Dear Reviewer,

We gratefully thank you for your constructive and insightful comments, which helped us improve the quality of our manuscript and dataset. Below the comments are answered point-by-point, with the original comments in **black** and our response in **blue**. The revised sentences in the manuscript are indicated in ***italics***.

Overview:

Review of "A Global Classification Dataset of Daytime and Nighttime Marine Low-cloud Mesoscale Morphology Based on Deep Learning Methods" by Wu et al [MS number: essd-2024-536]

This study produces a global dataset of daytime and nighttime low-cloud mesoscale morphologies (categorized into six types) using a convolutional neural network through a combination of MODIS infrared radiance data and machine-learning-retrieved cloud optical thickness. Leveraging this novel dataset, the authors analyzed the day-night contrast in climatology, seasonal cycles, and cloud properties of cloud morphologies. One of notable findings is the significant diurnal variation in the occurrence frequency of closed MCC and suppressed Cu. The primary contribution of this work lies in the generation of nighttime low-cloud morphology data, which complements the well-established daytime morphology datasets from prior studies. This advancement would inspire and enable more downstream research like understanding the diurnal cycle of cloud morphology and cloud-longwave-radiation-climate feedback. The manuscript is overall well-written and well-organized, with nice presentation of figures. However, my major concern pertains to the limitations in the model's training and validation processes, which could impact the dataset's reliability. Addressing these issues would significantly strengthen the study's contribution to the marine low cloud research community. I'd like to recommend a major revision before this manuscript is considered for publication in ESSD.

Thank you for your valuable comments and suggestions. We are pleased that the value of our nighttime dataset has been recognized. As you mentioned, this dataset can be used for various follow-up studies, such as cloud morphology diurnal cycles, the controlling factors of cloud type transitions, and cloud-longwave-radiation-climate feedbacks. Beyond these, we can also investigate the impact of climate change and anthropogenic emissions reductions on the long-term trends of cloud morphology, providing a more comprehensive evaluation of cloud radiative feedbacks associated with type transitions. We hope this research will encourage greater attention to the indirect climate effects of marine low-cloud morphology, which is also the primary purpose of our dataset: to be widely used by the community and to offer valuable insights.

Regarding the limitations you pointed in the model's training and validation process, we fully agree with your concerns. In the revised manuscript, we have further clarified the model's training and validation processes, validated the global applicability of the regionally trained model, and fixed some minor issues. Please see the response below for further details.

At last, your thoughts actually align closely with ours. Many of the suggestions you mentioned, such as investigating the physical reasons behind the diurnal variation of cloud morphology and the impact of morphology transitions on shortwave and longwave radiation at the TOA, are already ongoing within our group. However, since this article is more foundational, many aspects were not fully presented. We look forward to sharing new results with you soon.

Major comments:

1. One of my primary concerns is the validity of applying a regionally trained deep learning (DL) model to global predictions. In this study, the authors developed their model using data from the SEP region only and then applied it to generate a global dataset. While the model demonstrates relatively high prediction accuracy over SEP (Figure 3), it is unclear whether this performance extends to global applications. Regarding this issue, the authors should first clarify the rationale for selecting SEP as the training region rather than using a global or other regional dataset. Was this choice subject to the limited availability of the data, or is there a similarity in morphology climatology between SEP and the global scale? If SEP is your best choice at the moment, it would be essential to evaluate whether using a regionally trained model for global predictions is reasonable. One approach to examine this would be to generate a global map of prediction accuracy for each cloud morphology type to check the model's global performance. Additionally, the authors could examine the differences in the PDFs of thermal radiance, COT, and cloud morphology between SEP and the global dataset. A smaller difference or larger overlaps would indicate less extrapolation by the model, enhancing the credibility of the global dataset.

