Training and Validation of ORAC Retrievals and Cloud Masking for FCI using EarthCARE L2 Data

Daniel Robbins [1,2,†], Gareth Thomas [1,2], Dora Hegedus [1], Elisa Carboni [1,2]

[1] RAL Space, Harwell Campus, UK [2] NCEO, UK

[†] daniel.robbins@stfc.ac.uk

1. Introduction

The Optimal Retrieval of Aerosol and Cloud (ORAC) algorithm uses optimal estimation to retrieve cloud and aerosol macro- and micro-physical properties from passive satellite imagers [McGarragh et al., 2018; Thomas et al., 2009]. The latest European geostationary satellite imager, called the Flexible Combined Imager (FCI), on-board the Meteosat Third Generation Imager-1 (MTG-I1) satellite provides full-disk data for 16 channels at up to 0.5km spatial resolution every 10 minutes. For this study, the ORAC cloud retrieval has been applied to FCI data which has been collocated with EarthCARE L2 data. As the EarthCARE L2 data is derived from active instruments, it is treated as truth and used to validate the ORAC retrievals. In addition, previous studies have demonstrated that collocating data from active instruments with passive instruments can provide effective training and validation datasets for machine learning algorithms [Min et al., 2020; Poulsen et al., 2020]. These can be trained to identify clouds and aerosols, cloud phase and physical properties, which can then be used as *a priori* information for the ORAC algorithm.

2. Collocating Data

EarthCARE L2 AC__TC__2B and ACM_CAP_2B products (baseline BC) from within the FCI field-of-view (FOV) across December 2025 and January 2026 are collocated on a pixel-by-pixel basis with FCI and regridded ERA5 data. Each identified layer in the EarthCARE profiles are parallax corrected to ensure accurate collocation with FCI pixels. As several EarthCARE profiles can be collocated with the same FCI pixel, the closest matching profile per FCI pixel is kept and the rest discarded in this analysis. The data from December 2025 is also collocated with ORAC cloud retrievals.

3. Developing Neural Networks

Several neural networks (NNs) have been developed for FCI using the collocated EarthCARE data as truth. A simple feed-forward NN cloud mask has been developed with ~93% accuracy from the validation set, while another feed-forward NN has been trained on only cloudy pixels to identify phase, with an accuracy of ~85%. This lower accuracy is likely due to the limit of FCI to identify phase compared with the EarthCARE AC__TC__2B product. Another set of gaussian-loss NNs, which provide a mean and uncertainty value, have been trained on cloudy pixels to find cloud top height (CTH) and cloud optical thickness (COT). An example of all these NNs can be seen in Fig. 1 for 7th October 2025 at 10:10 UTC, which is outside the training time.

4. Validating ORAC and Neural Networks

ORAC cloud retrievals and NN outputs are compared to EarthCARE for cloudy pixels (as identified by EarthCARE) in Fig. 2, and an example retrieval is shown in Fig. 3. The collocated data is filtered to better match the passive and active retrievals, with only single-layer clouds with COT > 0.2 considered. In addition, COT values are only compared for FCI solar zenith angles < 70°, i.e. under daylight conditions. CTH values from both ORAC and the GLNN match the EarthCARE values well, with ORAC having a bias due to underestimation of CTH for higher clouds, while the GLNN seems to struggle with lower, potentially boundary layer height, clouds. The ORAC COT values do not match EarthCARE as well, although this may be due to a limitation of the ACM_CAP_2B product, with both ORAC and the GLNN values having a large value of RMSE.

