

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 **Hadley** Centre led Integrated **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 **Hadley** Centre and **Climatic Research Unit** Humidity 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 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.”

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 there 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 U_{gb} woven into the uncertainty estimate?

We think that the text was not clear that U_{gb} is just a generic variable name when in fact the equation given is used for each of the five uncertainty sources (U_i , U_m , U_w , U_c and U_h). 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.