JPL D-109746

MAIA Surface Monitor Data Product Specification

Version 1.0.0

Sina Hasheminassab
David J. Diner
Scott Gluck
James McDuffie
Sebastian Val

Apr 14, 2025

CONTENTS

CHAPTER

ONE

INTRODUCTION

1.1 Purpose

The Multi-Angle Imager for Aerosols (MAIA) Surface Monitor Data Product comprises ground-based measurements of total PM2.5, total PM10, and chemically speciated PM2.5 (including sulfate, nitrate, elemental carbon (EC), black carbon (BC), organic carbon (OC), and dust) at various locations within MAIA's globally distributed target areas. Surface monitors are fixed-location, ground-based stations that provide accurate information on local ambient particulate matter (PM) levels and composition. Surface monitor measurements serve as ground truth for transforming MAIA satellite data into estimates of near-surface PM mass concentrations. MAIA leverages existing PM monitoring networks where available and supplements them with additional equipment in select target areas. More details on the sources and types of PM monitors used in the MAIA project can be found here: https://maia.jpl.nasa.gov/mission/#surface_networks

The Surface Monitor Data Product is provided in NetCDF format and organized into individual files by PM type (total or speciated), with one file per day for each type. The file contents for a given MAIA target area contains measurements from all PM monitoring sites within that target area. Each file includes both raw and quality-controlled data. This document describes the file format, naming conventions, and data fields of the MAIA Surface Monitor Data Product, which will be used in Level 2 (L2) and Level 4 (L4) PM processing for the MAIA investigation. The Surface Monitor Data Product may also be used for independent research.

1.2 MAIA Data Products

The MAIA Science Data System (SDS) produces data products that support the mission's science objectives. These products enable epidemiological studies associating PM exposure with health impacts. The MAIA Science Computing Facility (SCF) at JPL provides a test and development environment, while the NASA Langley Research Center (LaRC) handles operational processing for the MAIA SDS. The Atmospheric Science Data Center (ASDC) is responsible for data archival and distribution as part of NASA's Earth Science Data and Information System (ESDIS).

1.3 Controlling Documents

[M-1] MAIA Level 2 Particulate Matter and Level 4 Gap-Filled Particulate Matter Retrieval Algorithm Theoretical Basis, D 101518, Rev. B.

CHAPTER

TWO

FILE NAMING CONVENTION

This product's filenames have the following naming convention:

```
MAIA_<ProductType>_<PMType>_<TargetShortName>_<DateTimeStamp>_<FormatVersion>_<DataVersion>_<FileVersion>.nc
```

Where:

- <*ProductType*> is always "ANC_SURFACEMONITOR" for this product.
- <*PMType*> indicates the type of particulate matter data:
  - "PM_TOTAL" for PM10 and PM2.5 Total Mass Concentration data.
  - "PM_2.5_SPECIES" for speciated PM2.5 data, which includes sulfate, nitrate, EC, BC, OC, and dust.
- <*TargetShortName*> refers to the abbreviated name of the target area, which consists of:
  - A three-letter country code (per ISO-3166-1).
  - The name of the representative city, separated by a dash (-).
  - A complete list of Primary Target Areas (PTA) and Secondary Target Areas (STA) <*TargetShortName*> values are available on the MAIA website: https://maia.jpl.nasa.gov/mission/#target_areas
- <*DateTimeStamp*> represents the timestamp of the PM measurement day in Coordinated Universal Time (UTC), formatted as YYYYMMDDThhmmssZ, where:
  - *YYYY* – Four-digit calendar year.
  - *MM* – Two-digit month (zero-padded if necessary).
  - *DD* – Two-digit day of the month (zero-padded if necessary).
  - *T* – Separator between date and time (mandatory in product filenames).
  - *hh* – Two-digit hour (24-hour format, UTC, zero-padded).
  - *mm* – Two-digit minute (UTC, zero-padded).
  - *ss* – Two-digit second (UTC, zero-padded).
- <*FormatVersion*> is a two-digit, zero-padded number representing the format version. The initial version will be "F01", with subsequent updates proceeding sequentially, such as "F02", "F03", and so on.
- <*DataVersion*> is a three-part version number, separated by a lowercase "p", following the structure V<major>p<minor>p<patch>. The initial version will be "V00p00p00", with updates incrementing based on major, minor, and patch releases as needed.
- <*FileVersion*> is a three-digit, zero-padded number representing the file version. The initial version will be "V001", with subsequent updates proceeding sequentially, such as "V002", "V003", and so on.

