Manuscript: Lenhardt, J., Quaas, J., Sejdinovic, D., and Klocke, D.: CloudViT: classifying cloud types in global satellite data and in kilometre-resolution simulations using vision transformers, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-2724, 2024.

Julien Lenhardt, Johannes Quaas, Dino Sejdinovic, and Daniel Klocke
19.01.2026

Response to Referee #2 on the manuscript "CloudViT: classifying cloud types in global satellite data and in kilometre-resolution simulations using vision transformers" submitted on 09 Nov 2025.

Dear anonymous referee,

We would like to thank you for the comments and the constructive discussion. Please find our response below, in which the review comments are in bold and followed by our response.

The complete edits can be viewed in the revised version of the manuscript which includes tracked changes. The line numbers referenced below correspond to the manuscript's lines before the revisions.

Best regards,
Julien Lenhardt on behalf of the authors

Suggestions for revision

In my opinion, the abstract is improved by acknowledging the limitations of the current model. The authors also included more relevant studies that deal with closely related subjects. I appreciate the authors' effort in making these changes and trying to improve the results. However, the abstract does not reflect this and still reads like the current state-of-the-art is schemes using 2-D histograms, which is not true. The approach in this article is very similar to the ones I mentioned in the last round. The specific deep learning technique/data source may differ somewhat but given the poor performance it is hard to justify the discussed implications in this paper.
The reviewer is right regarding that, despite some relevant references being added to the core of the text, the abstract was not adapted to reflect this. To this extent, we have adapted the abstract to read the following (L18-19):
"*Cloud types have traditionally been defined using a simplified partition of a two-dimensional space, e.g., cloud top pressure and optical thickness, but recently using deep learning.*"

On the point of applying the trained model to high resolution model simulations, the question I raised last time was not answered satisfactorily. This exercise basically tries to apply a model

trained on MODIS data and ground observation data, with poor performance, to high-resolution model simulation whose data may be quite different from MODIS, without any validation or even ground truth to compare with. Surely, it begs the question of how to justify it and what we should make of the results because applying a non-performant model to a different dataset that may be quite different from the MODIS data involves two risky steps. Yet, no validation is provided. This 'next step' should be paused because results from it may be highly unreliable and hard to interpret.

A really performant model that can be taken as capable of 'understanding' cloud classification could be used in this way assuming the data from the high-resolution model are not wildly different from MODIS. Even then, this step has to be trodden carefully because we know that though high-resolution models have many great features, they can still have shortcomings when compared to actual observations.

I hope that the authors understand that this is a real concern I as a reviewer have. Honestly, I appreciate the authors' effort in trying this method. I would have expected that this framework to work nicely for the same reasons the authors discussed in the manuscript, but the reality is that it did not. However, this paper is best published as a valuable exercise for the community to learn from. I suggest the authors discuss more on why the model does not perform well on the data, what the authors tried to improve the performance, and how they worked or not. Such information would be really valuable for researchers in our community. Packaging it as a powerful new way of studying high-resolution models is misguided given the current results. A simple thought experiment is that if it turns out that we could not train a performant model with this framework in the manuscript, it would suggest that the current framework is not viable. The implication should then be that researchers should avoid this seemingly viable option. The authors should either make it work, or document what they tried, how the results look, and what the implications are.

We thank the reviewer for the constructive comments regarding the goals/objectives and the framing of the manuscript. To further adapt the manuscript to reflect these points, we have proceeded with a number of modifications, mainly in the text, but also in the structure: for example with the addition of a section on limitations and challenges (now section 6), and with the original section 5.2 about the application to climate model simulation data moved out of the results section.

We agree that the actual cloud classification results on the climate model simulation should not be part of the results section of the manuscript as such. However, we find the display of the technical application of such a method, the discussion and comments regarding the actual implementation, and the current need for cloud type diagnostics applicable to climate model outputs, to be interesting enough to be mentioned in further details in the new section on challenges and limitations, and in an appendix as a whole. To this extent, the content regarding this application has been moved from section 5.2 to Appendix D, and the limitations and challenges previously highlighted are kept general enough and do not report on any potential scientific results that could derive from it. The title of the manuscript has also been adapted to reflect how we are taking into account this point (see below). For a detailed overview of this reorganization, please have a look at the new version of the manuscript or the track changes.

You will find in the list below the most relevant sections that were adapted, and we refer to the accompanying track changes for a complete list of modifications, including minor ones:

- L1-2: "*CloudViT: exploring cloud type classification with vision transformers in global satellite data.*"

- L25-27: "Potential improvements emerge in the reduction in mismatches between data sources, the extension of the colocated dataset, and the refinement of the classification model. *While the application of the method in its current state comes with apparent uncertainties due to limited performance, it raises relevant challenges and limitations, from which the community can benefit from discussing for the development of similar methods. To foster future advancements, the dataset and model are available from Zenodo.*"
- L114-116: "*The results in section 5 focus on the extension to a global distribution of cloud types. Challenges, limitations, and lessons learned from CloudViT's development are highlighted in the following section 6, with the guiding idea of making cloud type classification with vision transformers reliable, capable of achieving notable performance, and potentially applicable to high-resolution climate model simulations. Eventually, we conclude over the presented method, and the challenges of cloud type classification.*"
- L563: Section 5 is renamed to "*Global cloud type distributions in MODIS data*"
- L636: Section 6 is now named "*Challenges and limitations*" and introduced with the following paragraph: "*The method and results highlighted in the previous sections provide useful material to further analyze the developed methodology, but also to be critical of its shortcomings. The following section aims to focus on several aspects that we feel are relevant for the community when developing cloud type classification methods similar to the one presented here, namely on the benefits of such methods, dataset curation and extension, and the potential application to climate model data.*". This section then contains relevant paragraphs from the conclusion L647-701 that were only slightly adapted to follow the new structure.
- L644-646: "*We evaluated the classification model by examining distributions of cloud properties in Section 4 and the global spatial distribution of cloud types in Section 5. Lastly, we pinned down some existing challenges, limitations, and lessons learned from the development of the method for cloud type classification.*"
- L702-710: "*In conclusion, the method presented here showcases and highlights a wide array of potential applications in the study of cloud types, their characteristics and evolution, and their past, current, and future effects on the Earth's climate: from the extension of sparse surface observations to global yearly predictions, and existing challenges and limitations in the design of vision-transformer-based models. Despite the relatively imbalanced performance assessment of the method which shows both great promise in capturing large scale characteristics of cloud types distributions but struggles to capture precisely the features in the training dataset, the design and development of CloudViT is an interesting study in the line of improving existing cloud classification methodologies. Identified challenges and limitations in this particular case can be useful to the community, both in terms of methodology development and caveat to be avoided. We recommend future advancements in cloud classification methods like CloudViT being firstly focused on data curation and followingly on model tuning once the performance has been raised to desirable levels.*"
- L769: The appendix D is now renamed to "*Exploring the technical feasibility of the application to a global storm-resolving model simulation*".