

# Illuminance Prediction through SVM Regression

F. Bellocchio and S. Ferrari and M. Lazzaroni

Dept. of Information Technology  
Università degli Studi di Milano  
Milan, Italy

Email: {francesco.bellocchio, stefano.ferrari,  
massimo.lazzaroni}@unimi.it

L. Cristaldi and M. Rossi

Dept. of Electrical Engineering  
Politecnico di Milano  
Milan, Italy

Email: loredana.cristaldi@polimi.it,  
marco4.rossi@mail.polimi.it

T. Poli and R. Paolini

Dept. of Building Environment  
Science and Technology  
Politecnico di Milano  
Milan, Italy

Email: tiziana.poli@polimi.it,  
ricc.paolini@gmail.com

**Abstract**—In a scenario where renewable energies will play a foreground role, a reliable forecast of the energy production of such sources, like solar radiation, is a requirement for managing smart grids. However, the ability to predict the possibility to produce sustainable energy in different climatic conditions can be very useful for many other purposes (e.g., for Climate Sensitive Buildings). This is particularly true when working with climatic data that are, as a matter of fact, highly unsteady. Nevertheless, the use of data collected in the past can help to face the daily and seasonal variability. An algorithm for illuminance prediction based on Support Vector Regression (SVR) is here proposed and the results are presented and discussed.

## I. INTRODUCTION

Several renewable energy sources have been studied and used over the last years. A consequence of this growing interest is highlighted by an increasing diffusion of power plants based on renewable power source. The increasing interest in the scientific community on this kind of approach to energy production leads to develop more efficient solutions for renewable power generation and managing. Among the others, the more interesting renewable energy source is perhaps the solar radiation. In fact, with nowadays technology and energy costs, this way to produce energy provides both a cost-effectively high energy density and low environmental impact.

Starting from this simple consideration, it is very easy to understand why the photovoltaic (PV) technology, based on photodiodes that convert radiation into an electric current, is actively studied and analyzed. The function and the performance of the system based on PV technology are influenced by many factors. In particular, in this paper, the influence of the solar radiation is considered.

Since the solar radiation is not available in a controlled mode, it is very interesting to know the solar radiation data both for evaluating the plant performance and, if possible as here suggested, for forecasting the energy availability.

In this scenario, solar radiation data can play an important and sometimes fundamental role in the design and operation, but also in the economic assessment, of solar power systems. For tasks such as plant dimensioning or building planning,

a long term prediction of the solar radiation is required [1]. On the other hand, although long term fluctuation of the solar radiation are possible, short term prediction is usually sought for tasks related to the managing of power plants or smart buildings. In this paper, the last kind of prediction is considered.

Two main aspects should be taken into account:

- Geographic availability: measurements of solar radiation are not available for all potentially interesting sites. Often, for a given plant position, when no radiation data have been collected, data from another site, or, better, a pool of nearby sites, are utilized. Then, a suitable technique can combine these data and provide a relationship for the solar radiation estimate of the site under consideration. A negative aspect concerning this approach, however, have to be considered. This method, in fact, does not give satisfying results when the considered area is characterized, for example, by large variations in elevation or in surface type [2].
- Time availability: when forecast on energy market have to be elaborated, data available at the processing time are not fully useful. Data, in fact, give the past situation but are not really able to give information on future values.

Depending on the application, short term prediction can range from few minutes to several hours ahead.

In [3], a comparison of short term prediction statistical techniques is carried out over different time horizon and different datasets. The techniques used for the comparison spans from linear regression to Autoregressive Integrated Moving Average (ARIMA), Unobserved Components models, neural networks and hybrid models (a combination of neural networks and statistical models). For all the datasets, the input data used for the prediction where the values at the past hour and at the previous day, although for some datasets the pool of the input data were enriched with additional measurements such as humidity, cloud cover, and measures of atmospheric turbulence. The ARIMA model predictions resulted to be the

most robust for almost every datasets.

