Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 158 (2019) 6378–6383

Energy Procedia

www.elsevier.com/locate/procedia

10th International Conference on Applied Energy (ICAE2018), 22-25 August 2018, Hong Kong, China

Spatio-temporal modelling and uncertainty estimation of hourly global solar irradiance using Extreme Learning Machines

Alina Walch[a*], Roberto Castello[a], Nahid Mohajeri[b], Fabian Guignard[c], Mikhail Kanevski[c], Jean-Louis Scartezzini[a]

[a]*Solar Energy and Building Physics Laboratory (LESO-PB), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland*
[b]*Sustainable and Urban Development Programme, Department for Continuing Education, University of Oxford, OX 12JA Oxford, UK*
[c]*Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne (UNIL), 1015 Lausanne, Switzerland*

Abstract

Solar photovoltaic (PV) is one of the most promising technologies for the transition from fossil fuels to renewable energy production. Accurate spatial and temporal modelling of solar irradiance is a key factor in the evaluation of PV technology potential for harvesting solar energy. We present here a data-driven approach based on an ensemble of Extreme Learning Machines using geographic and topographic features in input to predict the global horizontal irradiance in Switzerland from coarse-resolution satellite measurements. This provides a precise mapping of hourly global solar irradiance for each (250 × 250) m^2 pixel of a grid covering the entire country. The uncertainty on predicted values is quantified through a variance-based analysis, able to distinguish between model and data uncertainty. The former amounts to 1%, whereas the latter is close to 15% of the predicted values. The presented methodology is scalable and applicable to any large environmental dataset. Our modelling of solar irradiance at hourly temporal resolution and of its uncertainty will allow for an estimate of hourly PV potential in Switzerland to facilitate a more efficient integration of solar photovoltaics into the built environment.

Keywords: Hourly solar irradiance; Extreme Learning Machines; Uncertainty; Satellite data

* Corresponding author. Tel.: +41 21 69 35553.
E-mail address: alina.walch@epfl.ch

1876-6102
10.1016/j.egypro.2019.01.219

1. Introduction

Current climate and environmental policies in Switzerland and worldwide aim at a strong reduction of CO_2 emissions in the next decades by transitioning from fossil fuels to renewable energy. Harvesting solar energy using photovoltaic (PV) and solar thermal technologies is one promising approach to achieve the ambitious emission targets. To determine the potential for large-scale deployment of solar technologies and to assess the requirements for a successful integration into the built environment, an accurate modelling of the spatial and temporal patterns of solar irradiance is essential. In this study, we present a methodology for modelling environmental variables at high spatial and temporal resolution by using large satellite datasets. We apply it to predict hourly global horizontal irradiance (GHI) on a (250 × 250) m^2 grid, in order to be able to estimate PV potential at the neighborhood scale in Switzerland.

Several data-driven methods exist to model solar irradiance. These include averaging the nearest neighbors [1], geostatistical methods such as kriging [2] as well as machine learning approaches such as Support Vector Machines [3], Random Forests [4] and neural networks [5,6]. As averaging tends to oversimplify the modelling, and kriging is computationally intensive and requires modelling of anisotropic spatial correlations and stationarity of the process, the data-driven machine learning algorithms have recently gained much attention due to their performance and speed [7]. Most studies however focus either on a high spatial or temporal resolution and frequently do not consider the uncertainty that is intrinsic to the modelling [8]. We use an Extreme Learning Machine (ELM) ensemble algorithm that allows to predict solar irradiance at an hourly time granularity for a high spatial resolution (250 × 250) m^2 grid over Switzerland and to estimate the uncertainty related to the model and the data. The main advantages of this algorithm are (i) its fast training time and (ii) its ability to learn and to model complex non-linear phenomena with the desired precision. Hence this algorithm is well suitable for the analysis and modelling of large datasets [6,9,10].

In this paper we present a methodology to accurately model hourly global horizontal irradiance at the scale of Switzerland using an ensemble of machine learning models. The methodology for training the ELM ensemble from topographic features and for estimating the modelling uncertainty is presented in Section 2. In Section 3, the input dataset and the high-resolution grid used for prediction are described. Finally, the resulting spatial and temporal prediction and model performance are discussed in Section 4 and the results are visualized in maps, along with the corresponding model and data uncertainties.

