PREFIRE Data User Guide

Level 2B Cloud Mask (2B-MSK)

Version R01
(20250331)

Cameron Bertossa
Brian H. Kahn
Erin Hokanson Wagner
Tim Michaels
Aronne Merrelli

Table of Contents

1 Introduction

This user guide contains information for the PREFIRE data collections PREFIRE_SAT1_2B-MSK and PREFIRE_SAT2_2B-MSK, which are archived by the Atmospheric Science Data Center (ASDC) at the NASA Langley Research Center. These collections contain cloud mask products presented as certainty for each scene collected by the PREFIRE Thermal Infrared Spectrometers (PREFIRE-TIRS).

1.1 Mission Overview

The Science Mission Directorate (SMD) at NASA Headquarters selected the Polar Radiant Energy in the Far InfraRed Experiment (PREFIRE) as an Earth System Science Pathfinder (ESSP) Earth Venture Instrument (EVI-4) class Mission of Opportunity. Through spectrally resolved observations of radiances spanning the radiatively significant portions of the Mid- and Far-InfraRed (MIR and FIR), PREFIRE addresses two complementary hypotheses:

1. Time-varying errors in both FIR surface emissivity and thermal radiation modulate estimates of energy exchanges between the surface and the atmosphere in the Arctic.
2. These terms are responsible for a large fraction of the spread in projected rates of change for Arctic surface, ocean, and atmosphere characteristics.

These hypotheses are addressed through five related objectives:

O1.1 Quantify snow and ice FIR emissivity spectra and their variability on seasonal scales;
O1.2 Quantify the FIR thermal radiation and its response to seasonal variations in cloud cover / water vapor;
O1.3 Quantify variability in Arctic spectral surface emission and the thermal radiation across the FIR owing to transient cloud and water vapor and sub-daily surface phase-change processes;
O2.2 Quantify thermal emission errors on projected rates of Arctic warming and sea ice loss;
O2.3 Determine the impact of improved surface emissivity on modeled ice sheet dynamic processes on hourly scales.

PREFIRE uses broadband infrared (> 75% of surface emitted thermal radiation) radiance measurements made from the separate orbiting platforms (CubeSats) to address the science objectives. The PREFIRE payloads are two stand-alone instruments built at JPL using heritage from the Mars Climate Sounder and the Moon Mineralogy Mapper. The PREFIRE instruments are thermal infrared imaging spectro-radiometers with more than 50 spectral bands. Each PREFIRE instrument uses ambient temperature thermopile detectors and operates in a pushbroom mode with a point and stare mirror for viewing nadir (Earth), space, and a calibration target. PREFIRE data are calibrated with data from views of the internal calibration target and of space, which are viewed multiple times per orbit.

Soon after launch, the orbit altitude was approximately 531 km for both satellites. However, the PREFIRE CubeSats do not have station-keeping abilities and so their altitudes decrease with time. The current satellite altitude is recorded within the (2B-MSK) data product files as the *sat_altitude* field (in the *Geometry* data group).

The PREFIRE project delivers space-based measurements of radiative fluxes, cloud masks, spectrally variant surface emissivity (ε_λ), and column water vapor (CWV). These are science products with the precision, resolution, and coverage needed to improve our understanding of polar energy balances and Earth-system effects over diurnal and seasonal cycles at scales that capture surface and cloud variability.

During its approximately one-year baseline mission, PREFIRE will capture the natural variability in Arctic and Antarctic CWV and surface temperature. PREFIRE reduces uncertainties in the surface and atmospheric components of the polar energy budget.

1.2 Data Overview

A neural network (NN) approach is used to estimate the likelihood of clouds within each TIRS pixel. The output is a cloud probability value between 0 and 1, where 1 indicates the highest likelihood of cloud presence.

