Predicting the Power Output of Distributed Renewable Energy Resources within a Broad Geographical Region

Athanasios Aris Panagopoulos[1] **Georgios Chalkiadakis**[1] **Eftichios Koutroulis**[1]

Abstract. In recent years, estimating the power output of inherently intermittent and potentially distributed renewable energy sources has become a major scientific and societal concern. In this paper, we provide an algorithmic framework, along with an interactive web-based tool, to enable short-to-middle term forecasts of photovoltaic (PV) systems and wind generators output. Importantly, we propose a generic PV output estimation method, the backbone of which is a solar irradiance approximation model that incorporates free-to-use, readily available meteorological data coming from online weather stations. The model utilizes non-linear approximation components for turning cloud-coverage into radiation forecasts, such as an MLP neural network with one hidden layer. We present a thorough evaluation of the proposed techniques, and show that they can be successfully employed within a broad geographical region (the Mediterranean belt) and come with specific performance guarantees. Crucially, our methods do not rely on complex and expensive weather models and data, and our web-based tool can be of immediate use to the community as a simulation data acquisition platform.

1 Introduction

In recent decades, with fossil fuel resources running out and environmental concerns regarding their use growing, the generation of power from renewable energy sources has been hailed as the way forward to an energy-sufficient future. Renewable energy sources, however, are inherently intermittent, as their power output depends on a variety of factors. This fact has lead research in engineering to develop numerous methods for estimating the power output of renewable energy generators. More recently, AI and multiagent systems research has been increasingly preoccupying itself with building intelligent systems for the Smart Grid [23]—and the efficient incorporation of renewable energy generators into the Smart Grid has emerged as a major challenge [16, 18]. The creation of Virtual Power Plants (VPPs), in particular, bringing together large numbers of heterogeneous Distributed Energy Resources (DERs) to create the impression of a single "conventional" power plant, has been suggested as a way to overcome the unpredictability of renewable energy generation [3, 17].

Now, forecasting PV systems output can, in many cases, be linked to the task of forecasting *solar irradiance* (or *radiation*) estimates. Though several such approximation methods have been proposed in the literature, they typically suffer from several drawbacks: *(a)* they rely on expensive meteorological forecasts; *(b)* they usually do not come with strict approximation performance guarantees; this is because *(c)* they are made up of components that have been evaluated only in isolation; or *(d)* their performance has been evaluated only in a narrow geographic region. Moreover, many such methods produce *clear sky* prediction models only. However, *the evaluation of prediction methods in a wide region* is important for the day-to-day operation of VPPs with regionally-distributed DER members, as they need to make decisions as to which members to employ for their daily production needs; in addition, it can be of value to VPPs or enterprises that need to plan where to recruit members from, or where to build renewable energy facilities; and, last but not least, to national or regional Grid operators, who need forecasts of solar and wind power to properly predict and balance supply with demand.

Against this background, in this paper we provide algorithmic tools to produce power output estimates coming from potentially distributed renewable energy resources (such as solar and wind generators). In a nutshell, we propose a generic method to come up with PV output estimates, the backbone of which is a solar irradiance approximation model that takes cloud coverage into account, makes use of free-to-use and readily available meteorological data, and comes with specific performance guarantees for a wide region of interest. Our solar irradiance model is built with components chosen after being carefully evaluated against each other in a broad geographic region—the Mediterranean belt (Med-Belt for short). The components in question are non-linear approximation methods for turning cloud-coverage into radiation forecasts, such as an MLP neural network with one hidden layer. Importantly, our tools use online data that can be downloaded for free from weather forecasting websites, and do not rely on complex and expensive weather models and data. By so doing, this paper is the first to present a *generic* but *low-cost* power output estimation method which is applicable within a wide geographical region. Our work also demonstrates how standard machine learning methods, like least-squares fitting and neural networks, can be effectively applied to predict the power output of solar plants in a wide region. Note that it is the use of "intermediate steps", such as using a solar irradiance model, that allows our method to be applicable outside narrow regions—as would be the case if we just trained a neural network over specific plants' production output data.

In more detail, our main contributions are as follows. *(a)* We propose novel non-linear approximation methods to estimate solar radiation falling on a surface given cloud coverage information, and evaluate them based on real data coming from across the whole Mediterranean belt. Moreover, we test the performance of those methods at specific locations within and outside that region. Our results suggest that one such method, an MLP neural network, significantly outperforms all others. *(b)* Our methods only require weather data that are readily available to all for free via weather websites. *(c)* We combine our solar irradiance model with existing models calculating various PV systems losses, and come up with a generic PV power output estimation model. *(d)* We estimate, via an error propagation procedure, the total error of our method for the Med-Belt. *(e)* By so doing, this paper is the first to provide *low-cost* power prediction estimates via

[1] Electronic and Computer Engineering, Technical University of Crete, Greece; emails: {apanagopoulos, gchalkiadakis, ekoutroulis}@isc.tuc.gr

a method applied to a wide region, via incorporating solar irradiance forecasts in the process. *(f)* We implemented a web-based, interactive DER power output estimation tool, *RENES*, that incorporates our PV power output estimation method, and also wind turbine output estimates, for any location in Europe. Our tool enables the user to enter equipment specifications, and derive power output estimates based on weather forecasts for the days of interest. *(g)* Our method and tool can be extended to incorporate any other "intermediate-step" techniques deemed appropriate for particular sub-regions (e.g., techniques that prove to perform better within a sub-region of interest). *(h)* Finally, our work provides the scientific community with a convenient user-interactive tool for simulations and experiments; this tool could also be of use, in the long term, to the operation of VPPs competing in the power market.

