A Novel Method for Predicting the Power Output of Distributed Renewable Energy Resources

by

© *Aris-Athanasios Panagopoulos*

Committee:

Assistant Professor Georgios Chalkiadakis (Supervisor)

Associate Professor Michail G. Lagoudakis

Assistant Professor Eftichios Koutroulis

A thesis submitted to the Department of Electronic and Computer Engineering

in partial fulfilment of the requirements for the degree of Diploma in Engineering

Department of *Electronic and Computer Engineering*

Technical University of Crete

June 2013

A. A. Panagopoulos Chania

Abstract

In recent years, estimating the power output of inherently intermittent and potentially distributed weather-driven and non-scheduled renewable energy resources has become a major scientific and societal concern. In this thesis, we provide an algorithmic framework, along with an interactive web-based tool, to enable short-to-middle term forecasts of Photovoltaic Systems (PVS) and Wind Turbine Generators (WTG) power output. Importantly, we propose a generic PVS power output estimation method, the backbone of which is a non-linear solar irradiance approximation model that incorporates free-to-use, readily available weather forecasting reports coming from online providers. We present a thorough evaluation of different approaches of the respective model, as well as a a thorough evaluation of our PVS power output estimation method. We show that our PVS power output estimation method 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 scientific community and society at large.

Περίληψη (Abstract in Greek)

Τα τελευταία χρονια, η εκτίμηση-πρόβλεψη της ενεργειακής απόδοσης των πιθανά διάσπαρτων και εξαρτώμενων από τις καιρικές συνθήκες Ανανεώσιμων Πηγών Ενέργειας (ΑΠΕ) έχει προκαλέσει έντονο κοινωνικό και επιστημονικό ενδιαφέρον. Στην παρούσα διπλωματική εργασία (α) αναπτύσουμε και προτείνουμε ένα καινοτόμο αλγοριθμικό πλαίσιο, και (β) αναπτύσουμε και παρέχουμε ένα σχετικό διαδραστικό διαδικτυακό εργαλείο, για την παραγωγή (βραχυπρόθεσμων και μεσοπρόθεσμων) προβλέψεων για την ενεργειακή απόδοση Φωτοβολταϊκών (Φ/Β) Συστημάτων και Ανεμογεννητριών. Κεντρική συνεισφορά της εργασίας αποτελεί η πρόταση μιας γενικής και οικονομικής μεθόδου για την πρόβλεψη της απόδοσης των διάσπαρτων Φ/Β Συστημάτων. Ο πυρήνας της μεθόδου αυτής είναι ένα μη-γραμμικό μοντέλο εκτίμησης της προσπίπτουσας ηλιακής ακτινοβολίας, βάσει μετεωρολογικών προβλέψεων που παρέχονται από ιστότοπους χωρίς χρηματικό αντίτιμο. Παρουσιάζουμε μια εκτενή αξιολόγηση διαφορετικών προσεγγίσεων του σχετικού μοντέλου, και μια εκτενή αξιολόγηση της συνολικής μεθόδου. Δείχνουμε ότι η μέθοδος αυτή μπορεί να χρησιμοποιηθεί σε έναν ευρύ γεωγραφικό χώρο (περιοχή της Μεσογείου), με συγκεκριμένες εγγυήσεις απόδοσης. Σε αντίθεση με άλλες προσεγγίσεις, οι μέθοδοί μας δεν απαιτούν τη χρήση σύνθετων ή/και ακριβών μετεωρολογικών μοντέλων ή δεδομένων, και το διαδικτυακό μας εργαλείο μπορεί να χρησιμοποιηθεί άμεσα από την επιστημονική κοινότητα - αλλά και από την ευρύτερη κοινωνία.

"It is better to read the weather forecast before we pray for rain"

Mark Twain

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my thesis supervisor Dr. Georgios Chalkiadakis for his enlightening guidance and inspiring instruction as well as for supporting and trusting me throughout this study.

Furthermore, I would like to thank Dr. Eftichios Koutroulis for his valuable and constructive suggestions during the planning and development of this work. I am also grateful to Dr. Emmanouil Alvizos (University of Warwick), and Dr. Alex Rogers, Dr. Luke Teacy (University of Southampton) for comments and/or fruitful discussions.

This work had quite high requirements in specific data and resources. I am grateful to all individuals, laboratories and service providers for their help in meeting these requirements. In particular, I would like to thank Intelligent Systems Laboratory and primarily Spyros Argyropoulos for offering me the necessary utility resources and facilities for this study. Spyros willingness to give his time so generously has been very much appreciated. Moreover, I would like to thank the National Observatory of Athens (NOA), and primarily Kostas Lagouvardos (Institute of Environmental Research, NOA), for the provision of archival weather data. Lastly, special thanks

goes to Weather Underground[1] for being a valuable online source of free archival weather data and weather forecasting reports.

At this point, I would like to express my very great appreciation to Dr. Dimitrios S. Patelis and to all the members of our philosophy group. I would like to thank them for inspiring me, supporting me, and delving me deeper into philosophy and scientific methodology. Also, I would like to thank them for providing me additional facilities for this work. Furthermore, I would like to thank the public educational system of Greece and the Greek taxpayers for granting me the opportunity to study at Technical University of Crete (TUC), Greece. Last but not least, I would like to express my love and gratitude to my family: Penelope Koulouri, Lysandros Panagopoulos, Orestis Panagopoulos and Euridiki Koulouri, for supporting me unconditionally.

I would like to dedicate this work to all people around the globe with no access to proper education.

[1] *http://www.wunderground.com*

Contents

List of Tables

List of Figures

List of Main Abbreviations & Symbols

A	Altitude
AI	Artificial intelligence
AM	Air mass coefficient
APE	Absolute percentage error
AU	Astronomical unit
Avg	Average
$B_{i,j}$	Regression coefficients array
$C_{i,j}$	Regression coefficients array
CRM	Cloud cover radiation model
D	Earth-Sun distance
D_0	Annually mean Earth-Sun distance
DER	Distributed energy resource
F_d	Diffuse fraction of horizontal radiation
$G_B^{arb}(N)$	Beam solar radiation incident on an arbitrarily oriented surface
$G_B^{hor}(N)$	Beam solar radiation incident on a horizontal surface
$G_D^{arb}(N)$	Sky-diffuse solar radiation incident on an arbitrarily oriented surface

$G_D^{hor}(N)$	Sky-diffuse solar radiation incident on a horizontal surface
G_{eff}	Amount of incident radiation actually absorbed by the PV module
G_L	Effective low irradiance
G_{on}	Extraterrestrial radiation
$G_R^{arb}(N)$	Ground-reflected solar radiation incident on an arbitrarily oriented surface
$G_R^{hor}(N)$	Ground-reflected solar radiation incident on a horizontal surface
G_{sc}	Solar constant
G_T	Incident radiation on the PV module
$G_T^{arb}(N)$	Total incident radiation on an arbitrarily oriented surface
$G_T^{hor}(N)$	Total incident radiation on a horizontal surface
K_d	Clearness index
MAE	Mean absolute error
$MAPE$	Mean absolute percentage error
MAS	Multiagent systems
MBE	Mean bias error
MLP	Multilayer perceptron

N	Cloud coverage level
$NomCapacity$	Nominal capacity
NWP	Numerical weather prediction
P_{eff}	PV system actual power output
P_m	PV module maximum power output
PV	Photovoltaic
PVS	Photovoltaic system
$RENES$	Renewable energy estimator
RH	Relative humidity
$RMSE$	Root mean square error
$rMAE$	Relative mean absolute error
SRC	Standard Reporting Conditions
T_a	Ambient temperature
T_c	PV module operating temperature
T_d	Dew point temperature
T_{d-a}	Dew point temperature minus ambient temperature
TSI	Total solar irradiance
V	Prevailing wind speed
VPP	Virtual power plant

$WDERs$	Weather-driven and non-scheduled energy resources
WTG	Wind turbine generators
β	Inclination angle of surface
γ	Maximum power correction factor for temperature
θ_s	Angle between the normal to the surface and the direction to the sun
θ_z	Solar zenith angle
ρ	Average reflectance of the ground
τ_b	Clear sky atmospheric transmittance coefficient for beam solar radiation
τ_{c_b}	Cloud transmittance coefficient for beam solar radiation
τ_{c_d}	Cloud transmittance coefficient for sky-diffuse solar radiation
τ_d	Clear sky atmospheric transmittance coefficient for sky-diffuse solar radiation
ω	Mounting coefficient

Chapter 1

Introduction

Up to now, solar radiation and wind remain, to a great extent, non-commoditized, public goods. This means that solar radiation and wind are freely accessible (although not in the same quality and/or quantity) to all people around the globe. The social use-value of these public goods is enormous and includes among other things their ability to be used as energy sources. In that particular ability, solar radiation and wind are known as renewable energy sources, since it seems impossible to exploit them faster than they are naturally replenished.[1] The electromagnetic energy of solar radiation and the kinetic energy of wind are commonly known as *solar*, and *wind* energy respectively. As the major current technology of converting solar and wind energy into a more convenient and desired form (i.e. electrical energy) is known to have no carbon emissions involved (at least during the very same process of conversion) [62], the energy produced is collectively well known as 'clean energy'.

Being *free*, *renewable*, *generating zero carbon emissions* (over their exploitation)

[1]The Sun is estimated to become a red giant in more than five billion years [17] hence solar radiation and wind can be expected to be naturally replenished for quite a long time.

and *relatively abundant*, solar radiation and wind share an advantage over other traditionally used sources of energy such as fossil fuels. Moreover, with the great advancements into the solar and wind energy exploitation technology and the great need for new sectors of investment, solar and wind energy represent a small but constantly increasing fraction of the global electrical power generation capacity.[2]

However, despite these advantages, solar and wind energy do posses certain disadvantages. Chief among them is their inherent unreliability; the power produced by *Photovoltaic (PV) Systems (PVSs)* and *Wind Turbine Generators (WTGs)* often suffer from fluctuations and intermittent operation due to their dependency on incident solar radiation and prevailing wind respectively. In the absence of dependable predictive methods, the power output of PVSs and WTGs can at best be considered as nondeterministic stochastic process. In this sense, PVSs and WTGs are at first glance seemingly uncontrollable and non-dispatchable power generators.

In general terms, the objectives of modern science are to accurately describe, explain and predict the movement of the world that surrounds us. The ability to predict allows us to plan our actions accordingly; allows us to act with prior conception of reality [51, 72]. In the case of the power output of *Weather-Driven and non-scheduled Energy Resources (WDERs for short)* such as PVSs and WTGs, prediction technology could make WDER reliability a reality.

[2]Despite the fact that the sources of information provided in *The REN21 Renewables 2011 Global Status Report* cannot always be held liable regarding their accuracy, the reader could get a taste on global electrical power generation capacity and its fractions on sources. The magazine is published online; *www.ren21.net*

1.1 Towards a Smarter Grid

The existing electricity Grid,[3] in most regions, is not a product of scientific planning but more of a product of an evolutionary process strictly connected to societal, economical, political, geographical, technological and other factors. Even if it had, originally, been a product of scientific planning, the unexpected changes in the production, transmission, conservation and/or consumption technologies would call for a paradigm shift. The evolution to a Smarter Grid is nowadays imminent, and concerns one of the greatest engineering challenges of our day.

In general terms, the move to a Smarter Grid is considered as the move from a centralized, producer-controlled network to a less centralized and more consumer-interactive one [26]. The main reasons that call for a radical reengineering of the Grid infrastructure and functionality include the growing demand caused by the electrification of transport and heating and the growing penetration of inherently intermittent and potentially distributed WDERs into the Grid.

In the context of power system operation, one of the greatest challenges is running a reliable supply-on-demand system. Historically this challenge led to an electricity Grid based on highly controllable supply in order to match a largely uncontrolled demand [55]. That said, the growing penetration of WDERs into the Grid will impact the system's reliability [46, 75, 25]. Therefore, one of the major challenges of the Smart Grid vision is the reliable integration of WDERs into the Grid while meeting the constantly growing demand [46].

[3]The electricity Grid is a vast, interconnected network for delivering electricity from suppliers to consumers; the term is also used broadly to refer to the network along with its operators and operating policies.

1.2 Grid Reliability and Power Output Prediction of Weather Driven Energy Resources

Combined with other technologies, such as energy storage,[4] WDER power output predicting deals with the issue stated in the previous section (1.1). Predictive technology could reinforce the reliable integration of WDERs into the electricity Grid as "forecasts of future requirements are essential to be able to prepare the controllable and flexible systems, such as those based on fossil fuels, to behave in the appropriate manner" [55, 61, 75].

More recently, *Artificial Intelligence (AI)* and *Multiagent Systems (MAS)* research has been increasingly preoccupying itself with building intelligent systems for the Smart Grid [71]. In the process, the efficient incorporation of WDERs into the Smart Grid has emerged as a major challenge [55, 57]. The term *Virtual Power Plants (VPPs)* corresponds to the notion of a large number of heterogeneous *Distributed Energy Resources (DERs)*, usually WDERs, joining forces and offering electricity to the Grid - while providing the guarantees of a single "conventional" power plant.[5] VPPs create the necessary synergies among DERs, so that the effective and efficient delivery of energy is assured, while still being able to utilize (the inherently intermittent and thus untrustworthy) WDERs [10, 56, 71]. However, several challenges arise in the formation and management of VPPs. In particular, the individual VPP members (many of which rely on WDERs), need to be able to come to an agreement

[4] "Energy storage is essential in ensuring the reliability of the Grid. Storage system can store excess energy and provide power when energy shortage. The existing energy storage technologies include battery, flywheels, super-capacitors and superconducting magnetic energy storage (SMES)" [31]

[5] Note that the term VPPs may also refer to "coalitions of consumers" offering energy consumption reduction services to the Grid [2]. We will not deal with this problem in this thesis.

in technical and financial terms [71]. In order for the VPP to be able to act effectively, it is essential that reliable predictions can be made regarding the power output of its various WDERs members.

The above illustrate the necessity for dependable WDER power output predictive technology and operational tools. This technology should be able to provide accurate WDER power output predictions within a wide geographical region. Moreover, it could be of great value, if this technology was generic but also of low-cost making it widely available.

1.3 Predicting the Power Output of Photovoltaic Systems

Now, forecasting PVS power 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.

1.4 Predicting the Power Output of Wind Turbine Generators

In the case of WTGs, the power prediction is tightly linked to the well investigated task of wind forecasting, with dependable wind forecasting reports being freely available from a great number of online providers. Hence, the WTG power output predictive technology has already shown its strong economic impact through integrated web tools and applications, despite its relatively young history [12, 73]. Although further improvement of the reliability and accuracy of existing models is an open issue, we will not concern ourselves with this problem in this thesis and we will rely on already available low-cost and generic methods. In particular we will rely on a simple physical approach, which turns predictions of prevailing wind speed (coming from online providers) into WTG power output predictions, through the so-called WTG power curve [1]. The related theory is discussed in Section 3.2.

1.5 Thesis Contributions

In this work we provide algorithmic tools to produce power output estimates coming from potentially distributed *Weather-Driven and non-scheduled Energy Resources (WDERs for short)*, such as *Photovoltaic (PV) Systems (PVSs)*, and *Wind Turbine Generators (WTGs)*. In a nutshell, we propose a generic method to come up with PVS power 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 that were 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—some carefully chosen from the literature and some novel, such as a *Multilayer Perceptron (MLP)* 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 work is the first to present a *generic* but *low-cost* power output estimation method that is applicable within a wide geographical region. Notice 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 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 PVS losses, and come up with a generic PVS 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, we provide *low-cost* power prediction estimates via a method applied to a wide region, via incorporating solar irradiance forecasts in the process. *(f)* We implemented a web-based, interactive *Distributed Energy Resource (DER)* power output estimation tool, *Renewable Energy Estimator (RENES)*, that incorporates our PVS power output estimation method, and also WTG power 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 here, and our user-interactive tool in particular, can be of use, in the long term, to the operation of *Virtual Power Plants (VPPs)* competing in the power market; in the short term, it provides the scientific community with a convenient tool for simulations and experiments.

Finally we note that a research paper based on this work, entitled "Predicting the Power Output of Distributed Renewable Energy Resources within a Broad Geo-

graphical Region" and co-authored by Aris-Athanasios Panagopoulos, Dr. Georgios Chalkiadakis and Dr. Eftichios Koutroulis, was awarded the best student paper award in the *Prestigious Applications of Intelligent Systems (PAIS)* track of the *2012 European Conference on Artificial Intelligence (ECAI 2012)* [50].

