A MACHINE LEARNING METHODOLOGY FOR ESTIMATING ROOF-TOP PHOTOVOLTAIC SOLAR ENERGY POTENTIAL IN SWITZERLAND

D. Assouline[1], N. Mohajeri[1]; J.-L. Scartezzini[1]

[1]*Solar Energy and Building Physics Laboratory (LESO-PB), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland*

ABSTRACT

Solar photovoltaic (PV) deployment on existing building roof-tops has proven to be one of the most viable large scale resources of sustainable energy for urban areas. While there have been many studies on roof-top integrated PV systems at building and neighborhood scale, estimating the PV energy potential through the available roof surface area in large scale, however, remains a challenge. This study proposes a methodology to estimate the roof-top solar PV potential for existing buildings at the commune level (the smallest administrative division) in Switzerland. In addition, while several studies suggest physical models to assess the photovoltaic solar energy potential, the current study proposes a computational data-based learning approach, using the following four steps: (1) monthly estimation of the main solar irradiance components (global horizontal, direct normal, diffuse horizontal) for the entire Switzerland, that is, the total amount of solar energy available from the Sun (2) processing of training and testing population and building data at a commune level, (3) estimation of the available roof-top surface (A_R), average roof tilted angle (β), and average shading coefficient (S) for buildings in the urban areas, for each commune, (4) combination of the available roof-top surface area for each commune with the solar potential along with other parameters so as to estimate the actual solar PV potential. The first step is achieved through a Support Vector Regression (SVR), using solar irradiance satellite data with an average RMSE of 1.68 W/m^2. In the second step, GIS have been used to aggregate the data at a commune level. In the third step, a supervised learning algorithm using SVR has been used so as to estimate A_R, β and S in existing buildings, which can be extremely difficult data to obtain at a large scale. We use (i) population and building density data, (ii) land use data, and (iii) building typologies as input features. Aggregated values for A_R, β and S for 46 communes in Geneva canton are used as a training output (labeled data) for the learning process. Finally, in the fourth step, by combining solar irradiance and building parameters predictions, we estimate the solar PV potential at the national scale.

Keywords: Rooftop Photovoltaics, Machine Learning, large scale solar potential, GIS, Support Vector Regression

INTRODUCTION

To lead to a better use and a more efficient management of renewable energies in urban environments, where the energy demand is often concentrated, the assessment of local energy potential is required. While many renewable resources are considered, one of them stands out for its almost unlimited power and everlasting presence: the sun. Consequently, solar PV panels on existing building roof-tops have proven to be an efficient and viable large scale resource of sustainable energy for urban areas. To estimate the roof-top solar potential, an assessment of the available solar energy coming to these urban areas as well as the surface area available on buildings (facades and roofs) for solar power plants is required. These assessments have been addressed at multiples scales, with various methods and using different parameters (e.g. building and population density, shade effect, number and orientation of buildings). However, very few studies explore the large scale solar PV potential. For example, study have been done in Spain at municipality level [1] and in Canada-Ontario at regional level [2].

This paper presents, for the first time, a machine learning methodology to estimate the rooftop PV potential of urban areas at large scale. A valuable characteristic of the method is its scalability, since it could be applicable from communal or regional to continental scales, depending on the level of aggregation of the available data. The methodology applies to Switzerland and estimates the rooftop PV potential of each commune, as the smallest Swiss administrative entity. It captures only urban areas defined by the CORINE Land Cover database (including residential, industrial and commercial buildings) enhanced for Switzerland [3]. This help us to capture accurately building density within the urban areas as one of the input variables. The data have been processed using a Geographical Information System (GIS); further processing including machine learning procedures have been performed in Python. All GIS data have been projected and processed in CH1903-LV03, the Swiss GIS coordinate system.

