**Response to Reviewer 2**

I thank Zachary for his detailed comments. Replies added in blue. Line numbers refer to the preprint.

Major

1) From Fig. 2 the filtered observations match a Laplacian shape to an impressive level, though I don't see such a match to any presented model for the autocorrelation (Fig. 3) or power spectrum (Fig. 4). There are statements about some of the models being more suitable than others here, but to me their level of bias appears similar. I'd like to lightly challenge Bernd to consider if the assessed measurements raise the need for a more suitable model.

Along these lines, the current models follow fits for the autocorrelation function (triangular or exponential), but it does not appear to have been noticed that the data suggests the power spectra to be more fundamental. As I show below (noting I was not able to perfectly replicate the data), assuming white noise up to frequency N (with zero at faster velocities) produces an autocorrelation spectrum that has sinusoidal elements and is far more similar to the SPB obs than either the triangular or exponential function. Could something like this be translated into a more suitable analytical model? And I am wondering if the manuscript's current interpretation of the wave elements in the autocorrelation function, which cites trapping at wave duct boundaries, is perhaps more complicated than the actual reason.

You are addressing a good point here.

Let me just briefly recap: The KP19 time series consists of a series of boxcars of width t_auto and with random amplitudes. Its autocorrelation is triangular. Because successive amplitudes are independent, the correlation drops to zero outside the triangle. Its power spectrum is the squared sinc function Eq.(3). Because the amplitudes in the time series have sharp boundaries, the power spectrum extends beyond the buoyancy frequency N (although with little power). Due to the low pass filtering of the original measurements, the power spectrum of the N-filtered data contains fewer frequency components above N than KP19. The filtering also causes the wavy pattern seen in the SPB time series (inset of Fig.1), which is smoother than the KP19 time series.

As you show, a power spectrum that is perfectly flat up to N corresponds to a sinc autocorrelation function with alternating positive and negative side lobes. You are right that this looks similar to the autocorrelation function of the balloon (SPB) data, which is no surprise given the data filtering. It doesn't appear to be more suitable, though. I agree that discussing the SPB autocorrelation behavior in terms of wave trapping may be an overinterpretation. The effect of low pass filtering the balloon data is discussed in the context of the power spectrum (lines 119-124), but not in the context of the autocorrelation function. I agree that the revision should catch up on this.

Because the true SPB autocorrelation function and power spectrum is not known due to the filtering, the simple-to-implement Laplacian colored noise model KP19 performs so well in the nucleation simulations, and the band-limited white noise autocorrelation function is not better than that from the other models, am skeptical about the need of another time series model. I'd rather think that better understanding the spectral behavior of real-world data near the buoyancy limit (role/ importance of wave trapping) and the associated increase in power is more relevant at this point.

I introduced the following changes in the revised manuscript.

Lines 93-97 containing the speculative link to wave trapping is removed, now reading: "This comparison suggests that the damped periodic sequence in the autocorrelation function of the SPB measurements may result from low pass filtering at N (section 2.1.1)." The revised Fig.3 additionally shows the sinc-autocorrelation function of white noise cut-off at N, and the rest of the text in this section has been revised to accommodate all changes, including a modification of line 121f in section 2.1.4: "… deviating from an idealized flat distribution extending up to N and corresponding to the sinc-autocorrelation function shown in Fig.3." Right before section 2.2 starts, a sentence is added: "Because the true autocorrelation function and power spectrum of vertical wind speed fluctuations at high frequencies are not known due to the filtering, an N-filtered white noise

source—identified as an alternative to the wave-cirrus parameterizations KP19 and FL21 discussed here—is not necessarily more realistic. At the very end of section 2.3.1, a sentence is added: “Because the amplitudes in the KP19 time series have sharp boundaries, the PSD extends beyond N.”.

2) The majority of the manuscript is in the “Data and Methods” section (Sec. 2), while the current results section (Sec. 3) is quite short. I think it would be easier to follow the flow of the article if there were, for example, a shorter section introducing the methods, then a first major results section on how the SPB measurements compare to statistical models, and then a second major results section on the implied ice nucleation distributions. I’m assuming that most of the material on the SPB statistics is novel enough to warrant presentation as a result. There is currently a daunting number of subsections in Sec. 2 and I think generally the organization of the manuscript can be improved.

Thank you for suggesting an alternative paper structure. I think the current structure, though a bit textbook-like, is meaningful and appropriate given the many individual components covered (two models, three variables, measurement data).