Thesis Outline

The rest of the thesis is structured as follows. We begin in Chapter 2 with a review of related work. Then, in Chapter 3, we present our approach to PVS and WTG power output estimation. In particular, in Section 3.1 we present our PVS power output estimation procedure—including an *all sky* solar irradiance estimation model incorporating cloud coverage information. In Section 3.2 we present our WTG power output estimation procedure. Chapter 4 describes our implementation; in Section 4.1 we evaluate our methods for turning cloud coverage to radiation, and select one of them for incorporation into our generic method and web-based tool, and in Section 4.2 we present some additional, extra, results of our study. In Chapter 5 we estimate the overall error of our PVS power output prediction model. There we also describe the error propagation method used to determine this results. Chapter 6 briefly introduces our web-based DER power output estimation tool, RENES and outlines its functionality. Finally, Chapter 7 concludes.

Chapter 2

Related Work

Here we provide a review of the work most relevant to ours. To begin, *neural networks* and *time-series models* have been extensively used to provide power output forecasts of *Photovoltaic (PV) Systems (PVSs) without* taking the intermediate step of estimating solar radiation (e.g., [44, 68, 3]). 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.

For instance, the work of [44] proposes two *Multilayer Perceptron (MLP)* network structures to predict the power output of a *Grid Connected Photovoltaic (GCPV)* plant. The first MLP structure proposed, estimates the power produced by the GCPV plant, taking as input the air temperature and the incident solar irradiance. The second MLP structure is a univariate model which takes as input the solar irradiance only. Although the method is quite accurate the MLP networks need to be trained on very specific data coming from a very specific power plant. Moreover, this work

assumes that solar irradiance predictions are already available.

The work of [68] proposes a low-cost method based on a *nonlinear autoregressive network with exogenous inputs (NARX)*, in order to provide hourly power output predictions for a PVS. This method utilizes a *clear sky* solar radiation model and meteorological forecasting reports coming from online providers. The method has been evaluated for a PVS in Denmark with quite accurate estimates. However the NARX network needs to be trained with data coming from specific PVS.

On the other hand, several *Cloud cover Radiation Models (CRM)* relating solar radiation with degrees of cloud coverage and clear sky radiation estimation methods have appeared in the literature over time (e.g., [28, 52, 47]). 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 which are based on analyzing hard to obtain ground-based or satellite images (e.g. [11, 19]) or global *Numerical Weather Prediction (NWP)* models which are based on complex raw meteorological data, are inappropriate for our work here (e.g. [36]).

As stated, this thesis 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 [8] which is nevertheless based on detailed forecasts from the European Center for *Medium-Range Weather Forecasts*

(ECMWF) [36], 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[1]. However, they do not come with an appropriate documentation of the forecasting method used.

In the case of *Wind Turbine Generators (WTGs)*, several software systems produce dependable power output estimates by using existing NWP data and live (SCADA[2]) WTGs' data as input (e.g., [27, 12]). Some of the NWP models used for this purpose are the European *High Resolution Limited Area Model (HiRLAM)* [39] and the *Global Forecast System (GFS)* from the *National Oceanic and Atmospheric Administration (NOAA)*. As wind forecasting reports can be found free of charge, online from several providers, our work here produces WTG power output predictions based on such reports. By so doing we are able to produce *low-cost* WTG power output forecasting reports for a *wide* geographical region. This kind of reports could be of great value especially for small producers and/or low-budget *Virtual Power Plants (VPPs)*.

[1] See, for instance, *http://www.solarserver.com/service/solar-Photovoltaic-power-forecast-for-worldwide-locations.html.*; *http://www.wunderground.com/calculators/solar.html*

[2] Supervisory Control And Data Acquisition

Chapter 3

A Power Output Prediction Method for Weather-Driven and Non-Scheduled Energy Resources

In this chapter we first, explicitly present our approach to the power output estimation model for *Photovoltaic (PV) Systems (PVSs)* (Section 3.1); and then, we present our approach to the power output estimation model for *Wind Turbine Generators (WTGs)* (Section 3.2).

3.1 Power Output Estimation Model for Photovoltaic Systems

The method for predicting the power output of PV systems presented in this thesis 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 (Section 3.1.1), there we also present 9 different approaches for turning cloud coverage into radiation; *(b)* estimating *the amount of incident radiation actually absorbed* by the PV module, G_{eff} (Section 3.1.2); *(c)* predicting the module's *operating temperature*, T_c (Section 3.1.3); *(d)* calculating the PV module's *maximum power output*, P_m (Section 3.1.4); and, finally, *(e)* predicting the PV system's *actual power output*, P_{eff} (Section 3.1.4). Figure 3.1 illustrates a flow chart of our method divided into the respective submodels/steps. Note that the output of the submodel *considering the effects of angle of incidence, soil and dirt*, which is G_{eff} (described in Section 3.1.2), is an input to both *PV Module Thermal* submodel and *PV Module Maximum Power Output* submodel. We now describe the aforementioned steps in detail, in a "bottom-up" order.

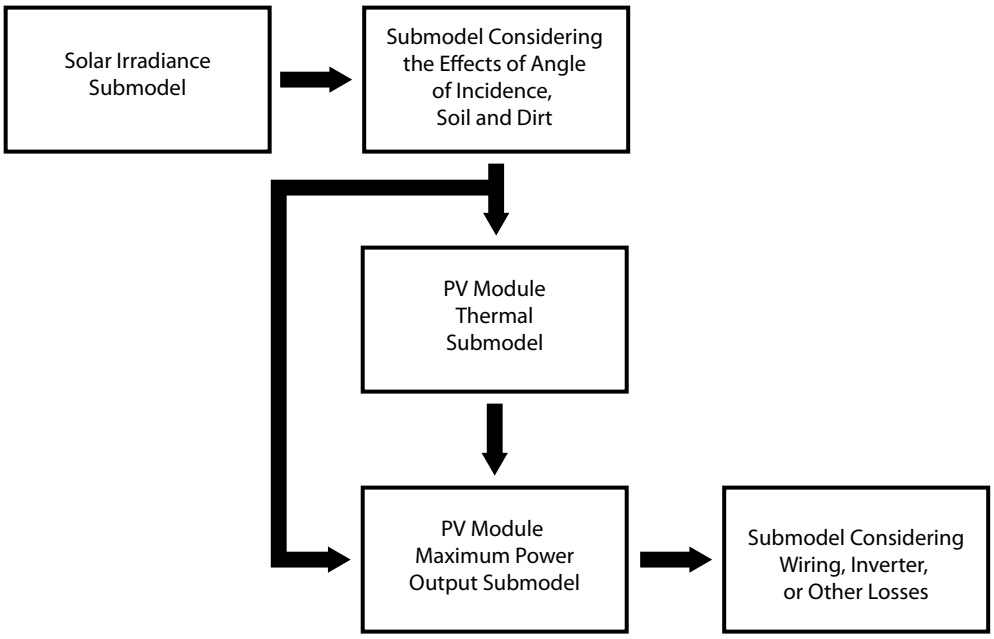

Figure 3.1: Flow chart describing the steps of our method

3.1.1 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., [29, 7, 24, 18]). Based on these, numerous models have been developed for the calculation of solar radiation under *cloudy conditions* as well (e.g., [23, 4, 54, 47]). 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:

- The beam cloud transmittance coefficient, τ_{c_b}
- The diffuse cloud transmittance coefficient, τ_{c_d}

Intuitively, these coefficients describe, respectively, the *"quantity" of beam and diffuse radiation allowed through certain degrees of cloudiness*. These coefficients need to be estimated in order to derive the solar radiation levels under different cloud coverage conditions. Our framework articulates a clear step-by-step methodology for estimating the respective cloud transmittance coefficients.

3.1.1.1 An All-Sky Solar Radiation Model

In order to start building our solar radiation model, the extraterrestrial solar radiation needs to be estimated. The *extraterrestrial* radiation, or, G_{on} 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} fluctuates mainly[1] due to two factors:

- Variation of solar activity
- Earth-Sun distance changing

The solar activity variation has periodic and aperiodic components. The main periodic component is the approximately 11-year solar cycle (or sunspot cycle), mainly related to the gradual fluctuation of the number of sunspots observed. The incident radiation on a surface located at the annually mean Earth-Sun distance (1-AU)[2] and oriented normally to the direction of the incoming solar radiation is known as *Total Solar Irradiance (TSI)*. Solar activity variation, in a period of few years, seems to result in quite of minor significance fluctuations in TSI (around 0.1%) as it can be seen from the readings in Figure 3.2 provided by the NASA-sponsored satellite mission SORCE.[3] Hence, a *solar constant*, or, G_{sc} can be defined as TSI's long-term average. This solar constant is valued at $1360.8 \pm 0.5 W/m^2$ based on recent estimations [30].

[1] We note that in our method of estimating G_T, we do not consider the effects due to the movement and/or the activity variation of any space object that emits, reflects and/or blocks light besides the sun. Although, for instance, a possible solar eclipse would highly affect G_{on}.

[2] An *Astronomical Unit (AU)* is a unit of length equal to about 149,597,870.7 kilometers (92,955,807.3 mi) [38] or approximately the yearly mean Earth-Sun distance.

[3] *The Solar Radiation and Climate Experiment (SORCE)* provides state-of-the-art measurements of incoming x-ray, ultraviolet, visible, near-infrared, and total solar radiation (*http://www.nasa.gov/*).

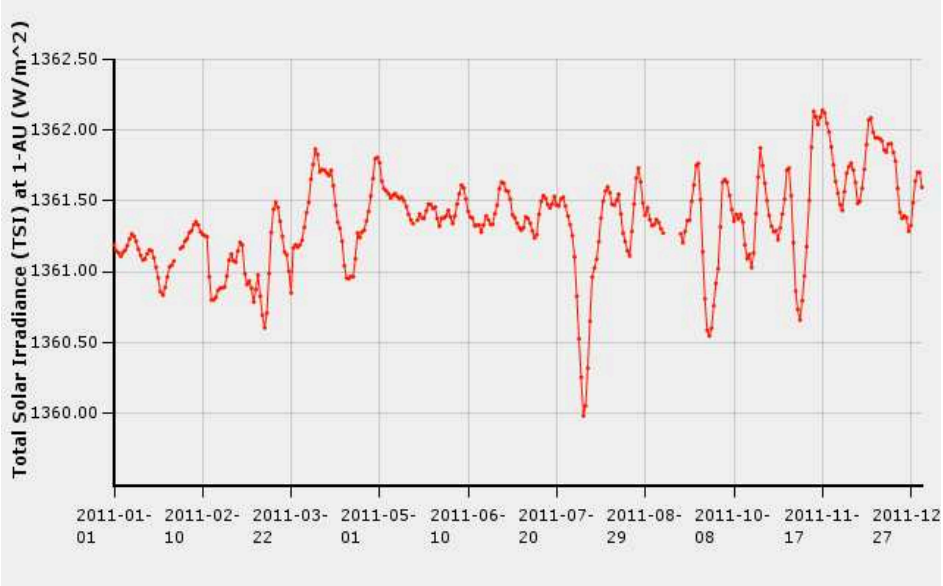

Figure 3.2: A typical TSI report chart provided by SORCE (*http://www.nasa.gov/*)

Since the earth's orbit is eliptic, the earth-sun distance varies during a year. With the distance variation being $\pm 1.7\%$ from the average, G_{on} is expected to vary significantly from the average value G_{sc}. Hence, by noting D_0 as the yearly mean Earth-Sun distance, D as the Earth-Sun distance and applying the inverse-square law,[4] we derive with Equation 3.1 [14].

$$G_{on} = G_{sc} \left(\frac{D_0}{D} \right)^2 \tag{3.1}$$

With G_{on}'s intra-day variations being considered negligible, day-to-day squared fraction $(D_0/D)^2$ is given by Equation 3.2 [14].

[4]In physics, an inverse-square law states that the magnitude of a physical quantity is inversely proportional to the square of the distance from the source of that physical quantity.

$$\left(\frac{D_0}{D}\right)^2 = 1.00011 + 0.034221cos(x) + 0.001280sin(x) + 0.000719cos(2x) + 0.000077sin(2x) \quad (3.2)$$

where,

$$x = \frac{360(n-1)}{365^\circ} \quad (3.3)$$

and n is the day of the year (starting from $1 = 1^{st}$ of January).

The above process is the process we use to derive the G_{on} estimates. Note though, that G_{on} can also be estimated with Equations 3.4 and 3.5 (found in [68] and [14] respectively). As illustrated in Figure 3.3, any of these three approaches can be used interchangeably, to estimate G_{on}.

$$G_{on} = G_{sc}\left(1 + 0.033cos\left(\frac{360n}{365}\right)^\circ\right) \quad (3.4)$$

$$G_{on} = G_{sc}\left(1 + 0.034cos\left(\frac{360n}{365.25}\right)^\circ\right) \quad (3.5)$$

For the evaluation of our G_{on} estimation methodology, derived G_{on} values have also been compared with corresponding values provided by *Ecole des Mines de Paris* through the *SODA (Solar Radiation Data)* online service.[5] The respective G_{on} values can be seen in Figure 3.4. The shift observed on the horizontal axis has been considered to be due to different G_{sc} valuing, as the value recommended some years ago was slightly higher [30].

[5] *http://www.soda-is.com/*

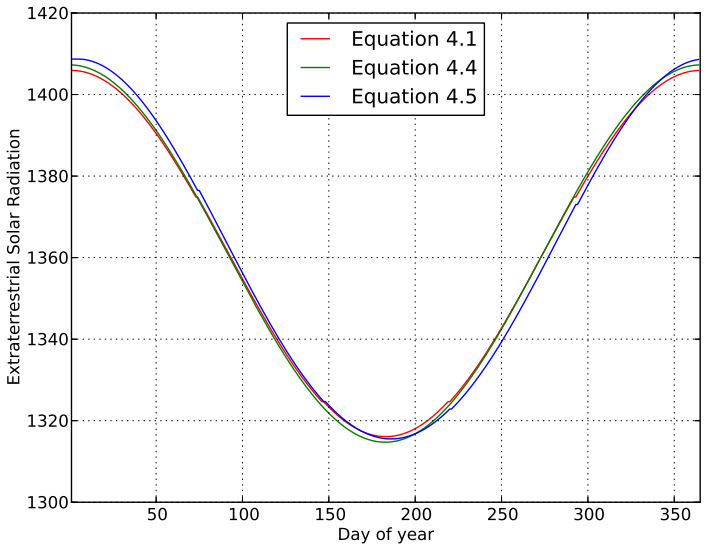

Figure 3.3: Different G_{on} estimation approaches

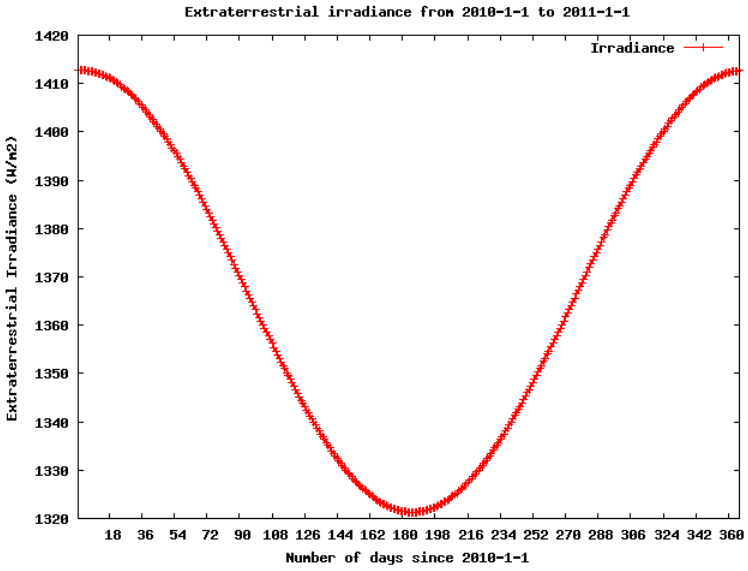

Figure 3.4: G_{on} estimates coming from SODA (*http://www.soda-is.com/*)

Having the G_{on} estimated and given a cloud coverage level N, the total incident radiation on an *arbitrarily oriented* (earth/terrestrial) surface, $G_T^{arb}(N)$, 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 as shown in Figure 3.5. Hence, $G_T^{arb}(N)$ can be calculated from Equation 3.6 [37].

$$G_T^{arb}(N) = G_B^{arb}(N) + G_D^{arb}(N) + G_R^{arb}(N) \tag{3.6}$$

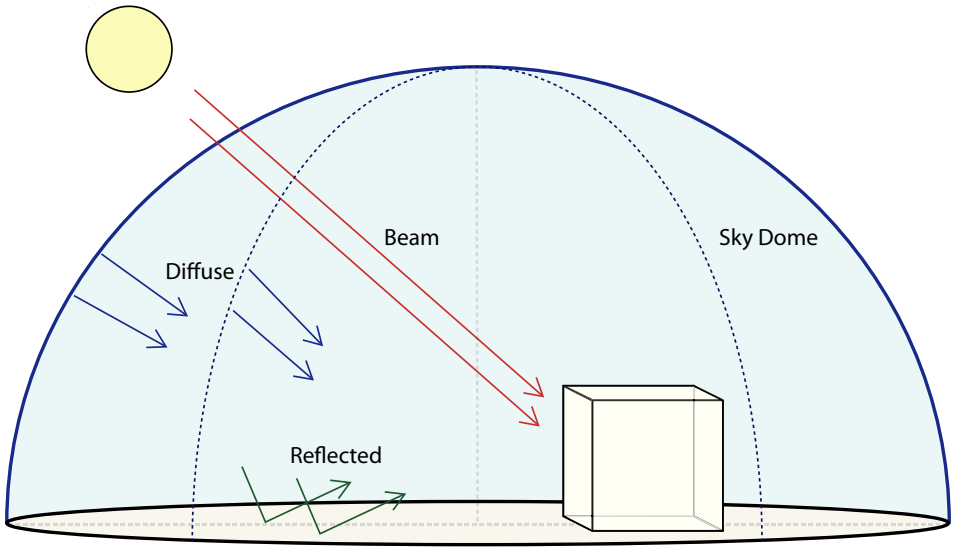

Figure 3.5: $G_T^{arb}(N)$ components

The first component, $G_B^{arb}(N)$ is calculated from Equation 3.7.

$$G_B^{arb}(N) = G_{on}\tau_b\tau_{c_b}cos\theta_s \tag{3.7}$$

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; and τ_b is the clear sky atmospheric transmittance coefficient for beam solar radiation.

