The Solar Forecast Similarity Method: a new method to compute solar radiation forecasts for the next day

Alexandre Boilley, Claire Thomas, Mathilde Marchand, Etienne Wey, Philippe Blanc

Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 91 (2016) 1018 – 1023

Energy Procedia

SHC 2015, International Conference on Solar Heating and Cooling for Buildings and Industry

The Solar Forecast Similarity Method: a new method to compute solar radiation forecasts for the next day

Alexandre Boilley[a], Claire Thomas[a], Mathilde Marchand[a], Etienne Wey[a], Philippe Blanc[b]

[a]*TRANSVALOR, Parc de Haute Technology, 694 Av. du Doc. Maurice Donat, 06255 Mougins*
[b]*MINES ParisTech, PSL-Research University, CS10207, 06904 Sophia Antipolis Cedex*

Abstract

The need for PV plant owners to plan what they are injecting in the electricity grid is more and more stringent to avoid endangering the whole supply in electricity. A new solar forecast algorithm, named Solar Forecast Similarity Method, has been developed to predict irradiance for the next day based on a statistical study of the long term HelioClim-3 irradiation database. This algorithm searches in the past for the most similar days compared to the day of interest and uses their following days to produce a forecast.
The model has been optimized against the database itself to compute the most adequate set of parameters over France and for the month of January 2014. With this configuration, the results are a null bias and a root mean square error of 48%. The algorithm outperforms the persistence by 20% and the error is similar to existing methods. An objective validation has then been carried out to compare the irradiance forecasts to high quality measurements from several Baseline Surface Radiation Network (BSRN) ground stations. The method is very promising since the comparison results are in line or lower than the one obtained with the first validation analysis performed on the HelioClim-3 database.
For high frequencies, however, predictions have a high error for rapidly varying weather. This demonstrates that the method provides information for the averaged production the following day but requires another input to reliably predict high frequency irradiance.

Keywords: Solar; Forecast; Similarity; BSRN; Validation; HelioClim-3;

1. Introduction

High penetration of renewable energy plant in electricity grid increases the need for accurate prediction of the amount of power injected in the grid. As for wind resource, solar resource prediction is required for regulation

doi:10.1016/j.egypro.2016.06.270

problems for example. Even if intra-hour prediction may be interesting for operational issues, this paper focuses on the day after prediction (D+1) for production and load scheduling.

Several methods for prediction may be found in [1]. In the following the k-nearest-neighbors method (k-NN) is applied to a satellite derived database: HelioClim-3. The method is applied for several locations in different weather patterns and over several years which is the first time to the author's knowledge.

The HelioClim-3 solar radiation database uses the images from MeteoSat Second Generation (MSG) satellite to provide a reliable value of the ground irradiation from 2004 up to the current day [2]. This long time series covers inter-annual as well as inter-seasonal variations. For that reason, knowing the hourly irradiation for one day, it is likely that a similar scenario has already occurred in the past. Based on this assumption the k-NN method is appropriate to provide a reliable prediction.

The algorithm seeks in the past the nearest neighbors in term of square distance to the day D before the day D+1 for which the prediction is requested. The days immediately following these nearest days found in the past are used to predict the irradiance for D+1. Only a prediction for D+1 is considered in the following. The algorithm has been first compared to the HelioClim-3 database itself to determine the best configuration and parameters. The method and the corresponding equations are presented in section 2.

Using the optimize set of parameters, the algorithm has been tested over a long period and compared to nine Baseline Surface Radiation Network (BSRN) ground measurement stations. Quality check applied to measurements data and comparison results are presented in section 3.

Optimization of the algorithm enables to set up a web-service on the SODA website under the name Solar Forecast Similarity Method (see appendix B). This service is already available.

2. Method

2.1. Principle and equations

The principle of the Solar Forecast Similarity Method algorithm is based on the k-NN method described in [1]. It searches in the HelioClim-3 past irradiance database (x) the period which is the closest to the considered sample (y). The closest period is defined by considering the minimum of the square distance. Once the nearest neighbor(s) is(are) determined, the following day(s) is(are) used as a prediction for the day following the sample.

