***Response to Reviewer 1:***

We thank the reviewer for the careful reading of the manuscript and for the constructive comments. We have revised the manuscript substantially, with particular attention to the clarity of the Young inversion methodology, the definition of the statistical sampling, and the interpretation of the Young-optimized subset. The methodology section has been reorganized into separate subsections describing lidar preprocessing, calibration and crosstalk correction, cloud detection, the Young inversion, and the selection of the Young-compatible subset. We have also clarified that the independent statistical unit is the detected cloud layer, whereas the points displayed in the two-dimensional histograms represent optically weighted samples within those layers and are not independent realizations. In addition, we have revised the discussion of the ice PSC and cirrus Young-optimized subsets, explicitly noting their limited statistical representativeness. A revised Table 1 now reports both the number of cloud layers and the fraction retained by the Young-optimization criteria.

*Ultimately, from the 144 measurement days that spanned the 3-y period, there were less than 400 individual clouds that were analyzed (Table 1). And line 219 states that "each detected cloud layer represents one independent realization for the lidar-ratio retrieval." However, there seems to be many, many more points that are displayed in the 2-d histograms in the figures. I can only conclude they are showing individual profiles that were measured in each of these clouds in those histograms, which brings up questions about the temporal correlations within a cloud (i.e., these are not independent samples). The authors need to be more clear about this sampling, and the impact of possible temporal correlation on their analyses.*

We thank both reviewers for raising this important point. We agree that the distinction between independent cloud-layer retrievals and the much larger number of points displayed in the two-dimensional histograms was not sufficiently clear in the original manuscript.

In our analysis, the independent retrieval unit is the detected cloud layer. Each cloud layer is defined as one contiguous cloudy altitude interval detected in a single lidar profile, and the Young inversion assigns one layer-mean LR to that layer. Therefore, the number of independent LR retrievals is the number of cloud layers reported in Table 1.

The points shown in the two-dimensional histograms are not independent LR retrievals. They correspond to range-time samples within the detected cloud layers. These samples are used to display where the optically dominant portions of the cloud population lie in the $(1-1/R, \delta_T)$ phase space and to compute optically weighted phase-space averages. Since all samples belonging to the same cloud layer share the same retrieved layer-mean LR, temporal and vertical correlations within a cloud do not increase the number of independent LR retrievals.

We have revised Sect. 4 and the caption of Table 1 to clarify this point explicitly. We now state that the two-dimensional histograms should be interpreted as optically weighted phase-space representations of the cloud population, not as collections of independent observations. Consequently, the statistical robustness of the LR distributions is governed primarily by the number of independent cloud layers, whereas the range-time samples only determine how each cloud is distributed in optical phase space. Moreover, we have added to the subsection "Statistical Analysis" the following text (also in response to Reviewer 3)

The dataset spans 144 measurement days in 2022, 90 in 2023, and 112 in 2024.

However, the statistical analyses presented in this work are based on individual cloud layers identified within each measurement, rather than on the number of measurement days. Here, a cloud layer is defined as a contiguous altitude interval classified as cloudy within one lidar profile after the cloud-boundary detection procedure. Each detected cloud layer is assigned a single layer-mean LR by the Young inversion and therefore represents one independent LR retrieval. The larger number of points displayed in the two-dimensional phase-space figures corresponds instead to range-time samples within these cloud layers, weighted according to their aerosol backscatter contribution. These samples are used to describe the optical distribution of the cloud population in the $(1-1/R, \delta_T)$ space, but they should not be interpreted as independent LR retrievals. Samples belonging to the same cloud layer are therefore temporally and vertically correlated and share the same retrieved layer-mean LR. They do not increase the effective number of independent observations, which remains the number of cloud layers reported in Table 1.

*The authors reported (line 135) that there are ancillary radiosondes and model output that provided temperature and pressure, from which ancillary molecular backscattering coefficients could be computed. However, these temperature data were not used in any of their analyses, either to select the height of the tropopause (I do recognize they performed a sensitivity test here, but why not use the actual value?), to evaluate some of their hypotheses (e.g., line 340 when they suggest the difficulty of isolating pure ice layers from mixed-phase structures; knowing if the temperature is above/below -40 Celsius would be useful, or e.g., line 353 when they discuss small vs large NAT crystals), or analyzing the LR values (LR does not depend on height like Fig 4 or Fig 8, but it can be related to temperature which helps constrain moisture).*

We thank the reviewer for this important comment. We agree that temperature provides a more physically meaningful coordinate than altitude alone for interpreting PSC and cirrus optical properties. In the original manuscript, temperature and pressure profiles from radiosondes or model fields were used primarily to compute the molecular backscatter and extinction profiles. We have now clarified this point and added a temperature-based discussion to better connect the retrieved LR values with the expected microphysical regimes.

