User manual for the SunPoint solar irradiance map product

J. Rabault, M. L. Sætra, A. Dobler, S. Eastwood, H. O. Hygen, E. Berge

October 18, 2024

Abstract

This is the user manual for the SunPoint solar irradiance map product. The solar irradiance is given as Global Horizontal Irradiance (GHI) values at the surface. The product contains machine learning (ML) optimized GHI maps over continental Norway, covering the period 1991-2020 with hourly, 0.03° (approximately 3.3 km) resolution.

1 Methodology

The product is obtained by combining data from the following sources:

- ERA5: the fifth generation global numerical reanalysis for climate and weather, produced by the European Center for Medium-Range Weather Forecasts (ECMWF) [1]. This provides data sources with 0.25° resolution (corresponding to approximately 30 km resolution) for atmospheric conditions.
- NORA3: a high-resolution numerical weather simulation produced at MET Norway that covers the Norwegian area with a resolution of 3 km [2]. Similarly to ERA5, this provides data sources for atmospheric conditions.
- CMSAF-SARAH3: a satellite-based solar irradiance dataset [3]. The dataset is obtained from geostationary satellites observations. This provides high-resolution (30 minute in time, 0.05° in space corresponding to approximately 5 km), direct observations of the solar irradiance. Since this is a geostationary satellite product, data are not available north of 65° north. Although this provides direct high-quality observations of solar irradiance, there are still uncertainties arising from, e.g., the effect of clouds and snow and ice.

In addition, we used data from pyranometers as ground truth to train and validate our model. Surface measurements of hourly global horizontal irradiance (GHI) were collected at 106 locations in Norway between 2016 and 2020. Most of these data are openly available through the Frost API of the Norwegian Meteorological Institute `frost.met.no`. These have also been supplemented with additional data from four locations collected within the SunPoint project.

Since the pyranometer data had not undergone prior quality control, visual inspection and quality flagging routines have been implemented to only use high-quality station and measurement data to train and validate our ML models [4]. This results in a final selection of 47 stations where the hourly GHI data have been collected in a NetCDF file. This dataset is freely available at `zenodo.org/records/8082726` [5] together with visualization and quality control plots, scripts, and report files that include flags and cleaned data, as well as the original raw data of the 106 stations.

The combination of data sources is used to generate the best estimate of the GHI. The different estimators used to combine the data sources have been trained and validated using the data from the pyranometer measurements. If all data sources are available at a given location and time, the best available model (i.e. a simple supervised Machine Learning (ML) model using a fully connected neural network (NN) estimator) is used to estimate the corresponding solar irradiance. If some data are not available for a given point in space or time to apply the best ML model, a simpler model is used instead for the corresponding time and location estimate. The models used are summarized in Table 1. For each location and time, the best (i.e., a higher "model index") data output that is available is used in the final product.

model_index	model description
6	nn_full: the best ML model, a simple fully connected NN, using all data sources
5	sklearn_full: a simpler linear regression model, using all data sources
4	SSI_value_cmsaf_sis_sarah: the CMSAF-SARAH data source alone
3	nn_nosat: a ML NN model, using only model (ERA5 + NORA3) data
2	sklearn_nosat: a simpler linear regression model, using only model (ERA5 + NORA3) data
1	rsds_era5: the ERA5 data source alone
0	rsds_nora3: the NORA3 data source alone

Table 1: The summary of models used. At each location in space and time, the best (i.e., highest model index) valid prediction for which all data sources are available is used.

North of 65 degrees north, CMSAF-SARAH3 data are not available since this data source relies on geostationary observations that cannot be performed close to the poles. Therefore, only model indexes 3 and lower are obtained in the northernmost part of Norway.

The ML NN model is the most sensitive to the input source data quality. Therefore, the nn_full model (model index 6) can only be produced when all data sources are available with good quality. Moreover, the nn_full model is only trained for GHI higher or equal than 50 W/m^2. Therefore, model_index 6 is only available south of 65N (that is, where geostationary satellite data are available), when all data sources have high quality, and the GHI is greater than or equal to 50 W / m^2. In all other cases, predictions of lower model_index are provided (i.e., we provide the output of the highest model_index that could be produced given available data). For a more complete description of the input data, the methodology, validations, and discussions, the reader is referred to additional materials linked to the Github repository.

2 Files resolution, location, and format

The data files of the product are hosted as netCDF-CF files on the thredds server of the Norwegian Meteorological Institute: https://thredds.met.no/thredds/catalog/sunpoint/ML-Optimized-Maps/catalog.html.

There are 3 types of data files provided:

- (1) solar radiation at ground, chunked to make it efficient to access a full map at a given point in time. These files provide the instantaneous solar irradiance at ground in W/m^2 at each point in space and 1-hour resolution. Both the solar irradiance value, as well as the model_index (see Table 1) that was used to produce it, are provided at each point. These files should be used to, e.g., plot the solar irradiance over a given area in space for a single point in time. See the files https://thredds.met.no/thredds/catalog/sunpoint/ML-Optimized-Maps/hourly/field_access/catalog.html and related.