These cloud probability values are then converted into a discrete cloud_mask with integer values from 0 to 4. The thresholds for this conversion are as follows:

- 0 if 0.0 ≤ *cldmsk_probability* < 0.2
- 1 if 0.2 ≤ *cldmsk_probability* < 0.4
- 2 if 0.4 ≤ *cldmsk_probability* < 0.6
- 3 if 0.6 ≤ *cldmsk_probability* < 0.8
- 4 if 0.8 ≤ *cldmsk_probability* ≤ 1.0

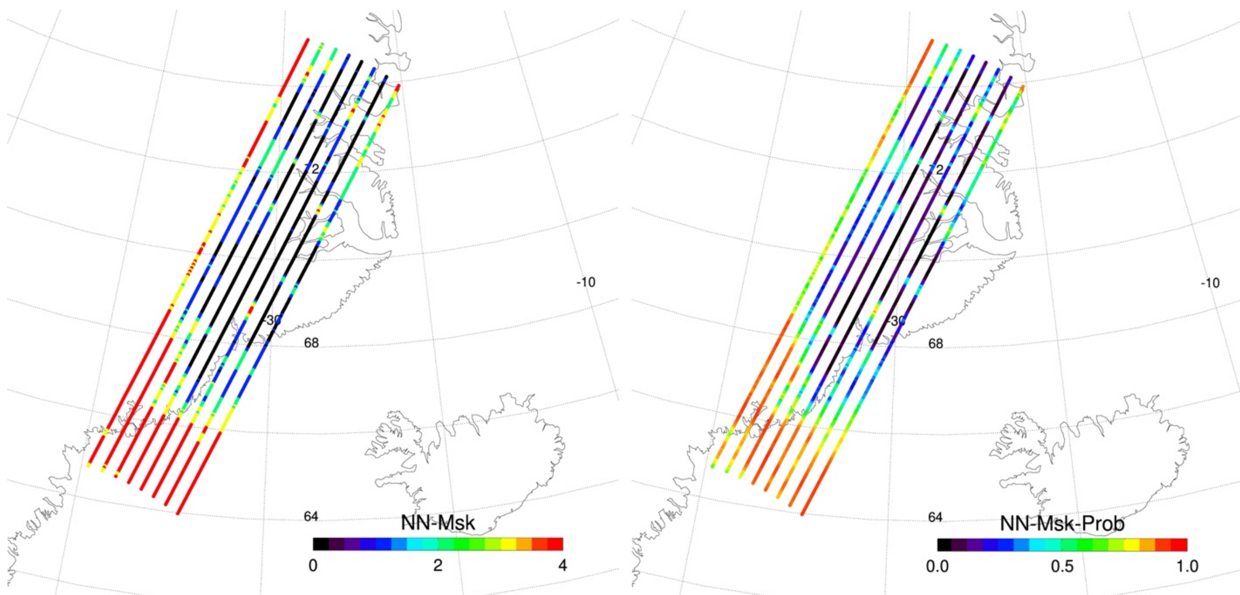

Figure 1-1. (left) Integer values of *cloud_mask* [0,1,2,3,4]. (right) Continuous values of *cldmask_probability* from 0.0 to 1.0. Scenes are from SAT2 granule ID 00659 (2024-07-07). Note that *cldmask_probability* more clearly depicts along-track spatial variability than the discrete *cloud_mask* values. For instance, the nearly uniform area of high-confidence *cloud_mask* in the southernmost parts of the tracks are more varied in *cldmsk_probability (*likely related to spatial variations in cloud altitude, optical thickness, or other characteristics).

By providing both the continuous *cldmsk_probability* and the discretized *cloud_mask*, users can tailor the data to their specific applications. For instance, they may choose to exclude low-confidence detections and retain only pixels with *cloud_mask* values of 0 or 4 (the highest-confidence classes), or alternatively, use the full range of values as a continuous proxy for cloud likelihood. An example overpass of the eastern side of Greenland is shown in Figure 1-1 and illustrates how *cldmsk_probability* and *cloud_mask* offer complementary information about cloud detection likelihood.

To account for differences in channel availability across scenes, a separate neural network is trained and applied for each TIRS scene. While the specific input features may vary depending on which channels are present, overlapping information among channels allows for reliable cloud classification even when some channels are missing.

