

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

Cameron Bertossa, Brian Kahn, Brian J. Drouin,  
Timothy I. Michaels, Erin Hokanson Wagner

R01 release, March/April 2025

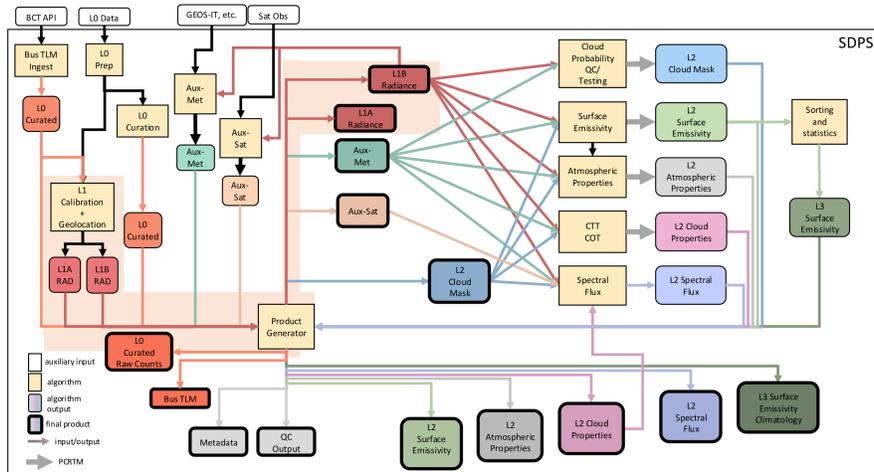


Figure 0.1: PREFIRE algorithm connectivity and flow.

# Contents

<b>1</b>	<b>Level-2 Cloud Mask</b>	<b>2</b>
1.1	Introduction . . . . .	2
1.2	PREFIRE mission requirements . . . . .	2
1.3	ML-MSK . . . . .	3
1.4	Using ML-MSK . . . . .	5
1.5	Initial Validation . . . . .	6
1.6	References . . . . .	7
<b>2</b>	<b>Appendix</b>	<b>8</b>
2.1	Table of variables and symbols . . . . .	8
2.2	Abbreviations and acronyms . . . . .	10
2.3	Figure listing with links . . . . .	11

## 1 Level-2 Cloud Mask

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

### 1.1 Introduction

The PREFIRE Level-2 algorithm suite requires that every Thermal InfraRed Spectrometer (TIRS, or TIRS-PREFIRE) scene be identified as either clear or cloudy. This determination is made through a cloud masking algorithm designed for TIRS. Not only should this algorithm work globally, but it should perform seamlessly in high latitudes where geophysical conditions make cloud detection especially challenging. Toward this end, the PREFIRE algorithm team developed a cloud mask based on a machine learning, neural network (NN) methodology (ML-MSK) that is described in Bertossa et al. (2023).

### 1.2 PREFIRE mission requirements

The PREFIRE mission has clearly defined requirements for the 2B-MSK product. (1) The cloud mask is required to detect 80-90% of clear-sky occurrences. The primary usage of the cloud mask is to identify clear scenes with high confidence such that other downstream PREFIRE algorithms can operate within clear-sky conditions to meet their own requirements. (2) The aforementioned requirement applies to clear-sky scenes with spatial scales of 15-50 km and larger extents. At these scales, clear-sky is mostly resolved by the nominal TIRS spatial footprint size. Clear-sky scenes with scales smaller than 15 km are smaller than the TIRS spatial footprint size and have no mission requirement for detection.