A similar study has been presented in [4] (using a different dataset) where the best model resulted to be the linear regression model which make use of 10 to 15 past data values.

Also the input data used for the prediction can differ. Since, the solar radiation at the ground is highly related to the meteorological situation, data and methods for weather forecast can be effectively used (e.g., satellite observations, [5]). However, while this kind of data covers a large region with a low resolution, a local measurement station (e.g., MeteoLab [2][6]) can provide a more accurate measurement for a given building. In this paper, a short term prediction based on local data through Support Vector Machines (SVM) is studied. In section II, the SVM model for regression is introduced, while the data used in the experiments are explained in section III, along with the methodological aspects of the experiments. The results are discussed in section IV, and some conclusions are drawn in section V.

## II. SUPPORT VECTOR REGRESSION

Support Vector Machines (SVM) is a powerful method for classification [7][8] and regression [9]. In the latter domain, the method is usually named Support Vector Regression (SVR). In its original formulation, the regression function is obtained as the linear combination of some samples, called Support Vectors (SV). The solution to the regression problem is hence the hyperplane that minimize a suitable loss function, which can be chosen such that the optimization problem results to be convex. This property guarantees that the optimal solution (which identifies the SVs and the corresponding coefficients) is unique. Since most of the real problems exhibits a non-linear relation between the observed and the predicted variables, a mapping can be applied to the observed variables. This transformation maps the variables into a higher dimensional space (called feature space) where the solution is computed as linear combination of the mapped SVs. The above mentioned scheme (known as “kernel trick”) do not increase the computational cost of the solution, since the explicit mapping is not required by the optimization procedure. In fact, only the inner product of (mapped) data pairs is required for computing the solution, which can be computed as a function called kernel.

The output of a SVR is computed as:

$$f_{\text{svr}}(x) = \sum_{i=1}^n \beta_i k(x, x_i) + b \quad (1)$$

where  $\beta_i$  and  $x_i$  are respectively the weight and the position of each SV,  $n$  is the number of SVs,  $b$  is the bias and  $k(\cdot, \cdot)$  is the kernel function.

The Gaussian kernel is one of the most used and its shape is regulated by the parameter  $\sigma$ , which represents the width of the Gaussian:

$$k(x, x_i) = G(\|x - x_i\|; \sigma) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right) \quad (2)$$

The optimization procedure for SVR is mainly regulated by two parameters,  $\varepsilon$  and  $C$ , which are used to estimate the

goodness of the solution and are called hyperparameters. In fact, the aim of SVR optimization is finding the solution,  $f_{\text{svr}}(\cdot)$ , that deviates no more than  $\varepsilon$  from the target data and that is as smoother as possible. The solution is hence computed as the function that minimize a functional that is composed by two terms: one that measures the closeness of the function to the data (controlled by  $\varepsilon$ ) and one that measures the smoothness of the solution. The trade off the two terms is determined by the constant  $C > 0$ . Due to the constraints that make the problem convex,  $C$  is also the maximum of the magnitude of the coefficients  $\beta_i$ . Moreover, when non-linear regression is required, also the choice of the kernel,  $k(\cdot, \cdot)$ , has to be operated.

The choice of the optimal values for the hyperparameters ( $\varepsilon$ ,  $C$ , and  $k(\cdot, \cdot)$ ) is usually operated by cross validation.

Several variants of SVM have been studied in literature (e.g., [10][11]) and commercial (e.g., [12]) as well as free (e.g., [13][14]) implementations are available.

## III. EXPERIMENTS

The dataset used in the experiments described in the present paper has been collected by the MeteoLab [2][6] between October 2005 and October 2007. MeteoLab measures:

- air temperature;
- relative humidity;
- global horizontal irradiance;
- diffuse horizontal irradiance;
- global horizontal illuminance.

The station samples the data every ten minutes, but the dataset used here considers only their hourly average.

