

Standard Time Series Data File Format

Aron P. Dobos

August 20, 2013

Abstract

This document describes in detail the standard comma-separated value (CSV) file format adopted by the System Advisor Model (SAM) and other NREL systems modeling tools to store hourly and sub-hourly time series solar and meteorological data. The format is designed to be flexible, extensible, easily editable with common desktop software like Microsoft Excel, and minimalist in size and overhead. The existing site-specific datasets provided by NREL such as TMY2, TMY3, Energy Plus, and PVWatts International have all been converted to this new standard CSV file format.

1 Overview

The flexible comma-separated value (CSV) text data file format documented here purports the following advantages over common fixed weather data file formats (TMY2, TMY3, EPW) that are currently in use.

- Directly editable by Microsoft Excel despite its quirks when reading/writing CSV files
- Does not require a fixed structure or ordering of header and data columns
- Allows additional data columns to be added as needed and as data is available
- Recognizes many standard header and column names, with and without units
- Provides options for interpolating data columns as appropriate for properly interpreting data
- Independent of data creation or amalgamation method: i.e. often TMY2 and TMY3 data file formats are used to store non-TMY data.
- Support for subhourly or multi-year datasets
- Allows a large data file such as a TMY3 file with 62 data columns to be stripped down to the 4 or 5 data columns that are required for a particular simulation - saves data storage space

The general data file must be an ASCII or UTF-8 text file that consists of:

1. One line of comma-separated header field names

2. One line of comma-separated header field values
3. One line of comma-separated data column names
4. Optional line of comma-separated unit names associated with the data column
5. An integer multiple of 8760 data lines each having comma-separated values associated with the data column

A basic data file might look like:

```
Source,Location ID,City,State,Latitude,Longitude,Time Zone,Elevation
TMY2,13962,ABILENE,TX,32.433333,-99.683333,-6,534
Year,Month,Day,Hour,GHI,DNI,DHI,Tdry,Tdew,RH,Pres,Wspd,Wdir,Snow Depth
1974,1,1,0,0,0,0,-8.3,-11.7,74,962,3.4,60,0
1974,1,1,1,0,0,0,-8.9,-11.8,77,962,3.7,60,0
1974,1,1,2,0,0,0,-9.4,-12.2,80,962,4.1,50,0
1974,1,1,3,0,0,0,-9.8,-12.3,80,962,4.5,40,0
1974,1,1,4,0,0,0,-10.2,-12.7,80,962,4.8,40,0
...
...
...
```

2 Headers

The file format recognizes several standard header sections. The order of header fields in the data file is unimportant, and the header field names are not case-sensitive. Additional header fields that are not one of the “standard” ones can be added to provide additional documentation or data in the file. For a PV system simulation, commonly required header fields include latitude, longitude, time zone, and site elevation. The other fields may be included as available and/or necessary as metadata.

Field	Units	Recognized Names
Latitude	degrees	<i>latitude,lat</i>
Longitude	degrees	<i>longitude,lon,long,lng</i>
Time zone	hours offset from GMT	<i>tz,timezone,time zone</i>
Site elevation	meters above sea level	<i>el,elev,elevation,site elevation</i>
Year	n/a	<i>year</i>
Location ID	n/a	<i>id,location,location id,station,station id,wban,wban#</i>
City	n/a	<i>city</i>
State	n/a	<i>state,province,region</i>
Country	n/a	<i>country</i>
Source	n/a	<i>source,src</i>
Description	n/a	<i>description,desc</i>
URL	n/a	<i>url</i>
Units flag	yes or no	<i>hasunits,units</i>
Interpolate flag	yes or no	<i>interpmet</i>

Table 1: Recognized header field names

3 Data Columns

The time series data section allows any number of data columns to be included in any order. Data columns recognized are listed below. Additional data columns may be added for specific uses.