This search for the most suited day could be very time consuming if a day to day difference is applied directly. A lot of work has been done to reduce this time by a factor up to 50 using convolutions. Optimization has been obtained as described in the following.

At first the equation of the distance for each day (i) is given by a difference for every hour of the day (h):

$$d_i^2 = \sum_h (x_h - y_{ih})^2 = \sum_h ({x_h}^2 - 2x_h y_{ih} + y_{ih}^2)$$

This equation can be written as 3 convolutions as follows:

$$d_i^2 = ({x_h}^F * \mathbf{1}_x) - 2(y_{ih} * {x_h}^F) + ({y_{ih}}^2 * \mathbf{1}_y)$$

where ^{F} is an operator that returns the array with the order of its elements reversed, $\mathbf{1}_A$ is an array constituted of value 1 only and of length A.

Writing the difference this way leads to speed up, therefore a comparison over long time period is possible. Once the Euclidian distance d_i is computed, the nearest days' indexes (z) are extracted and the predicted irradiance (I) is made:

$$\mathrm{I} = \frac{\sum_z \alpha_z\, y_{z+1}}{\sum_z \alpha_z}$$

where α is a weight that may be dependent of the distance d.

Behind these equations, different parameters had to be adjusted, as explained in section 2.2.

2.2. Compute a set of optimized parameters

Many parameters can be tuned to obtain the best prediction:

- The length of the sample (y), 1 day, 2 days, etc
- The length K of the database used for the computation of the square distance, from 1 up to 10 years
- The time window (year independent) before and after the sample date (y). This parameter is set to look for days with the same day duration
- The number of past days and the weight used to realize the prediction.

A comparison between the predicted irradiance and the HelioClim-3 database itself over France for the month of January 2014 has been conducted to determine the optimal parameters. The month of January has been considered for the rapidly varying weather conditions. Indeed the prediction shall be easier for constant weather conditions.

The length of 1 day for the sample has been determined to be optimal. A two consecutive day's sample prediction generates an error 10 % higher than for a 1 day sample prediction.

Using the entire 11 years of the HelioClim-3 database leads to a worse prediction than the one obtained using only the most recent years; according to our tests, a day selected further in the past is less representative of the current situation surely due to inter annual variation and climate change. A 4-years database length has been determined as optimal for the prediction. Considering a 15 days period before and after the sample date in this 4-years length is a good compromise in term of day duration differences between the candidate days and the day of interest. Reducing the time window worsen the prediction (less choices available) whereas increasing it do not conduct to a significant improvement.

The number of selected days to predict the irradiance is by far the most important parameter. When using the 3 closest days instead of the closest only, the error is reduced by 20%. Considering the 10 closest days, the error is reduced by 27%. Beyond 10, the error is not reduced much further, so the irradiance obtained for these 10 selected days is averaged to generate the prediction. A weighting function depending on the square distance to the current day has also been tested, but doesn't improve the prediction results.

In brief, a temporal window of 15 days before and after the date and for the previous 4 years is considered for the prediction. The days following the 10 selected days in the past are averaged to produce the prediction. With this configuration, the bias is null and the algorithm outperforms the persistence by 20% and the error is similar to existing methods [3] compared to HelioClim-3 database.

3. Validation using Baseline Solar Radiation Network (BSRN) station measurements

To perform an objective validation of the database, the forecasts of the Solar Forecast Similarity method have been compared to the measurements from nine BSRN stations: Tamanrasset (Algeria), Brasilia and Sao Martinho da Serra (Brazil), Toravere (Estonia), Palaiseau and Carpentras (France), Sede Boqer (Israel), Cabauw (Netherlands) and Payerne (Switzerland). Measurements of these stations cover more than the 4 years defined in the previous section as an optimal period to obtain a reliable prediction. As a consequence, the prediction results are available for the period of measurement excluding the first 4 years.