Similarly, the authors would have to be careful when extending the daytime-trained model to nighttime predictions, as this may also introduce potential extrapolation issues. The authors provided only a single example to illustrate the model's success at nighttime, which is insufficient to establish its statistical reliability. To address this concern, additional cases should be analyzed to validate the model's nighttime performance. Alternatively, examining the differences in the PDFs of thermal radiance, COT, and cloud morphology between daytime and nighttime could help assess the extent of extrapolation and ensure the robustness of the predictions.

We totally agree with your ideas. Applying a model trained on regional datasets to global dataset need careful validation.

First, we give the rationale for selecting the SEP region as the training dataset:

1) We believe that SEP region encompass all the cloud types around the world, and can provide sufficient samples for each type.

2) Meanwhile, we are limited by the available training data. We only labeled this portion of the data initially, and re-labeling the global dataset and regenerating the cloud dataset would take a considerable amount of time.

However, your suggestions provided effective methods for our model validation. Following your suggestion, we generated the plots of probability density functions (PDFs) of COT, thermal radiance, and cloud morphology. One plot compares our training dataset with a global full-year dataset (Figure R1), while the other compares a nighttime dataset (Figure R2).

Both figures show large overlap in the PDFs of COT and thermal radiance. So, we have added some statements in the manuscript to validate the reliability of our model and the nighttime results: "*The representativeness of this dataset was validated as the probability density functions (PDFs) of thermal radiance data and cloud optical thickness show large overlap with those of the global and full-year dataset (Fig. S1)*" (Lines 120-122)

"*In addition, we further examined the differences in the PDFs of the thermal radiance data and the TIR-CNN-based COT between our training dataset (daytime) and nighttime dataset. As depicted in Fig. S4, these PDFs nearly overlapped, which means less extrapolation will be introduced when the model is generalized to nighttime data. And it also illustrates the credibility of our nighttime classification results.*" (Lines 145-148)

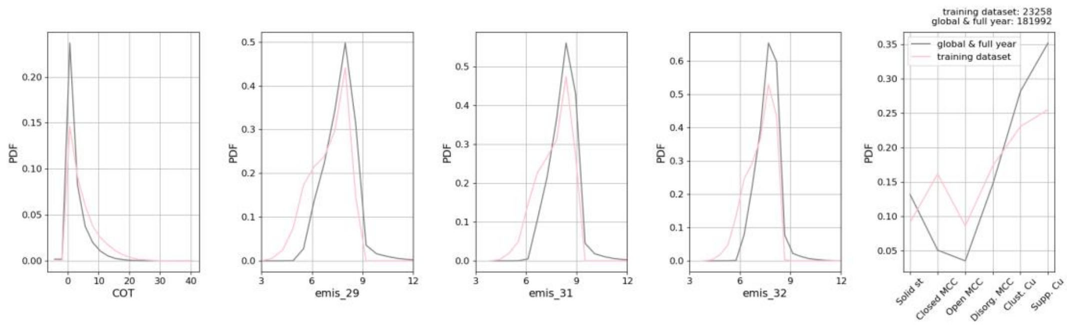

Figure R1. The comparison of probability density functions (PDFs) between our training dataset and a global full-year dataset. (a) PDFs of cloud optical thickness (COT); PDFs of radiance data from infrared channels: (b) 29, (c) 31, (d) 32; (e) PDFs of cloud morphology.