2. Methodology

2.1. Extreme Learning Machine Ensemble algorithm

Extreme Learning Machines, as proposed by Huang *et al.* [11], are single-layer feed-forward neural networks which are trained using a single-step optimization. The training process of ELM is up to hundreds of times faster than the training of other machine learning algorithms [11]. For our model we aggregate M ELM to an ensemble, as shown in Fig. 1. This is known to improve the generalization performance by reducing the risk of overfitting [6,9,12], and allows to estimate the uncertainty as described in Section 2.2. At the hidden layer of each ELM, the training points $\mathbf{x}_i$ of dimensionality d are multiplied by the randomly chosen input weights $\mathbf{w}$ and added to the random bias b. An activation function $f(\,)$ is applied to the resulting random projections for modelling nonlinear behaviour of the data. We use a sigmoid activation function in this study, which is a common choice for regression problems [12]. The hidden layer is multiplied by the output weights β and summed to give the model output $\hat{y}_i$ (see Eq.1). During training, only β needs to be optimized.

For ensemble training, we apply a bootstrap-aggregating (bagging) approach [13] where each ensemble member is trained on one bootstrapped resample of the training data. Each bootstrap replicate is obtained by resampling the N training samples N times uniformly and with replacement. On average, every replicate contains 63.2% of the original data, with some duplicated samples. At the output, the predictions $\hat{y}_i^m$ of each ELM are averaged to give the final estimation $\hat{y}_i$ as shown in Eq. 2. The generation of M bootstrap replicates comes at a computational cost. For the implementation of each ELM, we therefore use the python toolbox *hpelm* with GPU acceleration [14]. This showed significant speed-up in performance compared to the standard CPU implementation. This implementation allows us to train and to test ELM ensembles with a large number of hidden neurons on our large environmental dataset.

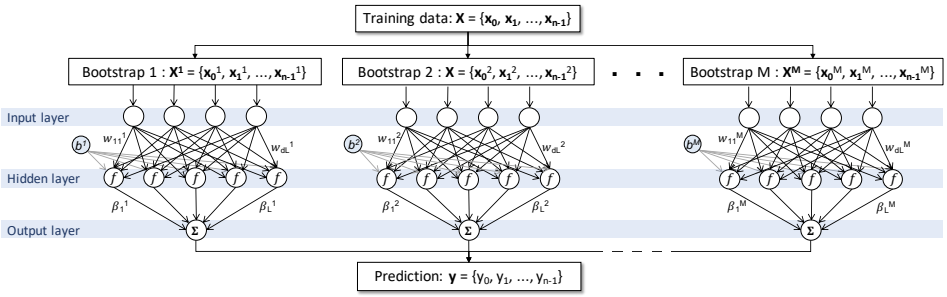

Fig. 1. Bootstrapped ELM ensemble showing 3 copies of an Extreme Learning Machine. Input weights $\mathbf{w}_j$ and biases b_j are selected randomly, output weights β_j are optimized during training. The outputs of all ELMs are averaged to give the predicted values $\hat{y}_i$.

$$\hat{y}_i^m = \sum_{j=1}^{L} \beta_j^m f(\mathbf{w}_j^m \mathbf{x}_i + b_j^m), \quad \mathbf{x}_i \in \mathbb{R}^d, \mathbf{w}^m \in \mathbb{R}^{L \times d}, \; i = 1,\ldots,N, \; m = 1,\ldots,M \tag{1}$$

$$\hat{y}_i = \frac{1}{M} \sum_{m=1}^{M} \hat{y}_i^m, \quad i = 1,\ldots,N \tag{2}$$