We retained the fixed 12 km climatological tropopause as an operational separator between tropospheric cirrus and PSCs because it provides a homogeneous and reproducible classification criterion for the full three-year dataset. Using profile-by-profile tropopause estimates from radiosondes or model fields would introduce an additional dependence on temporal sampling, vertical resolution and the availability of ancillary data. Moreover, the sensitivity test in which the tropopause height was shifted by ±2 km showed that the resulting LR statistics and cloud-class assignments are not significantly modified. We have now explained this rationale more explicitly in the manuscript.

Following the reviewer's suggestion, we also expanded the interpretation of the LR results in terms of temperature. In particular, we now discuss that ice PSC layers are expected to occur close to or below the frost-point temperature, whereas NAT and STS occur in warmer but still PSC-favourable regimes. This helps clarify why isolating pure ice PSC layers is difficult: at temperatures close to the transition between NAT/STS and ice, mixed or vertically layered particle populations are likely. Similarly, the discussion of NAT LR variability is now linked more explicitly to the temperature dependence of particle growth and possible changes in NAT size and asphericity. For cirrus, we clarified that LR is not expected to depend uniquely on height; temperature provides a more physically relevant coordinate, although the observed variability remains large.

We have added text in the Statistical analysis and Discussion sections to clarify how the ancillary temperature information is used and how it supports the interpretation of PSC and cirrus LR variability.

In the Statistical Analysis:

The fixed 12 km tropopause is therefore used here as a reproducible operational separator between tropospheric cirrus and PSCs over the full three-year dataset. Profile-by-profile tropopause estimates from radiosondes or model fields were not used for the main classification in order to avoid introducing an additional dependence on the temporal sampling and vertical resolution of the ancillary meteorological data. Nevertheless, the temperature profiles used to compute the molecular atmosphere provide important physical context for interpreting the retrieved LR values. For each cloud layer, temperature was therefore used diagnostically to relate the LR statistics to the expected thermodynamic regimes of STS, NAT, ice PSC and cirrus formation.

In the discussion:

Temperature provides an additional physical coordinate for interpreting these results. Altitude alone should not be interpreted as the controlling variable for LR; rather, altitude acts as a proxy for the thermodynamic regime sampled by the cloud. For PSCs, the relevant temperature thresholds are those controlling STS growth, NAT formation and ice nucleation. Ice PSC are expected near or below the frost-point temperature, whereas NAT and STS can occur at warmer temperatures within the PSC stability range. This is important for the interpretation of the LR distributions: layers sampled close to the NAT/ice transition may contain externally mixed or vertically interleaved particle populations, so that the retrieved layer-mean LR represents an effective value rather than a pure ice or pure NAT optical signature. Similarly, the increase of NAT LR with altitude is interpreted here primarily as a temperature-related effect, reflecting colder conditions that favour larger or more aspherical NAT particles, rather than as a direct dependence of LR on geometric height.

*Line 466 suggests that analyzing the subset of Young-optimized retrievals results in only a modest shift in median LR. However, for LS_ice, the value went from 52 sr to 32 sr; this is not modest. And given there are only 9 vs 2 clouds of ice PSC (Table 1), it seems these statistics are extremely uncertain (especially due to temporal autocorrelation; see first comment). Please provide more discussion here.*

We agree with the reviewer. The expression “modest shift” was inappropriate, especially for ice PSC. In the full dataset, the median LR for ice PSC is about 52 sr, whereas the few ice PSC layers retained in the Young-compatible subset yield values around 32-38 sr. This change should not be interpreted as a robust microphysical shift.

We have revised the Discussion to make this point explicit. For STS and NAT, the Young-compatible filtering mainly narrows the LR distributions and provides a useful benchmark for layers that better satisfy the Young assumptions. For ice PSC, however, only 2 layers are retained out of 9 in the full dataset. This number is too small to define a representative median LR, especially considering that samples within each layer are temporally and vertically correlated and do not increase the effective number of independent observations.

We now state that the low LR obtained for the Young-compatible ice PSC subset is most likely a selection effect. The Young criteria preferentially reject vertically structured, mixed-phase, or multilayer ice PSC, retaining only a few optically simple or marginal cases. Therefore, the Young-compatible ice PSC statistics are not used as representative of typical ice PSC conditions. The robust result is instead the strong reduction of ice PSC cases under the Young-compatibility criteria, which indicates that most ice PSC observed in this dataset violate the vertical homogeneity assumptions required by the Young inversion. In the discussion the text has been modified accordingly:

For ice PSC, however, the number of retained layers becomes too small to define a representative lidar-ratio statistic. The apparent decrease of the ice-PSC median LR in the Young-compatible subset should therefore not be interpreted as a robust microphysical result. It more likely reflects the preferential exclusion of vertically structured and mixed-phase ice PSC layers, leaving only a few optically simple or marginal cases. This limitation is particularly important because only 2 ice PSC layers are retained in the Young-compatible subset, compared with 9 in the full dataset, so that the apparent median value is highly uncertain and should not be treated as statistically representative.