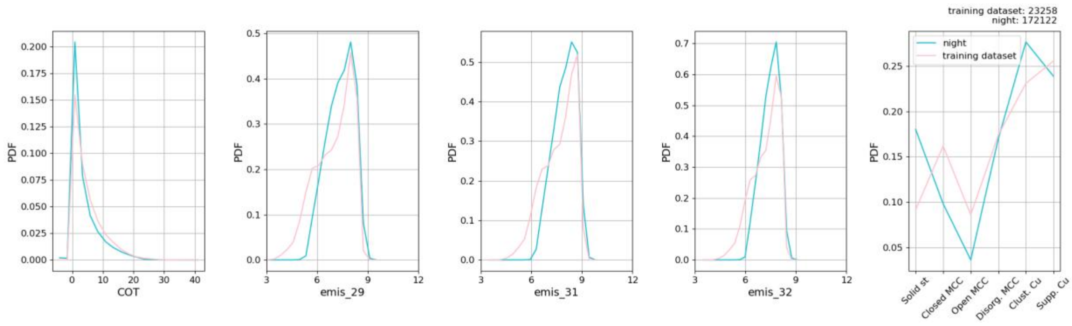

Figure R2. The comparison of probability density functions (PDFs) between our training dataset and a nighttime dataset. (a) PDFs of cloud optical thickness (COT); PDFs of radiance data from infrared channels: (b) 29, (c) 31, (d) 32; (e) PDFs of cloud morphology.

2. Regarding the model training, validation, and testing, the data-splitting strategy is unclear. For instance, was the dataset split randomly or manually into the 6:2:2 ratio? Furthermore, the validation method used to assess the model's predictions has not been described. The authors should clarify these aspects to improve the robustness of their results.

We fully agree that clarifying our dataset splitting method and model validation method is necessary. Indeed, the original annotated dataset was randomly split according to a 3:1:1 ratio, so we clarified it in the manuscript: *"These scenes were then randomly partitioned into three mutually independent datasets for training, validation, and testing, with a distribution ratio of 3:1:1 respectively."* (Lines 131-132)

Based on your suggestions, we have added the following content regarding to the validation method in the "2.4 Method" Section: *"After each training epoch, the validation dataset was used to evaluate the trained model's performance, which allows us to monitor the model's success. When the training process was completed, the test dataset was used for the final evaluation of the model's performance. We used the optimal model to predict each sample in the test dataset, compared the model's predictions with the true labels, and assessed the accuracy using metrics such as accuracy, F1-score, and recall."* (Lines 229-232)

And we have also added the content regarding model test results in Section '3.1 Model Performance': *"Our model achieves an average precision of approximately 91%, an F1-score of 90.6%, and a recall of 90.8%, demonstrating its strong generalization capability and robustness."* (Lines 241-243)

3. Given the critical role of cloud morphologies in Earth's radiation budget, the authors could consider including a climatological analysis of shortwave and longwave radiation at the TOA for the six cloud morphology types. Adding such an analysis would significantly enhance the insights and scientific value of this study.

Thank you for your insightful and constructive suggestions! Your idea aligns perfectly with our thoughts. We are currently conducting a long-term trend study on cloud morphology using this cloud dataset, as well as exploring the impact of aerosols and climate change on these long-term trends. We aim to conduct a systematic study of cloud morphology and cloud type transitions. Based on this, our next step will be to investigate radiative effects. However, some issues are existing in the current radiative datasets: there is a lack of instantaneous clear sky albedo. We need to address some fundamental issues to make the radiative data more solid, which will take time. Moreover, since our cloud scenes are instantaneous rather than monthly, matching the five years of radiative data will be slow, and this could eventually become a separate paper.

In earlier research, Mohrmann et al. (2021) have assessed the radiative properties of the six cloud types, using data from the Clouds and the Earth's Radiant Energy System (CERES), specifically SYN 1-degree hourly data (daytime). They analyzed the net cloud radiative effect (CRE) for each cloud type at different spatial scales and found that Solid MCC and Disorganized MCC exhibit the strongest climatological average cloud radiative effects (Figure R3c, square symbol, about -48 Wm^{-2}). However, they did not account for global radiative impacts associated with long-term changes in cloud morphology, which is the question we aim to address in future work.