The rest of the paper is structured as follows: We begin in Sec. 2 with a brief review of related work. Then, we present our PV output estimation procedure—including an *all sky* solar irradiance estimation model incorporating cloud coverage information—in Section 3. In Section 4 we evaluate our methods for turning cloud-coverage-to-radiation, and select two of them for incorporation into our generic method and web-based tool. There, we also present the overall error of the complete PV output estimation procedure. Section 5 briefly outlines our work on providing wind turbine power output estimates, and the rest of the RENES functionality. Finally, Section 6 concludes.

2 Related Work

Here we provide a brief review of the work most relevant to ours. To begin, *neural networks* and *time-series models* have been extensively used to provide PV systems output forecasts *without* taking the intermediate step of estimating solar radiation (e.g., [13, 22]). However, such methods are restricted to providing predictions for a specific PV system, or systems within a small region (as they have to be trained on data related to the particular system in question). Moreover, time-series models require access to online statistical performance data.

On the other hand, several *cloud-cover radiation (CRM)* models relating solar radiation with degrees of cloud coverage and clear sky radiation estimation methods have appeared in the literature over time (e.g., [7, 15]). These models are quite generic, but have not been thoroughly evaluated against each other, for the most part. Nevertheless, they can incorporate simple cloud coverage data as the ones provided by free weather websites, and therefore can potentially be utilized for the acquisition of short-to-medium term (24 to 48 hours) forecasting in a wide region. We thus incorporate such models in our method. By contrast, very short term (up to 6 hours) forecasting methods, or global numerical weather prediction (NWP) models, which are based on analyzing hard to obtain satellite images or complex raw meteorological data are inappropriate for our work here.

As stated, this paper is the first to provide a regionally-applied, *low-cost* power prediction estimation method, incorporating solar irradiance forecasts in the process. The only other work we are aware of that uses irradiance forecasts to produce *regional* renewable energy output estimates, is that of [2, 10], which is nevertheless based on detailed forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), that are in general provided to member state organizations only, or under a fee.

Finally, we note that web-tools for PV power output estimates have begun to appear in commercial websites[2]. However, they do not come with an appropriate documentation of the forecasting method used.

3 A PV Output Estimation Model

The method for predicting the energy output of PV systems presented in this paper consists of a series of relatively independent estimation steps that include: *(a)* developing a solar irradiance model to predict *the incident radiation*, G_T, on the PV module; *(b)* estimating *the amount of incident radiation actually absorbed* by the PV module, G_{eff}; *(c)* predicting the module's *operating temperature*, T_c; *(d)* calculating the PV module's *maximum power output*, P_m; and, finally, *(e)* predicting the PV system's *actual power output*, P_{eff}. We now describe the aforementioned steps in detail, in a "bottom-up" order.

3.1 A solar irradiance prediction model

There is a variety of *clear sky* models that have been developed for the calculation of solar radiation in optimum weather conditions (see, e.g., [8, 11]). Based on these, numerous models have been developed for the calculation of solar radiation under *cloudy conditions* as well (e.g., [7, 15]). As mentioned above, however, in general such models are evaluated in a specific region only, they use monthly-averaged rather than the more finely grained hour-by-hour data, and depend on hard to find meteorological information.

Our prediction model utilizes a number of formulas reported in the clear sky models literature, extending them to include two cloud transmittance coefficients, τ_{c_b} and τ_{c_d}, which need to be estimated in order to derive the solar radiation levels under different cloud coverage conditions. Intuitively, these coefficients describe the "quantity" of beam and diffuse radiation allowed through certain degrees of cloudiness. Our framework articulates a clear step-by-step methodology for estimating the relevant cloud transmittance coefficients.