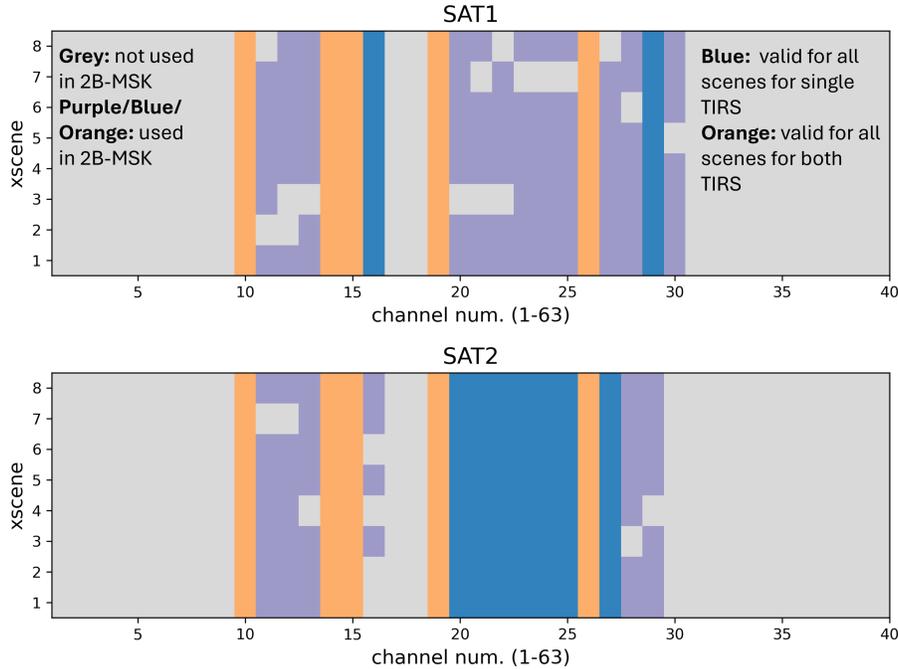


Figure 1.1: Channels used for ML-MSK NNs.

### 1.3 ML-MSK

This section outlines the design and training procedure for the machine learning-based cloud mask. The overall methodology closely follows the approach described in Bertossa et al. (2023), with some modifications to account for the specific characteristics of TIRS data found after launch. At its core, this approach exploits the ability of neural networks to model nonlinear relationships between inputs (in this case, multispectral radiance data) and expected outputs (cloud presence).

Training and evaluation of the cloud mask are performed using co-located VIIRS (Visible Infrared Imaging Radiometer Suite) cloud mask scenes, which provide a high-quality reference due to their use of both visible- and thermal-wavelength channels. ‘Co-locations’ are instances in which both instruments pass over the same location within at least a 60-minute window of one another. Because each TIRS footprint spans multiple VIIRS pixels, a thresholding strategy is applied: a TIRS footprint is labeled as “cloudy” only if at least 50% of the overlapping VIIRS pixels are classified as cloudy.

Only a selected subset of TIRS channels are used as NN inputs (see 1.1). This subset is determined by 1) the 1B-RAD `detector_bitflags` mask and 2) empirical testing designed to remove channels with especially large striping. This restriction helps mitigate noisy cloud mask fields caused by striping or noise

in the radiance fields. In addition to TIRS radiances, the neural networks also incorporate auxiliary meteorological variables from the PREFIRE AUX-MET data product — specifically, skin temperature and total column water vapor. These variables provide essential context for interpreting the thermal signal and enhance the NN’s ability to distinguish clouds from surface conditions within different latitude regimes.

The machine learning cloud mask is trained using approximately 200 granules drawn from the first ten days of July 2024 and January 2025. These dates were selected to capture the extremes of polar conditions observed during the initial year of PREFIRE operations. Only scenes poleward of 60 degrees latitude are included in the training set. During polar night, the VIIRS cloud mask relies exclusively on thermal channels; accordingly, the thermal-only VIIRS mask is used as the reference (“truth”) in those cases. Although this may result in a less reliable truth label, omitting polar night scenes during training has been found to significantly degrade model performance under those conditions. Therefore, their inclusion is essential. Future versions may adopt the use of more sensors to help compensate for the limited performance of VIIRS in the polar night, and/or a more targeted sampling strategy to better capture a wider range of meteorological variability.

The neural network is trained by minimizing a weighted binary cross-entropy loss function, defined as:

$$\text{Loss} = -\frac{1}{|\mathbf{Y}^*|} \sum \left[ (1 - \beta)\mathbf{Y}^* \log_{10}(\hat{\mathbf{Y}}) + \beta(1 - \mathbf{Y}^*)(\log_{10}(1 - \hat{\mathbf{Y}})) \right] \quad (1.1)$$

where  $\mathbf{Y}^*$  is a binary set representing ‘truth’. Clear is defined to be the null condition (0) and cloudy the alternate condition (1).  $\hat{\mathbf{Y}}$  is the corresponding set of predicted probabilities by the NN for the cloudy class. Finally,  $\beta = \frac{\sum \mathbf{Y}^*}{|\mathbf{Y}^*|}$ , adjusts for imbalances between the number of cloudy versus clear scenes in the training set. For each individual prediction, the first term in Eq. 1.1 is equal to 0 if the scene is clear and the second term is equal to 0 if the scene is cloudy. This loss function designates that, statistically, the detection of clear and cloudy scenes should have equal effect on the overall model skill.

