Polar Radiant Energy in the Far-Infrared Experiment (PREFIRE) Algorithm Theoretical Basis Document (ATBD) for the 2B-CLD data product

Nathaniel B. Miller, Brian J. Drouin, Aronne Merrelli, Timothy I. Michaels, Erin Hokanson Wagner

R01 release, March/April 2025

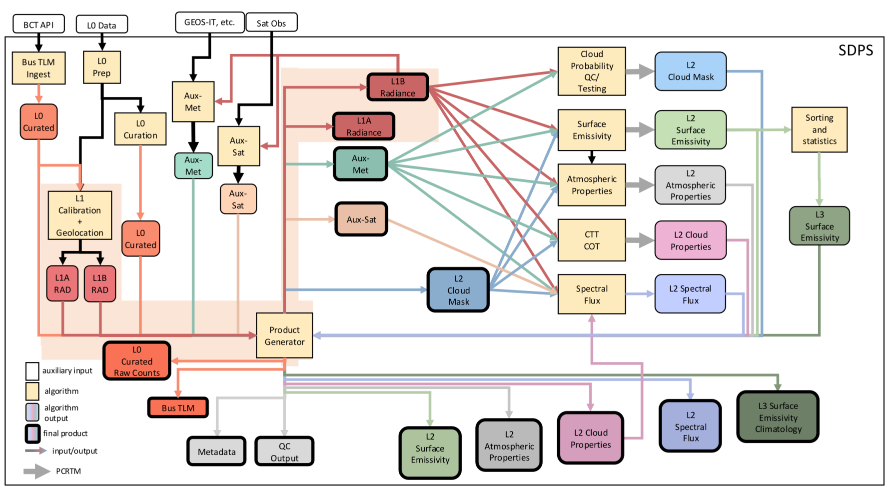

Figure 0.1: PREFIRE algorithm connectivity and flow.

Contents

1 Level-2 Cloud Properties

1.1 Introduction

This Algorithm Theoretical Basis Document (ATBD) describes the algorithms used to produce the 2B-CLD (cloudy atmosphere) product for the Polar Radiant Energy in the Far InfraRed Experiment (PREFIRE). The 2B-CLD algorithm uses data from the PREFIRE AUX-MET (Auxiliary Meteorological) and Cloud Mask (2B-MSK) data products as prior information.

1.2 Instrument overview

A detailed instrument overview can be found in the 2B-ATM Algorithm Theoretical Basis Document (ATBD). The 2B-ATM ATBD includes depictions of the Spectral Response Functions (SRFs), which illustrate PREFIRE's moderate spectral resolution. Pertinent details of the PREFIRE instrument from the 2B-ATM ATBD are provided below.

The spectrometer for PREFIRE, the Thermal Infrared Spectrometer (TIRS, or TIRS-PREFIRE), collects spectral radiance measurements across a wavelength range of approximately 5 to 54 μm with a spectral sampling of 0.84 μm. The light is dispersed by a grating onto a 64$\times$8 element detector array that measures 8 simultaneous spectra along the spectrometer slit. The first four channels respond to shortwave radiance ($< 3\,\mu$m) and are not planned to be part of the calibrated 1B-RAD dataset, as there will be no calibration system for these wavelengths and no expectation of instrument performance. Due to the instrument design, there are two-channel gaps at approximately 7, 15, and 30 μm, at the boundaries of the order-sorting filters used to select for specific grating diffraction orders. The layout of the filters results in 54 usable channels covering most of the thermal infrared range. The actual flight detectors have individual bad detector elements which will imply a different number of valid channels between the 8 cross-track spectra.

Due to unpowered flight, the slowly decreasing altitude of the PREFIRE orbits will result in gradually smaller observational footprints throughout the mission. The initial ground footprint shapes of the 8 TIRS scenes are quadrilaterals, approximately 11.8 km x 34.8 km (cross-track by along-track) in size, with the 8 scenes separated cross-track by 24.2 km gaps between them. The temporal sampling rate of TIRS (0.7007 s) at the initial orbit altitude results in an along-track translation of only about 5.3 km, so that more than 6 consecutive measurements overlap. For the baseline 2B-CLD algorithm, no attempt is made to combine these observations in any way. In other words, each spectrum is treated as an entirely independent measurement.