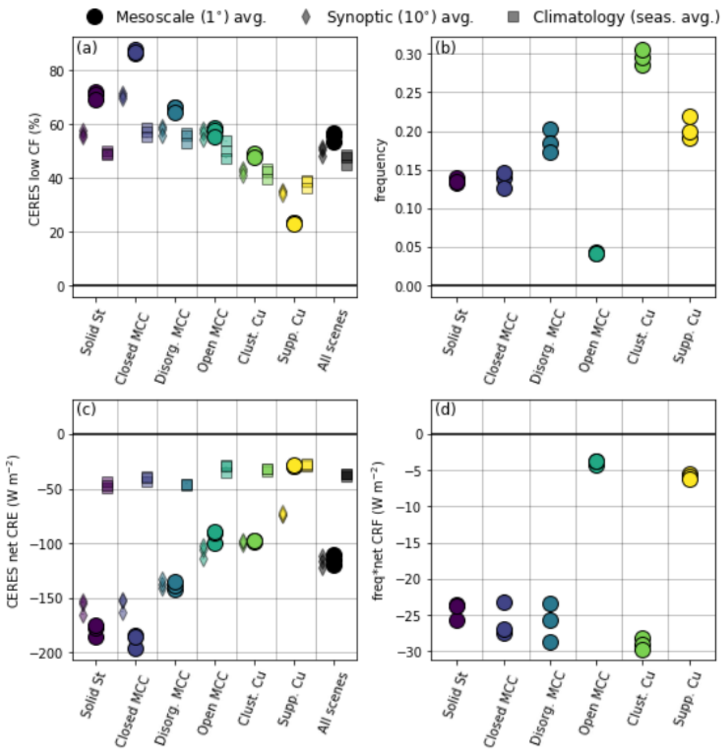

Figure R3. Cloud radiative properties by cloud type from Mohrmann et al. (2021): (a) CERES cloud fraction, (b) cloud frequency of occurrence, (c) average CERES net CRE per cloud type, (d) frequency-weighted net CRE. Each set of three symbols is for the 3 years (2014-2016) used. For panels (a) and (c), the mesoscale, synoptic, climatological averages are shown using circular, diamond, and square symbols, respectively.

Minor comments:

L30: longwave warming effects are more significant for high clouds, which might not be so for low clouds.

The original sentence was changed to: *"They exert a strong radiative cooling on the planet as the residual of a larger cooling effect and a positive warming effect (Klein and Hartmann, 1993; Eytan et al., 2020)."* (Lines 30-32)

L67: What is the major difference between the six-type classification of this study and the four type one here?

The major differences are: (1) Scale: The four types here require a larger scale to be fully observed, at least a 10° by 10° field of view. (2) Classification Basis: Our six-type classification emphasizes the underlying physical processes, while their four-type classification focuses primarily on the external appearance and morphological features.

We have incorporated some key information into the original sentence, and it now reads as follows: *"Moreover, Schulz et al. (2021) developed an object detection model to classify four larger scale ($10^o \times 10^o$) cloud morphologies in trade wind regions of North Atlantic. These morphologies were vividly named as "sugar," "gravel," "flowers," and "fish" mainly based on their visual appearances." (Lines 66-69)*

L81: Do you mean the decline in the *long-term* trend?

Yes, fixed.

L83: "how much they contribute to … remain unclear" to "how nighttime cloud cover varies under different cloud morphology types remain unclear."

Done.

L90: Please clarify the temporal and spatial resolution.

Done.

L97: "created" to "driven"

Done.

L119: Please clarify the temporal resolution of the training dataset.

Done.

L121-122: Have you excluded middle clouds (i.e., those situated between 3 and 6 km)? These clouds are prevalent over midlatitude oceans, and they also contaminate low cloud observations.

We are sorry for not eliminating the influence of middle clouds in our training process. In future work, we plan to refine our methodology by re-screening middle clouds to improve the accuracy of our model.

L174: Please clarify the level of the divergence used.

Done.

L199: I'd suggest labeling the input variables (three channels and COT) and the output variables (six cloud morphology types) in Figure 2a to improve its clarity and readability.

Done.