The final cloud mask skill is found to be not highly sensitive to the specific NN architecture used. However, for completeness, the specific structure employed is detailed in Table 1.1. The model includes batch normalization and dropout layers, which are standard techniques used to improve generalization and reduce overfitting during training.

To account for differences between sensors, a total of 16 unique NNs are trained — one for each of the eight sensors onboard the two TIRS instruments. All NNs are trained using the same set of granules, and the model architectures are identical except for the input layer, which is customized to accept only the valid channels available for that specific sensor configuration. Each NN is only used to evaluate its corresponding sensors’ scenes, and their outputs are combined to produce the final cloud mask product.

Table 1.1: The neural network structure for the ML cloud mask. Layer levels may be used as a reference for how inputs progress through the NN. Layer types and output shape are listed for each layer. The input shape takes the form of nchannels, referring to the number of ‘valid’ channels for that particular cross-track scene, with two additional inputs from the AUX-MET data products.

Baseline NN Structure		
Layer Level	Type	Output Shape
1	Input	(None, nchannel+2)
	BatchNormalization	(None, 52)
2	Dense	(None, 256)
	Dropout(0.2)	(None, 256)
3	Dense	(None, 256)
	Dropout(0.3)	(None, 256)
4	Dense (softmax activation)	(None, 2)

Because NN training is inherently stochastic, each of the 16 networks is independently retrained five times, starting from different random initializations. The version with the lowest loss on an independent validation set (i.e., granules not seen during training) is selected for operational use. This approach ensures robust performance and minimizes the impact of training variability.

#### 1.4 Using ML-MSK

The ability to detect clouds is highly dependent on cloud type and the atmospheric and surface variability within a given scene. The primary purpose of 2B-MSK is to confidently identify 80-90% of clear-sky scenes for scales of 15-50 km in the Arctic; thus the cloud detection has been optimized for this need. To ensure clear-sky detection with a high level of confidence, it is recommended that one uses 2B-MSK’s measure of confidence as needed.

Explicitly, a continuous estimate of cloud probability (0-1; `cldmask_probability`) is offered. Probabilities closer to one indicate that the neural network model is more confident that a cloud is present, and probabilities nearer to zero indicate higher confidence that the given TIRS scene is clear. For convenience, these probabilities are also discretized into five integer values corresponding to distinct probability bins:

- 0 [clear] if  $0.0 \leq \text{cldmask\_probability} < 0.2$
- 1 [likely clear] if  $0.2 \leq \text{cldmask\_probability} < 0.4$
- 2 [uncertain] if  $0.4 \leq \text{cldmask\_probability} < 0.6$
- 3 [likely cloud] if  $0.6 \leq \text{cldmask\_probability} < 0.8$

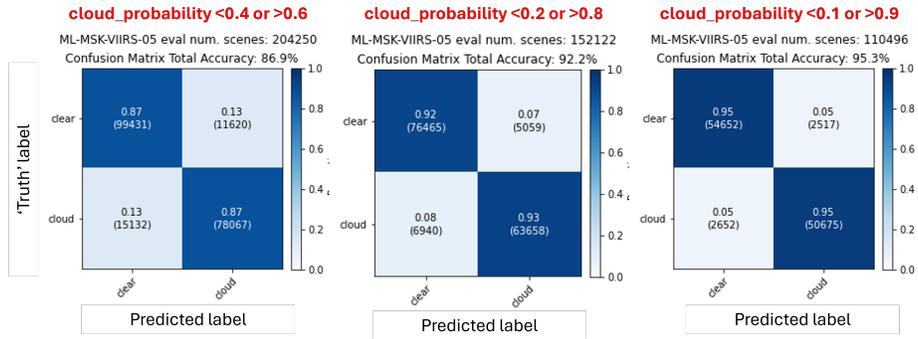


Figure 1.2: Confusion matrix of VIIRS cloud mask labels versus predicted labels from PREFIRE scenes for three different cloud probability thresholds. Left:  $<0.4$  (`cloud_mask = 0` or `1`) or  $>0.6$  (`cloud_mask = 3` or `4`), middle:  $<0.2$  (`cloud_mask = 0`) or  $>0.8$  (`cloud_mask = 4`), right:  $<0.1$  or  $>0.9$  (must be screened by users with the `cldmask_probability` field)

