A new Adaptive methodology of Global-to-Direct irradiance based on clustering and kernel machines techniques

Martín Gastón, Íñigo Pagola, Carlos M. Fernández-Peruchena, Lourdes Ramírez, Fermin Mallor

A NEW ADAPTIVE METHODOLGY OF GLOBAL-TO-DIRECT IRRADIANCE BASED ON CLUSTERING AND KERNEL MACHINES TECHNIQUES

Martín Gastón[1] and Iñigo Pagola[1] and Carlos Fernández[2] and Lourdes Ramírez[3] and Fermín Mallor[4]

[1] CENER (National Renewable energy Centre) Solar Thermal Energy Department. Address: c/ Ciudad de la Innovación 7, Sarriguren, 31621, Navarra (Spain). Phone: 948 25 28 00. E-mail: mgaston@cener.com

[2] CENER (National Renewable energy Centre) Solar Thermal Energy Department. Address: C/Isaac Newton 4 – Pabellón de Italia, 41092 Sevilla, Spain.

[3] CENER (National Renewable Energy Centre). Address: C/Somera 7-9, 28026 Madrid, Spain.

[4] UPNA. (Public University of Navarre). Statistic and Operational Research Department. Address: Campus Arrosadia 31006, Pamplona, Spain.

Abstract

Nowadays, there is a growing need of data related to beam solar radiation. The market situation of the solar technologies as well as the premium prices in countries like Spain drives to the necessity to characterize the expected DNI in a specific location with short measurement campaign.

Due to it is impossible to wait ten or more years for measure the DNI in a specific location, methodologies for deriving DNI representative series are needed. To estimate beam radiation from global radiation, there can be used information derived from satellite images, or data from well known data bases of global solar radiation. In both cases global-to-beam irradiance models are needed.

This work presents a novelty methodology to transform hourly global irradiance registers in a concrete emplacement, into beam irradiance data. In front of classical global to direct models where general curves are fitting to simulate the direct irradiance behavior, the present methodology allows us to fit the “better curve” to each available observation.

To implement this method, a set of ground measurements of Beam and Global irradiance were available and different mathematical techniques have been used. Concretely, topics of measure theory, cluster methods and kernel machines learning are included in the present methodology. So, this work is presented using public access data which allow any contrast of the methodology results.

Keywords: Global Irradiance, Beam Irradiance, k nearest neighbor, Machine learning, kernel methods, SVM.

1. Introduction

A new methodology to transform global irradiance data into beam irradiance is presented in this work. It is based on machine learning techniques and, therefore, training sets were necessary. The goal is to implement a process which allows us to obtain registers of beam irradiance in periods with only global data were available. Thinking in the existence of local behaviors we only use historical data in the concrete emplacement to fit the models. The proposed methodology predicts each beam irradiance data by mean of two steps. First, it selects the most suitable training subset using a clustering methodology. After that, a prediction model is fitted with this subset to finish generating DNI forecasts.

The next section describes the data used in this work to show the suitability of the proposed methodology. Section 3 collects a description of the methodology joint to a brief explanation of the foundations of the techniques used in the different steps. Section 4 shows the obtained results and the comparison with classical global to direct models ([6] to [10]). To finish, section 5 enumerates conclusions and future researches.

2. Available data

Ground measurements of beam and global irradiance in a set of four meteorological stations are available. These data coming from the Estatal Association of Meteorology of Spain (AEMET) and they have hourly frequency covering time periods up of two years. Only right data have been used, filtering wrong records by mean of typical criterion.

The meteorological stations used in this work are sited covering the Spanish geography but a general methodology is presented. The explained procedure can be implemented to transform global into beam irradiance independently of the interest site location. However, the adaptive procedure allows the inclusion of new real data to improve the accuracy of the model results.

In this methodology we suppose an historical set of ground measurements of global and beam irradiance. The objective magnitude is the direct irradiance and to train respective models the extraterrestrial irradiance will be used joint to ground measurements.

3. Adaptive methodology based on machine learning

The first goal of this methodology is the use of the best model to predict each data. So, in the first step, it searches the most suitable training set among all available data. Once it is done, a nonlinear prediction model is fitted to forecast the new beam irradiance register. In this two stages machine learning techniques are implemented.

Machine learning collects all those algorithms that try to solve problems learning from experience. A typical scheme of machine learning can be seen in figure 1. Problems as classification, clustering, regression or novelty detection can be included in this scheme. In our case we combine two machine learning, one of them to classify the suitable data set of training and other one to predict the beam irradiance.