An All-Sky Solar Radiation Model. The total incident radiation on an *arbitrarily oriented* (earth/terrestrial) surface, $G_T^{arb}(N)$, given a cloud coverage level N, is calculated with the following procedure:

In general, $G_T^{arb}(N)$ consists of the beam $G_B^{arb}(N)$, sky-diffuse $G_D^{arb}(N)$ and ground-reflected $G_R^{arb}(N)$ components [11]:

$$G_T^{arb}(N) = G_B^{arb}(N) + G_D^{arb}(N) + G_R^{arb}(N) \tag{1}$$

$G_B^{arb}(N)$ is calculated from equation 2.

$$G_B^{arb}(N) = G_{on}\tau_b\tau_{c_b}cos\theta_s \tag{2}$$

where θ_s is the angle between the normal to the surface and the direction to the sun; τ_{c_b} is the cloud transmittance coefficient for beam solar radiation; τ_b is the clear sky atmospheric transmittance coefficient for beam solar radiation; and G_{on}, or *extraterrestrial* radiation, is the incident radiation on a surface located immediately outside the earth atmosphere and oriented normally to the direction of the incoming solar radiation. G_{on} . With its intra-day variations being considered negligible, day-to-day G_{on} is given by:

$$G_{on} = G_{sc}\left(\frac{D_0}{D}\right)^2 \tag{3}$$

where G_{sc} is the average solar radiation at a distance of 1 AU[3] from the sun. This *solar constant* is valued at $1360.8 \pm 0.5W/m^2$ based on recent estimations. D_0 is the yearly mean Earth-Sun distance (1 AU), and D the Earth-Sun distance in a given day. Then, $G_D^{arb}(N)$ is given by Eq. 4, which assumes that every point of the celestial sphere emits light with equal radiance [11].

[2] See, for instance, *http://www.wunderground.com/calculators/solar.html*

[3] Astronomical Unit = 149,597,870.7 km (92,955,807.3 mi)

$$G_D^{arb}(N) = G_{on} cos\theta_z \tau_d \tau_{c_d} \frac{1 + cos\beta}{2} \quad (4)$$

where θ_z is the solar zenith angle, τ_{c_d} is the *cloud transmittance* coefficient for diffuse solar radiation, τ_d is the *clear sky* atmospheric transmittance coefficient for diffuse solar radiation, and β is the inclination angle of the surface.

The third component, $G_R^{arb}(N)$, is calculated by Eq. 5, which assumes that the ground is horizontal, of infinite extent, and reflects uniformly to all directions [11].

$$G_R^{arb}(N) = \rho G_T^{hor}(N) \frac{1 - cos\beta}{2} \quad (5)$$

where $G_T^{hor}(N)$ stands for the total incident radiation on a horizontal surface, and ρ is the average reflectance of the ground.

Now, note that, when considering $G_R^{arb}(N)$ on a horizontal surface, $\beta = 0$ and thus $G_R^{hor}(N) = 0$. As a consequence, the total incident radiation *on a horizontal surface*, $G_T^{hor}(N)$, is:

$$G_T^{hor}(N) = G_B^{hor}(N) + G_D^{hor}(N) \quad (6)$$

The clear sky atmospheric transmittance coefficient for *beam* solar radiation (τ_b) is estimated in accordance with standard procedures [6]. Subsequently, τ_d is approximated as $\tau_d = 0.271 - 0.294\tau_b$. The θ_z and θ_s angles are estimated through known methods [19].

Estimating the Cloud Transmittances. Given the model above, it is clear that, what is missing in order to calculate $G_T^{arb}(N)$, is estimating the values of the cloud transmittance coefficients τ_{c_b} and τ_{c_d}. These coefficients depend on the level of cloud coverage, but, intuitively, have a value of 1 under clear sky conditions (where all light is allowed to go through). Hence, one can easily see that Equations 2 and 4 can be expressed for a *horizontal* surface as

$$G_B^{hor}(N) = G_B^{hor}(0)\tau_{c_b} \quad (7)$$

$$G_D^{hor}(N) = G_D^{hor}(0)\tau_{c_d} \quad (8)$$

(since, for instance, $G_B^{hor}(0) = G_{on}\tau_b 1 cos\theta_s$).

Solving Eq. 7 and 8 for τ_{c_b} and τ_{c_d} would allow for the calculation of the beam and diffuse cloud transmittance coefficients for any level of cloud coverage, via Eq. 2 and 4. Now, $G_B^{hor}(0)$ and $G_D^{hor}(0)$ can be estimated via Eq. 2 and 4 by assuming a horizontal orientation instead of an arbitrary one, and replacing the cloud transmittance coefficients with the value of 1. Unfortunately, there is no direct way to calculate $G_B^{hor}(N)$ and $G_D^{hor}(N)$; and, moreover, measurements of those quantities are non-existent or very hard to obtain.

To overcome this difficulty, and since $G_T^{hor}(N)$ (i.e., horizontal-surface radiation under a given degree of cloud coverage) measurements are relatively commonplace, we *(i)* develop a *cloud-cover radiation (CRM) model* to predict estimates of the total $G_T^{hor}(N)$ irradiance on a horizontal surface, given relevant past measurements under cloud coverage degree N. Our CRM model can employ several approximation algorithms, such as using the least squares method to fit various non-linear models we introduce to approximate the $G_T^{hor}(N)/G_T^{hor}(0)$ ratio, or using an MLP neural network, as we detail below. Note that such regression and function approximation techniques have long been applied in the field of machine-learning and AI. Then, we *(ii)* decompose the estimated $G_T^{hor}(N)$ back to $G_B^{hor}(N)$ and $G_D^{hor}(N)$. For this step, we employ a readily available *diffuse ratio model* developed specially for our region of interest [4].