METHODOLOGY

There are several approaches to estimate the solar potential on rooftops. One method has been developed based on a hierarchical approach, as presented in [1]. It is composed of three main potential steps: (i) the physical potential is the solar power available, that is, the solar irradiance coming to the zones of interest; (ii) the geographical potential is the portion of the physical potential captured over the restricted area, more specifically here the available area for PV installation on roofs; and (iii) the technical potential is the actual electricity generated by the PV panel (considering its efficiency, varying between 10% and 40%) based on the geographical potential. Considering the legal issues and socio-economics aspects can be added as another step, that is, the societal potential.

To achieve the potential steps, we adopted a data-based learning approach, known as Machine Learning (ML). In ML, a certain amount of data is given to the machine so that it learns how this type of data is structured, and build directly models from data instead of calibrating pre-built models with the data. This help us to predict similar unknown data. Supervised learning (SL), a machine learning task, is used to derive a model from labeled training data. Here we are interested in finding a function f that relates an input to an output vector, $\mathbf{x}$ and $\mathbf{y}$, so that $\mathbf{y} = f(\mathbf{x})$, given a training data of observed - known - points (x_i, y_i).

One of the most popular SL algorithms is the Support Vector Machine algorithm [4]. It is often superior to classic interpolation since it can handle higher dimension data, often offers a lower computation time and a lower RMSE (Root Mean Square Error). Support Vector Regression, the SVM regression algorithm, has been used in this study. When using SVR, 75% of the known data is used as the training set , and 25% as the testing set (to test models built in the training set), for which the test RMSE is computed. K-fold cross validation is performed on the training set to tune the hyperparameters. The next following parts will present the application of the first two potential steps to our case, the technical and societal steps being not addressed here.

APPLICATION TO SWITZERLAND DATA

Physical potential (solar irradiance maps)

Solar monthly maps have been derived for the solar global horizontal (GHI), and horizontal diffuse and normal direct irradiances (respectively DHI and DNI), in order to account for the tilted angle of the roofs in our calculation. Hourly values of DNI computed on a virtual moving plane following the sun trajectory perpendicularly have been used. In addition, hourly GHI and DHI values from the HC3 (HelioClim-3) database have been used. These values have been extracted from 100 different sites in Switzerland available in the SoDa web service [5]. The data span from February 2004 to May 2015. Monthly values were extracted from the hourly series

by averaging the irradiances over 24 hours for each day of the month - taking the null values of the night into account. This helps us to finally obtain an average value of the solar irradiance received (in W/m^2) per month.
In order to test the efficiency of various algorithms, a yearly GHI map was computed, using SVR and several classical spatial interpolation techniques. These include Inverse Distance Weighting (IDW), Radial Basis Functions (RBF), General Regression Neural Networks 2D (GRNN) and Simple Kriging (SK). One of the SVR derived map is showed in Figure 1. Table 1 presents the RMSE for each technique, showing the best results with SVR.