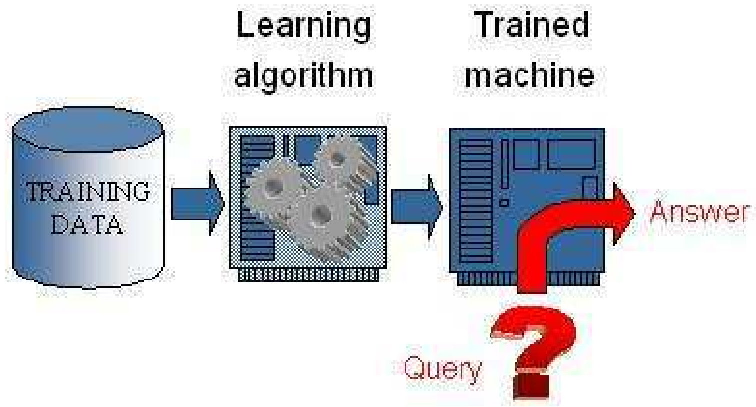

Fig. 1. Machine learning scheme.

3.1. K nearest neighbor

The k-nearest neighbor is amongst the simplest machine learning algorithms [11]. It is a classification algorithm which classifies an object by a majority vote of its neighbors. To implement this algorithm it is necessary a training set and a criteria to determine how to order the data, that is, a distance between objects which arrange them around the given object. In multivariate case, examples of distances are the Euclidean distance, the Manhattan distance or the Hamming distance. The selection of the more suitable distance depends of the nature of the problem. It is usual to design a distant ad-hoc for the currently problem.

In our case we don't want to classify our object, we only search the set of nearest neighbors. Our data have two variables, the measured global irradiance and the difference between global and extraterrestrial. These two dimensional vectors are used to select de training set of the next machine learning. This training set consists of the most similar data, respecting global and difference with extraterrestrial, amongst all the historical data containing ground measures of global and beam irradiance.

3.2 Support Vector Machines

Support vector machines (SVM) are a group of supervised learning methods that can be applied to classification or regression. They represent an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik [4].

In this section we give a brief introduction to SVM. First, we present the linear separable case which contains the original idea of SVM, after that the nonlinear version and the regression algorithm of SVM will be presented.

3.2.1. Linear separable case. Which separating hyperplane to use?

Suppose that we are given data points each of which belongs to one of two classes; the goal is to determine some way of deciding the class of a new data. In the case of support vector machine, we view a data point as a p + 1−dimensional vector, and we want to know if we can separate them with a p−dimensiona hyperplane, see figure 2.

There are many hyperplanes that might classify the data. However, we are interested in finding out if we can achieve maximum separation between the two classes (margin). By this we pick the hyperplane so that the distance from the hyperplane to the nearest data point is maximized. Now, if such hyperplane exists, it is clearly of interest and it is known as a maximum margin classifier.

The hope is that, the larger the margin or distance between these parallel hyperplanes, the better the generalization error of the classifier will be.

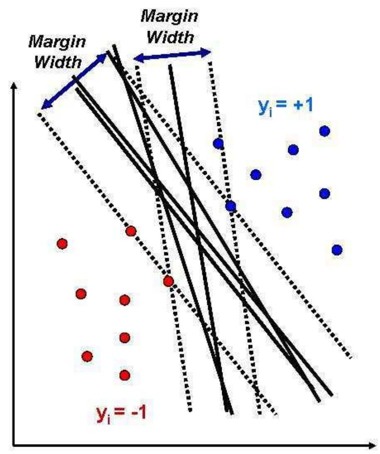

Fig. 2. Two linear separable classes.

3.2.2. Maximizing the margin

We are given some training data. A set of points of the next form: $D=\{(x_i,y_i) \mid x_i \in R^p, y_i \in \{-1,1\}\}$where the y_i is either 1 or −1, indicating the class to which the point belongs. We want to give the maximum-margin hyperplane which divides the points having yi = 1 from those having yi = − 1. Any hyperplane can be written as the set of points satisfying this equation: $<w, x> + b=0$. Rescaling w and b such that the point(s) closest to the hyperplane expiring the following expression: $|<w, x_i> + b|=1$we obtain a canonical form (w, b) of the hyperplane, satisfying this inequality: $y_i (<w, x_i> + b) \geq 1 \ \forall\ i$ (So, we prevent data points falling into the margin). We find that the distance between these two hyperplanes is given with the next quotient: $2/||w||$ and if we put this information together we get the optimization problem which we must solve:

$$\text{Choose } w, b \text{ to minimize } ||w||$$

$$\text{Subject to } y_i (<w, x_i> + b) \geq 1 \ \forall\ i$$

Data points closest to the boundary are called support vectors. Figure 3 shows the graphical explanation of the problem.

