Create Full Year EnergyPlus Weather Files with Real Historical Weather Data

Version 0.6 BETA

Brian Woo-Shem

Last updated: 2021-08-11

Get historical weather data for any recent year for many locations around the U.S. and world, process it into hourly averages when data is more frequently reported, and paste into the Energy Plus weather file for simulations.

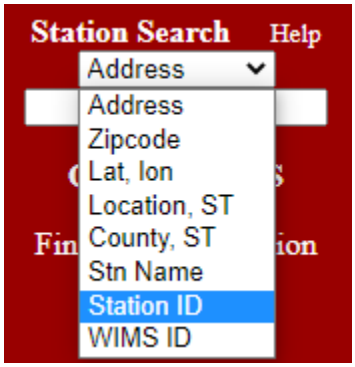
This process works but is only 50% automated. Weather for 1 year at 1 location takes 30 minutes to 1 hour to set up.

1. Get Source Weather Data

Start with an epw weather file from:
http://climate.onebuilding.org/WMO_Region_4_North_and_Central_America/USA_United_States_of_America/index.html#IDCA_California-

Go to MesoWest: https://mesowest.utah.edu/

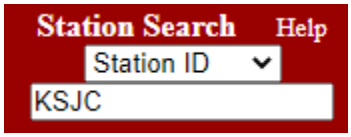
On MesoWest, change search type to station id



Search the airport name, matching the one the EPW is from.

Note: If needed, find the airport codename (3 letters) from here:
https://www.airportcodes.us/us-airports-by-name.htm

Then add a “K” in front. For example, San Jose International Airport is codenamed “SJC”, so we would search using “KSJC”

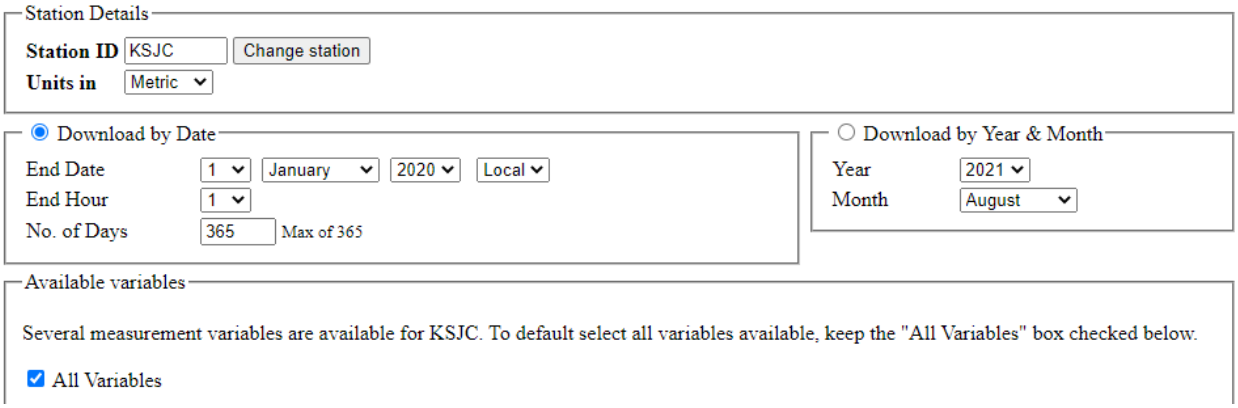


On left panel, look for Download Data

To download an entire year of data, you need to make a free account

Use the following settings:

1. DO NOT change Station ID
2. Switch Units to Metric
3. The interface is clunky and asks for END date. Data is on hours 0-23.
4. To download an entire year at once
    a. End Date = the first day of January for the year AFTER the one you want. (ex: to get the entirety of 2019, set End Date = 1 January 2020.)
    b. End hour = 1.
    c. Number of days = 365
5. Make sure Download by Date is selected (should be by default)
6. Keep “All Variables” selected



Scroll to the bottom and Retrieve Data. Beware it can take several minutes.

However, the airports don't have Solar Radiation. Use the bottom left menu to select Nearby Stations.

Look for a station that is of type **APRSWXNET/CWOP**

List of stations near KSJC that have reported data at least once in the last 90 days.

