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
06.03.2026

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 21 Feb 2026.

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

The changes that the authors made are great towards matching the results and significance discussion of the method. I appreciate the effort and like the discussion section. The application to satellite data is also interesting. I do find the discussion section a bit hard to follow. I give a few examples in the following.
We thank the reviewer for the constructive comments on the discussion section. The section has been adapted in several paragraphs. In summary, the discussion section has been streamlined to improve the flow and get rid of not so relevant sentences. The modifications are not reported in detail here for simplicity, but are visible in the attached track changes.

1. **'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.'**
   It is not clear what this sentence tries to express. Please make sure to write it with clarity. For example, if you meant to say 'While our method provides a good foundation to be further developed, it does have its limitations.' That would be clear and easy to follow.
   Agree, the sentence (L598-599) has been updated to be more readable.

2. **‘Spatially-resolved cloud properties provide usable context for the CloudViT model to improve the cloud classification, as shown in the comparison to the baseline method with limited spatial information. Introducing this new transformer model architecture additionally improves the classification skill over the CNN backbone mentioned in Lenhardt et al. (2024a). Overall, CloudViT achieves passable performance even on sparsely represented classes for both cases of 4 and 10 cloud types.’**
   It is true that this new method beats the baseline as shown in the result section. This is indeed progress and has been rightfully highlighted in previous sections. The main point of my previous comments should be highlighted here. The issue is that while the method achieves progress relative to previous methods, its absolute performance is surprisingly low. I say this because previous CCN-based methods (ViT should beat them usually) often achieves metrics that are much better than the current one. Discussions should be made towards this point.
   The relevant sentences (L602-605) have been updated to expand on this point, also following the more general review in comment 4.

3. **‘The limited colocated dataset proves to be a hurdle for the proper training and evaluation of the method on labelled samples but the generation of an extensive global dataset allows deeper investigation into the cloud types. Improvements could come from a more extensive training dataset which would encompass a larger variety of cloud type samples to certainly enhance the classification’s performance both for the training and testing metrics. The subsequent evaluation exhibits interesting results despite the limited performance on the colocated dataset.’**
   Authors say that limited collocated dataset ‘proves’ to be a hurdle. Where was the proof, or it is the author’s presumed cause? It is important to show this with evidence. Why ‘but the generation of an extensive global dataset allows deeper investigation into the cloud types’ is the case? The authors should spell out why applying the method to a global dataset allows deeper investigation into the cloud types. Applying a model that achieves around 0.43 f1 score to a global dataset should produce a dataset that contains a lot of errors and uncertainties.
   The relevant sentences (L605-610) have been updated to expand on this point, also following the more general review in comment 4.

4. **I feel ‘The subsequent evaluation exhibits interesting results despite the limited performance on the colocated dataset.’ is not warranted. The reason remains that applying a model that is relatively performant (as shown in this study, relative to old methods), but non-performant in absolute terms does not provide additional insights even if the patterns can be seemingly interesting. Again, due to the limited performance of the model, we do NOT know the true distribution for the global data. Simply having the global pattern seems not an asset. That remains the fundamental limitation and the authors should discuss this point. As a reviewer, these comments do not mean to take away the progress the paper makes. The authors did a good job at highlighting the potential power of the method and their achievements in the result section. Here it is important to focus on the current**

limitations and trying to provide insights how to improve the absolute performance metrics further so that we can obtain global data that is closer to the true distribution.
Combined with some modifications done in relation to the previous comments #2 and #3 (as these comments relate to the same paragraph (L602-617), the paragraph has been adapted by further focusing the discussion on current limitations and improvement insights.

5. **'Cloud type diagnostics could be a resourceful addition to the panel of assessment methods for model data (Kuma et al., 2023; Kaps et al., 2023) given improvements to achieve remarkable performance in the classification ability as previously described.' It is not clear what the authors mean here. Please clarify.**
   Indeed the sentence is not very clear. The sentence has been re-written to clarify the intended meaning which connects to the review comment #8.

6. **The next paragraph is a great example of a discussion. The emphasis is on increasing training data. 'To improve the spatial coverage of the CloudViT predictions, the direct application to granules from MODIS TERRA would technically not require much more work as the instruments are similar and provide the same cloud properties. An additional benefit would come after the upcoming decommissioning of the CloudSat mission which was providing cloud type retrievals along its track aligned with MODIS. We would then be able to still offer information about cloud types over the same areas even though no vertical information is available and used from our predictions on MODIS level 2 data. As for other satellite cloud products, the main difference would arise, similarly to climate model data, from the potentially different distributions and ranges in the input cloud properties which would need either retraining of the vision transformer or careful scaling to match the distributions seen in MODIS data. Some limitations due to satellite retrieval shortcomings should be taken into account when applying the described method to certain areas. Indeed, since MODIS data is collected through near-nadir scanning, observations in high-latitude regions become oblique, leading to distortions and potential errors in cloud property retrievals, such as cloud top height and optical thickness.'**
   This is very speculative and deviates from the main focus of the study. The current results are relatively performant but not super performant in absolute terms, yet the authors start to speculate on a completely different idea here. I suggest to remove. As I stated before, the authors do a great job at introducing the method, outlining the potential strength of the method, and present their study setup. They should also do a great job at identifying the existing weakness and discussing potential ways to improve like what's been done in the previous paragraph. I wish the authors to continue discussing ways that may cause the low absolute performance (I hope the authors agree with this assessment. If not, please let me know why they think the metrics in the paper are good).
   This part of the discussion has been adapted and the more speculative discussion on application to other satellite data has been removed.

7. **', the input scaling is crucial to ensure proper portability of a method to this other data sourceAs mentioned in more details in Appendix D.' Should be 'to other data source'.**
   Thanks for spotting this, it has been corrected.

8. **It appears that the authors are interested in applying this to global model data. I think that would be a nice move. However, it is imperative to acknowledge that the first step remains to achieve great performance before such exercise. The result is only as meaningful as the capability of the model/method. Please clarify this point in the discussion. Again, if the authors disagree with this point, please help to enlighten me so that I would not insist on it. I want to see the authors to succeed and provide wonderful tool to analyze global simulation data, but at the same time to note that the current results can only be described as in-progress.**

   The relevant discussion sections about the application to global model data (L616-619; L643-657) have been updated to insist further on the need for a performant method beforehand. They have also been aggregated in the same section instead of being spread in the discussion (cf. also review comment #5).