1.3 Overview

The 2B-CLD algorithm is a physical retrieval implemented with a standard optimal estimation approach (Rodgers, 2000), with a Levenberg-Marquardt parameter to adjust the weighting of the *a priori* and measurement information during iteration. The state vector consists of Cloud Top Pressure (CTP), Cloud Optical Depth (COD), and Cloud-particle Effective Diameter (CED). The COD is considered the visible-wavelength cloud optical depth as it is referenced to a visible wavelength at 550 nm (Liu et al., 2009). The remaining relevant geophysical properties are taken from the values in the *a priori* datasets, and are assumed to be fixed values. The 2B-CLD algorithm is intended to be run for cloudy-sky conditions only.

The forward model utilized in the retrieval algorithm, described in Liu et al. (2006), is the Principal Component-based Radiative Transfer Model (PCRTM). PCRTM contains ice and water cloud spectral emissivity models, although 2B-CLD performs the retrieval specifically using only the ice cloud model due to a current lack of *a priori* microphysical retrievals or coincident measurements. Parameters from the PREFIRE AUX-MET auxiliary data product, such as profiles of various atmospheric constituents, temperature profiles, and surface temperature are used as input into PCRTM, and a surface emissivity of 1.0 is assumed. All of these geophysical scene parameters are held constant during the retrieval.

In addition to the AUX-MET product, 2B-CLD also utilizes the 1B-RAD radiances and the L2 Cloud Mask (2B-MSK) products. Figure 1.1 depicts the overall data flow, including the three PREFIRE input data products and the static *a priori* data. The retrieval is only performed in cases where the cloud mask determines that there is cloud present. The threshold for this cloudy determination requires a cloud probability greater than 0.6, classified as “likely cloud” or “confidently cloud” in the 2B-MSK product.

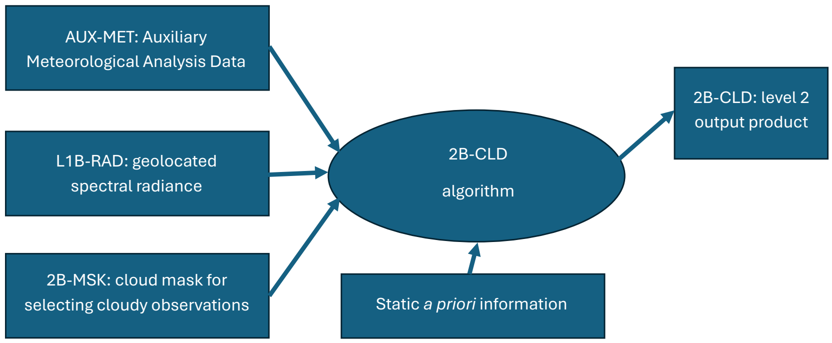

Figure 1.1: Cloud properties algorithm flowchart.

1.4 Spectral radiance forward model

The forward model used in the 2B-CLD algorithm is the PCRTM V3.4 (Liu et al., 2006). PCRTM includes ice and water cloud spectral emissivity models, although for the 2B-CLD retrieval the input assumes ice phase only because of a lack of prior microphysical cloud particle knowledge. By assuming the scattering by clouds is isotropic, the effective transmittance and reflectance are parameterized from the COD, CED, and viewing angle (Liu et al., 2009). PCRTM calculations were performed assuming that the modeled cloud is confined to one layer.

The forward model spectral radiance is found by convolving SRFs, for either TIRS1 or TIRS2, with a high resolution PCRTM output. In order to maximize speed and efficiency, PCRTM uses pre-computed Principal Components to define a sampling grid. Before the SRFs are applied to create the TIRS channel radiances, the full spectral resolution PCRTM output covers the wavelength range $50 - 2760\,\mathrm{cm}^{-1}$ on a $0.5\,\mathrm{cm}^{-1}$ sampling grid.