The second component, $G_D^{arb}(N)$ is given by the "isotropic" Equation 3.8, which assumes that every point of the celestial sphere emits light with equal radiance [32].

$$G_D^{arb}(N) = G_{on}cos\theta_z\tau_d\tau_{c_d}\frac{1+cos\beta}{2} \tag{3.8}$$

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 and last component, $G_R^{arb}(N)$, is calculated by the "isotropic" Equation 3.9, which assumes that the ground is horizontal, of infinite extent, and reflects uniformly to all directions [37].

$$G_R^{arb}(N) = \rho G_T^{hor}(N)\frac{1-cos\beta}{2} \tag{3.9}$$

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

Figure 3.6 illustrates the aforementioned solar angles (i.e θ_s and θ_z) along with the aforementioned inclination angle of the surface, β.

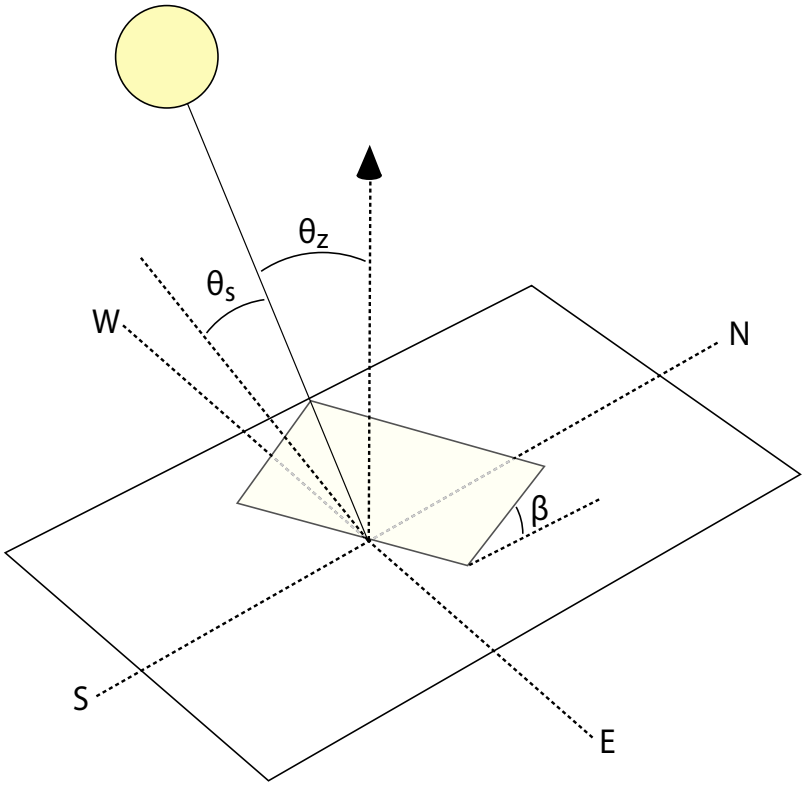

Figure 3.6: Solar position and surface orientation angles

Note that, when considering $G_R^{arb}(N)$ on a horizontal surface, β is zero ($\beta = 0$) and hence so is $G_R^{hor}(N)$ ($G_R^{hor}(N) = 0$). As a consequence, the total incident radiation *on a horizontal surface*, $G_T^{hor}(N)$, can be calculated through Equation 3.10.

$$G_T^{hor}(N) = G_B^{hor}(N) + G_D^{hor}(N) \tag{3.10}$$

The clear sky atmospheric transmittance coefficient for *beam* solar radiation (τ_b) is estimated from the Equation 3.11 proposed by Hottel [22].

$$\tau_b = \alpha_0 + \alpha_1 exp(-k/cos\theta_z) \tag{3.11}$$

where, α_0,α_1 and k are constants which are calculated from Equations 3.12, 3.13 and 3.14 respectively [22].

$$\alpha_0 = r_0[0.4237 - 0.00821(6 - A)^2] \tag{3.12}$$

$$\alpha_1 = r_1[0.5055 + 0.00595(6.5 - A)^2] \tag{3.13}$$

$$k = r_k[0.2711 + 0.01858(2.5 - A)^2] \tag{3.14}$$

where, A is the altitude of the location in km, and r_0, r_1 and r_k are correction factors for four different climate types put forth by Hottel [22]. In our implementation, we utilize the correction factors for two particular climate types (midlatitude summer and midlatitude winter).

Subsequently, the clear sky atmospheric transmittance coefficient for diffuse solar radiation τ_d can be estimated from Equation 3.15 [16].

$$\tau_d = 0.271 - 0.294\tau_b \tag{3.15}$$

The θ_z and θ_s angles are estimated through the methods described in the "*Solar Position Algorithm for Solar Radiation Applications*" technical report.[6] [58, 59] A web-based application for calculating the solar position, based on this report, is provided online[7] by *NREL's Measurement and Instrumentation Data Center*. This application has been utilized in verifying our implementation.

[6] In "*Solar Position Algorithm for Solar Radiation Applications*" is described "a procedure for a Solar Position Algorithm (SPA) to calculate the solar zenith and azimuth angle with uncertainties equal to ±0.0003 in the period from the year −2000 to 6000".

[7] *http://www.nrel.gov/midc/solpos/spa.html*

3.1.1.2 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 3.7 and 3.8 can be expressed for a *horizontal* surface as

$$G_B^{hor}(N) = G_B^{hor}(0)\tau_{c_b} \tag{3.16}$$

$$G_D^{hor}(N) = G_D^{hor}(0)\tau_{c_d} \tag{3.17}$$

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

Solving Equations 3.16 and 3.17 to 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 Equations 3.7 and 3.8. Now, $G_B^{hor}(0)$ and $G_D^{hor}(0)$ can be estimated via Equations 3.7 and 3.8 by assuming an 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 follow the two step procedure:

1. First we develop a *Cloud cover Radiation Model (CRM)* 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 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.

2. Then, we have to decompose the estimated $G_T^{hor}(N)$ back to $G_B^{hor}(N)$ and $G_D^{hor}(N)$. For this second step, we employ a readily available, hourly-valued [8] *Diffuse Ratio Model (DRM)* developed especially for our region of interest [13]. The underlying concept of this model is based on the propositions of Liu and Jordan [33] which consists of establishing empirical correlation between the clearness index (Equation 3.18) and the diffuse fraction of horizontal radiation (Equation 3.19).

$$K_d = \frac{G_{Thor}(N)}{G_{on} cos\theta_z} \tag{3.18}$$

$$F_d = \frac{G_{Dhor}(N)}{G_{Thor}(N)} \tag{3.19}$$

The empirical correlations of the utilized DRM are stated below:

- for $K_d \leq 0.21$:

$$F_d = 0.995 - 0.081 K_d \tag{3.20}$$

- for $0.21 < K_d \leq 0.76$:

$$F_d = 0.724 + 2.738 K_d - 8.32 K_d^2 + 4.967 K_d^3 \tag{3.21}$$

- for $K_d > 0.76$:

$$F_d = 0.180 \tag{3.22}$$

[8] Radiation over an hour (in Wh/m^2) is numerically equal to the mean radiation during this hour (in W/m^2) so radiation values at the instant time scale can be, to a certain extent, replaced by hourly ones [37].

Hence, $G_{Dhor}(N)$ is estimated by Equation 3.19 and $G_{Bhor}(N)$ is consequently estimated by Equation 3.10.

We now detail our approaches to developing a CRM (step *one* above).

3.1.1.3 Non-Linear Equation Models

Here we describe the non-linear equation models we test-evaluated, with the purpose of adopting one for our CRM. These models attempt to approximate $G_T^{hor}(N)/G_T^{hor}(0)$ ratio, which is known to be independent of the season and solar elevation [28]. (Note that $G_T^{hor}(0)$ quantities can be easily calculated by our all-sky radiation model, via Equation 3.10 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 the work of Kasten & Czeplak [28] which was 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 3.1 summarizes the various observable sky conditions along with their corresponding quantitative expression.

Table 3.1: Sky condition (provided in [66])

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 Kasten & Czeplak proposal after equipping it with an additional regression (correction) coefficient:

$$G_T^{hor}(N)/G_T^{hor}(0) = 1 + B_{0,0}(N/8)^{B_{0,1}} + B_{0,2} \tag{3.23}$$

We then use least-squares fitting to estimate the $B_{i,j}$ 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 3.24 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 3.1). The second method we propose is a third-degree polynomial, described in Equation 3.25; 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 3.26. These models are shown in the following equations, where $G_T^{hor}(N)/G_T^{hor}(0)$ is the dependent variable, and 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 irradiance measurements we accumulated, as we detail in Section 4.1.2.

$$\begin{aligned} G_T^{hor}(N)/G_T^{hor}(0) &= B_{1,0}(N/8)^4 + B_{1,1}(N/8)^3 \\ &+ B_{1,2}(N/8)^2 + B_{1,3}(N/8) \\ &+ B_{1,4} \end{aligned} \tag{3.24}$$

$$\begin{aligned} G_T^{hor}(N)/G_T^{hor}(0) &= B_{2,0}(N/8)^3 + B_{2,1}(N/8)^2 \\ &+ B_{2,2}(N/8) + B_{2,3} \end{aligned} \tag{3.25}$$

$$G_T^{hor}(N)/G_T^{hor}(0) = \frac{1}{1 + e^{-B_{3,0}(N/8+B_{3,1})}} \tag{3.26}$$

3.1.1.4 Informed Non-Linear Equation Models

In order to estimate $G_T^{hor}(0)$ we made the assumption, based on the work of Hottel [22], that the clear sky atmospheric transmittance coefficients (i.e. τ_b and τ_d) differ only due to the altitude differences throughout our region of interest. However, clear

sky atmospheric transmittance coefficients differ due to some other factors which cannot be considered to be the same throughout our region of interest. The reduction in the transparency of air results mainly from the scattering of light by tiny suspended particles (water droplets, ice crystallites, and dust and smoke particles) and from the absorption of light by water vapor. As such, clear sky atmospheric transmittance coefficients depend on place and time.

It has been noted in the literature [34] that air transparency depends on dew point temperature, T_d. This phenomenon has been suggested to be associated with the synoptic weather pattern as "different origins of air mass will bring in different amounts of water vapor in the vertical column"[34, 53].

Thus, the incorporation of T_d in our model is expected to improve our estimates. However, the dew point temperature, T_d is highly dependent on location and season/time; and so is the case for the ambient temperature, T_a. Therefore, we expect that by taking the difference, T_{d-a}, between these two quantities, as seen in Equation 3.27, we can end up with a quantity that is less dependent on location and season/time—and we can then use this quantity for training.

$$T_{d-a} = T_d - T_a \tag{3.27}$$

However, not all weather stations provide forecasting reports for T_d. Therefore, a method for the calculation of T_d from the more commonly provided T_a and *Relative Humidity*, or, RH is utilized.

The method utilized is the August-Roche-Magnus formula [5, 15], considered valid

for:

$$0°C < T_a < 60°C$$

$$1\% < RH < 100\%$$

$$0°C < T_d < 50°C$$

Here we assume that there will not be significant inaccuracies outside the above ranges neither. Formally, T_d is calculated from T_a and RH via Equation 3.28.

$$T_d(T_a, RH) = \frac{\nu_2 \lambda(T_a, RH)}{\nu_1 - \lambda(T_a, RH)} \tag{3.28}$$

where,

$$\lambda(T_a, RH) = ln(RH/100) + \frac{\nu_1 T_a}{\nu_2 + T_a} \tag{3.29}$$

and $\nu_1 = 17.271$, $\nu_2 = 237.7$.

In order to incorporate T_{d-a} quantity in our model we test-evaluated four (4) additional non-linear equation models. These "informed" non-linear equation models were build on top of the derived non-linear equation models (described in Section 3.1.1.3). These models are collectively shown in Equation 3.30.

$$\frac{G_T^{hor}(N)}{G_T^{hor}(0)} = F_i(N) + C_{i,3} T_{d-a}^3 + C_{i,2} T_{d-a}^2 + C_{i,1} T_{d-a} + C_{i,0} \tag{3.30}$$

where, the quantity $(G_T^{hor}(N)/G_T^{hor}(0) - F_i(N))$ is the dependent variable, T_{d-a} is the independent one, and $F_i(N)$ stands for the $(G_T^{hor}(N)/G_T^{hor}(0)$ estimates produced by its one of the derived (already fitted) non-linear equations described in the previous paragraph (i.e., Equations 3.23 - 3.26). The various $C_{i,j}$ coefficients have been estimated by employing least-squares fitting on irradiance measurements we accumulated, as it is detailed in Section 4.1.3.

3.1.1.5 Development of an MLP Network

In addition to evaluating the predictive performance of the non-linear equations above (both informed and not), we also trained an MLP network with one hidden layer [20, 41]. 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 T_a and RH parameters are direct inputs to the MLP network as any appropriate relationship between T_a and RH could be developed from the MLP. The use of T_a and RH inputs for network training was also inspired by [60], which suggests that temperature and relative humidity data can be utilized to replace missing irradiance measurements in a dataset.

3.1.1.6 Our Nine (9) CRM Approaches

In the previous paragraphs we presented nine (9) different CRM approaches that we evaluate-test, with the purpose of adopting one for our CRM in our region of interest. Summing these up, the nine (9) CRM approaches are:

- Four (4) non-linear equation models; Equations 3.23, 3.24, 3.25 and 3.26
- Four (4) informed non-linear equation models, trained on top of the "simple" non-linear equation models; collectively shown in Equation 3.30
- An MLP network; described in Section 3.1.1.5

3.1.2 Effects of Angle of Incidence, Soil and Dirt

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 the PV panel depends on their 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—including $G_T^{arb}(N)$'s components—are considered in the estimation of the panel's *effective incident radiation*, G_{eff}. To estimate G_{eff}, we follow the procedures detailed in [42, 37]. 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 types of modules can be incorporated into our model in a straightforward manner [42].

