

Thank you for your valuable comments. Please find below a point-by-point reply to the questions you have raised in your comment.

- (1) It is not clear if CLAAS-3 is a reliable reference against which the CTH retrievals with Šoljan et al. 2024 can be compared. Especially because CLAAS-3 retrieves CTH using a Neural network based algorithm and does not represent ground truth. Why not use ground-based ceilometer measurements as reference ?**

Active sensors - particularly profiling lidar - provide the most accurate reference for CTH validation in the scientific community. While no absolute 'ground truth' exists for cloud detection, CloudSat-CALIPSO lidar-radar data offer the most reliable assessment of retrieval methods and are frequently used to train and validate approaches like CLAAS-3. This study uses CloudSat-CALIPSO as the primary reference, supplemented by imaging instruments, and treats CLAAS-3 only as a benchmark and reference - never as 'ground truth' (hence the term does not appear in the paper).

Regarding ground-based ceilometer measurements, their use is physically impractical for two fundamental reasons: (1) ceilometers sense cloud base height, whereas this study focuses on cloud top height; (2) the study examines optically thick deep convective clouds, which prevent a lidar signal from penetrating from the Earth's surface to the cloud top and returning. Therefore, spaceborne lidar-radar observations remain the only physically viable reference for validating CTH retrievals of deep convective clouds.

- (2) Why is ICAO standard atmosphere, which is a static profile under an idealized condition, used for converting cloud top pressure to height ? Could it be more accurate or more closer to reality when profiles from NWP or Reanalysis are used ?**

Absolutely true. The ICAO atmosphere can be replaced by any other atmospheric model of interest, including reanalysis, numerical weather prediction (NWP), or rawinsonde observations to represent more realistic and site-specific conditions. There are two reasons the ICAO standard atmosphere is used here. First, employing a single predefined atmospheric profile simplifies and accelerates the computational process, especially when the algorithm is applied to large datasets (e.g., multidecadal time series of full-disk Meteosat imagery). Second, the ICAO atmosphere was originally implemented by Šoljan et al. (2024), who focused on fast CTH retrieval for aviation applications. The goal of this study was to validate the Šoljan et al. (2024) approach; therefore, the original setup was maintained. Future work can test whether using a local atmospheric model impacts the accuracy of CTH estimation.

- (3) There is a lack of reliable ground truth for assessing the accuracy of the parallax correction using the different CTH sources. Would it make sense to look at some surface parameters such as precipitation, direct solar irradiance, lightning, etc. that are directly impacted by the location of the DCC ?**

The suggested approach would be quite interesting if manageable. Unfortunately, all the mentioned weather phenomena and variables (precipitation, direct solar irradiance, lightning) come with some degree of uncertainty in their spatial location (and co-location). For instance, precipitation estimates may be derived from satellite imagery and then validated against ground-based radar or rain gauges (e.g. Roebeling and Holleman, 2009; 10.1029/2009JD012102). However, radar products are known to have limitations related to accuracy and reliability that decrease with distance from the radar location (e.g., Sharif et al. 2002, 10.1029/2001WR000525; Green et al. 2024, 10.5194/hess-28-4539-2024), while rain gauges offer sparse coverage compared to the dense grid of satellite pixels. As mentioned in the answer to comment (1), space-borne lidar cloud-top height (CTH) is acknowledged as the most reliable CTH dataset currently available and provides the most accurate relationship between Šoljan et al. (2024) estimates and actual CTH. Other CTH sources are used

only to demonstrate whether it is possible to achieve the same level of parallax correction with Šoljan et al. (2024) method as would be achievable with operational products (especially geostationary-based CLAAS-3).

- (4) **Section 5.1. Lines 350-355. The mean displacement vector length for parallax may not be very informative. As the author himself noted, parallax displacement is heavily dependent on the viewing satellite zenith angle (SZA). Therefore, it could be more intuitive to have the mean displacement vector and standard deviation for different SZA ranges.**

Thank you for that comment and for pointing out the limited use of parallax displacement expressed in metric length. It is correct to conclude that the satellite zenith angle makes the interpretation of parallax displacement challenging. This is why the metric lengths have been normalized by SZA - indirectly, using pixel size (a SZA-dependent measure) to report parallax displacement in pixels rather than meters (sec. 5.2 vs sec. 5.1). Obviously, as the SZA increases, the mean metric length of a parallax displacement per SZA bin also increases.

