistening shown in specific humidity and dew point temperature (Fig. 13 panels a, b and e, f) is widespread. The majority of gridboxes are considered to be statistically significant in that the 90th percentile confidence interval around the trend magnitude is the same sign as the trend and does not encompass zero. The largest increases in specific humidity are in the lower latitudes where as the largest increases in dew point temperature are more spread out with a tendency towards the extratropics and mid-latitudes. There are a few regions where there are clusters of gridboxes with drying trends. These are generally consistent between the specific humidity and dew point temperature, especially in the few cases where these negative trends are significant such as the central Pacific, the east coast of Brazil, the southern coast of Australia and around New Zealand.

Marine air temperature shows widespread and significant warming, in agreement with HadNMAT2 (Kent et al., 2013). Very few of the gridboxes with a negative trend are significant. In some cases they are in similar locations to the drying trends seen in specific humidity and/or dew point temperature e.g., the coast south of Australia around Tasmania, the east coast of Brazil. The warming is stronger in the northern mid--latitudes with the Baltic, Mediterranean and Red Seas showing particularly strong warming consistent with strongly increasing dew point temperature and specific humidity.

Whilst relative humidity is more sensitive to methodological choices and observational errors, the broad spatially coherent structures to the regions of increasing and decreasing saturation, with broadscale significance, are very encouraging in terms of data quality. Furthermore, the drying trends tend to be around the mid-latitudes while the increasing saturation trends are more around the tropics, as seen over land. We still urge caution in the use of marine relative humidity but these results collectively suggest that decreasing saturation might be a real feature.

5 Code and data availability

HadISDH.marine is available as 5° by 5° gridded fields of monthly means and anomalies along with a 1981-2010 climatology and uncertainty estimates at the gridbox scale. The data begin in January 1973 and continue to December 2018 (at time of writingprint) and will be updated annually. HadISDH.marine is publicly available from www.metoffice.gov.uk/hadobs/hadisdh/ under an Open Government license (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/) as netCDF and text files. Processing code (Python) can also be made available on request. HadISDH.marine data, derived diagnostics and plots can be found at www.metoffice.gov.uk/hadobs/hadisdh/indexMARINE.html (Willett et al., 202019). It should be cited using this paper and the following: Willett, K. M., Dunn, R. J. H., Kennedy, J. J. and Berry, D. I.: 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, 202019.

Commented [WK2]: To be updated with the CEDA archive link prior to publication.

Commented [WK3]: This should actually be Centre for Environmental Data Analysis, CEDA-link-to-DOIdata but this is not ready yet – it will be prior to publication.

This product forms one of the HadOBS (www.metoffice.gov.uk/hadobs) climate monitoring products and will be blended with the HadISDH.land product to create a global land and marine humidity monitoring product. Updates and exploratory analyses are documented at http://hadisdh.blogspot.co.uk and through the Met Office HadOBS twitter account @metofficeHadOBS.

6 Discussion and conclusions

Marine humidity data are susceptible to a considerable number of biases and sources of error that can be large in magnitude. We have cleaned the data where possible by applying quality control for outliers, supersaturation, repeated values and neighbour inconsistency which has removed up to 25 % of our initial selection in some years. We have also applied adjustments to account for biases arising from un-aspirated instrument types and differing observation heights over space and time. Care has also been taken to avoid diurnal and seasonal sampling biases as far as possible when building the gridded fields and the use of gridbox mean climate anomalies reduces remaining random error through averaging.

Spatial coverage of HadISDH.marine differs year to year. The coverage is generally poorer than seen for variables such as SST which benefit significantly from drifting buoy observations. Any further decline in observation and transmission of humidity from ships is of concern to our ability to robustly monitor surface humidity over oceans. Future versions may be able to make more use of humidity data from buoys but their proximity to the sea surface and difficulty of regular maintenance can lead to poor quality observations. The provision of digital metadata significantly improves our ability to quantify and account for biases in the data. Hence, the continuity of this metadata beyond 2014, and ideally an increase in quantity, also strongly affects our ability to robustly monitor ocean surface humidity. Given the current availability of ship data and metadata, and necessarily strict selection criteria and quality control, the resulting spatial coverage is good over the Northern Hemisphere outside of the high latitudes. There is very poor coverage over the Southern Hemisphere, especially south of 20° S. This means that our 'global' analyses are biased to the Northern Hemisphere. Care should be taken to account for different spatial coverage when comparing products. However, when comparing HadISDH to masked and unmasked ERA-Interim fields differences were surprisingly small.

We have shown that the observations are warm and moist relative to ERA-Interim reanalysis for the majority of the observed globe apart from the northwestern Pacific. This is despite ERA-Interim fields representing 2 m above the surface compared to the general observation heights of 10-30 m above the surface. Differences are largest around coastlines, particularly in the Red Sea and Persian Gulf. There is insufficient spatial coverage to produce a high resolution climatology from the data themselves, hence our use of ERA-Interim initially and then interpolated observation based fields. However, the lower resolution (5° by 5°) monthly mean climatologies from the final HadISDH.marine.1.0.0.2018f version show expected spatial patterns and have good spatial consistency, providing evidence that our data selection methods have resulted in reasonably high quality data.

The quality control and bias adjustment procedures have made small differences to the global average anomaly timeseries for specific humidity, dew point temperature and air temperature. This overall agreement in the global average timeseries between versions, and also between the daytime, night time and combined versions, increases confidence in the overall signal of increased moisture and warmth over oceans. These features show widespread spatial consistency in the HadISDH.marine.1.0.0.2018f gridbox decadal trends which also adds confidence. Hence, we can conclude that the ICOADS data are a useful source of humidity data for climate monitoring. However, we expect differences to be larger on smaller spatial scale analyses. HadISDH.marine shows consistency with other products in terms of long-term linear trends in the global averages. There are some differences year to year, with ERA-Interim showing warmer and moister anomalies prior to the early 1990s, and hence, smaller trends overall.

