Dear Reviewer,

We thank you for your comments and suggustions to improve the manuscript. We have taken the oppurtunity to add several things, some wanted by you some by other reviewers. The main changes in the manuscript are:

- We changed the title to "**EMADDC: high volume, high quality, and timely wind and temperature observations from aircraft surveillance data (Mode-S EHS)**
- We comparison with AMDAR data to the results section
- We included Vector RMS statistics in the results section
- We explain in more detail how the seperate steps for temperature corrections
- We revised the conclusion and added an outlook section
- Lukas Strauss has been added as co-author to give him the credits for inventing the Mach-Indicated Airspeed improvement.

We hope to have answered all remaining questions as good as possible.

Thank you sincerly for your time and effort, Best regards, Siebren de Haan and Co-authors

**Answers**

Given the importance that observations derived from MODE-S EHS data are taking on, both for operational meteorology and for research, and the amount of algorithm development and data processing techniques that have been necessary to achieve the ability to reliably produce so much useful data, the publication of such an article is fully justified.

The article follows a logical order, provides a synthesis between work already described, which is referenced and placed in the context of the present work, and more recent aspects, such as aircraft-dependent heading correction (§7.1). It ends with insightful characterisation of the data produced versus NWP and radiosondes.

Nevertheless:

the choice of journal remains, in my views, open to question. For example, could Earth System Science Data (ESSD) be a more appropriate choice? In its current form, the article needs to be corrected or reworked before it can become a solid reference for all future work using these data, which it ultimately deserves.

In several places, the article lacks precision of description, and relies on the reader's implicit understanding. This needs to be corrected in a definitive publication. In particular:

ok in §8 Processing Infrastructure, the text let the reader think that the duplicate removal process applies to calculated observations, and not to input (Mode-S and ADS-B) messages. This would means that each data supply channel is treated individually, and that the duplicate removal process is applied at the end, when all the groups of observations produced are merged. Is this really the case? As the article states that "A processing job starts by gathering all data available in the time window of interest." (line 212), it seems that there is room for de-duplication of input data (for example, same MODE-S message received by 2 receivers). Is this carried out, or not ? This should be clarified.

ok are the whitelistings described in §5.3 "Output control" and whitelisting at the end of §8 "Processing Infrastructure" (around line 230) the same? If so, it could be better described in one place, and simply referred to in the other one ("observations are within three times standard deviation of the measurement with NWP model equivalents" does not make much sense to me).

ok Also : some processing techniques depend on assumptions (for example, magnetic declination tables or the form of corrections for static pressure, Mach number or airspeed). Overall, the final results on the quality of the measurements produced validate the work carried out and the assumptions made, but for this article to give the reader a full understanding of the measurement and processing techniques, quantified indicators should be given for the various stages. For example: are there any aircraft for which minimisation of the cost function for magnetic declination (eq. 18) does not converge? Is so, do the authors have any clues about these aircraft (particularly old or recent, or else)? What is the typical amplitude of the true air speed correction mentioned in §7.3?

*we added a figure on this... around at most a few m/s* In addition to the ongoing research mentioned to develop a more physical method, did the authors try to check that this correction was indeed uncorrelated with a spatial characteristic, or some bias in the model, or else (for example, simply by drawing maps of typical values, or scatter plots, . . . ) ? What is the typical percentage of aircraft that are whitelisted? Is it evolving over time ? Is it possible to learn anything from the list of aircraft that are rejected ? Are they simply aircraft that transmit incorrect data, or are they particular types of aircraft for which other assumptions and calculation methods could be used? This could be an avenue for development if these aircraft fly where others do not.

