Solar Energy 141 (2017) 278–296

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Quantifying rooftop photovoltaic solar energy potential: A machine learning approach

Dan Assouline *, Nahid Mohajeri, Jean-Louis Scartezzini

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

ARTICLE INFO

Article history:
Received 13 June 2016
Received in revised form 24 October 2016
Accepted 25 November 2016
Available online 16 December 2016

Keywords:
Machine learning
Support vector regression
Solar photovoltaics
Large-scale solar potential
Geographic information systems

ABSTRACT

The need for reduction in CO_2 emissions to mitigate global warming has resulted in increasing use of renewable energy sources. In urban areas, solar photovoltaic (PV) deployment on existing rooftops is one of the most viable sustainable energy resources. The present study uses a combination of support vector machines (SVMs) and geographic information systems (GIS) to estimate the rooftop solar PV potential for the urban areas at the commune level (the smallest administrative division) in Switzerland. The rooftop solar PV potential for a total 1901 out of 2477 communes in Switzerland is estimated. Following a 6-fold cross validation, the root-mean-square error (also normalized) is used to estimate the accuracy of the different SVM models. The results show that, on average, 81% of the total ground floor area of each building corresponds to the available roof area for the PV installation. Also considering the total available roof area for PV installation, that is, 328 km^2 and roof orientations within ±90° of due south, the annual potential PV electricity production for the urban areas in Switzerland is estimated at 17.86 TW h (assumed 17% efficiency and 80% performance ratio). This amount corresponds to 28% of Switzerland's electricity consumption in 2015.

1. Introduction

The Swiss Federal Council's "Energy Strategy 2050", initiated in 2011 partly as a consequence of the Fukushima nuclear accident, proposes a phasing-out of nuclear energy by the year 2035, currently generating 40% of the national electricity demand (http://www.bfe.admin.ch/). To compensate for the loss of nuclear energy, the federal Council's Energy Strategy anticipates not only the improvement of energy efficiency, but also the increase in the use of renewable energy and associated development of grid and storage capacity. In addition, the Swiss climate policy aims at a drastic reduction of the country's GHG emissions, including 20–30% reduction of the country's CO_2 emissions from the 1990 level by the year 2020, according to the revised federal CO_2 Act, and a possible 50–80% reduction by 2050. Buildings have the largest share in energy demand in Switzerland: heating, ventilation, and air conditioning account for roughly 40% of the overall energy demand; 32% of the national electricity demand is also caused by buildings (HVAC, lighting, space heating). Therefore, the goals of the "Energy Strategy 2050" and the Swiss climate strategy can only meet when buildings become much more energy efficient compared to today's situation. To reach these goals, the remaining demand must primarily be met by renewable energies.

Solar photovoltaic (PV) panels on the existing building rooftops have proven to be an efficient and viable large scale resource of sustainable energy for urban areas (Wittmann et al., 1997; International Energy Agency IEA, 2002; Izquierdo et al., 2008; Wiginton et al., 2010; Hernandez et al., 2015; Yuan et al., 2016). In addition, solar panels can have an important role in integration of decentralised renewable energy resources in a neighbourhood (Mavromatidis et al., 2015; Wegertseder et al., 2016). In Switzerland, only 1.2% electricity generated in 2014 comes from PV (Kemmler et al., 2015) but is likely to increase to 13.4% by 2050 (Berg and Real, 2006). The feasibility of solar PV installations is of importance not only for individual property owners, but also the local governments and municipalities (Strzalka et al., 2012).

Depending on the availability of data, regional characteristics, as well as scale of study, several methodologies have been suggested to determine rooftop PV potential (e.g. Byrne et al., 2014; Yuan et al., 2016; Ramirez Camargo et al., 2015). At the scale of Europe as a whole (27 EU members), studies show that there is a large building–integrated photovoltaics (BIPV) potential, 840 TW h annually, which equals more than 22% of the expected European electricity demand by 2030 (Šúri et al., 2007; Defaix et al., 2012; Byrne et al., 2014). At national and regional scales,

* Corresponding author.
E-mail address: dan.assouline@epfl.ch (D. Assouline).

http://dx.doi.org/10.1016/j.solener.2016.11.045
0038-092X/

studies show significant values for urban (rooftop) PV applications in many countries. These include the USA (Denholm and Margolis, 2008; Paidipati et al., 2008), Israel (Vardimon, 2011), Canada (Pelland and Poissant, 2006), and Spain (Ordonez et al., 2010), where urban PV deployment could potentially cover 15–45% of national electricity consumption. At regional scale, Lopez et al. (2012) provide a GIS-based methodology for all the states of USA and their technical rooftop potential. Several studies explore the PV potential for buildings at the city and neighbourhood scale (e.g. Hofierka and Kaňuk, 2009; Bergamasco and Asinari, 2011; Strzalka et al., 2012; Peng and Lu, 2013; Byrne et al., 2014). For Hong Kong (Peng and Lu, 2013), as an example, the estimated potential of rooftop PV is 5981 GW h which can account for 14.2% of the city's 2011 electricity use. Another example is Seoul in South Korea where deployment of rooftop distributed photovoltaic systems can cover 30% of the city's annual electricity consumption.