For relative humidity, differences between the versions can be large for any one year but the overall decreasing saturation trend appears to be robust. We conclude this because the trend is consistent across all processing steps, apparent in ERA-Interim fields and also has spatial consistency across the extratropics and mid-latitudes. This is a somewhat surprising result and one that should be treated cautiously. Theoretical and mModel--based analysis of changes in relative humidity over ocean under a warming climate suggest negligible or small positive changes (Held and Soden, 2006; Schneider et al., 2010; Byrne and O'Gorman, 2013, 2016, 2018). The temporal patterns in global average relative humidity are quite different to those over land whereas specific humidity shows similarity with the HadISDH.land timeseries, largely driven by the El Niño related peaks. The pre-1982 data have previously been noted as having a moist bias and our processing steps do not appear to have removed this feature. The trend excluding this earlier period (1982-2018) is no longer a significant decreasing trend and therefore more consistent with expectation. Removal of whole number flagged data appeared to exacerbate the pre-1982 bias and make the negative trends larger. Further work to assess the physical mechanisms that might lead to such trends is needed.

There are known issues with ERA-Interim in terms of its stability. For example, sea surface temperatures cooled around mid-2001 due to a change in the SST analysis product used (Simmons et al., 2014). This is very likely to affect humidity over the ocean surface in ERA-Interim. Similarly, changes in satellite streams over time can also affect the long-term stability of ERA-Interim, even in the surface fields. Also, the assimilated ship data are not adjusted for biases in the ERA-Interim assimilation. Clearly, there are various issues affecting both in-situ based monitoring products and reanalysis products such that neither one can be easily identified as the more accurate estimate. Analyses should take into account all available estimates and their strengths and weaknesses. Comparison of HadISDH.marine with satellite-based estimates of humidity over ocean will be an important next step.

We have attempted to quantify uncertainty in HadISDH.marine. The uncertainty analysis comprises observation uncertainty at the point of measurement which is then propagated through to gridbox averages taking correlation in space and time into account where relevant. Sampling uncertainty at the gridbox level due to uneven sampling across the gridbox in space and time is assessed. We have also provided uncertainty estimates in regional and global averages including coverage uncertainty. The propagation of gridbox observation and sampling uncertainty to large scale averages does not explicitly take into account correlation in these uncertainty quantities in space and time. As this is a first version monitoring product this simple method is seen as an appropriate first attempt to assess uncertainty. The ranges presented should be seen as a lower limit on the uncertainty. Overall, uncertainty in the global average is dominated by the coverage uncertainty for all variables except relative humidity and dew point depression. The total observation uncertainty is larger at the beginning, and especially the end of the record, where digital metadata are fewer or non-existent (post-2014). Overall, the uncertainty is small relative to the magnitude of long-term trends with the exception of relative humidity. We suspect that this is an overestimate at the gridbox level owing to assumptions of complete correlation in the height adjustment, instrument adjustment and climatology uncertainty components, and an underestimate at the regional average level given assumptions of no correlation. This is a first attempt to comprehensively quantify 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.

Berry, D. I. and Kent, E. C. : A new air-sea interaction gridded dataset from ICOADS with uncertainty estimates, Bulletin of the American Meteorological Society, 90, 645-656, 2009.

Berry, D. I. and Kent, E. C.: Air–Sea fluxes from ICOADS: the construction of a new gridded dataset with uncertainty estimates, Int. J. Climatol., 31, 987–1001, 2011.

BIPM, 2008: Evaluation of measurement data – Guide to the expression of uncertainty in measurement. JCGM 100:2008. https://www.bipm.org/en/publications/guides/gum.html

Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014: The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. Bull. Amer. Meteor. Soc., 95, 1431–1443, https://doi.org/10.1175/BAMS-D-13-00047.1

Bosilovich, M. G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P., Wargan, K., Chao, W., Reichle, R., Takacs, L., Vikhliaev, Y., Bloom, S., Collow, A., Firth, S., Labow, G., Partyka, G., Pawson, S., Reale, O., Schubert, S. D. and Suarez, M. : MERRA-2: Initial Evaluation of the Climate, Technical Report Series on Global Modeling and Data Assimilation, Volume 43, NASA/TM–2015-104606/Vol. 43, pp. 136. http://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/, 2015.

Buck, A. L. : New equations for computing vapor pressure and enhancement factor, J. Appl. Meteor., 20, 1527–1532, 1981.

Byrne, M. P. and P. A. O'Gorman, 2013: Link between land-ocean warming contrast and surface relative humidities in simulations with coupled climate models. Geophysical Ressearch Letters, 40 (19), 5223-5227, https://doi.org/10.1002/grl.50971.

Byrne, M. P. and P. A. O'Gorman, 2016: Understanding decreases in land relative humidity with global warming: conceptual model and GCM simulations. Journal of Climate, 29, 9045-9061, DOI: 10.1175/JCLI-D-16-0351.1.

Byrne, M. P. and O'Gorman, P. A.: Trends in continental temperature and humidity directly linked to ocean warming, Procedings of the National Academy of Sciences USA. 115(19), 4863-4868. doi: 10.1073/pnas.1722312115, 2018.

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), February 2019. https://cds.climate.copernicus.eu/cdsapp#!/home

Dai, A. : Recent climatology, variability, and trends in global surface humidity, J. Climate., 19, 3589–3606, 2006.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L. J., Bidlot, L., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Ebita, A., and co-authors, : The Japanese 55-year reanalysis "JRA-55": An interim report. SOLA, 7, 149-152, doi:10.2151/sola.2011-038, 2011.

Elliott, W. P., Ross, R. J. and Schwartz, B. : Effects on climate records of changes in National Weather Service humidity processing procedures, Journal of Climate, 11, 2424-2436, 1998.