*we hope to answer these questions thoroughly by the foreseen research*

Here are some more specific remarks, along the text :

ok Line 15 : "For many years, aircraft observations form the backbone of the global observing system". The wording "form the backbone" appears a bit overstated. Associated references support the value and importance of aircraft observation, but they do not assess with certainty that it is central and structuring. Could be rephrased as "aircraft observations are an essential component of the global observation system"

ok Line 17: De Haan, 2013 is an article which deals with Assimilation of GNSS ZTD and radar radial velocity for the benefit of very-short-range regional weather forecast. It recalls the importance of aircraft-based observation, but only marginally demonstrate it. I don't think that its brings much here.

ok Line 33: Since there is a causal relationship, the wording 'However, as (or since) some airlines have continued to fly' seems more appropriate to me than 'However, whilst some airlines have continued to fly' (but I'm not a native English speaker).

ok Line 58-59 : You might want to quote a technical article on these decoding techniques, to give the interested reader the information they need to find out more about these difficulties and the techniques for overcoming them. For example : J. Sun, H. Vû, J. Ellerbroek and J. M. Hoekstra, "pyModeS: Decoding Mode-S Surveillance Data for Open Air Transportation Research," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 7, pp. 2777-2786, July 2020, doi: 10.1109/TITS.2019.2914770.

ok Line 66: missing reference at the end on the sentence "[. . . ] transmitted frequently and could be used in data assimilation (?)." . For example, Bruce Ingleby mentioned such a technique in is poster at the 2023 International Symposium on Data Assimilation (ISDA-2023), titled "ECMWF use of Mode-S winds and changes to aircraft thinning."

ok Line 80: Since reference is made to a personal communication, details of the calculation should be given in the present text.

ok Line 83 : "the timestamp is supplied by the receiver and not by the aircraft" is not the root cause of the difference between receiver and radar data, since "The timestamp is created at the moment of arrival of the information", as stated in §5.1. The advantage of the radars probably rather comes from the synchronisation between the positioning, carried out when the echo from the aircraft is received, and the reception of the Mode-S message, practically simultaneously

ok Line 95, eq.1 : it could be worth noting that the numerical constant used here are valid for dry air, and later, for example in in §7.3, consider the possibility of controlling the applied correction in areas known to be particularly humid (boundary layer in the Mediterranean or the Canaries)

ok Line 104 : given the prior presentation of the difference between receiver and radar data, and even if I agree with the need to choose "The (most relevant) parameters", I would expect table 1 to contain two different lines for positioning : (latitude-longitude) from ADS-B at 0.5-2 seconds period, and (range-azimuth) from radar at 5s - 20s period. Or EMADCC never uses radar positioning and mixes radar and ADS-B receiver to assign a position to an observation ? Also, in table 1, the headings of the 'frequency' and 'reported accuracy' columns are reversed.

ok Line 112: "check the input for obvious errors," could be completed by "or measurements in conditions where calculation is not possible".

ok Line 146, eq.6 : Even if their meanings can be guessed, I believe that "V" and "d" have not been formally defined before, and it would be more rigorous to do so. Please also recall quickly the hypothesis behind this formula (this is a 2D formula, not valid at large values of roll and pitch angles, which justifies the criteria roll <2.5% in table 2, and this formula assumes that the airspeed is aligned with the axis of the aircraft (the heading), and therefore that sideslip is zero, which is mostly true for airliners, but not necessarily during the aircraft's rapid manoeuvring phases)

ok Line 207 : after the sentence beginning with "For receiver data. . . ", the reader expects another one describing what is done for radar/tracker data. Or does this sentence apply to both receiver and radar data ?

ok Line 241 and further ; Although I know that this is common practice, the choice of the word ‘error’ to designate the difference between an observation and a model analysis is, to say the least, debatable. There are cases where the RMS time series of these deviations have changed significantly, without any change in the observation system, but when the model version was updated. Especially since the article later shows that “the comparison between radiosonde and Mode-S EHS show to have a standard deviation lower than that of the comparison is model and Mode-S EHS or radiosonde”, which suggests that part of the variance in the MODE-S/model difference is due to a discrepancy between the model and the reality.

ok Figure 4 legend : an “i” is missing in “Observatons”

ok Line 262 : “858” is probably a typo for “850”

ok Line 271 : I don’t fully understand the sentence “Note that, although the data is corrected using ECMWF forecast, the data is independent because a forecast lead time of minimal 9 hours is used”. Is the forecast lead time of 9 hours used for the computation of the magnetic declination table, or the True airspeed correction mentioned in §7.3? Wouldn’t this sentence be better placed closer to the correction description ? Does it imply that the impact of Mode-S assimilation in the model forecast does not extend beyond 9 hours?
*Added words in heading correction section "Derived wind measurements"*