3.1.3 Estimating PV Module Operating Temperature

A further factor that has to be taken into consideration when estimating power output of a PV system, is the module's operating temperature. Given the particular nuances of current PV technology, lower operating temperatures favor their 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 a variety of factors, such as the module-specific attributes, as well as 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 [65]. For the needs of our work, we utilize the model put forth by Skoplaki [63], which ties T_c to the panel's effective incident radiation, G_{eff}, the prevailing wind speed, V, and the ambient temperature T_a. The model is described in Equation 3.31 and utilizes a mounting coefficient, ω for the appropriate adjustments in respect to the PV module's mounting type. The respective values of coefficient ω are stated in Table 3.2.

$$T_c = T + \omega \left(\frac{0.32}{8.91 + 2.0V} \right) G_{eff} \tag{3.31}$$

Table 3.2: PV array mounting type coefficient

PV array mounting type	ω
Free standing	1.0
Flat roof	1.2
Sloped roof	1.8 (1.0-2.7)
Building integrated	2.4 (2.2-2.6)

3.1.4 Predicting PVS Power Output

Based on T_c, G_{eff}, and a PV module's characteristics, a number of conceptual and empirical estimation models have been developed for the calculation of a PV module's maximum power output, P_m. [64, 37]. Here, based on a comparison of Predictive Models for Photovoltaic Module Performance performed by the National Renewable Energy Laboratory (NREL) [40], we adopt the *PVForm* model [45], 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 [49]; when such measurements are available, our web-based tool automatically utilizes the *Improved PV* model [40], which is able to incorporate them successfully.

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),[9] and negligible wiring, inverter, or other losses. In order to account for such losses, an empirical "efficiency" factor, k_e has been used for the effective power output, P_{eff} to be computed as $P_{eff} = k_e P_m$.[10]

In more detail, the PV module's maximum power output models incorporated in our work are the following:

PVForm Model The *PVForm* model [45] for $G_{eff} > 125W/m^2$ considers the classic *power-temperature coefficient* model as stated in Equation 3.32, while for

[9] An MPPT is a high efficiency electronic controller that varies a PV module's electrical operating point in order to maximize power output.

[10] The value of k_e is user-provided, and should correspond to the inverter efficiency factor, if an inverter is used—adjusted to best fit the system.

$G_{eff} \leq 125W/m^2$, Equation 3.33 is used:

- If $G_{eff} > 125W/m^2$:

$$P_m = \frac{G_{eff}}{G_{SRC}} P_{mSRC}[1 + \gamma(T_c - T_{cSRC})] \tag{3.32}$$

- If $G_{eff} \leq 125W/m^2$:

$$P_m = \frac{0.008G_{eff}^2}{G_{SRC}} P_{mSRC}[1 + \gamma(T_c - T_{cSRC})] \tag{3.33}$$

where, γ is the maximum power correction factor for temperature (in $^\circ C^{-1}$) and SRC subscripts denote performance at *Standard Reporting Conditions*. These conditions define performance at an incident sunlight, G_{SRC} of $1000W/m^2$, a cell temperature, T_{cSRC} of $25^\circ C(77^\circ F)$ and an *air mass coefficient (AM)*[11] of 1.5.

Improved PV Model The *Improved PV* model [40] is stated below:

- If $G_{eff} > 200W/m^2$:

$$P_m = P_{mSRC}\left[\frac{G_{eff}}{G_{SRC}}[1 + \gamma(T_c - T_{cSRC})] - k_p\frac{G_{SRC} - G_{eff}}{G_{SRC} - 200}\right] \tag{3.34}$$

- If $G_{eff} \leq 200W/m^2$:

$$P_m = P_{mSRC}\left[\frac{G_{eff}}{G_{SRC}}[1 + \gamma(T_c - T_{cSRC})] - k_p\left[1 - \left(1 - \frac{G_{eff}}{200}\right)^4\right]\right] \tag{3.35}$$

The irradiance correction factor, k_m, is determined with Equation 3.36.

$$k_p = \frac{P_m(G_L, T) - P_{meas}(G_L, T)}{P_{mSRC}} \tag{3.36}$$

where:

[11] The air mass coefficient, or AM, defines the ratio of the direct optical path length through Earth's atmosphere over the path length vertically upwards.

- G_L = effective low irradiance, $\sim 200W/m^2$
- $P_m(G_L, T) = P_m$ from Equation 3.32 for G_L and T conditions
- $P_{meas}(G_L, T)$ = measured P_m for G_L and T conditions.

3.2 Power Output Estimation Model for Wind Turbine Generators

Part of our system's functionality is predicting the power output of *Wind Turbines Generators (WTGs)* at specified locations. Wind-based generation prediction is done based on a standard estimation method; we use the fact that the predicted energy output depends on the so-called *power curve* of each turbine, determining its output based on predicted wind speeds. Power curves are typically sigmoid [70], as seen in Figure 3.7: at low wind speeds, the power generated is low, and then increases rapidly for increasing wind speeds, until eventually leveling-off for high wind speeds [21, 9].

Moreover, a WTG produces power only for wind speeds within the bounds of *cut-in* and *cut-out wind speed* limits. In particular, *cut-in Wind Speed* stands for the lowest wind speed value that exerts sufficient torque on the turbine blades to make them rotate, and cut-out wind speed stands for the lowest wind speed value at which the WTG's braking system is employed to bring the rotor to a standstill in order to avoid potential damage to the rotor. These values are found in almost every WTG data sheet.

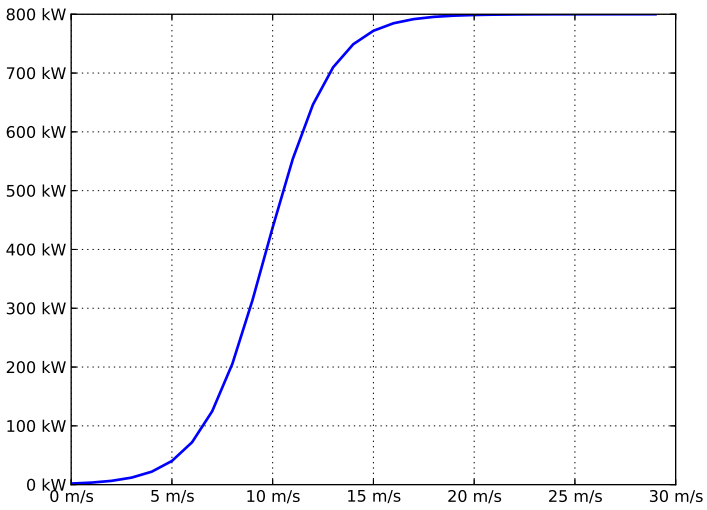

Figure 3.7: A typical WTG power curve modeled with Equation 3.37 ($\alpha_p = 0.625$, $\beta_p = 9.7$, $NomCapacity = 800$)

Formally, the energy generated by producer i at period t is given by Equation 3.37, when the prevailing wind speed is within the cut-in and cut-out wind speed limits. Otherwise the produced power is zero.

$$prod_{i,t}(V_t) = \frac{NomCapacity_i}{1 + e^{\alpha_p * (\beta_p - V_t)}} \tag{3.37}$$

Here, $NomCapacity$ is the turbine's nominal capacity, describing its maximum power (and, subsequently, energy per hour) output for "optimal" wind speed; V_t represents the predicted wind speed at t (and e is Euler's number); and, finally, the α_p and β_p parameters are user-specified or given default values depending on turbine type.

Chapter 4

Implementation

In this chapter we present the implementation details of our work. In Section 4.1 we describe the process of training and evaluating our methods for turning Cloud coverage-to-radiation and present the respective results, and in Section 4.2 we present some additional, extra, results of our study.

4.1 Evaluation of the PVS Power Output Estimation Model

In this section, we first describe the process we used to build a Mediterranean belt-specific dataset of weather observations for the training and evaluation of our models (Section 4.1.1). Then, we describe how we used this dataset to determine the coefficients of our proposed non-linear approximation equations (both informed and not) for our area of interest (Section 4.1.2 and Section 4.1.3), and train our neural network (Section 4.1.4). Following that, we evaluate all our nine irradiance under cloud

coverage estimation models; and derive and report the final power output prediction performance of our approach (Section 4.1.5).

4.1.1 Building the Observations Dataset

For the purposes of our research, archival meteorological data was drawn from the *Weather Underground* database for 9 regions in the Mediterranean belt, and 1 region in Northern Europe.[1] Specifically, we drew data for *sky condition*[2] (qualitative observations), *solar radiation* (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 4.1 and Figure 4.1.

To build our final dataset, observations with solar radiation out of bounds [0, $1.2G_{on}cos\theta_z$] [47] were excluded. Furthermore, observations with unusually high/low temperature readings (over/under the regional historical high/low 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 dataset ranges (of no more than 20,000 observations) within regions, we reduced the larger dataset ranges by progressively retaining every other observation. Subsequently the data was collated and the total

[1] In the case of Chania, Greece the respective archival meteorological data have been provided by the National Observatory Lab of Athens.

[2] There were many cases where the sky condition's observations provided were not actual but only estimated "observations". In such cases there have been observed sudden changes of sky condition (from clear to overcloud or overcloud to clear), especially during very early in the morning or late at night; during the sunlight and sunset of the sun respectively (such phenomena have been though to occur due to a systematic divergence of the clear sky model expected to be utilized by the weather station-provider). Such weather stations have been excluded and this process can be by itself considered as an initial quality control test of the related data.

observations were distributed in two sets: a training and a testing set. The training set was used to estimate the coefficient parameters of our equation models, as well as to train the MLP network. The testing set was used to evaluate the respective goodness-of-fit of all nine approaches (in the case of the MLP network, where *early stopping* [20] is applied, half of the testing set was used for validation and half for evaluation purposes).

Table 4.1: The final experimental dataset

		Stations				
Country	Location	Latitude	Longitude	Altitude	Range	Data[a]
Spain	Gava, Barcelona	N 41° 18' 22"	E 1° 59' 42"	120 ft.	2009 - 2011	14275
	Pantano de Cubillas, Albolote, Granada	N 37° 17' 8"	W 3° 40' 12"	2224 ft.	2011 - 2012	15520
	Patraix, Valencia	N 39° 27' 45"	E 0° 23' 43"	102 ft.	2010 - 2011	17498
Greece	Chania, Crete	N 35° 31' 56"	E 24° 4' 11"	400 ft.	2010 - 2011	15252
	Kato Pylea, Thessaloniki	N 40° 33' 59"	E 22° 58' 28"	7 ft.	2010 - 2011	13836
France	Montauroux, Provence	N 43° 37' 9"	E 6° 46' 59"	1250 ft.	2011 - 2012	17662
	Orange,Provence Alpes Cote d'Azur	N 44° 7' 49"	E 4° 49' 40"	154 ft.	2011 - 2012	17600
Italy	Mezzana Bigli, Lombardia	N 45° 3' 37"	E 8° 50' 43"	249 ft.	2011 - 2012	18642
Portugal	Lordelo do Ouro, Porto	N 41° 9' 22"	W 8° 39' 45"	284 ft.	2011 - 2012	18612
Denmark	Lake Arresoe, Ramloese, Helsinge	N 56° 0' 59"	E 12° 5' 15"	50 ft.	2011 - 2012	45087

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

Figure 4.1: The final experimental dataset (Google Map 2012)

4.1.2 Least-Squares Fitting of the Non-Linear Equation Models

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), as seen in Table 3.1, 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 Equation 3.23, $B_{0,0} = -0.6287$, $B_{0,1} = 1.1653$ and $B_{0,2} = 0.034$; for Equation 3.24, $B_{1,0} = 1.63$, $B_{1,1} = -3.047$, $B_{1,2} = 1.531$, $B_{1,3} = -0.7411$ and $B_{1,4} = 1.037$; for Equation 3.25, $B_{2,0} = 0.198$, $B_{2,1} = -0.4371$, $B_{2,2} = -0.3865$ and $B_{2,3} = 1.033$; and for Equation 3.26, $B_{3,0} = -3.6772$ and $B_{3,1} = -0.8665$ and the derived equations respectively are:

$$G_T^{hor}(N)/G_T^{hor}(0) = 1 - 0.6287(N/8)^{1.1653} + 0.034 \tag{4.1}$$

$$\begin{aligned} G_T^{hor}(N)/G_T^{hor}(0) &= 1.63(N/8)^4 - 3.047(N/8)^3 \\ &+ 1.531(N/8)^2 - 0.7411(N/8) \\ &+ 1.037 \end{aligned} \tag{4.2}$$

$$\begin{aligned} G_T^{hor}(N)/G_T^{hor}(0) &= 0.198(N/8)^3 - 0.4371(N/8)^2 \\ &- 0.3865(N/8) + 1.033 \end{aligned} \tag{4.3}$$

$$G_T^{hor}(N)/G_T^{hor}(0) = \frac{1}{1 + e^{+3.6772(N/8 - 0.8665)}} \tag{4.4}$$

4.1.3 Least-Squares Fitting of the Informed Non-Linear Equation Models

In order to fit our proposed informed non-linear curves to our dataset, the $\langle N, G_T^{hor}(N)/G_T^{hor}(0) - F_i(N)\rangle$ pairs where used to estimate the vector of $C_{i,j}$ coefficients using the least square fitting technique. Here, $F_i(N)$ stands for the $G_T^{hor}(N)/G_T^{hor}(0)$ estimates produced by its one of the derived (already fitted) non-linear equations presented in the previous paragraph (i.e. Equations 4.1- 4.4), and $N, G_T^{hor}(N)/G_T^{hor}(0)$ are derived by the same procedure described in Section 4.1.2. The derived informed equations for each one of the non-linear equations (as presented in Section 4.1.2) are:

$$\frac{G_T^{hor}(N)}{G_T^{hor}(0)} = F_0(N) - 0.00003T_{d-a}^3 - 0.00187T_{d-a}^2 - 0.03405T_{d-a} - 0.14446 \quad (4.5)$$

where, $F_0(N)$ stands for the right-hand side of Equation 4.1.

$$\frac{G_T^{hor}(N)}{G_T^{hor}(0)} = F_1(N) - 0.00003T_{d-a}^3 - 0.00183T_{d-a}^2 - 0.03367T_{d-a} - 0.14158 \quad (4.6)$$

where, $F_1(N)$ stands for the right-hand side of Equation 4.2.

$$\frac{G_T^{hor}(N)}{G_T^{hor}(0)} = F_2(N) - 0.00003T_{d-a}^3 - 0.00185T_{d-a}^2 - 0.0338T_{d-a} - 0.1435 \quad (4.7)$$

where, $F_2(N)$ stands for the right-hand side of Equation 4.3.

$$\frac{G_T^{hor}(N)}{G_T^{hor}(0)} = F_3(N) - 0.00003T_{d-a}^3 - 0.0019T_{d-a}^2 - 0.03711T_{d-a} - 0.15046 \quad (4.8)$$

where, $F_3(N)$ stands for the right-hand side of Equation 4.4.

4.1.4 Training the MLP Network

In order to train our neural network the testing set was divided into two equal parts by retaining every other observation; the validation set and a new testing set. The neural networks architecture comprises one hidden layer with five input and one output node as seen in Figure 4.2.

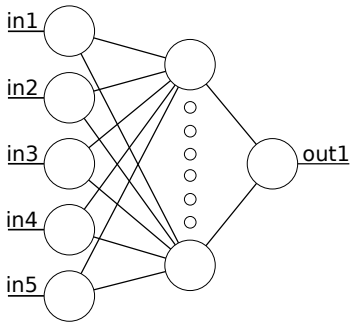

Figure 4.2: Neural network architecture

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 stopping* neural network training technique [20].

4.1.5 Evaluating the CRM (Cloud Cover Radiation Model)

By now we have developed nine (9) different CRM approaches that we evaluate-test, with the purpose of adopting one for our CRM in our region of interest. Collectively our nine (9) CRM approaches are illustrated in Table 4.2 with appropriate abbreviations and references to the respective equations (when needed).

Table 4.2: Our nine (9) CRM approaches

Abbreviation	Details	Equation
K&C-Ext	Equation based on the extended Kasten & Czeplak formula	4.1
4^{th}-degr.Pol.	4^{th}-degree polynomial equation	4.2
3^{rd}-degr.Pol.	3^{rd}-degree polynomial equation	4.3
Sigmoid	Sigmoid equation	4.4
Informed K&C-Ext	Informed equation utilizing K&C-Ext	4.5
Informed 4^{th}-degr.Pol.	Informed equation utilizing 4^{th}-degr.Pol.	4.6
Informed 3^{rd}-degr.Pol.	Informed equation utilizing 3^{rd}-degr.Pol.	4.7
Informed Sigmoid	Informed equation utilizing Sigmoid	4.8
MLP	Multilayer perceptron network	–

For the evaluation of our nine CRM approaches, we calculated the following quantities:

- *Mean Absolute Percentage Error*:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{F_i - A_i}{A_i}\right| 100 \tag{4.9}$$

- *Mean Absolute Error*:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - A_i| \tag{4.10}$$

- *relative Mean Absolute Error*:

$$rMAE = \frac{MAE}{1/n\sum_{i=1}^{n} A_i} 100 \tag{4.11}$$

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 [68]. 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 behavior outside the region of interest. Finally, we note that in all cases, the observations where the solar radiation model $G_T^{hor}(0) = 0$ were excluded from the Evaluation. The evaluation results for the *four CRM least square-fitted curves* are reported in Table 4.3 while a plot of the equations can be found in Figure 4.3. We note that in Figure 4.3 K&C-Med stands for K&C-Ext.

