

We thank the editor and Anonymous Referee #5 for their time and careful reading of our manuscript, and for their constructive comments, which helped us improve it. Below, the referee's comments are reproduced in *italics*, followed by our responses. Changes to the manuscript are highlighted in blue in the revised version.

- **Comment 1:** *Line 105: if so, I am curious how long is appropriate for robust statistical analysis?*

We thank the reviewer for this question. What matters is not the total duration of the timeseries but more the number of independent cloud-lifetime realizations it can provide.

Unlike nucleation events, which last a few minutes and can be sampled from simulations started every few minutes along a balloon time series (e.g., Jensen et al., 2026), each lifetime simulation needs a continuous forcing time series at least as long as the cloud lifetime itself. In our simulations, lifetimes can reach 40 h or more (Figs. 4, 7). Therefore, simulations started at closely spaced times along the same record would share most of their forcing history and would not clearly be independent. A 241-day record therefore provides only on the order of 100–200 independent realizations, assuming roughly 1–2 days of forcing per realization. This is insufficient for the part of the distribution that matters most here: the long-lived clouds (> 12 h), which represent a small amount of occurrences but contribute disproportionately to the cloud coverage (Table 2).

Indeed, with $N \approx 120$ independent realizations (241 days of balloon data at about 2 days of forcing per realization), a fraction of long-lived clouds of $p \approx 6\%$ (Table 2) would carry a relative uncertainty of $\sqrt{(1-p)/(pN)} \approx 35\%$. Reducing this uncertainty to 10% would require about 1,600 independent realizations, i.e., roughly 9 years of continuous TTL balloon observations.

Two further considerations favor the parameterization: first, the observed record combines different geographical regions and gravity wave regimes (standard deviation of vertical wind varies considerably between 5-day segments; Jensen et al., 2026), so even its full length would mix non-stationary conditions. Second, the parameterization allows us to set σ_w for a given regime, as in our Geo-ocean and Geo-continent scenarios.

We have expanded the sentence accordingly.

- **Comment 2:** *Is it possible that the GW parameterization includes other lower frequency waves?*

We thank the reviewer for this question, which was also raised by Anonymous Referee 4. The stochastic parameterization of Kärcher and Podglajen (2019) represents the gravity wave background between the inertial frequency f and the Brunt–Väisälä frequency N . Within this range, all intrinsic frequencies contribute equally to the vertical wind power spectrum. The resulting spectrum was computed and compared with Stratéole-2 superpressure balloon observations in Jensen et al. (2026, their Fig. 7).

Lower-frequency waves, with periods of several days, are also effectively excluded as KP19 includes a low-frequency damping that bounds the spectrum near f and keeps the temperature variance consistent with balloon observations. Without it, we would expect low frequencies to dominate T' .

We have added a description of the KP19 spectrum to Section 2.2 and a reference to Jensen et al. (2026, Fig. 7).

Jensen, E. J., Corcos, M., and Kärcher, B.: Influence of variations in gravity wave properties on cirrus ice number concentrations produced by homogeneous freezing in the tropical tropopause layer, *J. Geophys. Res. Atmos.*, 131, e2025JD044467, <https://doi.org/10.1029/2025JD044467>, 2026.