The neural networks used here are trained using co-located VIIRS scenes over the polar regions. Input features include radiances from available TIRS channels, as well as pixel-level estimates of column water vapor and skin temperature from the PREFIRE AUX-MET data product. These additional inputs from AUX-MET help contextualize the radiance data by accounting for atmospheric conditions that influence the thermal signal, improving the model's ability to distinguish clouds from surface features.

Only scenes where the *observation_quality_flag* is set to 0 ("good") or 1 ("uncertain") in the PREFIRE 1B-RAD granules are used for training and evaluation. No retrieval is performed for pixels where that flag is set to 2 ("bad").

For more details about the neural network architecture, training, and evaluation procedures, refer to the PREFIRE 2B-MSK Algorithm Theoretical Basis Document (ATBD).

1.2.1 Spatial characteristics

The PREFIRE-TIRS instruments collect data continuously in a pushbroom mode, with an integration time of 0.7 seconds for each data frame. Each data frame contains a spectral measurement from each cross-track scene collected simultaneously. Within this continuous data collection, there are planned interruptions due to calibration cycles or data downlinks, and there are also occasional interruptions due to unplanned instrument operations changes or outages. Each calibration cycle takes ~18.7 seconds for PREFIRE-TIRS-1 and ~9.7 seconds for PREFIRE-TIRS2, which implies a gap of approximately 27 and 14 data frames, respectively. Data downlinks create data gaps of up to 13 minutes, and the exact length varies.

Within the orbital swath there are eight distinct tracks of data associated with the eight separate spatial scenes for each PREFIRE-TIRS. The approximate scene footprint sizes are 11.8 km x 34.8 km (cross-track x along-track), with gaps between each scene of approximately 24.2 km. The swath itself is ~264 km across. Note that the scene footprint and swath sizes quoted here are for the orbit altitude soon after launch. However, the footprint size will slowly become smaller as the orbit altitude decreases with time. Do not assume constant footprint or swath dimensions.

PREFIRE-TIRS spatial footprints overlap each other in the along-track dimension. Assuming that no data are missing, any given point along the orbit swath will be observed by up to about 7 overlapping footprints in the along-track direction. The number of footprints that overlap a given footprint will slowly become smaller during the mission, as the satellites' orbital altitudes decrease. Do not assume an integer number of overlapping footprints.

A single data file or granule consists of data collected during approximately one orbit, beginning and ending near the equator to avoid granule borders over the polar regions. Data files are NetCDF4 format and approximately 4 MB in size. These data collections are archived at the ASDC DAAC and can be found at https://asdc.larc.nasa.gov/project/PREFIRE/PREFIRE_SAT1_2B-MSK_R01 and https://asdc.larc.nasa.gov/project/PREFIRE/PREFIRE_SAT2_2B-MSK_R01.