2.2. Uncertainty estimation

Applying bagging to the ELM ensemble increases the quality of the prediction and it also allows to compute the model uncertainty $\hat{\sigma}_M$, derived from the variance $\hat{\sigma}_M^2$ of the M model outputs $\hat{y}_i^m$ (see Eq. 3) [9,15]. Afterwards, remaining residuals are derived by subtracting the model variance from the squared difference between the model outputs and the targets t_i, as visible in Eq. 4. We compute these residuals for the training data, which are used to train a second ELM ensemble. The output is the data noise variance $\hat{\sigma}_D^2$, and the standard deviation $\hat{\sigma}_D$ quantifies the data uncertainty. Adding model and data noise variance provides a qualitative estimate of the uncertainty on the predictions.

$$\hat{\sigma}_M^2(\mathbf{x}_i) = \frac{1}{M} \sum_{m=1}^{M} \left(\hat{y}_i^m - \hat{y}_i \right)^2, \quad i = 1,\ldots,N \tag{3}$$

$$\hat{\sigma}_D^2(\mathbf{x}_i) = \max\left\{ \left(t_i - \hat{y}_i \right)^2 - \hat{\sigma}_M^2(\mathbf{x}_i), \; 0 \right\} \tag{4}$$

3. Dataset description

To model solar irradiance in Switzerland, we combine meteorological data from satellite measurements with topographic features from a digital elevation model (DEM). The satellite data of hourly GHI is available through the Swiss Federal Office of Meteorology and Climatology (www.meteoswiss.admin.ch**)** in raster format. Each cell covers 1.25° of longitude and latitude, or approximately (1.6 × 2.3) km^2, giving a total of 11,243 locations. The dataset includes historical measurements of GHI for 12 years (2004-2015), which are averaged to provide a smoother input to the model. Figure 2(a) shows the spatial distribution of the yearly sum of the averaged GHI, while Fig. 2(b) shows their temporal distribution. Noticeably, both spatial and temporal features are essential to accurately model GHI.

We merge the meteorological data with topographic features extracted from the DEM, namely altitude, x- and y-coordinates. The DEM is accessible through the Swiss Federal Office of Topography (www.swisstopo.admin.ch) and is aggregated to a raster of (250 × 250) m^2 for the modelling of GHI, defining a grid of about 640,000 pixels covering the entire Switzerland. Other topographic features that may be derived from the DEM, such as slope or orientation, meteorological variables that are considered in related studies [3,4,8] are not used here, as adding these variables would require a prepended modelling step, introducing significant additional uncertainty. The resulting input dataset is split into 12 smaller subsets, one for each month, allowing a customized model training and thus better accounting

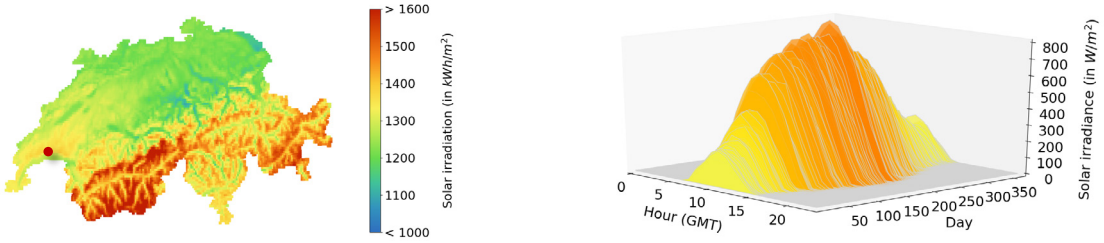

Fig. 2. (a) Yearly solar irradiation from satellite data, (b) hourly distribution of solar irradiance at selected location, averaged for 2004-2015.

for seasonal variations in predicting the irradiance. The training data for each subset has 4 input features (x-coordinate, y-coordinate, altitude, hour) and a single output target (hourly GHI). During data pre-processing, all hours with constantly zero GHI measurements are removed (i.e. night hours). The remaining 10-17 day hours result in ~3-6 million data samples per month, depending on the month. From the 11,243 locations, 80% (≈ 9,000) are used as training locations, while the rest are kept for the test set (TST). The measurements from the training locations are split into two random subsets, giving the training set (TRN) and validation set (VAL) with proportions of 80% and 20% respectively. We normalize each variable to zero mean and unit variance as requested by the algorithm structure [14].