Example for the file containing PM10 and PM2.5 Total Mass Concentration data:

```
MAIA_ANC_SURFACEMONITOR_PM_TOTAL_USA-LosAngeles_20190531T000000Z_F01_V01p01p01_V001.nc
```

Example for the file containing PM2.5 Species data:

```
MAIA_ANC_SURFACEMONITOR_PM_2.5_SPECIES_USA-LosAngeles_20190531T000000Z_F01_V01p01p01_
↪V001.nc
```

CHAPTER THREE

DIMENSIONS

Each PM type or species is associated with one or both of the following dimensions:

- Monitor Dimension: Represents the number of surface monitoring stations in a target area that provide data for a specific PM type or species.
- Measurement Dimension: Represents the number of individual measurements used in the daily average for each PM type or species.

The following table describes the dimension short codes used in the document to describe the shape of datasets.

Table 1: Dataset dimensions for the Surface Monitor Stage 0 files

Dim name.	Description
pm10_monitor	Dimension for number of PM10 monitors in the target area
pm10_measurement	Dimension for number of PM10 measurements used in the daily average
pm25_monitor	Dimension for number of PM2.5 monitors in the target area
pm25_measurement	Dimension for number of PM2.5 measurements used in the daily average
sulfate_monitor	Dimension for number of Sulfate monitors in the target area
sulfate_measurement	Dimension for number of Sulfate measurements used in the daily average
nitrate_monitor	Dimension for number of Nitrate monitors in the target area
nitrate_measurement	Dimension for number of Nitrate measurements used in the daily average
oc_monitor	Dimension for number of Organic Carbon monitors in the target area
oc_measurement	Dimension for number of Organic Carbon measurements used in the daily average
ec_monitor	Dimension for number of Elemental Carbon monitors in the target area
ec_measurement	Dimension for number of Elemental Carbon measurements used in the daily average
bc_monitor	Dimension for number of Black Carbon monitors in the target area
bc_measurement	Dimension for number of Black Carbon measurements used in the daily average
dust_monitor	Dimension for number of monitors in the target area used in the “Dust” calculation
aluminum_monitor	Dimension for number of monitors reporting aluminum used in the daily average “Dust” calculation
aluminum_measurement	Dimension for number of aluminum measurements used in the daily average “Dust” calculation
calcium_monitor	Dimension for number of monitors reporting calcium used in the daily average “Dust” calculation
calcium_measurement	Dimension for number of calcium measurements used in the daily average “Dust” calculation
silicon_monitor	Dimension for number of monitors reporting silicon used in the daily average “Dust” calculation
silicon_measurement	Dimension for number of silicon measurements used in the daily average “Dust” calculation
titanium_monitor	Dimension for number of monitors reporting titanium used in the daily average “Dust” calculation
titanium_measurement	Dimension for number of titanium measurements used in the daily average “Dust” calculation
iron_monitor	Dimension for number of monitors reporting iron used in the daily average “Dust” calculation
iron_measurement	Dimension for number of iron measurements used in the daily average “Dust” calculation

CHAPTER

FOUR

ATTRIBUTES

The following file level attributes are used to identify the file type and algorithm used for processing:

Table 2: Product identification metadata

Attribute Name	Value
title	MAIA Surface Monitor Stage 0
source	Surface Monitor Database
institution	MAIA products are produced by the MAIA Science Team, using processing and storage facilities of the NASA Langley Research Center ASDC.
instrument	MAIA
acknowledgment	Support for this research was provided by NASA
processing_level	Stage 0

CHAPTER

FIVE

FIELDS COMMON TO ALL PM TYPES

The following sections describe fields that are common to all Surface Monitor Data Product files, including both "PM_TOTAL" (PM10 and PM2.5) and "PM_2.5_SPECIES" (sulfate, nitrate, EC, BC, OC, and dust). Some parameters in the Surface Monitor Data Product files are directly used as inputs to the Geostatistical Regression Models (GRM) [M-1], while others are ingested into the Geographic Information Visualization Tool (GIVT) (https://maia.jpl.nasa.gov/mmgis/) for visualization purposes and assessment of the health of the MAIA surface monitoring network.