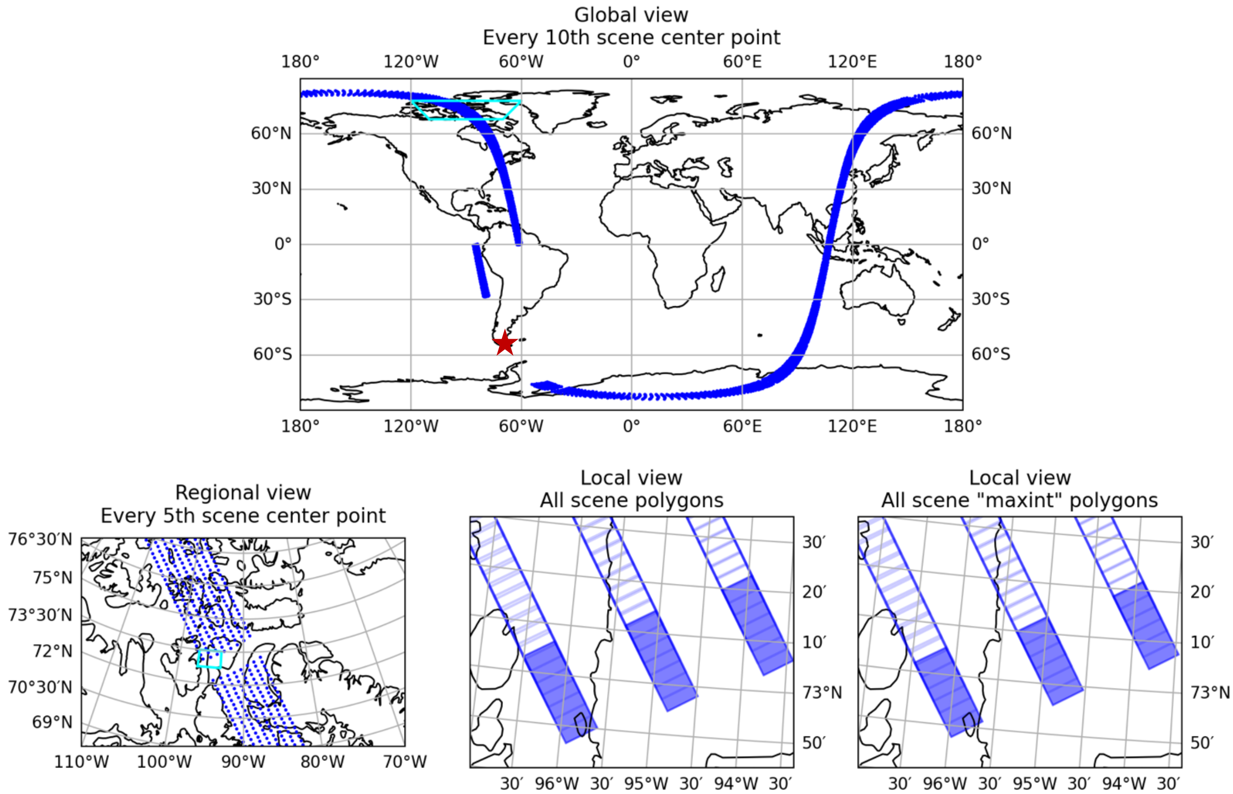

Figure 1-2. An example geolocated orbit (top panel) and focused regional and local plots (bottom panels). The global plot was selected to illustrate a data gap due to a data downlink at the Punta Arenas, Chile ground station, from approximately -70°S to -30°S on the ascending pass at the end of the granule. The zoomed in regional view (lower left) shows the data within the small cyan box in the global plot, and illustrates a smaller data gap due to instrument calibration. The local views (lower middle and right) show the actual scene ground footprint polygons, for the cyan box denoted in the regional view. The first scene's polygon is filled blue, to illustrate the shape of the full field of view (FOV) for one data integration. During the 0.7 second integration time, the satellite moves along track slightly more than 5 km, which means the leading and trailing edges of the instantaneous FOV have translated forward by the same amount. The lower right plot shows the "max integration" footprint polygon, which includes the interior portion of the scene footprint that was within the sensor field of view for the entire integration period.

1.3 Purpose

The PREFIRE Level 2B Cloud Mask (2B-MSK) data provide cloud detection information to users and to downstream PREFIRE algorithms that require knowledge of sky/cloud conditions.

2 Product Description

The Level 2B Cloud Mask (2B-MSK) product provides cloud condition information in the form of continuous probabilities and a discretized cloud mask. The input data for these collections are also available at the ASDC DAAC. They are as follows:

1. PREFIRE Level 1B Radiance (collections PREFIRE_SAT1_1B-RAD and PREFIRE_SAT2_1B-RAD
2. PREFIRE Auxiliary Meteorology (collections PREFIRE_SAT1_AUX-MET and PREFIRE_SAT2_AUX-MET)

2.1 File Specifications

2.1.1 File naming convention

File names for this collection follow the following convention:

PREFIRE_SAT*<satID>_<product ID>_<collection version>_<internal product version_<YYYYMMDDhhmmss>_<granule-ID>*.nc

For example, a representative Level 2B Cloud Mask (2B-MSK) product granule collected by PREFIRE-SAT1 on June 1, 2024 would have the following filename:

PREFIRE_SAT1_2B-MSK_R01_P00_20240601185321_00123.nc

2.1.2 File format

PREFIRE 2B-MSK data product files are created in NetCDF4 format with standard metadata. These files can be read with standard NetCDF libraries available in all popular scripting languages and many data visualization programs.