The Jacobians are calculated numerically using PCRTM via small perturbations in the state variables. In order to minimize computational time per retrieval iteration, the state variables are perturbed only once from the baseline forward run and then subtracted from the baseline value to calculate the Jacobian ($\mathbf{K}$). The state variable ($\mathbf{x}$) is perturbed by adding $\delta\mathbf{x}$ and then the forward model ($\mathcal{F}$) is run in order to calculate $\mathbf{K}$, according to Equation 1.1.

$$\mathbf{K} = (\mathcal{F}(\mathbf{x} + \delta\mathbf{x}) - \mathcal{F}(\mathbf{x}))/\delta\mathbf{x} \tag{1.1}$$

The COD and CED are perturbed by 10 percent of the current iteration value ($\delta\mathbf{x} = 0.1\mathbf{x}$). The CTP variable is usually perturbed by adding 1 hPa. The PCRTM computes the transmittance from the cloud top to TOA at whole-layer boundaries. Thus, if the CTP perturbation happens to cross the mid-point of the layer, the modeled radiance would have an unexpectedly large change because of the large change in the cloud top to TOA transmittance. This large

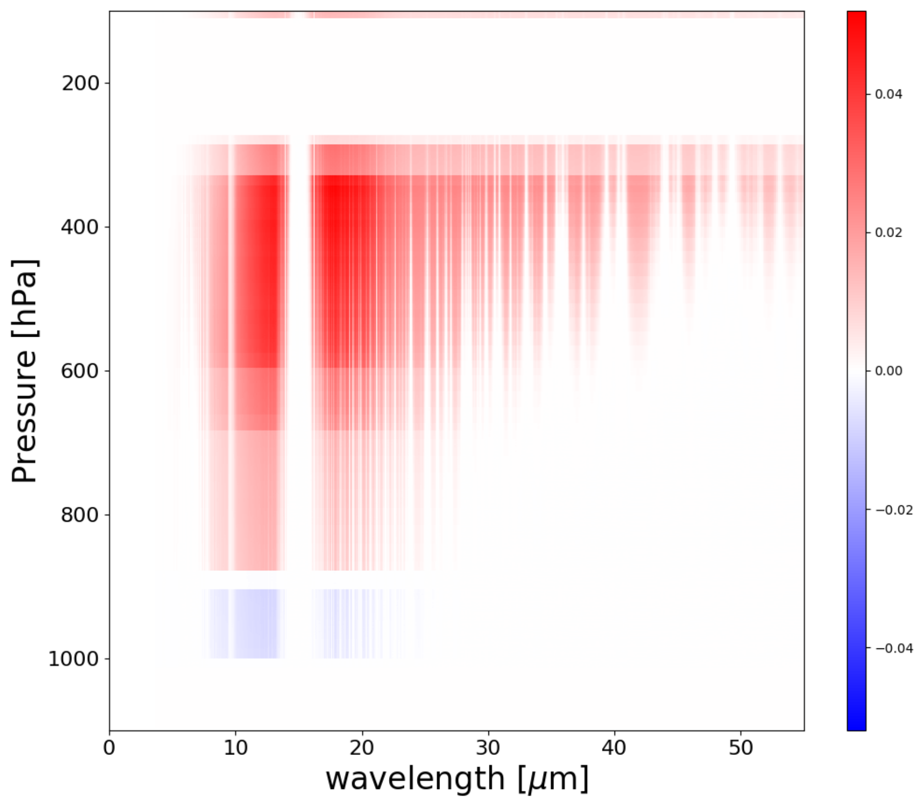

Figure 1.2: Cloud top pressure Jacobian, with units of $\mathrm{W\,(m^2\,sr\,\mu m)^{-1}\,(hPa)^{-1}}$, at the full PCRTM spectral resolution of $0.5\,\mathrm{cm^{-1}}$ for a sub-Arctic winter standard profile.

radiance change would introduce a large discontinuity in the CTP Jacobian. Hence, if adding 1 hPa were to cross the midpoint of a PCRTM level, then alternatively Equation 1.2 is used to calculate the Jacobian.

$$\mathbf{K} = (\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{x} - \delta\mathbf{x}))/\delta\mathbf{x} \tag{1.2}$$

While this will lead to small discontinuities in the radiance changes when the CTP is within 1 hPa of the PCRTM half level, the discontinuity is much smaller than the feature that would arise from the layer transmittance change described above. The CTP Jacobian for the full PCRTM spectral resolution is shown in Figure 1.2 for a standard sub-Arctic winter profile, assuming a CED of $40\,\mu$m and COD of 1.0.

