
LEVERAGING SYNTHETIC DATA AND MACHINE LEARNING FOR CLOUD OPTICAL THICKNESS ESTIMATION

Aleksis Pirinen¹, Nosheen Abid², György Kovács², Ronald Scheirer³, Chiara Ceccobello⁴, Nuria Agues Paszkowsky¹, Thomas Ohlson Timoudas¹, Anders Persson⁵, Marcus Liwicki²

¹ RISE Research Institutes of Sweden; aleksis.pirinen@ri.se

² Luleå University of Technology, Machine Learning, Sweden; nosheen.abid@ltu.se

³ Swedish Meteorological and Hydrological Institute; ronald.scheirer@smhi.se

⁴ AI Sweden

⁵ The Swedish Forest Agency

ABSTRACT

Cloud formations often obstruct the effectiveness of optical satellite monitoring, imposing limitations on Earth observation (EO) tasks such as land cover mapping, ocean color analysis, and cropland monitoring. While machine learning (ML) methods have improved EO tasks, challenges persist, primarily the dependence on annotated data for ML training, especially in EO contexts like cloud optical thickness (COT) estimation. To address the scarcity of COT data, we propose a synthetic dataset simulating top-of-atmosphere radiances for 12 spectral bands of the MSI sensor on Sentinel-2 platforms, and encompassing various cloud types, COTs, and environmental conditions. Extensive experimentation on training ML models to predict COT from spectral band reflectivities demonstrates the utility of the proposed dataset. Generalization to cloud cover mapping on real data is verified on two satellite image datasets. The data, code and models have been made available at <https://github.com/aleksisp/ai/ml-cloud-opt-thick>.

1 INTRODUCTION

Space-based EO significantly enhances data collection worldwide, encompassing not only atmospheric parameters from remote locations but also underlying layers. Examples include land use and cover classification (Abid et al., 2021b), damage assessment for natural disasters (Mateo-Garcia et al., 2019; Abid et al., 2021a), biophysical parameter retrieval (Camps-Valls et al., 2006), urban growth monitoring (Gomez-Chova et al., 2006), and crop yield estimation (Wolanin et al., 2020). These applications rely heavily on satellite optical sensors, and cloud coverage poses a hindrance to signal exploitation (Gómez-Chova et al., 2007). Precise cloud cover estimation is thus crucial for maintaining the effectiveness of these applications.

Cloud detection and estimation in multi-spectral images (MSI) encompass rule-based statistical methods and advanced deep learning approaches. Rule-based thresholding, exemplified by FMask (Zhu et al., 2015) and Sen2Cor (Louis et al., 2016), utilizes physical cloud properties across different spectral bands for cloud masking in Landsat and Sentinel-2 MSI. A recent surge in the literature has applied machine learning (ML) approaches to tackle the challenge of cloud detection and estimation (Mateo-García et al., 2020; Zhang et al., 2020; Kanu et al., 2020; Jeppesen et al., 2019). These methods, which employ CNNs coupled with extensive manually annotated satellite image datasets, outperform traditional rule-based techniques.

Clouds exhibit natural inhomogeneity, and their spatial inhomogeneity, particularly in *cloud optical thickness* (COT), impacts both remote sensing imagery and atmospheric radiations. The prevalent method for COT estimation is independent pixel analysis (IPA), which assumes homogeneity within a pixel and lacks information about neighboring pixels. Statistical approaches (Liu, 1995; Zinner and Mayer, 2006; Iwabuchi and Hayasaka, 2002) have explored factors influencing COT, and introduced parameters to mitigate their effects. Deep learning advancements (Krizhevsky et al., 2012; He et al.,

2016) have led to neural network-based COT approaches (Okamura et al., 2017; Sde-Chen et al., 2021), which typically require neighboring pixel information and substantial annotated spatial data.

Both statistical and ML-based methods require benchmark datasets for performance evaluation and improvement. Each pixel in the satellite images must typically be labeled manually or by a superior instrument (e.g. an active LIDAR (Håkansson et al., 2018)) as cloudy or cloud-free (or using even finer-grained labels). Cloud complexity intensifies the task, as it often requires time-consuming expert labeling, which results in limited publicly available datasets. Among the few available benchmarking datasets, we have chosen to consider the KappaZeta dataset (Domnich et al., 2021); see §4.