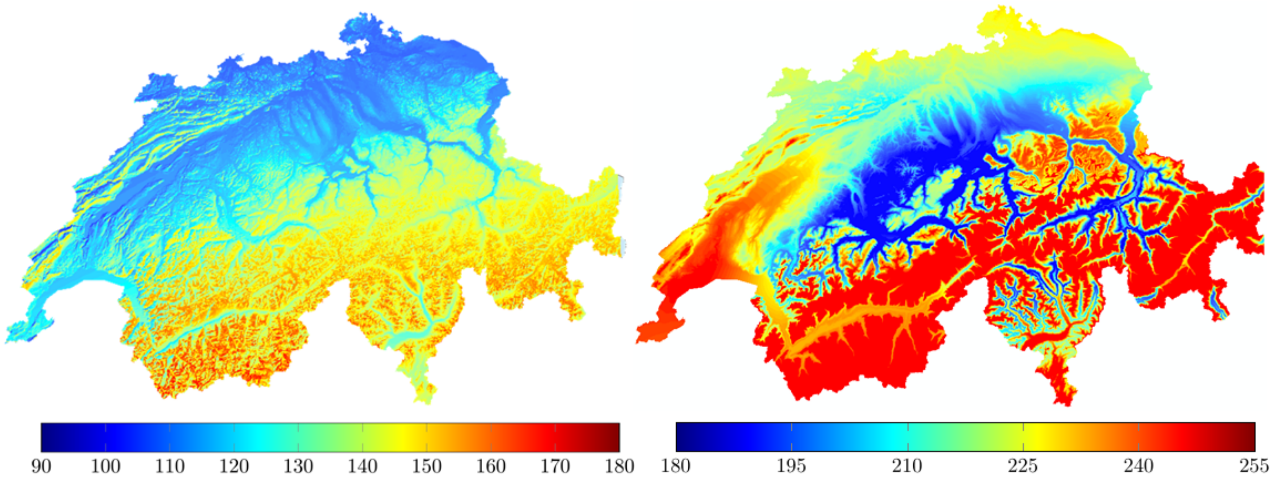

Figure 1: Left: Solar annual GHI map of Switzerland, in W/m^2, obtained with SVR, using spatial coordinates and terrain parameters derived from DEM's. Right: Solar DNI map for July, in W/m^2, using spatial coordinates and weather parameters.

	IDW	RBF	GRNN2D	SK	SVR2D	SVR3D	SVRTerrain
Test RMSE (W/m^2)	2.65	2.10	2.42	1.85	1.37	1.65	1.72

Table 1: RMSE summary of interpolation techniques for deriving a swiss GHI solar irradiance map.

Since the SVR algorithm is superior to the other 2D interpolating methods, it has been chosen for monthly mappings of DNI, DHI and GHI. An example of DNI map is showed in Figure 1. Latitude, longitude, altitude and weather parameters are chosen as input features. Temperature, precipitation, sunshine duration, and cloud cover data were gathered to account for the weather behavior, tremendously impacting on solar prediction. The raster basis for the prediction maps was a Digital Elevation Model (DEM) that covers all Switzerland (RIMINI DEM available in the Swisstopo website, available at http://www.swisstopo.admin.ch) and re-sampled to a resolution of 200m $\times$ 200m. The global tilted irradiance (GTI) on a non horizontal surface (i.e. a tilted roof) is given by the classical following equations:

$$GTI = R_d D_h + B_n \cos(\theta) + \rho R_r G_h \tag{1}$$

$$\cos(\theta) = \sin(\theta_z)\sin(\beta)\sin(\gamma_s - \gamma) + \cos(\theta_z)\sin(\beta) \tag{2}$$

Where D_h, B_n and G_h are respectively DHI, DNI, and GHI, θ is the angle of incidence of the sun rays on the tilted plane, R_d and R_r are diffuse and reflected factors (given by physical isotropic or anisotropic models), ρ is the foreground's albedo (taken here as 0.2), and θ_z , γ_s, β, and γ are respectively the sun zenith angle, sun azimuth angle, tilt angle of the surface, and the azimuth angle of the surface. Thus, all these parameters require to be aggregated at the communal level. The azimuth of a tilted surface is defined as the angle between the North and the horizontal projection of normal vector to the plane. To obtain an azimuth value for the average building in each commune, we made the following assumptions: (i) the footprint is rectangular and symmetric, (ii) the roof consists in two tilted surfaces with the same slope, and the A_R is equally shared between the two sides of the roof, and (iii) the roofs are likely to follow the direction of the longest walls, meaning that the azimuths of one of the two roof tilted surface are given by the smallest walls direction which is perpendicular to them. The rooftop azimuths are aggregated in each commune considering the main direction - most frequent azimuth from a histogram with 20 degree bin width. We computed monthly raster maps for the suns altitude and azimuth with SVR, based on hourly data for various points in Switzerland, considering again the suggested days per month in [6], at 1:00 PM. Raster maps offer much more resolution than the aggregated maps that only offer one value per commune, thus we decided to clip these sun position raster values within the CORINE urban areas, and average the clipped cells within these areas. Slope estimation, needed to obtain the GTI, will be addressed in the next part.

