Response to Reviewers' Comments

Overall response

We sincerely thank the reviewer for the careful re-evaluation of our manuscript, the constructive suggestions, and the positive assessment of the improvements.

In addition to the revisions addressing the reviewer's comments, we have updated the preprocessing description in Appendix A1. We clarified the order of range-squared correction and natural-logarithmic transformation and specified that range is expressed in kilometres. These changes make the methodological description consistent with the implementation used to obtain the reported results. We have also adjusted the layout of several figures to better accommodate the publication format of AMT.

Reviewer 1

Technical correction 1

Reviewer comment

Fig 1 map: I still don't see the need to include country borders in general into the map for any atmospheric research. But I leave the decision up to Copernicus. Nevertheless, I cannot find the statement concerning the used map in the caption as you have promised in the reply to my previous review ("… and clarified in the caption that these elements have no additional analytical meaning in this study.").

Author response

Thank you for pointing this out. We acknowledge that the clarification promised in our previous response was missing from the revised caption. We have now added an explicit statement to the caption of Fig. 1 explaining that country borders and the South China Sea inset, including its dashed boundary line, are base-map elements only, have no additional analytical meaning, and are not used for site selection, dataset partitioning, or result analysis.

Location of revision

Caption of Fig. 1.

Technical correction 2

Reviewer comment

Fig 5. is completely different in terms of data content compared to the Fig. 5 of the initial manuscript. Please clarify what has been changed. Why are there no outliers any more?

Author response

Thank you for highlighting this substantial change. Fig. 5 was recalculated using a revised layer-evaluation procedure. We acknowledge that our previous response did not adequately explain these changes, which involve layer matching, layer eligibility, model checkpoints, and VDE-mask generation.

In the original evaluation script, reference and predicted layers were sorted by height, and layers were paired according to their order: the first reference layer with the first predicted layer, followed by the second reference layer with the second predicted layer. Only the first two layers were evaluated. A corresponding predicted layer was considered detected whenever it existed, without checking whether the paired layers overlapped vertically or were reasonably close. Consequently, missed layers, additional thin layers, or layer splitting could shift the layer ordering and produce comparisons between geometrically unrelated layers.

The revised evaluation uses global one-to-one assignment based on the Hungarian algorithm (Kuhn, 1955). Vertically overlapping layer pairs are prioritized using vertical intersection over union, with layer-center separation as a secondary criterion. For non-overlapping layers, a candidate pair is permitted only when the minimum vertical separation does not exceed 300 m. All reference and predicted layers satisfying the minimum thickness requirement of 90 m are evaluated, rather than only the first two layers. Boundary errors are calculated for successfully matched pairs, while unmatched reference layers are counted as missed detections.

To examine the effect of matching separately, we held the current reference layers, predicted masks, and layer-thickness criteria fixed and changed only the pairing procedure. The resulting detectable cloud-top MAEs were:

Method	Height-order pairing	Revised physical pairing
Pretrained model	215.6 m	82.2 m
Direct supervised model	326.9 m	94.0 m
VDE	1346.1 m	250.7 m

This controlled comparison demonstrates that the pairing procedure substantially affects the boundary-error statistics. However, it uses the current inputs and should not be interpreted as a complete decomposition of the changes between the original and revised manuscripts.

Regarding the apparent disappearance of outliers, no successfully matched pairs were removed because their boundary errors exceeded 300 m. The 300 m criterion applies to the minimum separation of non-overlapping candidate layers, not to the boundary error of an accepted pair. The revised evaluation still contains 778, 865, and 2045 matched layers with absolute cloud-top errors greater than 300 m for the pretrained model, direct supervised model, and VDE, respectively.

The revised assignment can associate a layer with a more appropriate counterpart or leave it unmatched when no valid candidate exists, thereby reducing comparisons between unrelated layers.

All reported statistics, including N, MAE, bias, RMSE, and Pearson R, are calculated from the complete set of successfully matched pairs. For visualization, a maximum of 2000 pairs per method are randomly plotted using a fixed seed of 42. The original plotting script used the same sampling procedure; this plotting choice is therefore not a new outlier-removal step.

We have expanded the manuscript to describe the current layer-extraction criteria, matching rules, and distinction between the complete statistical sample and the plotted subset.

Location of the revision: *Sections 2.6 and 4.2.2, the caption of Fig. 5, and the reference list.*

Technical correction 3

Reviewer comment

Please describe all symbols used in Equations, especially in the Appendix this is often missing.

Author response

Thank you for this suggestion. We have supplemented the missing symbol definitions in Appendix A1 and A4. Specifically, we now explain the channel, temporal, and vertical indices and normalization quantities; the symbols used in range-squared correction and logarithmic compression; and the total losses, individual loss components, and weighting coefficients for Stages 1 and 2. These explanations are placed immediately after the relevant equations.

Location of revision

Appendix A1 and A4.

Technical correction 4

Reviewer comment

Line 467 of tracked changed manuscript: "Temporal continuity features are extracted using 3×1, which integrate..." I believe the word "kernel" is missing after "3x1".

Author response

Thank you for pointing out this omission. We have added "convolutional kernels" after "3 × 1" to clarify the operation described in this sentence.

Location of revision

Section 3.2.

Technical correction 5

Reviewer comment

Line 470 of tracked changed manuscript: The sentence sounds a bit strange, please check. E.g. what is bottleneck layer?

Author response

Thank you for this comment. We have rewritten the sentence to clarify which network components use each convolutional kernel size. We have also defined the bottleneck as the central network block connecting the encoder and decoder at the coarsest vertical resolution.

Location of revision

Section 3.2.

Technical correction 6

Reviewer comment

Line 532 of tracked changed manuscript: what is "multi-wavelength averaging"? Please rephrase or describe

Author response

Thank you for highlighting this unclear expression. We have replaced "multi-wavelength averaging" with an explicit description of the operation: at each time–height pixel, the VDE-derived cloud-probability scores from the 355, 532, and 1064 nm elastic channels are averaged to form the soft supervisory target used during Stage 2.

Location of revision

Section 3.3, description of Stage 2.

Technical correction 6

Reviewer comment

Line 1035 - 1041: The explanation of the vertical resampling procedure is much clearer now. But I still miss a statement about which pooling approached is used. Is it Max-pooling, Average-pooling, Min-pooling, or any other pooling based approach?

Author response

Thank you for identifying this missing implementation detail. We have clarified that the input vertical resampling uses average pooling along the height dimension. In the middle and upper regions, each non-overlapping group of four and eight adjacent range bins, respectively, is replaced by its arithmetic mean. The lower region retains

its original vertical resolution, and the temporal and channel dimensions remain unchanged.

Location of revision

Appendix A1, description of vertical resampling.

Reviewer 2

Reviewer comment

The revised manuscript is accepted, but some technical corrections should be made.
The figure name should appear at the beginning of sentences as "Figure" and in the middle of sentences as "Fig."
The same should be done for "section" and "sect."
For example, line 192 should be written as " Figure 1 shows ..." , because the sentence starts with the figure name.

Author response

Thank you for this helpful suggestion. We have carefully checked the manuscript and standardized the use of figure references throughout the text. Specifically, "Figure" is used when the figure reference appears at the beginning of a sentence, whereas "Fig." is used within sentences. The section reference format was also checked and found to be consistent with the journal style.

Location of revision

The usage of figure and section references has been corrected throughout the revised manuscript