Several studies propose a hierarchical approach for estimating the rooftop potential of PV on a national scale (Hoogwijk, 2004; Izquierdo et al., 2008; Ordóñez et al., 2010; Wiginton et al., 2010; Schallenberg-Rodríguez, 2013). For example, Izquierdo et al. (2008) use a sampling technique to estimate the available roof surface. Wiginton et al. (2010) use a GIS-based Feature Analyst (FA) tool to compute the overall rooftop area. Then a sampling technique with additional variables is used to explore the relation between population density and PV potential. The International Energy Agency (IEA, 2002) use statistical information to estimate the building area (roofs and facades) and to obtain the potential for solar energy. Several other studies use aerial images (Suzuki et al., 2007) and ArcGIS together with LiDAR (Light Detection and Ranging) data to determine roof geometries and associated roof areas for the PV potential (Tarsha-Kurdi et al., 2007; Carneiro et al., 2009, 2010; Brito et al., 2012; Lukac et al., 2014; Gooding et al., 2015; Verso et al., 2015). Advanced cartographic information and high-resolution images derived from remote sensing technologies such as LiDAR is expected to produce high accuracy results for PV potential (Schallenberg-Rodríguez, 2013; Byrne et al., 2014). Recently, machine learning algorithms (e.g. support vector machines, artificial neural networks) have been extensively used for predicting solar radiation on horizontal and tilted surfaces (Bouzerdoum et al., 2013; Ramedani et al., 2014; Yadav and Chandel, 2014; Ramli et al., 2015; Lauret et al., 2015) as well as for geospatial environmental data modelling (Kanevski and Maignan, 2004; Kanevski et al., 2009). However, using machine learning algorithms so as to estimate urban characteristics for solar prediction on building roofs, including the available roof area and roof geometries, has not been previously explored.

This paper uses Support Vector Machines (SVMs), a kernel-based machine learning technique, to estimate the solar PV potential of building rooftops in the greater part of the urban areas of Switzerland. The data is aggregated at the commune level (the smallest administrative division) in Switzerland. Using a combination of machine learning and GIS methods, the main aims of this study are as follows: (1) to estimate the monthly global horizontal solar radiation and global tilted solar radiation on tilted roofs for the entire Switzerland, (2) to estimate the shadowing effects on the roofs using LiDAR data, (3) to estimate roof slopes and roof aspects, as well as the available roof surface area, for PV installation in the urban areas, and (4) to estimate the technical potential of rooftop PV electricity production in Switzerland.

2. Methodology

We apply a hierarchical methodology which has been used in several studies (Izquierdo et al., 2008; Wiginton et al., 2010) to estimate the rooftop solar PV potential in Switzerland. The hierarchical methodology consists of three steps: (i) The physical potential, which is based on the assessment of the total energy received from the sun by the urban areas, (ii) the geographic or urban potential, which reflects the constraints on the locations where the solar energy can be captured and used for PV installations, and (iii) the technical potential, which relates to the transformation of the solar energy received by the available roof area into electrical energy using the technical characteristics of the PV technology (e.g. efficiency and the performance). For a complete assessment of the solar energy available on rooftops, however, the social potential and the economic potential must be considered (Branker et al., 2011; Bazilian et al., 2013; Lang et al., 2015; Luka et al., 2016). The social and economic potentials are beyond the scope of this study and will be considered in a future study. A combination of machine learning and geographic information systems have been used in order to estimate various variables in the three levels mentioned above and, finally, to determine the total potential for rooftop solar PV electricity production.

2.1. Support vector regression

Support vector machines (SVMs) are a set of related supervised learning algorithms that were initially introduced by Vladimir Vapnik in the middle of the 1990s (Vapnik, 1995). While the classical methods of statistical learning are based on the minimisation of the error in the training data (empirical risk), the main advantages of SVMs are to offer a good generalisation performance by minimising both the training error and the confidence interval (Vapnik, 1995, 1998; Smola and Schölkopf, 2004). SVMs are kernel-based machine learning techniques and usually used in both classification and regression problems (Scholkopf and Smola, 2002).

The basic idea of a support vector regression (SVR) is as follows (Vapnik, 1995; Scholkopf and Smola, 2002; Smola and Schölkopf, 2004; Kelleher et al., 2015): Given a training data $\{(\boldsymbol{x}_i, y_i)\}$ ($i = 1, 2, \ldots, N$), where $\boldsymbol{x}_i \in R^d$ is the input vector of dimension d (the number of features) and $y_i \in R$ is the desired output value (called the label) for point i, the goal of a regression model is to develop a function f so that $f(x_i) \approx y_i$ for all $1 \leqslant i \leqslant N$ in the training data, and to use f for further predictions. The SVR algorithm approximates a linear function in the form:

$$f(x) = \langle w, x \rangle + b \tag{1}$$

where w is defined as the weight vector, that is, an unknown vector in R^d to be optimised, x is a vector in R^d (input space), $\langle ., . \rangle$ is the inner product, and b is a constant. Given a positive real number ε, that is, a margin of tolerance, a linear function f is built and has the following characteristics:

- $f(x_i)$ does not deviate from y_i by more than ε. In other words, the errors will be ignored as long as they are less than ε. Any deviation, however, larger than ε will not be accepted.
- f should be as flat as possible, which means that it does not follow the fluctuations of the target values too closely, so as to prevent overfitting. Overfitting happens when the model learns to follow the particular trend of the training data so closely that it will have trouble adapting to a new dataset, and therefore performs poorly for new predictions.

To ensure the flatness of the function f, the Euclidean norm of w, $\|w\|$, has to be minimised, namely:

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}\|w\|^2 \\ \text{subject to} & |y_i - \langle w, x_i \rangle - b| \leqslant \varepsilon, i = 1, 2, \ldots, N \end{array} \tag{2}$$

This strict optimization problem is not always feasible due to the outliers. Outliers are defined as the data points that are beyond ε, the margin of tolerance. Thus, the concept of soft margin is used