5.1 Core Measurement Fields

"date_LST" is the date of the surface monitor measurement in YYYYMMDD format, based on local standard time (LST). No adjustment of the local standard time for daylight savings is made in any of the target areas.

"pm_mass_concentration" is the daily average PM mass concentration in micrograms per cubic meter (µg/m3). The daily average is calculated over a 24-hour time window, which extends from either 00:00 to 24:00 or 09:00 to 09:00 LST, depending on the target area [M-1].

"measurement_id" is a unique identification string for each individual surface monitor measurement, constructed as follows:

```
<PMType>_<TargetShortName>_<DateTimeStamp>
```

Where:

- *<PMType>* is either "TOTAL" or "SPECIES"
- *<TargetShortName>* refers to the abbreviated name of the target area as described in section 2. The list of *<TargetShortName>* values is available on the MAIA website: https://maia.jpl.nasa.gov/mission/#target_areas.
- *<DateTimeStamp>* is the timestamp of the PM measurement day in Coordinated Universal Time (UTC), formatted as `YYYYMMDDThhmmssZ` as described in section 2.

5.2 Monitor Information

The “Monitor_Information” group contains metadata about the surface monitoring stations included in the file. These fields are essential for scientific analysis, validation, and quality assurance of data products that rely on surface monitors.

“maia_surface_monitor_id” is a unique identification string assigned to each surface monitor for tracking and reference within the MAIA project. It is constructed as follows:

```
<TargetShortName>-<Latitude>_<Longitude>_<DataSourceCode><POC>
```

Where:

- *<TargetShortName>* refers to the abbreviated name of the target area as described in section 2. The list of *<TargetShortName>* values is available on the MAIA website: https://maia.jpl.nasa.gov/mission/#target_areas
- *<Latitude>* and *<Longitude>* represent the coordinates of the surface monitoring site in decimal format. Decimal latitude and longitude values are typically positive or negative and do not require a cardinal direction. However, because surface monitor IDs do not support negative values, a cardinal direction is used instead:
  - Latitude: Only positive values are used. “N” is used for locations north of the equator, and “S” is used for locations south of the equator.
  - Longitude: Only positive values are used. “E” is used for locations east of the prime meridian, and “W” is used for locations west of the prime meridian.
- *<DataSourceCode>* is a unique identifier for each surface PM data source.
- *<POC>* is the Parameter Occurrence Code (POC), which identifies distinct instruments measuring the same parameter at a given site. The POC ensures that multiple instruments measuring the same pollutant at a location can be differentiated in the dataset.

Example 1:

```
USA-LosAngeles-0001-33.8306N_117.9385W_1000
```

Refers to a monitor in the USA-LosAngeles target area, located in Anaheim, with coordinates 33.8306º latitude, -117.9385º longitude.

Example 2:

```
ITA-Rome-0001-41.8940N_012.4754E_1200
```

Refers to a monitor in the ITA-Rome target area, located in Arenula, with coordinates 41.8940º latitude, 12.4754º longitude

Definition of other fields under the “Monitor_Information” group is provided below for total PM2.5, but, as mentioned above, these fields are common across all PM types.