Search for stations in a 25 mi Radius Change Radius

Location QC Flag: Ok, Caution, Suspect

Station	Info	Dist/Dir
E6095 **EW6095 San Jose**	APRSWXNET/CWOP 55 ft	1 mi E
SJCC1 **SAN JOSE**	CNRFC 67 ft	1 mi ESE
E6873 **EW6873 San Jose**	APRSWXNET/CWOP 82 ft	2 mi SE
A4544 **San Jose - Jackson St.**	AIRNOW 62 ft	2 mi ESE

In this case, either the first or third entries will work. Pick the nearest one

On the page with tons of data, look to see how often the station is reporting, what variables it has, and if it is up to date. A good station reports at least every 15 minutes reliably, and has Solar Radiation values.

This is an example of a good station to pick: 5 minute timesteps, and within 1 mile

Weather Conditions for E6095

Current Time: 08/06/2021 13:31 PDT

Most Recent Weather Conditions at: 08/06/2021 13:25 PDT

Graphical Links	13:25	Max Since 0:00 (PDT)	Min Since 0:00 (PDT)	24 Hour Maximum	24 Hour Minimum
Temperature	83.0° F	83.0 at 13:25	57.0 at 7:05	83.0 at 13:25	57.0 at 7:05
Dew Point	60.7° F	61.5 at 11:25	53.2 at 6:00	61.5 at 11:25	53.2 at 6:00
Wet bulb temperature	68.1° F	68.4 at 13:20	54.8 at 6:00	68.4 at 13:20	54.8 at 6:00
Relative Humidity	47%	89 at 7:55	47 at 13:25	89 at 7:55	47 at 13:25
Wind Speed	1.0 mph	3.0 at 11:55	0.0 at 11:20	5.0 at 15:10	0.0 at 11:20
Wind Gust	8.0 mph	13.0 at 12:45	0.0 at 10:00	22.0 at 17:00	0.0 at 10:00
Wind Direction	N	-	-	-	-
Pressure	29.95 in	29.99 at 0:25	29.95 at 13:25	30.01 at 22:55	29.95 at 13:25
Altimeter	30.01 in	30.05 at 0:25	30.01 at 13:25	30.07 at 22:55	30.01 at 13:25
Solar Radiation	810.0 W/m*m	1225.0 at 12:35	0.0 at 6:15	1270.0 at 15:35	0.0 at 6:15

***Note:** Observations above in yellow indicate that they are older than the last row of observations below.

Tabular Listing of 299 Observations from 08/05/2021 12:35 PDT to 08/06/2021 13:25 PDT:

Time (PDT)	Temperature °F	Dew Point °F	Wet bulb temperature °F	Relative Humidity %	Wind Speed mph	Wind Gust mph	Wind Direction	Pressure in	Altimeter in	Solar Radiation W/m*m	Precipitation 24hr in	Precipitation since local midnight in	Quality Control
13:25	83.0	60.7	68.1	47	1.0	8.0	N	29.95	30.01	810.0	0.00	0.00	OK
13:20	83.0	61.3	68.4	48	1.0	7.0	N	29.95	30.01	811.0	0.00	0.00	OK
13:15	82.0	60.3	67.6	48	2.0	6.0	N	29.95	30.01	818.0	0.00	0.00	OK
13:10	82.0	60.3	67.6	48	2.0	11.0	N	29.96	30.02	823.0	0.00	0.00	OK
13:05	82.0	60.3	67.6	48	2.0	10.0	N	29.96	30.02	810.0	0.00	0.00	OK
13:00	81.0	61.1	67.7	51	1.0	6.0	N	29.96	30.02	801.0	0.00	0.00	OK
12:55	81.0	61.1	67.7	51	2.0	8.0	N	29.96	30.02	826.0	0.00	0.00	OK
12:50	81.0	61.1	67.7	51	2.0	8.0	N	29.96	30.02	881.0	0.00	0.00	OK
12:45	80.0	60.8	67.2	52	2.0	13.0	WSW	29.96	30.02	869.0	0.00	0.00	OK
12:40	80.0	60.2	66.9	51	1.0	10.0	W	29.97	30.03	1040.0	0.00	0.00	OK
12:35	80.0	60.2	66.9	51	2.0	8.0	W	29.97	30.03	1225.0	0.00	0.00	OK

If the nearest one lacks good data, try the next nearest, and repeat until you find one that has decent data.
Repeat the download process for the same year of data from this station.

Note: Leap years are poorly supported. EPW does not have leap day, but downloading 365 days of a leap year from Mesowest will include leap day and skip Jan 1. To do a leap year, download an extra 1-2 days around Jan 1 separately, manually append it to the main input csv with the rest of the year, and manually delete Feb 29.