L210: It looks like the improvement is limited. Have you examined the COT retrieval uncertainty? If it is greater than the improved accuracy, it would be unnecessary to include the COT into the predictors.

Thank you for your insightful suggestion. The result of a sensitivity experiment indicates that our model based on retrieved COT demonstrates identical predictive accuracy to the model trained on daytime MODIS COT (Figure R4), which confirms the reliability of COT retrieval product.

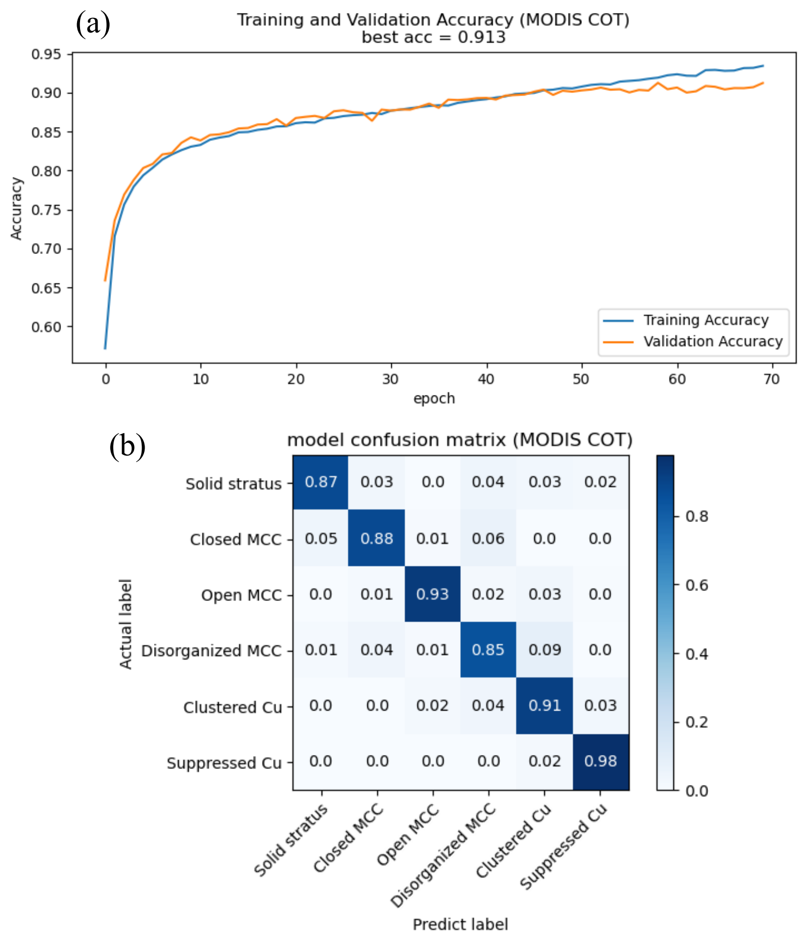

Figure R4. Model training results based on MODIS COT. (a) the model's accuracy on the training and validation datasets. (b) the confusion matrix of the model.

L210: Typo: “Yuan et al (2020)_due to” to “Yuan et al (2020) due to”

Fixed.

L212: Which is it relative to?

We have completed the sentence as follows: *“ Although it is a bit less accurate compared to the visible light model from Yuan et al. (2020), it is undeniable that this model has achieved a relatively high accuracy level when compared to other TIR model (Lang et al., 2022), and can effectively accomplish the classification tasks we proposed.”* (Lines 237-240)

L219: “clustered Cu” to “clustered Cu or closed MCC”

Done.

L250: “n denotes” to “with n denoting”

Done.

L305: “its seasonal variation” to “the peak in summer”

Done.

L332: do you mean “decrease by 2 microns on average”?

Yes, fixed.

L333: Please clarify whether the LWP mentioned here represents the in-cloud value or the grid box mean value.