Table 3: /PM_2.5/Total/Monitor_Information

Field Name / Description	Dimension List	Data Type	Units
site_name Name of the surface monitor site	pm25_monitor	str	N/A

continues on next page

Table 3 – continued from previous page

Field Name / Description	Dimension List	Data Type	Units
station_country The country in which the station resides. Country names should be harmonized in case different spellings are used in various original databases.	pm25_monitor	str	N/A
station_state_or_province The state or province in which a station resides.	pm25_monitor	str	N/A
elevation Ground elevation of the surface monitor station in meters above sea level.	pm25_monitor	float64	m
height Height of the surface monitor station in meters above ground level	pm25_monitor	float64	m
station_location_type Station characterization or land use type in the area surrounding the station (available categories: urban, suburban, rural, remote/background, traffic, industrial, other, and unknown).	pm25_monitor	str	N/A
source_name Short name of the surface monitor data source or data provider.	pm25_monitor	str	N/A
source_url Base url where surface monitor data was accessed	pm25_monitor	str	N/A
latitude Latitude of surface monitor measurement in decimal degrees. Allowed values are from -90 to +90. Negative latitudes are in the Southern hemisphere.	pm25_monitor	float64	°N
longitude Longitude of surface monitor measurements in decimal degrees. Allowed values are from -180 to +180.	pm25_monitor	float64	°E
maia_surface_monitor_id MAIA identification string of surface monitor location.	pm25_monitor	str	N/A

5.3 Source Input Data

The “Source_Input_Data” group contains information about the raw PM measurements used for quality screening and the calculation of daily averaged values [M-1]. For “PM_2.5/Total”, “PM_10/Total”, and “PM_2.5/Black_Carbon”, raw measurements are obtained from time-resolved continuous monitors. These measurements undergo various quality screening steps (as described in [M-1]) before daily averages are calculated. In contrast, PM2.5 speciation monitors provide data as a single daily average value. In these cases, the “pm_mass_concentration” values in the species group and the “Source_Input_Data” group are identical. For example:

- “PM_2.5/Elemental_Carbon/pm_mass_concentration” = “PM_2.5/Elemental_Carbon/Source_Input_Data/pm_mass_concentratio

This applies specifically to the following PM2.5 species:

- “PM_2.5/Elemental_Carbon”
- “PM_2.5/Nitrate”
- “PM_2.5/Organic_Carbon”
- “PM_2.5/Sulfate”

The “Source_Input_Data” group contains the following fields for “PM_2.5/Total”, “PM_10/Total”, and all PM2.5 species, except “PM_2.5/Dust”.

Table 4: /PM_2.5/Total/Source_Input_Data

Field Name / Description	Dimension List	Data Type	Units
maia_surface_monitor_id MAIA identification string of surface monitor location	pm25_measurement	str	N/A
date_time_UTC Datetime of surface monitor measurement based on UTC time, in format YYYYMMDDThhmmssZ, where Z indicates that the time value is the time in Greenwich, England or UTC time.	pm25_measurement	str	N/A
date_time_LST Datetime of surface monitor measurement based on local solar time (LST) with UTC offset, in format YYYYMMDDThhmmss+|-hhmm, where + indicates the time zone offset east of Greenwich, England and - indicates the timezone offset west of Greenwich, England.	pm25_measurement	str	N/A
pm_mass_concentration ‘Raw’ PM mass concentration	pm25_measurement	float64	µg / m ** 3

The PM2.5 dust mass concentration is calculated as a weighted sum of aluminum, silicon, calcium, iron, and titanium [M-1]. However, some monitors may not report all five metals used in the dust calculation. According to [M-1], any sample missing both aluminum and iron abundances is excluded from data processing. If silicon, calcium, and/or titanium abundances are unavailable but aluminum and iron abundances are present, the concentrations of the missing elements are estimated based on scaling of the aluminum concentrations, as described in [M-1].

The “PM_2.5/Dust/Source_Input_Data” group is subdivided into its elemental subgroups:

- “Aluminum”
- “Calcium”

- "Iron"
- "Silicon"
- "Titanium"

Each of these elemental subgroups includes the fields described in Table 4, along with the "Monitor_Information" fields outlined in Table 3. The following tables present the fields for "Aluminum", which are similar to those used for other dust components.