2. Process the CSV data from MesoWest

Use [mesoweatherepw.py]

Change the variables in the top section before running the program.

1. Sourcefile must be the name of the input csv
2. Outputfile is the file it will write to
3. Station tells it what type of station. To process the airports (ie KSJC), use 'NWS'. For the stations for solar radiation, use 'APR'
4. Start year, month, day, and hour must match the first data row on the input csv
5. Change numdays if you want anything other than 1 year of data

```
# PARAMETER VARIABLES TO CHANGE ------------------------------------------
# CSV input file
sourcefile = "KSJC_2019_Metric.csv"

outputfile = "SJ_KSJC_2019_Processed_1.csv"

station = 'NWS'

# List of columns from source to look at. Col A is indexed as 0, B = 1, ..., Z = 25
# For NWS/FAA Type Stations (eg KSJC, KSFO)
if station == 'NWS':
    getcols = [3,5,6,30,37]
    # Headers for first row. Leave first four entries as-is!!!
    # First output row will be headers, all others are data
    headers = ['Year','Month','Day','Hour',"Drybulb","RH","Windspeed","Dewpoint","Pressure"]

# For APRSWXNET/Citizen Weather Observer Program stations (eg https://mesowest.utah.edu/cgi-bin/droman/meso_base_dyn.cgi?stn=E6095)
elif station == 'APR':
    getcols = [3,4,5,9,12,16]
    headers = ['Year','Month','Day','Hour',"Drybulb","RH","Windspeed","SolarRad","Dewpoint","Pressure"]
# For custom
elif station == 'custom':
    getcols = [3,5,6,11,13]
    headers = ['Year','Month','Day','Hour',"Drybulb","RH","Windspeed","Dewpoint","Pressure"]


# When to start - must match input data
startyear = 2019
startmonth = 1
startday = 1
starthour = 0

# length of output = header + 365 days * 24 hours per day
numdays = 365 # How many days of data; often want 365
tsperday = 24 #Timesteps per day; usually want 24
```

Run the program in Python. It may take several minutes.

Warning: Data sets often have missing data or anomalies that will cause the program to crash. These must be resolved manually by editing the csv to something correct. Of 4 csvs, 3 had fatal errors initially.

- If input csv has a missing hour, the code can correct by filling in with the closest available data.
- If multiple consecutive hours are missing, the result will be inaccurate
- If input has duplicated time data, it will not crash and instead skip it; ~~you need to manually delete some of it so the times are correct.~~ This error usually indicates that an hour is duplicated. Read the input above to figure out where in the file
-
```
Traceback (most recent call last):
  File "weathercsv1.py", line 219, in <module>
    dat[s,0:4] = 0
IndexError: index 24 is out of bounds for axis 0 with size 24
```
- If input is too short, it will give a warning. If it only gives the warning for the last hour of the year, this is expected and okay.

3. Merge into EPW

Download the free open-source Elements program from
https://bigladdersoftware.com/projects/elements/
For help with Elements, see:
https://bigladdersoftware.com/projects/elements/docs/user-guide/document-window.html

Open the processed csv files for the airport and the APRSWXNET/CWOP stations.

Airport:
Next to Pressure create a new column. Call this Pressure[kPa] and set it to divide the pressure in Pa by 1000. Ctrl+D to do this for the entire length of the pressure values.

	Pressure	Pressure[kPa]
2	101670.368571429	=I2/1000
3	101693.019230769	101.693019230769
5	101778.81	101.77881
7	101864.585384615	101.864585384615
	101926.97	101.92697
6	101981.560769231	101.981560769231