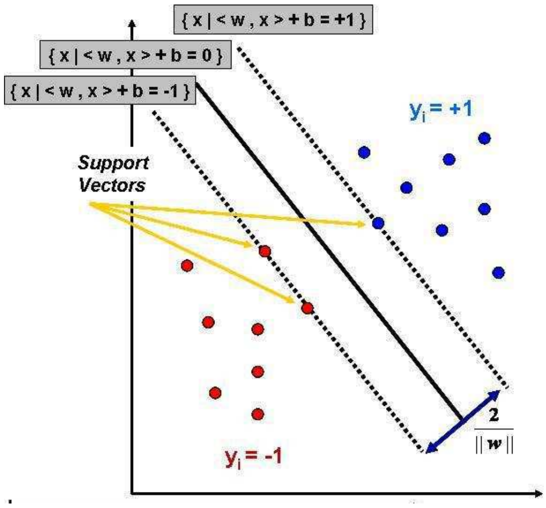

Fig. 3. Problem resolution.

3.2.3. Optimization problem

The optimization problem presented above is hard to optimize. The reason is that, in mathematical terms, it is a non-convex optimization problem which is known to be much more difficult to solve. Fortunately it is possible to alter the equation by substituting $||w||$ with $0.5||w||^2$ without changing the solution. This is a quadratic programming (QP) optimization problem. More clearly,

$$\text{Minimize } 0.5||w||^2$$

$$\text{Subject to } y_i (\langle w, x_i \rangle + b) \geq 1 \ \forall\ i$$

The factor of 0.5 is used for mathematical convenience. This problem can now be solved by standard quadratic programming techniques and algorithms. Writing the classification rule in its unconstrained dual form reveals that the maximum margin hyperplane, and therefore the classification task, is only a function of the support vectors, (this is due to Karush-Kuhn-Tucker condition). The dual of the SVM can be shown with this equation:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle x_i , x_j \rangle$$

$$\text{Subject to} \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \qquad \alpha_i \geq 0 \qquad \forall i$$

Where the α terms constitute a dual representation for the weight vector in terms of the training set:

$$w = \sum_{i=1}^{m} \alpha_i y_i x_i$$

3.2.4. Non-linear Support Vector Classifiers

If SVM were able to handle only linearly separable data, theirs usefulness would be quite limited. In machine learning, the kernel trick [2] is a method for using a linear classifier algorithm to solve a nonlinear problem by mapping the original observations into a higher-dimensional space where training set becomes separable. This is done using Mercer’s theorem which states that if a kernel function is continuous, symmetric and positive semidefinite there exits a Hilbert space defined on X and a feature map such that the kernel function can be expressed as an inner product in a higher-dimensional space.

$$\Phi : X \rightarrow H$$
$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle$$

Wherever an inner product is used, it is replaced with the kernel function.

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$

Subject to $\sum_{i=1}^{m} \alpha_i y_i = 0 \qquad \alpha_i \geq 0 \qquad \forall i$

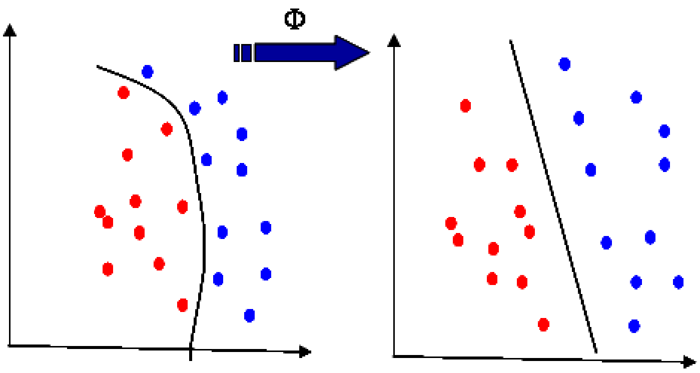

Fig. 4. Nonlinear problem.

3.2.5. Soft Margin

In practice, a separating hyperplane need not exist; and even if it does, it is not always the best solution to the classification problem (this is due to the over-fitting). An algorithm which can tolerate a certain fraction of outliers would be interesting. The Soft Margin method will choose a hyperplane that splits the examples as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples. The method introduces slack variables, ζ_i which measure the degree of misclassification of the datum x_i and uses these relaxed separation constraints: $y_i (<w, x_i> + b) \geq 1 - \zeta_i \ \forall\ i$

The objective function is then increased by a function which penalizes non-zero ζ_i, and the optimization becomes a trade off between a large margin, and a small error penalty.