Available communal roof area, average slope and shading estimation

Precise data on various parameters of buildings are very rarely available. Some of these parameters directly influence the solar potential and are necessary for a geographical potential study: for example, the available area for PV installation on rooftops (A_R), the average slope of the building rooftops (the slanting angle of the roof from a horizontal plane), and the shading effect caused by surrounding buildings. While most studies arbitrarily choose correction coefficients to account for the first and the third parameter, we evaluated these coefficients based on the building characteristics and properties of the urban areas. We use machine learning methods to estimate an aggregated value of these unknown buildings parameters in every Swiss commune by extrapolating them from the Geneva canton, where accurate data is available through SITG open access website (http://ge.ch/sitg/). This database contains available roof surface for PV panels and the average slope for each building in the canton. The shading factor is extracted for each commune in Geneva from a DOM (Digital Orthophoto Map), a DEM that takes buildings and vegetation into account. This computes the hillshade raster map from the DOM with the Spatial Analyst toolbox in ArcMap. As hillshade depends on the sun altitude and azimuth, we repeated the DOM extraction for each month, by taking into account a monthly sun position, corresponding to typical days per month [6], and calculating the hillshade for this position.

The portions of the hillshade corresponding to the building rooftops were then clipped in order to extract an aggregated shading value within each commune in Geneva. The input features for the SVR algorithm are as follows: (i) building density (D_b), (ii) population density (D_p), (iii) average building footprint area (A_f), (iv) total building footprint area, and (v) various building typologies ratios (e.g. period of construction, number of stories, main energy resource used etc.). D_b and D_p as well as the building areas were extracted from a 2D footprint building data which covers all Switzerland, VEC25 from the Swisstopo website. The building typologies data were taken from the swiss Federal Office of Statistics (OFS). All these features have been aggregated within the urban areas defined by CORINE Land Cover in different communes in order to obtain one value for each Swiss commune. The average roof slope and A_R have been extracted from the SITG database. These values are also aggregated within the urban area along with the shading factor from the DOM in order to present one value for each commune in Geneva. A correction coefficient c_R has then been computed from A_R ($c_R = A_R/A_f$) for each commune in Geneva. Average roof slope, shading coefficient and c_R are the outputs in the ML algorithm in order to predict the same variables for every Swiss commune. SVR have been used for all three variables separately. The RMSE obtained for each of them are presented in Table 2, along with RMSE obtained for GHI, DNI, and DHI predictions. It resulted in one A_R map (calculated as $c_R A_f$ in each commune), one slope map, and 12 monthly shading maps. The slope aggregated values are used to finalize the computation of GTI in each CORINE area.