In this work we introduce a new IPA dataset simulating top-of-atmosphere (ToA) radiances for 12 MSI spectral bands.¹ Acknowledging the systematic error introduced by IPA due to 3D effects (Zinner et al., 2010), particularly for high spatial resolution applications, we recognize the value of derived COT for flexible data use, such as clear-conservative vs cloud-conservative cloud masks. Generated using RTTOV v13, the dataset comprises 200,000 data points. Our analysis of various ML models for cloud detection and COT estimation on this dataset suggests that multi-layer perceptron (MLP) outperform alternatives, such as linear regression. Furthermore, we evaluate our models on two real satellite image datasets – KappaZeta as well as a national forest agency dataset that we introduce in this work – demonstrating their generalization to cloud cover mapping on real-world data.

2 SYNTHETIC CLOUD OPTICAL THICKNESS DATASET

Our novel synthetic dataset for COT estimation is created by connecting surface and atmospheric properties to ToA reflectivities observed by the MSI on Sentinel-2 satellites. The dataset is generated using the fast Radiative Transfer for TOVs (RTTOV) v13 model (Saunders et al., 2018). The simulations rely on atmospheric profiles, randomly chosen from a dataset of 10,000 profiles provided by the European Centre for Medium-range Weather Forecasts (ECMWF) (Chevallier et al., 2006). These profiles are used to represent diverse atmospheric conditions. The surface is treated as a Lambertian reflector (Koppal, 2014). Spectral reflectance for soil, rocks, and vegetation are obtained from the ECOSTRESS spectral library (Baldrige et al., 2009; Meerdink et al., 2019), with surface types chosen randomly. The instrument-specific reflection is derived by convolving the spectral surface reflection with the spectral response function of the corresponding MSI channel.

The dataset comprises 200,000 simulated data points resembling individual pixels observed by a satellite instrument, covering various cloud types, COTs, ground surface characteristics, and atmospheric profiles. It is divided into four equal-sized parts, with one dedicated to clear sky and the remaining three featuring water clouds, ice clouds, and mixed clouds. Each data point is comprising simulated measured reflectivities for 12 spectral bands, along with optional features such as satellite zenith angle, sun zenith angle, azimuth difference angle, gas optical thickness, vertically integrated water vapour, and surface profile. The ground truth for each data point includes the non-negative COT and cloud type. This study focuses on using the 12 band reflectivities as ML model input, as they are least affected by aerosols. The ground truth COT is within the range $[0, 50]$, where the radiative transfer model determines the upper limit. For the training, validation, and testing of ML models, we randomly allocate 160,000 data points for training, 20,000 for validation, and 20,000 for testing.

3 MACHINE LEARNING MODELS

This section offers a concise overview of the ML models used for COT estimation. Given the dataset’s pixel-level format with no spatial relationships, we predominantly utilized MLP models. Preliminary experiments explored various approaches, including Random Forests, but MLPs demonstrated superior performance. Extensive model validations indicated that a five-layer MLP with ReLU activations and a hidden dimension of 64 (consistent across all layers) yielded the optimal results. The proposed model operates on a per-pixel basis due to the limitations in the synthetic dataset. So, to enhance spatial consistency when working with real imagery $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$, we employ a simple post-processing trick on the resulting COT prediction map $\mathbf{C} \in \mathbb{R}^{H \times W}$. This involves a sliding window of size $M \times M$ with a stride of 1 over \mathbf{C} , computing an average among M^2 values at each location. We found $M = 2$ to produce robust results.

¹The aerosol (B1) band is not included, as it vastly increases computational complexity to the simulation.

Model training. We perform standardization on the input, applying channel-wise zero mean and unit variance based on training data statistics, as more sophisticated normalization techniques did not yield improvements. The ML models are trained using the Adam optimizer (Kingma and Ba, 2014) with an L2-loss for 2×10^6 batch updates, employing batch size 32 and learning rate 3×10^{-4} . To improve model robustness, independent Gaussian noise is introduced to each input, characterized by a zero-mean and a standard deviation equivalent to 3% of the average magnitude of each input feature. The models are very lightweight and training can be done without a GPU in approximately an hour.

