



**HAL**  
open science

## Machine Learning methods for solar radiation forecasting: a review

Cyril Voyant, Gilles Notton, S. Kalogirou, Marie Laure Nivet, Christophe Paoli, Fabrice Motte, Fouilloy Alexis

### ► To cite this version:

Cyril Voyant, Gilles Notton, S. Kalogirou, Marie Laure Nivet, Christophe Paoli, et al.. Machine Learning methods for solar radiation forecasting: a review. *Renewable Energy*, 2017, 105, pp.569-582. 10.1016/j.renene.2016.12.095 . hal-01426321

**HAL Id: hal-01426321**

**<https://univ-corse.hal.science/hal-01426321v1>**

Submitted on 4 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Machine Learning methods for solar radiation forecasting: a review

Cyril Voyant<sup>1,2</sup>, Gilles Notton<sup>1</sup>, Soteris Kalogirou<sup>3</sup>, Marie-Laure Nivet<sup>1</sup>, Christophe Paoli<sup>1,4</sup>, Fabrice Motte<sup>1</sup>, Alexis Fouilloy<sup>1</sup>

1- University of Corsica/CNRS UMR SPE 6134, Campus Grimaldi, 20250 Corte – France

2- CHD Castelluccio, radiophysics unit, B.P85 20177 Ajaccio- France

3- Department of Mechanical Engineering and Materials Science and Engineering, Cyprus University of Technology, P.O. Box 50329, Limassol 3401, Cyprus

4- Galatasaray University, Çırağan Cad. No: 36, 34349 Ortaköy/İstanbul – Turkey

\*corresponding author; tel +33495293666, fax +33495293797, [voyant@univ-corse.fr](mailto:voyant@univ-corse.fr)

## Abstract

Forecasting the output power of solar systems is required for the good operation of the power grid or for the optimal management of the energy fluxes occurring into the solar system. Before forecasting the solar systems output, it is essential to focus the prediction on the solar irradiance. The global solar radiation forecasting can be performed by several methods; the two big categories are the cloud imagery combined with physical models, and the machine learning models. In this context, the objective of this paper is to give an overview of forecasting methods of solar irradiation using machine learning approaches. Although, a lot of papers describes methodologies like neural networks or support vector regression, it will be shown that other methods (regression tree, random forest, gradient boosting and many others) begin to be used in this context of prediction. The performance ranking of such methods is complicated due to the diversity of the data set, time step, forecasting horizon, set up and performance indicators. Overall, the error of prediction is quite equivalent. To improve the prediction performance some authors proposed the use of hybrid models or to use an ensemble forecast approach.

Keywords: Solar radiation forecasting, machine learning, artificial neural networks, support vector machines, regression.

31

## 32 1. Introduction

33

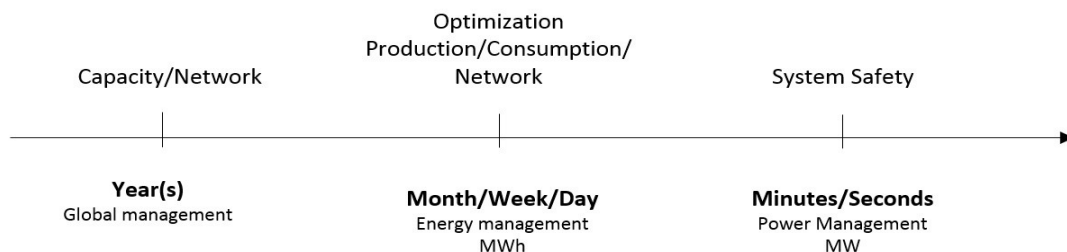
34 An electrical operator should ensure a precise balance between the electricity production and  
35 consumption at any moment. This is often very difficult to maintain with conventional and  
36 controllable energy production system, mainly in small or not interconnected (isolated) electrical grid  
37 (as found in islands). Many countries nowadays consider using renewable energy sources into their  
38 electricity grid. This creates even more problems as the resource (solar radiation, wind, etc.) is not  
39 steady. It is therefore very important to be able to predict the solar radiation effectively especially in  
40 case of high energy integration [1].

41

### 42 1.1. The necessity to predict solar radiation or solar production

43

44 One of the most important challenge for the near future global energy supply will be the large  
45 integration of renewable energy sources (particularly non-predictable ones as wind and solar) into  
46 existing or future energy supply structure. An electrical operator should ensure a precise balance  
47 between the electricity production and consumption at any moment. As a matter of fact, the operator  
48 has often some difficulties to maintain this balance with conventional and controllable energy  
49 production system, mainly in small or not interconnected (isolated) electrical grid (as found in  
50 islands). The reliability of the electrical system then become dependent on the ability of the system to  
51 accommodate expected and unexpected changes (in production and consumption) and disturbances,  
52 while maintaining quality and continuity of service to the customers. Then, the energy supplier must  
53 manage the system with various temporal horizons (see Fig. 1).



54

55 Figure 1. Prediction scale for energy management in an electrical network [2]

56

57 The integration of renewable energy into an electrical network intensifies the complexity of the grid  
 58 management and the continuity of the production/consumption balance due to their intermittent and  
 59 unpredictable nature [1, 2]. The intermittence and the non-controllable characteristics of the solar  
 60 production bring a number of other problems such as voltage fluctuations, local power quality and  
 61 stability issues [3, 4]. Thus forecasting the output power of solar systems is required for the effective  
 62 operation of the power grid or for the optimal management of the energy fluxes occurring into the  
 63 solar system [5]. It is also necessary for estimating the reserves, for scheduling the power system, for  
 64 congestion management, for the optimal management of the storage with the stochastic production and  
 65 for trading the produced power in the electricity market and finally to achieve a reduction of the costs  
 66 of electricity production [1, 3, 6, 7]. Due to the substantial increase of solar power generation the  
 67 prediction of solar yields becomes more and more important [8]. In order to avoid large variations in  
 68 renewable electricity production it is necessary to include also the complete prediction of system  
 69 operation with storage solutions. Various storage systems are being developed and they are a viable  
 70 solution for absorbing the excess power and energy produced by such systems (and releasing it in peak  
 71 consumption periods), for bringing very short fluctuations and for maintaining the continuity of the  
 72 power quality. These storage options are usually classified into three categories:

- 73 - Bulk energy storage or energy management storage media is used to decouple the timing of  
 74 generation and consumption.
- 75 - Distributed generation or bridging power - this method is used for peaks shaving - the storage  
 76 is used for a few minutes to a few hours to assure the continuity of service during the  
 77 energy sources modification.
- 78 - The power quality storage with a time scale of about several seconds is used only to assure the  
 79 continuity of the end use power quality.

80 Table 1 shows these three categories and their technical specifications. As shown, every type of  
 81 storage is used in different cases to solve different problems, with different time horizon and quantities  
 82 of energy.

83 Table 1. The three categories of storage and their technical specification

Category	Discharge power	Discharge Time	Stored Energy	Representative Application
Bulk energy	10-1000 MW	1-8 h	10-8000 MWh	Load levelling, generation capacity
Distributed generation	0.1-2 MW	0.5-4 h	50-8000 kWh	Peak shaving, transmission deferral
Power quality	0.1-2 MW	1-30 s	0.03-16.7 kWh	End-use power quality/reliability

84

85 Table 1 shows that the electricity storage can be widely used in a lot of cases and applications as a  
 86 function of the time of use and the power needs of the final user. Finally, it shows that the energy  
 87 storage acts at various time levels and their appropriate management requires the knowledge of the  
 88 power or energy produced by the solar system at various horizons: very short or short for power  
 89 quality category to hourly or daily for bulk energy storages. Similarly, the electrical operator needs to  
 90 know the future production (Figure 1) at various time horizons from one to three days, for preparing  
 91 the production system and to some hours or minutes for planning the start-up of power plants (Table  
 92 2). Starting a power plant needs between 5 min for a hydraulic one to 40 hours for a nuclear one.  
 93 Moreover, the rise in power of the electrical plants is sometimes low, thus for an effective balance  
 94 between production and consumption an increase of the power or a starting of a new production needs  
 95 to be anticipated sometimes well in advance.

96 Table 2. Characteristics of electricity production plants [9]

Type of electrical generator	Power size MW	Minimum power capacity percentage of peak power	Rise speed in power per min percentage of peak power	Starting time hours
Nuclear Power Plant	400–1300 per reactor	20%	1%	40 h (cold)-18 h (hot)
Steam thermal plant	200–800 per turbine	50%	0.5%-5%	11-20 h (cold)-5 h (hot)
Fossil-fired power plants	1–200	50% - 80%	10%	10 min-1 h
Combined-cycle plant	100–400	50%	7%	1-4 h
Hydro power plant	50–1300	30%	80%-100%	5 min
Combustion turbine (light fuel)	25	30%	30%	15-20 min
Internal combustion engine	20	65%	20%	45-60 min

97  
 98 Furthermore, the relevant horizons of forecast can and must range from 5 minutes to several days as it  
 99 was confirmed by Diagne et al. [6]. Elliston and MacGill [10] outlined the reasons to predict solar  
 100 radiation for various solar systems (PV, thermal, concentrating solar thermal plant, etc.) insisting on  
 101 the forecasting horizon. It therefore seems apparent that the time-step of the predicted data may vary  
 102 depending on the objectives and the forecasting horizon. All these reasons show the importance of  
 103 forecasting, whether in production or in consumption of energy. The need for forecasting lead to the  
 104 necessity to use effective forecasting models. In the next section the various available forecasting  
 105 methodologies are presented.

106

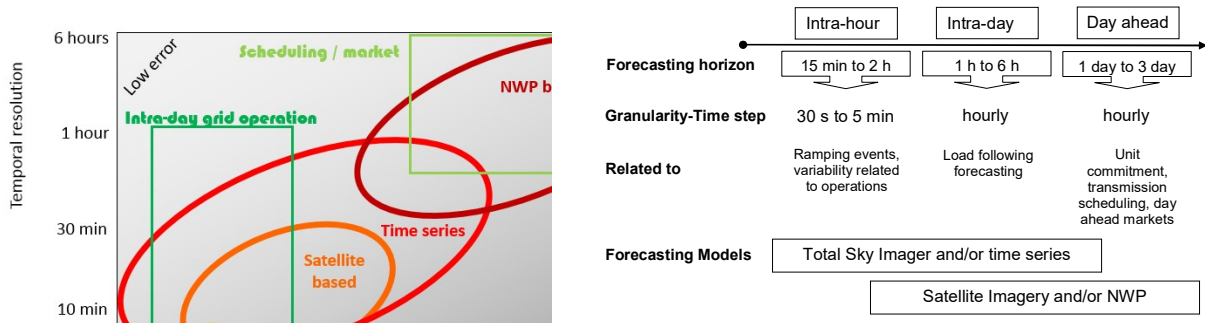
107 **1.2. Available forecasting methodologies**

108

109 The solar power forecasting can be performed by several methods; the two big categories are the cloud  
 110 imagery combined with physical models, and the machine learning models. The choice for the method  
 111 to be used depends mainly on the prediction horizon; actually all the models have not the same  
 112 accuracy in terms of the horizon used. Various approaches exist to forecast solar irradiance depending  
 113 on the target forecasting time. The literature classifies these methods in two classes of techniques:

- 114 - Extrapolation and statistical processes using satellite images or measurements on the ground  
 115 level and sky images are generally suitable for short-term forecasts up to six hours. This  
 116 class can be divided in two sub-classes, in the very short time domain called “Now-  
 117 casting” (0–3 h), the forecast has to be based on extrapolations of real-time measurements  
 118 [5]; in the Short-Term Forecasting (3–6 h), Numerical Weather Prediction (NWP) models  
 119 are coupled with post-processing modules in combination with real-time measurements or  
 120 satellite data [5, 11].
- 121 - NWP models able to forecast up to two days ahead or beyond [12, 13] (up to 6 days ahead  
 122 [13]). These NWP models are sometimes combined with post-processing modules and  
 123 satellite information are often used [2].

124 Figures 2a and 2b [6, 14] summarize the existing methods versus the forecasting horizon, the objective  
 125 and the time step.