Freeman, E.,; Woodruff, S. D.,; Worley, S. J.,; Lubker, S. J.,; Kent, E. C.,; Angel, W. E.,; Berry, D. I.,; Brohan, P.,; Eastman, R.,; Gates, L., ;-Gloeden, W.,; Ji, Zaihua,; Lawrimor, J.,; Rayner, N. A.,; Rosenhagen, G.,; Smith,

S. R., : ICOADS Release 3.0: a major update to the historical marine climate record, International Journal of Climatology, 37 (5). 2211-2232.10.1002/joc.4775, 2017.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M. and Zhao, B., : The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate, 30, 5419–5454,https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

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.

Held, I. M. and Soden, B. J., : Robust responses of the hydrological cycle to global warming. Journal of Climate, 19, 5686-5699, 2006.

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.

Santer, B. D., Thorne, P. W., Haimberger, L., Taylor, K. E., Wigley, T. M. L., Lanzante, J. R., Solomon, S., Free, M., Gleckler, P. J., Jones, P. D., Karl, T. R., Klein, S. A., Mears, C., Nychka, D., Schmidt, G. A., Sherwood, S. C. and Wentz, F. J.: Consistency of modelled and observed temperature trends in the tropical troposphere. Int. J. Climatol., 28, 1703-1722. doi:10.1002/joc.1756, 2008.

Schneider, T., O 'Gorman, P. A. and Levine, X. J.,: Water vapor and the dynamics of climate changes, Reviews in Geophysics, 48, RG3001, doi:10.1029/2009RG000302, 2010.

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. M., Dunn, R. J. H., Kennedy, J. J. and Berry, D. I.: 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, 2020.

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.

Commented [WK4]: This will actually be 'Centre of Environmental Data Analysis, link-to-ceda-DOIdata. This hasn't been set up yet but will be prior to publication.

Tables

Table 1. Description of quality control tests.

Test	**Description**	**1st and 2nd Iteration**	**3rd Iteration and Bias Adjusted**	**% of Observations Removed**
day / night	values likely to be affected by the solar heating of a ship where the sun was above the horizon an hour before the observation (based on the month, day, hour, latitude and longitude; Kent et al. (2013)) are flagged as 'day'	flagged	flagged	NA
climatology	T and T_d must be within a specified threshold of the nearest 1° by 1° pentad climatology	removed	removed	T = 2.39 and T_d = 5.14
supersaturation	T_d must not be greater than T (only T_d removed)	removed	removed	0.54
track	The distance and direction travelled by the ship must be plausible and consistent with the time between observations, normal ship speeds and observation locations before and after.	removed	removed	0.86
repeated value	A T or T_d value must not appear in more than 70 % of a ship track where there are at least 20 observations.	removed	removed	T = 0.04 and T_d = 0.06
repeated saturation	Saturation ($T_d = T$) must not persist for more than 48 hours within a ship track where there are at least 4 observations (only T_d removed).	removed	removed	0.54
buddy	T and T_d must be within a specified threshold of the average of nearest neighbours in space and time.	not applied	removed	T = 7.16 and T_d = 9.47
whole number	A T or T_d value must not appear as a whole number in more than 50 % of a ship track where there are at least 20 observations.	flagged	flagged	T = 11.73 and T_d = 8.20

Table 21. Description of the uncertainty elements affecting marine humidity. All uncertainties are assessed as 1σ uncertainty.

<table>
<tr><th colspan="2">Uncertainty Source</th><th>Description</th><th>Type</th><th>Formula</th><th>Correlation</th></tr>
<tr><td rowspan="2">U_i</td><td rowspan="2">Non-aspirated instrument adjustment uncertainty. Expressed as q (g kg^{-1}) and then propagated to other humidity variables</td><td>Adjusted poorly aspirated instrument: 0.2 g kg^{-1} in terms of q (following Berry and Kent, 2011 standard uncertainty assessment).</td><td rowspan="2">sStandard</td><td>0.2</td><td rowspan="2">Space and time,
r = 1</td></tr>
<tr><td>Partially adjusted unknown instrument: 0.2 g kg^{-1} + the full adjustment amount in terms of q.</td><td>$0.2 + 100\left(\frac{abs(q - q_{adj})}{55}\right)$</td></tr>
<tr><td>U_h</td><td>Observation height adjustment</td><td>Height adjusted ship and valid SST:</td><td>nNormally distributed</td><td>$\frac{xH_{max} - xH_{min}}{2}$</td><td>Space and time,
r = 1</td></tr>
</table>

	uncertainty. Expressed as T (°C) and q (g kg^{-1}) and then propagated to other humidity variables	assessed using the range of adjustments from a 1σ uncertainty in the height estimate.			
		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.	normally distributed	x_{adj} 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}).	standard	$\frac{T_{(adj)} - SST}{2}$ $\frac{q_{(adj)} - q_{sf}}{2}$ $q_{sf} = 0.98 q_{sat} f(SST)$	
		Height adjustment or uncertainty range not resolved, no valid SST: 0.1 °C in terms of T and 0.007q.	standard	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.	sStandard	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	uUniformly distributed	$\frac{0.5}{\sqrt{3}}$	None, r = 0

		deck in Table S4.			
		If both T and T_d are offending whole numbers then RH, T_w and DPD have a combined uncertainty.		$\frac{1}{\sqrt{3}}$	
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.	sStandard	$\frac{\sigma_{clim}}{\sqrt{N_{obs}}}$	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.	sStandard	$\sqrt{{U_i}^2 + {U_h}^2 + {U_m}^2 + {U_w}^2 + {U_c}^2}$	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 'station' variance, the mean inter-site correlation and the number of 'stations' contributing to the gridbox.	sStandard	$\frac{\left(\bar{s}_i^2 \bar{r}(1-\bar{r})\right)}{(1+(N_s-1)\bar{r})}$	Space and time to some extent, decreasing with space and time
U_{fg}	Full uncertainty of the gridbox	All gridbox uncertainty sources are combined, assuming no correlation between sources.	sStandard	$\sqrt{{U_{og}}^2 + {U_{sg}}^2}$	Space and time to some extent, decreasing with space and time