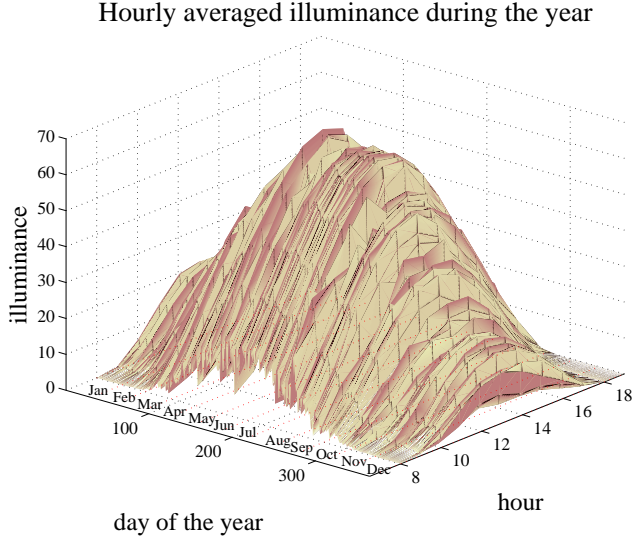
The illuminance varies both on daily and seasonal basis. The surface reported in Figs. 1a–b shows this behaviour. It has been obtained by averaging the illuminance samples measured in the same hour of the same day of the year. Although there is a clear trend, the variability of the illuminance (which depends also by fast changing meteorological phenomena) make the resulting surface very rough.

### A. Dataset Pre-Processing

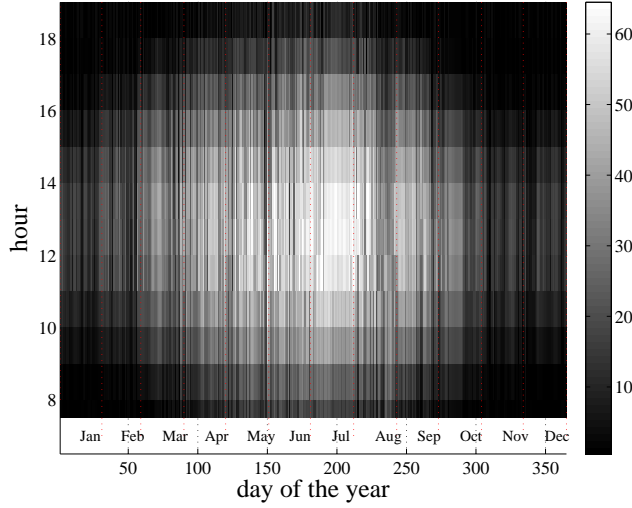
For this preliminary work, we focused only on the global horizontal illuminance. Besides, only the daily hours range ([8, 19]) has been considered. The dataset used for the experiments has been formed by records with the following fields:

- $d_i$ , day of the year
- $h_i$ , hour
- $lh_i$ , illuminance of one hour before  $h_i$
- $ld_i$ , average illuminance of the day before the hour  $h_i$  of the day  $d_i$
- $lw_i$ , average illuminance of the week before the hour  $h_i$  of the day  $d_i$
- $l_i$ , illuminance at  $h_i$

We used the first 5 fields as input variables for predicting the value of the last field. Hence, we have a dataset  $S = \{(x_i, y_i) | x_i = [d_i, h_i, lh_i, ld_i, lw_i], y_i = l_i, i = 1, \dots, n\}$  and the prediction problem consists in finding a regression function,  $f : R^5 \rightarrow R$  such that  $y_i \approx f(x_i)$ .



(a)



(b)

Fig. 1. The illuminance samples belonging to the same day of the year and to the same hour have been averaged and plotted as a surface. It is evident that there is a trend, but with an high variability which results in a non-smooth surface.

The dataset is composed of 8978 samples. It has been randomly partitioned in training, validation, and testing set, which are composed of 33%, 33%, and 34% of the total samples, respectively (i.e., 2963, 2963, and 3052 samples). Five different randomizations have been used for the experiments.