*Line 479 leads to an intriguing thought: if the Young criteria can be used to identify cloud inhomogeneity, then it seems that it would be useful to indicate that fraction of each observed cloud type that could be considered homogeneous.*

*This could be added to Table 1. But more than just adding an additional column to that table, please see if there is supporting literature that supports the 'estimated fraction that are homogeneous' for each class.*

We thank the reviewer for this useful and insightful suggestion. We agree that reporting the fraction of layers satisfying the Young-compatible criteria provides valuable information on how often each cloud class approaches the ideal conditions required by the Young inversion. We have therefore added an additional column to Table 1 reporting the fraction of cloud layers retained in the Young-compatible subset relative to the full dataset.

However, we have deliberately labelled this quantity as the "Young-optimized fraction'' rather than as a homogeneous-cloud fraction. The reason is that the selection criteria are applied to the external reference windows above and below each cloud layer. A failure of these criteria may indicate internal cloud inhomogeneity, because a vertically heterogeneous cloud cannot be described by a single layer-mean LR and may produce an inconsistent extinction correction. However, this is not the only possible cause. Similar failures may also result from residual aerosol or thin undetected layers in the external reference windows, vertical gradients in the background atmosphere, uncertainties in cloud-boundary detection, or reduced signal-to-noise ratio.

We have therefore revised the Discussion to clarify that the Young-optimized fraction is an operational measure of how often each cloud class satisfies the combined cloud-background conditions required by the Young inversion. It should not be interpreted as a direct or universal estimate of the fraction of internally homogeneous clouds.

To our knowledge, directly comparable homogeneous-cloud fractions based on the same Young top-bottom consistency and external-window homogeneity criteria have not been reported in the literature. We have therefore discussed our results qualitatively in relation to previous PSC studies showing that PSCs, especially ice PSC and mixed-phase layers, often exhibit strong vertical structure and compositional variability. The low Young-optimized fractions reported here are consistent with this picture, but they should be interpreted as compatibility with the assumptions of the Young inversion rather than as an independent climatological estimate of cloud homogeneity.

The table has been updated with a fourth column, and commented in the text as:

The additional Young-optimized fraction reported there provides an operational estimate of the fraction of layers satisfying the symmetry, smoothness and external-window homogeneity criteria used in this study. This quantity should not be interpreted as a universal homogeneous-cloud fraction, because the criteria are applied outside the cloud layer and depend on the adopted thresholds, the signal-to-noise ratio, the accuracy of cloud-boundary detection and the uniformity of the atmosphere immediately above and below the cloud. Rather, it quantifies how often each class satisfies the combined cloud-background conditions required for a robust Young inversion.

Moreover, in the Discussion we have changed the text following line 479 as:

Although the Young-optimized criteria are applied to regions outside the cloud, they are sensitive to any departure from the ideal conditions required by the Young inversion. If the cloud is a single, vertically homogeneous layer embedded in a uniform background, the extinction correction based on a trial LR modifies the attenuated signal in a smooth and self-consistent manner, yielding nearly identical values of the corrected backscatter ratio above and below the cloud. In contrast, if the cloud contains internal sublayers with different extinction-to-backscatter ratios, no single LR can adequately describe the entire layer. The inversion may then overcorrect one part of the cloud and undercorrect another, with these errors propagating into the extinction-corrected signal outside the cloud (Ansmann et al., 1990). As a result, the corrected backscatter ratio may become asymmetric above and below the cloud and may exhibit enhanced variance or gradients in the external control windows.

However, a failure of the Young-optimized criteria cannot be attributed uniquely to internal cloud inhomogeneity. Similar signatures may also be produced by residual aerosol or thin undetected layers in the external reference windows, vertical gradients in the background atmosphere, uncertainties in cloud-boundary detection, or reduced signal-to-noise ratio. For this reason, the Young-optimized fraction reported in Table 1 should be interpreted operationally: it quantifies the fraction of cloud layers for which the combined cloud-background system satisfies the symmetry, smoothness and homogeneity conditions required for a robust Young inversion. It should not be interpreted as a direct or universal estimate of the fraction of internally homogeneous clouds. This interpretation is consistent with previous PSC studies showing that polar stratospheric clouds often display substantial structural and compositional variability. Ground-based and spaceborne lidar observations have shown that PSC fields may contain externally mixed or vertically layered STS, NAT and ice particle populations, and that ice PSC in particular are often associated with sharp gradients, mesoscale temperature perturbations and complex vertical structure (Pitts et al., 2018, Snels et al. 2021, Di Liberto et al. 2024). To our knowledge, however, directly comparable Young-optimized fractions based on the same top-bottom consistency and external-window homogeneity criteria have not been reported previously. The fractions reported here should therefore be regarded as dataset- and method-specific indicators of compatibility with the Young inversion assumptions. The apparently larger Young-optimized fraction for ice PSC should not be overinterpreted, because it is based on only nine ice PSC layers in the full dataset. The more robust information is the very small absolute number of retained ice PSC cases, which confirms that the Young-optimized ice PSC subset is not statistically representative.