$$\min \frac{1}{2} \| w \|^2 + C \sum_i \zeta_i$$
$$\text{s.a} \quad y_i (\langle w, x \rangle + b) \geq 1 - \zeta_i$$
$$\zeta_i \geq 0$$

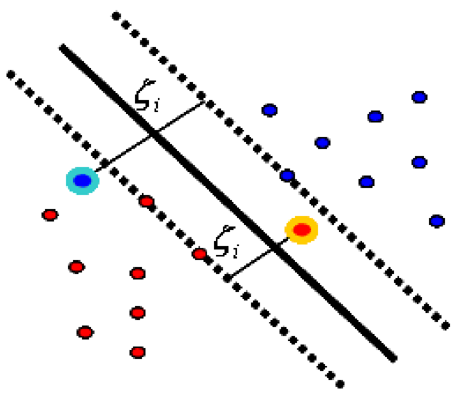

Fig. 5. Soft margin. Slaks variables.

3.2.6. A version of SVM for regression

To generalize the SV algorithm to the regression case, an analogue of the soft margin is constructed in the space of the target values y (note that we now have $y \in R$) by using Vapnik's ε−insensitive loss function. This quantifies the loss incurred by predicting f(x) instead of y as the next function: $|y-f(x)|_{\varepsilon} := \max\{0, |y-f(x)| - \varepsilon\}$

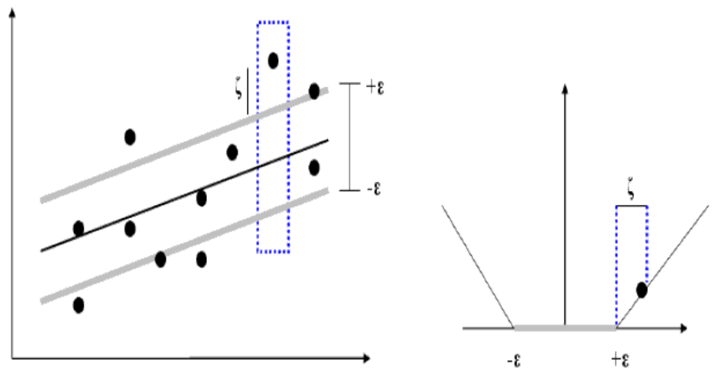

Fig. 6. SVM for regression.

To estimate a regression f(x) = <w, x>+b, one minimizes the next expression $\frac{1}{2}\|w\|^2 + C\sum_i |y - f(x)|_{\varepsilon}$

We can transform this into a constrained optimization problem by introducing slack variables, akin to the soft margin case. The optimization problem is now given by:

$$\min r(w,\zeta) = \frac{1}{2}\|w\|^2 + C\sum_i (\zeta_i + \zeta_i^*)$$

$$\text{s.a} \quad \begin{aligned} & f(x_i) - y_i \leq \varepsilon + \zeta_i \\ & y_i - f(x_i) \leq \varepsilon + \zeta_i^* \\ & \zeta_i^* \geq 0 \quad \zeta_i \geq 0 \end{aligned}$$

4. Global to direct transformation.

In our case, we train a regression SVM by mean of the selected set in the first stage (knn). The objective variable is the DNI and the global one plays the role of explained variable. To avoid the problems caused by the nonlinearity of this relationship Gaussians kernels are used.

Figure 7 shows an example of the first stage of the methodology, all available data are painted in cyan color, the green point is the new real data for which a new beam record we must generate. The selected training data by knn has been represented in red color. The blue point corresponds to the new prediction.

On the other hand, figure 8 collects the respective SVM model: red points are the training set; the green points are the solutions of the fitted model applied to the training data. To finish, pink and blue points are real and predicted data respectively.

Note that this is made ever data that we want generate the direct radiation from global measure. It is easy to observe the nonlinearity of the model and how the using of the similar points respecting to global and difference with extraterrestrial could help in the final prediction. We see a local behavior that is capable to be modeled.

To finish, table 1 collects the obtained results respecting to RMSE of the new model joint to the results obtained by some classical global to direct models. It can be see a general better accuracy of the new methodology comparing with the classical ones. The paper 11724 contains the complete benchmarking.