Figures

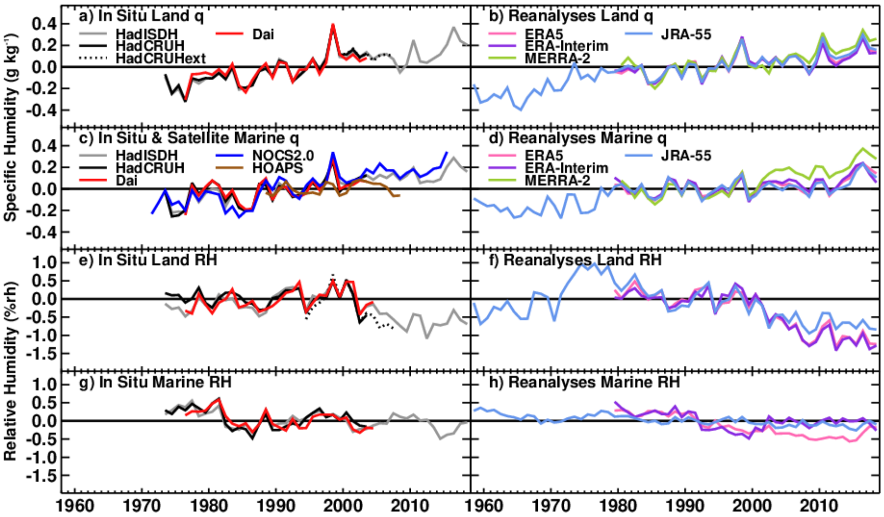


Figure 1 Global average surface humidity annual anomalies (base period: 1979–2003). For in-situ datasets, 2-m surface humidity is used over land and ~10-m over the oceans. For the reanalysis, 2-m humidity is used across the globe. For ERA-Interim and ERA5, ocean-only points over open sea are selected and background forecast values are used as opposed to analysis values to avoid incorporating biases from unadjusted ship data. All data have been given a mean of zero over the common period 1979–2003 to allow direct comparison, with HOAPS given a zero mean over the 1988–2003 period. [Sources: HadISDH (Willett et al., 2013, 2014); HadCRUH (Willett et al., 2008); Dai (Dai 2006); HadCRUHext (Simmons et al. 2010); NOCSv2.0 (Berry and Kent, 2009, 2011); HOAPS (Fennig et al. 2012), ERA-Interim (Dee et al., 2011), ERA5 (C3S 2017, Hersbach et al., 2018), MERRA-2 (Gelaro et al. 2017; Bosilovich et al. 2015) and JRA-55 (Ebita et al. 2011). Adapted from Willett et al., 2019.

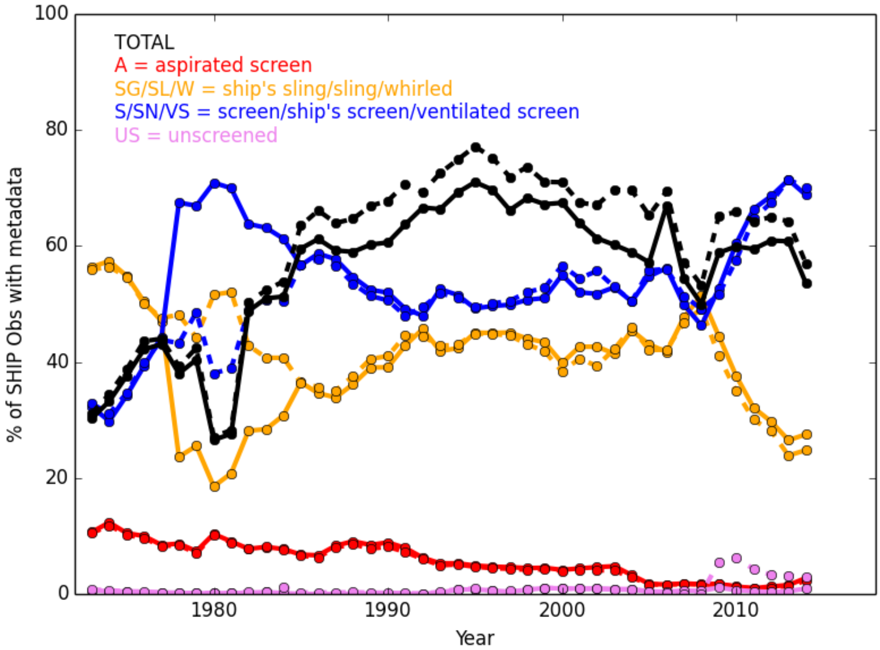


Figure 2 Availability of instrument exposure information (black) for ships (platform (PT) = 0, 1, 2, 3, 4, 5) for the hygrometer (EOH, SOLID) and thermometer (EOT, DASHED) for each year. All ICOADS 3.0.0/3.0.1 observations passing 3rd iteration quality control are included. The percentage of EOHs/EOTs in each exposure category is also shown. Aspirated (A) screens are shown in red. Handheld instruments (ship's sling [SG], sling [SL], whirling [W]) are shown in orange. Unaspirated/unventilated screens (S) and ship's screens (SN) are shown in blue. Additionally, ventilated screens (VS) are also shown in blue as these are generally not artificially aspirated. Unscreened (US observations are shown in violet.