Table 5: /PM_2.5/Dust/Source_Input_Data/Aluminum

Field Name / Description	Dimension List	Data Type	Units
maia_surface_monitor_id MAIA identification string of surface monitor location	aluminum_measurement	str	N/A
date_time_UTC Datetime of surface monitor measurement based on UTC time, in format YYYYMMDDThhmmssZ, where Z indicates that the time value is the time in Greenwich, England or UTC time.	aluminum_measurement	str	N/A
date_time_LST Datetime of surface monitor measurement based on local solar time (LST) with UTC offset, in format YYYYMMDDThhmmss+|-hhmm, where + indicates the time zone offset east of Greenwich, England and - indicates the timezone offset west of Greenwich, England.	aluminum_measurement	str	N/A
aluminum_pm_mass_concentration Raw aluminum PM2.5 mass concentration	aluminum_measurement	float64	µg / m ** 3

Table 6: /PM_2.5/Dust/Source_Input_Data/Aluminum/Monitor_Information

Field Name / Description	Dimension List	Data Type	Units
site_name Name of the surface monitor site	aluminum_monitor	str	N/A
station_country The country in which the station resides. Country names should be harmonized in case different spellings are used in various original databases.	aluminum_monitor	str	N/A
station_state_or_province The state or province in which a station resides.	aluminum_monitor	str	N/A
elevation Ground elevation of the surface monitor station in meters above sea level.	aluminum_monitor	float64	m

continues on next page

Table 6 – continued from previous page

Field Name / Description	Dimension List	Data Type	Units
height Height of the surface monitor station in meters above ground level	aluminum_monitor	float64	m
station_location_type Station characterization or land use type in the area surrounding the station (available categories: urban, suburban, rural, remote/background, traffic, industrial, other, and unknown).	aluminum_monitor	str	N/A
source_name Short name of the surface monitor data source or data provider.	aluminum_monitor	str	N/A
source_url Base url where surface monitor data was accessed	aluminum_monitor	str	N/A
latitude Latitude of surface monitor measurement in decimal degrees. Allowed values are from -90 to +90. Negative latitudes are in the Southern hemisphere.	aluminum_monitor	float64	°N
longitude Longitude of surface monitor measurements in decimal degrees. Allowed values are from -180 to +180.	aluminum_monitor	float64	°E
maia_surface_monitor_id MAIA identification string of surface monitor location.	aluminum_monitor	str	N/A

CHAPTER

SIX

QUALITY METRICS

This section describes the "Quality_Metrics" fields reported for PM measurements. Separate subsections detail the quality metrics for "PM_TOTAL" and "PM_SPECIES" products.

6.1 Quality Metrics for PM10 and PM2.5 Total Mass Concentration

The quality metrics for Total PM2.5 and PM10 apply to data collected within a 24-hour time window, which is defined as either 00:00 to 24:00 or 09:00 to 09:00 LST, depending on the target area [M-1].

Details of quality metric calculations are provided in [M-1]. The following tables describe the "Quality_Metrics" fields for "PM_2.5/Total," which are identical for "PM_10/Total" as well.

Table 7: /PM_2.5/Total/Quality_Metrics

Field Name / Description	Dimension List	Data Type	Units
completeness Data completeness metric. Fractional value computed from the number of valid data points divided by the number of expected data points.	pm25_monitor	float64	N/A

Table 8: /PM_2.5/Total/Quality_Metrics/Count_Per_Monitor

Field Name / Description	Dimension List	Data Type	Units
total_samples Total number of prescreened samples including those from previous and subsequent days.	pm25_monitor	int64	N/A
actual_samples Total number of prescreened samples on the reporting day.	pm25_monitor	int64	N/A
expected_samples Number of expected samples on the reporting day.	pm25_monitor	int64	N/A
duplicate_samples Number of samples screened due to duplicate time and PM values.	pm25_monitor	int64	N/A

continues on next page

Table 8 – continued from previous page

Field Name / Description	Dimension List	Data Type	Units
duplicate_time Number of unscreened samples with duplicate time stamps but different PM values.	pm25_monitor	int64	N/A
stuck Number of samples screened due to stuck PM values.	pm25_monitor	int64	N/A
negative Number of samples screened due to negative PM values below the established threshold.	pm25_monitor	int64	N/A
anomalous Number of samples screened due to anomalous PM rate of change.	pm25_monitor	int64	N/A
unphysical Number of samples screened due to PM2.5 > PM10.	pm25_monitor	int64	N/A
outlier Number of samples screened due to being labeled as outliers.	pm25_monitor	int64	N/A
valid Number of valid (screened) samples.	pm25_monitor	int64	N/A

6.2 Quality Metrics for PM2.5 Species

This section describes the "Quality_Metrics" fields reported for measurements of PM2.5 species, including sulfate, nitrate, EC, BC, OC and dust.