2.1.3 Quality flag and bitflags conventions

Table 2-3 describes the *msk_quality_flag* and *msk_qc_bitflags* that contain FOV-specific information about the overall quality of the cloud mask determination.

2.1.4 Variables

The variable specifications for this collection are described below, with one table devoted to each top-level data group in the NetCDF4 file: *Geometry*, *Msk*. Note that the *Geometry* group, including all variables, is propagated to every downstream Level 2 data product from the Level 1B Radiance product (1B-RAD).

2.1.5 Variable dimensions

A summary of all array dimensions is given in Table 2-1. The *xtrack* dimension is equal to the number of cross-track scenes (8, for both instruments) and the *atrack* dimension is equal to the number of along-track Earth observation data frames in the product. The number of along-track frames varies from orbit to orbit, depending on the timing of downlink contacts, calibration data, and other rarer events. Generally, the maximum is around 7700–7900 frames in one product file, with substantially fewer in granules containing downlinks or unplanned instrument/spacecraft events.

Table 2--1

Dimension	**Abbreviation**
Along-track	*atrack*
Cross-track	*xtrack*
UTC parts	*UTC_parts* (= 7)
FOV (footprint) vertices	*FOV_vertices* (= 4)
Dimension label	**Definition (C-order)**
1D	(*atrack*)
1Dx	(*xtrack*)
2D	(*atrack*, *xtrack*)
2Du	(*atrack*, *UTC_parts*)
3Dv	(*atrack*, *xtrack*, *FOV_vertices*)

2.1.5.1 *Geometry* group

The *Geometry* data group consists of all timing, observation geometry, and geolocation variables produced during Level-1B processing (see Table 2-2). This data group and its contents will be replicated within any relevant downstream product (e.g., Level-2 data products), rather than stored as a separate geometry file.

Users of NetCDF software packages that try to automatically decode times should be aware that these packages may incorrectly interpret the *ctime* variable as a UTC time. The *ctime* variable is a count of total fractional SI seconds since the epoch 2000-01-01T00:00:00 UTC (i.e., no leap second adjustments since that epoch), while the UTC time standard is adjusted to account for all leap seconds. For example, when the PREFIRE *Geometry* group is read by the open_dataset function of the Python xarray package using the default decode_times=True argument, the resulting *ctime* values (with datetime64 data type) will differ from the *time_UTC_values* variable by the number of leap seconds that occurred between 2000-01-01T00:00:00 UTC and the observation time. Users of xarray and other packages that exhibit this behavior are recommended to use *ctime* along with *ctime_minus_UTC* to calculate UTC times if desired, and/or consult *time_UTC_values* to verify the correct UTC timestamps of PREFIRE observations.

For example, for an xarray dataset, a datetime64 DataArray could be computed as follows:

```
import xarray as xr
ds = xr.open_dataset({path_to_1B-RAD_product}, group='Geometry')
ds['UTC_dt64'] = ds['ctime'] – ds['ctime_minus_UTC']
```

Further details on the handling of leap seconds in the Climate and Forecast (CF) NetCDF Metadata Conventions can be found in Section 4.4.1 of the CF-1.9 Conventions: https://cfconventions.org/Data/cf-conventions/cf-conventions-1.9/cf-conventions.html#calendar.

Table 2--2

Variable Name	**Type**	**Dimen-sion**	**Units**	**Description**
obs_ID	int64	2D		unique integer identifier for each TIRS look YYYYMMDDhhmmsstbd, composed of UTC date (YYYYMMDD) and time (hhmmss) at TIRS image integration midpoint, t = tenths of seconds [0–9], b = satellite number [1–2], d = scene number [1–8]
ctime	float64	1D	seconds	continuous time since the epoch 2000-01-01T00:00:00 UTC (i.e., similar to TAI) at the midpoint of each TIRS image integration
ctime_minus_UTC	int8	1D	seconds	continuous time minus UTC (i.e., leap seconds since the ctime epoch) at the midpoint of each TIRS image integration
time_UTC_values	int16	2Du	various	UTC datetime at the midpoint of each TIRS image integration, represented as an integer array. Array parts: year, month, day, hour, minute, second, millisecond
latitude	float32	2D	degrees_north	topography-corrected latitude of FOV centroid
longitude	float32	2D	degrees_east	topography-corrected longitude of FOV centroid
vertex_latitude	float32	3Dv	degrees_north	topography-corrected latitude for each of the four vertices/corners (arranged counterclockwise starting at the trailing-left corner) of a 4-sided polygon that closely approximates the geolocated FOV (orbital motion taken into account)
vertex_longitude	float32	3Dv	degrees_east	topography-corrected longitude for each of the four vertices/corners (arranged counterclockwise starting at the trailing-left corner) of a 4-sided