3) I like the concept of linking the statistics of the vertical wind observations to those of available models, though am a bit concerned the ice nucleation section (Sec. 3) is very similar to analyses already done in Jensen et al (2026), especially that Fig. 6 here is extremely similar to

Fig. 8 of J26. I'd like expressed in the manuscript why this analysis warrants a new iteration, including whether the use of DICE here is an improvement versus CARMA in the previous study. I appreciate that there is benefit of having the statistics (Sec. 2) and the ice nucleation outcomes from such representations (Sec. 3) together in one study, so I wonder if the latter (currently quite short) can be a bit expanded to link better to the preceding sections.

The present study has a different focus than the study of Jensen et al. (2026) (lines 50-53). The ice nucleation simulations are central inasmuch as they serve as a very stringent test of the WCPs because of the strong sensitivity of homogeneous freezing to vertical winds. It came quite handy that Jensen et al. (2026) performed such simulations, because this made me think about introducing a mini model intercomparison (outcome summarized in lines 334-336). At any rate, the present work is not about judging whether one code performs better than the other.

4) There's a stated aim that this study "offers pointers" for modelers, but being myself a modeler I'd appreciate some more concrete info and guidance. For instance: How advanced are the current approaches in standard GCMs compared to what is shown here? Do the Laplacian forms here represent a substantial path toward improvement over simpler Gaussian representations in models? If so, how exactly would models ideally be altered? Conversely, could it be worth keeping Gaussian (or other simple) representations but

tuning these to the SPB obs? Is there sufficient confidence that scaling distributions relative to Strateole-2 measurements are representative of cirrus in other regions and at other altitudes (noting a GCM is unlikely to have multiple cirrus schemes at once), or does there continue to be a large gap even if recommendations here were to be followed?

Please note that I have stated in lines 47+48: “and thus provides guidance for future developments of more advanced WCPs.” Nonetheless, I agree to include the GCM standpoint, as it is mentioned in the Conclusions section and upon request of reviewer #1.

The preprint states four recommendations in section 4.2:
- generalize/combine WCPs that work at all distances to gravity wave sources and are thus more source-specific (lines 361-365)
- better understand wave trapping effects on amplitudes and frequency composition of vertical winds (lines 366-374)
- capturing effects of small-scale turbulence that is often associated with GWs (lines 374-378)
- analyze the current state regarding high resolution atmospheric GCMs (lines 379-383)

There is a section on “Implications” but it doesn’t address these. Also, coarse 100km-scale GCMs are gradually becoming dated, so it could be worth briefly expressing how implications translate to km-scale models (i.e. global storm-resolving or limited-domain convection-resolving models).

Higher resolution models will capture a greater portion of high frequency gravity waves (though still not fully at km-scales). Therefore, I think the last bullet point above makes sense before embarking on a specific parameterization.

Section 4.2 is renamed to “Discussion and implications” in the revised manuscript.

Regarding more concrete info on:
- current approaches in standard GCMs drive ice nucleation parameterization with tuned turbulence kinetic energy (TKE), i.e., they may get the mean ice crystal number concentrations right by tuning a mean or standard deviation to values associated with gravity waves
- the Laplacian distribution is an important element of the wave forcing
- using TKE and a Gaussian distribution would cirrus formation on the wrong physical process, produce too narrow ice frequency distributions, and ignore intermittency effects
- gravity waves are never not present, so that application of current WCPs outside the tropics and polar regions (where SPB measurements were made) is plausible; one microphysical study applied gravity wave forcing to investigate their impact on midlatitude cirrus (https://doi.org/10.1002/jgrd.50421), however, would be favorable to have direct observational evidence also at midlatitudes

Actually, much of the above has been discussed in Kärcher & Podglajen (2019), so in the revision, it should suffice to address the first point. Adding to line 384:

"Most global climate models drive ice nucleation parameterizations with tuned turbulence kinetic energy. While this approach may get the mean nucleated concentrations right, intermittency effects are not accounted for and associated statistics of ice concentrations are too narrow. In the case of orographic gravity waves, this led to a lack of a representation of lee wave cirrus in global model simulations (Dean et al., 2005)."

5) I tried to replicate Figs. 2-4 but found the paper does not currently have all needed info:

a) Which specific version of the Herzog 2019 dataset is used? (v3b?)

Yes.

b) I don't see w' in the Herzog 2019 files, so how exactly all the w' values calculated? Is any other dataset used in this process?

The w'-values have been calculated from the pressure measurements in the data set cited. This is described e.g., in Jensen et al. (2026).

c) What quality control is applied to get to 695,104 data points from the larger Herzog 2019 files? (the raw data across all flights has about 710,000 points, so maybe small segments are removed?)

Data points obtained during depressurization events and during periods when the balloons were still rising were removed prior to analysis.

d) How are gaps treated, and are all flights simply combined as if a single continuous dataset?

All individual TTL time series were stitched together int one data set.