Sorry for the lack of information provided. We have added the following clarification in Lines 356-357: *“All of the cloud microphysical properties represent the in-cloud mean value within a 1°×1° grid.”*

L349: Why is there a westward shift at night? Also, for stratocumulus clouds, LTS is usually higher at night. Why does it decline for closed MCC at night?

Sorry for the confusion. We originally hypothesize that the observed decline in LTS at night may be related to the movement of clouds. Since our statistical analysis covers a relatively large spatial scale, the meteorological conditions associated with specific cloud types could change as the clouds move, potentially leading to the observed decrease in LTS. However, we apologize for the previous implication that this movement is necessarily westward. So, we removed the mention of "westward" in the revised text. We believe that further investigation is required to fully understand the reasons behind the LTS decline in closed MCCs at night.

L351: It would be more interesting to discuss their physical reason.

Thank you for pointing the direction for us, it is the next part of our work. It is difficult to determine the key cloud-controlling factors for each cloud type based solely on the statistical analysis in this study, and many other environmental conditions have not been included. Therefore, we plan to conduct a more detailed and comprehensive investigation into their physical causes in our next work. Furthermore, since the main focus of this article is to introduce a cloud dataset, we feel that including an analysis of cloud-controlling factors here would dilute the main theme.

L367: Why are the results shown here only for SEP, while Figure 10 presents global results?

Cloud morphology is controlled by multiple meteorological factors (Liu et al., 2024). When we study the influence of one controlling factor, cloud morphologies can be affected by the variability in other factors if the study is conducted in a global scale. For instance, while exist within the same LTS environment, clouds in mid-to-high latitude regions and those near the equator have distinctly different sea surface temperatures. Therefore, by restricting the region, we can facilitate the day-night comparison of the primary controlling factors while excluding the interference from other variables. Clouds properties are the final result of all meteorological conditions, and their properties show little differences across different regions.

References

Eytan, E., Koren, I., Altaratz, O., Kostinski, A. B., and Ronen, A.: Longwave radiative effect of the cloud twilight zone, Nature Geoscience, 13, 669-673, https://doi.org/10.1038/s41561-020-0636-8, 2020.

Klein, S. A. and Hartmann, D. L.: The Seasonal Cycle of Low Stratiform Clouds, Journal of Climate, 6, 1587-1606, https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2, 1993.

Lang, F., Ackermann, L., Huang, Y., Truong, S. C. H., Siems, S. T., and Manton, M. J.: A climatology of open and closed mesoscale cellular convection over the Southern Ocean derived from Himawari-8 observations, Atmos. Chem. Phys., 22, 2135-2152, https://doi.org/10.5194/acp-22-2135-2022, 2022.

Liu, J., Zhu, Y., Wang, M., and Rosenfeld, D.: Cloud Susceptibility to Aerosols: Comparing Cloud-Appearance vs. Cloud-Controlling Factors Regimes, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-4059, https://doi.org/10.5194/egusphere-egu24-4059, 2024.

Mohrmann, J., Wood, R., Yuan, T., Song, H., Eastman, R., and Oreopoulos, L.: Identifying meteorological influences on marine low-cloud mesoscale morphology using satellite classifications, Atmos. Chem. Phys., 21, 9629-9642, https://doi.org/10.5194/acp-21-9629-2021, 2021.

Schulz, H., Eastman, R., and Stevens, B.: Characterization and Evolution of Organized Shallow Convection in the Downstream North Atlantic Trades, Journal of Geophysical Research: Atmospheres, 126, e2021JD034575, https://doi.org/10.1029/2021JD034575, 2021.

Yuan, T., Song, H., Wood, R., Mohrmann, J., Meyer, K., Oreopoulos, L., and Platnick, S.: Applying deep learning to NASA MODIS data to create a community record of marine low-cloud mesoscale morphology, Atmos. Meas. Tech., 13, 6989-6997, https://doi.org/10.5194/amt-13-6989-2020, 2020.