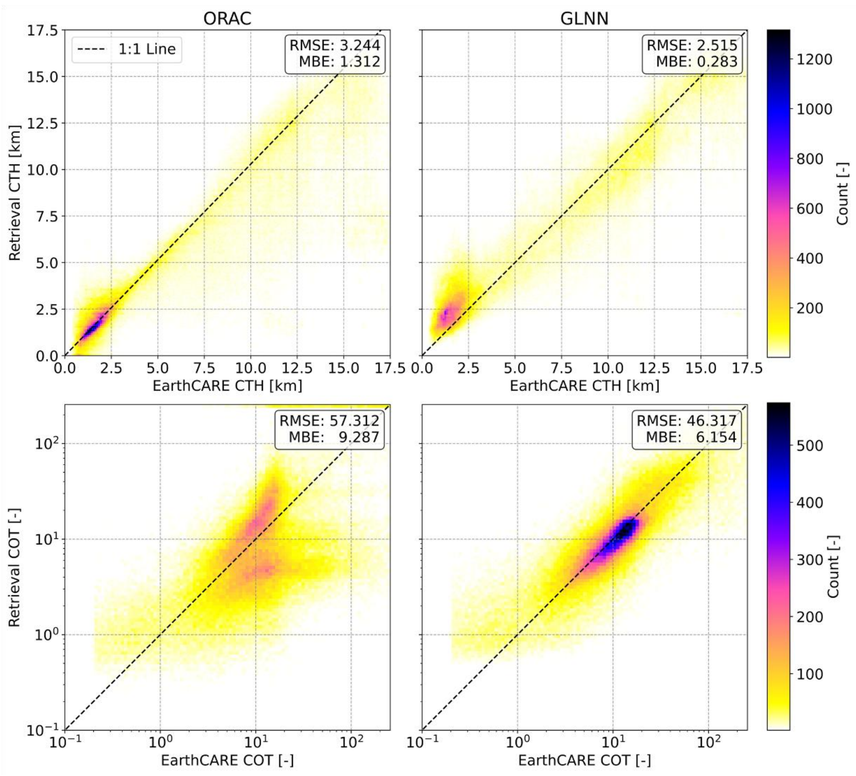

Figure 2: Heatmaps of (left column) ORAC and (right column) Gaussian-Loss NNs (GLNNs) versus EarthCARE values for (upper row) cloud top height and (lower row) cloud optical thickness. Counts are normalised per row in the figure. Root mean squared error (RMSE) and mean bias error (MBE) are given per panel. COT values are for day only, i.e. solar zenith angles > 70°.