All missing infos are added to the revised section 2.1.1.

e) Is all needed info on the filtering approach included?

Yes.

f) Is it possible to reproduce the ice nucleation plots (Figs. 5 & 6)? Is DICE publicly available online, and the needed details stated in the manuscript?

No, as now stated in the combined code and data availability statement.

Minor

- Lines 10-11: It is advertised as a significant result that, as the abstract states, “the causes of small discrepancies in ice concentrations are traced back to filtering of the balloon data necessary to remove non-geophysical artifacts and differences”, but this seems to be speculative, I cannot follow the logic, and “small” feels a bit overforgiving. Is it possible for some visual

support to be added to the manuscript?

I believe the small differences in the ice frequency statistics shown in Fig.6 have been adequately discussed (lines 297-318). I can't see the speculation.

- Line 32: The "due mainly to" fragment doesn't logically follow the beginning of the sentence.

The part of the sentence "…, due mainly to limited knowledge of factors controlling their lifecycles, including the formation stage" is deleted.

- Line 37: "Due to their smaller horizontal wavelengths […] high-frequency GWs are closely related to deep convection" – I don't follow a causal effect here.

The part of the sentence "Due to their smaller horizontal wavelengths and larger vertical group velocities, …" is deleted.

- Line 39-41: Is there an explanation for why GWs of relatively high frequencies, but not more than N, are most important? Considering both the speed of an updraft and its duration, has anyone already formulated why this appears to be a 'sweet spot' and a citation can be added? This feels like a quite interesting and relevant problem worth addressing.

The citation Jensen et al. (2026) is added at the end of the sentence in line 39, as the most recent one. The underlying mechanism for the sensitivity of

homogeneously nucleated ice numbers on high frequency oscillations is the strong sensitivity of the freezing process on the cooling rate.

- Line 57: Section 2 feels like part methods and mostly results, and most figures are associated with this section. I'd suggest breaking off the statistical analyses of the SPB observations and their comparisons to simple fittings as a results section, and reorganizing the methods section accordingly.

See reply to major comment above.

- Fig. 1. Can the times shown in the inset be stated for replication purposes?

Unfortunately, I have not taken note of the time window the segment has been taken from. I vaguely remember that it was from the first day, but can't be sure. I could have chosen any time window and avoided picking an instant with a super high fluctuation.

- Lines 77-78: Can it be explained physically why w' closely follows a Laplacian? Also, is this a very specific outcome of the TTL? I would think in much of the tropical troposphere there is asymmetry with stronger updrafts and more extensive weaker downdrafts.

The distribution contains effects of intermittency. It is not specific to the TTL (see e.g., Podglajen et al., 2016). A distribution comprised on long-term measurements like the balloon data should be expected to be symmetric,

contrary to the limited flight legs sampled during an aircraft campaign, where the mean value may show a bias.

- Lines 92-97: I'm wondering, along the lines of my first major comment, if the autocorrelation function for the filtered observations is perhaps more explainable than the speculation here conveys. I think it is important to get this right, because it is clear that the shape is fundamentally neither Laplacian nor triangular and so appears to signify a need for an improved model.

I agree that this deserves a better explanation along the lines of my reply to your above comment. However, because this is the autocorrelation function of the filtered data set, and the true version is not known, a model based on band-listed white noise that better replicates the black curve in Fig.3 does not improve the situation or yield a model that would be fundamentally superior to KP19.

- Line 157: replace "short-term autocorrelation" to "triangular autocorrelation" to link better to the models shown in the preceding sections?

Done.

- Lines 194-195: How does OU30 compare to current GCM approaches? I would think even having a Gaussian distribution with hooks from this to ice nucleation is currently relatively advanced. Also, it might be better for the arguments of the paper to additionally

show a cruder model than this if one is standard in GCMs, in order to point out more effectively what ability for improvement already exists.

I think many GCMs just fix a mean vertical wind speed fluctuation in the TKE framework (Lohmann & Kärcher, 2002) and do not go into any of the details discussed here. I have commented on deficiencies of a Gaussian approach in reply to your comment 4) above. With specific regard to section 2.3.3, it is worth noting that a Gaussian model with exponential autocorrelation will have power spectrum that contains too much power at frequencies above N, similar to the red curve shown in Fig.4 (lines 203+204)

- Line 205: I don't see this parameterization type used anywhere so (if I haven't missed something) am wondering if it warrants its own section. Also, generally it would be beneficial to know which ice nucleation schemes (and which GCMs that use them) are in each basket. And does Kärcher & Lohmann 2002 (and related versions) connect well to any of the categories presented here?

Equation (6) follows from Bartosch (2001). The citation of this study in line 199 is moved to line 201 to make the connection clearer.

