Supplement of Atmos. Meas. Tech., 13, 4947–4961, 2020
https://doi.org/10.5194/amt-13-4947-2020-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

A new Orbiting Carbon Observatory 2 cloud flagging method and rapid retrieval of marine boundary layer cloud properties

Mark Richardson et al.

Correspondence to: Mark Richardson (markr@jpl.nasa.gov)

1. Selection of classifier

We considered 13 combinations of observables and 6 classifier setups, including *k*-nearest neighbours (KN with 12- or 25 neighbours), random forest (RF) and multi-layer perceptron (MLP with 1-, 2- or 3-layers), taking the default sklearn class properties unless specified otherwise. The observations are combinations of the continuum radiances or super-channel A-band ratios as described in Supplementary Table 1, which also provides the input dataset names and the selected classifier information.

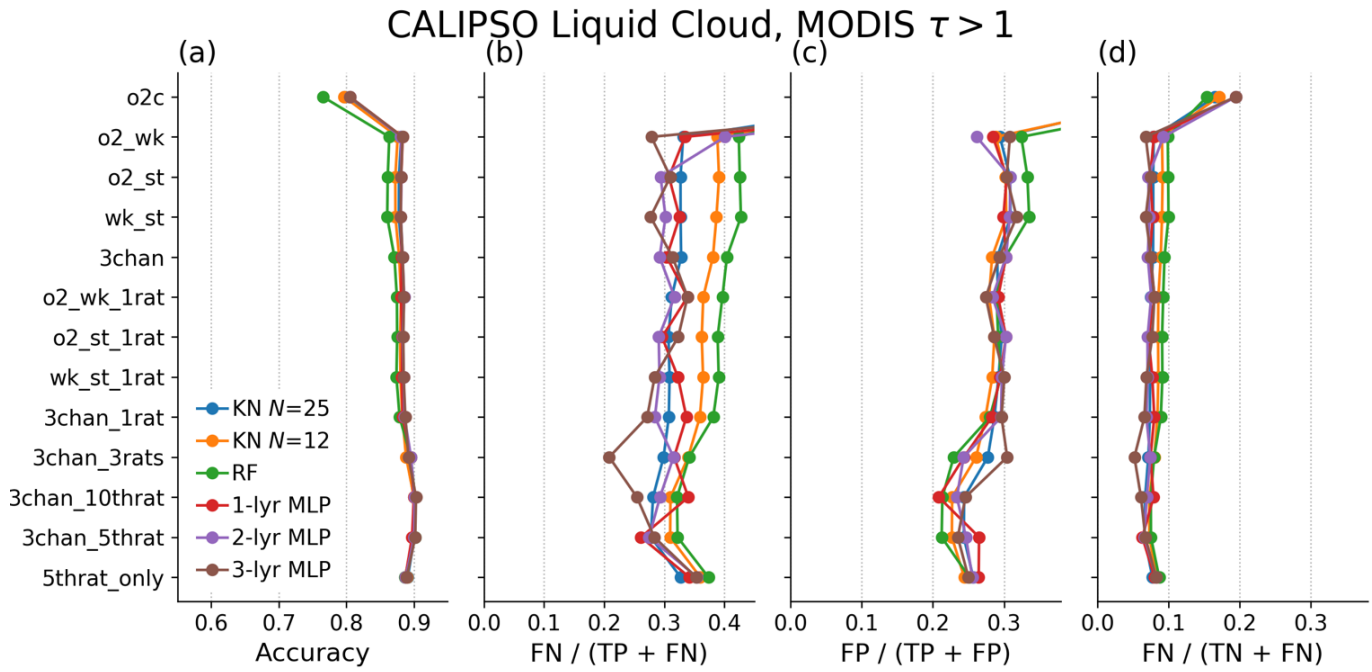
Performance metrics are from the standard confusion matrix elements: true positive (TP), false positive (FP), false negative (FN) and true negative (TN) as described in the main manuscript. The accuracy score is the trace of the confusion matrix, i.e. TP + TN, and we also inspect the following combinations:

(i) FN/(TP + FN), i.e. the fraction of real cases that are missed,

(ii) FP/(TP+FP), i.e. the fraction of classifier positives which are not true clouds,

(iii) FN/(TN+FN), i.e. the fraction of classifier negatives which are not truly negatives,

We wish to minimise (i)—(iii), with (i) representing the loss of potentially good retrievals, (ii) is related to wasted computational resources on footprints which are unlikely to be good retrievals and (iii) quantifies the fraction of negatives which are truly likely to be cloud free. These statistics are calculated for each classifier and observable combination for the independent (non-training) 250,000 footprint sample and are displayed in Supplementary Figure 1, which makes clear that the inclusion of the band ratios improves the accuracy score for all classifiers. Panels (b)—(d) show a range of performances and we selected the 3 layer neural network with the 3chan_10thrat observations. This shows among the lowest rate of missed cases without a large spike in false positives, although there is no clear objectively best choice.



