

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	1 st and 2 nd Iteration	3 rd 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 T_d 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 $u = (xH_{max} - xH_{min})/2$. 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 - correct?

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/q_s but these are clearly not equivalent. Clarify if you use e/es or q/q_s.

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/q_s 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 $\mu = 16 \text{ m} + \text{the linear trend in mean HOP/HOB/HOT height to the date of observation}$, $\sigma = 4.6 \text{ m} + \text{the linear trend in standard deviation HOP/HOB/HOT height to the date of observation}$ "

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 show 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 (-0.27 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.