126 Figure 2 a) Forecasting error versus forecasting models (left) [6,14]. b) Relation between forecasting  
 127 horizons, forecasting models and the related activities (right) [6, 14]

128 The NWP models predict the probability of local cloud formation and then predict indirectly the  
 129 transmitted radiation using a dynamic atmosphere model. The extrapolation or statistical models  
 130 analyse historical time series of global irradiation, from satellite remote sensing [15] or ground  
 131 measurements [16] by estimating the motion of clouds and project their impact in the future [6, 13,  
 132 17]. Hybrid methods can improve some aspects of all of these methods [6, 14]. The statistical  
 133 approach allows to forecast hourly solar irradiation (or at a lower time step) and NWP models use  
 134 explanatory variables (mainly cloud motion and direction derived from atmosphere) to predict global  
 135 irradiation N-steps ahead [15]. Very good overviews of the forecasting methods, with their limitations  
 136 and accuracy can be found in [1, 5, 6, 10, 12, 14, 18]. Benchmarking studies were performed to assess

137 the accuracy of irradiance forecasts and compare different approaches of forecasting [8, 13, 17, 19–  
138 21]. Moreover, the accuracy evaluation parameters are often different; some parameters such as  
139 correlation coefficient and root mean square error are often used, but not always adapted to compare  
140 the model performance. Thus the time period used for evaluating the accuracy varies widely. Some of  
141 them analysed the model accuracy over a period of one or several years, whereas some others over a  
142 period of some weeks introducing a potential seasonal bias. In these conditions, it is not easy to make  
143 comparisons and the accuracy of the results produced, as shown in this paper, must be carefully  
144 evaluated in selecting the right method to use. As part of COST Action ES1002. (European  
145 Cooperation in Science and Technology) [22] on Weather Intelligence for Renewable Energies  
146 (WIRE) a literature review on the forecasting accuracy applied to renewable energy systems mainly  
147 solar and wind is carried out. In this paper an overview on the various methodologies available for  
148 solar radiation prediction based on machine learning is presented. A lot of review papers are available,  
149 but it is very rare to find a paper which is totally dedicated to the machine learning methods and that  
150 some recent prediction models like random forest, boosting or regression tree be integrated. In the  
151 next section the different methodologies used in the literature to predict global radiation and the  
152 parameters used for estimating the model performances are presented.

153

## 154 **2. Machine learning methods**

155

156 Machine learning is a subfield of computer science and it is classified as an artificial intelligence  
157 method. It can be used in several domains and the advantage of this method is that a model can solve  
158 problems which are impossible to be represented by explicit algorithms. In [23] the reader can find a  
159 detailed review of some machine learning and deterministic methods for solar forecasting. The  
160 machine learning models find relations between inputs and outputs even if the representation is  
161 impossible; this characteristic allow the use of machine learning models in many cases, for example in  
162 pattern recognition, classification problems, spam filtering, and also in data mining and forecasting  
163 problems. The classification and the data mining are particularly interesting in this domain because  
164 one has to work with big datasets and the task of preprocessing and data preparation can be undertaken  
165 by the machine learning models. After this step, the machine learning models can be used in  
166 forecasting problems. In global horizontal irradiance forecasting the models can be used in three  
167 different ways [24]:

- 168 - structural models which are based on other meteorological and geographical parameters;
- 169 - time-series models which only consider the historically observed data of solar irradiance as  
170 input features (endogenous forecasting);

171 - hybrid models which consider both, solar irradiance and other variables as exogenous  
172 variables (exogenous forecasting).

173 As already mentioned machine learning is a branch of artificial intelligence. It concerns the  
174 construction and study of systems that can learn from data sets, giving computers the ability to learn  
175 without being explicitly programmed. In the predictive learning problems, the system consists of a  
176 random “output” or “response” variable  $y$  and a set of random “input” or “explanatory” variables  $x =$   
177  $\{x_1, \dots, x_n\}$ . Using a “training” sample  $\{y_i, x_i\}_1^N$  of known  $(y, x)$ -values, the goal is to obtain an  
178 estimate or approximation  $f(x)$ , of the function  $f^*(x)$  mapping  $x$  to  $y$ , that minimizes the expected  
179 value of some specified loss function  $L(y, f(x))$  over the joint distribution of all  $(y, x)$ -values:

$$180 \quad E_{yx}(L(y, f(x))) = \underset{f(E_x(E_y(L(y, f(x))))|x)}{\operatorname{argmin}} \quad (1)$$

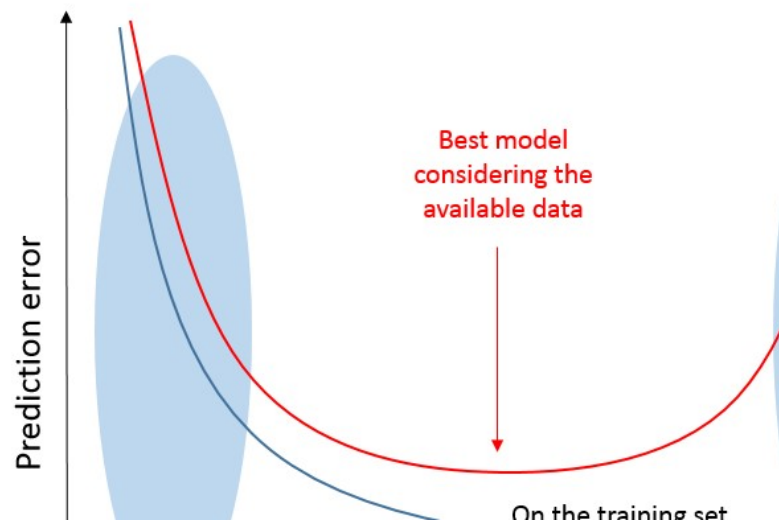
181 The frequently employed loss functions include squared-error  $(y - f(x))^2$  and absolute error  
182  $|y - f(x)|$  for regression and negative binomial log-likelihood for classification. A common

183 procedure is to restrict  $f$  to be a member of a parameterized class of functions  $F(x; P)$ , where  $P$   
184  $= \{P_1, P_2, \dots\}$  is a finite set of parameters whose joint values identify individual class members.  
185 Usually all the methods dedicated to the machine learning, especially the supervised cases, are  
186 confronted to bias-variance tradeoff (see Figure 3). This is the problem of trying to minimize two  
187 sources of error simultaneously which prevent supervised learning algorithms from generalizing  
188 outside their training set:

- 189 - The bias is the deviation (error) from erroneous assumptions made in the learning algorithm.  
190 High values of bias can cause an algorithm to lose its ability to establish relations between  
191 actual and target outputs (under-fitting).
- 192 - The variance is the error created from actually capturing small fluctuations in the training set.  
193 It should be noted that high variance can cause overfitting which results in modeling the  
194 random noise in the training dataset, rather than the intended output.

195 The breakdown of bias-variance relationship is a way of investigating the expected generalization  
196 error of a learning algorithm for a particular problem that is the sum of three terms, the bias, variance,  
197 and irreducible error, which result from noise in the problem itself.





198

199

Figure 3: Bias variance tradeoff

200

201 In this part we present the different machine learning models used in forecasting, initially the models  
 202 for classification and data preparation, secondly the supervised learning models, thirdly the  
 203 unsupervised learning models and finally the ensemble learning models.

204

## 205 **2.1. Classification and data preparation**

206

207 Machine learning algorithms learn from data. It is therefore critical to choose the right data and  
 208 prepare them properly to enable the problem to be solved effectively.

### 209 **2.1.1. Discriminant analysis and Principal Component Analysis (PCA)**

210

211 The principal component analysis (PCA) is a statistical method which uses an orthogonal  
 212 transformation to transform a set of observations of probably correlated variables into a set of values  
 213 of linearly uncorrelated variables which are called principal components [25]. The number of principal  
 214 components created in the process, is lower or equal to the number of original variables. Such  
 215 transformation is defined in such a way so as the first principal component has the largest variance  
 216 possible, i.e., to account for the maximum variability in the data, and each subsequent component to  
 217 have the highest variance possible under the restriction that it is orthogonal to the previous  
 218 components. As a result, the resulting vectors form an uncorrelated orthogonal basis set. It should be

219 noted that the principal components are orthogonal as they are the eigenvectors of the covariance  
220 matrix, which is symmetric. Moreover, PCA is sensitive to the relative scaling of the original variables  
221 [26].

222

### 223 **2.1.2. Naive Bayes classification and Bayesian networks**

224

225 In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on  
226 applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive  
227 Bayes classifiers are highly scalable, requiring a number of parameters proportional to the number of  
228 variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by  
229 evaluating a closed-form expression, which takes linear time, rather than by expensive iterative  
230 approximation as used for many other types of classifiers [27]. A Bayesian network, also called Bayes  
231 network, Bayesian model, belief network or probabilistic directed acyclic graphical model is a  
232 probabilistic graphical model, which is a type of statistical model that represents a set of random  
233 variables and their conditional dependencies via a directed acyclic graph (DAG).

234

### 235 **2.1.3. Data mining approach**

236

237 A data mining consists of the discovery of interesting, unexpected or valuable structure in large data  
238 sets that can be called with the slogan Big Data [28]. In other words, data mining consists of extracting  
239 the most important information from a very large data set. Indeed, the classical statistical inference has  
240 been developed for processing small samples. In the presence of very large databases, all the standard  
241 statistical indexes become significant and thus interesting (e.g. for 1 million of data, the significance  
242 threshold of correlation coefficient is very low reaching 0.002, ...). Additionally, in data mining, data  
243 collected are analyzed for highlighting the main information before to use them in the forecasting  
244 models. Rather than opposing data mining and statistics, it is best to assume that data mining is the  
245 branch of statistics devoted to the exploitation of large databases. The techniques used are from  
246 different fields depending on classical statistics and artificial intelligence [29]. This last notion was  
247 defined by “*The construction of computer programs that engage in tasks that are, for now, more*  
248 *satisfactorily performed by humans because they require high-level mental processes such as*  
249 *perceptual learning organization memory and critical thinking*”. There is not really a consensus of  
250 this definition, and many other similar ones are available.

251

## 252           **2.2.Supervised learning**

253

254   In supervised learning, the computer is presented with example inputs and their desired outputs, given  
255   by a "teacher", and the goal is to learn a general rule that maps inputs to outputs[23]. These methods  
256   need an “expert” intervention. The training data comprise of a set of training examples. In supervised  
257   learning, each pattern is a pair which includes an input object and a desired output value. The function  
258   of the supervised learning algorithm is to analyze the training data and produce an inferred function.

259

### 260           **2.2.1. Linear Regression**

261

262   Early attempts to study time series, particularly in the 19th century, were generally characterized by  
263   the idea of a deterministic world. It was the major contribution of Yule (1927) which launched the idea  
264   of stochasticity in time series by assuming that every time series can be regarded as the realization of a  
265   stochastic process. Based on this simple idea, a number of time series methods have been developed  
266   since that time. Workers such as Slutsky, Walker, Yaglom, and Yule first formulated the concept of  
267   autoregressive (AR) and moving average (MA) models [30]. Wold’s decomposition theorem [31] led  
268   to the formulation and solution of the linear forecasting problem of Kolmogorov in 1941. Since then, a  
269   considerable amount of literature is published in the area of time series, dealing with parameter  
270   estimation, identification, model checking and forecasting; see, for example ref. [32] for an early  
271   survey.

272

### 273           **2.2.2. Generalized Linear Models**

274

275   Generalized linear model (GLM) in statistics, is a flexible generalization of ordinary linear regression  
276   which allows for response variables that have error distribution models other than a normal  
277   distribution. GLM generalizes linear regression by permitting the linear model to be related to the  
278   response variable through a link function and by considering the magnitude of the variance of each  
279   measurement to be a function of its predicted value[33]. Some studies improve the regression quality  
280   using a coupling with other predictors like Kalman filter [34].

281

### 282           **2.2.3. Nonlinear Regression**

283

284 Artificial Neural Networks (ANN) are being increasingly used for nonlinear regression and  
 285 classification problems in meteorology due to their usefulness in data analysis and  
 286 prediction[35]. The use of ANN is particularly predominant in the realm of time series forecasting  
 287 with nonlinear methods. Actually the availability of historical data on the meteorological utility  
 288 databases and the fact that ANNs are data driven methods capable of performing a non-linear mapping  
 289 between sets of input and output variables makes this modelling software tool very attractive.