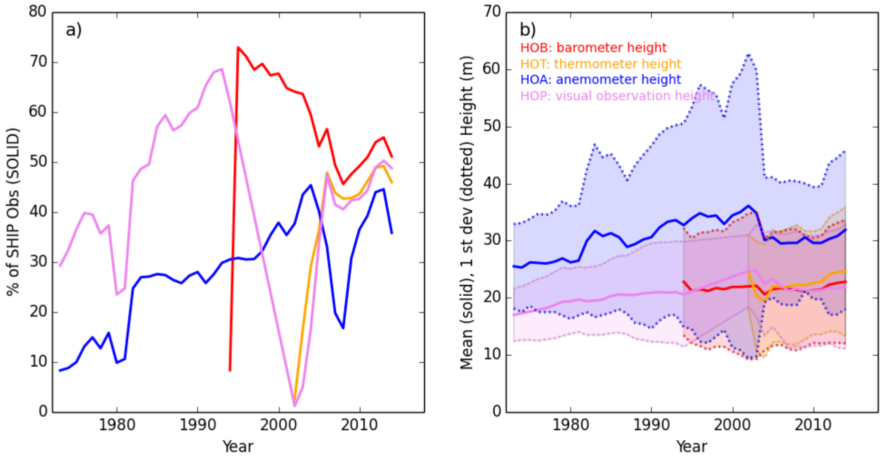


Figure 3 a) Availability of instrument height information for ships (platform (PT) = 0, 1, 2, 3, 4, 5) for the barometer (HOB), thermometer (HOT), anemometer (HOA) and visual observing platform (HOP) with b) mean heights (solid lines) and standard deviations (dotted lines) for each year. All ICOADS 3.0.0/3.0.1 observations passing 3rd iteration quality control are included.

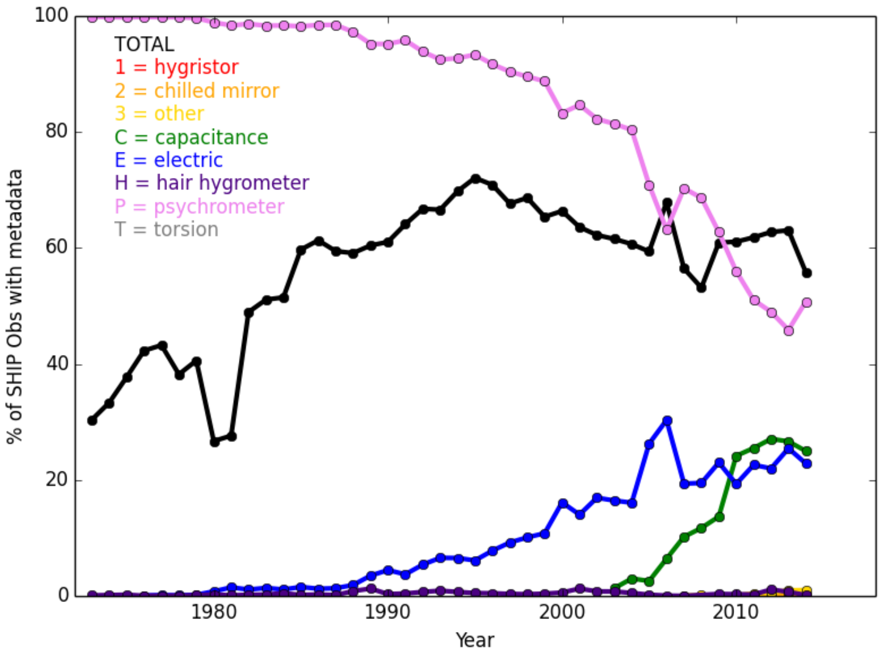


Figure 4 Availability of instrument type information (black) for ships (platform (PT) = 0, 1, 2, 3, 4, 5) for the hygrometer (TOH) for each year. All ICOADS 3.0.0/3.0.1 observations passing 3rd iteration quality control are included. The percentage of TOHs in each type category is also shown.

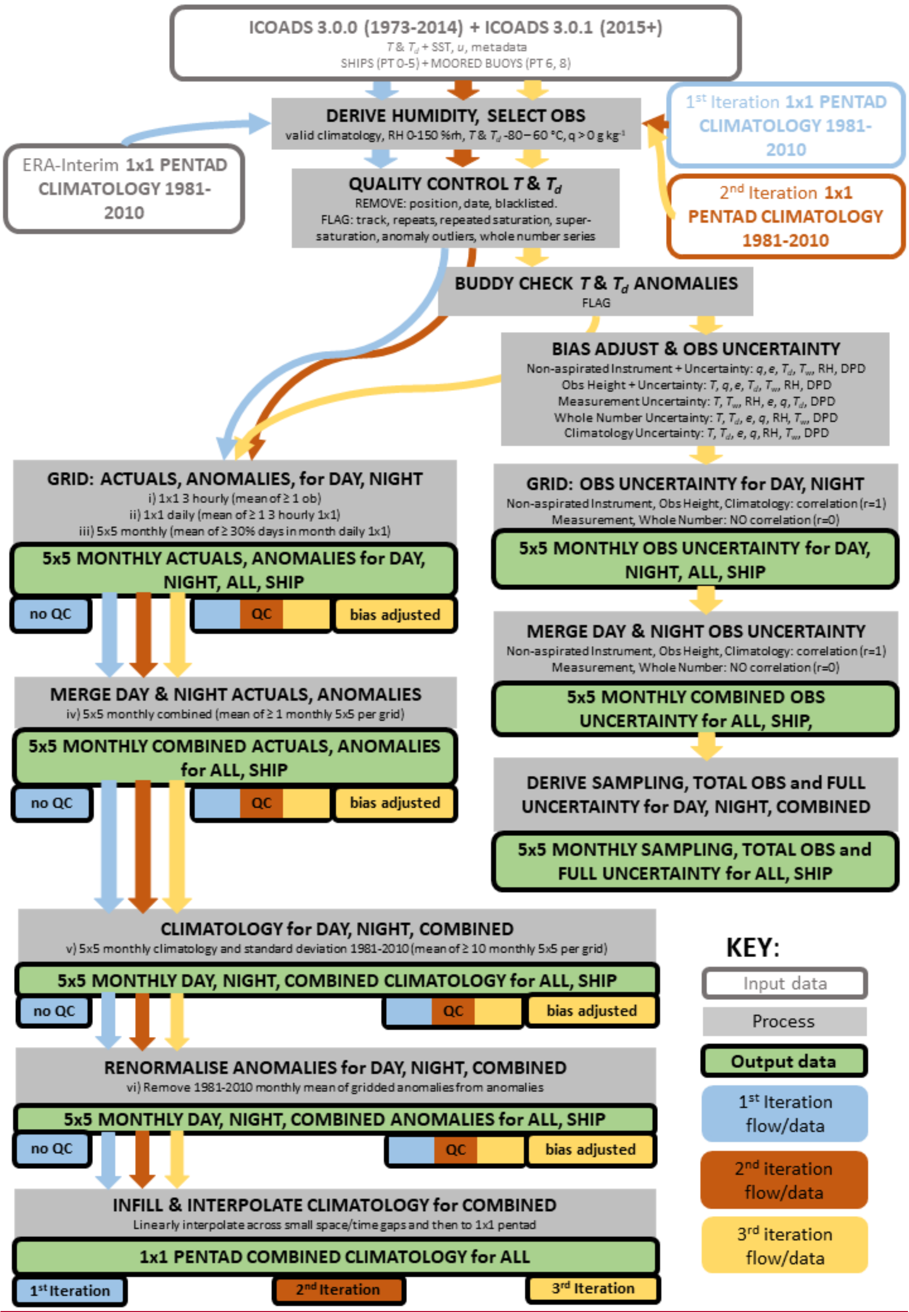
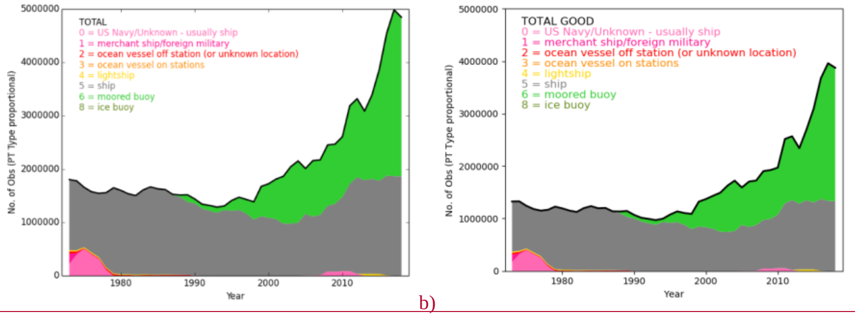