4. Results and Discussion

4.1. Model structure selection

Model structure selection is concerned with finding appropriate hyper-parameters for the machine learning model in order to maximize the generalization performance of the model and to reduce overfitting from data. The two main hyper-parameters for our ELMs ensemble are the number of hidden nodes (L) and the size of the ensemble (M). While the number of nodes significantly impacts the generalization performance of the algorithm, an increased ensemble size improves the stability of the algorithm and hence M is mainly limited by computational time.

The hyper-parameters are tuned by minimizing the mean-squared error (MSE) calculated over the validation dataset. A k-fold cross-validation procedure, as suggested in [10], is not applied here as it is beneficial for relatively small data sizes but does not bring significant improvements for large datasets as used here [15]. The tuning of ELM hyper-parameters based on the MSE minimization on VAL dataset is performed by scanning M ranging from 10 to 100 and L from 50 to 10000. Preliminary studies showed that this is a reasonable range, given that higher values come at high computational cost without resulting in significant improvements. Depending on L, the wall clock time needed for training a single model can range from few seconds up to 15-20 minutes, as shown in Fig. 3(a), on a GPU accelerated machine using one K40 NVIDIA card. The optimal number of M and L is hence a trade-off between the computational time required by the model and the exhibited performance on the VAL sample.

Results of the model structure selection for the month of June are reported in Fig. 3(a) and Fig. 3(b). The computational time is dominated by the evaluation of the chosen model on the dense spatial grid used for the GHI predictions (PRED). The MSE for VAL, representing the model performance, steeply falls before 500 nodes and then stabilizes on low values without significant improvement when further increasing L, suggesting a minimum model size of L = 500. Considering also the computational time, which constantly increases with L, we choose the model hyper-parameters as L = 2000 and M = 30, as a trade-off between performance and computational efficiency. The model performance on TST shows a-posteriori that this choice is appropriate.

Table 1. Mean squared error (MSE) for data modelling (L = *2000*, M = *30*).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Satellite data	0.097	0.067	0.073	0.057	0.057	0.045	0.064	0.049	0.124	0.105	0.106	0.057
VAL	0.100	0.067	0.070	0.056	0.057	0.042	0.060	0.049	0.117	0.096	0.103	0.058
TST	0.103	0.068	0.070	0.056	0.057	0.043	0.061	0.050	0.118	0.096	0.104	0.061

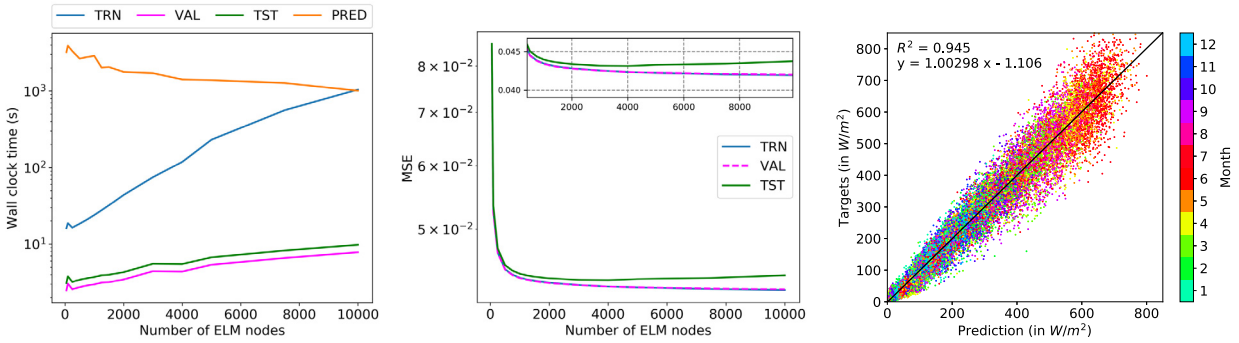

Fig. 3. (a) Wall clock time per model, (b) data MSE as a function of number of nodes, (c) accuracy of ELM model per month. Accuracy is measured from the test data (TST) and is shown together with the linear regression curve fitted on targets and predicted values.