1.5 State and measurement vectors

The retrieved state vector for the 2B-CLD algorithm includes the cloud top pressure, **CTP**, the cloud effective particle diameter **CED**, and the logarithm

Parameter	Variance
Cloud Top Pressure	$(200 \text{ hPa})^2$
ln(Cloud Optical Depth)	$(1.15)^2$
Cloud effective diameter	$(20\ \mu\text{m})^2$

Table 1.1: Parameters defining the *a priori* covariance matrix, which includes cloud top pressure, cloud optical depth, and cloud particle effective diameter.

of the cloud optical depth, ln(**COD**) as follows:

$$\mathbf{x} = [\mathbf{CTP}; \mathbf{CED}; \ln(\mathbf{COD})] \tag{1.3}$$

The measurement vector is the measured spectral radiance. TIRS spectra will contain up to 54 valid spectral channels, less any identified bad detector elements as determined by the 1B-RAD fields. If in the `detector_bitflags` variable in the 1B-RAD product, bit flags b0, b1, b3, b4, or b5 are set, then these channels are not used in the retrieval. In addition, if the `observation_quality_flag` is set to 2 ("bad") in 1B-RAD then no retrieval is performed.

1.6 *A priori* data

Before the OE algorithm iterates, the *a priori* covariance matrix and mean values are needed to describe the expected probability distribution of the state vector. The diagonal elements of the covariance matrix are provided in Table 1.1. The *a priori* mean values are the same as the first guess for the 3 state vectors. Due to a lack of relevant information, predetermined first guess values of 600 hPa, 40 μm, and 5 are given for the CTP, CED and COD, respectively. The OE solver goes out of range less frequently if it starts with an optically thick cloud, as a CTP is more easily determined – thus a larger value is used as the first guess for the COD in order to increase the number of physically realistic solutions.

PCRTM requires estimates of geophysical variables to model the spectral radiance for a given scene. The surface spectral emissivity is fixed to a value of 1.0 across all wavenumbers. Profiles of CO and N_2O are fixed as given from the standard sub-Arctic winter atmosphere (McClatchey et al. 1972). Profiles of temperature, water vapor mixing ratio, and O_3 are taken from the the PREFIRE AUX-MET auxiliary data product, derived from spatially- and temporally-interpolated profiles of the GEOS-IT analysis data stream from NASA GMAO (Lucchesi, 2015). A fixed volume mixing ratio profile is used for CO_2 and CH_4 as described in the 2B-ATM ATBD.

1.7 Inversion method

The following are pertinent details of the OE inversion method, which is explained in more detail in the 2B-ATM ATBD, as the 2B-CLD retrieval framework is essentially the same.

The inversion method used in the 2B-CLD algorithm is a standard Bayesian non-linear optimal estimation (OE) approach. Starting from an initial guess, the algorithm iterates the state vector, $\mathbf{x}$ value, recomputing the forward modeled spectral radiance and Jacobians at each step. The state vector updates at each iteration are the standard linear cost-function minimization steps. The method is similar to the standard Newton's method, with an additional Levenberg-Marquardt parameter to adjust the weighting of the *a priori* and measurement information during iteration (Rodgers 2000). This implementation closely follows the OE solver method used in the NASA OCO-2 L2 algorithm (Crisp et al., 2021).