Figure 5 Flow chart of the build process from raw hourly observations to gridded fields. Note that the blue 'no QC' output boxes are produced during the 1st iteration by selecting all data rather than those passing quality control.



a) b)

Figure 65 Annotual observation count for the initial selection (a) and only those observations passing the final 3rd iteration quality control (b), broken down by platform (PT) type.

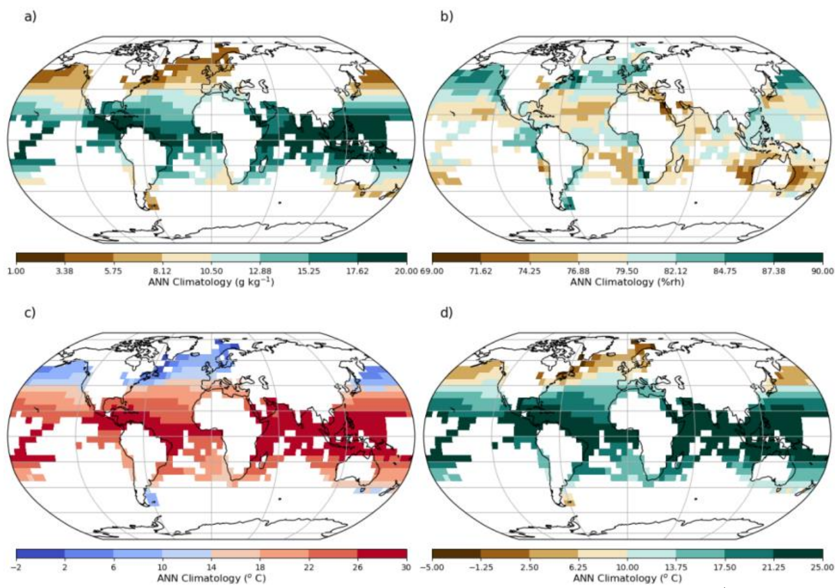
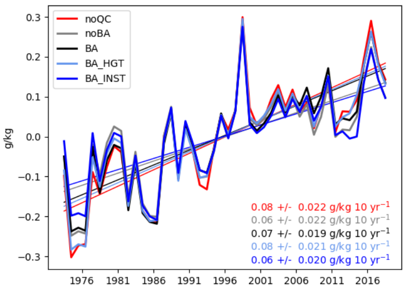
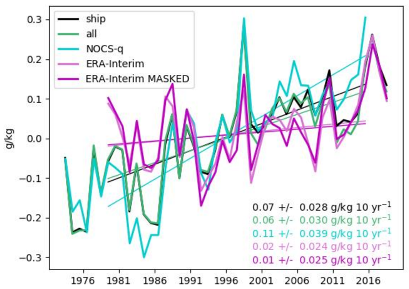


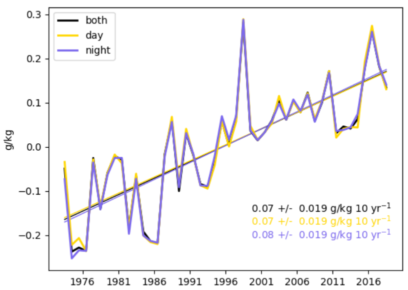
Figure 7 Annual mean climatologies relative to 1981-2010 for a) specific humidity (g kg^{-1}), b) relative humidity (%rh), c) air temperature (° C) and d) dew point temperature (° C) for 3rd iteration quality-controlled and bias-adjusted ship version. Climatological means are calculated for gridboxes 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.



a)

