MEaSUREs Precipitation Phase and Rate from Passive Microwave Observations and Earth System Models using Transfer Learning, Version 1

USER GUIDE

How to Cite These Data

As a condition of using these data, you must include a citation:

Subedi, B., M. Garshasbi, and A. Ebtehaj. 2026. *MEaSUREs Precipitation Phase and Rate from Passive Microwave Observations and Earth System Models using Transfer Learning, Version 1*. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/L4S3PU4C9XU7. [Date Accessed].

FOR QUESTIONS ABOUT THESE DATA, CONTACT NSIDC@NSIDC.ORG

FOR CURRENT INFORMATION, VISIT https://nsidc.org/data/NSIDC-0804

National Snow and Ice Data Center

TABLE OF CONTENTS

1 DATA DESCRIPTION

1.1 Summary

This data set, part of the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Program, contains estimates of precipitation phase (rain, snow, or clear) and total precipitation rate at approximately 5.1 km × 13.2 km resolution along the GPM[1] Microwave Imager (GMI) S1 swath.

The data are generated using GMI brightness temperatures (T_bs) and the TLP-R2S[2] framework, a supervised machine learning algorithm that retrieves precipitation phase and rate by transferring prior information from reanalysis data to satellite passive microwave observations.

Data files also include the GMI T_bs and ERA5[3] ancillary data used to train TLP-R2S, plus precipitation rates from ERA5 and the Goddard Profiling Algorithm (GPROF) for comparison.

1.2 Parameters

Precipitation phase (clear, rain, or snow)
Total surface precipitation rate (mm/h)

1.3 File Information

1.3.1 Format

NetCDF-4

1.3.2 File Contents

As shown in Figure 1 below, NetCDF data files are organized into a hierarchy of folders and subfolders. Latitudes and longitudes at the centers of grid cells plus along-track scan times are stored at the top level, along with the “mapping” variable which contains a complete description of the projection.

The contents of the “TLP-R2S,” “Input,” and “Reference” folders are described in the following sections.

[1] Joint NASA/Japan Aerospace Exploration Agency Global Precipitation Measurement (GPM) Core Observatory
[2] Transfer Learning of global Precipitation from Reanalysis data to Satellite observations
[3] European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5

1.3.2.1 TLP-R2S

This folder contains the primary variables of interest—precipitation phase (“Phase”), stored as 0 (clear), 1 (rain), or 2 (snow), and total precipitation rate at Earth’s surface (“SurfacePrecip”) in mm/h.

1.3.2.2 Input

Input data to TLP-R2S are stored within three subfolders. The “ERA5” subfolder contains the ERA5 reanalysis variables listed in Table 1.

GMI T_bs are stored in the “GMI” subfolder. The nine lower GMI frequency channels that comprise the S1 swath are located in “TbS1”; The four higher GMI frequency channels from the S2 swath are in “TbS2”. GMI T_b variable names and descriptions (i.e., frequencies and polarizations) are listed in Table 2, where “h” and “v” in the variable names indicate horizontal or vertical polarization.

The third subfolder, “GTOPO30,” contains elevation (“Elevation”) and aspect (“Aspect”) data derived from the USGS Global 30 Arc-Second Elevation (GTOPO30) digital elevation model.

1.3.2.3 Reference

ERA5 mean snowfall rates (“msr”) and mean total precipitation rates (“mtpr”) at Earth’s surface are stored in the “ERA5” subfolder. Similarly, the “GPROF” subfolder contains solid-phase precipitation rates (“FrozenPrecip”) and total precipitation rates (“SurfPrecip”) at the Earth's surface retrieved by the Goddard Profiling Algorithm. These data are provided for comparison to the TLP-R2S output.