Table 4.3: Evaluation results of the fitted non-linear curves

	Mediterranean			Denmark		
Equation	MAPE	rMAE	MAE	MAPE	rMAE	MAE
K&C-Ext	23.727	21.441	75.904	34.538	37.051	98.938
4^{th}-degr.Pol.	23.825	21.585	76.414	34.611	37.109	99.091
3^{rd}-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

MAPE & rMAE in %, MAE in W/m^2.

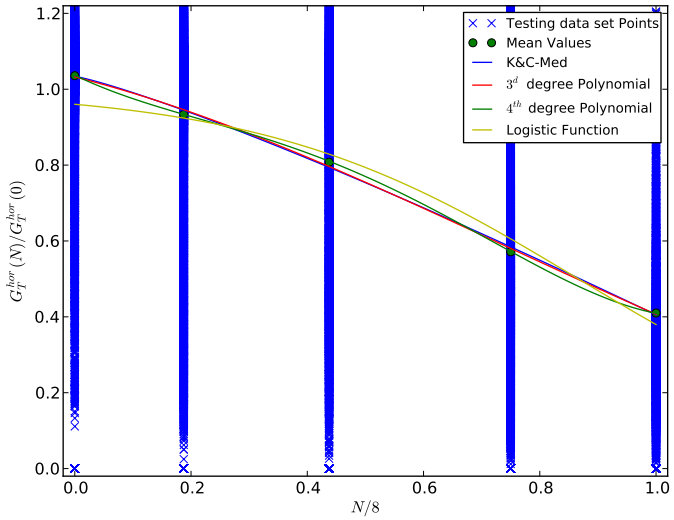

Figure 4.3: Non-linear least-squares fitted curves (K&C-Med stands for K&C-Ext)

We ran a standard one-way ANOVA [43] 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 4^{th} & 3^{rd} 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 that of the sigmoid function (i.e., the sigmoid is significantly *worse* than the others—*cf.* Table 4.3).

The evaluation results for the *four informed CRM approaches* are reported in Table 4.4.

Table 4.4: Evaluation results of the fitted informed non-linear curves

	Mediterranean			Denmark		
Equation	MAPE	rMAE	MAE	MAPE	rMAE	MAE
Informed K&C-Ext	23.133	20.39	72.185	33.981	35.953	96.004
Informed 4^{th}-degr.Pol.	23.193	20.441	72.363	34.097	36.009	96.156
Informed 3^{rd}-degr.Pol.	23.107	20.356	72.063	34.031	35.987	96.096
Informed Sigmoid	23.641	20.497	72.564	34.786	36.504	97.477

MAPE & rMAE in %, MAE in W/m^2.

Paired T-tests showed there is statistical significance (with 95% confidence) among the CRM least square-fitted curves and the informed ones except among the 3^{rd} degree polynomial and the informed one where there is a statistical significance with slightly less than 95% confidence, being 94.8%.

Then, our results for the *MLP* show *it is a winner* when compared with the eight other CRM models as seen in Table 4.5. Subsequent paired T-tests confirmed its error is indeed lower in a statistical significant manner from the CRM non-linear curves. However, there has not been statistical significant variances in the error when this has been compared with the error related to the informed non-linear equations. Moreover,

the MLP network's performance deteriorates considerably outside the Med-Belt, as it is trained on Med-Belt data.

Thus, RENES incorporates the MLP network as its CRM model inside the Med-Belt, but uses the 3^{rd} degree polynomial outside the Med-Belt (due to its simplicity and slightly better performance there when compared to the MLP and the other non-informed equations).

Table 4.5: Evaluation results of all nine (9) fitted/trained CRM approaches

	Mediterranean			Denmark		
Equation	MAPE	rMAE	MAE	MAPE	rMAE	MAE
K&C-Ext	23.727	21.441	75.904	34.538	37.051	98.938
4^{th}-degr.Pol.	23.825	21.585	76.414	34.611	37.109	99.091
3^{rd}-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
Informed K&C-Ext	23.133	20.39	72.185	33.981	35.953	96.004
Informed 4^{th}-degr.Pol.	23.193	20.441	72.363	34.097	36.009	96.156
Informed 3^{rd}-degr.Pol.	23.107	20.356	72.063	34.031	35.987	96.096
Informed Sigmoid	23.641	20.497	72.564	34.786	36.504	97.477
MLP	22.946	19.456	68.69	46.171	39.762	106.149

MAPE & rMAE in %, MAE in W/m^2.

4.2 Local Training and Evaluation

In the context of the present work, we also trained and tested our nine CRM approaches *separately* on datasets for all our 9, specified above, locations. The procedure followed is almost the same as the procedure described in the previous section (4.1). The only difference is that, for all our 9 locations, the entire datasets have been used without the prior reduction procedure of the larger dataset ranges as there is no need for equivalent dataset ranges among the 9 locations. Up to this point, the major necessity is to have, in all locations separately, representative and large enough dataset ranges. By avoiding this procedure we only reinforce these goals. The dataset ranges,[3] the derived non-linear curves (both informed, and not) coefficients, the derived MLP attributes and the derived evaluation results are stated separately for all our 9 specific locations in the tables that follow.

Conclusions

This procedure confirmed the expected superiority of the MLP network for very specific locations. Moreover we observed that certain regions are more suitable for accurate solar radiation forecasting reports than others (at least using our approaches). This result was also expected as different weather pattern enable more or less accurate modeling. For instance predicting the next day solar radiation for a location in Sahara desert could be held as a trivial process; using one day's weather pattern as a forecast for the next day could be adequate in most cases.

Although the specific location CRM approaches are not being used in our work

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

here, we envisage that they can be of value for other research work and/or applications which might need higher accuracy for very specific locations.

Table 4.6: Local CRM approaches and evaluation results for *Barcelona* ($Data^a = 28607$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Ext	-0.8186	0.8993	0.1354	–	–	30.881%	25.64%	$91.145 W/m^2$
4^{th}-degr.Pol.	-0.7155	2.292	-1.9364	-0.4161	1.1222	30.207%	24.519%	$87.159 W/m^2$
3^{rd}-degr.Pol.	0.8674	-1.0725	-0.5718	1.1241	–	30.328%	24.667%	$87.685 W/m^2$
Sigmoid	-4.2464	-0.7684	–	–	–	33.077%	27.793%	$98.796 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Ext	0.0	-0.0005	-0.0224	-0.1153		30.402%	24.902%	$88.522 W/m^2$
Informed 4^{th}-degr.Pol.	0.0	0.0001	-0.0126	-0.0868		29.696%	23.961%	$85.176 W/m^2$
Informed 3^{rd}-degr.Pol.	0.0	0.0001	-0.013	-0.0878		29.821%	24.087%	$85.623 W/m^2$
Informed Sigmoid	-0.0	-0.0013	-0.0411	-0.1755		31.075%	24.326%	$86.474 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
14						29.027 %	23.397 %	$83.187\ W/m^2$

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

Table 4.7: Local CRM approaches and evaluation results for *Granada* ($Data^a = 124314$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Ext	-0.4395	0.7285	-0.0388	–	–	25.077%	22.654%	$88.363 W/m^2$
4^{th}-degr.Pol.	-2.2015	5.7376	-4.257	0.3526	0.9446	21.641%	19.072%	$74.391 W/m^2$
3^{rd}-degr.Pol.	1.3541	-1.5986	-0.1264	0.9504	–	22.298%	19.83%	$77.35 W/m^2$
Sigmoid	-2.0336	-0.9976	–	–	–	25.662%	22.679%	$88.461 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Ext	-0.0	-0.0019	-0.0466	-0.2431		21.342%	17.346%	$67.66 W/m^2$
Informed 4^{th}-degr.Pol.	-0.0	-0.0017	-0.0405	-0.2202		20.057%	16.701%	$65.143 W/m^2$
Informed 3^{rd}-degr.Pol.	-0.0	-0.0017	-0.0408	-0.2205		20.207%	16.709%	$65.175 W/m^2$
Informed Sigmoid	-0.0	-0.0021	-0.0502	-0.2603		21.566%	17.094%	$66.676 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
14						19.67 %	15.927 %	$62.242\ W/m^2$

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

Table 4.8: Local CRM approaches and evaluation results for *Valencia* ($Data^a = 70044$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Ext	-0.7087	0.8965	0.0411	–	–	26.905%	24.727%	$91.327 W/m^2$
4^{th}-degr.Pol.	0.365	-0.0186	-0.4311	-0.5899	1.0319	26.599%	24.139%	$89.154 W/m^2$
3^{rd}-degr.Pol.	0.7081	-0.8718	-0.5105	1.0309	–	26.552%	24.078%	$88.932 W/m^2$
Sigmoid	-3.5179	-0.7541	–	–	–	28.78%	26.556%	$98.083 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Ext	0.0001	0.0023	-0.0065	-0.1845		25.467%	23.212%	$85.733 W/m^2$
Informed 4^{th}-degr.Pol.	0.0001	0.0024	-0.0046	-0.1799		25.249%	22.72%	$83.914 W/m^2$
Informed 3^{rd}-degr.Pol.	0.0001	0.0024	-0.0047	-0.1806		25.205%	22.671%	$83.734 W/m^2$
Informed Sigmoid	0.0001	0.0013	-0.0235	-0.2303		25.189%	22.26%	$82.215 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
26						24.463 %	20.974 %	77.695 W/m^2

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

Table 4.9: Local CRM approaches and evaluation results for *Chania* ($Data^a = 15252$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Ext	-0.694	2.5387	0.0939	–	–	23.432%	17.292%	$70.877 W/m^2$
4^{th}-degr.Pol.	-3.1254	6.1772	-4.6539	0.9283	1.0718	23.505%	17.389%	$71.275 W/m^2$
3^{rd}-degr.Pol.	-0.0459	-0.88	0.2482	1.0801	–	23.219%	17.077%	$69.996 W/m^2$
Sigmoid	-7.0817	-0.9358	–	–	–	22.958%	16.112%	$66.042 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Ext	-0.0002	-0.0083	-0.113	-0.4197		23.755%	17.985%	$73.72 W/m^2$
Informed 4^{th}-degr.Pol.	-0.0002	-0.0082	-0.1121	-0.4222		23.782%	18.083%	$74.122 W/m^2$
Informed 3^{rd}-degr.Pol.	-0.0002	-0.0081	-0.1109	-0.4144		23.617%	17.907%	$73.398 W/m^2$
Informed Sigmoid	-0.0002	-0.0086	-0.1224	-0.4305		24.133%	17.214%	$70.558 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
26						20.696 %	15.157 %	62.387 W/m^2

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

Table 4.10: Local CRM approaches and evaluation results for *Thessaloniki* ($Data^a = 110752$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Ext	-0.7718	1.5008	0.15	–	–	23.172%	21.38%	$74.194 W/m^2$
4^{th}-degr.Pol.	1.2777	-2.4189	0.8996	-0.5359	1.1564	23.061%	21.158%	$73.425 W/m^2$
3^{rd}-degr.Pol.	0.1251	-0.6432	-0.2579	1.153	–	23.173%	21.339%	$74.053 W/m^2$
Sigmoid	-5.7361	-0.8905	–	–	–	25.257%	23.214%	$80.558 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Ext	0.0001	0.0021	-0.0008	-0.0914		23.336%	21.451%	$74.441 W/m^2$
Informed 4^{th}-degr.Pol.	0.0001	0.0022	0.0004	-0.0883		23.169%	21.138%	$73.356 W/m^2$
Informed 3^{rd}-degr.Pol.	0.0001	0.0021	-0.0006	-0.0914		23.319%	21.382%	$74.2 W/m^2$
Informed Sigmoid	0.0001	0.0028	0.0015	-0.0875		24.201%	21.117%	$73.282 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
4						22.702 %	19.835 %	68.797 W/m^2

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

Table 4.11: Local CRM approaches and evaluation results for *Montauroux* ($Data^a = 70731$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Local	-0.6598	0.6253	0.1268	–	–	20.348%	21.809%	$68.933 W/m^2$
4^{th}-degr.Pol.	0.8946	-2.5265	2.6091	-1.6401	1.1269	20.336%	21.778%	$68.836 W/m^2$
3^{rd}-degr.Pol.	-0.7453	1.5288	-1.4454	1.1245	–	20.147%	21.522%	$68.028 W/m^2$
Sigmoid	-3.1741	-0.8954	–	–	–	22.282%	23.427%	$74.05 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Local	-0.0001	-0.0043	-0.0581	-0.1879		19.391%	19.872%	$62.812 W/m^2$
Informed 4^{th}-degr.Pol.	-0.0001	-0.0043	-0.0577	-0.1873		19.362%	19.841%	$62.715 W/m^2$
Informed 3^{rd}-degr.Pol.	-0.0001	-0.0043	-0.0575	-0.1872		19.152%	19.61%	$61.983 W/m^2$
Informed Sigmoid	-0.0001	-0.0035	-0.053	-0.1689		19.888%	19.549%	$61.791 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
4						15.925 %	15.027 %	47.553 W/m^2

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

Table 4.12: Local CRM approaches and evaluation results for *Orange* ($Data^a = 70533$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Local	-0.549	0.8152	0.1333	–	–	20.083%	17.109%	$57.717 W/m^2$
4^{th}-degr.Pol.	2.1836	-3.6217	1.7923	-0.8672	1.1271	19.617%	16.699%	$56.334 W/m^2$
3^{rd}-degr.Pol.	0.7262	-0.8445	-0.392	1.1213	–	19.33%	16.297%	$54.977 W/m^2$
Sigmoid	-3.3857	-1.0662	–	–	–	20.501%	17.574%	$59.285 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Local	-0.0001	-0.0016	-0.0068	0.0004		20.563%	17.704%	$59.724 W/m^2$
Informed 4^{th}-degr.Pol.	-0.0001	-0.0018	-0.0088	-0.0097		20.056%	17.331%	$58.465 W/m^2$
Informed 3^{rd}-degr.Pol.	-0.0001	-0.0018	-0.0086	-0.0093		19.751%	16.976%	$57.269 W/m^2$
Informed Sigmoid	-0.0001	-0.0017	-0.0141	0.0034		20.777%	16.918%	$57.072 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
26						16.318 %	13.415 %	$45.305\ W/m^2$

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

Table 4.13: Local CRM approaches and evaluation results for *Lombardia* ($Data^a = 74614$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Local	-0.5867	0.6887	0.0294	–	–	29.765%	30.852%	$87.702 W/m^2$
4^{th}-degr.Pol.	11.0467	-22.3717	14.0838	-3.3564	1.0414	28.442%	28.975%	$82.367 W/m^2$
3^{rd}-degr.Pol.	-0.3765	0.7448	-0.9525	1.0122	–	29.58%	30.437%	$86.524 W/m^2$
Sigmoid	-2.8417	-0.8504	–	–	–	28.891%	29.109%	$82.747 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Local	-0.0002	-0.0068	-0.0773	-0.1985		28.25%	27.499%	$78.17 W/m^2$
Informed 4^{th}-degr.Pol.	-0.0002	-0.0064	-0.0709	-0.1798		27.451%	26.278%	$74.702 W/m^2$
Informed 3^{rd}-degr.Pol.	-0.0002	-0.0066	-0.075	-0.2001		28.215%	27.328%	$77.687 W/m^2$
Informed Sigmoid	-0.0002	-0.0059	-0.0678	-0.2017		27.979%	26.775%	$76.114 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
26						25.116 %	21.714 %	$61.754\ W/m^2$

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

Table 4.14: Local CRM approaches and evaluation results for *Porto* ($Data^a = 37280$)

Equation	$B_{i,0}$	$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	$B_{i,4}$	MAPE	rMAE	MAE
K&C-Local	-0.8067	1.1279	0.1699	–	–	23.798%	19.291%	$74.601 W/m^2$
4^{th}-degr.Pol.	1.3804	-2.6439	1.3632	-0.9091	1.1738	23.961%	19.517%	$75.473 W/m^2$
3^{rd}-degr.Pol.	0.1047	-0.3037	-0.6087	1.1701	–	23.824%	19.317%	$74.7 W/m^2$
Sigmoid	-5.0645	-0.8536	–	–	–	26.762%	22.344%	$86.407 W/m^2$
Informed Equation	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$		MAPE	rMAE	MAE
Informed K&C-Local	-0.0004	-0.009	-0.0518	-0.0736		24.312%	20.199%	$78.11 W/m^2$
Informed 4^{th}-degr.Pol.	-0.0004	-0.0091	-0.0515	-0.0701		24.481%	20.389%	$78.846 W/m^2$
Informed 3^{rd}-degr.Pol.	-0.0004	-0.0091	-0.052	-0.0738		24.336%	20.219%	$78.189 W/m^2$
Informed Sigmoid	-0.0005	-0.0112	-0.0881	-0.1043		24.831%	19.304%	$74.65 W/m^2$
MLP Network's Number of Hidden Layer Nodes						MAPE	rMAE	MAE
26						24.27 %	18.213 %	70.547 W/m^2

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

Chapter 5

Final Photovoltaic System Power Output Prediction Performance Guarantees

For the evaluation of our tool, we employ an error propagation methodology [6], in order to accumulate each individual submodel's error and calculate the total error. The *Mean Absolute Error (MAE)* and *relative Mean Absolute Error* (rMAE) were calculated for *Photovoltaic (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).