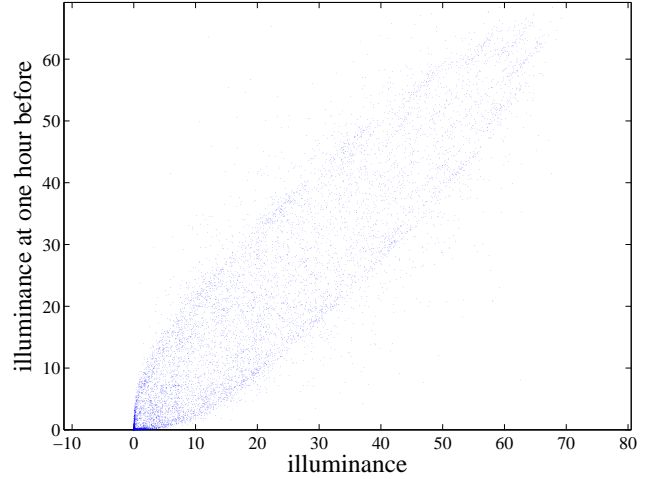
The prediction error has been measured by means of the average of the absolute error achieved on the testing set data:

$$\text{Err}(f) = E(|y_i - f(x_i)|) \quad (3)$$

and its relative value:

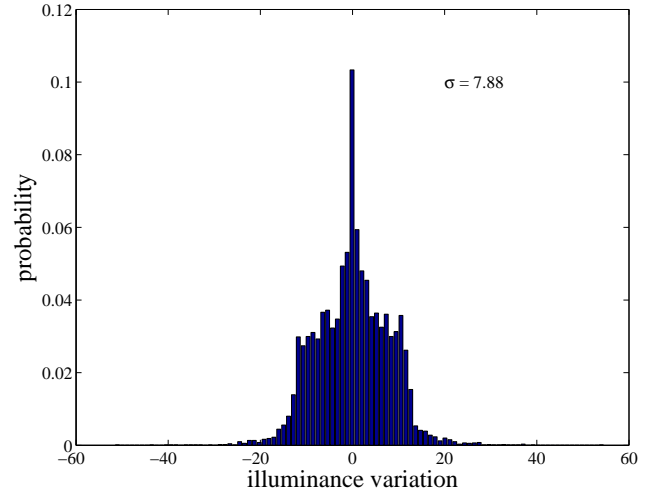
$$\text{Rel}(f) = E\left(\frac{|y_i - f(x_i)|}{y_i}\right) \quad (4)$$

One hour illuminance variation



(a)

One hour illuminance variation distribution



(b)

Fig. 2. A naive predictor can make direct use of the illuminance value measured one hour before. In panel (a), the relationship between the two illuminance measurement made with one hour one from the other is shown. Although the samples populate the region around the identity line, an evident dispersion is shown. In panel (b), the estimated probability density function of the variation (which standard deviation is 7.88).

### B. Simple predictors

In order to assess the effectiveness of the prediction using SVR based models, we used some naive predictors for estimating the baseline. Since it is reasonable that the illuminance at a given hour is similar to the illuminance at the previous hour, the simpler lazy predictor,  $f_h(\cdot)$ , can be formalized as:

$$f_h(h, lh) = lh \quad (5)$$

where the illuminance at a given hour is estimated as the value of the illuminance measured one hour before.

The error achieved has been  $\text{Err}(f_h) = 6.09$ , which is in line with the results reported in Fig. 2. Since no training is involved, the whole dataset has been used to estimate  $\text{Err}(f_h)$ .

TABLE I  
TEST ERROR ACHIEVED BY THE SVR PREDICTOR.

