

Format Description for Daily Meteorological Data Files Used in GLERL Procedures

GLERL's standard programs rely on the station data files conforming to a predefined formatting convention. These specifications must be followed exactly or else the programs will fail to process the data correctly. One of the most significant dangers here is that the programs that process these files have varying levels of error detection and error handling built into them. Therefore, depending upon the nature of the formatting error in the data files, the programs may process the data without recognizing the error in them (thereby producing erroneous results) or may fail with cryptic error messages that do not pinpoint the issue with the input data. To aid in preventing these situations, here is a very detailed description of the data formatting requirements. Please pay special attention to the exact columns for each data item, as the most common error that is made when producing these files is to have data in the wrong columns. The programs that read these files require complete compliance with the column specifications.

The file format descriptions follow on the next few pages.

File type: M???????.DAT

These files contain daily values for maximum air temperature, minimum air temperature and total precipitation. The data units are dependent upon whether the station is a U.S. station (data are in english units) or a station from Canada (data are in metric units). Stations from other parts of the world are not currently directly supported, although we have made use of them by typically assigned a manufactured station ID that starts with a '6' and used metric units for the data. The programs use the first character in the station ID to determine the location of the station (and data units). A '0' (zero) as the first character indicates a U.S. station, while a '6' indicates Canadian station. There are four lines of header information followed by the data lines. Each of the data lines contains the daily values (Tmax, Tmin, Precip) for a single day, in chronological order. Note that EVERY day must be represented by a data line. Missing data values are represented by a value of -999 for that data element. The format is described as follows:

Line 1: Station header information.

Columns	Description	Notes
2-8	Station ID	Alphanumeric (letters and numbers only). First character of 0 indicates data in English units. Anything else indicates data in metric units. A 6 usually means Canadian data.
10-18	Latitude	Latitude in decimal degrees north of the equator (e.g. 42.123); right-justified
20-28	Longitude	Longitude in decimal degrees east of the prime meridian (e.g. a station in Michigan might be -87.340); right-justified
30-80	Station Name	Optional field. Alphanumeric – any characters

Line 2: Station data start date

Columns	Description	Notes
1-4	“From”	Optional indicator to reader that this is the start date
6-9	Year	Year of the first data line
11-12	Month	Month of the first data line; right-justified
14-15	Day	Day of the first data line; right-justified

Line 3: Station data end date

Columns	Description	Notes
1-4	“To”	Optional indicator to reader that this is the end date
6-9	Year	Year of the last data line
11-12	Month	Month of the last data line; right-justified
14-15	Day	Day of the last data line; right-justified

Line 4: Count of data lines

Columns	Description	Notes
4-9	Number of lines	Count of the number of days in the file. Must match both the number of days indicated by the header as well as the number of data lines that follow.

Line 5-end: Station data

Columns	Description	Notes
1-4	Max air temperature	Right-justified integer value; Degrees Fahrenheit for U.S. stations Tenths of degrees Celsius for Canadian stations
5-8	Min air temperature	Right-justified integer value; Degrees Fahrenheit for U.S. stations Tenths of degrees Celsius for Canadian stations
9-12	Precipitation	Right-justified integer value; Hundredths of inches for U.S. stations Tenths of millimeters for Canadian stations
13-??	Optional comments	This is not used by the programs at all, but we often use it to insert helpful comments for human readability. Commonly we will put the year and month on the first day of each month to serve as a “waypoint” marker for a reader.

File type: E????????.DAT

These files contain daily values for drybulb air temperature, dewpoint air temperature, windspeed and cloud cover. The data units are dependent upon whether the station is a U.S. station (data are in english units) or a station from Canada (data are in metric units). Stations from other parts of the world are not currently directly supported, although we have made use of them by typically assigned a manufactured station ID that starts with a '6' and used metric units for the data. The programs use the first character in the station ID to determine the location of the station (and data units). A '0' (zero) as the first character indicates a U.S. station, while a '6' indicates Canadian station. There are four lines of header information followed by the data lines. Each of the data lines contains the daily values (Air, Dew, Wind, Cloud) for a single day, in chronological order. Note that EVERY day must be represented by a data line. Missing data values are represented by a value of -999 for that data element. The format is described as follows:

Line 1: Station header information.