290 Artificial neural network with d inputs, m hidden neurons and a single linear output unit defines a non-  
 291 linear parameterized mapping from an input vector x to an output y is given by (unbiased form):

$$292 \quad (2)$$

293 Each of the m hidden units are usually related to the tangent hyperbolic function  
 294  $f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$ . The parameter vector  $\mathbf{w} = (\{w_j\}, \{w_{ji}\})$  governs the non-linear mapping  
 295 and is estimated during a phase called the training or learning phase. During this phase, the ANN is  
 296 trained using the dataset  $\mathcal{D}$  that contains a set of n input and output examples. The second phase,  
 297 called the generalization phase, consists of evaluating, on the test dataset  $\mathcal{D}_*$ , the ability of the ANN to  
 298 generalize, i.e., to give correct outputs when it is confronted with examples that were not seen during  
 299 the training phase.

300 For solar radiation the relationship between the output  $\bar{k}^*(t+h)$  and the inputs  
 301  $\{k^*(t), k^*(t-1), \dots, k^*(t-p)\}$  has the form given by:

$$302 \quad \bar{k}^*(t+h) = \sum_{j=1}^m w_j f \left( \sum_{i=0}^p w_{ji} k^*(t-i) \right) \quad (3)$$

303 As shown by the preceding equation, the ANN model is equivalent to a nonlinear autoregressive (AR)  
 304 model for time series forecasting problems. In a similar manner as for the AR model, the number of  
 305 past input values p can be calculated with the auto-mutual information factor [36].

306 Careful attention must be put on the building of the model, as a too complex ANN will easily overfit  
 307 the training data. The ANN complexity is in relation with the number of hidden units or conversely the  
 308 dimension of the vector  $\mathbf{w}$ . Several techniques like pruning or Bayesian regularization can be  
 309 employed to control the ANN complexity. The Levenberg-Marquardt (approximation to the Newton's  
 310 method) learning algorithm with a max fail parameter before stopping training is often used to  
 311 estimate the ANN model's parameters. The max fail parameter corresponds to a regularization tool  
 312 limiting the learning steps after a characteristic number of prediction failures and consequently is a  
 313 means to control the model complexity [18,37]. Note that hybrid methods such as master optimization  
 314 by conjugate gradients for selecting ANN topology allow ANNs to perform at their maximum

315 capacity. A lot of studies show the impact of the coupling between ANN and other tools like for  
 316 example Kalman Filter or the fuzzy logic, which verify that often the gain is very interesting [38]

317

#### 318 **2.2.4. Support Vector Machines / Support Vector Regression**

319

320 Support vector machine is another kernel based machine learning technique used in classification tasks  
 321 and regression problems introduced by Vapnik in 1986 [39]. Support vector regression (SVR) is based  
 322 on the application of support vector machines to regression problems [18]. This method has been  
 323 successfully applied to time series forecasting tasks. In a similar manner as for the Gaussian Processes  
 324 (GPs), the prediction calculated by a SVR machine for an input test case  $\mathbf{x}_*$  is given by:

$$325 \quad \bar{y} = \sum_{i=1}^n \alpha_i k_{rbf}(x_i, \mathbf{x}_*) + b \quad (4)$$

326 With the commonly used RBF kernel defined by:

$$327 \quad k_{rbf}(x_p, x_q) = \exp\left[-\frac{(x_p - x_q)^2}{2\sigma^2}\right] \quad (5)$$

328 The parameter b (or bias parameter) is derived from the preceding equation and some specific  
 329 conditions. In the case of SVR, the coefficients  $\alpha_i$  are related to the difference of two Lagrange  
 330 multipliers, which are the solutions of a quadratic programming (QP) problem. Unlike ANNs, which  
 331 are confronted with the problem of local minimum, here the problem is strictly convex and the QP  
 332 problem has a unique solution. In addition, it must be stressed (that unlike GPs), not all the training  
 333 patterns participate to the preceding relationship. Indeed, a convenient choice of a cost function  
 334 (Vapnik's insensitive function) in the QP problem allows obtaining a sparse solution. The latter  
 335 means that only some of the coefficients  $\alpha_i$  will be nonzero. The examples that come with non-  
 336 vanishing coefficients are called Support Vectors.

337 One way to use the SVR in prediction problem is related to the fact that given the training dataset  
 338  $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^n$  and a test input vector  $\mathbf{x}_*$ , the forecasted clear sky index can be computed for a  
 339 specific horizon, h, like:

$$340 \quad \widehat{\mathbf{k}}^*(\mathbf{t} + \mathbf{h}) = \sum_{i=1}^n \alpha_i k_{rbf}(\mathbf{x}_i, \mathbf{x}_*) + b \quad (6)$$

341

### 2.2.5. Decision tree learning (Breiman bagging)

342  
343  
344  
345  
346  
347  
348  
349  
350  
351

The basic idea is very simple. A response or class  $Y$  from inputs  $X_1, X_2, \dots, X_p$  is required to be predicted. This is done by growing a binary tree. At each node in the tree, a test to one of the inputs, say  $X_i$  is applied. Depending on the outcome of the test, either the left or the right sub-branch of the tree is selected. Eventually a leaf node is reached, where a prediction is made. This prediction aggregates or averages all the training data points which reach that leaf. A model is obtained by using each of the independent variables. For each of the individual variables, mean squared error is used to determine the best split. The maximum number of features to be considered at each split is set to the total number of features [40–42].

352

### 2.2.6. Nearest neighbor

353  
354

Nearest neighbor neural network (k-NN) is a type of instance-based learning, where a function is only approximated locally and all computation is delayed until classification [37]. The k-NN algorithm is one of the simplest machine learning algorithms. For both classification and regression, it can be useful to assign a weight to the contributions of the neighbors, so that the nearest neighbors contribute more to the average than the distant ones. For example, in a common weighting arrangement, each neighbor is given a weight of  $1/d$ , where  $d$  is the distance to the neighbor [43].

361

### 2.2.7. Markov chain

362  
363

In forecasting domain, some authors have tried to use the so-called Markov processes, specifically the Markov chains. A Markov process is a stochastic process with the Markov property, which means that given the present state, future states are independent of the past states[44]. Expressed differently, the description of the present state fully captures all the information that could affect the future evolution of the process. In this, future states are reached through a probabilistic process instead of a deterministic one. The proper use of these processes needs to calculate initially the matrix of transition states. The transition probability of state  $i$  to the state  $j$  is defined by  $p_{i,j}$ . The family of these numbers is called the transition matrix of the Markov chain  $R$  [27].

372

## 373           **2.3. Unsupervised learning**

374

375   In contrary with supervised learning model, an unsupervised learning model does not need an “expert”  
376   intervention and the model is able to find hidden structure in its inputs without knowledge of outputs  
377   [45]. Unsupervised learning is similar to the problem of density estimation in statistics. Unsupervised  
378   learning however, also incorporates many other techniques that seek to summarize and explain the key  
379   features of the data. Many methods normally employed in unsupervised learning are based on data  
380   mining methods used to pre-process data.

381

### 382           **2.3.1. k-Means and k-Methods Clustering**

383

384   k-means clustering is a method of vector quantization, originally derived from signal processing,  
385   which is popular for cluster analysis in data mining. k-means clustering aims to partition n  
386   observations into k clusters in which each observation belongs to the cluster with the nearest mean,  
387   serving as a prototype of the cluster. k-Means algorithms are focused on extracting useful information  
388   from the data with the purpose of modelling the time series behaviour and find patterns of the input  
389   space by clustering the data. Furthermore, nonlinear autoregressive (NAR) neural networks are  
390   powerful computational models for modelling and forecasting nonlinear time series [46]. A lot of  
391   methods of clustering are available; the interested reader can see [47] for more information.

392

### 393           **2.3.2. Hierarchical Clustering**

394

395   In data mining and statistics, hierarchical clustering (also called hierarchical cluster analysis) is a  
396   method of cluster analysis which seeks to build a hierarchy of clusters. Hierarchical clustering creates  
397   a hierarchy of clusters which can be represented in a tree structure called “*dendrogram*” which  
398   includes both roots and leaves. The root of the tree consists of a single cluster which contains all  
399   observations, whereas the leaves correspond to individual observations. Algorithms for hierarchical  
400   clustering are generally either agglomerative, in which the process starts from the leaves and  
401   successively merges clusters together; or divisive, in which the process starts from the root and  
402   recursively splits the clusters [48]. Any function which does not have a negative value can be used as a  
403   measure of similarity between pairs of observations. The choice of which clusters to merge or split, is  
404   determined by a linkage criterion that is a function of the pairwise distances between observations. It  
405   should be noted that cutting the tree at a given height will give a clustering at a selected precision.

406

407

### 2.3.3. Gaussian Mixture Models

408

409 Gaussian Processes (GPs) are a relatively recent development in non-linear modelling [49]. A GP is a  
 410 generalization of a multivariate Gaussian distribution to infinitely many variables. A multivariate  
 411 Gaussian distribution  $\mathbf{d}$  is fully specified by a mean vector  $\mu$  and covariance matrix  $\Sigma$ , e.g.  $\mathbf{d} \sim \mathcal{N}(\mu, \Sigma)$ .  
 412 The key assumption in GP modelling is that the data  $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$  can be represented as a sample from a  
 413 multivariate Gaussian distribution e.g. the observations  $\mathbf{y} = (y_1, y_2, \dots, y_n) \sim \mathcal{N}(\mu, \Sigma)$ . In order to better  
 414 introduce GPs, the case is often restricted to one scalar input variable  $x$ . As the data are often noisy  
 415 usually from measurement errors, each observation  $y$  can be thought of as an underlying function  $f(x)$   
 416 with added independent Gaussian noise with variance  $\sigma_n^2$ , i.e.,  $y = f(x) + \mathcal{N}(0, \sigma_n^2)$ . As a GP is an  
 417 extension of a multivariate Gaussian distribution, it is fully specified by a mean function  $m(x)$  and a

418 covariance function  $k_f(x, x')$ . Expressed in a different way, the function can be modelled by  
 419 a GP  $f(x) \sim \mathbf{GP}(m(x), k_f(x, x'))$ . The setting of a covariance function permits to relate one  
 420 observation  $y_p$  to another one  $y_q$ . A popular choice of covariance function is the squared exponential

421  $k_{se}(x_p, x_q) = \sigma_f^2 \exp\left[-\frac{(x_p - x_q)^2}{2l^2}\right]$ . As predictions are usually made using noisy measurements, the

422 covariance between two observations can be stated as

423  $cov(y_p, y_q) = k_{se}(x_p, x_q) + \delta_{pq}\sigma_n^2 = \sigma_f^2 \exp\left[-\frac{(x_p - x_q)^2}{2l^2}\right] + \delta_{pq}\sigma_n^2$ ,  $\delta_{pq}$  is the Kronecker delta.  $\sigma_f^2$  and  $l$

424 are called hyperparameters of the covariance function and they control the model complexity and can  
 425 be learned (or optimized) from the training data at hand [49]. For instance, in prediction studies, given  
 426 the training database  $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ , the vector of  $n_*$  forecasted irradiation for horizon  $h$  for new test  
 427 inputs  $\mathbf{X}_*$  is given by the mean of the predictive Gaussian distribution predictions.

428

429

### 2.3.4. Cluster Evaluation

430

431 Typical objective functions in clustering formalize the goal of attaining high intra-cluster similarity  
 432 and low inter-cluster similarity [50]. This is an internal criterion for the quality of a clustering. Good  
 433 scores on an internal criterion do not necessarily mean a good effectiveness in an application. An



434 alternative to internal criteria is the direct evaluation of the application of interest. For search result  
435 clustering, the amount of the time a user is required to find an answer with different clustering  
436 algorithms may be required. This is the most direct evaluation, but it is time consuming, especially if  
437 large number of studies are necessary.

438

## 439 **2.4. Ensemble learning**

440

441 The basic concept of ensemble learning is to train multiple base learners as ensemble members and  
442 combine their predictions into a single output that should have better performance on average than any  
443 other ensemble member with uncorrelated error on the target data sets [51]. Supervised learning  
444 algorithms are usually described as performing the task of searching through a hypothesis space to  
445 find a suitable hypothesis that can perform good predictions for a particular problem. Even if the  
446 hypothesis space contains hypotheses that are very well-matched for a particular problem, it may be  
447 very difficult to find which one is the best. Ensembles combine multiple hypotheses to create a better  
448 hypothesis. The term ensemble is usually used for methods that generate multiple hypotheses using the  
449 same base learner. Fast algorithms such as decision trees are usually used with ensembles, although  
450 slower algorithms can also benefit from ensemble techniques. Evaluating the prediction accuracy of an  
451 ensemble typically requires more computation time than evaluating the prediction accuracy of a single  
452 model, so ensembles may be considered as a way to compensate for poor learning algorithms by  
453 performing much more computation. The general term of multiple classifier systems covers also  
454 hybridization of hypotheses that are not induced by the same base learner. The interested reader can  
455 see [52] for more details about ensemble learning.