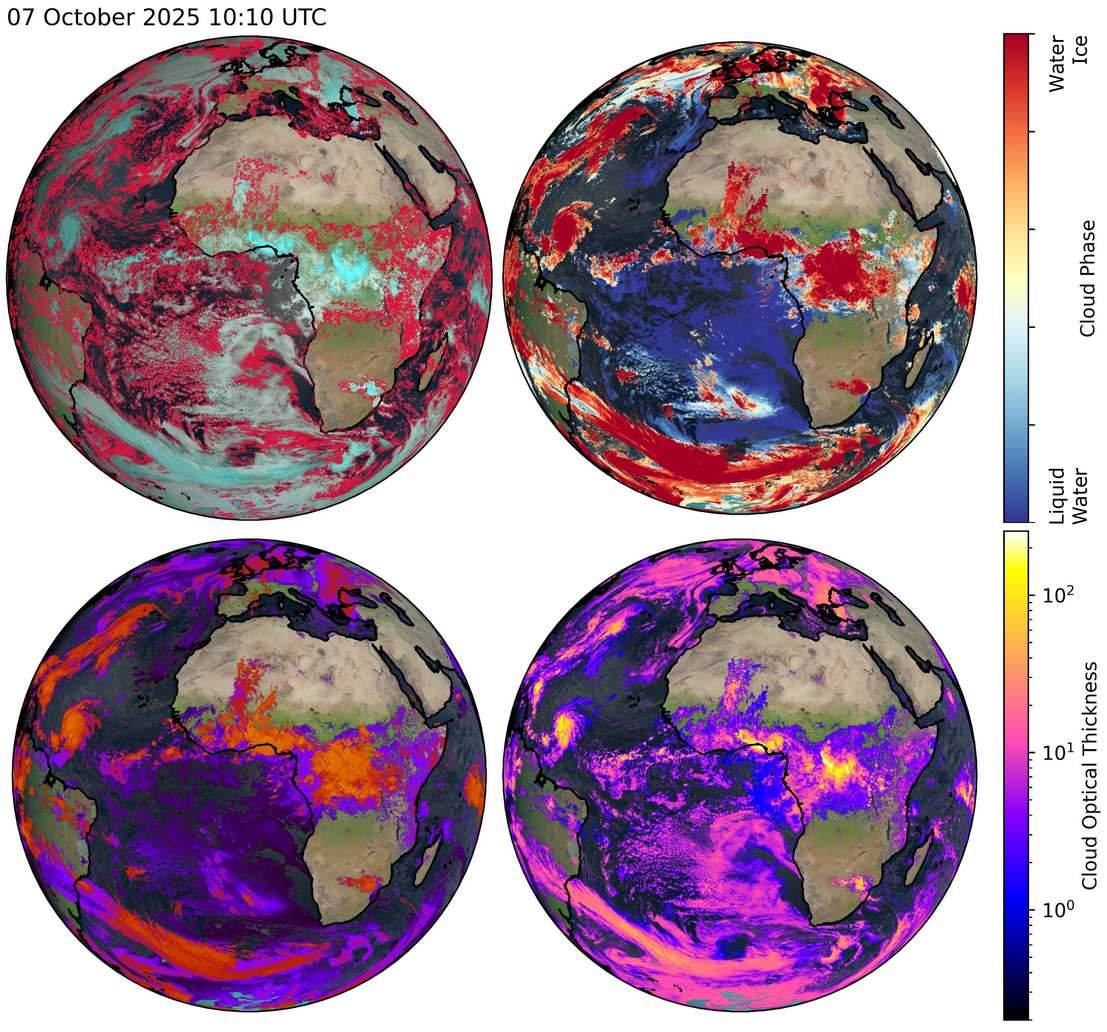

Figure 1: An FCI scene from 7th October 2025 at 10:10 UTC with natural colour images and (upper left) the cloud mask outlining clouds in red, (upper right) the cloud phase for the identified cloud pixels, (lower left) the cloud top height from the NN and (lower right) the cloud optical thickness from the NN.