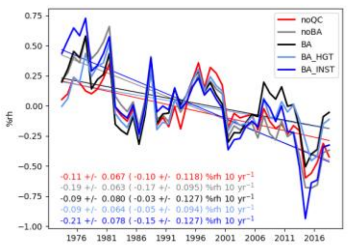


b)

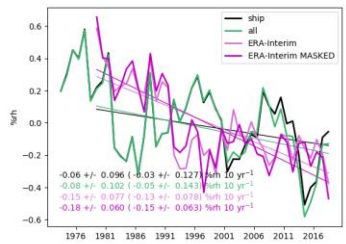


c)

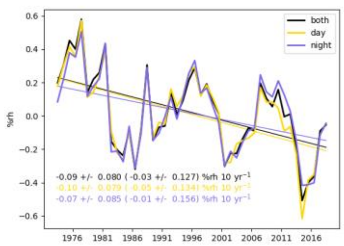
Figure 8 Global ~~(70 °S to 70 °N)~~ annual average anomaly timeseries and decadal trends (+/- 90~~0~~ % confidence interval) for specific humidity. a) Processing comparison for ships only: raw data (noQC), 3rd iteration quality-controlled with no bias adjustment (no~~N~~BA~~C~~), 3rd iteration quality-controlled and bias-adjusted (BA~~Clocal~~), 3rd iteration quality-controlled and bias-adjusted for ship height only (BA_~~Clocal~~HGT), 3rd iteration quality-controlled and bias-adjusted for instrument ventilation only (BA_~~Clocal~~INST). b) Platform and alternative product comparison: 3rd iteration quality-controlled and bias-adjusted ship-only (ship), 3rd iteration quality-controlled and bias-adjusted for ships and moored buoys (all), NOCSv2.0 in-situ quality-controlled and bias-adjusted product based on ships only (NOCS-q), ERA-Interim reanalysis 2m fields using complete ocean coverage at the 1° by 1° scale (ERA-Interim), ERA-Interim reanalyses 2m fields using complete ocean coverage at the 1° by 1° scale and masked to HadISDH.marine spatio-temporal coverage (ERA-Interim MASKED). Trends cover the common 1979 to 2015 period. 1979 to 2018 trends for ERA-Interim are 0.03~~4~~ ± 0.028~~98~~ and 0.03~~29~~ ± 0.027~~098~~ for the full and masked versions respectively. c) Time of observation comparison for 3rd iteration quality-controlled and bias-adjusted ship-only: all times (both), daytime hours only (day), night time hours only (night). Linear trends were fitted using ordinary least squares regression with AR(1) correction applied when calculating confidence intervals (Santer et al., 2008)~~the median of pairwise slopes~~.



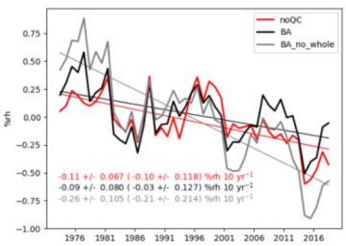
a)



b)

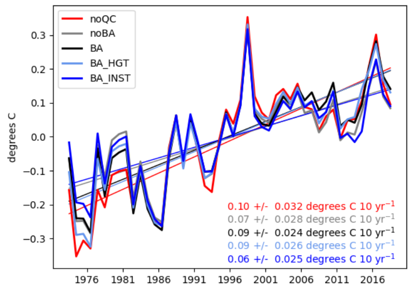


c)

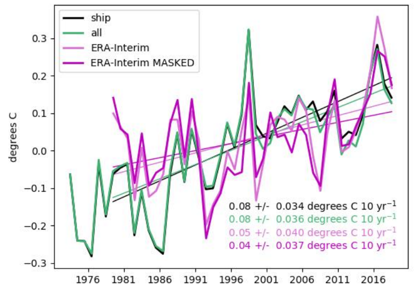


d)

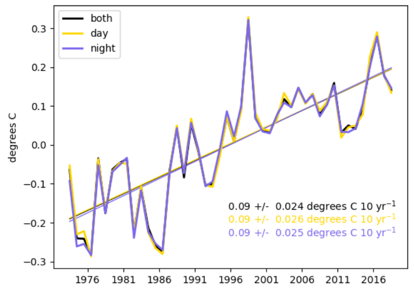
Figure 9 Global ~~(70 °S to 70 °N)~~ annual average anomaly timeseries and decadal trends (+/- 90~~0~~ % confidence interval) for relative humidity. See Figure 8 caption for details. In addition, panel d) shows the timeseries from the bias adjusted data with removal of any data with a whole number flag set (BA_no_whole). Trends in b) cover the common 1979 to 2018 period and all trends in parentheses cover the 1982-2018 period.



a)

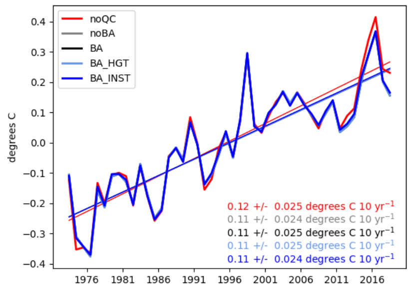


b)

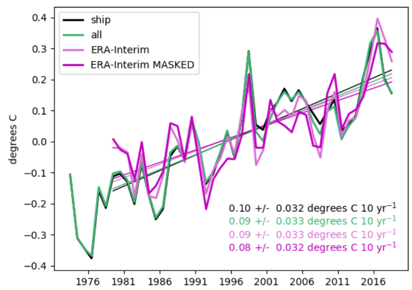


c)

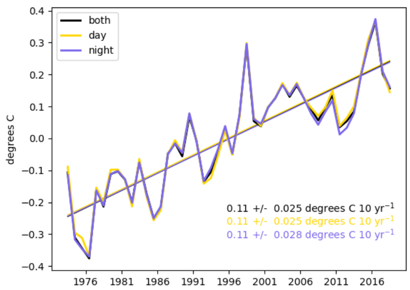
Figure 10 Global (70 °S to 70 °N) annual average anomaly timeseries and decadal trends (+/- 900 % confidence interval) for dew point temperature. See Figure 8 caption for details. Trends in b) cover the common 1979 to 2018 period.



a)



b)



c)

Figure 11 Global ~~(70 °S to 70 °N)~~ annual average anomaly timeseries and decadal trends (+/- 9~~0~~0 % confidence interval) for marine air temperature. See Figure 8 caption for details. Trends in b) cover the common 1979 to 2018 period.

