Reply to RC3

Thank you very much for your reviewing our manuscript and providing us with valuable comments and suggestions. We reprocessed the IMS-100-GDP because we found some minor issues in the processing data used in the previous manuscript. Furthermore, we applied the updated screening of data following your review comment. The number of samples has increased from 57 to 59. However, the main message of the manuscript has unchanged.
Hereafter, the referee's comments are numbered as Cx and Rx represents the reply to Cx.

C1) [line 18] It would be useful to state the weight of the RS-92 sonde so that the relevant lightness of the RS-11G sonde (85g) is clear.
R1) Added.

C2) [line 31] expand first use of 'LC'
R2) Done.

C3) [line 56] change 'ant' for 'and'
R3) Corrected.

C4) [line 80] and Table 2/3 – tables 2 and 3 identifies the uncertainty sources in the temperature and RH products and these include identification of correlated and uncorrelated sources – but this is the only reference I could see in the paper to this separation of uncertainty source classes. Discussion should be included in the main text about this separation and the implication for the expected level of agreement in the intercomparison of multiple flights.
R4) In the revised manuscript, a brief explanation is added for the classification of correlated and uncorrelated uncertainties with examples.

C5) [lines 81 – 101] The section on the RH correction could be clearer in terms of the processing steps made as well as the justification (and robustness) of the assumed values of the key parameters. So, for example three parameter values are given without source information [lines 91-92] and then only two are apparently used in the calculation [lines 93-95].
R5) The description of RH correction processing steps is detailed in the revised manuscript.

C6) [lines 130-132] The parameters in the text are not consistent with the parameters given in Figure 7 and Tpend is undefined.
R6) The explanation of parameters is added to the revised manuscript.

C7) [line 165] missing comma after ‘plastic’
R7) We meant “corrugated plastic board” by “plastic cardboard” (i.e., “plastic cardboard” is a one single material name, not two materials). Revised.

C8) [line 195] according to Fig 9 the RH for iMS-100 does decrease in the stratosphere, just not as rapidly as for the RS-92 results.
R8) For RS92, there is little need to consider potential freezing of its humidity sensor since it is heated. On the other hand, the very slow RH decreasing after passage through supercooled layers (red shaded in the figure) and (relatively) high RH values in the stratosphere for iMS-100 are not explained by the hysteresis, but probably due to icing. The sentences are revised as: “For RS92, there is little need to consider potential freezing of its humidity sensor since it is heated. On the other hand, the very slow RH decreasing after passage through a supercooled layer (red shaded in Fig. 10 (b)) and relatively high RH values in the stratosphere for IMS-100-GDP are not explained by the hysteresis, but probably due to contamination and changes in the RH sensor specifications related to icing or freezing during passage through supercooled droplet clouds.”

C9) [line 205] should ‘logistic regression’ be ‘logarithmic regression’?
R9) No, ‘logistic regression’ , which is used for the probability prediction. The reference is added to the revised manuscript.

C10) [line 207] more explanation of the source of this expression would be useful, for example the reason for the altitudes of ST1 and ST2 and why RH and VMR values are both included.
R10) To determine the variables for logistic regression analysis, the variable selection from ISSR parameters and RH and VMR values at several altitudes with decision tree was tested. But describing the detail of the variable selection is not the main point of this article, so only a brief explanation is added. The sentence is revised as: “For screening of ice-contaminated profiles, the probability of icing, Pr_{ice}, is derived using logistic regression analysis after variable selection from the length of ISSR, RH and the volume mixing ratio at several levels.”

C11) [lines 211 – 220] I had a number of questions over the uncertainty screening section. Firstly, the purpose of the uncertainty screening should be clarified in terms of what data issues it aims to address - as the GDP processing aims to provide detailed uncertainty information this information should, in itself, define how many points would lie outside the uncertainty bounds for a given confidence limit. This is fully expected within a data distribution and outliers should not be eliminated on this basis, particularly if a later assessment looks at the level of agreement between two data sets. Secondly, it is

stated that the coefficients are determined empirically based on a 90% criterion – was this done of the complete data set (and so, by default excludes 10% of the data)? Finally, since the uncertainties for T and RH are significantly different for day-time and night-time flights it would seem sensible for separate screening criteria to be used for the two cases.

