Typical Meteorological Year generation service methodology report

Authors: Claire Thomas*, Etienne Wey*, Philippe Blanc**

V1.0 - 13 October 2017

*** Transvalor S.A.**
**** Mines ParisTech**

Outline

Glossary

- PV : Photo Voltaic
- CPV : Concentrated Photo Voltaic
- CSP : Concentrated Solar Power
- GHI : Global Horizontal Irradiation
- GTI : Global Tilted Irradiation (in-plane irradiation for a fixed PV panel with a given tilt and azimuth)
- BNI (or BTI) : Beam Normal Irradiation (direct irradiation in a plane tracking the sun position)
- BSRN : Baseline Surface Radiation Network (BSRN)
- TMY : Typical Meteorological Year
- P50 : 50% percentile of a distribution (median)
- FS50 : 50% median of a distribution based on the Finkelstein-Schafer distance between two Cumulative Distribution Functions (CDF)
- P90 : 90% percentile of a distribution (90% of the values will be above this threshold)
- LT: long-term

FS50 (P50) TMY creation methodology

Introduction

A Typical Meteorological Year is one year of meteorological parameter(s), which is representative of a given situation (median Percentile 50, pessimistic e.g. Percentile 75 or optimistic e.g. Percentile 10).

For computing a typical yearly production for a solar installation, customer needs a Typical Meteorological Year (TMY) P50 (median scenario).

The method for the TMY (P50) generation exploits the HelioClim-3v5 for the radiation data, and the MERRA-2 reanalysis from NASA for the other meteorological parameters. This method has been developed and validated within the framework of the European research project named ENDORSE (funded by the Seventh Framework Programme (FP7) of the EU, number of agreement n°262892). A full description of the method is available at http://www.endorse-fp7.eu/pre-market-services/tmy-generation/service.

There are no missing data in a TMY, thus a gap-filling method is required.

Why using a TMY instead of using the long-term (LT) dataset of meteorological parameters?

- “Widely-used” simulation software in solar energy (e.g. PVSYST, System Advisor Model) require one year of meteorological data <=> at the same time, industrials generally use TMYs for their (pre)feasibility studies of solar energy project
- It can be useful for comparing two different locations (even if the temporal coverage of the data are different)
- For “not-so-long LT”, TMY’s analysis increases the combinatorial possibilities (monthly data blocks) 10 years LT => ~ 62 billion (12^10) possible years of data
- Gain of time when you have many sites to proceed.

A TMY is then a synthetic year of data built from a selection of successive temporal segments (in general months) picked from a multiple years set of hourly data as summarized in Fig 1.

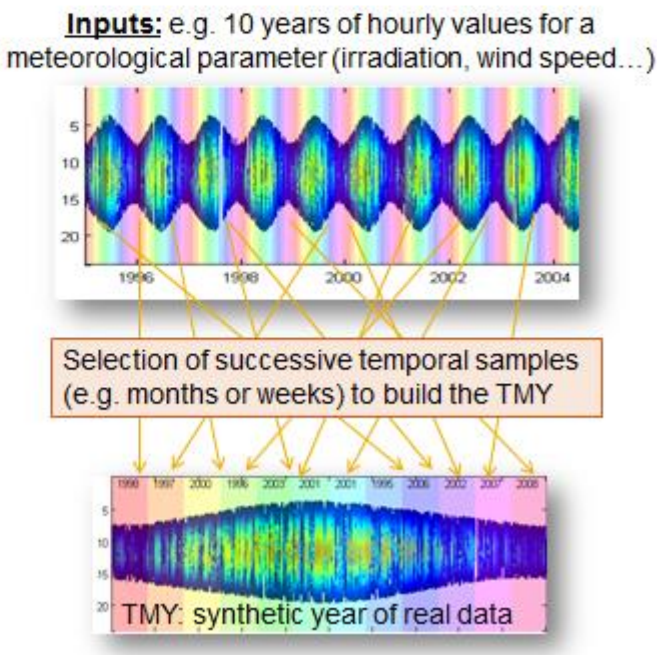


Figure 1: from a long-term time series (at least 10 years) to a one representative year: the TMY

State of the art

The starting point of the service is the NREL (National Renewable Energy Laboratory) method, published in Hall et al. (1978) and Kalogirou (2003).

Here are the specifications of this TMY method:

- Long term series of hourly values for all meteorological data are used
- Data-blocks of monthly meteorological datasets are the basis blocks of the TMY
- TMY is based on the linear weighted combination of the Filkenstein-Schafer distance, where weights are chosen depending on the application.