Fine-tuning with weaker labels. The KappaZeta dataset has pixel labels for 'clear', 'opaque cloud', and 'semi-transparent cloud', respectively. Let τ^{semi} and τ^{opaque} (where $0 < \tau^{\text{semi}} < \tau^{\text{opaque}}$) denote the semi-transparent and opaque COT thresholds, respectively. Then, we refine a model using a loss \mathcal{L} which satisfies the following criteria. If p denotes the prediction of a pixel which is labeled as

- 'clear', then

$$\mathcal{L}(p) = \begin{cases} 0 & \text{if } p \leq \tau^{\text{semi}}, \\ \frac{1}{2}(p - \tau^{\text{semi}})^2 & \text{if } p > \tau^{\text{semi}}, \end{cases}$$

- 'opaque cloud', then

$$\mathcal{L}(p) = \begin{cases} 0 & \text{if } p \geq \tau^{\text{opaque}}, \\ \frac{1}{2}(p - \tau^{\text{opaque}})^2 & \text{if } p < \tau^{\text{opaque}}, \end{cases}$$

- 'semi-transparent cloud', then

$$\mathcal{L}(p) = \begin{cases} 0 & \text{if } \tau^{\text{semi}} \leq p \leq \tau^{\text{opaque}}, \\ \frac{1}{2}(p - \tau^{\text{semi}})^2 & \text{if } p < \tau^{\text{semi}}, \\ \frac{1}{2}(p - \tau^{\text{opaque}})^2 & \text{if } p > \tau^{\text{opaque}}. \end{cases}$$

4 EXPERIMENTAL RESULTS

In this section we extensively evaluate various ML models for COT estimation, as well as COT-based cloud masking in real satellite imagery. The models include: **(i) MLP-k**, a k -layer MLP (the main model has $k = 5$), where results are averaged among 10 identical k -layer MLPs, differing only in the random seed for network initialization; **(ii) MLP-k-ens-n**, an ensemble of n k -layer MLPs, each identically trained but with a unique random network initialization; **(iii) MLP-k-no-noise** is equivalent to $MLP-k$ but trained without additional noise in the training data (note that default training includes 3% additive noise); and **(iv) lin-reg**, a linear regression model.

Results on the proposed synthetic dataset. Table 1 reports the mean absolute error (MAE) between the predicted outputs and corresponding ground truths on unseen test data. Training on data with artificially added noise improves model robustness significantly. Ensembling only marginally improves performance. MLPs significantly outperform linear regression models.

Results on KappaZeta. The publicly available KappaZeta dataset (Domnich et al., 2021) is used to assess our models on real satellite imagery. Due to the absence of COT values, we focus on pixel categories 'clear', 'semi-transparent cloud' and 'opaque cloud'. A subset is set aside (*April, May*,

Table 1: MAE values on different variants of the test set of our synthetic dataset. *Test-x%* refers to the test set with $x\%$ added noise. Ensemble methods marginally improve over single-model ones. Models trained with additive input noise yield significantly better average results. Linear regression performs worst by far. Note that for the single-model variants MLP-5, MLP-5-no-noise and Lin-reg, we show the mean MAE over 10 different network parameter initializations and standard deviations.

	MLP-5	MLP-5-ens-10	MLP-5-no-noise	MLP-5-no-noise-ens-10	Lin-reg
Test-0%	1.63 ± 0.01	1.56	1.05 ± 0.01	0.92	6.49 ± 0.00
Test-1%	1.68 ± 0.01	1.61	1.59 ± 0.01	1.46	6.51 ± 0.00
Test-2%	1.82 ± 0.01	1.75	2.51 ± 0.01	2.34	6.55 ± 0.00
Test-3%	2.04 ± 0.01	1.97	3.42 ± 0.02	3.20	6.63 ± 0.00
Test-4%	2.32 ± 0.01	2.25	4.21 ± 0.03	3.95	6.74 ± 0.01
Test-5%	2.63 ± 0.01	2.56	4.90 ± 0.04	4.58	6.88 ± 0.01
Average	2.02 ± 0.01	1.95	2.95 ± 0.02	2.74	6.63 ± 0.00

Table 2: Test set results on KappaZeta. Our top MLP approach (MLP-5-ens-10-tune) exhibits lower performance compared to a U-net, as U-net integrates information from neighboring pixels, a capability inherently lacking in MLPs. Among the MLP approaches, fine-tuning on the KappaZeta training set enhances results (column 4 vs 3).

