|  |  |
| --- | --- |
| C:\Users\rudy-\Desktop\Challenge_Data\Logos\logo_challenge.png | Challenge proposal |

I) Provider description

The Center Observation, Impacts, Energy (OIE) is a research center of MINES ParisTech, within the Department Energy and Processes (DEP). The scientific objectives of OIE are to assess and to analyze spatial and temporal aspects of renewable energies resources and the environmental impacts of energy systems. It develops research activities in the domains of renewable resources evaluation (solar, wind, etc.), of meteorology for energy, of environmental impacts of energy uses, of interoperability and dissemination of information through databases and Web services. Among others, OIE is currently working on the modelling, the assessment and the forecasting the solar radiation, its spectral distribution, its angular distribution on the sky vault from in situ measurements, satellite images and outputs of numerical models for meteorological and chemical-transportation in the atmosphere such as CAMS (Copernicus Atmosphere Monitoring Service).

II) Contact

*Name and email of the people responsible for the project*

Prof. Thierry Ranchin, [thierry.ranchin@mines-paristech.fr](mailto:thierry.ranchin@mines-paristech.fr)  
Prof. Philippe Blanc, [philippe.blanc@mines-paristech.fr](mailto:philippe.blanc@mines-paristech.fr)  
Dr. Yves-Marie Saint-Drenan, [yves-marie.saint-drenan@mines-paristech.fr](mailto:yves-marie.saint-drenan@mines-paristech.fr)

III) Problem definition

In the domains of Solar Energy and energy meteorology, there is a need for accurate intraday — named hereinafter short-term — solar forecasting. Indeed, short-term forecasts allow a better integration of photovoltaic (PV) systems by anticipating the variability of solar radiation in space and time. This is particularly important in electric systems with a high penetration of solar energy, where the dispatch of generation units to match electricity production and consumption at each time is particularly challenging. This need not only holds for large scale electric grid PV integration but also in the special case of off-the-grid electricity supply systems.

Geostationary satellites, notably thanks to algorithms such as Heliosat, are a source of spatially and temporally resolved of surface solar irradiance (SSI), the “fuel” of PV systems (unit: W/m2). In the framework of Copernicus Atmospheric Monitoring Services (CAMS), the multispectral images acquired by Meteosat Second Generation (MSG) at the longitude 0° are used to provide, in near-real-time basis, every 15 min, images of SSI and SSI under clear-sky condition at 3 km resolution. These services –resp. CAMS Rad and CAMS McClear– are operated and maintained by Transvalor Innovation SoDA ([www.soda-pro.com](http://www.soda-pro.com))[[1]](#footnote-1), in collaboration with the DLR, the German Aerospace. This source of time series of SSI images is notably used to provide short-term (up to 2 hours) solar forecasting. The state-of-the-art of such satellite-based short-term solar forecasting is based on cloud motion vector (CMV) using optical flow or block-matching techniques.

The aim of this challenge is to propose machine learning approaches on sequences of images to provide better short-term forecast of future image of SSI on horizontal plan, noted GHI for Global Horizontal Irradiance, for time horizon ranging from 15 minutes to 1 hour, with a time resolution of 15 min and a spatial resolution of 3 km.

More precisely, we are interested in a square region of interest (RoI) of size 51 pixels x 51 pixels (approx. 150 km). With an assumption of max cloud speed of 10+ m/s, and considering solar forecasting up to 1 hour ahead, the observation region (OR) encompassing the RoI have a size of 81 pixels x 81 pixels (approx. 240 km).



Figure 1. Definition of the Observation Region (81 pixels x 81 pixels) and the Region of Interest (51 pixels x 5 pixels)

At a given time t, one hour after the sunrise and one hour before the sunset, considering the sequence of the 4 previous GHI images on the OR every 15 min, the solar forecasting aims at predicting the GHI images on the RoI for the next times ahead, ranging from the next 15 min up to the next hour with a time step of 15 min. This forecast of GHI for the location , done at the time for the future time is noted:

The learning phase is done on one year of data and the test phase is done on a separate year.