The cost function is from the standard OE formalism, following from an assumed *a priori* state vector mean ($\mathbf{x_a}$) and covariance ($\mathbf{S_a}$), a measurement vector ($\mathbf{y}$) and measurement error covariance ($\mathbf{S}_\epsilon$), and a forward model function ($\mathcal{F}$). For a particular iteration where the state vector value is $\mathbf{x_i}$, the cost function (c) is given by:

$$c = (\mathbf{y} - \mathcal{F}(\mathbf{x_i}))^T \mathbf{S}_\epsilon{}^{-1} (\mathbf{y} - \mathcal{F}(\mathbf{x_i})) + (\mathbf{x_i} - \mathbf{x_a})^T \mathbf{S_a}{}^{-1} (\mathbf{x_i} - \mathbf{x_a}) \tag{1.4}$$

At each iteration, the forward model returns the modeled measurement ($\mathcal{F}(\mathbf{x_i})$) as well as the Jacobian at the state vector value ($\mathbf{K_i}$). These are used to compute the state update, $\mathbf{dx_{i+1}}$. The actual state update is computed using a linear matrix solver (the `linalg.solve` function in NumPy, which utilizes LAPACK). The state update equation is given by:

$$[(1+\gamma)\mathbf{S_a}{}^{-1} + \mathbf{K_i}{}^T \mathbf{S}_\epsilon{}^{-1} \mathbf{K_i}] \mathbf{dx_{i+1}} = [\mathbf{K_i}{}^T \mathbf{S}_\epsilon{}^{-1} (\mathbf{y} - \mathcal{F}(\mathbf{x_i})) + \mathbf{S_a}{}^{-1} (\mathbf{x_i} - \mathbf{x_a})] \tag{1.5}$$

Due to a lack of prior information about the cloud properties, there are times when the state variables are updated to a non-physical value and thus cannot be used as input to update the forward model. Thus, limits were implemented on the state variables and if any of the 3 state variables step out of their range, then the retrieval is halted and flagged accordingly. The range for CTP is between 50 hPa and the surface pressure (as given by the PREFIRE AUX-MET auxiliary data product). The range for COD is between 0.0001 and 18.0, and for CED it is between 0.5 μm and 162 μm. To limit these out of range occurrences, the OE retrieval is initiated with a gamma parameter of 100 such that the prior values have more weight at the start, and thus the algorithm will be less likely to take a large step to a non-physical value.

1.8 Output processing

The output of the 2B-CLD optimal estimation algorithm has several status variables that are used for quality assessment. A summary integer flag (`cld_quality_flag`) with categorical values is provided, as well as a separate bit flags field (`cld_qc_bitflags`) that includes more detailed status information. The summary integer flag should be sufficient for most uses, with the additional detail that is contained in the bit flags available for more advanced analysis of the product. In the future, the

status value	Description
0	Best quality, converged retrieval
1	Retrieval converged but did not meet quality check
2	Retrieval did not converge
3	Retrieval went out of range
-99	Retrieval not attempted (missing value)

Table 1.2: Description of 2B-CLD integer quality flag

bit number	Status description
0	reduced χ^2 threshold exceeded
1	retrieval exceeded iteration count limit
2	retrieval exceeded diverging step count limit
3	retrieval went outside allowable state vector range
4	retrieval solver crashed
12	retrieval not attempted, due to cloud mask
13	retrieval not attempted due to latitude constraint
14	retrieval not attempted due to 1B-RAD status

Table 1.3: Descriptions of 2B-CLD quality bitflags

bit flags will likely include additional status conditions as more conditions are identified.

A summary integer flag records the four main status conditions for the 2B-CLD algorithm within each TIRS observation and a fifth condition in which no retrieval was attempted. If the retrieval was attempted, then the status outcomes could be: a good-quality converged retrieval, a poorer-quality converged retrieval, an unconverged retrieval, or the retrieval stepped out of range for the state variables. These are listed in table 1.2 with the corresponding integer values.

The quality bit flags give more detail regarding the status of iteration convergence, which can fail because either the diverging step or iteration count limits were reached, or an unphysical state vector value was reached during iteration (see Section 1.7). The maximum divergence count threshold is set to 5 and the iteration count limit is set to 20. The reduced χ^2 threshold is set conservatively to a value of 20. Table 1.3 lists the bit numbers and descriptions of the bit flags for the attempted retrievals (b0-b4) and retrievals that were not attempted (b12-b14). For example, initially the 2B-MSK product is only available at latitudes greater than 60°, such that non-polar data will have a (`cld_quality_flag`) value of -99 and the (`cld_qc_bitflags`) b13 bit set. Other reasons for the retrieval not being attempted are if the observation quality flag is set to 2 ("bad") in the 1B-RAD producy or if 2B-MSK does not classify the scene as being suitably cloudy.