Unlike total PM data, raw data for most of speciated PM2.5 are provided as daily averages that have already undergone quality control by the data providers (e.g., chemical laboratories) before entering the MAIA science data processing system. However, since the data are merged from multiple measurement networks, an additional quality assessment is performed to flag values that are unusual compared to prior years. Details of these assessments are provided in [M-1].

Some data sources report BC, while others report EC, or both. However, the MAIA GRMs use only EC in their processing. Therefore, all BC measurements are converted to EC before ingestion into the GRM training system. Additional details and descriptions of the conversion algorithms are provided in [M-1].

Unlike other PM2.5 species, some BC measurements are time-resolved at sub-daily scales (e.g., 1-minute or hourly resolution). As a result, the quality metrics applied to time-resolved total PM monitors are also applied to those BC data which are reported at sub-daily temporal resolution. The time averaging and quality screening for BC follow the same procedures as those for "PM_TOTAL" products (i.e., "PM_2.5/Total" and "PM_10/Total"). Specifically, time averaging for BC is performed on data collected within a 24-hour window, defined as either 00:00 to 24:00 or 09:00 to 09:00 LST, with the start time associated with the corresponding calendar date.

The tables below illustrate the two different layouts for the "Quality_Metrics" group within the "PM_2.5_SPECIES" files:

1. One layout for dust, which is identical to the quality metrics for other species, including sulfate, nitrate, EC, and OC (Table 9)
2. One layout for BC (Tables 10 and 11)

Table 9: /PM_2.5/Dust/Quality_Metrics

Field Name / Description	Dimension List	Data Type	Units
mean_concentration_prior_year Mean concentration from same month in prior years.	dust_monitor	int64	μg / m ** 3
standard_deviation_prior_year Concentration standard deviation from the same month in prior years.	dust_monitor	int64	μg / m ** 3
expected_range_flag Flag indicating whether data are within expected range.	dust_monitor	int64	N/A

Table 10: /PM_2.5/Black_Carbon/Quality_Metrics

Field Name / Description	Dimension List	Data Type	Units
completeness Data completeness metric. Fractional value computed from the number of valid data points divided by the number of expected data points.	bc_monitor	float64	N/A

Table 11: /PM_2.5/Black_Carbon/Quality_Metrics/Count_Per_Monitor

Field Name / Description	Dimension List	Data Type	Units
total_samples Total number of prescreened samples including those from previous and subsequent days.	bc_monitor	int64	N/A
actual_samples Total number of prescreened samples on the reporting day.	bc_monitor	int64	N/A
expected_samples Number of expected samples on the reporting day.	bc_monitor	int64	N/A
duplicate_samples Number of samples screened due to duplicate time and PM values.	bc_monitor	int64	N/A
duplicate_time Number of unscreened samples with duplicate time stamps but different PM values.	bc_monitor	int64	N/A
stuck Number of samples screened due to stuck PM values.	bc_monitor	int64	N/A
negative Number of samples screened due to negative PM values below the established threshold.	bc_monitor	int64	N/A
anomalous Number of samples screened due to anomalous PM rate of change.	bc_monitor	int64	N/A
unphysical Number of samples screened due to PM2.5 > PM10.	bc_monitor	int64	N/A
outlier Number of samples screened due to being labeled as outliers.	bc_monitor	int64	N/A

continues on next page

Table 11 – continued from previous page

Field Name / Description	Dimension List	Data Type	Units
valid Number of valid (screened) samples.	bc_monitor	int64	N/A