				polygon that closely approximates the geolocated FOV (orbital motion taken into account)
land_fraction	float32	2D		land_area / total_area (remainder is water_area) within the FOV, according to the Digital Elevation Model (DEM)
elevation	float32	2D	m	mean topographic elevation within the FOV
elevation_stdev	float32	2D	m	standard deviation of topographic elevation within the FOV
viewing_zenith_angle	float32	2D	degrees	viewing zenith angle at the FOV centroid
viewing_azimuth_angle	float32	2D	degrees	viewing azimuth angle at the FOV centroid (zero is north, clockwise-positive looking down from the zenith)
solar_zenith_angle	float32	2D	degrees	solar zenith angle at the FOV centroid
solar_azimuth_angle	float32	2D	degrees	solar azimuth angle at the FOV centroid (zero is north, clockwise-positive looking down from the zenith)
solar_distance	float64	2D	km	distance from FOV centroid to the solar barycenter
subsat_latitude	float32	1D	degrees_ north	sub-satellite latitude
subsat_longitude	float32	1D	degrees_ east	sub-satellite longitude
sat_altitude	float32	1D	km	satellite altitude above the reference ellipsoid (at the sub-satellite point)
sat_solar_illumination_flag	int8	1D		flag specifying whether the spacecraft is illuminated by the sun; 0=no, 1=partial, 2=full
geoloc_quality_bitflags	uint16	2D		integer composed of bit flags that contain info about the quality of the overall geolocation of each along-track frame of scenes
maxintgz_verts_lat	float32	3Dv		latitude (topography-corrected) for each of the four vertices/corners (arranged counterclockwise starting at the trailing-left corner) of a 4-sided polygon that closely approximates the geolocated zone with the maximum TIRS image integration time
maxintgz_verts_lon	float32	3Dv		longitude (topography-corrected) for each of the four vertices/corners (arranged counterclockwise starting at the trailing-left corner) of a 4-sided polygon that closely approximates the geolocated zone with the maximum TIRS image integration time
orbit_phase_metric	float32	1D	degrees	orbit phase angular metric (range of 0-360 degrees, varying approximately

				linearly with time), defined as 0 deg at the ascending node (northward equator crossing) of the satellite orbit, 180 deg at the descending node, and so on
satellite_pass_type	int8	1D		flag specifying which type of satellite pass each frame is mostly/all part of. −1 = descending, 1 = ascending

2.1.5.2 *Msk* group (Cloud Mask)

The specifications for the cloud mask, cloud probability, quality flag, and bitflags are provided in Table 2-3.

Table 2--3

Variable Name	**Type**	**Dimension**	**Units**	**Description**
cloud_mask	int8	2D		consists of integers indicating cloud conditions in graduated confidence categories integer_meanings = [0] clear [1] likely clear [2] uncertain [3] likely cloud [4] cloud
cldmask_probability	float32	2D		Cloud probability in a continuous fractional value in the range [0-1]
msk_quality_flag	int8	2D		flag specifying the overall quality of the cloud mask determination(s) for each FOV flag_values = 0b, 1b flag_meangs = [0] nominal [1] not attempted due to radiance_quality_flag value
msk_qc_bitflags	uint16	2D		integer composed of bit flags that contain additional detail about the quality of the cloud mask determination(s)/fields value_meaning = [b0] based on best-quality radiances [b1] based on uncategorized radiances [b2] cloud mask determination not attempted due to *radiance_quality_flag* value

3 Updates Since Previous Version

None – this is the initial version.