456

### 457 **2.4.1. Boosting**

458

459 An ensemble model uses decision trees as weak learners and builds the model in a stage-wise manner  
460 by optimizing a loss function [34, 35]. Boosting emerged as a way of combining many weak  
461 classifiers to produce a powerful “committee”. It is an iterative process that gives more and more  
462 importance to bad classification. Simple strategy results in dramatic improvements in classification  
463 performance. To do so, a boosting autoregression procedure is applied at each horizon on the residuals  
464 from the recursive linear forecasts using a so-called weak learner, which is a learner with large bias  
465 relative to variance.

466  
467

### 2.4.2. Bagging

468 Bootstrap aggregating, also called bagging used in statistical classification and regression, is a  
469 machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine  
470 learning algorithms. The algorithm also reduces variance and helps to prevent overfitting. Although it  
471 is generally applied to decision tree methods, it can be used with any type of learning method.  
472 Bagging is a special case of the model averaging approach. Bagging predictors is generally used to  
473 generate multiple versions of a predictor and using them to get an aggregated predictor. The  
474 aggregation averages all the versions when predicting a numerical result and does a plurality vote to  
475 predict a class. The multiple versions are formed by making bootstrap replicates of the learning set and  
476 using them as new learning sets [55].

477

478  
479

### 2.4.3. Random Subspace

480 The machine learning tool that is used in the proposed methodology is based on Random Forests,  
481 which consists of a collection, or ensemble of a multitude of decision trees, each one built from a  
482 sample drawn with replacement (a bootstrap sample) from a training set, is the group of outputs.  
483 Furthermore, only a random subset of variables is used when splitting a node during the construction  
484 of a tree. As a consequence, the final nodes (or leaves), may contain one or several observations. For  
485 regression problems, each tree is capable of producing a response when presented with a set of  
486 predictors, being the conditional mean of the observations present on the resulting leaf. The  
487 conditional mean is typically approximated by a weighted mean. As a result of the random  
488 construction of the trees, the bias of the forest generally slightly increases with respect to the bias of a  
489 single non-random tree but, due to the averaging its variance decreases, frequently more than  
490 compensating for the increase in bias, hence yielding an overall better model. Finally, the responses of  
491 all trees are also averaged to obtain a single response variable for the model, and here as well a  
492 weighted mean is used [56]. Substantial improvements in classification accuracy were obtained from  
493 growing an ensemble of trees and letting them vote for the most popular class. To grow these  
494 ensembles, often random vectors are generated which govern the growth of each tree in the ensemble.  
495 One of the first examples used is bagging, in which to grow each tree a random selection (without  
496 replacement) is made from the examples contained in the training set [57–60].

497

498 **2.4.4. Predictors ensemble**

499

500 Current practice suggests that forecasts should be composed either by a number of simple- say  
501 “conventional” forecasts- or produce a simple forecast from other simple forecasts (not only point  
502 forecasts, but also probabilistic). This leads to gains in performance, relative to the contributing  
503 forecasts. In the case of statistical models, realizations coming from the same technology (for example  
504 the same neural network architecture) trained multiple times, or using different samples of the dataset;  
505 or different technologies. Once “first stage forecasts” are available, different combination approaches  
506 are possible. The simplest approach is averaging of results given by different methods. A more general  
507 approach assigns a weight to each of the contributing methods, for each time horizon, depending on  
508 different criteria and with different weighting policies. Simple forecasts can be seen as different  
509 perceptions of the same true state. In this way, approaches of imperfect sensor data fusion should also  
510 be valid to perform a combination of forecasts. Ensemble-based artificial neural networks and other  
511 machine learning technics have been used in a number of studies in global radiation modeling and  
512 provided better performance and generalization capability compared to conventional regression  
513 models [28, 42].

514

515 **3. Evaluation of model accuracy**

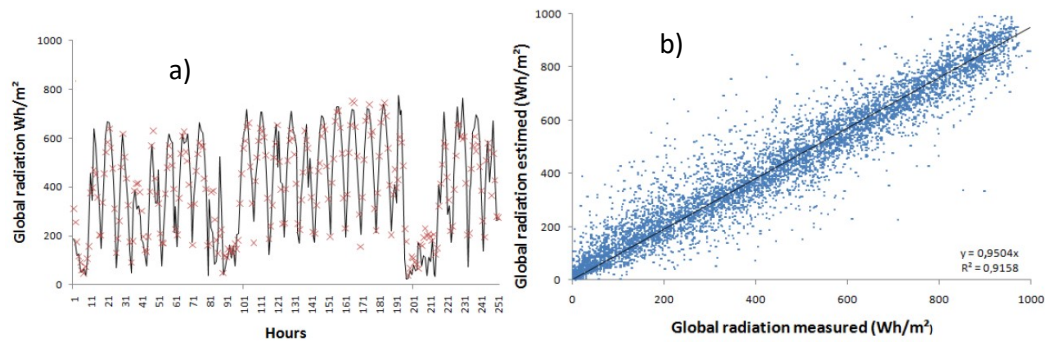
516

517 Evaluation, generally, measures how good something is. This evaluation is used at various steps of the  
518 model development as for example during the evaluation of the forecasting model itself (during the  
519 training of a statistical model for example), for judging the improvement of the model after some  
520 modifications and for comparing various models. As previously mentioned, this performance  
521 comparison is not easy for various reasons such as different forecasted time horizons, various time  
522 scale of the predicted data and variability of the meteorological conditions from one site to another  
523 one. It works by comparing the forecasted outputs  $\hat{y}$  (or predicted time series) with observed data  $y$   
524 (or observed or measured time series) which are also measured data themselves linked to an error (or  
525 precision) of a measure.

526 Graphic tools are available for estimating the adequacy of the model with the experimental  
527 measurements such as:

- 528 - Time series of predicted irradiance in comparison with measured irradiance which allows to  
529 visualize easily the forecast quality. In Fig. 4a, as an example, a high forecast accuracy in  
530 clear-sky situations and a low one in partly cloudy situations can be seen.

- 531 - Scatter plots of predicted over measured irradiance (see an example in Fig. 4b) which can
- 532 reveal systematic bias and deviations depending on the irradiance conditions and show the
- 533 range of deviations that are related to the forecasts.
- 534 - Receiver Operating Characteristic (ROC) curves which compare the rates of true positives and
- 535 false positive.



536

537

538 Figure 4: a) Time series of predicted and measured global irradiance for 2008 in Ajaccio (France); b)

539 Scatter plot of predicted vs. measured global irradiance in Ajaccio (France); c) Example of ROC curve

540 (an ideal ROC curve is near the upper left corner).

541 No standard evaluation measures are accepted, which makes the comparison of the forecasting

542 methods difficult. Sperati et al. [62] presented a benchmarking exercise within the framework of the

543 European Actions Weather Intelligence for Renewable Energies (WIRE) with the purpose of

544 evaluating the performance of state of the art models for short term renewable energy forecasting. This

545 study is a very good example of reliability parameter utilization. They concluded that: “More work

546 using more test cases, data and models needs to be performed in order to achieve a global overview of

547 all possible situations. Test cases located all over Europe, the US and other relevant countries should

548 be considered, in an effort to represent most of the possible meteorological conditions”. This paper

549 illustrates very well the difficulties of performance comparisons.

550 The usually used statistics include the following:

551 The mean bias error (MBE) represents the mean bias of the forecasting:

$$552 \quad MBE = \frac{1}{N} \times \sum_{i=1}^N (\hat{y}(i) - y(i)) \quad (7)$$

553 with  $\mathcal{Y}$  being the forecasted outputs (or predicted time series),  $y$  the observed data (or observed or  
554 measured time series) and  $N$  the number of observations. The forecasting will under-estimate or over-  
555 estimate the observations. Thus, MBE is not a good indicator for the reliability of a model because  
556 the errors compensate each other but it allows to see how much it overestimates or underestimates.

557 The mean absolute error (MAE) is appropriate for applications with linear cost functions, i.e., where  
558 the costs resulting from a poor forecast are proportional to the forecast error:

$$559 \quad MAE = \frac{1}{N} \times \sum_{i=1}^N |\hat{y}(i) - y(i)| \quad (8)$$

560 The mean square error (MSE) uses the squared of the difference between observed and predicted  
561 values. This index penalizes the highest gaps:

$$562 \quad MSE = \frac{1}{N} \times \sum_{i=1}^N (\hat{y}(i) - y(i))^2 \quad (9)$$

563 MSE is generally the parameter which is minimized by the training algorithm.

564 The root mean square error (RMSE) is more sensitive to big forecast errors, and hence is suitable for  
565 applications where small errors are more tolerable and larger errors cause disproportionately high  
566 costs, as for example in the case of utility applications [22]. It is probably the reliability factor that is  
567 most appreciated and used:

$$568 \quad RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^N (\hat{y}(i) - y(i))^2} \quad (10)$$

569 The mean absolute percentage error (MAPE) is close to the MAE but each gap between observed and  
570 predicted data is divided by the observed data in order to consider the relative gap.

$$571 \quad MAPE = \frac{1}{N} \times \sum_{i=1}^N \left| \frac{\hat{y}(i) - y(i)}{y(i)} \right| \quad (11)$$

572 This index has a disadvantage that it is unstable when  $y(i)$  is near zero and it cannot be defined for  
573  $y(i)=0$ .

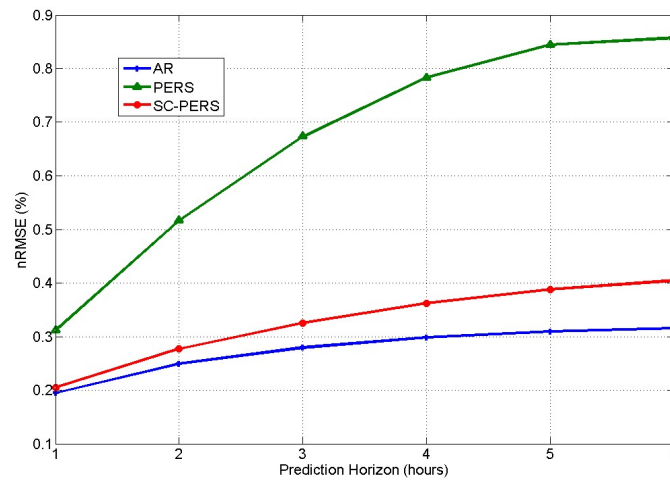
574 Often, these errors are normalized particularly for the RMSE; as reference the mean value of  
 575 irradiation is generally used but other definitions can be found:

$$576 \quad nRMSE = \frac{\sqrt{\frac{1}{N} \times \sum_{i=1}^N (\hat{y}(i) - y(i))^2}}{\bar{y}} \quad (12)$$

577 With  $\bar{y}$  being the mean value of  $y$ . Other indices exist and can be used as the correlation coefficient  $R$   
 578 (Pearson Coefficient), or the index of agreement ( $d$ ) which are normalized between 0 and 1.

579 As the forecast accuracy strongly depends on the location and time period used for evaluation and on  
 580 other factors, it is difficult to evaluate the quality of a forecast from accuracy metrics alone. Then, it is  
 581 best to compare the accuracy of different forecasts against a common set of test data [63]. “Trivial”  
 582 forecast methods can be used as a reference [22], the most common one is the persistence model  
 583 (“things stay the same”, [64]) where the forecast is always equal to the last known data point. The  
 584 persistence model is also known in the forecasting literature as the naive model or the RandomWalk (a  
 585 mathematical formalization of a path that consists of a succession of random steps). The solar  
 586 irradiance has a deterministic component due to the geometrical path of the sun. This characteristic  
 587 may be added as a constraint to the simplest form of persistence in considering as an example, the  
 588 measured value of the previous day or the previous hour at the same time as a forecast value. Other  
 589 common reference forecasts include those based on climate constants and simple autoregressive  
 590 methods. Such comparison with referenced NWP model is shown in Figure 5. Generally, after 1 h the  
 591 forecast is better than persistence. For forecast horizons of more than two days, climate averages show  
 592 lower errors and should be preferred.