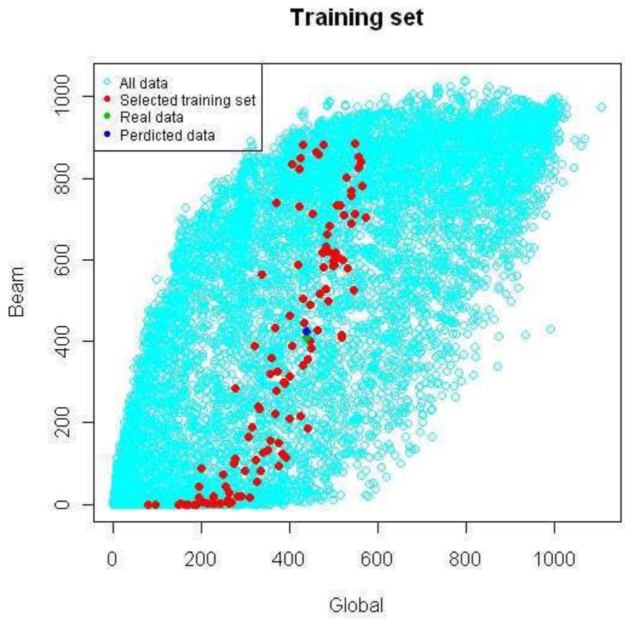

Fig. 7. All data and subset selected by knn.

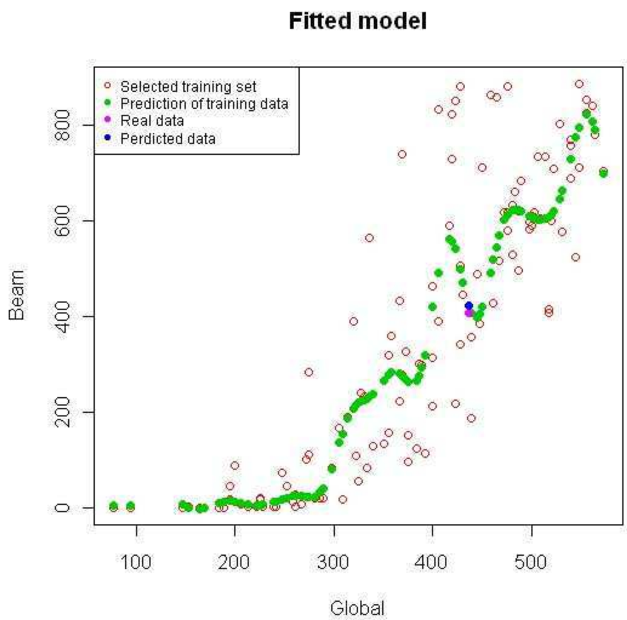

Fig. 8. SVM prediction in the training set.

Site	Orgill H	ERBS	REINDL	LOUCHE	CENER's model
Madrid	113	112	105	107	80
Cáceres	133	131	121	125	96
Murcia	101	98	94	95	83
Sevilla	83	81	79	77	72

Table 1. Comparison of obtained RMSE.

5. Conclusions and future works

This paper describes a new methodology to generate beam irradiance data when only global is available. The methodology combines some machine learning techniques, knn to select the training set and SVM to model direct irradiance. It has been applied in data from four Spanish stations and it shows better results that classical methods.

The proposed methodology suggests the possibility to include other classification techniques and other way to model the beam radiation.

We are working in the study of other classification methods and searching improvement in the SVM models to obtain a major accuracy of the final prediction of DNI.

References

[1] I. Guyon, N. Matic, V. Vapnik, Discovering informative patterns and data cleaning Advances in Knowledge Discovery and Data Mining, (1996) 181-203. G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors.

[2] B. Scholkopf , A. J. Smola. Learning with Kernels. MIT Press, 2002.

[3] A. J. Smola, B. Scholkopf. On a kernel–based method for pattern recognition, regression, approximation and operator inversion. Algorithmica, 22:211–231, 1998. Technical Report 1064, GMD FIRST, April 1997.

[4] V. Vapnik. The Nature of Statistical Learning Theory. Springer, N.Y., 1995.

[5] I. Steinwart D. Hush and C. Scovel. A classification framework for anomaly detection. Journal of Machine Learning Research, 6(211–232):22, February 2005.

[6] Ineichen P., (2008). Comparison and validation of three global-to-beam irradiance models against ground measurements. Solar Energy, 82, 501-512.

[7] Orgill J.F., Hollands K.G.T., (1977). Correlation equation for hourly diffuse radiation on a horizontal surface. Solar Energy, 19, 357-359.

[8] Erbs D.G., Klein S.A., Duffie J.A., (1982). Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Solar Energy, 28, 293-304.

[9] Reindl D.T., Beckman W.A., Duffie J.A., (1990). Diffuse fraction correlations. Solar Energy, 45, 1-7.

[10] Louche A., Notton G., Poggi P., Simonnot G., (1991). Correlations for direct normal and global horizontal irradiation on French Mediterranean site. Solar Energy, 46, 261-266.

[11] Dasarathy Belur V, (1991). Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. ISBN 0-8186-8930-7