- Lines 210-254: The "notes" feel a bit detached from the rest of the manuscript and breaks the flow. I'd suggest either incorporating them more closely or moving them to an appendix.

I like to keep the notes separated, as 2.4 addresses some common aspects of the PSDs from all models and 2.5 looks into the link to temperature fluctuations. Both are to m taste too important to be delegated into appendices.

- Line 268: While vertical winds are certainly a major control, do the simulations make assumptions on other controls worth mentioning? E.g. an assumed temperature, or an initial relative humidity?

The simulations do not make assumptions on controlling factors other than contained in the initial conditions. Temperature follows from the undamped adiabatic relationship justified by the short simulation time per freezing event and relative humidity evolves accordingly additionally accounting for a sink term due to deposition growth.

- Line 273: What is “one ensemble, based on 39,038 data points”? This is many fewer than the 695,104 data points used elsewhere.

A single nucleation simulation does not use all TTL data points, but begins at a random time as described in section 2.6 and ends as described in App. B. To produce robust statistics, DICE is called 80,000 times per ensemble (line 258). About half of the calls end up with persistent warming and wouldn’t quickly lead to a freezing event. Those simulations are discarded, hence 39,038 successfully analyzed nucleation events are

used. The revised text in section 3.1 (lines 270+271) clarifies this point.

- Line 279; isn't median < mean "right" skewed?

To avoid confusion, the sentence in line 279 now reads: "Due to the large number of low updraft speeds in the GW forcing, the media concentration is much smaller than the mean values; …"

- Line 283: 144 cm-3 isn't really "far off" scale given the axes are logarithmic - "off" would suffice.

Done.

- Fig. 6: Why not include OU30 in the inset? And please in answer to the relevant major comment reconcile the similarity to Jensen et al (2026) Fig. 8 with how this new version is warranted.

Note that the Gaussian model is only shown for comparison to judge differences in assumed probability distributions of w' (lines 194+195). The OU30 result is not shown in the inset because it misrepresents the observed w' statistic and should thus not be involved in the discussion of the difference between the model and SPB data-based statistics. See modified sentence in the paragraph right before section 4.

Regarding the similarity to the Jensen et al. study, refer my reply to comment 3) above.

- Line 291: Can it be clarified if these are scaled to match the SPB obs and exactly what parameters are used for this?

They match the filtered SPB data in terms of autocorrelation time and total variance (line 155+156). This information is added to line 270 for clarity in the revised manuscript,

- Line 319: the first Conclusion paragraph seems like it could be combined with the Summary subsection.

While the first paragraph summarized what has been done, the subsequent Summary subsection recalls the results, hence, the separation. Section 4.2 now starts with: “Synthetic time series from stochastic parameterization schemes were employed to drive a microphysical model simulating the formation of cold TTL cirrus. The schemes were constrained by the fluctuation variance and autocorrelation time inferred from the measurement data.”

- Line 347-349: The causes of discrepancies would seem quite important for this work but we seem to be left with speculation, whereas I think these causes might not be so mysterious (c.f. major comment 1, and I’m sure other analyses would be able to pin it down further).

Please see my reply to your comment 1) above. The proposed changes to the manuscript should take care of this issue.

- Line 357: Fig. 3 looks more promising for the the exponential fitting in FL21 than the triangular one of KP19, because – even though the shape is neither exponential nor triangular – it does decline first sharply and then less sharply. And the ice nucleation simulations to me appear to have similar biases with KP19 and FL21. To say KP19 is more supported seems to me a slightly surprising conclusion.

The reason why FL21 is judged less realistic than KP19 is stated in lines 357+358: It produces too much power above N, which is incompatible with GW propagation physics (lines 213+214) and is less straightforward to implement in models (line 186).

- Line 366-373: Can some evidence be shown for the information in this paragraph about trapping and filtering, if there is confidence in the material?

I am not sure what you are referring to. Observational evidence is given in Kottayil et al. (2025). As homogeneously nucleated ice numbers scale with updraft speed, enhance of w' due to trapping increases ice numbers. Trapping signatures are filtered out of the data set because they occur near N. And it is necessary to learn about the detailed frequency signal near N beyond the sheer increase in power due to trapping to reliably model the associated effect. All that is stated here and seems evident given what has been stated before in the manuscript.

- Line 480 – The data availability statement here hasn't been filled in, even though the manuscript file has a separate data availability footnote that mentions Herzog 2019.

See my reply to your comment 5) above. The footnote originally referred to the reference list, and the revised statement now reads: "The Stratéole-2 campaign was sponsored by CNES, CNRS/INSU, and NSF and the measurement data set v3b is available from https://doi.org/10.14768/7396b9ec-582f-4be2-8c37-c34ef41247a7.