593



594

595 Figure 5: Relative RMSE of forecasts(persistence, auto regression, and scaled persistence) and of  
 596 reference models depending on the forecast horizon [18]

597 Classically, a comparison of performance is performed with a reference model and to do it, a skill  
 598 factor is used. The skill factor or skill score defines the difference between the forecast and the  
 599 reference forecast normalized by the difference between a perfect and the reference forecast [18]:

$$600 \text{ SkillScore} = \frac{\text{Metric}_{\text{forecasted}} - \text{Metric}_{\text{reference}}}{\text{Metric}_{\text{perfectforecast}} - \text{Metric}_{\text{reference}}} = 1 - \frac{\text{MSE}_{\text{forecastd}}}{\text{MSE}_{\text{reference}}} \quad (13)$$

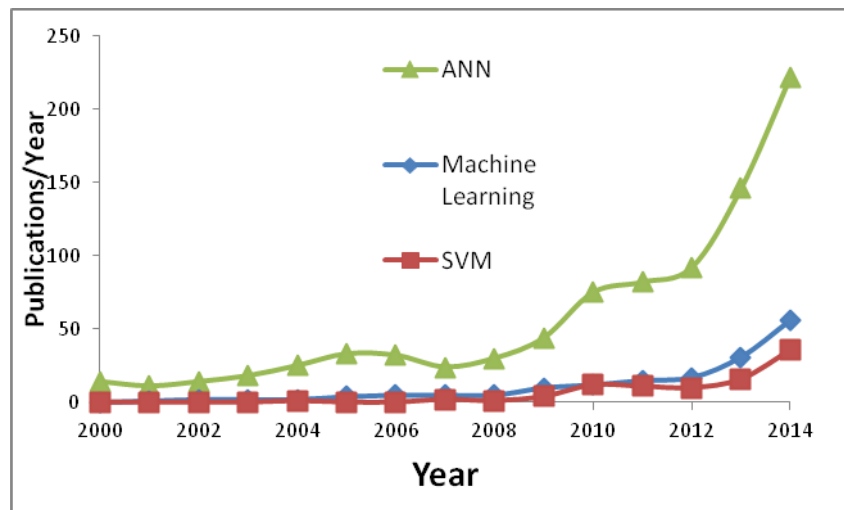
601 Its value thus ranges between 1 (perfect forecast) and 0 (reference forecast). A negative value indicates  
 602 a performance which is even worse compared to the reference. Skill scores may be applied not only  
 603 for comparison with a simple reference model but also for inter-comparisons of different forecasting  
 604 approaches (improvement scores). As an example, Bacher et al. [65] reported an improvement in  
 605 RMSE by 36% with respect to persistence, then the RMSE skill score with respect to persistence was  
 606 equal to 0.36. Benchmarking can also be used to identify conditions under which forecasts perform  
 607 relatively well. Numerous benchmarking were realized in US [17], Canada and European countries [8]  
 608 and in Italy [62]. Note that solar forecasting methods in literature go beyond point forecasts only.  
 609 Probabilistic forecasts are also widely used and are often more practical solutions to solar energy  
 610 needs. The evaluation of probabilistic/prediction interval forecasts is different and metrics used are not  
 611 limited to the presented one (see use of prediction intervals [66,67]).

612

#### 613 4. Machine learning forecasters' comparison

614

615 Before presenting the results related to the machine learning method in order to predict the global  
 616 radiation, Figure 6 shows the number of times the term ANN, machine learning and SVM/SVR are  
 617 referenced in the five main journals of solar energy prediction (*Solar Energy*, *Energy*, *Applied Energy*,  
 618 *Renewable Energy* and *Energy Conversion and Management*).



619

620 Figure 6: Number of time the ANN, machine learning and SVM terms have been used in the original  
621 articles

622

623 All three terms are more and more used in literature. It can be seen the ANN is the mostly used  
624 method in global radiation forecasting.

625

#### 626 **4.1. The ANN case**

627

628 Reviews of this kind of prediction methods are presented in [28, 47]; the interested reader can find  
629 more information in these papers. Neural networks have been studied on many parts of the world and  
630 researchers have shown the ability of these techniques to accurately predict the time series of  
631 meteorological data. It is essential to distinguish between two types of studies; modeling with  
632 multivariate regression and time series prediction. Indeed, MLP are quite regularly used for their  
633 property of "universal approximation", capable of non-linear regression. In 1999, for the first time an  
634 author presented the prediction of global solar radiation time series via MLP. Kemmoku [69] uses a  
635 method based on MLP to predict the irradiation for the next day. The results show a prediction error  
636 (MAPE) of 18.5% in summer and 21.8% in winter. From all the articles related to ANN [28, 47], the  
637 errors associated with predictions (monthly, daily, hourly and minute) are between 5% and 15%. In  
638 Mellit and Kalogirou review article [68], can be seen that 79% of Artificial Intelligence (AI) methods  
639 used in weather prediction data are based on a connectionist approach (ANN). The use of fuzzy logic  
640 (5%), Adaptive neuro fuzzy inference system (ANFIS) account for 5% of the papers, networks  
641 coupling wavelet decomposition and ANN for 8% and mix ANN/ Markov chain for a small 3%.  
642 Summing up the use of ANN, especially the MLP represents a large majority of research works. This  
643 is the most commonly used technique. Other methods are used only sporadically. According to  
644 published literature, the parameters that influence the prediction are many, so it is difficult to adopt the  
645 results from other studies. Taking into account this fact, it may be interesting to test methods or  
646 parameters even though they have not necessarily been proven in other studies. Based on what is said  
647 above, all parameters inherent to the MLP or ARMA method must be studied for each tested site.  
648 Even if MLP seems better than ARMA, in a few cases the reverse is presented. It is in this context that  
649 other machine learning methods have been studied. Note that forecast performance in terms of  
650 statistical metrics (RMSE, forecast skills) depend not only on the weather conditions (variability) but  
651 also on the forecast horizons.

652



653 **4.2. Single machine learning method**

654

655 Table 3 shows the results of machine learning methods used in global solar radiation prediction, direct  
 656 normal irradiance and diffuse irradiance. A lot of papers are not cited here. The interested reader can  
 657 see some well written review papers related to this topic [26, 47, 49].

658 Table 3. List of representative papers related to the global solar radiation forecasting using single  
 659 machine learning methods

References	Location	Horizon	Evaluation criteria	Dataset	Results
[71]	Canada	18 hours	NA	exogenous	Regression tree>linear regression
[20]	Greece	1 hour	NA	endogenous	ANN=AR
[72]	Argentina	1 day	NA		Generalized regression is useful
[73]	China	NA	NA	Endogenous and influents factors	Regression tree > ANN > linear regression
[27]	France	D+1	nRMSE=21%	exogenous	ANN>AR>k-NN>Bayesian>Markov
[74]	USA	24 hours	nRMSE=17.7 %	exogenous	ANN>persistence
[75]	Spain	1 day	NA	exogenous	ANN=generalized regression
[76]	Benchmark	various	MAPE=18.95 %	exogenous	MIMO-ACFLIN strategy (lazy learning) is the winner
[77]	Turkey	10 min	nRMSE=18%	exogenous	k-NN>ANN
[78]	Italia	1 hour	NA	exogenous	SVM>ANN>k-NN>persistence
[79]	Japan	1 hour	NA	exogenous	Regression tree interesting to select variables
[80]	Nigeria	NA	nRMSE=24%	exogenous	ANN=regression tree
[81]	Spain	1 day	NA	exogenous	SVM>persistence
[59]	Australia	1 min	MAPE=38%	exogenous	Random forest>linear regression
[82]	Canada	1 hour	NA	exogenous	SVM>NWP
[83]	Macao	1 day	MAPE=11.8 %	exogenous	ANN>SVM>k-NN>linear regression
[84]	Spain	1 day	NA	exogenous	Extreme machine learning is useful
[85]	Benchmark	10 min	nRMSE=10%	exogenous	Random Forest>SVM>generalized regression>boosting>bagging>persistence
[56]	Spain	24 hours	nMAE=3.73% - 9.45%	exogenous	Quantile regression forests coupled with NWP give a good accuracy for PV prediction
[86]	Germany	1 hour	nRMSE=6.2%	exogenous	SVR>k-NN
[18]	French islands	1 hour	nRMSE=19.6 %	exogenous	ANN=Gaussian=SVM>persistence
[87]	Italia	1 hour	NA	exogenous	SVR>ANN>AR>k-NN>persistence
[88]	benchmark	1 hour	nRMSE=13%	exogenous	Regression tree>NWP
[43]	USA	30 min	Skill over persistence =23.4%	Exogenous	k-NN>persistence

660

661 In this list, it can be seen that papers using SVR, ANN, k-NN, regression tree, boosting, bagging or  
 662 random forests give systematically better results than classical regression methods. ANN and SVM  
 663 give similar results in term of prediction, but it can be concluded that SVM is more easy to use than  
 664 ANN; the optimization step is automatic while it is very complex in the ANN case. Therefore, it is

665 maybe preferable to use SVM rather than ANN. All the methods related to the use of regression trees  
666 or similar methods (boosting, bagging or random forest) are rarely used but give excellent results. It is  
667 not easy at this stage to draw a conclusion, but in the next 5 years, it is probable that these methods  
668 may become the reference in term of irradiation prediction. An alternative to all the previous methods  
669 is certainly k-NN, but some more publications are necessary to conclude that it is a good forecaster.  
670 Actually, it is very difficult to propose a ranking of machine learning methods although SVR,  
671 regression trees, boosting, bagging and random forest seem the most efficient. To overcome this  
672 problem of ranking, some authors do not hesitate to combine single predictors.

673

### 674 4.3. Ensemble of predictors

675

676 There are a lot of solutions which combine predictors as it is shown in Table 4. In these papers we can  
677 see that often it is ANN who is used in order to construct ensemble of predictors (>70% of the cases).

678

679 Table 4. list of representative papers related to the global radiation forecasting combining machine  
680 learning methods

References	Location	Horizon	Evaluation criteria	Dataset	Results
[41]	Japan	1 day	nMAE=1.75%	exogenous	{regression tree-ANN}>ANN
[89]	China	1 hour	R <sup>2</sup> =0.72	exogenous	{ANN-wavelet}>ANN
[16]	USA	1 hour	nRMSE=26%	endogenous	{ARMA}>ANN
[90]	Japan	1 hour	MAPE=4%	exogenous	{ANN}>ANN
[91]	Spain	1 hour	NA	exogenous	{SVM-k-NN}>climatology
[92]	USA	1 hour	NA	exogenous	Bayesian>{SVM-ANN}
[93]	Australia	6 hour	NA	exogenous	{ANN-least median square}>least median square>ANN>SVM
[94]	Italia	10 min	nRMSE=9.4%	exogenous	{SARIMA-SVM}>SARIMA>SVM
[95]	USA	10 min	Skill over persistence=20%	exogenous	{GA-ANN}>ANN
[96]	Czech rep	10 min	NA	exogenous	{ANN-SVM}>SVM>ANN
[24]	USA	1 day	NA	exogenous	{ANN-linear regression}>ANN>linear regression
[97]	USA	1 day	NA	exogenous	{LSR-ANN}>regularized LSR=ordinary LSR
[61]	UAE	10 min	rRMSE=9.1%	exogenous	{ANNs}>ANN
[98]	USA	30 min	NA	exogenous	{PCA-gaussian process}>NWP
[99]	Singapore	NA	NA	endogenous	{GA-kmean-ANN}>ANN>ARMA
[100]	Malaysia	1 hour	nRMSE=5%	exogenous	{GA-SVM-ANN-ARIMA}>SVM>ANN>ARIMA
[101]	Taiwan	1 day	nMAE=3%		{ANN-SVM}>SVM>ANN
[102]	USA	10 min	Skill over persistence=6%	exogenous	{GA-ANN}>persistence
[103]	Italia	1 day	MAPE=6%	exogenous	SVM>linear model
[104]	USA	1 hour	nRMSE=22%	exogenous	{ANN-SVM}>ARMA

[105]	USA	1 hour	NA	exogenous	{SVR}>SVR>{SVR-PCA}>ARIMA>linear regression
-------	-----	--------	----	-----------	---------------------------------------------

681

682 As it can be seen, systematically the ensemble of predictors gives better results than single predictors  
683 but again the best methodology of hybridization is not really defined. A lot of more works are  
684 necessary in order to propose a robust method, or maybe to prove that all the methods are equivalent,  
685 but certainly this cannot be done with the limited cases presented here.