4 Known Issues

Geolocation:

The GPS receiver on PREFIRE-SAT1 has performed poorly since launch, and the GPS receiver on PREFIRE-SAT2 ceased to function well at the end of August 2024. Because of the lack of continuously reliable GPS position and time data, the time-dependent orbital position and velocity vectors used for geolocation are based on orbital reconstructions. This uses publicly available orbit element sets (e.g., Two-Line Element sets (TLEs) based on ranging observations by the United States Space Force and other entities. The precision and accuracy of the orbit reconstruction is currently undergoing evaluation. In addition, residual uncertainties exist due to pointing offsets from lack of precise knowledge of the spectrometer slit orientation relative to the spacecraft. These uncertainties will be addressed after the orbit reconstruction is evaluated and optimized. The current best estimate is that individual geolocated scenes could have along-track geolocation errors of up to 50 km with an average of approximately 30 km (less than the along-track dimension of a ground footprint). The cross-track geolocation error has not been quantified, but the error is likely to be less than the cross-track scene width (approximately 12 km), based on favorable spatial correlations with co-located geostationary imagery collected in the MIR atmospheric window.

As more PREFIRE-TIRS data are collected and analyzed, the quantification of the geolocation biases will improve. Further refinements of the geolocation algorithm are planned, which will reduce these errors in future 1B-RAD and downstream data product releases.

Electronic pattern noise:

Electrical cross talk between adjacent FPA detectors was largely mitigated by alternating the wiring polarity in the readout integrated circuits. However, residual pattern noise has been noted in both the raw data and the calibrated radiances. This noise is highly temporally correlated and impacts all spectral channels.

This electrical noise manifests in two primary ways. First, “sawtooth-like” patterns can be visible in an individual spectral observation, where the even and odd spectral channels have different radiometric biases. These patterns are generally visible in spectral residuals (observation – modeled radiance). Due to the temporal correlation this pattern could be visible in multiple consecutive frames. Second, “striping” is visible when data from a selected channel are viewed spatially, where specific spatial scenes are clearly biased relative to the other scenes. Again, due to the temporal correlation these stripes will continue along track for some time. No data flagging is performed related to this pattern noise effect, but future developments in the calibration algorithm are planned to further reduce this noise.

5 Resources

The Algorithm Theoretical Basis Document (ATBD) can be found at https://prefire.ssec.wisc.edu/Documents/PREFIRE_2B-MSK_ATBD.pdf. For more information, contact Erin Hokanson Wagner at prefire-sdps.admin@office365.wisc.edu.

6 References

Bertossa, Cameron, et al. "A neural network–based cloud mask for PREFIRE and evaluation with simulated observations." *Journal of Atmospheric and Oceanic Technology* 40.4 (2023): 377-396.

L'Ecuyer, T.S., Drouin, B.J., Anheuser, J., Grames M., Henderson, D., Huang, X., Kahn, B.H., Kay, J.E., Lim, B.H., Mateling, M., Merrelli, A., Miller, N.B., Padmanabhan, S., Peterson, C., Schlegel, N.-J., White, M.L., Xie, Y., "The Polar Radiant Energy in the Far-InfraRed Experiment: A New Perspective on Polar longwave Energy Exchanges," *Bulletin of the American Meteorological Society (BAMS*), 102(7), E1431–E1449, 2021.

Padmanabhan, S., Drouin, B., L'Ecuyer T., White, M., Lim. B., Kenyon, M., Mariani, G., McGuire J., Raouf, N., De Santos, O., Bendig, R., "The Polar Radiant Energy in the Far-InfraRed Experiment (PREFIRE)," *IGARSS 2019 – 2019 IEEE International Geoscience and Remote Sensing Symposium*

Citation
DOI:

- PREFIRE_SAT2_2B-MSK_R01: 10.5067/PREFIRE-SAT2/PREFIRE/MSK_L2B.R01
- PREFIRE_SAT1_2B-MSK_R01: 10.5067/PREFIRE-SAT1/PREFIRE/MSK_L2B.R01