We now detail our approaches to completing step *(i)* above.

Non-linear equation models. Here we describe the non-linear models we test-evaluated, with the purpose of adopting one for our CRM model. These models attempt to approximate the $G_T^{hor}(N)/G_T^{hor}(0)$ ratio, which is known to be independent of the season and solar elevation [7]. (Note that $G_T^{hor}(0)$ quantities can be easily calculated by our all-sky radiation model, via Equation 6 and after estimating the $G_B^{hor}(0)$ and $G_D^{hor}(0)$ quantities.) We eventually derived the parameters of our models via the well-known least-squares fitting technique.

The first of our models, is based on a commonly used formula put forth by Kasten & Czeplak [7] witch was originally based on 10 years of measurements from Hamburg, Germany. To relate $G_T^{hor}(N)$ with $G_T^{hor}(0)$ and cloud coverage N, they propose a parameterized formula of the form: $G_T^{hor}(N)/G_T^{hor}(0) = 1 + B_{0,0}(N/8)^{B_{0,1}}$. The $1/8$ in the model comes from the fact that the "sky condition" qualitative attribute is reported by weather forecasting agencies as a simple cloud coverage estimate (usually considering five levels of cloud coverage), and then takes a quantitative expression in "eighths". Table 1 summarizes the various observable sky conditions along with their corresponding quantitative expression.

Table 1. Sky Conditions (table provided in [24])

Reported Sky Condition	Meaning	Summation Amount of Layer (X / 8)
SKC or CLR	Clear	0
FEW	Few	1/8 - 2/8
SCT	Scattered	3/8 - 4/8
BKN	Broken	5/8 - 7/8
OVC	Overcast	8/8

To better approximate the Med-Belt regional characteristics, our first model uses their proposal after equipping it with an additional regression (correction) coefficient:

$$G_T^{hor}(N)/G_T^{hor}(0) = 1 + B_{0,0}(N/8)^{B_{0,1}} + B_{0,2} \quad (9)$$

We then use least-squares fitting to estimate the B parameters. Note that, though well-known, this model is evaluated in the Mediterranean region for the first time in our work here.

We also developed three additional non-linear models. The first of them is a fourth-degree polynomial, described in Equation 10 below; intuitively, a polynomial of degree 4 is expected to best-fit data with 5 levels of cloud coverage, which is the number of cloud coverage levels normally found in the online data provided by weather websites (see Table 1). The second method proposed is a third-degree polynomial, described in Equation 11; we chose to evaluate this method in order to test the hypothesis that a polynomial of degree 3 would be able to fit data with 5 levels of cloud coverage quite well, while being better at avoiding potential "overfitting" effects. Furthermore, after observing that our data-points approximately take a sigmoid shape, we decided to also attempt to fit it with a regular sigmoid (logistic) curve, described in Equation 12. These models are shown in the following equations, where $G_T^{hor}(N)/G_T^{hor}(0)$ is the dependent variable, N is the independent one (corresponding to levels of cloud coverage). We estimated the actual values of the various $B_{i,j}$ coefficients by employing least-squares fitting on accumulated irradiance measurements, as we detail in the next section.

$$\begin{aligned} G_T^{hor}(N)/G_T^{hor}(0) &= B_{1,0}(N/8)^4 + B_{1,1}(N/8)^3 \\ &+ B_{1,2}(N/8)^2 + B_{1,3}(N/8) \\ &+ B_{1,4} \quad (10) \end{aligned}$$

$$G_T^{hor}(N)/G_T^{hor}(0) = B_{2,0}(N/8)^3 + B_{2,1}(N/8)^2 + B_{2,2}(N/8) + B_{2,3} \quad (11)$$

$$G_T^{hor}(N)/G_T^{hor}(0) = \frac{1}{1+e^{-B_{3,0}(N/8+B_{3,1})}} \quad (12)$$

Development of an MLP network. In addition to evaluating the predictive performance non-linear equations above, we also trained a multilayer perceptron (MLP) neural network with one hidden layer [5]. The network computes the $G_T^{hor}(N)$ quantity given the level of cloud coverage, N; the estimated $G_T^{hor}(0)$ quantity; the environmental temperature T_a; and the relative humidity, RH.The use of the T_a and RH parameters for network training was inspired by [20], which suggests that temperature and relative humidity data can be utilized to replace missing irradiance measurements in a dataset.

3.2 Estimating the final output of the PV system

The procedures presented in the previous section enable us to estimate the PV module's (total) incident solar radiation $G_T^{arb}(N)$. However, not all of this radiation is absorbed by the module.

First of all, absorption depends on the *angle of incidence* of solar radiation, as the reflectance and transmittance of optical materials changes along with it. As such, the optical input of a PV panel depends on its orientation to the sun. Another factor affecting radiation absorption concerns sediments of soil and dirt that are deposited on a functioning PV on a daily basis.