30 **Supplementary Figure 1. Performance statistics of each tested classifier for each observable. Classifiers are labelled in legend, for *k*-nearest neighbours (KN, with *N* neighbours), Random Forest (*RF*) and multi-layer perceptron (*MLP*, with 1-, 2- or 3 hidden layers). Observations refer to $\mu_0^{-1}I$ continua for the O_2 A-band (o2), weak CO_2 band (wk) and strong CO_2 band (st) while "rat" refers to the radiance ratio corrected for μ_0 as described in Supplementary Table 1.**

Supplementary Table 1. Properties used in the classifier along with sources

Property	Value or short name	Data source or calculation method
Hidden layer sizes	(100, 50, 25)	hidden_layer_sizes in sklearn.neural_networks.MLPClassifier
Cloud Classification flag	1	CALIPSO 01kmCLay Feature_Classification_Flag = 2 (i.e. liquid), MODIS MYD061KM Cloud_Optical_Thickness > 1, only one retrieved CALIPSO layer
	0	CALIPSO non-liquid or not retrieved, OR MODIS τ < 1 or not retrieved
Classifier inputs	SZA	SoundingGeometry/sounding_solar_zenith
	O2 continuum $I_{O2,c}$	RadianceMeasurements/rad_continuum_o2
	Weak CO2 continuum $I_{wk,c}$	RadianceMeasurements/rad_continuum_wco2
	Strong CO2 continuum $I_{st,c}$	RadianceMeasurements/rad_continuum_sco2
	A-band ratio $-\ln\left(\frac{I}{I_c}\right)(\mu_0+1)^{-1}$	RadianceMeasurements/radiance_o2, ranked from brightest, non-overlapping 10 channel mean, then every tenth index from 35 inclusive (first index = 1).

2. Determination of surface albedo

To estimate the equivalent Lambertian surface we used refractor's surface albedo estimator, which takes instrument information, the band continuum radiance and solar irradiance to provide a footprint surface albedo. To obtain this by SZA, we selected 1,000 random footprints in each 5° SZA bin from 20—45° inclusive, provided that they were over ocean and that they were flagged as confidently clear by MODIS. The median refractor estimate of each of these 1,000 values was used, and the values are in Supplementary Table 2. The albedos were interpolated from this table onto each LUT SZA, with the SZA=20° LUT using the 22.5° value and the 45° LUT using the 42.5° value.

Supplementary Table 2. Median retrieved surface albedo over ocean and the centre of each 5° solar zenith angle bin.

	SZA [°]				
	22.5	27.5	32.5	37.5	42.5
O_2 A-band	0.0541	0.0357	0.0271	0.0278	0.0230
Weak CO_2 band	0.0421	0.0256	0.0172	0.0181	0.0122
Strong CO_2 band	0.0330	0.0204	0.0139	0.0132	0.0096

3. Lookup table input values

Supplementary Table 3. Input values used for the Nakajima-King tables.

Input Property	**Short Name**	**Values**
Solar zenith angle	SZA	20, 25, 30, 35, 40, 45
Cloud optical depth,	τ	1, 2, 4, 6, 8, 10, 12, 14, 16, 20, 35, 50
Droplet effective radius [μm]	r_e	4, 6, 8, 10, 12, 14, 16, 20, 24, 28, 32
Cloud top pressure [hPa]	P_{top}	650, 690, 730, 770, 810, 850, 910, 950