- (2) solar radiation at ground, chunked to make it efficient to access the timeseries for a single point in space over the whole time coverage of the dataset. These are the exact same data as (1) and the units and fields are identical, except that the chunking, i.e. the memory layout of the data, are optimized for a timeseries access pattern. These files should be used to, e.g., extract timeseries of solar irradiance at a selection of points. See the files https://thredds.met.no/thredds/catalog/sunpoint/ML-Optimized-Maps/hourly/point_access/catalog.html and related.

- (3) monthly accumulated values data files; these files provide the accumulated solar energy in kWh/m^2, for each point in space and each month between 1991 and 2020. See the file https://thredds.met.no/thredds/catalog/sunpoint/ML-Optimized-Maps/monthly/catalog.html.

We provide an example of how to access and plot the data in the following Python3 Jupyter notebook: https://github.com/metno/2024_Sunpoint_solar_irradiance_map_over_Norway_1991-2020. This makes heavy use of the xarray package to open netcdf-CF datasets.

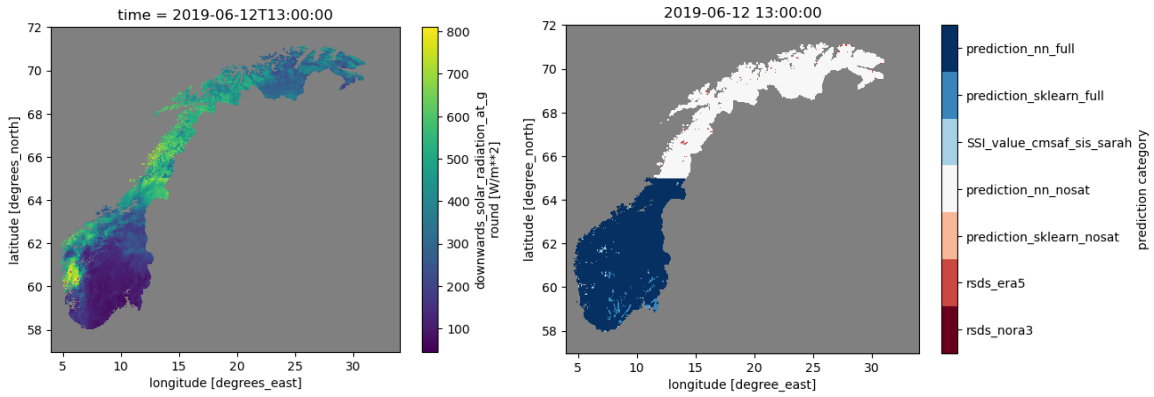

Figure 1: Illustration of the data produced. Solar irradiance at ground is predicted over continental Norway with a spatial resolution of 0.03° (approximately 3.3 km) and a time resolution of 1 hour. Files contain both the best estimate for the irradiance at ground, and an information flag about what model was used to produce the data at each point in space and time.

3 Note about the data

Due to challenges with the quality of some of the data sources over the period 2012-05 to 2012-08, only ERA5 data (corresponding to model_index 1) are used over the corresponding period.

4 Links summary

- GitHub repository: please visit the GitHub repository for examples of how the product can be used, as well as updates and further information: https://github.com/metno/2024_Sunpoint_solar_irradiance_map_over_Norway_1991-2020. This includes an example Jupyter notebook showing how to use the data.
- Product data files on thredds.met.no: main entry point: https://thredds.met.no/thredds/catalog/sunpoint/ML-Optimized-Maps/catalog.html.

5 Contact information

Please take contact us either through the GitHub issue tracker at https://github.com/metno/2024_Sunpoint_solar_irradiance_map_over_Norway_1991-2020, or by writing to post@met.no.

6 Data License

These data are released under the Norwegian license for public data (NLOD) and Creative Commons 4.0 BY, see https://www.met.no/en/free-meteorological-data/Licensing-and-crediting.

Acknowledgements

This document and the GHI maps have been produced in the context of the SunPoint project, funded by The Research Council of Norway (grant number 320750).

References

[1] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, *et al.*, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[2] H. Haakenstad, Ø. Breivik, B. R. Furevik, M. Reistad, P. Bohlinger, and O. J. Aarnes, "NORA3: A nonhydrostatic high-resolution hindcast of the North Sea, the Norwegian Sea, and the Barents Sea," *Journal of Applied Meteorology and Climatology*, vol. 60, no. 10, pp. 1443–1464, 2021.

[3] U. Pfeifroth, S. Kothe, J. Drücke, J. Trentmann, M. Schröder, N. Selbach, and R. Hollmann, "Surface Radiation Data Set - Heliosat (SARAH) - Edition 3," 2023.

[4] A. Dobler, E. Berge, H. O. Hygen, S. Grini, and H. Nygard Riise, "Quality control of radiation data for solar resource mapping," *MET Norway Report*, no. 8, 2024.

[5] A. Dobler and E. Berge, "Quality controlled observations of hourly incoming shortwave radiation data at the surface for solar resource mapping in norway (2016-2020)," 2024.