Prior the comparison a thorough quality check procedure as recommended by [4] and [5] has been applied onto the BSRN data. The major steps can be summarized as follows:

- Set night, sunrise and sunset values to zero,
- Set to "Not a Number" (NaN) the values of the estimates when the references are missing, and reverse,
- Discard values beyond "extremely rare limits" and "physical possible limits",
- Perform the consistency checks when the three radiation components are available.

Then, a temporal aggregation is performed to generate the values at different summarizations. This procedure is of utmost importance since it directly impacts the validation results. Our approach is as follows:

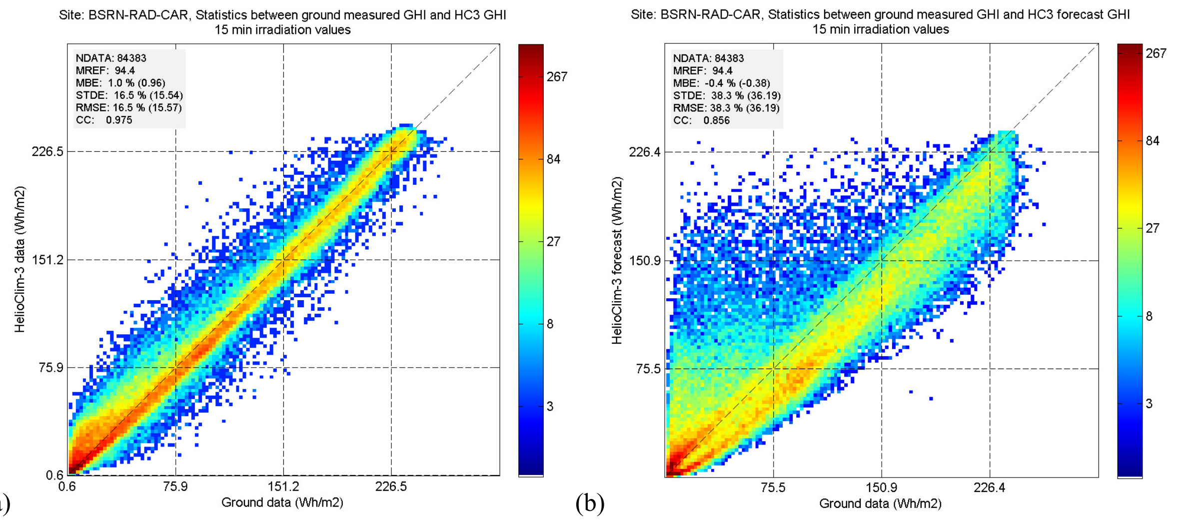
- Generate the 15 min irradiation from the 1 min BSRN measurements if at least 85% of the slots are available. Then apply an "intelligent interpolation" taking into account the sun position at each minute to synthesize the 15 min data to fill gaps. Compute the quantities summarizing the deviation at 15 min.

- Generate the hourly, daily and monthly irradiation by summing up the 15 min irradiation if at least respectively 75%, 65% and 50% of the slots are available. No temporal interpolation is applied, leading to partial sums. Compute the quantities summarizing the deviation at the hourly, daily and monthly time steps.

The predicted irradiation is given at a time step of 15 min. To compare at different summarizations, predicted irradiation is sum to obtain 1h, 1day and 1 month values. Statistics quantities are then computed: bias, Root Mean Square Error (RMSE) and correlation coefficient. Definitions of these quantities can be found in [6].

Results for the nine different stations at different summarizations are presented in Appendix A. At first, results are accurate for monthly values, which means large patterns are well predicted. As soon as the comparison is made for higher frequencies, discrepancies can be detected for stations in the Northern hemisphere with RMSE up to 58 % for Toravere station at 15 min time step. This is due to rapidly varying weather conditions. Indeed, for Carpentras for example, the algorithm often misses highly cloudy days after a sunny day. It means that a sunny day is most of the time followed by a sunny day in the past above Carpentras. The algorithm is not constrained enough to determine a sharp variation in the irradiation.