Metric	MLP-5-tune	MLP-5-ens-10	MLP-5-ens-10-tune	U-Net
F1-avg	0.51	0.49	0.52	0.66
F1-clear	0.54	0.53	0.54	0.72
F1-semi	0.30	0.25	0.31	0.49
F1-opaque	0.70	0.70	0.71	0.78
mIoU	0.46	0.43	0.47	0.54
IoU-clear	0.63	0.60	0.63	0.62
IoU-semi	0.20	0.15	0.21	0.35
IoU-opaque	0.54	0.54	0.56	0.65

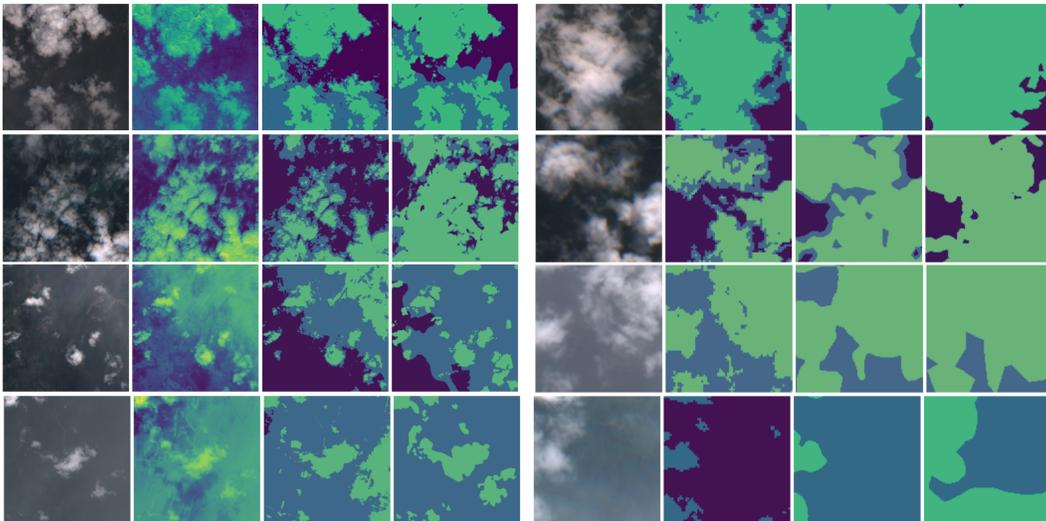


Figure 1: **Left:** Examples of our main MLP approach (ensemble of ten 5-layer MLPs) on unseen KappaZeta test data. Column C1: Input image (only RGB is shown). C2: COT estimates (relative intensity scaling, to more clearly show the variations). C3: Pixel-level cloud type predictions based on thresholding the COTs in column 2. C4: KappaZeta ground truth. Dark blue is clear sky, lighter blue is semi-transparent cloud, and turquoise is opaque cloud. **Right:** Similar to the left, but the 2nd column shows the thresholded model predictions (instead of the COT estimates), and the 3rd column is the U-net prediction. A failure case for both models is shown on the bottom row.

June; 3543 images) to find COT thresholds corresponding to semi-transparent and opaque clouds. These thresholds are respectively found to be 0.75 and 1.25 on the KappaZeta training data.

The MLP models, pretrained on synthetic data, are evaluated using the KappaZeta dataset with 2511 from the *July*, *August*, and *September* subsets. In Table 2, MLP-5-ens-10-tune obtains a lower mIoU and F1-score compared to the U-net (Wada, 2017) (0.47 vs 0.54 and 0.52 vs 0.66, respectively). The expected superiority of U-net is due to its incorporation of spatial connectivity in images inherently lacking in MLPs. Fine-tuning on the KappaZeta training set improves test set results (column 4 vs column 3), benefiting from domain consistency. Additionally, model ensembling only marginally improves results (column 4 vs column 2).