In Table 1, we report the model performance in the form of MSE for each month and we compare it with the mean squared difference between the raw satellite data points and the corresponding average GHI per hour of each month. The test errors are very close to the satellite ones, which indicates that the ELM ensemble is appropriate for modelling the spatial and temporal patterns of monthly GHI. The high accuracy is also shown by the linear relationship between predicted values and targets, shown in Fig. 3(c). The goodness-of-fit is confirmed by the R^2 value which is close to unity. The observed spread represents the intrinsic data uncertainty which we measure using the residuals. For modelling these residuals, an analysis equivalent to the one presented above is conducted, giving optimal model hyper-parameters of $L = 500$ and $M = 50$.

4.2. Prediction and uncertainties

After defining the model parameters and training the model on the satellite data, we predict the monthly-mean-hourly GHI on a dense grid over Switzerland and estimate the model uncertainty from the ELM ensemble variance and the data uncertainty from the second ELM ensemble trained on the remaining residuals. Table 2 shows the monthly mean predictions and satellite data, as well as model and data uncertainty. All values were summed over the respective time span (month or year) and averaged across all locations. We observe that the predicted monthly mean values are slightly above the satellite data. However, the total difference amounts to 0.2% of yearly predicted GHI, which is negligible compared to the yearly model uncertainty of 0.94% of GHI. Overall, data uncertainty dominates with a total of 14.45% of GHI. Figure 4 shows the high-resolution spatial prediction and the spatial distribution of the uncertainties, summed to yearly values. Note that the scale for the prediction is equivalent to Fig. 2(a). The spatial patterns follow the ones observed in the satellite data, but with much higher precision. In the low-altitude regions of Switzerland, spanning from the west to the north-east of the country, the total potential is low, and so is the model uncertainty. In the high-altitude regions of the Swiss alps with high predicted irradiance, as well as near the borders we can observe a higher model uncertainty, with some peaks at the summits of high mountains. These peaks may be due to spatial extrapolation and a lack of data at these altitudes, as the satellite data is the mean over a pixel. The data uncertainty shows some correlation with the altitude profile of Switzerland and is largest in the south-western part of the country. This can be an indication that the weather in this region is least predictable.

Table 2. Monthly mean solar irradiation (in kWh/m^2) and monthly mean estimation of uncertainties (in % of prediction values).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Satellite data (kWh/m^2)	44.8	66.0	113.8	144.8	167.6	180.1	182.3	150.5	115.0	76.90	45.40	36.50	1323.7
Prediction (kWh/m^2)	44.9	66.2	114.2	145.0	167.8	180.3	182.4	150.6	115.2	77.10	45.60	36.60	1325.9
Relative $\widehat{\sigma}_M$ (%)	1.20	0.96	0.78	0.84	0.86	0.91	1.03	0.91	1.05	1.00	0.95	1.03	0.94
Relative $\widehat{\sigma}_D$ (%)	16.19	14.70	15.41	13.50	13.82	11.22	13.47	13.27	18.85	18.13	18.89	14.05	14.45

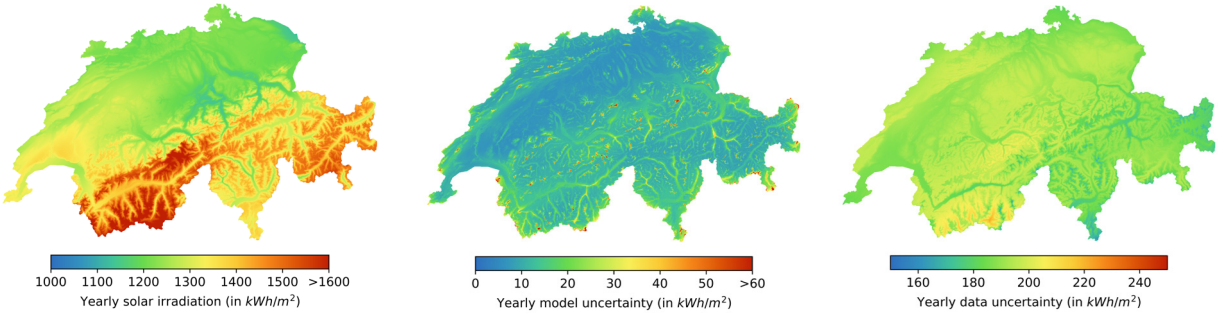

Fig. 4. Maps of the yearly sum of (a) predicted irradiation, (b) model uncertainty, and (c) data uncertainty. Model uncertainty sums to ∼1% of the predicted irradiation, while data uncertainty is close to 15% of the yearly irradiation.