In general, the algorithm mainly underestimates irradiation (see Fig.1). This is due to the way prediction is made from indexes aggregation; 10 days are averaged to produce the prediction leading to a smooth synthetic day. Weighted mean could be thought as a better solution. Weighted mean gives however priority to the day after the nearest neighbor leading to a worse prediction when it is far from truth.



(a) (b)

Fig. 1 (a) 2-D histogram of the 15 min HC3v4 GHI values versus the measurements of the BSRN station of Carpentras; (b) Solar Forecast Similarity against Carpentras BSRN station measurements.

Taking into account another parameter seems to be mandatory to improve prediction. Considering pressure may help to determine a weather change and remove 1 degree of freedom to the algorithm. An attempt with the pressure from numerical weather prediction model (MERRA) has been conducted. The spatial resolution of the model was not high enough to capture local rapidly changing pressure. The algorithm was not driven enough and the prediction was not improved.

4. Conclusion and perspectives

A new method to compute the irradiance values for the day after based on the statistical investigation of the HelioClim-3 database has been presented. The objective validation campaign conducted on the measurements from nine high quality BSRN stations located in the MSG coverage has demonstrated a same level of error than other forecast methods, and overtakes the models based on persistence.

The work conducted to determine the optimal parameters has been conducted only for one location and one month. Optimization may be done on larger geographical area and longer time period to verify the validity of the parameters used in this paper. It has also been demonstrated that the algorithm is not constrained enough and may predict completely wrong irradiation some days when weather is rapidly varying.

Statistics for northern hemisphere over performed persistence but error remains important. Other parameters have to be used to improve the prediction. Taking into account 2 consecutives may constrain the algorithm for example; considering the irradiance prediction from numerical weather prediction model can be investigated.

In a near future and in the perspective to reach better results, an extension of the service to the statistical study in the vicinity surrounding the point of interest will be test. Using high resolution numerical weather prediction for the location considered may also be a perspective. Using the distribution of the nearest neighbors can be used to give error to the user or to construct the prediction based on conditional probability.

Appendix A. Validation results

This appendix proposes four tables (Table 1 to Table 4) which give the validation results of the Similarity Forecast method values against the measurements of the nine BSRN stations located in the coverage of the HC3v4 database, and respectively for the monthly, daily, hourly and 15 minute summarizations.

Table 1. Monthly sum of 15 min irradiation forecast values. Units are kWh m^{-2}.

Station	Number of values	Mean BSRN	Bias HC3v4 (relative in %)	Bias Forecast (relative in %)	RMSE HC3v4 (relative in %)	RMSE Forecast (relative in %)	Correl. coeff. HC3v4	Correl. coeff. Forecast
Toravere	50	86.9	-2.8 (-3.2%)	-2.4 (-2.8%)	8.3 (9.6%)	11.1 (12.8%)	0.987	0.974
Cabauw	65	88.2	-3.1 (-3.5%)	-3.7 (-4.2%)	7.3 (8.3%)	12.0 (13.6%)	0.994	0.980
Palaiseau	54	103.7	3.5 (3.4%)	2.6 (2.5%)	6.0 (5.8%)	11.9 (11.5%)	0.998	0.981
Payerne	29	107.8	-8.5 (-7.8%)	-11.0 (-10.2%)	10.1 (9.4%)	16.2 (15.0%)	0.996	0.985
Carpentras	61	130.0	1.3 (1.0%)	-0.4 (-0.3%)	3.8 (2.9%)	9.2 (7.1%)	0.999	0.990
Sede Boqer	44	166.9	-11.7 (-7.0%)	-12.9 (-7.8%)	13.9 (8.3%)	14.2 (8.5%)	0.992	0.995
Tamanrasset	65	159.9	0.1 (0.1%)	0.2 (0.1%)	8.1 (5.1%)	10.6 (6.6%)	0.973	0.953
Brasilia	27	144.8	5.7 (4.0%)	8.5 (5.8%)	7.7 (5.3%)	12.6 (8.7%)	0.989	0.958
Sao Martinho da Serra	32	141.2	-1.9 (-1.3%)	-3.5 (-2.5%)	4.2 (3.0%)	10.9 (7.8%)	0.997	0.978

Table 2. Daily sum of 15 min irradiation forecast values. Units are kWh m^{-2}.