686

687 **5. Conclusions and outlook**

688

689 As shown in the present paper, many methods and types of methods are available. There are a lot of  
690 methods to estimate the solar radiation, some are often used (ANN, ARIMA, naive methods), others  
691 begin to be used (SVM, SVR, k-mean) more frequently and other are rarely used (boosting, regression  
692 tree, random forest, etc.). In some cases, one is the best and in others it is the reverse. As a conclusion,  
693 it can be said that the ANN and ARIMA methods are equivalent in term of quality of prediction in  
694 certain variability conditions, but the flexibility of ANN as universal nonlinear approximation makes  
695 them more preferable than classical ARIMA. Generally, the accuracy of these methods depends on the  
696 quality of the training data. The three methods that should be generally used in the next years are the  
697 SVM, regression trees and random forests, as the results given are very promising and some  
698 interesting studies will certainly be produced the next few years. Actually, considering the published  
699 papers, these methods yield similar error statistics. The implementation of the methods may have more  
700 to do with the errors reported in the literature than the methods themselves. For example, when the  
701 autocorrelation of the error is reduced white noise for the same inputs, SVM, SVR, regression trees or  
702 random forests perform very similarly, with no statistical differences between them. The second point  
703 which can be seen from Table 4 is the fact that the predictor ensemble methodology is always better  
704 than simple predictors. This shows the way that the problem must be studied whereas the simple  
705 forecast methodology using only one stochastic method (above all ANN and ARIMA) should tend to  
706 disappear. In the present paper the deep learning, which is a branch of machine learning based on a set  
707 of algorithms that attempt to model high-level abstractions in data by using model architecture, with  
708 complex structures or otherwise, composed of multiple non-linear transformations, are not taken into  
709 account. This research area is very recent and there is not enough experience, but in the future this  
710 kind of methodology may outperform conventional methods, as is already the case in other predicting  
711 domains (air quality, wind, economy, etc.). As a consequence, forecasts reached through various  
712 methods can be calculated in order to satisfy the various needs. The question then arises how they will  
713 be put together. The answer is clearly not trivial because the various resulting forecasts show

714 differences on many points. Moreover, some of them will be associated with confidence intervals  
715 which should also be merged.

## 716 **Acknowledgement:**

717 This work was carried out in the framework of the Horizon  
718 2020 project (H2020-LCE-2014-3 - LCE-08/2014 - 646529) TILOS “Technology  
719 Innovation for the Local Scale, Optimum Integration of Battery Energy  
720 Storage”. The authors would like to acknowledge the financial support of the  
721 European Union.

## 722 **Bibliography**

723

- 724 [1] B. Espinar, J.-L. Aznarte, R. Girard, A.M. Moussa, G. Kariniotakis, Photovoltaic Forecasting: A  
725 state of the art, in: OTTI - Ostbayerisches Technologie-Transfer-Institut, 2010: p. Pages 250-255-  
726 ISBN 978-3-941785-15-1. [https://hal-mines-paristech.archives-ouvertes.fr/hal-](https://hal-mines-paristech.archives-ouvertes.fr/hal-00771465/document)  
727 [00771465/document](https://hal-mines-paristech.archives-ouvertes.fr/hal-00771465/document) (accessed March 4, 2015).
- 728 [2] V. Lara-Fanego, J.A. Ruiz-Arias, D. Pozo-Vázquez, F.J. Santos-Alamillos, J. Tovar-Pescador,  
729 Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain), *Sol.*  
730 *Energy*. 86 (2012) 2200–2217. doi:10.1016/j.solener.2011.02.014.
- 731 [3] A. Moreno-Munoz, J.J.G. De la Rosa, R. Posadillo, F. Bellido, Very short term forecasting of solar  
732 radiation, in: 33rd IEEE Photovolt. Spec. Conf. 2008 PVSC 08, 2008: pp. 1–5.  
733 doi:10.1109/PVSC.2008.4922587.
- 734 [4] D. Anderson, M. Leach, Harvesting and redistributing renewable energy: on the role of gas and  
735 electricity grids to overcome intermittency through the generation and storage of hydrogen,  
736 *Energy Policy*. 32 (2004) 1603–1614. doi:10.1016/S0301-4215(03)00131-9.
- 737 [5] M. Paulescu, E. Paulescu, P. Gravila, V. Badescu, *Weather Modeling and Forecasting of PV*  
738 *Systems Operation*, Springer London, London, 2013. [http://link.springer.com/10.1007/978-1-](http://link.springer.com/10.1007/978-1-4471-4649-0)  
739 [4471-4649-0](http://link.springer.com/10.1007/978-1-4471-4649-0) (accessed March 4, 2015).
- 740 [6] H.M. Diagne, P. Lauret, M. David, Solar irradiation forecasting: state-of-the-art and proposition  
741 for future developments for small-scale insular grids, in: n.d. [https://hal.archives-](https://hal.archives-ouvertes.fr/hal-00918150/document)  
742 [ouvertes.fr/hal-00918150/document](https://hal.archives-ouvertes.fr/hal-00918150/document) (accessed March 4, 2015).
- 743 [7] A. Hammer, D. Heinemann, E. Lorenz, B. Lückehe, Short-term forecasting of solar radiation: a  
744 statistical approach using satellite data, *Sol. Energy*. 67 (1999) 139–150. doi:10.1016/S0038-  
745 [092X\(00\)00038-4](https://doi.org/10.1016/S0038-092X(00)00038-4).
- 746 [8] E. Lorenz, J. Remund, S.C. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J.A. Ruiz-Arias, V.L.  
747 Fanego, L. Ramirez, M.G. Romeo, others, Benchmarking of different approaches to forecast  
748 solar irradiance, in: 24th Eur. Photovolt. Sol. Energy Conf. Hambg. Ger., 2009: p. 25.  
749 [http://task3.iea-shc.org/data/sites/1/publications/24th\\_EU\\_PVSEC\\_5BV.2.50\\_lorenz\\_final.pdf](http://task3.iea-shc.org/data/sites/1/publications/24th_EU_PVSEC_5BV.2.50_lorenz_final.pdf)  
750 (accessed March 4, 2015).
- 751 [9] M. Saguan, Y. Perez, J.-M. Glachant, L’architecture de marchés électriques : l’indispensable  
752 marché du temps réel d’électricité, *Rev. Déconomie Ind.* (2009) 69–88. doi:10.4000/rei.4053.
- 753 [10] B. Elliston, I. MacGill, The potential role of forecasting for integrating solar generation into the  
754 Australian national electricity market, in: *Sol. 2010 Proc. Annu. Conf. Aust. Sol. Energy Soc.*,  
755 2010. [http://solar.org.au/papers/10papers/10\\_117\\_ELLISTON.B.pdf](http://solar.org.au/papers/10papers/10_117_ELLISTON.B.pdf) (accessed March 4, 2015).
- 756 [11] E. Lorenz, J. Kühnert, D. Heinemann, Short term forecasting of solar irradiance by combining  
757 satellite data and numerical weather predictions, in: *Proc. 27th Eur. Photovolt. Sol. Energy*  
758 *Conf. Valencia Spain, 2012*: pp. 4401–440.
- 759 <http://meetingorganizer.copernicus.org/EMS2012/EMS2012-359.pdf> (accessed March 4, 2015).

- 760 [12] D. Heinemann, E. Lorenz, M. Girodo, Forecasting of solar radiation, *Sol. Energy Resour. Manag.*  
761 *Electr. Gener. Local Level Glob. Scale Nova Sci. Publ. N. Y.* (2006). [https://www.uni-](https://www.uni-oldenburg.de/fileadmin/user_upload/physik/ag/ehf/enmet/publications/solar/conference/2005/Forecasting_of_Solar_Radiation.pdf)  
762 [oldenburg.de/fileadmin/user\\_upload/physik/ag/ehf/enmet/publications/solar/conference/200](https://www.uni-oldenburg.de/fileadmin/user_upload/physik/ag/ehf/enmet/publications/solar/conference/2005/Forecasting_of_Solar_Radiation.pdf)  
763 [5/Forecasting\\_of\\_Solar\\_Radiation.pdf](https://www.uni-oldenburg.de/fileadmin/user_upload/physik/ag/ehf/enmet/publications/solar/conference/2005/Forecasting_of_Solar_Radiation.pdf) (accessed March 4, 2015).
- 764 [13] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renné, T.E. Hoff, Validation of short and  
765 medium term operational solar radiation forecasts in the US, *Sol. Energy*. 84 (2010) 2161–2172.  
766 doi:10.1016/j.solener.2010.08.014.
- 767 [14] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, Review of solar irradiance forecasting  
768 methods and a proposition for small-scale insular grids, *Renew. Sustain. Energy Rev.* 27 (2013)  
769 65–76. doi:10.1016/j.rser.2013.06.042.
- 770 [15] E. Lorenz, A. Hammer, D. Heinemann, Short term forecasting of solar radiation based on  
771 satellite data, in: *EUROSUN2004 ISES Eur. Sol. Congr.*, 2004: pp. 841–848. [https://www.uni-](https://www.uni-oldenburg.de/fileadmin/user_upload/physik/ag/ehf/enmet/publications/solar/conference/2004/eurosun/short_term_forecasting_of_solar_radiation_based_on_satellite_data.pdf)  
772 [oldenburg.de/fileadmin/user\\_upload/physik/ag/ehf/enmet/publications/solar/conference/200](https://www.uni-oldenburg.de/fileadmin/user_upload/physik/ag/ehf/enmet/publications/solar/conference/2004/eurosun/short_term_forecasting_of_solar_radiation_based_on_satellite_data.pdf)  
773 [4/eurosun/short\\_term\\_forecasting\\_of\\_solar\\_radiation\\_based\\_on\\_satellite\\_data.pdf](https://www.uni-oldenburg.de/fileadmin/user_upload/physik/ag/ehf/enmet/publications/solar/conference/2004/eurosun/short_term_forecasting_of_solar_radiation_based_on_satellite_data.pdf) (accessed  
774 March 4, 2015).
- 775 [16] G. Reikard, Predicting solar radiation at high resolutions: A comparison of time series forecasts,  
776 *Sol. Energy*. 83 (2009) 342–349. doi:10.1016/j.solener.2008.08.007.
- 777 [17] R. Perez, M. Beauharnois, K. Hemker, S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, G. Van  
778 Knowe, Evaluation of numerical weather prediction solar irradiance forecasts in the US, in:  
779 *Prep. Be Present. 2011 ASES Annu. Conf.*, 2011.  
780 <http://www.asrc.albany.edu/people/faculty/perez/2011/forc.pdf> (accessed March 4, 2015).
- 781 [18] P. Lauret, C. Voyant, T. Soubdhan, M. David, P. Poggi, A benchmarking of machine learning  
782 techniques for solar radiation forecasting in an insular context, *Sol. Energy*. 112 (2015) 446–  
783 457. doi:10.1016/j.solener.2014.12.014.
- 784 [19] R. Perez, K. Moore, S. Wilcox, D. Renné, A. Zelenka, Forecasting solar radiation—Preliminary  
785 evaluation of an approach based upon the national forecast database, *Sol. Energy*. 81 (2007)  
786 809–812.
- 787 [20] G. Mihalakakou, M. Santamouris, D.N. Asimakopoulos, The total solar radiation time series  
788 simulation in Athens, using neural networks, *Theor. Appl. Climatol.* 66 (2000) 185–197.
- 789 [21] J. Remund, R. Perez, E. Lorenz, Comparison of solar radiation forecasts for the USA, in: *Proc*  
790 *23rd Eur. PV Conf.*, 2008: pp. 1–9. [http://www.w.iew-](http://www.w.iew-shc.org/data/sites/1/publications/Comparison_of_USA_radiation_forecasts.pdf)  
791 [shc.org/data/sites/1/publications/Comparison\\_of\\_USA\\_radiation\\_forecasts.pdf](http://www.w.iew-shc.org/data/sites/1/publications/Comparison_of_USA_radiation_forecasts.pdf) (accessed  
792 March 4, 2015).
- 793 [22] COST | About COST, (n.d.). [http://www.cost.eu/about\\_cost](http://www.cost.eu/about_cost) (accessed July 31, 2016).
- 794 [23] R.H. Inman, H.T.C. Pedro, C.F.M. Coimbra, Solar forecasting methods for renewable energy  
795 integration, *Prog. Energy Combust. Sci.* 39 (2013) 535–576. doi:10.1016/j.pecs.2013.06.002.
- 796 [24] S.K. Aggarwal, L.M. Saini, Solar energy prediction using linear and non-linear regularization  
797 models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction  
798 Contest, *Energy*. 78 (2014) 247–256. doi:10.1016/j.energy.2014.10.012.
- 799 [25] Z. Şen, Ş.M. Cebeci, Solar irradiation estimation by monthly principal component analysis,  
800 *Energy Convers. Manag.* 49 (2008) 3129–3134. doi:10.1016/j.enconman.2008.06.006.
- 801 [26] M. Zarzo, P. Martí, Modeling the variability of solar radiation data among weather stations by  
802 means of principal components analysis, *Appl. Energy*. 88 (2011) 2775–2784.  
803 doi:10.1016/j.apenergy.2011.01.070.
- 804 [27] C. Paoli, C. Voyant, M. Muselli, M.-L. Nivet, Forecasting of preprocessed daily solar radiation  
805 time series using neural networks, *Sol. Energy*. 84 (2010) 2146–2160.  
806 doi:10.1016/j.solener.2010.08.011.
- 807 [28] D.J. Hand, Data Mining: Statistics and More?, *Am. Stat.* 52 (1998) 112–118.  
808 doi:10.1080/00031305.1998.10480549.
- 809 [29] G. Saporta, *Probabilités, analyse des données et statistique*, Editions TECHNIP, 2011.
- 810 [30] P.J. Brockwell, R.A. Davis, *Time series: theory and methods*, 2nd ed., Springer-Verlag, New York,  
811 1991.