In the following sections; we first explicitly introduce our error propagation methodology (Section 5.1); then we describe the methods for calculating/estimating the ab-

solute error at the output of each submodel (Section 5.2); we describe how we derived with the overall evaluation results (Section 5.3); we describe how we supplemented our dataset with additional, necessary for our evaluation, archival meteorological data (Section 5.4); and finally we demonstrate our evaluation results and discuss them (Section 5.5).

5.1 The Error Propagation Methodology

The method for predicting the power output of *Photovoltaic Systems (PVSs)* presented in this thesis consists of a series of relatively independent estimation steps described in Section 3.1 and illustrated in Figure 3.1. In our error propagation procedure, the absolute error produced from each submodel was calculated for each combination of PV module manufacturing technology, mounting type and tilt angle configurations. Then, that error was propagated through the "chain" of submodels, 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.

For each data point contained in our dataset, the absolute error produced from each submodel was *computed in a straightforward manner* (if we had access to the data required), or *estimated via utilizing evaluation results coming from the literature*. That said, in order to be able to calculate the absolute error produced from a submodel, either respective data or appropriate evaluation results were needed. Due to a lack of both respective data and appropriate evaluation results, *two (2) of our method's submodels have not been accounted in our evaluation procedure.*

The *first submodel we did not account* in our evaluation procedure is the last submodel of our PVS power output estimation method and in particular the *submodel considering wiring, inverter and other secondary losses* (described in Section 3.1.4). Note that we did not just set the error of the submodel negligible but we did not account the respective submodel at all. Hence, our evaluation is conducted with respect to PV modules's maximum power output (P_m) only.

The *second submodel we did not account for* in our evaluation procedure is the submodel considering *the effects of solar radiation's angle of incidence, as well as the effects of soil and dirt* (Section 3.1.2). Once again, we did not just set the error of the submodel negligible but we did not account the respective submodel at all. Hence, we have considered *the effects of solar radiation's angle of incidence, and the effects of soil and dirt* negligible. This was done by talking into consideration: 1) the relatively low participation rate of these effects in the PV modules's maximum power output (P_m), and 2) the necessity for not accounting this submodel in order to be able to utilize valuable evaluation results coming from the literature, as we detail below.

The submodel considering *the effects of solar radiation's angle of incidence, and the effects of soil and dirt* estimates the amount of incident radiation actually absorbed by the PV module; G_eff, receiving as input the *specific components* (i.e Beam, Diffuse, Reflected) of the *total incident radiation on an arbitrarily oriented surface* ($G_T^{arb}(N)$). Estimates of these components come from our *Solar Irradiance* submodel (Section 3.1.1). However, by considering *the effects of solar radiation's angle of incidence, and the effects of soil and dirt* negligible (i.e $G_{eff} = G_T^{arb}(N)$), as we do in our evaluation here, we are only interested in $G_T^{arb}(N)$ *as a whole - and not in its specific components.*

Hence, the procedure followed in our *Solar Irradiance* submodel (Section 3.1.1) can now be considered equivalent to a method which, when a horizontal orientation is considered, directly outputs the *total incident radiation on a horizontal surface* ($G_T^{hor}(N)$) as predicted from our *Cloud cover Radiation Model (CRM)*. Otherwise, it (i) decomposes the estimated $G_T^{hor}(N)$ to its components via utilizing the *Diffuse Ratio Model, DRM* (described in Section 3.1.1.2); then (ii) directly estimates $G_T^{arb}(N)$'s components from $G_T^{hor}(N)$'s components; and finally (iii) calculates $G_T^{arb}(N)$.

In more detail, $G_T^{arb}(N)$'s *beam* and *diffuse* components can be calculated directly from $G_T^{hor}(N)$'s respective components through Equations 5.1 and 5.2 respectively.

$$G_B^{arb}(N) = \frac{G_B^{hor}(N)}{cos\theta_{s(hor)}} cos\theta_{s(arb)} \tag{5.1}$$

$$G_D^{arb}(N) = G_D^{hor}(N) \frac{1 + cos\beta_{(arb)}}{2} \tag{5.2}$$

where "(arb)" and "(hor)" angles' subscripts, denote angles with respect to *arbitrary* (non-horizontal) and *horizontal* PV module's orientation respectively. Furthermore, the *reflected* $G_T^{arb}(N)$'s component can be calculated via Equation 3.9 which enables us to, finally, calculate the entire $G_T^{arb}(N)$ via Equation 3.6.

As such, the entire *Solar Irradiance* submodel can be divided into two internal submodels. The first submodel, or, submodel *HorRad* consists of the CRM approach used in order to estimate $G_T^{hor}(N)$. In particular, it consists of our *Multilayer Perceptron (MLP)* network (described in Section 3.1.1.5) since this is the one we employ in the region of the *Mediterranean Belt (Med-Belt for short)*. The second submodel, or, submodel *HorToInc* is in sequence connected to submodel *HorRad* and includes the aforementioned estimation steps in order to derive $G_T^{arb}(N)$. The submodel *HorToInc*

is used only if the PV module's orientation is *non-horizontal*.

Now, note that the absolute error produced by submodel *HorRad* can be directly computed for each data row in our dataset in a straightforward manner, as we have access to the required data. Moreover, submodel *HorToInc* have been evaluated in the work of [48]. Hence, we are able to evaluate our method with an error propagation procedure, which *would not* be feasible if we have accounted the submodel considering *the effects of solar radiation's angle of incidence, and the effects of soil and dirt* in our evaluation.

Figure 5.1 illustrates a flow chart of our method without considering the submodels we have not accounted in our evaluation.

5.2 Estimating the Absolute Error at the Output of Each Submodel

In our error propagation procedure the absolute error at the output of each submodel was calculated for each data point contained in our dataset. Here, we explicitly describe the methods for calculating the absolute error at the output of each submodel.

5.2.1 Solar Irradiance Submodel

The entire *Solar Irradiance* submodel consists of two internal submodels as seen in Figure 5.1. The first submodel, or, submodel *HorRad* estimates $G_T^{hor}(N)$. The second submodel, or, submodel *HorToInc* is used only if the PV arbitrary orientation consists a non-horizontal one. Hence in our experiments with horizontally oriented

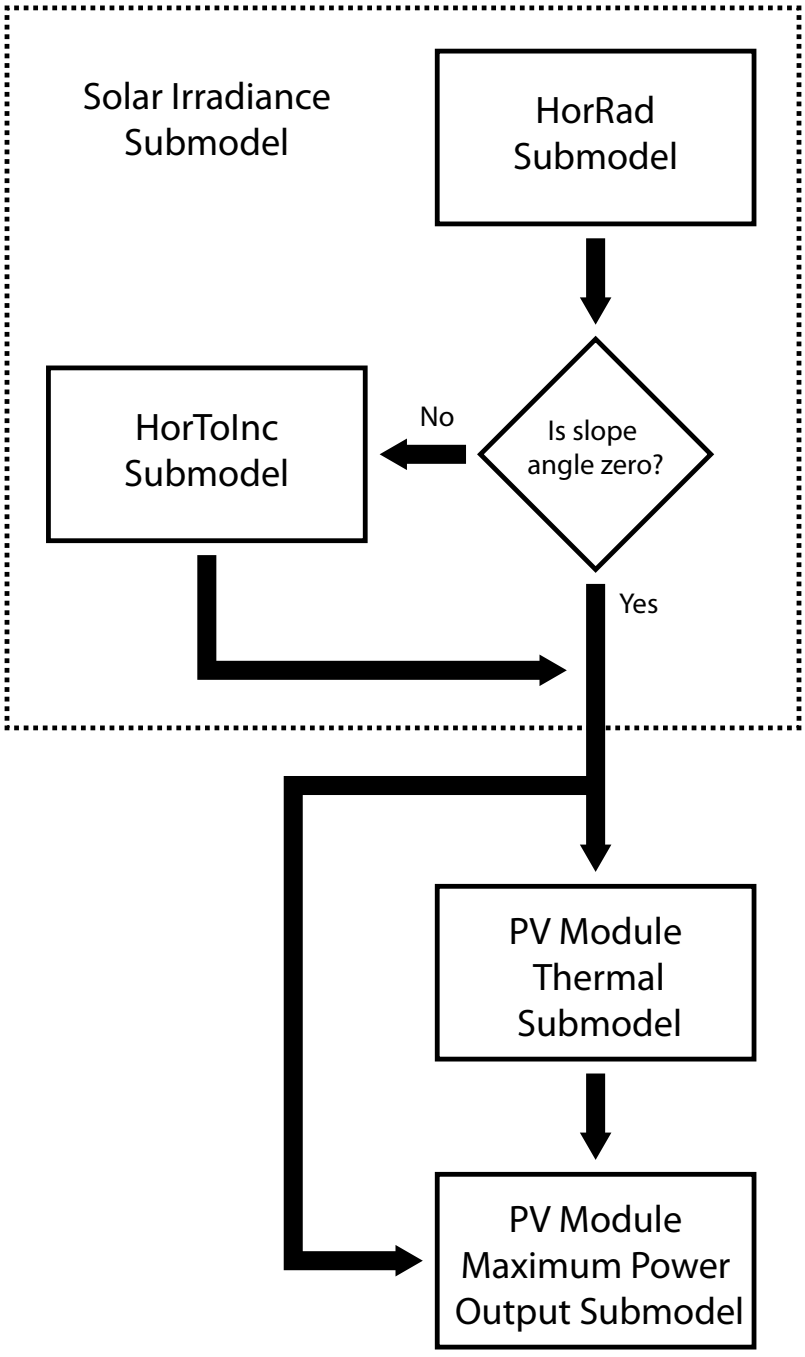

Figure 5.1: Our error propagation process

PV modules the submodel *HorToInc* has not been used/"activated" at all.

5.2.1.1 Submodel HorRad

The submodel *HorRad* is the first submodel of our PVS power output prediction method. Hence, for each data point contained in our dataset, the absolute error produced by submodel *HorRad* is also directly the absolute error at the output of the submodel.

In the case of *horizontal* orientation, we had access to the data required for irradiance measurements on a horizontal surface. Hence, the absolute error produced by submodel *HorRad* (i.e our MLP network) has been calculated for each data point contained in our dataset in a straightforward manner through $|F_i - A_i|$, where A_i represents a data-point coming from the actual (historical data) and F_i represents the corresponding forecasted (estimated) one, with i ranging from 1 to n within the dataset.

5.2.1.2 Submodel HorToInc

In the case of *not horizontally* oriented surface, the absolute error at the output of submodel *HorRad* (calculated as described in Section 5.2.1.1) was propagated through submodel *HorToInc* and was added to the absolute error produced by submodel *HorToInc*. As such, the absolute error at the output of submodel *HorToInc* was calculated for each data point contained in our dataset.

Due to a lack of required data with respect to irradiance measurements at non-zero slope angles, we utilized evaluation results coming from the literature in order to estimate the absolute error produced by submodel *HorToInc*. Specifically, we

have utilized the work of [48] where evaluation results (i.e $RMSE$[1], MBE[2], Avg[3]) regarding a typical south-facing PV panel with 45° slope angle are provided.

However, it has to be noted that submodel *HorToInc* has been evaluated in the work of [48] for the region of *the Faculty of Agriculture, University of Tehran, Karaj, Iran*. Moreover, the location dependent component of submodel *HorToInc* (i.e the DRM) has been developed for the region of the Med-Belt [13]. Hence, the evaluation results provided in [48] can, at best, be considered as upper limit guaranties for the submodel *HorToInc*'s performance inside the region of the Med-Belt. As such, our evaluation results for the power output estimates of *not horizontally* orientated PVSs can, subsequently, at best, be considered as an upper limit of the "true" error inside the Med-Belt.

Now, in order to calculate the absolute error produced by submodel *HorToInc* with respect to our dataset, we made the working assumption that the absolute error of the submodel follows a Gaussian distribution, defined by its mean, μ, and the standard deviation, σ.

When a discrete random variable X takes random values from a finite dataset $x_1, x_2, ..., x_n$, with each value having the same probability, we can estimate μ and σ, through Equations 5.3 and 5.4 respectively.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5.3}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2} \tag{5.4}$$

[1]Root Mean Square Error
[2]Mean Bias Error
[3]Mean Output Value

In order to define the Gaussian distribution of the absolute error produced by submodel *Beta* with respect to [48], μ and σ need to be obtained. It is trivial to estimate μ:

$$\mu = MAE \tag{5.5}$$

Moreover, we can easily show that σ is now estimated with Equation 5.6 (for full details see Appendix B):

$$\sigma = \sqrt{RMSE^2 - MAE^2} \tag{5.6}$$

where

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2} \tag{5.7}$$

Here, A_i represents a data-point coming from the actual (historical) data, and F_i represents the corresponding estimated (forecasted) one, with i ranging from 1 to n within the dataset.

Hence, in order to estimate σ, we only need the respective MAE and *Root Mean Square Error (RMSE)* estimates. Unfortunately a MAE estimation was not provided in [48], therefore, we used the provided $RMSE$ as MAE. That was done by taking into account that $RMSE$ is MAE's upper limit [74]. Hence, the distribution's parameters are: $\mu = RMSE$ and $\sigma = 0$. By so doing, our evaluation results for the power output estimates of not horizontally orientated PV modules can only be considered as an upper limit of the "true" error.

Now, note that the evaluation results provided in [48] come with respect to the dataset they used. In order to best describe the absolute error produced by submodel *HorToInc* with respect to *our* dataset, a normalization procedure have been followed. In particular, we divided the μ value of the Gaussian distribution by the average

output of the submodel *HorToInc* as evaluated in [48]. Subsequently this value is multiplied with the average output of the submodel *HorToInc* with respect to our dataset. Finally, the derived μ has been used as the absolute error produced by submodel *HorToInc* for each data point in our dataset.

5.2.2 PV Module Thermal Submodel

The submodel utilized for the estimation of *the PV module's operating temperature*; T_c, is provided with certain performance guarantees (i.e. MAE) over the wind speed range $[1 - 15m/s]$ [63]. As an RMSE estimate was not provided, we were not able to estimate σ. Moreover, due to lack of information regarding the average T_c output of *PV Module Thermal* submodel with respect to the evaluation in [63], we used the provided MAE as the absolute error produced by the submodel without any normalization with respect to our dataset. This is appropriate, since the values in our dataset range mostly inside the bounds of the dataset used for the aforementioned evaluation.

Now, for each data point in our dataset, the absolute error at the output of *Solar Irradiance* submodel (calculated as described in Section 5.2.1) was propagated through the *PV Module Thermal* submodel. This value was added to our estimation of the absolute error produced by the *PV Module Thermal* submodel. As such the absolute error at the output of the *PV Module Thermal* submodel was calculated for each data point in our dataset.

It has to be noted that the *PV Module Thermal* submodel varies slightly for different *PV module mounting types*, as it accumulates a relative coefficient (as noted

in Section 3.1.3). As such the propagated solar radiation's absolute error will be slightly different when propagated through the different versions of *PV Module Thermal* submodel. To combat this, we calculated *separately* the total error of the entire methodology for each *PV module's mounting type*.

Finally, the *PV Module Thermal* submodel receives also as input the prevailing wind speed. Since, archival meteorological data of prevailing wing speed have not been already included in our dataset, we supplemented our dataset with respective data as we further detail in Section 5.4.