Station	Number of values	Mean BSRN	Bias HC3v4 (relative in %)	Bias Forecast (relative in %)	RMSE HC3v4 (relative in %)	RMSE Forecast (relative in %)	Correl. coeff. HC3v4	Correl. coeff. Forecast
Toravere	1481	2795.3	-82.1 (-2.9%)	-56.4 (-2.0%)	445.3 (15.9%)	1140.4 (40.8%)	0.978	0.843
Cabauw	1946	2940.1	-103.6 (-3.5%)	-123.1 (-4.2%)	378.1 (12.9%)	1106.9 (37.6%)	0.987	0.868
Palaiseau	1639	3412.4	114.2 (3.3%)	101.1 (3.0%)	332.4 (9.7%)	1144.7 (33.5%)	0.991	0.867
Payerne	877	3558.8	-278.9 (-7.8%)	-360.5 (-10.1%)	467.7 (13.1%)	1263.8 (35.5%)	0.987	0.862
Carpentras	1824	4359.2	43.5 (1.0%)	-14.4 (-0.3%)	319.3 (7.3%)	1168.1 (26.8%)	0.992	0.881
Sede Boqer	1287	5647.8	-398.2 (-7.1%)	-418.3 (-7.4%)	553.4 (9.8%)	741.6 (13.1%)	0.981	0.949
Tamanrasset	1723	5624.6	6.3 (0.1%)	25.3 (0.4%)	566.3 (10.1%)	902.5 (16.0%)	0.908	0.762
Brasilia	761	5073.7	194.7 (3.8%)	306.4 (6.0%)	583.5 (11.5%)	1046.0 (20.6%)	0.911	0.662
Sao Martinho da Serra	970	4721.2	-67.2 (-1.4%)	-148.5 (-3.1%)	405.3 (8.6%)	1742.7 (36.9%)	0.986	0.672

Table 3. Hourly sum of 15 min irradiation forecast values. Units are kWh m^{-2}.

Station	Number of values	Mean BSRN	Bias HC3v4 (relative in %)	Bias Forecast (relative in %)	RMSE HC3v4 (relative in %)	RMSE Forecast (relative in %)	Correl. coeff. HC3v4	Correl. coeff. Forecast
Toravere	18095	221.9	-6.0 (-2.7%)	-3.3 (-1.5%)	60.3 (27.2%)	126.5 (57.0%)	0.958	0.800
Cabauw	22886	249.6	-9.0 (-3.6%)	-10.8 (-4.3%)	56.7 (22.7%)	126.9 (50.8%)	0.968	0.821
Palaiseau	19680	285.1	9.3 (3.3%)	6.6 (2.3%)	51.4 (18.0%)	134.1 (47.1%)	0.978	0.830
Payerne	10167	306.7	-24.3 (-7.9%)	-31.6 (-10.3%)	67.3 (22.0%)	144.3 (47.1%)	0.970	0.839
Carpentras	21959	362.3	3.5 (1.0%)	-1.9 (-0.5%)	47.2 (13.0%)	133.7 (36.9%)	0.986	0.878
Sede Boqer	14417	505.6	-35.3 (-7.0%)	-39.0 (-7.7%)	69.2 (13.7%)	98.0 (19.4%)	0.982	0.957

Tamanrasset	20996	479.2	0.7 (0.1%)	0.4 (0.1%)	71.7 (15.0%)	120.6 (25.2%)	0.974	0.926
Brasilia	9428	417.1	16.4 (3.9%)	24.2 (5.8%)	110.4 (26.5%)	154.2 (37.0%)	0.933	0.860
Sao Martinho da Serra	11567	396.3	-5.8 (-1.5%)	-13.0 (-3.3%)	73.3 (18.5%)	190.6 (48.1%)	0.971	0.787

Table 4. 15 min irradiation forecast values. Units are kWh m^{-2}.