#trial	Err (std)	Err( $f_{\text{SVR}}$ )	Rel (std)	Rel( $f_{\text{SVR}}$ )
1	2.26 (3.66)		0.279 (0.820)	
2	2.34 (3.79)		0.286 (0.770)	
3	2.45 (4.02)	2.34	0.319 (0.829)	0.297
4	2.32 (3.73)		0.309 (0.750)	
5	2.31 (3.68)		0.292 (0.602)	

Another simple predictor,  $f_{\text{incr}}(h, lh)$ , can estimate, for each hour, the average increment of illuminance with respect to one hour before. In this case, the predictor needs of a training phase for estimating the increments, but, since the training procedure does not require hyperparameters, the validation sets has been added to the training sets. The average error resulting from 5 different randomizations of the data has been  $\text{Err}(f_{\text{incr}}) = 3.98$ .

### C. Prediction through SVR

In order to train a SVR predictor, the hyperparameters that regulate the optimization procedure, have to be set to the proper value. Since the optimal values cannot be estimated a-priori, several combinations have to be tried and their effectiveness have to be assessed by cross validation.

The hyperparameters values that we challenged are:

- accuracy,  $\varepsilon$ : we tested the values in  $\{0.01, 0.1, 1\}$ ;
- regularization trade off,  $C$ : we tested the values in  $\{0.1, 1, 10, 100\}$ ;
- kernel function,  $k(\cdot, \cdot)$ : we tested the linear (i.e., without mapping the data in the feature space), and the Gaussian kernel, with the width  $\sigma$  in  $\{0.1, 0.2, 0.4, 0.8, 1.6, 3.2\}$ ; it should be noticed that the regression is operated on normalized data.

The best average validation error has been obtained for  $\varepsilon = 0.01$ ,  $C = 100$ , with a Gaussian kernel with  $\sigma = 3.2$ . In Table I, the test error by SVR models trained with the above mentioned hyperparameters is reported for all the trials, along with the standard deviation of the error. On the average, the test error of the best configuration,  $\text{Err}(f_{\text{SVR}})$ , is 2.34, for a relative error,  $\text{Rel}(f_{\text{SVR}})$ , of 0.297.

## IV. DISCUSSION

The SVR model allows to achieve a prediction error well below that of the simple predictors described in section III-B. It can exploit the information on context in which the prediction is required. In this sense, it is able to adapt the prediction to the trend of the period of time before the prediction time. The achieved relative error, 0.297 (cfr. Table I), well compares with the results obtained in [3], where the relative error obtained by the ARIMA model has been of 0.264, 0.289, and 0.236 for three datasets composed by data similar to those here used, although ten times larger in terms of observations.

Since the best hyperparameters combination ( $\varepsilon = 0.01$ ,  $C = 100$ , and  $\sigma = 3.2$ ) results to be composed of extremal values for each hyperparameters, the need of experimenting a larger set of values for the hyperparameters can be questioned.