These factors are considered in the estimation of the panel's *effective incident radiation*, G_{eff}. To estimate G_{eff}, we follow the procedures detailed in [11]. Also, given these factors' relatively small variations across different modules, our implementation considers them w.r.t. a typical monocrystalline silicon module. However, (corrective) values for other module types can be incorporated into our model in a straightforward manner.

A further factor to consider when estimating PV power output, is the PV module's operating temperature, as lower operating temperatures improve its ability to convert solar radiation into electricity. The solar cell operating temperature T_c of a PV module depends on the ambient temperature, as well as on the heat produced by the module, and the heat lost to the environment. The heat exchange between the module and its environment, in turn, depends on various factors, such as module-specific attributes, and on the prevailing heat transfer mechanisms (i.e conduction, convection and radiation).

A variety of conceptual and empirical estimation models have been developed for the calculation of the PV module's operating temperature.For the needs of our work, we utilize the model of [21], which ties T_c to the panel's effective incident radiation, G_{eff}, the prevalling wind speed, V, and the ambient temperature T_a.

Taking such characteristics into account, a number of conceptual and empirical estimation models have been developed for the calculation of a PV module's maximum power output, P_m. Here, based on a comparison of Predictive Models for Photovoltaic Module Performance performed by the National Renewable Energy Laboratory (NREL) [12], we adopt the *PvForm* model [14], which can account for reductions in the PV module's efficiency due to low irradiance levels. However, in recent years manufacturers have begun to provide measurements of such performance reductions. When such measurements are available, our web-based tool automatically utilizes the *Improved PV* model [12], which successfully incorporates them.

The module's *maximum power output*, P_m, corresponds to the final PV system's power output, assuming the utilization of an optimally regulated maximum power point tracker (MPPT),[4] and negligible wiring, inverter, or other losses. In order to account for such losses, an empirical "efficiency" factor, k has been used so the effective power output, P_{eff} is computed as $P_{eff} = kP_m$.[5]

4 Evaluation and Performance Guarantees

In this section, we first describe the process we used to build a Mediterranean belt-specific dataset of weather observations for training and evaluating our models. Then, we describe how we used this dataset to determine the coefficients of our proposed non-linear approximation equations for our area of interest, and train our neural network. Following that, we evaluate all our five irradiance under cloud coverage estimation models; and derive and report the final power output prediction performance of our approach.

Building the observations dataset For the purpose of our research, archival meteorological data was drawn from the *Weather Underground* database for 9 regions in the Med-Belt[6], and 1 region in Northern Europe. Specifically, we drew data for *sky condition* (qualitative observations), *solar radiation* (i.e., $G_T^{hor}(N)$ in W/m^2), *ambient temperature* (°C), and *relative humidity* (%). At least one year worth of observation data during 2009-2012 was collected in each city. The locations (and corresponding datasets) are seen in Table 2.

To build our final dataset, observations with solar radiation out of bounds $[0, 1.2G_{on}cos\theta_z]$ [15] were excluded. Furthermore, observations with unusually high or low temperature readings (given the regional historical extremes); unusually high nightly radiation readings; as well as unusually low (∼zero) midday radiation readings were also excluded (as possible anomalies or "maintenance" incidents). To derive homogeneous and equivalent datasets for the Med-Belt regions, we reduced the larger datasets by progressively retaining every second observation. Then, all Med-Belt sets were *collated* and the resulting "global" observations dataset was divided in two sets: a training and a testing set. These subsets were derived from the global one through an iterative process of distributing its data-points to each subset in an alternating fashion. The whole process ensured there were no regularities present in the datasets. The training set was used to estimate the $B_{i,j}$ coefficient parameters above, as well as to train the MLP network. The testing set was used to evaluate the respective goodness-of-fit of all five approaches (in the MLP's case, where *early stopping* [5] is applied, half of the testing set was used for validation and half for evaluation purposes, as we later explain).

Table 2. The final experimental dataset

Country	Location	Range[a]
Spain	Gava, Barcelona	14275
	Pantano de Cubillas, Albolote, Granada	15520
	Patraix, Valencia	17498
Greece	Chania, Crete	15252
	Kato Pylea, Thessaloniki	13836
France	Montauroux, Provence	17662
	Orange,Provence Alpes Cote d'Azur	17600
Italy	Mezzana Bigli, Lombardia	18642
Portugal	Lordelo do Ouro, Porto	18612
Denmark	Lake Arresoe, Ramloese, Helsinge	45087

[a] Number of valid observations after all quality control tests

[4] An MPPT is a high efficiency electronic controller that varies a PV module's electrical operating point in order to maximize power output.
[5] The value of k is user-provided, and should correspond to the inverter efficiency factor, if an inverter is used—adjusted to best fit the system.
[6] In the case of Chania, Greece the respective archival meteorological data have been provided by the National Observatory of Athens.