R11)

1) The screening described in Sect. 4.1 is based on the idea that data with uncertainties exceeding the criteria (thus, the uncertainty is the outlier) are of questionable reliability and need to be verified individually (so they are classified as "checked" for RS92-GDP.2). In the revised manuscript, a figure for the case which is excluded by screening by LC but adopted in this study is added.
2) The data set used to determine the threshold formula for IMS-100-GDP was not the data set used for the dual-flight intercomparison but for routine (i.e., single) observations from April to November 2018. The distribution of the uncertainties was tabulated per appropriate interval of values (T, RH, and P), and the envelope of data that would not be an outlier was determined by regression. The result is Equation 18. The "ratio criteria is set to 90%" means that if more than 90% of the whole profile has data whose uncertainty does not exceed the threshold, the IMS-100-GDP is adopted. Some examples are shown in the revised manuscript.
3) As you pointed out, the temperature uncertainties for daytime observations become larger than nighttime due to radiation correction. However, since most of the uncertainties, for both daytime, and nighttime, are within the thresholds indicated in Eq. 18, we do not consider it necessary to distinguish between them. On the other hand, we have reconsidered your point about RH uncertainties to use the respective discrimination formulas for daytime and nighttime. As a result, two cases in the daytime were added to the verification.

C12) [line 231] some parameters in this equation are undefines (REGSEE and s)

R12) RESGEE is expanded and revised equation (change s to t).

C13) [line 237] is this criteria correct, or should it say that profiles with more than 10% of abnormal points are excluded?

R13) The latter is what we meant. The sentence is revised as: “Profiles with > 10 % of abnormal data points …”

C14) [line 239] what was the criteria for ‘abnormal wind data’ in this context?

R14) No screening for wind was done (for both GDPs).

C15) [line 251] should this be the sum over j from 1 to M (as in eq 18) ?

R15) Corrected.

C16) [line 255] ‘MVD’ is undefined.
R16) Expanded expression for MVD (mean vector difference) is added to the revised manuscript.

C17) [lines 297-298] The evaluation of ‘consistent’ and ‘in agreement’ should take into account the effect of the uncertainty screening and correlated/uncorrelated uncertainties (see previous comments)
R17) In this study, the ratio of consistency check results are used for simultaneous data using the total uncertainties of GDPs. Thus, a brief discussion of the factors involved in uncertainty increases/decreases is added to the revised manuscript.

C18) [line 309-310] clarify what is meant by ‘differences...are small’ – point by point differences or systematic differences over multiple flights? Small compared to expected difference given uncertainties (which would imply uncertainties are over-estimated) or consistent with expected uncertainties? And similarly for ‘seasonal variations are large’, with possible reference to the earlier point about correlated and uncorrelated uncertainties.
R18) For line 309–310, the systematic differences are discussed, not considering consistency.

C19) [line 322] same comment as for lines 297-298
R19) As mentioned in A17, only factors involved in uncertainty increases/decreases are discussed.

R20) [line 343] is this the standard or expanded uncertainty? Also, the uncertainty on both barometers should be considered when comparing the results.
R20) This is the standard uncertainty. The phrase “for k=1” is added.

C21) [line 365] what does ‘small enough’ actually mean ? See previous comments on evaluating consistency.
R21) The result and discussion for wind speed and direction consistency are added in the revised manuscript.

C22) [line 376] see previous comments on evaluating consistency.
A22) A brief discussion about consistency with the SKYDEW measurements is added to the revised manuscript. But please note that the uncertainty for the SKYDEW has not been fully evaluated and thus not calculated here.

C23) [lines 378 – 405] the summary should be updated based on the points raised above.

R23) The discussion about consistency for wind is added.

C24) [line 412] GRUAN data product (rather than processing)
R24) Corrected.

C25) [Figure 3] as SEA is an acronym should it be capitalised?
R25) Corrected.

C26) [Figure 7] wonder if the first box should be modified as current content implies that the wind info is derived from a single lat and long value?
R26) “lat0” and “lon0” mean the set of initial latitude and longitude values, not a single value. The main text is revised.