Columns	Description	Notes
2-8	Station ID	Alphanumeric (letters and numbers only). First character of 0 indicates U.S. data in English units. First character of 6 indicates Canadian data in metric units.
10-18	Latitude	Latitude in decimal degrees north of the equator (e.g. 42.123); right-justified
20-28	Longitude	Longitude in decimal degrees east of the prime meridian (e.g. a station in Michigan might be -87.340); right-justified
30-80	Station Name	Optional field. Alphanumeric – any characters

Line 2: Station data start date

Columns	Description	Notes
1-4	“From”	Optional indicator to reader that this is the start date
6-9	Year	Year of the first data line
11-12	Month	Month of the first data line; right-justified
14-15	Day	Day of the first data line; right-justified

Line 3: Station data end date

Columns	Description	Notes
1-4	“To”	Optional indicator to reader that this is the end date
6-9	Year	Year of the last data line
11-12	Month	Month of the last data line; right-justified
14-15	Day	Day of the last data line; right-justified

Line 4: Count of data lines

Columns	Description	Notes
4-9	Number of lines	Count of the number of days in the file. Must match both the number of days indicated by the header as well as the number of data lines that follow.

Line 5-end: Station data

Columns	Description	Notes
1-4	Drybulb air temperature	Right-justified integer value; Degrees Fahrenheit for U.S. stations Tenths of degrees Celsius for Canadian stations
5-8	Dewpoint temperature	Right-justified integer value; Degrees Fahrenheit for U.S. stations Tenths of degrees Celsius for Canadian stations
9-12	Windspeed	Right-justified integer value; Miles per hour for U.S. stations Meters per second for Canadian stations
13-16	Cloud cover	Right-justified integer value; Tenths; e.g. 50% cloud cover is shown by a 5 in this field.
17-??	Optional comments	This is not used by the programs at all, but we often use it to insert helpful comments for human readability. Commonly we will put the year and month on the first day of each month.

File type: MET_<station id>.TXT

These files contain daily values for the various meteorological variables. There will be a column header for each data type to denote the contents of the file. This is a new format that was designed in 2013 in order to address issues with the previous M*.DAT and E*.DAT formats. We were running into two major issues with those files:

- 1) The station ID was fixed at 7 characters, but station IDs are now often longer than that.
- 2) Units of the data (metric vs english) were implied by the station ID. This is also not a workable situation with the varied station IDs we now get.

Additionally, fixed-width formatting has become a rarity these days, replaced by comma-delimited files as a common paradigm. By moving to comma-delimited formatting, we can probably avoid most future issues with field lengths. It will also make our files more “user-friendly” to other software such as Excel or R. With this format, the software which reads the data will need to read the header lines (5 & 6) to determine what data is in the file.

Line 1: Station header information.

Field #	Description	Notes
1	Station ID	Alphanumeric (letters and numbers only, no imbedded spaces).
2	Station Name	Optional field. Alphanumeric – any characters except comma

Line 2: Station location information.

Field #	Description	Notes
1	“Lat & Long”	Indicator that this is latitude and longitude
2	Latitude	Latitude in decimal degrees north of the equator (e.g. 42.123)
3	Longitude	Longitude in decimal degrees east of the prime meridian (e.g. a station in Michigan might be -87.340)

Line 3: Station data start date

Field #	Description	Notes
1	“Starts (YMD):”	Indicator to reader that this is the start date
2	Year	Year of the first data line
3	Month	Month of the first data line
4	Day	Day of the first data line

Line 4: Station data end date

Field #	Description	Notes
1	“Ends (YMD):”	Indicator to reader that this is the end date
2	Year	Year of the first data line
3	Month	Month of the first data line
4	Day	Day of the first data line

Line 5: Column Header 1

Field#	Description	Notes
1	Blank	Blank; used in header line 2
2-n	Data Type	Type of data to be found in this column; Valid entries are: AIRTEMPMAX = maximum air temperature for the day AIRTEMPMIN = minimum air temperature for the day AIRTEMP = mean air temperature for the day DEWPOINT = mean dew point temp for the day WINDSPEED = mean wind speed for the day CLOUD = mean cloud cover for the day PRECIP = total precipitation for the day

Line 6: Column Header 2

Field#	Description	Notes
1	“YYYYMMDD”	Explains column 1 is date in YYYYMMDD format
2-n	Data Units	Units for the data found in this column; Valid entries are: DEGC = degrees Celsius DEGF = degrees Fahrenheit INCH = inches CM = centimeters MM = millimeters M/S = meters per second % = percent FRACTION = decimal fraction; 0.0 to 1.0

Lines 7-end: Station data

Field #	Description	Notes
1	YearMonthDay	Date of the observation (YYYYMMDD)
2-n	Data values	Data values corresponding to this date and placed in the correct column to match header lines. Missing data values are denoted by any of the following: Blank entry, -9.9e9, “N/A”