5. Conclusions

We develop a model to forecast the global solar horizontal irradiance at hourly time granularity and high spatial resolution (250×250) m^2 over Switzerland, using an ensemble of Extreme Learning Machines. Additionally, we propose a method, based on variance estimation, to compute the uncertainty associated to the predicted values, able to distinguish between different sources of uncertainty. We provide model uncertainty for each location and timestamp, being on average 0.9% of the predicted values. Likewise, we quantify the statistical noise from data to be on average 14.4% of the predictions. Combining together model and data uncertainties gives a preliminary assessment of the uncertainty associated to the output of our predictive model. Our methodology can be applied to other environmental variables (e.g. diffuse and direct irradiance) to assess the potential for an efficient integration of solar PVs into the built environment. Future work also includes the comparison of this method with other predictive models (e.g. Random Forest) as well as further research into handling, comparing and visualizing uncertainty.

References

[1] Klauser D. Solarpotentialanalyse für Sonnendach.ch. 2016.
[2] Alsamamra H, Ruiz-Arias JA, Pozo-Vázquez D, Tovar-Pescador J. A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern Spain. Agric For Meteorol 2009;149:1343–57. doi:10.1016/j.agrformet.2009.03.005.
[3] Assouline D, Mohajeri N, Scartezzini J-L. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. Sol Energy 2017;141:278–96. doi:10.1016/j.solener.2016.11.045.
[4] Assouline D, Mohajeri N, Scartezzini J-L. Large-scale rooftop solar photovoltaic technical potential estimation using Random Forests. Appl Energy 2018;217:189–211. doi:10.1016/j.apenergy.2018.02.118.
[5] Hocaoğlu FO, Gerek ÖN, Kurban M. Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks. Sol Energy 2008;82:714–26. doi:10.1016/j.solener.2008.02.003.
[6] Şahin M, Kaya Y, Uyar M, Yıldırım S. Application of extreme learning machine for estimating solar radiation from satellite data. Int J Energy Res n.d.;38:205–12. doi:10.1002/er.3030.
[7] Kanevski M, Pozdnoukhov A, Timonin V. Machine learning for spatial environmental data theory, applications and software. Lausanne, Switzerland; Boca Raton, Fla.: EPFL Press ; Distributed by CRC Press; 2009.
[8] Zhang J, Zhao L, Deng S, Xu W, Zhang Y. A critical review of the models used to estimate solar radiation. Renew Sustain Energy Rev 2017;70:314–29. doi:10.1016/j.rser.2016.11.124.
[9] Wan C, Xu Z, Pinson P, Dong ZY, Wong KP. Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine. IEEE Trans Power Syst 2014;29:1033–44. doi:10.1109/TPWRS.2013.2287871.
[10] Leuenberger M, Kanevski M. Extreme Learning Machines for spatial environmental data. Comput Geosci 2015;85:64–73. doi:10.1016/j.cageo.2015.06.020.
[11] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: Theory and applications. Neurocomputing 2006;70:489–501. doi:10.1016/j.neucom.2005.12.126.
[12] Huang G, Huang G-B, Song S, You K. Trends in extreme learning machines: A review. Neural Netw 2015;61:32–48. doi:10.1016/j.neunet.2014.10.001.
[13] Breiman L. Bagging predictors. Mach Learn 1996;24:123–40. doi:10.1007/BF00058655.
[14] Akusok A, Bjork K-M, Miche Y, Lendasse A. High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications. IEEE Access 2015;3:1011–25. doi:10.1109/ACCESS.2015.2450498.
[15] Heskes T. Practical Confidence and Prediction Intervals. Adv. Neural Inf. Process. Syst. 9, MIT press; 1997, p. 176–182.