Qualitative results. We further examine qualitative results in Fig. 1. The left side focuses on the MLP results with estimated COT values (column 2), while the right side features comparisons to the U-net. We can see that the ground truth segmentation masks exhibit a bias towards spatial connectivity among pixels. This bias emphasizes spatial consistency over finer-grained contours, leading to potential inaccuracies. The U-net architecture, capable of leveraging spatial connectivity, accommodates this bias effectively. However, per-pixel models, such as the MLPs

Table 3: Results on test data from the Swedish Forest Agency (SFA). It is worth noticing that the various MLP approaches were *not* trained on any SFA data. Our MLP-5-ens-10 model yields results comparable to ResNet18-clc (which was trained on the SFA training set), and it outperforms the SCL by ESA. Note that MLP-5 represents the average result of ten MLP-5 models that were trained using different network initializations, whereas MLP-5-ens-10 is an ensemble of those ten different models.

Metric	MLP-5 (ours)	MLP-5-ens-10 (ours)	ResNet10-clc	ESA-SCL
F1-avg	0.73	0.88	0.90	0.68
F1-clear	0.81	0.94	0.94	0.88
F1-cloudy	0.65	0.82	0.86	0.48

in our case, may face challenges due to this bias. Interestingly, in some instances, the MLP predictions visually appear more accurate than the ground truth when compared with the input images.

Results on data from the Swedish Forest Agency (SFA). The Swedish Forest Agency (SFA) is a national authority in charge of forest-related issues in Sweden. Their main function is to promote management of Sweden’s forests, enabling the objectives of forest policies to be attained. Among other things, the SFA runs change detection algorithms on satellite imagery, e.g., to detect if forest areas have been felled. For this, they rely on ESA’s scene classification layer (SCL), which also includes a cloud probability product. The SFA’s analyses require cloud-free images, but the SCL layer is not always accurate enough. Therefore, we applied models that were trained on our synthetic dataset on the SFA’s data in order to classify a set of images as ‘cloudy’ or ‘clear’.

To achieve this, the SFA provided 432 Sentinel-2 Level 2A images of size 20×20 (corresponds to $200 \times 200 \text{ m}^2$) that they had labeled as cloudy or clear (120 cloudy, 312 clear), where an image was labeled as clear if no pixel was deemed to be cloudy. Note that the cirrus (B10) band was not included, so when working with this dataset, we re-trained our MLP models on the synthetic dataset after excluding this band from that data. The 432 images were randomly split into a training, validation, and test split, such that the respective splits have the same ratio between cloudy/clear images (i.e., roughly 28% cloudy and 72% clear images per split). The training, validation, and test sets contain 260, 72 and 100 images, respectively.

The results on the test set are shown in Table 3. For our MLPs, we use the validation data to set a COT threshold above which a pixel is predicted as cloudy (0.5 was found to be best). For the SCL, a pixel is predicted to be cloudy if the SCL label is ‘cloud medium probability’, ‘cloud high probability’, or ‘thin cirrus’. For both approaches, if a pixel is predicted as cloudy, the overall image is predicted as cloudy. We also compare with a ResNet-18 model He et al. (2016) for binary image classification (cloudy or clear), trained on the union of the training and validation sets. Left-right and bottom-up flipping were used for data augmentation. From Table 3, we see that our main model (an ensemble of ten five-layer MLPs) is on par with the dedicated classification model, *despite not being trained on a single SFA data point* (except for COT threshold tuning), and despite the fact that the synthetic training data represents top-of-atmosphere data (i.e., Level 1C, not Level 2A as the SFA data). We also see that model ensembling is crucial (MLP-5-ens-10 vs. MLP-5), and that our MLP-5-ens-10 model significantly outperforms the SCL (average F1-score 0.88 vs. 0.68 for the SCL). Finally, we note that even using only a single MLP-5 model yields a higher average F1-score than the SCL.