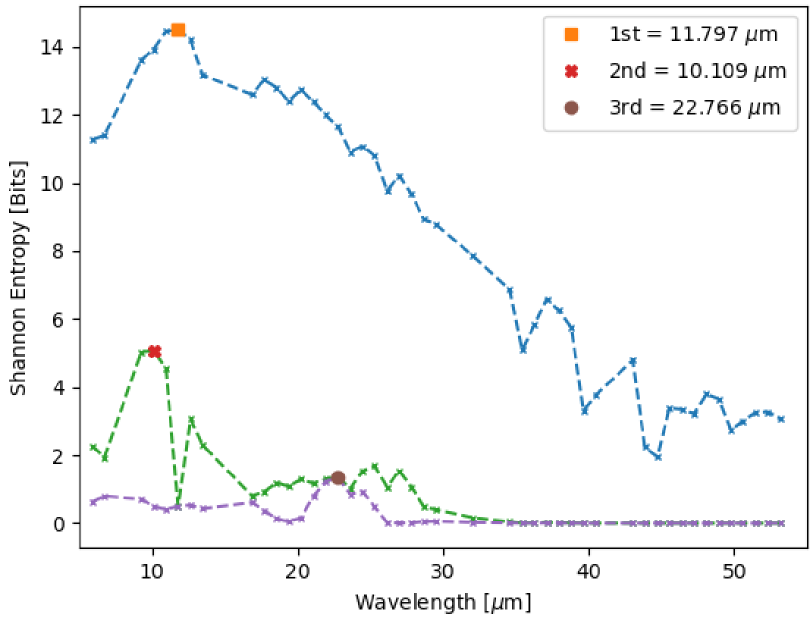

Figure 1.3: The information content (IC) of the 2B-CLD product assuming an ice cloud at 500 hPa with an optical depth of 1.0 in a standard sub-Arctic winter profile for SAT2 scene 1. The orange square indicates the 1^{st}-ranked channel selected with the maximum of the IC spectrum (blue line). The green line is the IC after factoring out the contribution from the 1^{st}-ranked channel and the red X indicates the selected 2^{nd}-ranked channel. The purple line is the IC after factoring out the contributions for the selected 1^{st}- and 2^{nd}-ranked channels and the brown circle indicates the 3^{rd}-ranked channel.

1.9 Retrieval analysis: information content

Information content analysis is useful for determining which TIRS channels have the greatest impact on 2B-CLD retrievals. Figure 1.3 illustrates the information content (IC) spectrum, using the method described by L'Ecuyer et al. (2006), of the various TIRS2 channels. Larger values of Shannon Entropy indicate which channels have higher information content.

For a sub-Arctic winter temperature and humidity profile with an ice cloud inserted at 500 hPa and an optical depth of 1.0, it is found that the 11.8 μm channel is ranked first for the TIRS2 instrument. This mid-infrared window channel is very useful for determining cloud top pressure. After removing the first-ranked channel, the analysis is repeated to find the second-ranked channel (10.1 μm) as well as the associated information content spectrum, associated with the cloud optical depth state variable. The third-ranked channel is found to be in the far-infrared (22.8 μm), and although it has a much lower information content value, it does have information regarding the effective cloud particle