Field	Units	Recognized names
Year	1950-2050	<i>year, yr</i>
Month	1-12	<i>month, mo</i>
Day	1-31	<i>day</i>
Hour	0-23	<i>hour, hr</i>
Minute	0-59	<i>min, minute</i>
Global horizontal irradiance	W/m2	<i>gh, ghi, global, global horizontal, global horizontal irradiance</i>
Beam normal irradiance	W/m2	<i>dn, dni, beam, direct normal, direct normal irradiance</i>
Diffuse horizontal irradiance	W/m2	<i>df, dhi, diffuse, diffuse horizontal, diffuse horizontal irradiance</i>
Ambient dry bulb temperature	C	<i>tdry, dry bulb, dry bulb temp, temperature, ambient, ambient temp</i>
Wet bulb temperature	C	<i>twet, wet bulb, wet bulb temperature</i>
Dew point temperature	C	<i>tdew, dew point, dew point temperature</i>
Wind speed	m/s	<i>wspd, wind speed</i>
Wind direction	deg	<i>wdir, wind direction</i>
Relative humidity	%	<i>rh, rhum, relative humidity, humidity</i>
Atmospheric pressure	millibar	<i>pres, pressure</i>
Snow cover	cm	<i>snow, snow cover, snow depth</i>
Ground reflectance (albedo)	0..1	<i>albedo, alb</i>
Aerosol optical depth	0..1	<i>aod, aerosol, aerosol optical depth</i>

Table 2: Hourly input data fields

4 Time Steps

By default, the weather file is assumed to contain an integer multiple of 8760 records from which the timestep is estimated by Eqn. 1.

$$TS_{min} = \frac{60 \text{ min}}{\text{int}(N/8760)} \tag{1}$$

If the weather file does not contain an integer number of records, the `hour` and `minute` data columns must exist, so that the data reader can determine the time step from the time difference between the first two records in the data file.

5 Interpolation

If the header field `InterpMet` exists and has a value of `yes`, the meteorological data that exists in the data file will be interpolated. The data columns interpolated are wind speed, wind direction, dry bulb temperature, wet bulb temperature, dew point temperature, relative humidity, pressure, snow cover, and albedo.

The interpolation method is as follows. For the first data record, a backwards interpolation of the first two records is done, per Eqn. 2. For subsequent data records, the average value of the current timestep and the previous one is used, per Eqn. 3.

$$D_0 = D_1 + 1.5 \cdot (D_0 - D_1) \tag{2}$$

$$D_i = 0.5 \cdot (D_i + D_{i-1}) \quad \text{for } i \geq 1 \tag{3}$$

6 Value Estimation

The weather file data reader implemented in the SAM Simulation Core (SSC) engine is able to estimate the value of certain data fields that do not appear in the file given the values of other fields. The data fields with estimation capability are listed in Table 3.

Estimated Field	Method	Requires
Wet bulb temperature	Martinez [6]	<i>Tdry, Rhum, Pres</i>
Dew point temperature	Magnus formulation [5]	<i>Tdry, Rhum</i>
Year	n/a	<i>data in header</i>
Month	Estimate from data index	<i>8760 records</i>
Day	Estimate from data index	<i>8760 records</i>
Hour	Estimate from data index	<i>8760 records</i>

Table 3: Estimatable data fields

References

- [1] Marion, W., Urban, K., *User's Manual for TMY2s*, 2012. <http://rredc.nrel.gov/solar/pubs/tmy2/titlepg.html>
- [2] Wilcox, S., Marion, W., *Users Manual for TMY3 Data Sets*, NREL Technical Report NREL/TP-581-43156, 2008.
- [3] Crawley, D., et al. *Energy Plus Weather File Documentation*, 1999, <http://apps1.eere.energy.gov/buildings/energyplus/pdfs/weatherdatainformation.pdf>
- [4] System Advisor Model (SAM). National Renewable Energy Laboratory, 2013. <http://sam.nrel.gov>
- [5] Dew point. *Wikipedia*. http://en.wikipedia.org/wiki/Dew_point, August 2013.
- [6] Martinez, A.; *On the evaluation of the wet bulb temperature as a function of dry bulb temperature and relative humidity*. Atmosfera, vol.7, 1994. accessed: <http://www.ejournal.unam.mx/atm/Vol07-3/ATM07304.pdf>