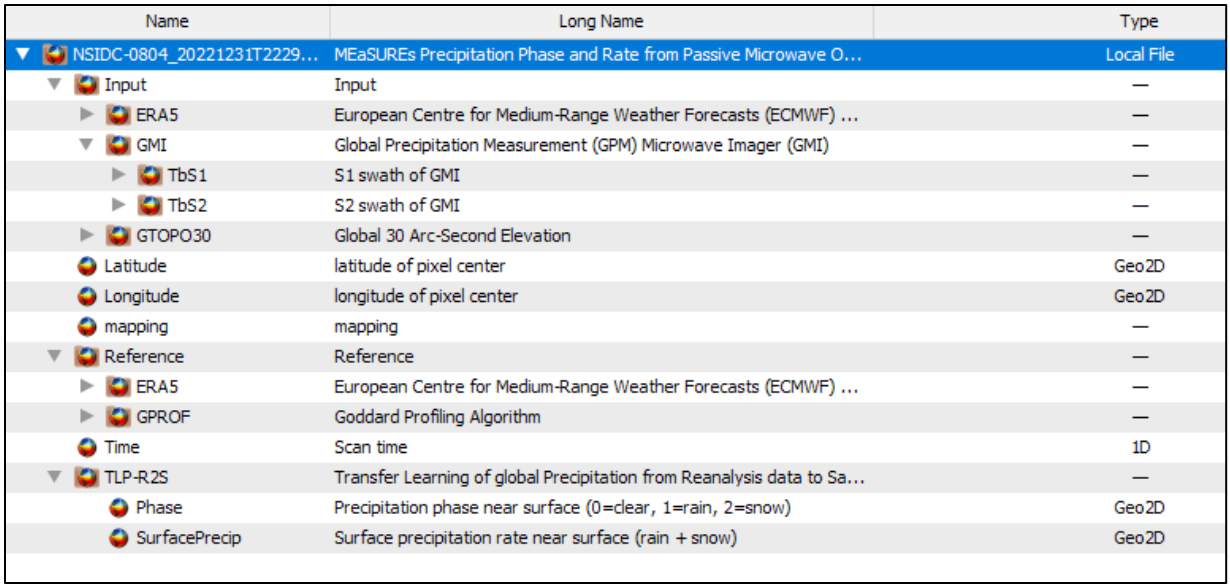

Figure 1. NetCDF File Internal Structure

ⓘ The units for the ERA5 "rsn" variable (see Table 1 below) are incorrectly specified in the "units" attribute as "m." The correct units for this variable are kg/m^3

Table 1. ERA5 Variable Names and Descriptions

Variable Name	Description
asn	Snow Albedo: Fraction of the solar radiation in the grid cell reflected by snow-covered surfaces, ranging from 0.0 – 1.0.
cape	Convective Available Potential Energy (J/kg)
cin	Convective Inhibition (J/kg)
lsm	Land-Sea Mask (1 = land, 0 = ocean)
rsn	Snow Density (kg/m3) **Note:** The "units" attribute for this variable incorrectly specifies the units as "m."
sd	Snow Depth in meters of water equivalent (kg/m2)
siconc	Sea Ice Area Fraction: Fraction of the grid cell area covered by sea ice, ranging from 0.0 – 1.0
skt	Skin temperature: Theoretical temperature required to satisfy the surface energy balance at the land or sea surface (K)
swvl1	Volumetric Soil Water Layer 1: Volumetric water content in the uppermost soil layer (0–7 cm depth), expressed as the volume of water per unit volume of soil (m^3/m^3)
t2m	Two meter temperature (K)
tciw	Total Column Cloud Ice Water: Vertically integrated mass of ice within clouds per unit area in the entire atmospheric column (kg/m^2)

Variable Name	Description
tclw	Total Column Cloud Liquid Water: Vertically integrated mass of liquid water droplets within clouds per unit area in the entire atmospheric column (kg/m^2)
tcslw	Total Column Supercooled Liquid Water: Vertically integrated mass of liquid water < 0° C in the entire atmospheric column (kg/m^2)
tcw	Total Column Water: Total mass of water in all phases (vapor, liquid, and ice) per unit area in the entire atmospheric column (kg/m^2)
tcwv	Total Column Water Vapor: Total mass of water vapor per unit area in the entire atmospheric column (kg/m^2)
u10	U Wind Component: Eastward (west-to-east direction) component of the 10 m wind vector (m/s)
v10	V Wind Component: Northward (south-to-north direction) component of the 10 m wind vector (m/s)