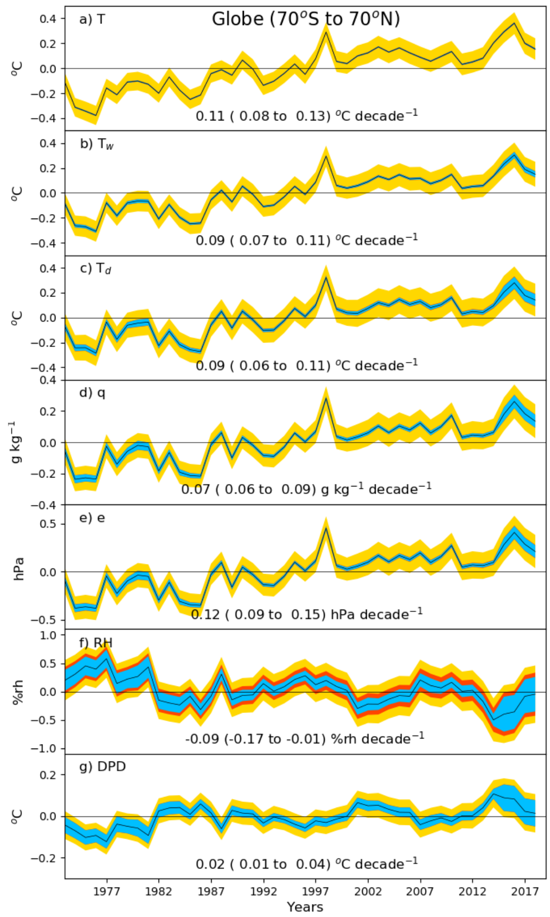


Figure 12 Global average timeseries of annual mean climate anomalies for all variables. The 2 sigma uncertainty ranges for total observation (blue), sampling (red) and coverage (gold) uncertainty contributions combined are shown. All series have been given a zero mean over the common 1981-2010 period. Decadal linear trends and 90th percentile confidence intervals (in parentheses) were fitted using ordinary least squares regression with AR(1) correction applied when calculating the confidence intervals (Santer et al., 2008).Linear trends were fitted using the median of pairwise slopes with the range representing the 900 % confidence interval in the trend.

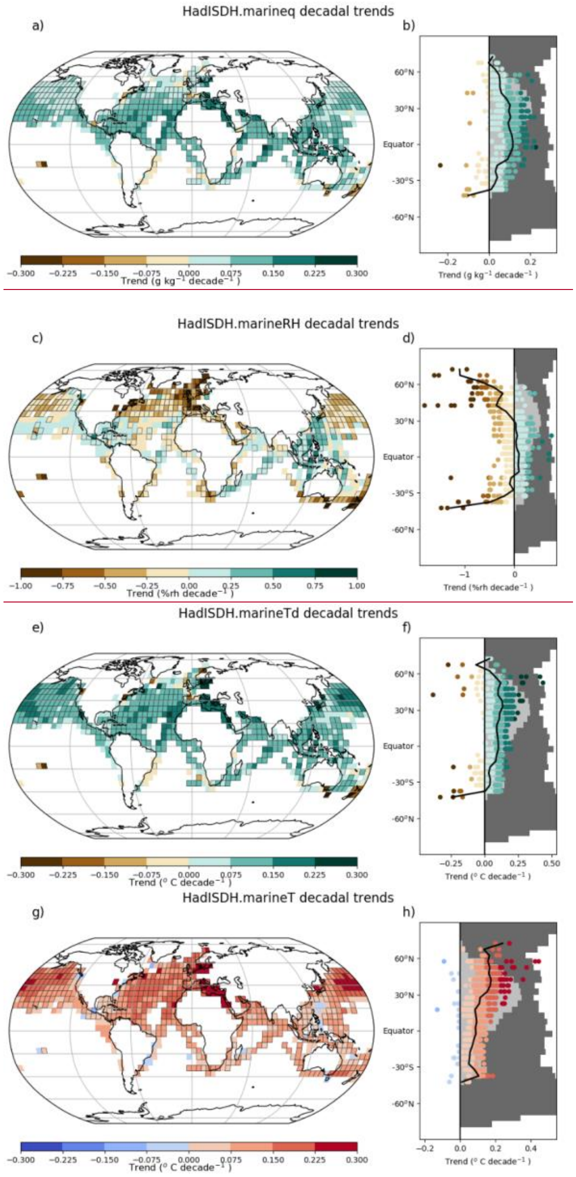


Figure 13 Linear decadal trends from 1973 to 2018 for a, b) specific humidity (g kg^{-1}), c, d) relative humidity (%rh), e, f) dew point temperature (° C) and g, h) air temperature (° C) for the 3rd iteration quality-controlled and bias-adjusted ships only. Decadal linear trends were fitted using ordinary least squares regression~~Trends have been fitted using the median of pairwise slopes~~ when there are at least 70 % percent of months present over the trend period. Gridboxes with boundaries show significant trends in that the 90 % confidence interval (calculated with AR(1) correction following Santer et al., 2008) around the trend magnitude is the same sign as the trend and does not encompass zero. The right-hand panels (b, d, f, h) show the distribution of gridbox trends by latitude with the mean shown as a solid black line. The dark grey shading shows the proportion of the globe at that latitude which is ocean. The light grey shading shows the proportion of the globe that contains observations.