Least-squares fitting of the non-linear curves In order to fit our proposed non-linear curves to our dataset above, we used the following procedure. First, given that each qualitative sky condition value usually corresponds to more than one "eighths" (e.g. FEW corresponds to $1/8 - 2/8$, SCT to $3/8 - 4/8$, and so on), we derived a "midpoint" unique corresponding quantitative value to characterize each cloud coverage level. That is, we characterize $\{CLR, FEW, SCT, BKN, OVC\}$ by the following respective values for N: $\{0, 1.5, 3.5, 6, 8\}$. We then used our training set to compute the sample mean of the corresponding $G_T^{hor}(N)/G_T^{hor}(0)$ for each of those values of N. The resulting $\langle N, G_T^{hor}(N)/G_T^{hor}(0)\rangle$ pairs then define five points on the Cartesian plane which where used to estimate the vector of $B_{i,j}$ coefficients of our least square fitting models. The derived $B_{i,j}$ coefficients are the following. For Eq. 9, $B_{0,0} = -0.6287$, $B_{0,1} = 1.1653$ and $B_{0,2} = 0.034$; for Eq. 10, $B_{1,0} = 1.63$, $B_{1,1} = -3.047$, $B_{1,2} = 1.531$, $B_{1,3} = -0.7411$ and $B_{1,4} = 1.037$; for Eq. 11, $B_{2,0} = 0.198$, $B_{2,1} = -0.4371$, $B_{2,2} = -0.3865$ and $B_{2,3} = 1.033$; and for Eq. 12, $B_{3,0} = -3.6772$ and $B_{3,1} = -0.8665$.

Training the MLP network To train our neural network the testing set was divided into two equal parts, the validation set and a new testing set (by adding to each the data-points of the original testing set in an alternating fashion). The neural networks architecture comprises one hidden layer with five input and one output nodes. After five experimental iterations of training the network with 3,4,5,7,8,14, and 26 hidden layer neurons, the MLP comprising of 4 nodes in the hidden layer was found to present the best network architecture. Normalized values in the range of $[-1, 1]$ for the quantities $T_a, RH, G_B^{hor}(0), G_D^{hor}(0), N$ constituted the networks five input nodes. Sigmoid activation functions were used for the hidden layer neurons, while linear functions were used for the output node. The MLP training used the back propagation learning algorithm with the batch method and uniform learning. Overfitting is avoided via the *early stoping* neural network training technique [5].

Evaluating the CRM (cloud-cover radiation) models For the evaluation of our five CRM approaches, we calculated their *Mean Absolute Percentage Error*: $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{F_i - A_i}{A_i}\right| 100$; *Mean Absolute Error*: $MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - A_i|$; and *relative Mean Absolute Error*: $rMAE = \frac{MAE}{1/n\sum_{i=1}^{n}A_i}100$. Here, A_i represents a data-point coming from the actual (historical data) $G_T^{hor}(N) \neq 0$ quantities, and F_i represents the corresponding forecasted (estimated) one, with i ranging from 1 to n within the dataset. Note that, for near-zero A_i values, the corresponding *absolute percentage error (APE)* will approach infinity, even if the error is small. For this reason, we excluded all the value-pairs of (A_i , F_i) with $A_i/max\{A_i\}_{i=1}^{n} < 0.1$ from the MAPE calculation, as is standard practice [22]. All CRM methods were evaluated on the appropriate testing sets described earlier, and on the dataset collected from Lake Arresoe in Denmark to test their behaviour outside the region of interest. The evaluation results for the four least square-fitted curves are reported in Table 3.

We ran a standard one-way ANOVA test on these methods, which showed that their APE errors are different in a statistical significant manner. However, follow-up paired T-tests showed there is no statistical significance (with 95% confidence) among the 4th & 3rd degree polynomials and the Kasten & Czeplak's Med-Belt formulation methods, while there is statistical significance between the error of each one of those methods and the error of the sigmoid function (i.e., the sigmoid is significantly *worse* than the others—*cf.* Table 3).

Table 3. Evaluation of the fitted non-linear curves.

	Mediterranean			Denmark		
Equation	MAPE	rMAE	MAE	MAPE	rMAE	MAE
K&C-Med	23.727	21.441	75.904	34.538	37.051	98.938
4th-degr.Pol.	23.825	21.585	76.414	34.611	37.109	99.091
3rd-degr.Pol.	23.692	21.396	75.744	34.554	37.059	98.958
Sigmoid	25.0	22.688	80.319	35.882	38.238	102.108

"K&C-Med" is Eq. 9. MAPE & rMAE in %, MAE in W/M^2.

Our results show that the MLP network *is a clear winner* when compared with the four other CRM models. Specifically, its MAPE, rMAE and MAE were 22.946%, 19.456% and $68.69W/M^2$ respectively for the Med-Belt, and subsequent paired T-tests confirmed its error is indeed lower in a statistical significant manner. Moreover, we trained it and tested it *separately* on datasets for all our 9 specific locations, and observed that its performance was significantly enhanced; for all of the cities, MAPE, rMAE, and MAE dropped to the levels of (approximately) 16%, 15%, and $45W/M^2$, respectively.