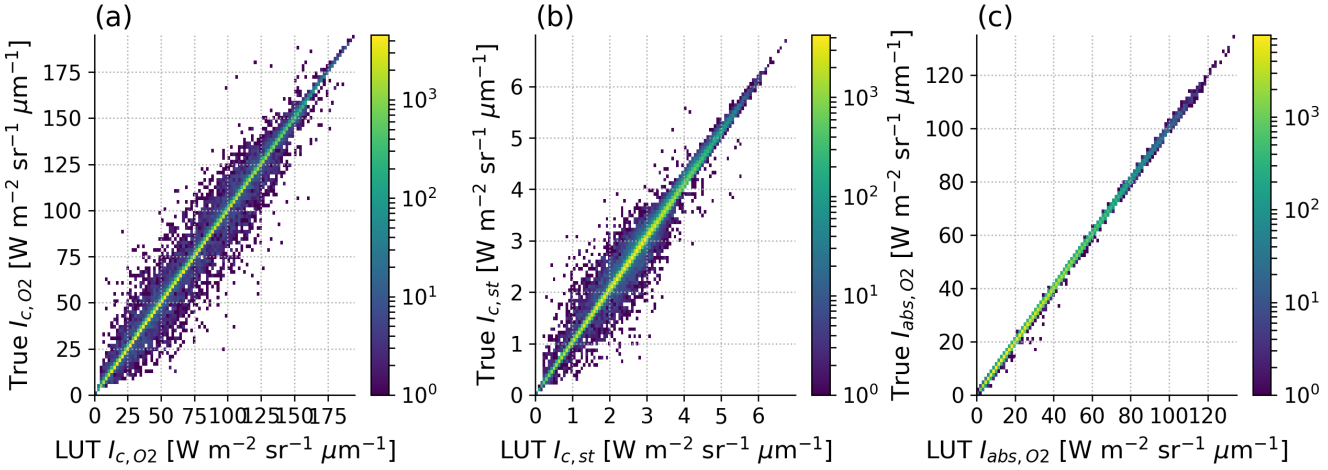
4. Channel selection algorithm and performance

We selected a single set of channels for the LUT that would be valid for all meteorological conditions, spacecraft motion, and sounding positions by testing against a sample consisting of every SZA < 45° MODIS $\tau > 1$ footprint from every 50th orbit (N~230,000).

The algorithm aimed to select channels whose mean minimised the sample root mean square error (RMSE) against the relevant "truth". For $I_{c,O2}$ the truth is the L1bSc SoundingMeasurements/rad_continuum_o2, for $I_{c,st}$ it is SoundingMeasurements/rad_continuum_sco2 and for $I_{abs,O2}$ it is the spectrum's 60th-ranked superchannel as described in the main manuscript.

The algorithm first calculates every individual channel's RMSE against the target and then selects the 30 channels with the lowest individual RMSEs. It then calculates the 30-channel mean and returns its RMSE, before removing each channel in turn and calculating the RMSEs of the new 29-channel means. The combination of channels with the lowest RMSE is taken as the 29-channel example, and this process is repeated iteratively down to 3 channels.

The comparison between the targets and the estimates using the selected channels is in Supplementary Figure 2, note that the colours represent logarithmic counts. There is excellent correlation for all three observables, even though each of the LUT values uses the mean of a fixed set of channels whereas the individual channels used in each of the truth values may vary between soundings. In particular, the excellent agreement in Supplementary Figure 2(c) compares with the individual channel spread in the main manuscript Figure 3 shows that the algorithm selected channels with sufficient anti-correlation to reliably reproduce our estimate of $I_{abs,O2}$. The median and 5—95 % ranges of the radiance differences are listed in Supplementary Table 4 and the list of channels are in Supplementary Table 5.



Supplementary Figure 2. 2d histograms of the target ("true") radiance properties as a function of the means estimated from the fixed channel sets ("LUT"), (a) for the O_2 A-band continuum, (b) for the strong CO_2 band continuum, (c) for the 60th-ranked super-channel, representing an O_2 A-band absorption channel.

Supplementary Table 4. Median and 5—95 % range of the differences between the radiance estimates from the mean of the channels used in the lookup table, and the targets.

	Percentiles of LUT minus true radiance differences		
Radiance property	**Median (%)**	**5th percentile (%)**	**95th percentile (%)**
$I_{c,O2}$	-0.2	-1.7	0.0
$I_{c,st}$	-2.3	-4.7	-0.5
$I_{abs,O2}$	-0.3	-2.1	1.3

Supplementary Table 5. Channel indices used for each radiance property in the LUT, counting from index 0.

Radiance property	Channels
$I_{c,O2}$	782, 827, 838, 839, 889
$I_{c,st}$	841, 842, 924, 927, 937
$I_{abs,O2}$	115, 113, 116, 246, 329, 338, 339, 380, 381, 576

75