- 4 [cloud] if  $0.8 \leq \text{cldmask\_probability} \leq 1.0$

Users may choose to apply this discretized cloud mask directly — e.g., filtering scenes using only the most confident classes (0 or 4) — or define custom thresholds using the continuous cloud probability field to suit their specific application needs.

## 1.5 Initial Validation

Initial validation of ML-MSK has been performed against the VIIRS cloud mask and CrIMSS (Crosstrack Infrared and advanced technology Microwave Sounder Suite) effective cloud fraction for orbital segments that are not used in the formation of the training data sets. As with the identification of ideal training set radiances, the most optimal comparisons are limited to small time differences (less than 60 minutes) between PREFIRE and VIIRS + CrIMSS data.

Figure 1.2 presents initial validation results for ML-MSK against the VIIRS cloud mask, using several granules from July 2024 and January 2025 that were excluded from the training set. The comparisons are displayed as confusion matrices, with true positives (top left), true negatives (bottom right), false positives (top right), and false negatives (bottom left) indicated. Each matrix reports both absolute counts and prediction-relative frequencies.

Even for lower-confidence classes (left), the average classification accuracy for both clear and cloudy scenes exceeds 87%. As expected, higher confidence thresholds (right) yield improved agreement with the VIIRS reference mask, at the cost of excluding a greater number of ambiguous scenes from the labeled output.

## 1.6 References

Bertossa, C., L'Ecuyer, T., Merrelli, A., Huang, X. and Chen, X., 2023. A neural network-based cloud mask for PREFIRE and evaluation with simulated observations. *Journal of Atmospheric and Oceanic Technology*, 40(4), pp.377-396.

Kahn, B. H., B. J. Drouin, and T. S. L'Ecuyer (2020), Assessment of sampling sufficiency for low-cost satellite missions: Application to PREFIRE, *J. Atmos. Ocean. Tech.*, 37, 2283–2298, doi: 10.1175/JTECH-D-20-0023.1.

## 2 Appendix

### 2.1 Table of variables and symbols

<b>A</b>	averaging kernel matrix
$\alpha$	angular resolution
$\beta$	azimuth angle
$B$	blackbody radiance
$BW$	spectral bandwidth
$\chi$	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
$\varepsilon$	emissivity
$\epsilon$	noise, error
$\phi$	longitude
$E$	irradiance
$F$	flux
$f$	focal length
$\mathcal{F}$	function
$\gamma$	<i>a priori</i> 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
<b>K</b>	Jacobian
$\lambda$	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
$\nu$	frequency
NEdT	Noise-Equivalent delta Temperature
$o$	offset
$\Omega$	solid angle
$p$	pressure
$P$	probability
PWV	Precipitable Water Vapor
$Q$	water vapor
$\rho$	reflection coefficient
$R$	radius, resistance, cost-function change
$\mathfrak{R}$	response function
$\wp$	responsivity
$\sigma_B$	Stefan-Boltzmann constant
$\mathbb{S}$	signal level in digitized counts
$S$	covariance
$SI$	Segmentation Index
SNR	Signal-to-Noise Ratio
SRF	Spectral Response Function
$\theta$	latitude, potential temperature, polar coordinate angle
$\tau$	transmission, optical depth
$T$	temperature
<b>TR</b>	Training Radiances
<b>TREM</b>	TRaining Eigenvector Matrices
$t$	time
$\phi$	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
$\zeta$	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
NE $\delta$ R	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.

Table of Contents	
0.1	PREFIRE algorithm connectivity and flow
Cloud Mask Algorithm	
1.1	Channels used for ML-MSK NNs
1.2	Cloud probability confusion matrices