5 CONCLUSIONS

In this work, we have introduced a novel synthetic dataset that can be used to train models for predicting cloud types of pixels (e.g., clear, semi-transparent, and opaque clouds) via thresholding cloud optical thicknesses (COTs). Several ML approaches were explored for this dataset, and it was found that ensembles of MLPs perform best. Despite our proposed synthetic dataset (and thus associated models) being inherently pixel-independent, the models show promising results on real satellite imagery. In particular, our MLP approach can seamlessly transition to real datasets without additional training. To showcase this, we directly applied (without further training) this MLP approach to the task of cloud classification on a novel real image dataset and achieved an F1 score that is on par with a ResNet model that was explicitly trained on the target dataset. Furthermore, our approach is superior to the ESA scene classification layer at classifying satellite imagery as clear or cloudy, and can flexibly generate cloud type segmentation masks via COT thresholding.

REFERENCES

- N. Abid, M. I. Malik, M. Shahzad, F. Shafait, H. Ali, M. M. Ghaffar, C. Weis, N. Wehn, and M. Liwicki. Burnt forest estimation from sentinel-2 imagery of australia using unsupervised deep learning. In *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pages 01–08. IEEE, 2021a. doi: 10.1109/DICTA52665.2021.9647174.
- N. Abid, M. Shahzad, M. I. Malik, U. Schwanecke, A. Ulges, G. Kovács, and F. Shafait. Ucl: Unsupervised curriculum learning for water body classification from remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, 105:102568, 2021b. doi: 10.1016/j.jag.2021.102568.
- A. Baldridge, S. Hook, C. Grove, and G. Rivera. The aster spectral library version 2.0. *Remote Sensing of Environment*, 113(4):711–715, 2009. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2008.11.007>. URL <https://www.sciencedirect.com/science/article/pii/S0034425708003441>.
- G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Vila-Francés, J. Amorós-López, and J. Calpe-Maravilla. Retrieval of oceanic chlorophyll concentration with relevance vector machines. *Remote Sensing of Environment*, 105(1):23–33, 2006. doi: 10.1016/j.rse.2006.06.004.
- F. Chevallier, S. Di Michele, and A. McNally. *Diverse Profile Datasets from the ECMWF 91-level Short-range Forecasts*. European Centre for Medium-Range Weather Forecasts, 2006.
- M. Domnich, I. Sünter, H. Trofimov, O. Wold, F. Harun, A. Kostjukhin, M. Järveoja, M. Veske, T. Tamm, K. Voormansik, et al. Kappamask: Ai-based cloudmask processor for sentinel-2. *Remote Sensing*, 13(20):4100, 2021. doi: 10.3390/rs13204100.
- L. Gomez-Chova, D. Fernández-Prieto, J. Calpe, E. Soria, J. Vila, and G. Camps-Valls. Urban monitoring using multi-temporal sar and multi-spectral data. *Pattern Recognition Letters*, 27(4): 234–243, 2006. doi: 10.1016/j.patrec.2005.08.004.
- L. Gómez-Chova, G. Camps-Valls, J. Calpe-Maravilla, L. Guanter, and J. Moreno. Cloud-screening algorithm for envisat/meris multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4105–4118, 2007. doi: 10.1109/TGRS.2007.905312.
- N. Håkansson, C. Adok, A. Thoss, R. Scheirer, and S. Hörnquist. Neural network cloud top pressure and height for modis. *Atmospheric Measurement Techniques*, 11(5):3177–3196, 2018. doi: 10.5194/amt-11-3177-2018. URL <https://amt.copernicus.org/articles/11/3177/2018/>.
- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.
- H. Iwabuchi and T. Hayasaka. Effects of cloud horizontal inhomogeneity on the optical thickness retrieved from moderate-resolution satellite data. *Journal of the atmospheric sciences*, 59(14): 2227–2242, 2002. doi: 10.1175/1520-0469(2002)059<2227:EOCHIO>2.0.CO;2.
- J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftgaard. A cloud detection algorithm for satellite imagery based on deep learning. *Remote sensing of environment*, 229:247–259, 2019. doi: 10.1016/j.rse.2019.03.039.
- S. Kanu, R. Khoja, S. Lal, B. Raghavendra, and C. Asha. Cloudx-net: A robust encoder-decoder architecture for cloud detection from satellite remote sensing images. *Remote Sensing Applications: Society and Environment*, 20:100417, 2020. doi: 10.1016/j.rsase.2020.100417.
- D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- S. J. Koppal. *Lambertian Reflectance*, pages 441–443. Springer US, Boston, MA, 2014. ISBN 978-0-387-31439-6. doi: 10.1007/978-0-387-31439-6_534. URL https://doi.org/10.1007/978-0-387-31439-6_534.