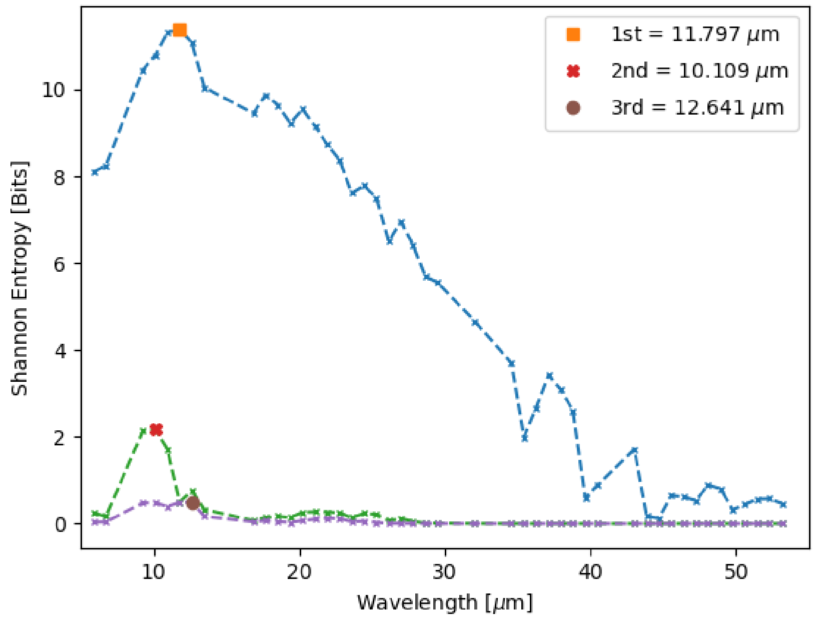

Figure 1.4: The information content (IC) of the 2B-CLD product assuming an ice cloud at 500 hPa with an optical depth of 0.1 in a standard sub-Arctic winter profile for SAT2 scene 1. The orange square indicates the 1^{st}-ranked channel selected with the maximum of the IC spectrum (blue line). The green line is the IC after factoring out the contribution from the 1^{st}-ranked channel and the red X indicates the selected 2^{nd}-ranked channel. The purple line is the IC after factoring out the contributions for the selected 1^{st}- and 2^{nd}-ranked channels and the brown circle indicates the 3^{rd}-ranked channel.

size. In investigating the TIRS1 information spectra for the same test case as in Figure 1.3, the top ranked channels were found to be at similar spectral locations – 12.2, 9.7, and 22.4 μm (not shown).

For test cases with higher amounts of total column water vapor, the third-ranked channel occurs at a shorter wavelength (17.7 μm) and has greater Shannon Entropy bit values. In contrast, for the sub-Arctic winter case with an ice cloud at 500 hPa but with a cloud optical depth of 0.1 there are effectively only 2 channels with independent pieces of information (Figure 1.4). This indicates that uncertainty of the 2B-CLD retrieval will be higher for optically-thin clouds due to a reduction in the amount of information available.

1.10 References

Crisp, D., O'Dell, C., Eldering, A., Fisher, B., Oyafuso, F., et al., 2021. OCO (Orbiting Carbon Observatory)-2 Level 2 Full Physics Retrieval Algorithm Theoretical Basis, Tech. Rep. OCO D-55207, NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, Version 3.0 - Rev 1, https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO_L2_ATBD.pdf

L'Ecuyer, T. S., P. Gabriel, K. Leesman, S. J. Cooper, and G. L. Stephens, 2006: Objective assessment of the information content of visible and infrared radiance measurements for cloud microphysical property retrievals over the global oceans. Part I: Liquid clouds. J. Appl. Meteor. Climatol., 45, 20–41, https://doi.org/10.1175/JAM2326.1

Liu, X., Smith, W.L., Zhou, D.K., Larar, A., 2006. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. Appl. Opt. 45, 201–209.

Liu, X., Zhou, D.K., Larar, A., Smith, W.L., Schluessel, P., Newman, S.M., Taylor, J.P, Wu, W., 2009. Retrieval of atmospheric profiles and cloud properties from IASI spectra using super-channels. Atmos. Chem. Phys. 9, 9121–9142, https://doi.org/10.5194/acp-9-9121-2009

Lucchesi, R., 2015: File Specification for GEOS-5 FP-IT. GMAO Office Note No. 2 (Version 1.4), 60 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes .

McClatchey, R. A., R. W. Fenn, J. E. A. Selby, F. E. Volz, and J. S. Garing,1 972: Optical properties of the atmosphere. Rep. AFCRL-72-0497, Hanscom Air Force Base, Bedford, Mass., 110 pp.579

Rodgers, C. (2000) Inverse Methods for Atmospheric Sounding: Theory and Practice. World Scientific Publishing Co Pte. Ltd.