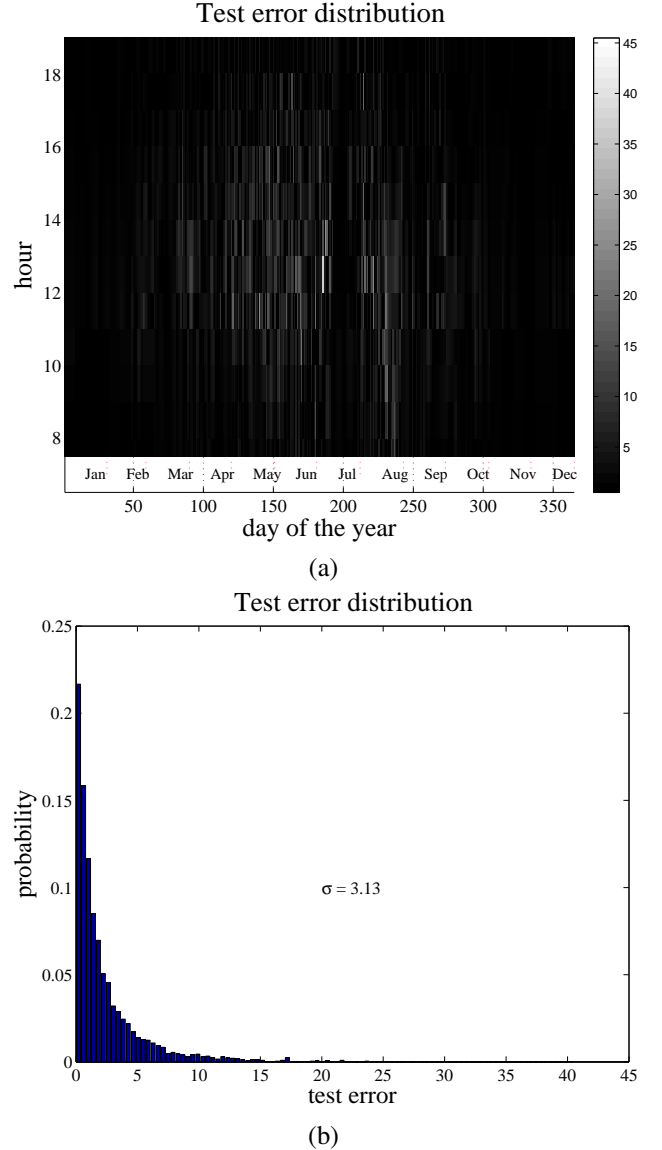


Fig. 4. Test error distribution. In panel (a), the test error computed in all the trial is reported with respect to the day of the year and the hour. The error follows the seasonal and daily variability. In panel (b), the estimated probability density function of the test error (which standard deviation is 3.13).

However, although it is possible that different hyperparameters values (e.g.,  $\varepsilon = 0.005$ ,  $C = 200$ ) can make the test error smaller, the improvement cannot worth the effort required. In fact, the computational time required by the SVR optimization greatly increases with  $C$  and is inversely proportional to  $\varepsilon$ . In Fig. 3, the test error achieved with several combination of values of the hyperparameters is reported. In particular, Fig. 3a-c depicts the test error with respect to  $\varepsilon$  and  $C$  values, and to the kernel, respectively. Moreover, for each hyperparameter, a line that connect the results relative to the configuration with the optimal values of the other two hyperparameters is reported. These lines shows a clear trend for the test error that hardly can decrease for values of  $\varepsilon$

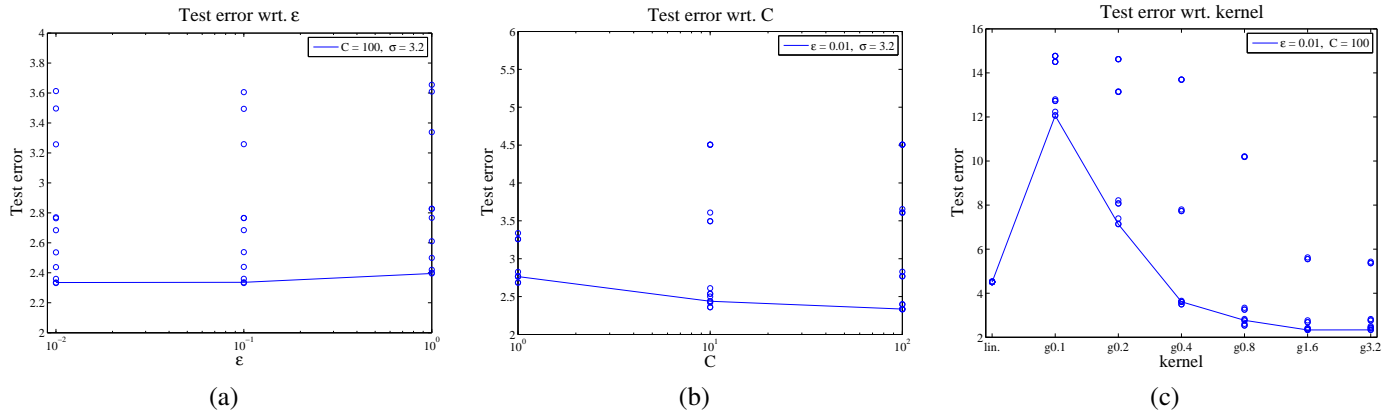


Fig. 3. Test error wrt. the hyperparameters.

smaller than 0.01 or values of  $C$  larger than 100.