Thank you for your comments. Please find below a point-by-point reply to the raised questions.

- Lines 187-189: The description of the DCC definition is methodologically incomplete. The author reviews several BTDD thresholds from the literature but does not clearly state which threshold(s) are actually applied in this study, nor how the regional differentiation (e.g., tropics vs. mid-latitudes) is implemented. Similarly, the CAPE threshold of 100 J kg^{-1} is introduced as “arbitrarily selected”. Please specify: (1) the exact BTDD threshold(s) and the spatial domain to which each applies; (2) whether the BTDD and CAPE conditions are combined with “AND”; and (3) a justification or sensitivity check for the CAPE threshold.**

(1) The BTDD thresholds were: 2.5 K for mid-latitudes only, 1.0 K for the tropics only, and 2.0 K when all latitudes were considered together; this has been clarified in the text as requested.

(2) The BTDD requirement and the CAPE requirement are both necessary conditions, so they are combined with a logical ‘AND’; this has been clarified in the text as requested.

(3) The CAPE threshold affects the absolute frequency of DCCs (Tab. R1), but not the CTH retrieval procedure or the parallax correction procedure. Raising the CAPE threshold would reduce the total number of DCC cases under evaluation while simultaneously increasing the proportion of colder, more extreme events. As a consequence, the results would shift towards what is currently observed for DCCs with the highest CTH - more extreme DCCs are also those that penetrate deeper into the troposphere, and therefore tend to have higher CTH and colder cloud tops. A discussion of this sensitivity has been added to the manuscript.

Tab. R1. Mean annual DCC frequency over Europe (10°W - 40°E , 30°N - 60°N) as a function of CAPE threshold (BTDD threshold fixed at 2.5 K).

CAPE threshold	Mean annual DCC frequency		
	Year 2007	Year 2013	Year 2023
No threshold, any CAPE	0.0098	0.0113	0.0099
>0 J/kg	0.0083	0.0094	0.0083
>100 J/kg	0.0027	0.0026	0.0024
>500 J/kg	0.0011	0.0011	0.0010
>1000 J/kg	0.0004	0.0005	0.0004

- Figure 1: Please verify the validity of the data presented in the figure. The statistical metrics shown (e.g., correlation coefficients) do not appear to be consistent with the visual results in the figure, nor do they align with the discussion in the main text.**

Verified and clarified. Both values are correct, but refer to different statistics, which turned out to be insufficiently described. In Fig. 1, the R^2 coefficient refers to the linear fit applied to CTH averaged over 1° latitude bands, to account for the significant latitudinal disproportion in DCC occurrence. The correlation coefficient in the text refers to the correlation computed between individual CTH retrievals (millions of points, with larger spread than the zonally averaged means). To avoid confusion, the captions of Fig. 1 and Fig. 2 have been updated to explicitly state that R^2 applies to the zonally averaged means.

- The author stated that “after applying a simple regression-based bias adjustment, the absolute error is reduced to below 1.2 km.” However, it should be noted that this regression model was trained on a specific sample from 2007. From my perspective, the adjustment was not dynamic. If this regression correction is applied directly to historical Meteosat**

MVIRI data from four decades ago, how robust is it? Is there a risk that the regression formula artificially smooths out genuine physical signals?

Fully agreed. The empirical bias adjustment was trained on 2007 data and may not be fully transferable to the older MVIRI record. It should be noted, however, that the key finding is that the CTH and parallax correction methods perform well in their unadjusted form — i.e. without bias correction: the method placed 84% of DCC pixels in the same location as the CLAAS-based correction (or 97% when accounting for uncertainty in the CLAAS dataset). Noted in the manuscript.