Station	Number of values	Mean BSRN	Bias HC3v4 (relative in %)	Bias Forecast (relative in %)	RMSE HC3v4 (relative in %)	RMSE Forecast (relative in %)	Correl. coeff. HC3v4	Correl. coeff. Forecast
Toravere	71550	61.6	-2.0 (-3.2%)	-1.7 (-2.8%)	20.4 (33.2%)	36.0 (58.5%)	0.935	0.779
Cabauw	88643	64.7	-2.3 (-3.5%)	-2.7 (-4.2%)	19.0 (29.4%)	35.3 (54.6%)	0.944	0.787
Palaiseau	76660	73.4	2.4 (3.3%)	1.8 (2.5%)	18.1 (24.7%)	37.5 (51.1%)	0.957	0.795
Payerne	39758	78.6	-6.2 (-7.8%)	-8.0 (-10.2%)	21.3 (27.1%)	39.2 (49.9%)	0.951	0.814
Carpentras	84383	94.4	1.0 (1.0%)	-0.4 (-0.4%)	15.6 (16.5%)	36.2 (38.3%)	0.975	0.856
Sede Boqer	56695	130.7	-9.2 (-7.0%)	-10.1 (-7.7%)	20.7 (15.8%)	27.2 (20.8%)	0.970	0.943
Tamanrasset	80956	128.4	0.1 (0.1%)	0.1 (0.1%)	23.0 (17.9%)	34.8 (27.1%)	0.958	0.903
Brasilia	36273	110.2	4.6 (4.2%)	6.7 (6.0%)	37.1 (33.7%)	45.7 (41.5%)	0.881	0.807
Sao Martinho da Serra	44186	104.3	-1.4 (-1.4%)	-3.2 (-3.1%)	24.7 (23.7%)	51.9 (49.8%)	0.949	0.749

Appendix B. Access to the Solar Forecast Similarity online service

The service can be access through the SODA web page at the following address:
http://www.soda-pro.com/soda-products/hc3-similarity-forecast
The methodology and the comparison results are available on this page. The service can be access for free and for unlimited requests until the end of 2015.

References

[1] Rich H. Inman, Hugo T.C. Pedro, Carlos F.M. Coimbra. Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science, Volume 39, Issue 6, 2013, Pages 535–576.

[2] Espinar B, Blanc P., Wald L, Gschwind B, Ménard L, Wey E., et al. HelioClim-3: a near-real time and long-term surface solar irradiance database. COST-WIRE-ES1002: Workshop May 22nd-23rd 2012, Roskilde, Denmark.

[3] Sengupta M, Habte A, Kurtz A, Dobos A, Wilbert S, Lorenz E, et al.. Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications. Technical Report NREL/TP-5D00-63112, 2015.

[4] Roesch, A., Wild, M., Ohmura, A., Dutton, E.G., Long, C.N., and Zhang, T.: Assessment of BSRN radiation records for the computation of monthly means, Atmos. Meas. Tech., 4, 339–354, doi:10.5194/amt-4-339-2011, 2011. Corrigendum, Atmos. Meas. Tech., 4, 973-973, doi:10.5194/amt-4-973-2011, 2011.

[5] WMO. Technical Note N° 172, WMO-No. 554, World Meteorological Organization, Geneva, Switzerland, 1981, p. 121-123

[6] Thomas C., Wey E., Blanc P., Wald L., Lefèvre M.. Validation of HelioClim-3 version 4, HelioClim-3 version 5 and MACC-RAD using 14 BSRN stations. SHC 2015, International Conference on Solar Heating and Cooling for Buildings and Industry.