In this challenge, we will only consider the cloud effects on GHI, assuming that the concomitant and collocated GHI under clear-sky condition (with no cloud) is perfectly known and noted GHIcls.

The figure below is an example of concomitant and collocated SSI (global horizontal irradiance) and GHIcls:

Chart, line chart

Description automatically generated

Figure 2. Example of SSI

Contextual information of interest are the corresponding solar zenith angles (SZA) , solar azimuth angle (SAA) , following the convention:

Diagram

Description automatically generated

Figure 3. Solar Zenith Angle (SZA) and Solar Azimuth Angle (SAA) from (Blanc et Wald, 2012)[[2]](#footnote-2).

IV) Data description

The input vector X represents the concatenation for a given time t of:

* a sequence of the 4 previous 15-min GHI images of size (81,81) on the OR
* a corresponding sequence of the 4 previous and 4 next modelled 15-min clear-sky (i.e. with no clouds) GHI images on the OR, noted GHIcls, of size (81,81)
* a corresponding sequence of the 4 previous and 4 next modelled 15-min SZA and SAA on the OR, both of size (81, 81)

More precisely, for a time t, X is composed of:

* 5 values: YYYY (year), MM (month), DD (day), HH (hour Universal Time), MIN (minute UT)
* The matrix GHI of size (81, 81, 4) for the times t-45min, t-30min, t-15min, and t
* The matrix GHIcls of size (81, 81, 8) for the times t-45min, t-30min, t-15min, t, t+15min, t+30min, t+45min, t+60min
* The matrix SZA of size (81, 81, 8) for the times t-45min, t-30min, t-15min, t, t+15min, t+30min, t+45min, t+60min
* The matrix SAA of size (81, 81, 8) for the times t-45min, t-30min, t-15min, t, t+15min, t+30min, t+45min, t+60min

These matrix are vectorized following the order of the dimensions: for a matrix of size , the vector index k corresponding to the triplet of 0-base indices is (warning: this is the vectorization order of Matlab, not the one of Python).

The output vector y represents the sequence of the 4 next 15-min GHI images on the RoI, corresponding to a matrix of size (51,51,4), for the 4 future times t+15min, t+30min, t+45min and t+60min.

Note: the OR and the RoI are concentric: with python-like index RoI = OR[15:66,15:66].

For illustration and validation purposes, the following figures represent the 100th element of Xtrain and Ytrain

Graphical user interface, chart

Description automatically generated

Figure 4. GHI matrix of the 100th element of Xtrain (OR)

Chart

Description automatically generated

Figure 5. GHI matrix of the 100th element of Ytrain (RoI).

With an 2-digit ASCII coding, without compression, each element X has a size of approx. 1.3 Mo and each element of Y has a size of approx. 133 ko.

The learning and testing dataset correspond each to one distinct year: there will be approx. 4 000 elements of X and Y, corresponding to approx. 5 Go.

V) Metrics

Different metrics can be used, but, for the challenge, two simple metrics can be considered: the RMSE (root mean square error) and the MAE (mean absolute error) between the forecasted and the real y. These metrics should be computed globally for the whole test dataset, with respect each time horizon, from 15 min to 1 hour, with a time step of 15 min (4 values).

To better contrast the metrics, it is usual to consider the forecasting skill-score (SC) for both RMSE and MAE, with respect the persistence (P) forecast (see next section): if M stands for RMSE or MAE,   
SCM=1-M/MP.

VI) Benchmark

Two simple forecasts methods will be provided for the benchmark:

* the persistence forecasting P: . This method of forecasting is used as a baseline, to compute the skill-score (SC)
* the CMV forecasting based on a state-of-the-art optical flow and CMV persistence.

1. *http://www.soda-pro.com/web-services/radiation/cams-radiation-service/info* [↑](#footnote-ref-1)
2. Blanc, P., & Wald, L. (2012). The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period. Solar Energy, 86(10), 3072–3083. <https://doi.org/10.1016/j.solener.2012.07.018> [↑](#footnote-ref-2)