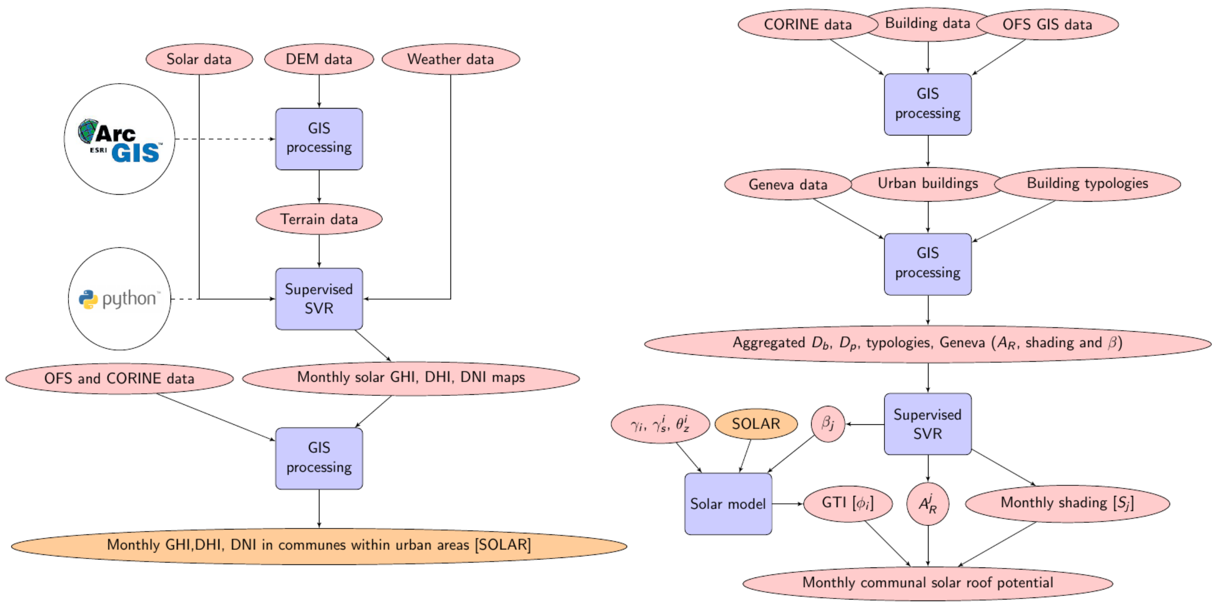

Figure 2: Methodology scheme. Left and right diagrams respectively describe physical and geographical potential steps. A_R^j and β_j are the available rooftop area for PV installation and the roofs average tilt angle in each commune, θ_z^i , γ_s^i, and γ_i are respectively the sun altitude and azimuth angles, and the main azimuth of the roofs in each CORINE area.

	GHI (W/m^2)	**DNI (W/m^2)**	**DHI (W/m^2)**	C_R (0 $\leq$**ratio**$\leq$ 1)	β (°)	**Shade**(0 $\leq$**coeff**$\leq$ 255)
Test RMSE	1.29	2.97	0.76	0.0164	0.54	2.14

Table 2: RMSE summary of various parameters predicted with SVR. In case of monthly derived parameters, the given value is the average of the monthly RMSE's.

Geographical potential (multi-layer approach)

The geographical potential has been obtained from the combination of aggregated GTI values with the previously presented predicted commune aggregated maps. A scheme of the complete methodology is showed in Figure 2.

We computed the solar potentials of the CORINE urban areas of a commune, and summed them to obtain the total commune potential. The values that are are already aggregated by the prediction (for example A_R) are taken the same for each urban area within the commune. For each month, the total available solar power on rooftops P_j for each commune j is computed as follows:

$$P_j = \sum_{i=1}^{N} b_i S_j \left(\frac{A_R^j}{2} \varphi_i\left(\gamma_i\right) + \frac{A_R^j}{2} \varphi_i\left(\gamma_i + 180\right) \right) \tag{3}$$

Where the sum is computed over the CORINE urban areas i, b_i is the number of buildings in area i, S_j is the shading factor predicted in j, A_f^j is the average available rooftop area in j, φ_i is the GTI clipped in area i, given by equation (1), as a function of γ_i, the predicted roof azimuth of CORINE urban areas i. We consider both γ_i and $(\gamma_i + 180)$ to account for both sides of the roof. It should also be noted that we used the Klucher model [7] to compute R_d and the classical isotropic sky model to compute R_r in equation (1). This is computed for every considered commune, for each month, resulting in twelve final potential maps. An example of the resulted maps is showed in Figure 3 for June, along with the obtained map for c_R. The general PV potential can be clearly seen from each commune, yet, it is highly dependent on the number of buildings. As a result, values have also been divided by CORINE and total footprint areas, and inhabitants of the commune, to obtain normalized maps.