- 812 [31] H.O.A. Wold, On Prediction in Stationary Time Series, *Ann. Math. Stat.* 19 (1948) 558–567.  
813 doi:10.1214/aoms/1177730151.
- 814 [32] J.G. De Gooijer, R.J. Hyndman, 25 years of time series forecasting, *Int. J. Forecast.* 22 (2006)  
815 443–473. doi:10.1016/j.ijforecast.2006.01.001.
- 816 [33] H.A.N. Hejase, A.H. Assi, Time-Series Regression Model for Prediction of Mean Daily Global  
817 Solar Radiation in Al-Ain, UAE, *Int. Sch. Res. Not.* 2012 (2012) e412471.  
818 doi:10.5402/2012/412471.
- 819 [34] H.-Y. Cheng, Hybrid solar irradiance now-casting by fusing Kalman filter and regressor, *Renew.*  
820 *Energy.* 91 (2016) 434–441. doi:10.1016/j.renene.2016.01.077.
- 821 [35] S. Kalogirou, Artificial neural networks in renewable energy systems applications: a review,  
822 *Renew. Sustain. Energy Rev.* 5 (2001) 373–401. doi:10.1016/S1364-0321(01)00006-5.
- 823 [36] C. Voyant, D. Kahina, G. Notton, C. Paoli, M.L. Nivet, M. Muselli, P. Haurant, The global  
824 radiation forecasting based on NWP or stochastic modeling: an objective criterion of choice, in:  
825 2013. <https://hal.archives-ouvertes.fr/hal-00848844/document> (accessed July 3, 2016).
- 826 [37] C. Voyant, C. Paoli, M. Muselli, M.-L. Nivet, Multi-horizon solar radiation forecasting for  
827 Mediterranean locations using time series models, *Renew. Sustain. Energy Rev.* 28 (2013) 44–  
828 52. doi:10.1016/j.rser.2013.07.058.
- 829 [38] M. Chaabene, M. Ben Ammar, Neuro-fuzzy dynamic model with Kalman filter to forecast  
830 irradiance and temperature for solar energy systems, *Renew. Energy.* 33 (2008) 1435–1443.  
831 doi:10.1016/j.renene.2007.10.004.
- 832 [39] *The Nature of Statistical Learning Theory | Vladimir Vapnik | Springer, (n.d.).*  
833 <http://www.springer.com/kr/book/9780387987804> (accessed July 19, 2016).
- 834 [40] H. Mori, A. Takahashi, A data mining method for selecting input variables for forecasting model  
835 of global solar radiation, in: *Transm. Distrib. Conf. Expo. T 2012 IEEE PES, 2012: pp. 1–6.*  
836 doi:10.1109/TDC.2012.6281569.
- 837 [41] N.K. H. Mori, Optimal regression tree based rule discovery for short-term load forecasting,  
838 (2001) 421–426 vol.2. doi:10.1109/PESW.2001.916878.
- 839 [42] A. Troncoso, S. Salcedo-Sanz, C. Casanova-Mateo, J.C. Riquelme, L. Prieto, Local models-based  
840 regression trees for very short-term wind speed prediction, *Renew. Energy.* 81 (2015) 589–598.  
841 doi:10.1016/j.renene.2015.03.071.
- 842 [43] H.T.C. Pedro, C.F.M. Coimbra, Nearest-neighbor methodology for prediction of intra-hour  
843 global horizontal and direct normal irradiances, *Renew. Energy.* 80 (2015) 770–782.  
844 doi:10.1016/j.renene.2015.02.061.
- 845 [44] M.C. Torre, P. Poggi, A. Louche, Markovian model for studying wind speed time series in  
846 corsica, *Int. J. Renew. Energy Eng.* 3 (2001) 311–319.
- 847 [45] V. Badescu, *Modeling Solar Radiation at the Earth’s Surface: Recent Advances*, Springer Science  
848 & Business Media, 2008.
- 849 [46] G. Zhang, Time series forecasting using a hybrid ARIMA and neural network model,  
850 *Neurocomputing.* 50 (2003) 159–175. doi:10.1016/S0925-2312(01)00702-0.
- 851 [47] C. Bouveyron, C. Brunet-Saumard, Model-based clustering of high-dimensional data: A review,  
852 *Comput. Stat. Data Anal.* 71 (2014) 52–78.
- 853 [48] F. Murtagh, P. Contreras, *Methods of Hierarchical Clustering*, ArXiv11050121 Cs Math Stat.  
854 (2011). <http://arxiv.org/abs/1105.0121> (accessed July 31, 2016).
- 855 [49] C.E. Rasmussen, *Gaussian processes for machine learning*, in: MIT Press, 2006.
- 856 [50] L. Gaillard, G. Ruedin, S. Giroux-Julien, M. Plantevit, M. Kaytoue, S. Saadon, C. Ménézo, J.-F.  
857 Boulicaut, Data-driven Performance Evaluation of Ventilated Photovoltaic Double-skin Facades  
858 in the Built Environment, *Energy Procedia.* 78 (2015) 447–452.  
859 doi:10.1016/j.egypro.2015.11.694.
- 860 [51] Y. Gala, Á. Fernández, J. Díaz, J.R. Dorronsoro, Hybrid machine learning forecasting of solar  
861 radiation values, *Neurocomputing.* 176 (2016) 48–59. doi:10.1016/j.neucom.2015.02.078.
- 862 [52] T.G. Dietterich, *Ensemble Methods in Machine Learning*, in: *Mult. Classif. Syst.*, Springer Berlin  
863 Heidelberg, 2000: pp. 1–15. doi:10.1007/3-540-45014-9\_1.

- 864 [53] R. Aler, R. Martín, J.M. Valls, I.M. Galván, A Study of Machine Learning Techniques for Daily  
865 Solar Energy Forecasting Using Numerical Weather Models, in: D. Camacho, L. Braubach, S.  
866 Venticinque, C. Badica (Eds.), *Intell. Distrib. Comput. VIII*, Springer International Publishing,  
867 2015: pp. 269–278. [http://link.springer.com.gate6.inist.fr/chapter/10.1007/978-3-319-10422-](http://link.springer.com.gate6.inist.fr/chapter/10.1007/978-3-319-10422-5_29)  
868 [5\\_29](http://link.springer.com.gate6.inist.fr/chapter/10.1007/978-3-319-10422-5_29) (accessed March 4, 2015).
- 869 [54] M. Azhar, Y. Waheeb, A. Zeyar, Probabilistic Solar Power Forecasting Derived from Ensemble of  
870 Point Forecast Regression Models, *Energies*. (in press).
- 871 [55] L. Breiman, Bagging Predictors, *Mach. Learn.* 24 (1996) 123–140.  
872 doi:10.1023/A:1018054314350.
- 873 [56] M.P. Almeida, O. Perpiñán, L. Narvarte, PV power forecast using a nonparametric PV model,  
874 *Sol. Energy.* 115 (2015) 354–368. doi:10.1016/j.solener.2015.03.006.
- 875 [57] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001) 5–32. doi:10.1023/A:1010933404324.
- 876 [58] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we Need Hundreds of Classifiers  
877 to Solve Real World Classification Problems?, *J. Mach. Learn. Res.* 15 (2014) 3133–3181.
- 878 [59] J. Huang, A. Troccoli, P. Coppin, An analytical comparison of four approaches to modelling the  
879 daily variability of solar irradiance using meteorological records, *Renew. Energy.* 72 (2014) 195–  
880 202. doi:10.1016/j.renene.2014.07.015.
- 881 [60] A. Liaw, M. Wiener, Classification and regression by randomForest, *R News.* 2 (2002) 18–22.
- 882 [61] M.H. Alobaidi, P.R. Marpu, T.B.M.J. Ouarda, H. Ghedira, Mapping of the Solar Irradiance in the  
883 UAE Using Advanced Artificial Neural Network Ensemble, *IEEE J. Sel. Top. Appl. Earth Obs.*  
884 *Remote Sens.* 7 (2014) 3668–3680. doi:10.1109/JSTARS.2014.2331255.
- 885 [62] S. Sperati, S. Alessandrini, P. Pinson, G. Kariniotakis, The “Weather Intelligence for Renewable  
886 Energies” Benchmarking Exercise on Short-Term Forecasting of Wind and Solar Power  
887 Generation, *Energies.* 8 (2015) 9594–9619. doi:10.3390/en8099594.
- 888 [63] S. Pelland, G. Galanis, G. Kallos, Solar and photovoltaic forecasting through post-processing of  
889 the Global Environmental Multiscale numerical weather prediction model, *Prog. Photovolt. Res.*  
890 *Appl.* 21 (2013) 284–296. doi:10.1002/pip.1180.
- 891 [64] J.R. Trapero, N. Kourentzes, A. Martin, Short-term solar irradiation forecasting based on  
892 Dynamic Harmonic Regression, *Energy.* 84 (2015) 289–295. doi:10.1016/j.energy.2015.02.100.
- 893 [65] P. Bacher, H. Madsen, H.A. Nielsen, Online short-term solar power forecasting, *Sol. Energy.* 83  
894 (2009) 1772–1783. doi:10.1016/j.solener.2009.05.016.
- 895 [66] M. David, F. Ramahatana, P.J. Trombe, P. Lauret, Probabilistic forecasting of the solar irradiance  
896 with recursive ARMA and GARCH models, *Sol. Energy.* 133 (2016) 55–72.  
897 doi:10.1016/j.solener.2016.03.064.
- 898 [67] C. Join, M. Fliess, C. Voyant, F. Chaxel, Solar energy production: Short-term forecasting and risk  
899 management, *ArXiv160206295 Cs Q-Fin.* (2016). <http://arxiv.org/abs/1602.06295> (accessed  
900 October 4, 2016).
- 901 [68] A. Mellit, S.A. Kalogirou, L. Hontoria, S. Shaari, Artificial intelligence techniques for sizing  
902 photovoltaic systems: A review, *Renew. Sustain. Energy Rev.* 13 (2009) 406–419.  
903 doi:10.1016/j.rser.2008.01.006.
- 904 [69] Y. KEMMOKU, S. ORITA, S. NAKAGAWA, T. SAKAKIBARA, DAILY INSOLATION FORECASTING  
905 USING A MULTI-STAGE NEURAL NETWORK, *Sol. Energy.* 66 (1999) 193–199. doi:10.1016/S0038-  
906 092X(99)00017-1.
- 907 [70] C. Voyant, C. Paoli, M. Muselli, M.-L. Nivet, Multi-horizon solar radiation forecasting for  
908 Mediterranean locations using time series models, *Renew. Sustain. Energy Rev.* 28 (2013) 44–  
909 52. doi:10.1016/j.rser.2013.07.058.
- 910 [71] W.R. Burrows, CART Regression Models for Predicting UV Radiation at the Ground in the  
911 Presence of Cloud and Other Environmental Factors, *J. Appl. Meteorol.* 36 (1997) 531–544.  
912 doi:10.1175/1520-0450(1997)036<0531:CRMFPU>2.0.CO;2.
- 913 [72] G.P. Podestá, L. Núñez, C.A. Villanueva, M.A. Skansi, Estimating daily solar radiation in the  
914 Argentine Pampas, *Agric. For. Meteorol.* 123 (2004) 41–53.  
915 doi:10.1016/j.agrformet.2003.11.002.