5.2.3 PV Module Maximum Power Output Submodel

The model utilized for predicting the PV modules's maximum power output, and in particular the improved model, is provided with certain performance guarantees (i.e. MBE, MAE, $RMSE$ and Avg) over different PV module's manufacturing technology [40]. As both the MAE and $RMSE$ values were provided the respective Gaussian distribution's parameters have been defined through equations 5.5 and 5.6. For each data point contained in our dataset, we sampled random values following this Gaussian distribution in order to derive with the absolute error produced by the *PV Module Maximum Power Output* submodel. In particular, in order to best describe the absolute error produced by the submodel with respect to our dataset, these values were divided by the average output of the submodel as evaluated in [40] (i.e. Avg) and multiplied with the average output of the submodel with respect to our dataset (Avg_0). However, the values derived take negative values occasionally. In order to avoid this, the returned values were bounded inside the range $[0, 2(MAEAvg_0/Avg)]$

by replacing every outranged value with the respective bound (e.g. if the value has been -1 it was replaced by 0, or if the value has been $2(MAEAvg_0/Avg) + 1$ it was replaced by $2(MAEAvg_0/Avg)$). Note, that, by so doing, the mean value of the distribution is not influenced, as the upper and lower bound has the same distance from the mean.

Now, for each data point in our dataset, the absolute error at the output of *Solar Irradiance* submodel (calculated as described in Section 5.2.1), and the absolute error at the output of *PV Module Thermal* submodel (calculated as described in Section 5.2.2), were propagated through the *PV Module Maximum Power Output* submodel. These values were added to our estimations of the absolute error produced by the *PV Module Maximum Power Output* submodel. As such, the absolute error at the output of the *PV Module Maximum Power Output* submodel was estimated for each data point in our dataset.

Finally it has to be also noted that the *PV Module Maximum Power Output* submodel is as already mentioned evaluated over different PV module manufacturing technology. As such, the absolute error produced by the submodel is different respecting different PV module's manufacturing technology. Therefore the total error of the entire methodology has been estimated *separately* for each PV module's manufacturing technology.

5.3 Deriving the Overall Evaluation Results

The MAE and rMAE of our method were estimated for each combination of PV module manufacturing technology, mounting type and tilt angle configurations. MAE

and rMAE were estimated via Equations 5.8 and 5.9 respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} AE_i \tag{5.8}$$

$$rMAE = \frac{MAE}{Avg_0}100 \tag{5.9}$$

Here, AE_i represents the *Absolute Error* at the output of our entire method (i.e. at the output of the *PV Module Maximum Power Output* submodel) calculated as described in Section 5.2 for each data point contained in our dataset, with i ranging from 1 to n within our dataset. Avg_0 stands for the average output of the *PV Module Maximum Power Output* submodel with respect to our dataset. Note that, formally, rMAE is the MAE normalized to the average of the actual values (and not the estimated ones). However as we did not have access to the respective data we estimated rMAE by normalizing MAE to the average output of our model. Intuitively, this rMAE corresponds to the MAE as a fraction of the prediction, which gives us an estimate of the prediction error.

5.4 Deriving the Appropriate Dataset

For the aforementioned purposes, our dataset has been supplemented with archival meteorological data of prevailing wing speed, V, drawn from the *Weather Underground* online database. In order to match the restrictions in the evaluation of *PV Module Thermal* submodel in [63], all missing V measurements and all V measurements which were outside the $[1 - 15m/s]$ bounds, have been populated/replaced with values sampled from a uniform distribution inside the aforementioned bounds. Finally, all data points where $G_T^{hor}(N) = 0$ has been excluded in our evaluation.

5.5 Results and Discussion

The derived overall method's power output prediction errors for *horizontal* orientation appear in Table 5.1. Due to a lack of required data with respect to irradiance measurements at non-zero slope angles within the Med-Belt and the restrictions described extensively in Section 5.2.1.2, we were only able to estimate what can be considered as *a worst-case* approximate bound for a typical south-facing, 45°, slope angle orientation. In all configurations examined, the respective rMAE was consistently in the order of 40%.

Table 5.1: Overall Output Prediction Error on Horizontal orientation

	Multi-crystal Si		Single-crystal Si	
	Nominal P_m : 35.16W		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

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 in Chapter 2. To the best of our knowledge, the only generic prediction methodology that has been applied in a wide area is that of [8, 36]—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 [35]: 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 in Chapter 2, 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.[4]

[4]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 [35].

Chapter 6

RENES: A Tool for Estimating the Power Output of Distributed Renewable Energy Resources

We incorporated our models for estimating the power output of *Photovoltaic (PV) Systems (PVSs)* and *Wind Turbine Generator (WTGs)* in a web-based, graphical, user-interactive, *renewable energy estimation* tool, *RENES (Renewable Energy Estimator)* which can be found at: *http://www.intelligence.tuc.gr/renes*. A screenshot of RENES is shown in Figure 6.1.

The tool currently provides accurate estimates (within the error guarantees for PV output estimates reported in Chapter 5) 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. 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.

RENES is not a commercial product, and its usage is available to all, free of charge. We envisage RENES to be employed by Artificial Intelligence and Multiagent Systems researchers, for activities such as accumulating data, designing experiments or evaluating new algorithms. To this end, we also provide a web-based *Application Program Interface (API)*, thus enabling the user to fully exploit the tool's requests.

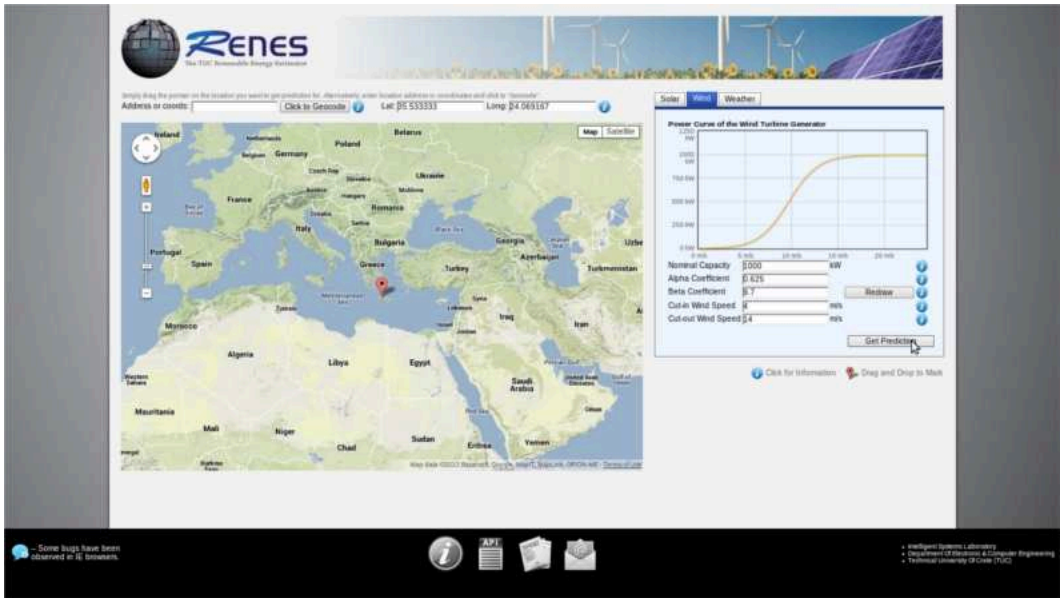

Figure 6.1: Screenshot of RENES

RENES Documentation

RENES is hosted online at: *http://www.intelligence.tuc.gr/renes*, and explicit instructions can be found at the appropriate information links. Regarding the re-

spective API an up to date documentation can be found online at all time at: *http://www.intelligence.tuc.gr/renes/fixed/fixed/api* (a brief documentation is also included in Appendix B).

Chapter 7

Conclusions and Future Work

In the context of power system operation, one of the greatest challenges is running a reliable supply-on-demand system. The growing penetration of inherently intermittent and potentially distributed *Weather-Driven and non-scheduled Energy Resources (WDERs)*, such as *Photovoltaic (PV) Systems (PVSs)* and *Wing Turbine Generators (WTGs)*, into the electricity grid is expected to impact the system's reliability. Among with the growing demand due to the electrification of transportation and heating, a radical reengineering of the infrastructures and functionality of the Grid is nowadays imminent. The notion of the Smart Grid captures the essence/concept of the next generation Grid.

Artificial Intelligence (AI) and MultiAgent Systems (MASs) research is increasingly preoccupying itself with building intelligent systems for the Smart Grid–and the efficient incorporation of WDERs into the Smart Grid has emerged as a major challenge. The creation of Virtual Power Plants (VPPs) (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 more efficiently incorporate WDERs into the Grid.

Now, dependable WDERs power output predicting technology could reinforce the reliable integration of WDERs into the Grid as it could: *(i)* enable the more flexible systems to prepare and behave accordingly, *(ii)* reinforce appropriate decision making by the VPP manager. This technology should be able to provide accurate WDERs' power output predictions within a wide geographical region. Moreover, it could be of great value, if this technology was *generic* but also of *low-cost* making it widely available.

In this thesis, we presented an appropriately *generic* and *low cost* PVS power output estimation method. The method is based on free weather readings from online websites, and it is evaluated with real data over the Mediterranean region. This method is incorporated in a web-based tool – *Renewable Energy Estimator (RENES)* [1] – that enables the user of predicting the output of distributed WDERs (PVSs and WTGs).

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, it is important that RENES is enhanced with more capabilities, and that a proper user evaluation study of the tool is conducted.

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

[1] RENES is hosted online at: *http://www.intelligence.tuc.gr/renes*.

related to VPP operation, or using machine learning techniques for optimal sun-tracking.

As a matter of fact, RENES is already being used in the development of a functional system prototype for a *Cooperative Virtual Power Plant (CVPP)* [10]; in particular, the system is to be using WDERs power output estimates (coming from RENES) as inputs to decision making algorithms employed during the scheduling of VPP power production.

Moreover, our method is already being used as a PVS model in simulation experiments for *Optimal Control* of *Dual-Axis* and *Single-Axis Solar Tracking PVSs*. The goal of this line of research is to estimate the optimal sun tracking policy,[2] which is to be found/approximated by the use of *reinforcement learning* techniques [67].

We believe that the fact that our method and tool are already being utilized in research in progress, testifies to the value of the work presented in this thesis.

[2]Taking into consideration both the energy consumption (due to the tracking) and the energy generation of the PVS.

Bibliography

[1] Enercon. Technical report, Enercon wind energy converters - product overview, 2010.

[2] P. Asmus. Microgrids, virtual power plants and our distributed energy future. *The Electricity Journal*, 23(10):72 – 82, 2010.

[3] P. Bacher, H. Madsen, and H. A. Nielsen. Online short-term solar power forecasting. *Solar Energy*, 83(10):1772 – 1783, 2009.

[4] S. Barbaro, S. Coppolino, C. Leone, and E. Sinagra. An atmospheric model for computing direct and diffuse solar radiation. *Solar Energy*, 22(3):225 – 228, 1979.

[5] A. Barenbrug, Transvaal, and O. F. S. C. of Mines. *Psychrometry and psychrometric charts*. Cape & Transvaal Printers Ltd., 1974.

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

[7] R. Bird, R. Hulstrom, S. E. R. Institute, and U. S. D. of Energy. *A simplified clear sky model for direct and diffuse insolation on horizontal surfaces*. SERI/TR. Solar Energy Research Institute, 1981.

[8] S. Bofinger and G. Heilscher. Solar electricity forecast approaches and first results. In *21st European Photovoltaic Solar Energy Conference*, number 9, pages 4–8, 2006.

[9] T. Burton, N. Jenkins, D. Sharpe, and E. Bossanyi. *Wind Energy Handbook*. Wiley, 2011.

[10] 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 Int. Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, 2011.

[11] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom. Intra–hour forecasting with a total sky imager at the uc san diego solar energy testbed. *Solar Energy*, 85(11):2881–2893, 2011.

[12] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725 – 1744, 2008.

[13] 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.

[14] J. F. K. Frank Kreith. *Principles of Sustainable Energy (Mechanical and Aerospace Engineering Series)*. CRC Press, first edition, 2010.

[15] D. Fredlund, H. Rahardjo, and D. Fredlund. *Unsaturated Soil Mechanics in Engineering Practice*. Wiley, 2012.

[16] D. M. Gates. *Biophysical Ecology (Dover Books on Biology)*. Springer-Verlang, first edition, 1980.

[17] S. A. Gregory and M. Zeilik. *Introductory Astronomy and Astrophysics (Saunders Golden Sunburst Series)*. Saunders Golden Sunburst Series, fourth edition, 1997.

[18] P. Grindley, W. Batty, and S. Probert. Mathematical model for predicting the magnitudes of total, diffuse, and direct-beam insolation. *Applied Energy*, 52(2-3):89 – 110, 1995. Fifth Arab International Solar Energy Conference.

[19] A. Hammer, D. Heinemann, E. Lorenz, and B. LG'Ockehe. Short–term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, 67(1–3):139–150, 1999.

[20] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1998.

[21] A. Hemami. *Wind Turbine Technology*. Delmar, Cengage Learning Series in Renewable Energies. Delmar, Cengage Learning, 2011.

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

[23] F. Ideriah. A model for calculating direct and diffuse solar radiation. *Solar Energy*, 26(5):447 – 452, 1981.

[24] M. Iqbal. *An introduction to solar radiation*. Academic Press, 1983.

[25] J. Kabouris and F. D. Kanellos. Impacts of large-scale wind penetration on designing and operation of electric power systems. *Sustainable Energy, IEEE Transactions on*, 1(2):107–114, 2010.

[26] S. Kaplan and T. Net. *Smart grid: modernizing electric power transmission and distribution; energy independence, storage and security; energy independence and security act of 2007 (EISA); improving electrical grid efficiency, communication, reliability, and resiliency; integrating new and renewable energy sources*. Government series. TheCapitol.Net, 2009.

[27] G. Kariniotakis, J. Mousafir, J. Usaola, I. Marti, I. Cruz, H. Madsen, C. Lac, G. Giebel, J. Halliday, U. Focken, and J. Toefting. ANEMOS: Development of a next generation wind power forecasting system for the large-scale integration of onshore & offshore wind farms. In *Proceedings of the EGS-AGU-EUG Joint Assembly*, 2003.

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

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

[30] G. Kopp and J. L. Lean. A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, 38(L01706), 2011.

[31] M. Liserre, T. Sauter, and J. Hung. Future energy systems: Integrating renewable energy sources into the smart power grid through industrial electronics. *Industrial Electronics Magazine, IEEE*, 4(1):18–37, 2010.

[32] B. Liu and R. Jordan. Daily insolation on surfaces tilted towards equator. *Solar Energy*, 10:53, 1961.

[33] B. Y. Liu and R. C. Jordan. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy*, 4(3):1 – 19, 1960.

[34] C. Liu and T. Feng. Atmospheric turbidity over taiwan. *Atmospheric Environment. Part A. General Topics*, 24(5):1303 – 1312, 1990.

[35] E. Lorenz, J. Remund, S. Miller, W. Traunmiller, Steinmaurer, D. G., J. Ruiz-Arias, V. Fanego, L. Ramirez, M. Romeo, C. Kurz, L. Pomares, and C. Guerrero. Benchmarking of different approaches to forecast solar irradiance. In *24th European Photovoltaic Solar Energy Conference*, 2009.

[36] 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.

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

[38] B. Luzum, N. Capitaine, A. Fienga, W. Folkner, T. Fukushima, J. Hilton, C. Hohenkerk, G. Krasinsky, G. Petit, E. Pitjeva, M. Soffel, and P. Wallace. The iau 2009 system of astronomical constants: the report of the iau working group on

numerical standards for fundamental astronomy. *Celestial Mechanics and Dynamical Astronomy*, 110:293–304, 2011. 10.1007/s10569-011-9352-4.

[39] B. Machenhauer. Hirlam final report. Technical Report 5, Copenhagen, Denmark, 1988.

[40] B. Marion. *Comparison of predictive models for photovoltaic module performance*, pages 1–6. Number 1. Ieee, 2008.

[41] S. Marsland. *Machine Learning: An Algorithmic Perspective*. Chapman & Hall/CRC, 1st edition, 2009.

[42] N. Martin and J. Ruiz. Calculation of the pv modules angular losses under field conditions by means of an analytical model. *Solar Energy Materials and Solar Cells*, 70(1):25 – 38, 2001.