The distribution of the test error evaluated for the models trained (for all the five randomizations) with the best combination of the hyperparameters is reported in Fig. 4a, while its normalized histogram is shown in Fig. 4b. It clearly exhibits a better behaviour than the naive predictor based on the hourly increments (cfr. Fig. 2). However, since it follows the seasonal and daily variability more information have to be involved in the learning process. This can be achieved both using the illuminance measurements over a period of time longer than that has been used in the presented experiments, and exploiting also the meteorological data from MeteoLab. For the ARIMA model in [3], this worth a relative error of 0.197, while [4] reports an error of 0.151 for a linear regression model that makes use of 15 past values. Hence, a similar improvement in the prediction accuracy can be hypotes for our model.

Besides, similar measurements from other sites could be integrated in order to model the climatic pattern of the region and forecast the almost random variations which are due to meteorological phenomena.

## V. CONCLUSION

In the present work, the prediction ability of the Support Vector Regression model has been challenged on a problem of forecasting the solar radiation quantity that reaches the ground.

In this application, the training database is composed of measures of environmental data and then they define the feature space on which the prediction is operated. Although the regression technique used is a basic machine learning model, the results are promising and compares well with similar studies found in the literature. Starting from these preliminary results, both the selection of a new pool of input data and of a more complex prediction model can be studied in order to improve the efficiency of the algorithm.

## REFERENCES

- [1] A. Ahmed, N. E. Korres, J. Ploennigs, H. Elhadi, and K. Menzel, "Mining building performance data for energy-efficient operation," *Advanced Engineering Informatics*, vol. 25, no. 2, pp. 341–354, 2011.
- [2] T. Poli, L. P. Gattoni, D. Zappalà, and R. Gottardi, "Daylight measurement in milan," in *Proc. of PLEA2006, Conf. on Passive and Low Energy Architecture*, 2006.

- [3] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, pp. 342–349, Mar 2009.
- [4] W. Ji, C. K. Chan, J. W. Loh, F. H. Choo, and L. H. Chen, "Solar radiation prediction using statistical approaches," in *Proceedings of the 7th international conference on Information, communications and signal processing*, ser. ICICS'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 646–650.
- [5] A. Hammer, D. Heinemann, E. Lorenz, and B. Lckhe, "Short-term forecasting of solar radiation: a statistical approach using satellite data," *Solar Energy*, vol. 67, pp. 139–150, 1999.
- [6] T. Poli, L. P. Gattoni, D. Zappalà, and R. Gottardi, "Daylight measurement in Milan," in *Clever Design, Affordable Comforta Challenge for Low Energy Architecture and Urban Planning*. Geneve - CH: Raphael Compagnon & Peter Haefeli and Willi Weber, 6 2006, pp. 429–433.
- [7] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
- [8] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
- [9] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, 2004.
- [10] S. Ferrari, F. Bellocchio, V. Piuri, and N. A. Borghese, "Multi-scale support vector regression," in *Proc. of IJCNN 2010 (IEEE Int. Joint Conf. on Neural Networks)*, Jul. 2010, pp. 2159–2164, (IJCNN Runner-up Best Paper Award for 2010).
- [11] M. Gönen and E. Alpaydın, "Localized multiple kernel regression," in *Proc. of the 20th IAPR Int. Conf. on Pattern Recognition*, 2010.
- [12] "Mosek," 2011. [Online]. Available: <http://www.mosek.com/>
- [13] T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999, ch. 11, pp. 169–184.
- [14] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, May 2011.