- 916 [73] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: A comparison of regression  
 917 analysis, decision tree and neural networks, *Energy*. 32 (2007) 1761–1768.  
 918 doi:10.1016/j.energy.2006.11.010.
- 919 [74] R. Marquez, C.F.M. Coimbra, Forecasting of global and direct solar irradiance using stochastic  
 920 learning methods, ground experiments and the NWS database, *Sol. Energy*. 85 (2011) 746–756.  
 921 doi:10.1016/j.solener.2011.01.007.
- 922 [75] A. Moreno, M.A. Gilabert, B. Martínez, Mapping daily global solar irradiation over Spain: A  
 923 comparative study of selected approaches, *Sol. Energy*. 85 (2011) 2072–2084.  
 924 doi:10.1016/j.solener.2011.05.017.
- 925 [76] S. Ben Taieb, G. Bontempi, A.F. Atiya, A. Sorjamaa, A review and comparison of strategies for  
 926 multi-step ahead time series forecasting based on the NN5 forecasting competition, *Expert*  
 927 *Syst. Appl.* 39 (2012) 7067–7083. doi:10.1016/j.eswa.2012.01.039.
- 928 [77] M. Demirtas, M. Yesilbudak, S. Sagiroglu, I. Colak, Prediction of solar radiation using  
 929 meteorological data, in: 2012 Int. Conf. Renew. Energy Res. Appl. ICRERA, 2012: pp. 1–4.  
 930 doi:10.1109/ICRERA.2012.6477329.
- 931 [78] S. Ferrari, M. Lazzaroni, V. Piuri, A. Salman, L. Cristaldi, M. Rossi, T. Poli, Illuminance prediction  
 932 through Extreme Learning Machines, in: 2012 IEEE Workshop Environ. Energy Struct. Monit.  
 933 *Syst. EESMS*, 2012: pp. 97–103. doi:10.1109/EESMS.2012.6348407.
- 934 [79] H. Mori, A. Takahashi, A data mining method for selecting input variables for forecasting model  
 935 of global solar radiation, in: *Transm. Distrib. Conf. Expo. T 2012 IEEE PES*, 2012: pp. 1–6.  
 936 doi:10.1109/TDC.2012.6281569.
- 937 [80] F. Olaiya, A.B. Adeyemo, Application of Data Mining Techniques in Weather Prediction and  
 938 Climate Change Studies, *Int. J. Inf. Eng. Electron. Business IJIEEB*. 4 (2012) 51.
- 939 [81] Á. Fernández, Y. Gala, J.R. Dorronsoro, Machine Learning Prediction of Large Area Photovoltaic  
 940 Energy Production, in: W.L. Woon, Z. Aung, S. Madnick (Eds.), *Data Anal. Renew. Energy Integr.*,  
 941 Springer International Publishing, Cham, 2014: pp. 38–53.  
 942 [http://link.springer.com/10.1007/978-3-319-13290-7\\_3](http://link.springer.com/10.1007/978-3-319-13290-7_3) (accessed March 3, 2015).
- 943 [82] P. Krömer, P. Musílek, E. Pelikán, P. Krč, P. Juruš, K. Eben, Support Vector Regression of multiple  
 944 predictive models of downward short-wave radiation, in: 2014 Int. Jt. Conf. Neural Netw.  
 945 *IJCNN*, 2014: pp. 651–657. doi:10.1109/IJCNN.2014.6889812.
- 946 [83] H. Long, Z. Zhang, Y. Su, Analysis of daily solar power prediction with data-driven approaches,  
 947 *Appl. Energy*. 126 (2014) 29–37. doi:10.1016/j.apenergy.2014.03.084.
- 948 [84] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sánchez, M. Sánchez-Girón, Daily global solar  
 949 radiation prediction based on a hybrid Coral Reefs Optimization – Extreme Learning Machine  
 950 approach, *Sol. Energy*. 105 (2014) 91–98. doi:10.1016/j.solener.2014.04.009.
- 951 [85] M. Zamo, O. Mestre, P. Arbogast, O. Pannekoucke, A benchmark of statistical regression  
 952 methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic  
 953 forecast of hourly production, *Sol. Energy*. 105 (2014) 792–803.  
 954 doi:10.1016/j.solener.2013.12.006.
- 955 [86] Björn Wolff, Elke Lorenz, Oliver Kramer, Statistical Learning for Short-Term Photovoltaic Power  
 956 Predictions (chapter), (2015) in print.
- 957 [87] M. Lazzaroni, S. Ferrari, V. Piuri, A. Salman, L. Cristaldi, M. Faifer, Models for solar radiation  
 958 prediction based on different measurement sites, *Measurement*. 63 (2015) 346–363.  
 959 doi:10.1016/j.measurement.2014.11.037.
- 960 [88] A. McGovern, D.J. Gagne II, L. Eustaquio, G. Titericz, B. Lazorthes, O. Zhang, G. Louppe, P.  
 961 Prettenhofer, J. Basara, T.M. Hamill, D. Margolin, Solar Energy Prediction: An International  
 962 Contest to Initiate Interdisciplinary Research on Compelling Meteorological Problems, *Bull. Am.*  
 963 *Meteorol. Soc. BAMS*. (2015). <http://orbi.ulg.ac.be/handle/2268/177115> (accessed March 8,  
 964 2015).
- 965 [89] J. Cao, X. Lin, Study of hourly and daily solar irradiation forecast using diagonal recurrent  
 966 wavelet neural networks, *Energy Convers. Manag.* 49 (2008) 1396–1406.  
 967 doi:10.1016/j.enconman.2007.12.030.



- 968 [90] A. Chaouachi, R.M. Kamel, K. Nagasaka, Neural Network Ensemble-Based Solar Power  
969 Generation Short-Term Forecasting., *JACIII*. 14 (2010) 69–75.
- 970 [91] M. Gastón, Í. Pagola, C.M. Fernández-Peruchena, L. Ramírez, F. Mallor, A new Adaptive  
971 methodology of Global-to-Direct irradiance based on clustering and kernel machines  
972 techniques, in: 15th SolarPACES Conf., Berlin, Germany, 2010: p. 11693. [https://hal.archives-](https://hal.archives-ouvertes.fr/hal-00919064)  
973 [ouvertes.fr/hal-00919064](https://hal.archives-ouvertes.fr/hal-00919064) (accessed March 15, 2015).
- 974 [92] P. Chakraborty, M. Marwah, M.F. Arlitt, N. Ramakrishnan, Fine-Grained Photovoltaic Output  
975 Prediction Using a Bayesian Ensemble., in: *AAAI*, 2012.  
976 <http://people.cs.vt.edu/naren/papers/chakraborty-aaai12.pdf> (accessed March 8, 2015).
- 977 [93] M.R. Hossain, A.M.T. Oo, A.B.M.S. Ali, Hybrid prediction method of solar power using different  
978 computational intelligence algorithms, in: *Univ. Power Eng. Conf. AUPEC 2012 22nd Australas.*,  
979 2012: pp. 1–6.
- 980 [94] M. Bouzardoum, A. Mellit, A. Massi Pavan, A hybrid model (SARIMA–SVM) for short-term  
981 power forecasting of a small-scale grid-connected photovoltaic plant, *Sol. Energy*. 98, Part C  
982 (2013) 226–235. doi:10.1016/j.solener.2013.10.002.
- 983 [95] Y. Chu, H.T.C. Pedro, C.F.M. Coimbra, Hybrid intra-hour DNI forecasts with sky image processing  
984 enhanced by stochastic learning, *Sol. Energy*. 98, Part C (2013) 592–603.  
985 doi:10.1016/j.solener.2013.10.020.
- 986 [96] L. Prokop, S. Misak, V. Snasel, J. Platos, P. Kroemer, Supervised learning of photovoltaic power  
987 plant output prediction models, *Neural Netw. World*. 23 (2013) 321–338.
- 988 [97] S.K. Aggarwal, L.M. Saini, Solar energy prediction using linear and non-linear regularization  
989 models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction  
990 Contest, *Energy*. 78 (2014) 247–256. doi:10.1016/j.energy.2014.10.012.
- 991 [98] I. Bilonis, E.M. Constantinescu, M. Anitescu, Data-driven model for solar irradiation based on  
992 satellite observations, *Sol. Energy*. 110 (2014) 22–38.
- 993 [99] J. Wu, C.K. Chan, Y. Zhang, B.Y. Xiong, Q.H. Zhang, Prediction of solar radiation with genetic  
994 approach combing multi-model framework, *Renew. Energy*. 66 (2014) 132–139.  
995 doi:10.1016/j.renene.2013.11.064.
- 996 [100] Y.-K. Wu, C.-R. Chen, H. Abdul Rahman, A Novel Hybrid Model for Short-Term Forecasting in PV  
997 Power Generation, *Int. J. Photoenergy*. 2014 (2014) e569249. doi:10.1155/2014/569249.
- 998 [101] H.-T. Yang, C.-M. Huang, Y.-C. Huang, Y.-S. Pai, A Weather-Based Hybrid Method for 1-Day  
999 Ahead Hourly Forecasting of PV Power Output, *IEEE Trans. Sustain. Energy*. 5 (2014) 917–926.  
1000 doi:10.1109/TSTE.2014.2313600.
- 1001 [102] Y. Chu, H.T.C. Pedro, M. Li, C.F.M. Coimbra, Real-time forecasting of solar irradiance ramps with  
1002 smart image processing, *Sol. Energy*. 114 (2015) 91–104. doi:10.1016/j.solener.2015.01.024.
- 1003 [103] M. De Felice, M. Petitta, P.M. Ruti, Short-term predictability of photovoltaic production over  
1004 Italy, *Renew. Energy*. 80 (2015) 197–204. doi:10.1016/j.renene.2015.02.010.
- 1005 [104] Z. Dong, D. Yang, T. Reindl, W.M. Walsh, A novel hybrid approach based on self-organizing  
1006 maps, support vector regression and particle swarm optimization to forecast solar irradiance,  
1007 *Energy*. (n.d.). doi:10.1016/j.energy.2015.01.066.
- 1008 [105] M. Samanta, B.K. Srikanth, J.B. Yerrapragada, Short-Term Power Forecasting of Solar PV  
1009 Systems Using Machine Learning Techniques, (n.d.).  
1010 <http://cs229.stanford.edu/proj2014/Mayukh%20Samanta,Bharath%20Srikanth,Jayesh%20Yerrapragada,Short%20Term%20Power%20Forecasting%20Of%20Solar%20PV%20Systems%20Using%20Machine%20Learning%20Techniques.pdf> (accessed March 15, 2015).
- 1011  
1012  
1013

1014	
1015	<b>List of Tables</b>
1016	Table 1. The three categories of storage and their technical specification
1017	Table 2. Characteristics of electricity production plants [9]
1018	Table 3. List of representative papers related to the global solar radiation forecasting using single
1019	machine learning methods
1020	Table 4. list of representative papers related to the global radiation forecasting combining machine
1021	learning methods
1022	
1023	<b>List of figures</b>
1024	Figure 1. Prediction scale for energy management in an electrical network [2]
1025	Figure 2 a) Forecasting error versus forecasting models (left) [6,14]. b) Relation between forecasting
1026	horizons, forecasting models and the related activities (right) [6, 14]
1027	Figure 3: Bias variance tradeoff
1028	Figure 4:a) Time series of predicted and measured global irradiance for 2008 in Ajaccio (France); b)
1029	Scatter plot of predicted vs. measured global irradiance in Ajaccio (France); c) Example of ROC curve
1030	(an ideal ROC curve is near the upper left corner).
1031	Figure 5: Relative RMSE of forecasts(persistence, auto regression, and scaled persistence) and of
1032	reference models depending on the forecast horizon [18]
1033	Figure 6: Number of time the ANN, machine learning and SVM terms have been used in the original
1034	articles
1035	
1036	
1037	