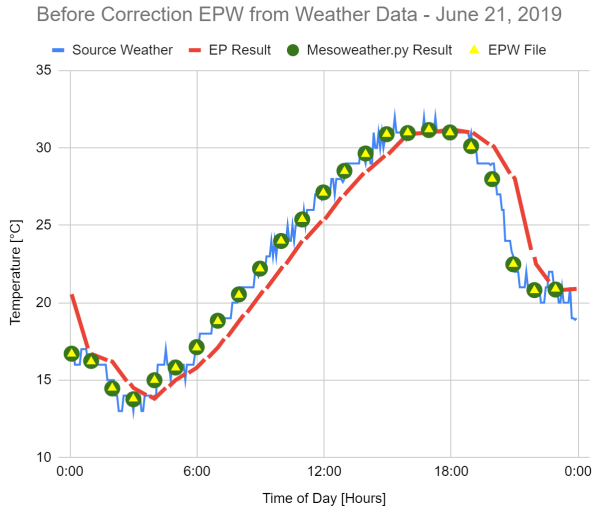
Open Elements

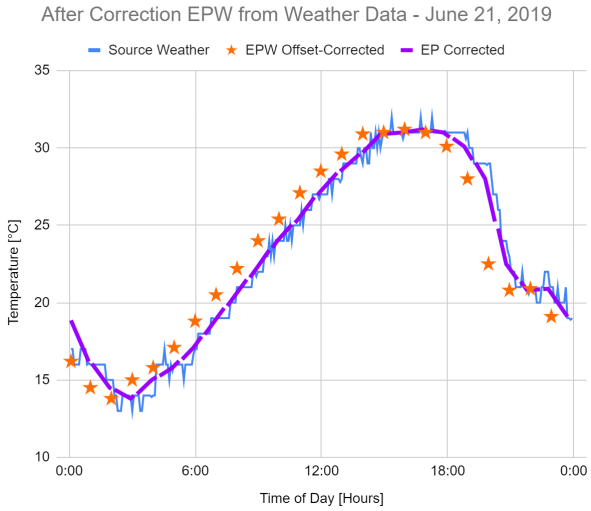
Copy paste the following columns from the processed csv to the Elements window by selecting the cells for one column that contain data, NOT the header and NOT any blank rows at the bottom.

- Drybulb
- Relative Humidity (RH)
- Wind Speed
- Dew Point Temp
- Pressure [kPa]

APRSWXNET/CWOP Station: copy the Solar Radiation to Normal Solar in Elements.

Update on weather and timesteps: EnergyPlus has an offset issue such that if you take the Mesoweather output, copy it to the EPW matching the correct hours, and use this EPW, the EnergyPlus output will be one hour delayed (see first graph). However, the Mesoweather and EPW will match the original dataset.





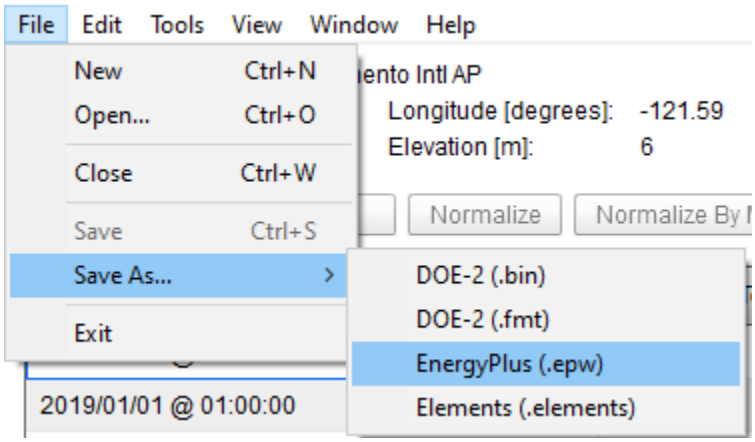
To fix this, skip the 0th hour on the first day of data in Mesoweather output when copying, and instead paste hour 1 into 00:00:00 in Elements. This makes the EPW data appear to be offset relative to the source data, but gives the correct result in the simulation.

Date/Time	Dry Bulb Temperature [C]
2019/01/01 @ 00:00:00	1.6
2019/01/01 @ 01:00:00	1.4
2019/01/01 @ 02:00:00	2.6
2019/01/01 @ 03:00:00	3
2019/01/01 @ 04:00:00	2.2
2019/01/01 @ 05:00:00	3
2019/01/01 @ 06:00:00	3.4

	1	Year	Month	Day	Hour	Drybulb
2	2019	1	1	0	2.86
3	2019	1	1	1	1.59
4	2019	1	1	2	1.39
5	2019	1	1	3	2.55
6	2019	1	1	4	2.98
7	2019	1	1	5	2.25
8	2019	1	1	6	3.02
9	2019	1	1	7	3.41

Using the “Header” button on top right and pop up window, change header to reflect new dataset. I often include the following:
Comments (line 1) : year used, “Data from MesoWest open weather database” and list the station codenames and which data I got from which station
Comments (Line 2) add in front, replacing any download site credits: “Developed by Santa Clara University School of Engineering, Dept. of Mechanical Engineering.”
Reference Year = year used
Data Source = MesoWest/SCU

When finished, Save As an .epw file



Test by running a local simulation such as GetWeatherSolar before doing a UCEF one in case there are any bugs.