The model defines that the selected month is the one, which minimizes the following expression for all the same months in the long-term series:

$$5\ d_{FS}(GHI) + 5\ d_{FS}(BNI) + 4\ d_{FS}(T) + 1\ d_{FS}(WS) + 4\ d_{FS}(DWT)$$

Where:

- GHI: Global Horizontal Irradiation
- BNI: Beam Normal Irradiation
- T: Temperature, WS: Wind Speed
- DWT: relative humidity
- $d_{FS}(XX)$: Filkenstein-Schafer distance for variable XX between the given month and the average for all the same months in the series

In the rest of this document, this expression, which is a function of different meteorological parameters is named **"driver"**.

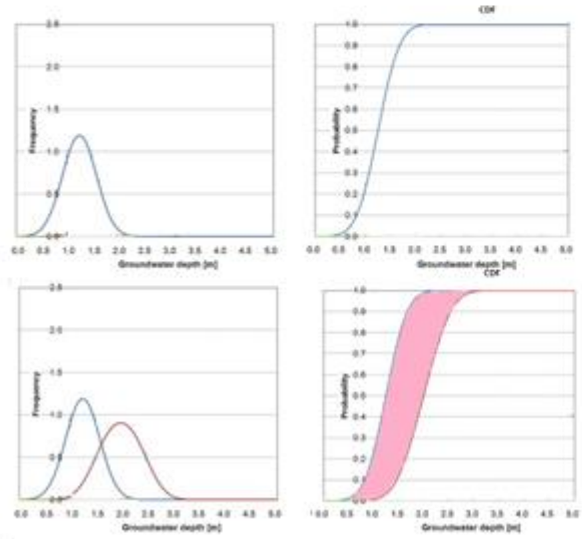


Figure 2: Top left: example of histogram. Top right: CDF (Cumulative Distribution Function) derived from the histogram. Bottom left: two histograms. Bottom right: the Filkenstein-Schafer distance is defined as the space between the two CDFs derived from the two histograms (d_{FS}), represented in pink

NB: Please note that the distance is based on the histogram. This implies to select a bin width to generate the histogram and thus the CDF, and a different bin width will provide different histograms and thus different TMYs.

Innovation coming out from the ENDORSE-fp7 R&D

The main innovation of the new generation method is to take into account the characteristics of the conversion system under concern for the selection of the representative months (or other block size), instead of a linear combination of the different meteorological parameters.

- Tilt of PV panels or solar technology
- Type of Solar Tracking (single axis, dual axis,…) and mechanical limits of the tracking system
- Security issues like safety position of a tracking system when the wind exceeds a given threshold

This resulted in the creation of a **"driver"**, as introduced in the previous section: the **driver** is a one-dimensional composite time series from the multidimensional meteorological long-term dataset more related (or more linearly correlated) to the energy production of the solar energy conversion system of interest (PV, CPV, CSP, etc...). As seen in Figure 3 the type of system has an extremely strong influence on the amount of radiation (energy) which is collected throughout the year.

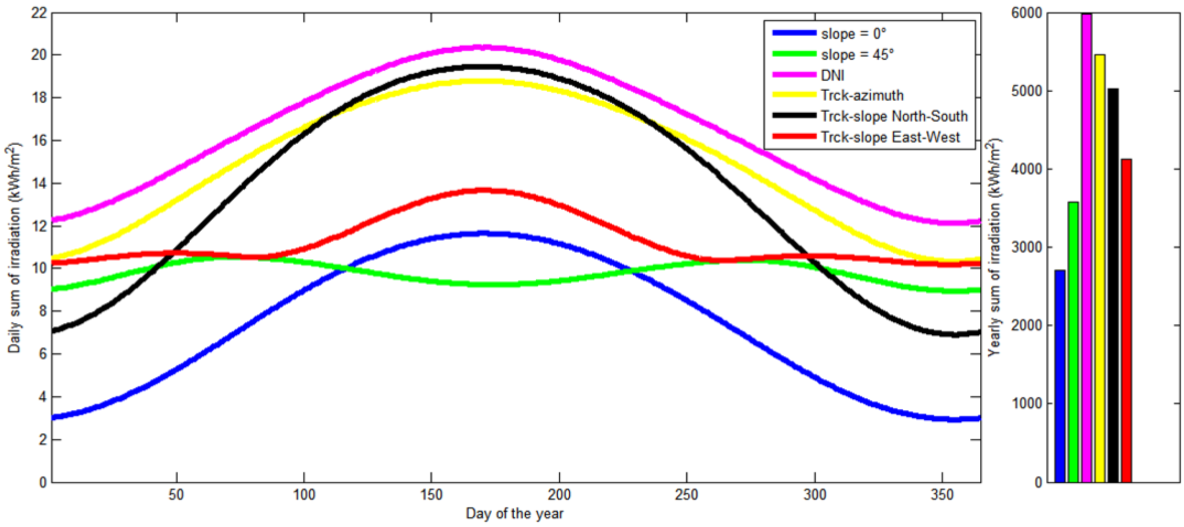


Figure 3: Influence of the types of system or plane orientations on the radiation received by the plane

To give an example of a driver, let us take the example of a one-axis sun tracking for a CSP parabolic trough system. The effective part of the BNI for this system is the cosine of the incident angle.