Table 2. GMI T_b Variable Names and Descriptions

Folder	Variable Name	Channel Central Frequency
TbS1	10.6h	10.65 GHz
	10.6v	10.65 GHz
	18.7h	18.7 GHz
	18.7v	18.7 GHz
	23v	23.8 GHz
	37h	36.5 GHz
	37v	36.5 GHz
	89h	89.0 GHz
	89v	89.0 GHz
TbS2	166h	166.5 GHz
	166v	166.5 GHz
	183±3	183.31 ± 3 GHz (vertical polarization)
	183±7	183.31 ± 7 GHz (vertical polarization)

1.3.3 File Naming Convention

Example

NSIDC-0804_20150101T012932-20150101T030203_V01.0.nc

NSIDC-0804_20150101T030204-20150101T043435_V01.0.nc

NSIDC-0804_20150101T043436-20150101T060707_V01.0.nc

Convention

NSIDC-0804_[start date]T[start time]_[end date]T[end time]_V01.0.nc

The following table describes the variables in the file naming convention:

Table 3. File Name Variables and Descriptions

Variable	Description
NSIDC-0804	“MEaSUREs Precipitation Phase and Rate from Passive Microwave Observations and Earth System Models using Transfer Learning” data set
[start date]T[start time]	Start date (YYYYMMDD), T (time follows), start time (HHMMSS)
[end date]T[end time]	End date (YYYYMMDD), T (time follows), end time (HHMMSS)
V01.0	Major and minor version number. V01.0 indicates Version 1.0
nc	NetCDF file extension

1.4 Spatial Information

1.4.1 Coverage

N: 69.41° N

S: 69.42° S

E: 180° E

W: 180° W

1.4.2 Resolution

5.1 km × 13.2 km

1.4.3 Geolocation

World Geodetic System 1984 (EPSG:4326)

1.5 Temporal Information

1.5.1 Coverage

7 April 2014 – 6 May 2025

1.5.2 Resolution

Approximately 90 minutes

2 DATA ACQUISITION AND PROCESSING

2.1 Acquisition

This data set was generated using satellite acquisitions and atmospheric reanalyses from the following sources:

- Global Precipitation Measurement (GPM) Microwave Imager (GMI)
- Global Precipitation Measurement (GPM) Dual-frequency Precipitation Radar (DPR)
- Cloud Profiling Radar (CPR)
- ECMWF Reanalysis v5 (ERA5)

2.2 Processing

In the TLP-R2S framework, a baseline model consisting of Extreme Gradient Boosting (XGBoost[4]) decision trees is first trained on ERA5 precipitation labels, by minimizing a multi-class cross-entropy loss function with hyperparameters optimized by a Bayesian optimization algorithm. The model is then expanded by adding additional decision trees, or boosters, and training them on coincident satellite retrievals by the GPM Dual-frequency Precipitation Radar (DPR) and the Cloud Profiling Radar (CPR) on board the NASA CloudSat satellite.

Once precipitation phase classification is complete, two XGBoost regression models are then used to estimate precipitation rates for rainfall and snowfall. Determining precipitation rates follows a similar two-stage incremental learning process as phase classification, but uses rainfall rates/snowfall rates from CPR/DPR and minimizes root mean square error.

[4] XGBoost is an open-source, optimized distributed gradient boosting library that implements machine learning algorithms under the Gradient Boosting framework. For more information, see XGBoost Documentation.

For a complete description of the TLP-R2S framework, see Subedi at al. (2026). Source code, training data, and interactive demos are available at the following GitHub repositories:

- https://github.com/Buddha-subedi/PMWPrecip_TLP-R2S
- https://github.com/MahyarGarshasbi/TLP-R2S-ATMS/

3 VERSION HISTORY

V1 (Initial release)

4 REFERENCES

Garshasbi, M., B. Subedi, A. Ebtehaj, M. Lisa, F. J. Turk, G. Huffman (2026). Incremental Learning for Passive Microwave Precipitation Retrievals using Advanced Technology Microwave Sounder. [Manuscript submitted for publication]

Subedi, B., M. Garshasbi, A. Ebtehaj, M. Lisa, N. Utsumi, G. Huffman (2026). An Incremental Learning Framework Linking Reanalysis and Radar Data for Passive Microwave Precipitation Retrievals. [Manuscript submitted for publication]

5 DOCUMENT INFORMATION

5.1 Publication Date

April 2026

5.2 Date Last Updated

April 2026