[43] S. Maxwell and H. Delaney. *Designing experiments and analyzing data: a model comparison perspective*. Number v. 1 in Inquiry and Pedagogy Across Diverse Contexts Series. Lawrence Erlbaum Associates, 2004.

[44] A. Mellit and A. M. Pavan. Performance prediction of $20kw_p$ grid-connected photovoltaic plant at trieste (italy) using artificial neural network. *Energy Conversion and Management*, 51(12):2431 – 2441, 2010.

[45] D. Menicucci and J. Fernandez. *User's manual for PVFORM: A photovoltaic system simulation program for stand-alone and grid-interactive applications*. Sandia National Laboratories, 1989.

[46] K. Moslehi and R. Kumar. Smart grid - a reliability perspective. In *Innovative Smart Grid Technologies (ISGT), 2010*, pages 1 –8, jan. 2010.

[47] 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.

[48] A. M. Noorian, I. Moradi, and G. A. Kamali. Evaluation of 12 models to estimate hourly diffuse irradiation on inclined surfaces. *Renewable Energy*, 33(6):1406 – 1412, 2008.

[49] C. E. C. B. . A. E. Office, R. E. P. (Calif.), and C. E. C. R. Committee. *Guidelines for California's Solar Electric Incentive Programs Pursuant to Senate Bill 1: Guidelines*. California Energy Commission, 2007.

[50] A. A. Panagopoulos, G. Chalkiadakis, and E. Koutroulis. Predicting the power output of distributed renewable energy resources within a broad geographical region. In *20th European Conference on Artificial Intelligence*, pages 981–986, 2012.

[51] D. Patelis. Socio-philosophical heritage and the logic of history. *Fragmentos de Cultura*, 21:389–414, 2011.

[52] R. Perez, K. Moore, S. Wilcox, D. Renné, and A. Zelenka. Forecasting solar radiation preliminary evaluation of an approach based upon the national forecast database. *Solar Energy*, 81(6):809–812, 2007.

[53] J. T. Peterson, E. C. Flowers, G. J. Berri, C. L. Reynolds, and J. H. Rudisill. Atmospheric turbidity over central north carolina. *Journal of Applied Meteorology*, 20(3):229 – 241, 1981.

[54] D. Pisimanis, V. Notaridou, and D. Lalas. Estimating direct, diffuse and global solar radiation on an arbitrarily inclined plane in greece. *Solar Energy*, 39(3):159 – 172, 1987.

[55] C. Potter, A. Archambault, and K. Westrick. Building a smarter smart grid through better renewable energy information. In *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, pages 1 –5, march 2009.

[56] 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.

[57] 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.

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

[59] I. Reda and A. Andreas. Corrigendum to "solar position algorithm for solar radiation applications" [solar energy 76 (2004) 577-589]. *Solar Energy*, 81(6):838 –, 2007.

[60] D. F. A. Riza, S. I. ul Haq Gilani, and M. S. Aris. Hourly solar radiation estimation using ambient temperature and relative humidity data. *International Journal of Environmental Science and Development*, 2(3):188–193, 2011.

[61] G. Shafiullah, A. M. T. Oo, D. Jarvis, A. S. Ali, and P. Wolfs. Potential challenges: Integrating renewable energy with the smart grid. In *Universities Power Engineering Conference (AUPEC), 2010 20th Australasian*, pages 1–6, December 2010.

[62] R. E. Sims, H.-H. Rogner, and K. Gregory. Carbon emission and mitigation cost comparisons between fossil fuel, nuclear and renewable energy resources for electricity generation. *Energy Policy*, 31(13):1315–1326, 2003.

[63] E. Skoplaki, A. Boudouvis, and J. Palyvos. A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting. *Solar Energy Materials and Solar Cells*, 92(11):1393 – 1402, 2008.

[64] E. Skoplaki and J. Palyvos. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Solar Energy*, 83(5):614 – 624, 2009.

[65] E. Skoplaki and J. Palyvos. Operating temperature of photovoltaic modules: A survey of pertinent correlations. *Renewable Energy*, 34(1):23–29, 2009.

[66] U. States. *Surface weather observations and reports*. Office of the Federal Coordinator for Meteorological Services and Supporting Research, Silver Spring, Md., 1998.

[67] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[68] C. Tao, D. Shanxu, and C. Changsong. Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement. In *Power Electronics for Distributed Generation Systems (PEDG), 2010 2nd IEEE International Symposium on*, pages 773 –777, june 2010.

[69] D. Thevenard and K. Haddad. Ground reflectivity in the context of building energy simulation. *Energy and Buildings*, 38(8):972 – 980, 2006.

[70] O. Uluyol, G. Parthasarathy, W. Foslien, and K. Kim. Power curve analytic for wind turbine performance monitoring and prognostics. In *Annual Conference of the Prognostics and Health Management Society, Montreal, Canada*, 2011.

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

[72] A. Vazioulin, Viktor. *The Logic of History (in Greek)*. Ellinika grammata, 2004.

[73] X. Wang, P. Guo, and X. Huang. A review of wind power forecasting models. *Energy Procedia*, 12(0):770 – 778, 2011. The Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011).

[74] C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, 30:79–82, 2005.

[75] A. Woyte, V. V. Thong, R. Belmans, and J. Nijs. Voltage fluctuations on distribution level introduced by photovoltaic systems. *Sustainable Energy, IEEE Transactions on*, 21(1):202–209, 2006.

Appendix A

Documentation of RENES Application Program Interface (API)

RENES *Application Program Interface (API)* is based on HTTP requests. The response messages are in Extensible Markup Language (XML).

A.1 Solar API

The HTTP requests for power output predictions of *Photovoltaic (PV) Systems (PVSs)* should be in the following format:

http://147.27.14.3:11884/solarAPI/-latitude-/-longitude-/-altitude-/-slope-/-azimuth-/-reflectance-/-pressure-/-temperature-/-powerpeak-/-temperatureCoefficient-/-dirtiness-/-extra-/-powerpeaklow-/-reduction-/-

mountingtype-/-efficiencyInv-

The fields noted with '-' characters should be replaced according to your needs with respect to each field name as stated bellow. It is noted that "solarAPI" is not a replaceable field and should be maintained as is.

A.1.1 Replaceable Fields

- "latitude" and "longitude"

  The "latitude" and "longitude" fields stand for the respective *Geographical Co-ordinates* (WGS84 datum). The coordinates should be expressed in a decimal form (e.g. 35.533333, 24.069167). It is noted that when the coordinates are expressed in a decimal form, northern latitudes should be positive, southern latitudes should be negative, eastern longitudes should be positive and western longitudes should be negative.

- "altitude"

  The "altitude" field stands for the respective *Above Mean Sea Level (AMSL)* which refers to the elevation (on the ground) or altitude (in the air) of any object, relative to the average sea level datum (zero AMSL). The provided AMSL should be in meters (m).

- "slope"

  The "slope" field stands for the respective *Slope Angle* of the PV module. The provided slope angle values should be in degrees (°), in the range $[0° - 90°]$, and

measured from the horizontal plane (e.g. horizontal plane: 0°, vertical plane: 90°).

- "azimuth"

  The "azimuth" field stands for the respective *Azimuth Angle* of the PV module. The provided azimuth angle values should be in degrees (°), in the range [0° − 360°), and measured clockwise starting from South (e.g. South: 0°, West: 90°, North: 180°, East: 270°)

- "reflectance"

  The "reflectance" field stands for the respective *Average Ground Reflectance* (also known as *Albedo*). Average ground reflectance is defined as the average fraction of incident radiation reflected by the ground. The respective values should be provided in percentage form without considering the percentage symbol (e.g. 0 stands for total absorption and 100 for total reflection).

  The following table [69] shows estimates of average reflectance values respecting different materials. In case of great uncertainty, the common value of 20 can be used.

Ground cover	Reflectivity
Water (large incidence angles)	7%
Coniferous forest (winter)	7%
Bituminous and gravel roof	13%
Dry bare ground	20%
Weathered concrete	22%
Green grass	26%
Dry grassland	$20 - 30\%$
Desert sand	40%
Light building surfaces	60%

- "pressure"

  The "pressure" field stands for the *annual average of the atmospheric pressure* of the respective location. This value is used to better approximate the solar angles via the Nrel's SPA algorithms and should be in $millibar(mb)$ units. In case of great uncertainty, the value '1012' can be used without great accuracy losses.

- "temperature"

  The "temperature" field stands for the *annual average of the ambient temperature* of the respective location. This value is used to better approximate the solar angles via Nrel's SPA algorithms and should be in Celsius degrees ($^{\circ}C$). In case of great uncertainty, the value '19.5' can be used without great accuracy losses.

- "powerpeak"

  The "powerpeak" field stands for the *Maximum Power (P_{max})*, or *Peak Power*, of the PV module which is a measure of the nominal power of the module under specific testing conditions. This value is found in almost every PV module data sheet. The provided value should be in $Watt(W)$ units.

- "temperatureCoefficient"

  The "temperatureCoefficient" field stands for the *Temperature Coefficient* of P_{max} $(\%/^{\circ}C)$ which is used to determine the effect of temperature on the PV module's power output. This value is found in almost every PV module data sheet.

- "dirtiness"

  The "dirtiness" field stands for the degree of *Dirtiness* of the PV module which is a qualitative attribute considering the sediments of soil and dirt deposited on the respective PV module's surface. The appropriate descriptive values are *'Clear'*,*'Low'*,*'Medium'* and *'High'*.

- "extra"

  The "extra" field is auxiliary and indicates whether the appropriate characteristic attributes in order to account the reduction in efficiency under low irradiance level is utilized. Specifically, the possible descriptive values are: *'no'*,*'pmlow'* and *'redlow'*. A possible *'no'* valuing indicates that non of the characteristic attributes "P_{max} at low Iradiance" or "Efficiency Reduction under Low Irradiance" is utilized. Moreover, *'pmlow'* and *'redlow'* values indicate the utilization

of "P_{max} at low Iradiance" or "Efficiency Reduction under Low Irradiance" respectively.

- "powerpeaklow"

  The "powerpeaklow" field stands for the P_{max} at low Irradiance which is the measured maximum power at effective low irradiance ($\sim 200W/m^2$ and $25°C$). This value is found in some PV modules data sheets. If the value for the field "extra" above is *'no'*, then this field's value is irrelevant.

- "reduction"

  The "reduction" field stands for the *Efficiency Reduction under Low Irradiance* which is the reduction of efficiency due to the drop of the incident irrandiance from $1000W/m^2$ to $200W/m^2$ (cell temperature $\sim 25°C$). This value is found in some PV modules data sheets. If the value for the field "extra" above is *'no'*, then this field's value is irrelevant.

- "mountingtype"

  The "mountingtype" field stands for the *Mounting type* of the respective PV module. The possible descriptive values are *'FreeStanding'*, *'FlatRoof'*, *'SlopedRoof'* and *'FacAdeintegrated'*.

- "efficiencyInv"

  The "efficiencyInv" field stands for the *Efficiency Factor* which is required in order to account for inverter, wiring or any other secondary losses. If an inverter is used the efficiency factor should correspond to the inverter efficiency factor -

adjusted to best fit the system (e.g. 100 efficiency factor stands for negligible secondary losses).

A.1.2 Example

A possible HTTP request: *http://147.27.14.3:11884/solarAPI/35.533333/24.069167/137/0/0/20/1012.3661/19.497649/70/-0.5/Clear/no/60/10/FreeStanding/90*

A.1.3 XML Response

We note that in the respective XML response document each tuple considers an estimation for every next hour. The estimated PV power output is in *Watt (W)* units; all solar irradiance estimates are in W/m^2; temperature is in $°C$; humidity and cloud coverage are in percentage form.

A.2 Wind API

The HTTP requests for power output predictions of *Wind Turbine Generators (WTGs)* should be in the following format:

http://147.27.14.3:11884/windAPI/-latitude-/-longitude-/-a-/-b-/-nomCapacity-/-cin-/-cout-

The fields noted with '-' characters should be replaced according to your needs with respect to each field name as stated bellow. It is noted that "windAPI" is not a replaceable field and should be maintained as is.

A.2.1 Replaceable Fields

- "latitude" and "longitude"

  The "latitude" and "longitude" fields stand for the respective *Geographical Coordinates* (WGS84 datum). The coordinates should be expressed in a decimal form (e.g. 35.533333, 24.069167). It is noted that when the coordinates are expressed in a decimal form, northern latitudes should be positive, southern latitudes should be negative, eastern longitudes should be positive and western longitudes should be negative.

- "a"

  The "a" field stands for the *Alpha Coefficient* which is a parametrization coefficient of the Wind Turbine Generator's (sigmoid) Power Curve. The respective Power Curve is:

  $$prod_{i,t}(V_t) = NomCapacity_i/(1 + e^{\alpha_p*(\beta_p - V_t)})$$

  Here, NomCapacity is the turbine's nominal capacity, V_t represents the predicted wind speed at time t (and e is Euler's number) and, finally, the Alpha (α_p) and Beta (β_p) parameters are user-specified or given default values depending on turbine type.

- "b"

  The "b" field stands for the *Beta Coefficient* which is a parametrization coefficient of the Wind Turbine Generator's (sigmoid) Power Curve.

- "nomCapacity"

The "nomCapacity" field stands for the *Nominal Capacity* of the Wind Turbine Generator which is the turbine's maximum power output for "optimal" wind speed. The respective value should be in $kiloWatt(kW)$ units.

- "cin"

  The "cin" field stands for the *Cut-in Wind Speed* which is the lowest wind speed value that exerts sufficient torque on the turbine blades to make them rotate. This value is found in almost every WTG data sheet.

- "cout"

  The "cout" field stands for the *Cut-out Wind Speed* which is the lowest wind speed value at which the WTG's braking system is employed to bring the rotor to a standstill in order to avoid potential damage to the rotor. This value is found in almost every WTG data sheet.

A.2.2 Example

A possible HTTP request: *http://147.27.14.3:11884/windAPI/35.533333/24.069167/0.625/9.7/1000/4/14*

A.2.3 XML Response

We note that in the respective XML response document each tuple considers an estimation for every next hour. The estimated WTG power output is in *kiloWatt (kW)* units; and wind speed is in m/s unit.

Appendix B

Gaussian Distribution of Absolute Error

A Gaussian distribution is defined by its mean, μ, and the standard deviation, σ.

When a discrete random variable X takes random values from a finite dataset $x_1, x_2, ..., x_n$, with each value having the same probability, we can estimate μ and σ, through Equations B.1 and B.2 respectively.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{B.1}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2} \tag{B.2}$$

In order to define the Gaussian distribution of Absolute Error, μ and σ need to be obtained. In the following paragraphs, we let A_i represents an actual value, and F_i represents the corresponding estimated one, with i ranging from 1 to n within our finite dataset. By replacing x_i with $|F_i - A_i|$ in Equation B.1, we can see that μ can

now be estimated through Equation B.3.

$$\mu = MAE \tag{B.3}$$

where,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - A_i| \tag{B.4}$$

Moreover, by replacing x_i with $|F_i - A_i|$ and μ with MAE in Equation B.2 we get:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(|F_i - A_i| - MAE\right)^2} \Longrightarrow$$

$$\Longrightarrow \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((F_i - A_i)^2 + MAE^2 - 2|F_i - A_i|MAE\right)} \Longrightarrow$$

$$\Longrightarrow \sigma = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}\left((F_i - A_i)^2\right) + \sum_{i=1}^{n}\left(MAE^2\right) - \sum_{i=1}^{n}\left(2|F_i - A_i|MAE\right)\right)} \Longrightarrow$$

$$\Longrightarrow \sigma = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left((F_i - A_i)^2\right) + \frac{1}{n}\sum_{i=1}^{n}\left(MAE^2\right) - \frac{2}{n}\sum_{i=1}^{n}\left(|F_i - A_i|MAE\right)\right)} \Longrightarrow$$

$$\Longrightarrow \sigma = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left((F_i - A_i)^2\right) + \frac{n}{n}MAE^2 - 2MAE\frac{1}{n}\sum_{i=1}^{n}\left(|F_i - A_i|\right)\right)} \Longrightarrow$$

$$\Longrightarrow \sigma = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left((F_i - A_i)^2\right) + MAE^2 - 2MAE\frac{1}{n}\sum_{i=1}^{n}\left(|F_i - A_i|\right)\right)} \tag{B.5}$$

Now, substituting the RMSE and MAE quantities in B.5 gives us Equation B.6, which allows us to estimate σ.

$$\sigma = \sqrt{RMSE^2 - MAE^2} \tag{B.6}$$

where,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2} \tag{B.7}$$