It exists power output thresholds related to:

- Min irradiance (min start-up temperature)
- Max irradiance (defocusing)
- Max wind speed (tracking security)

In that situation, we derive the driver from the hourly BNI multiplied with the cosine of the incident angle, and the values would be equal to zero (or set to lower values) when the temperature, irradiance and wind speed reach their respective thresholds.

For a simple fixed tilted PV system, the driver is simply the Global Tilted Irradiation (GTI), which is the amount of energy effectively collected by the system.

Validation of the "driver" approach

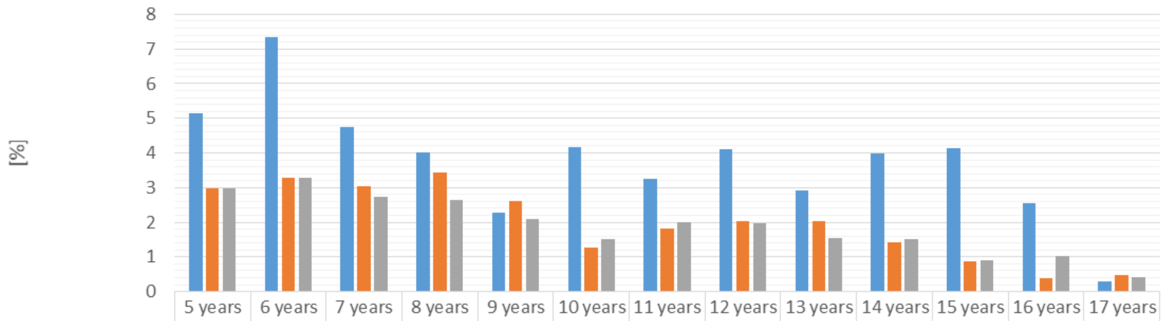
The test site is Carpentras, and the ground station measurements belong to the Baseline Surface Radiation Network (BSRN).

- Study for NEOEN done by Ms. Ana Maria Realpe (SOLAIS) in collaboration with ARMINES

- CPV system: SOITEC System
  (Module: CX-M500, Structure: CX-S540, Tracker: CX-T040, Exosun)
- Site: BSRN CARPENTRAS dataset from 1997 to 2013 (17 years of hourly data)
- Three tested TMY generation approaches:
  - The Sandia method
  - The TMY with the driver = BNI (simple driver)
  - The TMY with the driver = CPV yield production simulation (filtered driver, taking into account the effects of Air Mass, modules temperature increase from irradiation, ambient temperature)

The CPV yield production obtained from the three tested TMY generations is then compared against the effective yield production computed from the long-term time series data for different sliding time windows between 5 to 17 years.

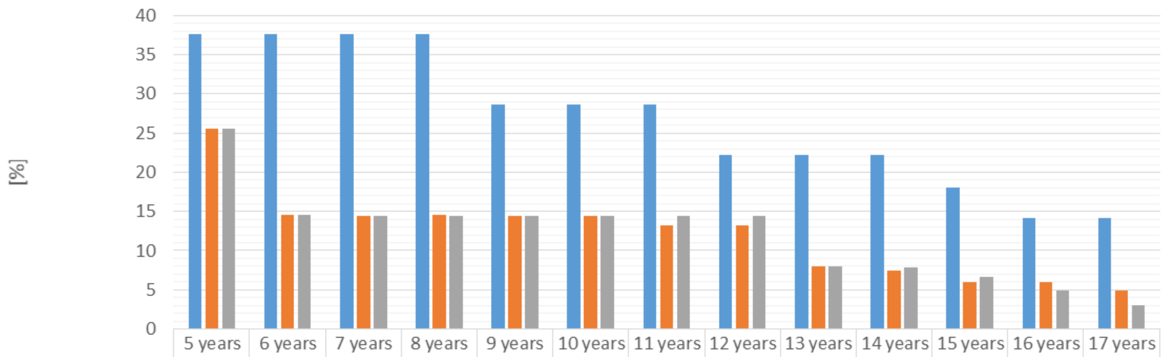
Figure 4 plots the results of this comparison for the maximum discrepancies on the YEARLY values. The simplified and filtered driver methods give very similar results, with a bias below 3% in all cases. The Sandia method can show more than 7% of bias.