However, the MLP network's performance deteriorates considerably outside the Med-Belt, as it is trained on Med-Belt data; specifically, for Arresoe, MAPE=46.171%, rMAE=39.762% and MAE=$106.149W/M^2$. Thus, RENES incorporates the MLP network as its CRM model inside the Med-Belt, but uses the 3rd degree polynomial outside the Med-Belt (due to its slightly better performance there when compared to the other methods).

Final power output prediction performance guarantees For the evaluation of our tool, we employ an error propagation methodology [1], in order to accumulate each individual sub-model's error and calculate the total error. The MAE and rMAE were calculated for PV modules of two different manufacturing technologies (i.e. multi-crystal and single-crystal Si) and four mounting configurations (i.e. stand-alone, flat roof, sloped roof and building-integrated). The PV modules were considered to be installed with either a $0°$ or a $45°$ tilt angle (in the latter case, south-facing). For each combination of PV module manufacturing technology, mounting type and tilt angle configurations, the error at the output of each sub-model was calculated. Then, that error was propagated through the "chain" of sub-models, being recursively added to the subsequent model's error, to estimate the overall error for each data point contained in our dataset. Finally, the MAE and rMAE of the method were estimated.

The derived overall method's power output prediction errors for *horizontal* orientation appear in Table 4.

Table 4. Overall Output Prediction Error on Horizontal orientation

	Multi-crystal Si Nominal $P_m : 35.16W$		Single-crystal Si Nominal $P_m : 74.34W$	
Mounting Type	MAE (W)	rMAE (%)	MAE (W)	rMAE (%)
Stand-Alone	2.527	22.494	5.451	21.891
Flat Roof	2.504	22.603	5.404	21.989
Sloped Roof	2.445	22.967	5.269	22.319
Building-Integrated	2.391	23.397	5.143	22.724

Due to a lack of required data with respect to irradiance measurements at non-zero slope angles within the Med-Belt, we were only able to estimate *a worst-case* approximate bound for the *inclined* orientation above (i.e., a typical south-facing, $45°$ slope angle), of around 40% relative mean absolute error (for all mounting types above). We defer the details to an extended version of this paper.

In terms of comparing our method's performance with related work, we note that most existing power output prediction work (e.g., using trained neural networks) refers to specific narrow geographical areas, as explained earlier. To the best of our knowledge, the only generic prediction methodology that has been applied in a wide area is that of [2, 10]—but their PV output prediction performance results are incomparable to ours, since they lie outside the Med-Belt. However, their method's error relies heavily on irradiance forecasting (which is also the main factor affecting our method's performance). This enables us to compare our irradiance forecasting error to theirs, as found in a paper reporting an application of their method in Southern Spain [9]: their results for that region have a relative MAE of approximately 12.5%. This is better than our MLP's rMAE of 19.456% (over the whole Med-Belt); however, as noted earlier, their methodology relies on global numerical weather predictions (NWP) provided by meteorological organizations, while ours is an inexpensive methodology based on free-for-all online weather data.[7]

5 A Web-Based DER Output Estimation Tool

We incorporated our PV power output estimation model in a web-based, graphical, user-interactive, *renewable energy estimation* tool, *RENES* witch can be found at *http://www.intelligence.tuc.gr/renes*. The tool currently provides accurate estimates (within the aforementioned error guarantees for PV output estimates) for the Med-Belt. Its operation is based on weather predictions from online weather websites (such as *Weather Underground*), and specifications for renewable generators for any location on a user-clickable map of Europe. Most essential parameters, such as longitude/latitude, or typical PV systems parameters, are automatically populated with values, but can also be filled in by the user. We note that RENES allows for the easy incorporation and extension of all the models discussed above, and different ones. It also provides a web-based application program interface (API), enabling the service of direct http request messages. Finally, part of the tool's functionality is predicting the power output of wind turbines at specified locations. Wind-based generation prediction employs a standard method, estimating production based on the so-called *power curve* of each turbine, which determines its output based on forecasted wind speeds (see, e.g., [3] for more details).

6 Conclusions and Future Work

In this paper, we presented a generic, low cost PV output estimation method, based on weather readings from online websites, and evaluated it with real data over the Mediterranean region. We incorporated this method in a web-based tool that enables the user of predicting the output of distributed energy renewable (solar and wind) energy generators. Our tool, RENES, can be of use to the research community for experiments and simulations (as it can be a convenient platform for "scrapping" online weather data). Moreover, it can be potentially of value to VPPs and the energy industry, or the wider public. To this end, impending work includes user-evaluating RENES and enhancing it with more capabilities.

Regarding future work, we plan to evaluate alternative algorithms for inclusion in our generic prediction method. Further, we aim to utilize our tool to gather data for Smart Grid and energy-related research, such as designing economic mechanisms related to VPP operation, or using machine learning techniques for optimal sun-tracking.