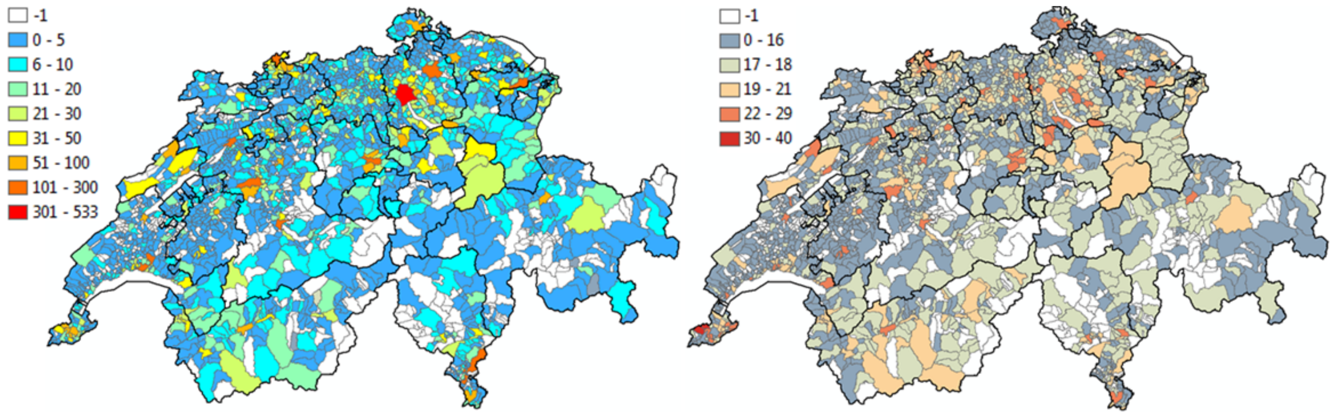

Figure 3: Left: Solar PV Potential map for June, in MW. Right: c_R map (portion of footprint area available on rooftop), in percentages (-1 is for unconsidered communes).

CONCLUSION

A Machine Learning methodology coupled with ArcGIS has been used to estimate rooftop solar PV potentials for 1895 communes in Switzerland. The ML methodology is very useful for considering a significant range of parameters when estimating solar irradiances and additional characteristics with reasonable errors. To the best of our knowledge, such a scale for estimating solar PV potentials has not been considered before for Switzerland. Given the level of aggregation (commune) and the the amount of data available, several assumptions and approximations had to be made. There are several limitations regarding the data and the methods of estimation which needs to be addressed in the future. Some of these limitation are as follows: (i) the inconstistency between different data sources for communes, buildings and population, resulting in several communes not considered, (ii) the lack of precise data for other cantons in Switzerland apart from Geneva, which causes the question of generalization to all other communes in Switzerland; (iii) here only the urban areas defined by CORINE Land Cover are considered, meaning that buildings in the rural areas might be missing in this study. While the aggregation results at commune level are very useful and have implications for planning and policy making, estimating the rooftop solar potentials at the neighborhood level is also necessary. In the future we plan to extend these results to the pixel level so as to make them more accurate.

ACKNOWLEDGEMENTS

This work is part of the SCCER Future Energy Efficient Buildings and Districts (FEEB&D) supported by the Commission for Technology and Innovation (CTI).

REFERENCES

1. Izquierdo, S., Rodrigues, M., and Fueyo, N.: A method for estimating the geographical distribution of the available roof surface area for large-scale photovoltaic energy-potential evaluations. *Solar Energy*, 82(10):929–939, Oct. 2008.
2. Wiginton, L., Nguyen, H., and Pearce, J.: Quantifying rooftop solar photovoltaic potential for regional renewable energy policy. *Computers, Environment and Urban Systems*, 34(4):345–357, 2010.
3. Steinmeier, C.: Corine land cover 2000/2006. 2013.
4. Cortes, C. and Vapnik, V.: Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
5. Ratto, P. T., Reise, C., Remund, C., Wald, L., Albuisson, M., Best, C., Delamare, C., Gaboardi, E., Hammer, A., Heinemann, D., Kift, R., Lefvre, M., Leroy, S., Martinoli, M., Mnard, L., Goot, E. V. D., Vanroy, F., and Webb, A. Soda: a web service on solar radiation.
6. Klein, S.: Calculation of monthly average insolation on tilted surfaces. *Solar energy*, 19(4):325–329, 1977.
7. Klucher, T. M.: Evaluation of models to predict insolation on tilted surfaces. *Solar energy*, 23(2):111–114, 1979.