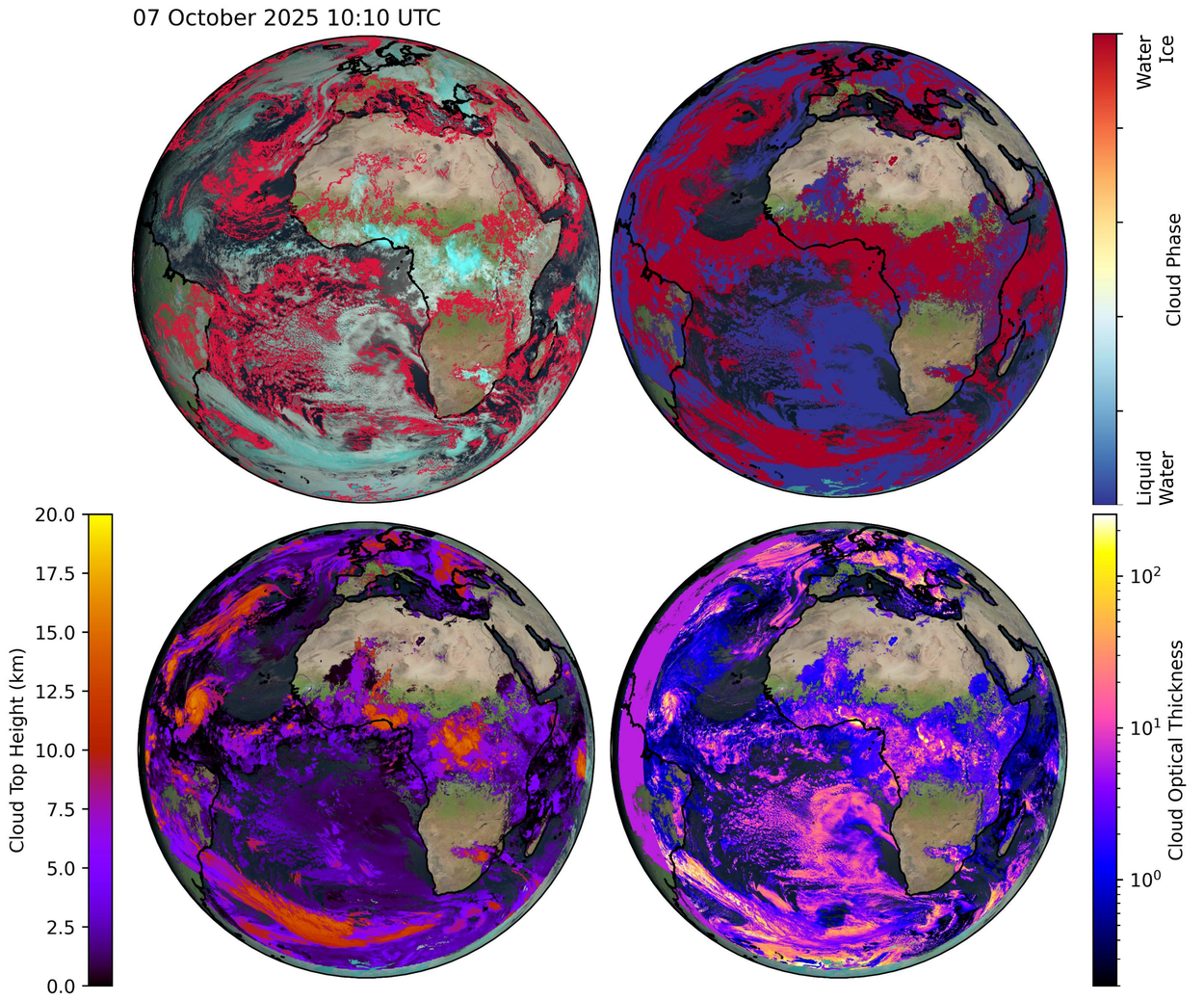

Figure 3: The same FCI scene and overall format as Figure 1, but for the ORAC retrieval.

5. Conclusions

FCI data has been collocated with EarthCARE L2 products and used to train a range of neural networks (NNs) for clouds in FCI, as well as validate ORAC retrievals. The NNs achieve a relatively high level of accuracy for cloud classification and compare well to EarthCARE values for cloud top height (CTH) and optical thickness (COT). ORAC retrievals of CTH compare well with EarthCARE, but show a bias for higher clouds, whilst ORAC COT values have a large bias and spread, although this may be due to the upper limit of the active instruments used in the EarthCARE ACM_CAP_2B product.

Acknowledgements

This work was funded by the National Centre for Earth Observation (NCEO) at RAL Space. FCI L1c data is provided by EUMETSAT and can be accessed via the EUMETSAT data store. EarthCARE L2 data is provided by ESA and can be accessed via the MAAP system.

References

McGarragh, G.R. et al. (2018) 'The Community Cloud retrieval for CLimate (CC4CL) – Part 2: The optimal estimation approach', Atmospheric Measurement Techniques, 11(6), pp. 3397–3431. Available at: https://doi.org/10.5194/amt-11-3397-2018.
Thomas, G.E. et al. (2009) "The GRAPE aerosol retrieval algorithm," Atmospheric Measurement Techniques, 2(2), pp. 679–701. Available at: https://doi.org/10.5194/amt-2-679-2009.
Min, M. et al. (2020) "Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms," Remote Sensing of Environment, 239, p. 111616. Available at: https://doi.org/10.1016/j.rse.2019.111616.
Poulsen, C. et al. (2020) "Evaluation and comparison of a machine learning cloud identification algorithm for the SLSTR in polar regions," Remote Sensing of Environment, 248, p. 111999. Available at: https://doi.org/10.1016/j.rse.2020.111999.