	5 years	6 years	7 years	8 years	9 years	10 years	11 years	12 years	13 years	14 years	15 years	16 years	17 years
Sandia	5.2	7.3	4.8	4.0	2.3	4.2	3.3	4.1	2.9	4.0	4.1	2.5	0.3
Simplified Driver	3.0	3.3	3.0	3.4	2.6	1.3	1.8	2.0	2.0	1.4	0.9	0.4	0.5
Filtered Driver	3.0	3.3	2.7	2.6	2.1	1.5	2.0	2.0	1.5	1.5	0.9	1.0	0.4

Figure 4: Maximum discrepancies on yearly average TMY for different time windows

If the same comparison is done on the MONTHLY values, the differences between the Sandia and the driver method is also very large as seen in Figure 5.



	5 years	6 years	7 years	8 years	9 years	10 years	11 years	12 years	13 years	14 years	15 years	16 years	17 years
Sandia	37.7	37.7	37.7	37.7	28.6	28.6	28.6	22.3	22.3	22.3	18.0	14.2	14.2
Simplified Driver	25.5	14.6	14.4	14.6	14.4	14.4	13.2	13.2	8.0	7.5	6.0	6.0	4.9
Filtered Driver	25.5	14.6	14.4	14.4	14.4	14.4	14.4	14.4	8.0	7.8	6.6	4.9	3.1

Figure 5: Maximum discrepancies on monthly average TMY for different time windows

The full article can be found there:

Ana Realpe, Christophe Vernay, Sébastien Pitaval, Camille Lenoir, Philippe Blanc. Benchmarking of Five Typical Meteorological Year Datasets Dedicated to Concentrated-PV Systems. *Energy Procedia*, Elsevier, 2016, 97, pp.108 - 115. 〈10.1016/j.egypro.2016.10.031〉. 〈hal-01406110〉

P90 TMY creation methodology

In addition to the median year scenario, customers often needs to assess a “bad year” scenario output for the solar installation and this is an important part of the “bankable report” for the solar project. The common approach is to compute the solar installation output based on a P90 TMY (pessimistic scenario). A P90 year of data is supposed to be exceeded in 90% of the cases.

The typical P90, as per its name (P stands for “percentile”), is based on the percentile approach, derived from a Gaussian statistical analysis. Classically, this type of statistical approach is valid for a very large set of independent data. This is not true in our case, as the long-term time series of meteorological data are classically 12 to 25 years long when the irradiation is retrieved from satellite observation. In addition to this already short period, most of generation of satellites last only 10 to 12 years without major changes in the instrument aboard. This instrument type change needs to calibrate the older satellite technology part of the time series with the most recent one, and this violates further the principle of independence of the data from one year to the other.

These very stringent conditions can be lifted if we consider that the Central Limit Theorem can be applied to our time series of data.

Wikipedia: In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficient large number of iterates of identically distributed independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normal distributed, regardless of the underlying distribution.

In other words, for the subsequent statistical analysis, we assume that the annual values of irradiation can be considered as a normal (Gaussian) distribution.

The inter annual variability of a radiation component is calculated from the unbiased standard deviation **STD** for the whole period of complete years available. For the HelioClim-3 database, this is 12 complete years from 2005 to 2016 (at the time of redaction of this document).

$$STD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

… with:

- N: total number of years

- x_i: is the i^{th} sample of the yearly irradiation value
- $\bar{x}$: is the average value of the yearly irradiation value over the whole period of data available.

The variability **Vn** for a number of consecutive year **n** is obtained from the unbiased standard deviation **STD in percent** with the formula:

$$Vn = \frac{STD}{\sqrt{n}}$$

In statistics, if we admit that the annual values follow a normal distribution, 80% of the values are contained in the interval +/- 1.28155**STD*. By extension, we compute the uncertainty with the following expression:

$$uncertainty = 1.28155 * v_n$$

From this, we can deduce the lower and upper boundaries of the 80% values zone which represent respectively the 90% and the 10% of exceedance, also named percentile *P*90 (90% of the values are exceeding the limit) and percentile *P*10 (10% of the values only are exceeding the limit):

$$lower\ bound\ (P90) = \bar{x} - uncertainty$$

$$upper\ bound\ (P10) = \bar{x} + uncertainty$$

Please note also that the P90 TMY is based on an annual P90 analysis. To define this P90 year,we proceed as follows.

We compute the standard deviation of the long-term annual values of the driver irradiation and we multiply it by 1.28155 to obtain a P90 estimate. Then, the P90 representative year is simply chosen as the year for which the annual sum of the driver irradiation is the closest to the P90 estimation.

Table of illustrations