-
- A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- D. Liu. Inferring optical depth of broken clouds from landsat data. *J. Climate*, 8:2620–2630, 1995. doi: 10.1175/1520-0442(1995)008<2620:IODOBC>2.0.CO;2.
- J. Louis, V. Debaecker, B. Pflug, M. Main-Knorn, J. Bieniarz, U. Mueller-Wilm, E. Cadau, and F. Gascon. Sentinel-2 sen2cor: L2a processor for users. In *Proceedings living planet symposium 2016*, pages 1–8. Spacebooks Online, 2016.
- G. Mateo-Garcia, S. Oprea, L. Smith, J. Veitch-Michaelis, G. Schumann, Y. Gal, A. G. Baydin, and D. Backes. Flood detection on low cost orbital hardware. *arXiv preprint arXiv:1910.03019*, 2019.
- G. Mateo-García, V. Laparra, D. López-Puigdollers, and L. Gómez-Chova. Transferring deep learning models for cloud detection between landsat-8 and proba-v. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160:1–17, 2020. doi: 10.1016/j.isprsjprs.2019.11.024.
- S. K. Meerdink, S. J. Hook, D. A. Roberts, and E. A. Abbott. The ecostress spectral library version 1.0. *Remote Sensing of Environment*, 230:111196, 2019. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2019.05.015>. URL <https://www.sciencedirect.com/science/article/pii/S0034425719302081>.
- R. Okamura, H. Iwabuchi, and K. S. Schmidt. Retrieval of optical thickness and droplet effective radius of inhomogeneous clouds using deep learning. *Atmos. Meas. Tech.*, 10:4747–4759, 2017. doi: 10.5194/amt-2017-154.
- R. Saunders, J. Hocking, E. Turner, P. Rayer, D. Rundle, P. Brunel, J. Vidot, P. Roquet, M. Matricardi, A. Geer, et al. An update on the rtov fast radiative transfer model;? xmltex\break?;(currently at version 12). *Geoscientific Model Development*, 11(7):2717–2737, 2018. doi: 10.5194/gmd-11-2717-2018.
- Y. Sde-Chen, Y. Y. Schechner, V. Holodovsky, and E. Eytan. 3DeepCT: Learning volumetric scattering tomography of clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5671–5682, 2021.
- K. Wada. pytorch-fcn: PyTorch Implementation of Fully Convolutional Networks. <https://github.com/wkentaro/pytorch-fcn>, 2017.
- A. Wolanin, G. Mateo-García, G. Camps-Valls, L. Gómez-Chova, M. Meroni, G. Duveiller, Y. Liangzhi, and L. Guanter. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environmental research letters*, 15(2):024019, 2020. doi: 10.1088/1748-9326/ab68ac.
- J. Zhang, X. Li, L. Li, P. Sun, X. Su, T. Hu, and F. Chen. Lightweight u-net for cloud detection of visible and thermal infrared remote sensing images. *Optical and Quantum Electronics*, 52:1–14, 2020. doi: 10.1007/s11082-020-02500-8.
- Z. Zhu, S. Wang, and C. E. Woodcock. Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote sensing of Environment*, 159:269–277, 2015. doi: 10.1016/j.rse.2014.12.014.
- T. Zinner and B. Mayer. Remote sensing of stratocumulus clouds: Uncertainties and biases due to inhomogeneity. *Journal of Geophysical Research: Atmospheres*, 111(D14), 2006. doi: 10.1029/2005JD006955.
- T. Zinner, G. Wind, S. Platnick, and A. S. Ackerman. Testing remote sensing on artificial observations: Impact of drizzle and 3-d cloud structure on effective radius retrievals. *Atmos. Chem. Phys.*, 10: 9535–9549, 2010. doi: 10.5194/acp-10-9535-2010.