Acknowledgements We thank the National Observatory of Athens (NOA), and primarily Kostas Lagouvardos (Institute of Environmental Research, NOA), for the provision of archival weather data. We are also grateful to Emmanouil Alvizos (Warwick University), and Alex Rogers, Luke Teacy (University of Southampton) for comments and fruitful discussions.

[7] We note that, interestingly, our solar irradiance forecasting MLP approach has a performance similar to that of most other such (solar irradiance prediction, but global NWP-based) methods reported in [9].

REFERENCES

[1] P.R. Bevington, *Data reduction and error analysis for the physical sciences*, McGraw-Hill, 1969.

[2] S. Bofinger and G. Heilscher, 'Solar electricity forecast - approaches and first results', in *Proc. of the 21st European Photovoltaic Solar Energy Conference*, (2006).

[3] G. Chalkiadakis, V. Robu, R. Kota, A. Rogers, and N. R. Jennings, 'Cooperatives of distributed energy resources for efficient virtual power plants', in *Proc. of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, (2011).

[4] A. de Miguel, J. Bilbao, R. Aguiar, H. Kambezidis, and E. Negro, 'Diffuse solar irradiation model evaluation in the north mediterranean belt area', *Solar Energy*, **70**(2), 143 – 153, (2001).

[5] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 1998.

[6] Hoyt C. Hottel, 'A simple model for estimating the transmittance of direct solar radiation through clear atmospheres', *Solar Energy*, **18**(2), 129 – 134, (1976).

[7] F. Kasten and G. Czeplak, 'Solar and terrestrial radiation dependent on the amount and type of cloud', *Solar Energy*, **24**, 177–189, (1979).

[8] R. King and R.O. Buckius, 'Direct solar transmittance for a clear sky', *Solar Energy*, **22**(3), 297 – 301, (1979).

[9] E. Lorenz, J. Remund, S.C. Mller, W. Traunmller, Steinmaurer, D. G., J.A. Ruiz-Arias, V.L. Fanego, L. Ramirez, M.G. Romeo, C. Kurz, L.M. Pomares, and C.G. Guerrero, 'Benchmarking of different approaches to forecast solar irradiance', in *Proc. of the 24th European Photovoltaic Solar Energy Conference*, (2009).

[10] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, 'Regional pv power prediction for improved grid integration', *Progress in Photovoltaics: Research and Applications*, **19**(7), 757–771, (2011).

[11] A. Luque and S. Hegedus, *Handbook of photovoltaic science and engineering*, Wiley, 2003.

[12] B. Marion, *Comparison of predictive models for photovoltaic module performance*, 1–6, number 1, 2008.

[13] A. Mellit and A. Massi Pavan, 'Performance prediction of $20kw_p$ grid-connected photovoltaic plant at trieste (italy) using artificial neural network', *Energy Conversion and Management*, **51**(12), (2010).

[14] D.F. Menicucci, J.P. Fernandez, and Sandia National Laboratories, *User's manual for PVFORM: a photovoltaic system simulation program for stand-alone and grid-interactive applications*, Accents Publication Service, 1988.

[15] T. Muneer and F. Fairooz, 'Quality control of solar radiation and sunshine measurements lessons learnt from processing worldwide databases', *Building Services Engineering Research And Technology*, **23**(3), 151–166, (2002).

[16] C. W. Potter, A. Archambault, and K. Westrick, 'Building a smarter smart grid through better renewable energy information', in *Power Systems Conference and Exposition*, pp. 1–5, (March 2009).

[17] D. Pudjianto, C. Ramsay, and G. Strbac, 'Virtual power plant and system integration of distributed energy resources', *IET Renewable Power Generation*, **1**(1), 10–16, (2007).

[18] S. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings, 'Putting the "smarts" into the smart grid: A grand challenge for artificial intelligence', *Communications of the ACM*, (2012).

[19] I. Reda and A. Andreas, 'Solar position algorithm for solar radiation applications', *Solar Energy*, **76**(5), 577 – 589, (2004).

[20] D. Firmanda Al Riza, S. Ihtsham ul Haq Gilani, and M. Shiraz Aris, 'Hourly solar radiation estimation using ambient temperature and relative humidity data', *International Journal of Environmental Science and Development*, **2**(3), 188–193, (2011).

[21] E. Skoplaki, A.G. Boudouvis, and J.A. Palyvos, 'A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting', *Solar Energy Materials and Solar Cells*, **92**(11), (2008).

[22] Cai Tao, Duan Shanxu, and Chen Changsong, 'Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement', in *Proc. of 2nd IEEE Intern. Symposium on Power Electronics for Distributed Generation Systems*, (June 2010).

[23] U.S. Department of Energy. Grid 2030: A national vision for electricity's second 100 years, 2003.

[24] U.S. Office of the Federal Coordinator for Meteorological Services and Supporting Research, *Surface weather observations and reports*, 1998.