2 Appendix

2.1 Table of variables and symbols

$\mathbf{A}$	averaging kernel matrix
α	angular resolution
β	azimuth angle
B	blackbody radiance
BW	spectral bandwidth
χ	convergence criterion
c	speed of light, cost function
CED	Cloud particle Effective Diameter
COD	Cloud Optical Depth
CTP	Cloud Top Pressure
CWP	Cloud Water Path
d	degree of freedom
ε	emissivity
ϵ	noise, error
ϕ	longitude
E	irradiance
F	flux
f	focal length
$\mathcal{F}$	function
γ	*a priori* weight
G	gravitational constant
g	gain
H	height
h	Planck's constant
I	radiance
IC	Information Content
IWC	Ice Water Content
IWP	Ice Water Path
j	counter
k	Boltzmann's constant, unknown
$\mathbf{K}$	Jacobian
λ	wavelength, Marquardt-Levenberg parameter
l	distance
L	radiance
LTS	Lower Tropospheric Stability
LWC	Liquid Water Content
LWP	Liquid Water Path
M	counter, mass
m	number of along-track frames
$\mathcal{M}$	matrix

N	counter
n	channel
$\mathcal{N}$	normal distribution
ν	frequency
NEdT	Noise-Equivalent delta Temperature
o	offset
Ω	solid angle
p	pressure
P	probability
PWV	Precipitable Water Vapor
Q	water vapor
ρ	reflection coefficient
R	radius, resistance, cost-function change
$\Re$	response function
$\wp$	responsivity
σ_B	Stefan-Boltzmann constant
$\mathbb{S}$	signal level in digitized counts
S	covariance
SI	Segmentation Index
SNR	Signal-to-Noise Ratio
SRF	Spectral Response Function
θ	latitude, potential temperature, polar coordinate angle
τ	transmission, optical depth
T	temperature
TR	Training Radiances
TREM	TRaining Eigenvector Matrices
t	time
ϕ	polar coordinate angle
V	voltage
v	velocity
x, y, z	position coordinates
z	convergence, standard deviation of scaled differences
$\mathbf{x}$	state vector
X	focal plane position
$\mathbf{y}$	measurement vector
Y	focal plane position
ζ	incidence angle

Table 2.1: Table of variables and symbols.

2.2 Abbreviations and acronyms

ADM	Angular Distribution Model
AIRS	Atmospheric Infrared Sounder
ATBD	Algorithm Theoretical Basis Document
CERES	Clouds and the Earth's Radiant Energy System
DEM	Digital Elevation Model
DOF	Degree of Freedom
ECI	Earth-Centered Inertial
ECMWF	European Centre for Medium-Range Weather Forecasts
EOF	Empirical Orthogonal Function
FIR	Far-InfraRed
FOV	Field Of View
FPA	Focal Plane Array
FWHM	Full Width at Half Maximum
GEOS-IT	Goddard Earth Observing System for Instrument Teams
GMAO	Global Modelling and Assimilation Office
IFOV	Instantaneous Field Of View
IFS	Integrated Forecasting System
LW	Longwave
MIR	Mid-InfraRed
NASA	National Aeronautics and Space Administration
NEP	Noise Equivalent Power
NEdR	Noise Equivalent delta spectral Radiance
OE	Optimal Estimation
OLR	Outgoing Longwave Radiation
PCRTM	Principal Component-based Radiative Transfer Model
PREFIRE	Polar Radiant Energy in the Far-InfraRed Experiment
ROIC	Read-Out Integrated Circuit
RMSE	Root Mean Square Error
SDPS	Science Data Processing System
SSF	Single Scanner Footprint
SRF	Spectral Response Function
TCWV	Total Column Water Vapor
TIRS	(TIRS-PREFIRE) Thermal InfraRed Spectrometer
TIRS1	Thermal InfraRed Spectrometer on PREFIRE-SAT1
TIRS2	Thermal InfraRed Spectrometer on PREFIRE-SAT2
TOA	Top of Atmosphere
UTC	Coordinated Universal Time
VZA	Viewing Zenith Angle
WV	Wavelength

Table 2.2: Abbreviations and acronyms.

2.3 Figure listing with links

Table 2.3: List of Figures in this ATBD.