On the second part of the question - whether the regression could artificially smooth out a genuine physical signal - it cannot. The calibration is a first-order (affine) transform and is therefore monotonic and one-to-one: it rescales and offsets the values but preserves the sign of any trend or gradient, and with fitted slopes ≥ 1 it amplifies rather than dampens CTH variability. We have added a sentence to Sect. 4.1 stating this explicitly.

4. **For V24-based CTH estimates from SEVIRI measurements, why did the author choose to compare the results with CLAAS CTH, which itself also has inherent biases in CTH retrieval? The direct comparison with CALIPSO/CloudSat CTH products would be more reliable.**

+

5. **The evaluation logic in Section 5.1 is confusing. The author compared the VŠ24 cloud top height (CTH) retrievals against three different reference datasets (SEVIRI, MODIS, CloudSat/CALIPSO), and then evaluated the calibrated results against CLAAS dataset using parallax displacement vectors. The CLAAS-derived parallax displacements were regarded as ground truth without discussing the uncertainties inherent to CLAAS CTH products. I recommend that the author needs to better elaborate: (1) justify why CLAAS is a suitable evaluation benchmark, including its uncertainty, and why not CALIPSO; and (2) detail the bias adjustment method explicitly.**

Questions 4 and 5 are closely related; I will therefore address both in a single reply.

CLASS as benchmark. The overall goal of this study is to evaluate the VŠ24 method for a specific purpose: to obtain a fast yet reliable CTH retrieval that provides parallax correction equivalent to what is currently achievable with the CLASS dataset, while requiring only IR brightness temperature data (supported by reanalysis). Since the intention is to apply the method to CTH estimation of geostationary data, CLASS was selected as the natural benchmark, as it serves as the primary CTH product for SEVIRI.

It is true that CLASS carries its own uncertainty — as do MODIS and CloudSat/CALIPSO. Only the latter is generally considered to provide the least biased CTH data, and is therefore used as an "absolute" reference in many studies. This paper is no exception: VŠ24 retrievals are compared with lidar–radar data in Sec. 4.1. However, the CloudSat/CALIPSO does not accommodate high viewing angles, which are characteristic of geostationary observations and unachievable in lidar–radar nadir geometry. This constitutes an additional, technical justification for the use of CLASS as the benchmark.

Uncertainties were addressed in terms of the per-pixel CLASS retrieval uncertainty provided by EUMETSAT. The paper considers not only a single ("central") CTH estimate, but also upper and lower estimates derived from the uncertainty range.

The bias adjustment method has been described explicitly in the revised manuscript, as suggested.

6. **Author states that “Despite a ~10% underestimation of CTH, the impact on parallax correction was minimal: 84% of pixel geolocations coincided with those derived from CLAAS-based corrections, increasing to 97% when CLAAS retrieval uncertainty was accounted for.” Satellite viewing zenith angles (VZA) over Europe are generally large due to its higher latitude, which should amplify the effect of height errors. Why does such height error lead to a small pixel misalignment after parallax correction?**

The validation of parallax correction results constitutes the main — and ultimate — goal of this study. We assume that parallax correction based on VŠ24 CTH retrieval is reliable and acceptable if DCC clouds are shifted in the Meteosat image by the same number of pixels and in the same direction as when using CLASS CTH. The effect of VZA is noted and addressed when the metric lengths of shift vectors for the proposed method and the CLASS-based method are compared. However, in operational use, metric vectors are immediately converted into pixel-based shift vectors, where metric length is normalized by pixel size (a function of viewing angle). Consequently, differences in metric length — whether overall or as a function of viewing angle — are not essential: what matters are the final pixel-level shifts. This is precisely why larger errors in the metric length of a shift vector can still produce accurate parallax correction: errors in length may be smaller than the viewing-angle-normalized pixel size. The results of Sec. 5.2 are therefore the most relevant for the study and deserve the closest attention.

Sec. 5.1 is provided for reference and could technically be removed (if deemed confusing) without affecting the results presented in Sec. 5.2. I leave this decision to the editor.