Karlsruher Institut für Technologie

Master's Thesis

Analytical Modeling Concept for Weather Phenomena as Renewable Energy Resource

Bruno Pousinho

November 08, 2017

Supervisors: PD Dr. Gudrun Thaeter
Prof. Dr. Thomas Hamacher
Franz Christange, M.Sc.

Institute for Applied and Numerical Mathematics,
Karlsruhe Institute of Technology

Chair of Renewable and Sustainable Energy Systems,
Technical University of Munich

Abstract

This thesis introduces developments for a general concept to model weather phenomena as a renewable energy resource. The concept presented is based on an existing weather generator model, which uses a high amount of parameters and a set of Fourier series to model weather phenomena. The methodology is altered in order to reduce the number of parameters required to describe the same weather phenomena. The weather phenomena considered are global horizontal irradiation (GHI), air temperature and wind speed. The model uses a simple analytic implementation to describe typical properties of weather phenomena, such as time-variance, rates-of-change and causality between them. The implementation is not site specific, which allows the model to be applied worldwide.

In order to reliably reproduce the weather properties of a site, the data set utilized must have a minimum length of 8 years. Two approaches are presented in this work, a parametric dynamic model for the weather and an approach to parametrize specific features of the weather. This model uses stochastic processes from which the features of the weather are added and emulate realistic behavior. The analysis concept is an approach to determine the necessary parameters to describe weather and it does so by solving a series of numerical optimization methods, like the Levenberg-Marquardt algorithm. The implementation is entirely done in MATLAB and a total of 12 sites throughout the world are used to validate the results of the model. Since stochastic processes are applied, the accuracy is evaluated for all modeled behaviors of weather. The determination of errors is done for each individual parameter. A general shape error is also determined. It describes how close the probability density function of each synthetic weather phenomenon matches the one of their correspondent measured weather data. Overall, the relative shape error for GHI, air temperature and wind speed is below 3%, 5% and 10% respectively.

Acknowledgements

This thesis is the result of the research done at the chair of Renewable and Sustainable Energy Systems of Technical University of Munich in 2017. In the first place, I would like to thank the people who collaborated and enabled this master's thesis to happen. I would like to thank Franz Christange for the opportunity of working with him. I appreciate the rigorous scientific push that he always aimed for. I can say, without a doubt, it also pushed me to become a better engineer. I would like to thank prof. Gudrun Thäther for the opened arms with which she received this project. Thanks to her, my mind was at ease knowing that I have a solid connection at KIT, someone who I could entirely trust to have by best well being at heart.

I would also like to thank my fellow colleagues at the chair of Renewable and Sustainable Energy Systems for the companionship that was developed. In particular, I would like to thank Soner Candaş and Anurag Mohapatra. It was with them that I often kept working long hours to the best of my ability. It was with them that so many interesting topics were discussed which often led us down the rabbit hole. I would like to thank Kais Siala for the help with the weather data that I absolutely needed to accomplish my goals.

Finally, I would like to thank Eleni Alevizou for the full support she gave me in my academic commitment. I can surely say that without her, I would have never been able to reach my full potential. While this thesis pushed my academic limits further, she provided me with all the love and happiness I could have wished for. In the end, if I knew what would happen in this journey, I would definitely do it again.

List of Symbols

Symbol	Unit	Description
n, n'	W/m^2	$n \in (I, T, V)$, synthetic and Measured global horizontal irradiation, air temperature, wind speed
t		time
T_s		time sampling
μ		mean of a normal distribution
σ		standard deviation of a normal distribution
k		weibull shape parameter
λ	m/s	weibull scale parameter
ω_d	rad	base frequency of one day
ω_y	rad	base frequency of one year
Ψ	W/m^2	maximum possible Irradiation ratio during day time
η		ratio between theoretical maximum and measured irradiation
ψ		latitude in radians
δ		solar declination
h		hour angle in radians
E_0	W/m^2	solar constant
ρ		atmospheric proportion
E_g	W/m^2	maximum possible global horizontal irradiation value
W_{max}		constant parameter in cost function original
χ_n		time dependent normally distributed noise signal
χ_n''		correlated noise signal
χ_n'		time-invariant standard normally distributed $N(0, 1)$ white noise
A_0		average value of a time series
A_d		amplitude of one day period
ϕ_d	rad	phase of a day long period
A_y		amplitude of one year period
ϕ_y	rad	phase of a year long period
A_s		amplitude for the daily shift over seasons
W		lambet W function
w_n		weight function for the n weather phenomenon
B_n		gain parameter
s		laplacian frequency parameter
T_n	h	time-correlation parameter
ζ_n		cross-correlation parameter
Tr_n		transformation from normal to natural distribution
ϵ		infinitesimal positive constant
p_{z_n}		parameter of a modeled behavior for the n weather phenomenon
e_{z_n}	%	error of a modeled behavior for the n weather phenomenon

Contents

List of Figures

List of Tables

1 Introduction

Energy is one of the fundamental needs of every society. Traditionally, energy systems have been based on fossil fuels, like oil, natural gas, coal and others. These sources allowed for reliable energy systems, but are limited and have been directly correlated to global warming. The challenges that arose focus on mitigating green house effects. A widely accepted solution is transitioning from fossil fuels based energy systems to renewable oriented energy systems. The most common renewable energy sources are hydro-power, photovoltaic (PV) and wind. Germany is setting an example for what a transition from traditional fossil fuels to renewable energy generation could look like. Between 1990 and 2016, Germany increased its renewable installed capacities from 4.168 GW to 104.024 GW, which corresponds to an annual power consumption share of 3.4% and 31.7% respectively [1]. The majority of this growth happened while renewable technologies were not cost-effective. Today, PV, wind and hydro are already economically competitive. The result is that many nations worldwide are starting to push for renewable energy systems (RES).

The growth of RES solves the availability of energy sources and pollution problems but introduces issues regarding energy security and reliability. The general drawback of renewable power generation (RPG), mainly PV and wind, is that they do not generate power 'on demand', it is necessary to have the right weather conditions. On top of this possible non-availability of RPG, the typical peaks at which RES generate the most electricity do not necessarily match the peaks in demand [2]. The reliability of RPG is therefore intrinsically linked with the fluctuations of weather phenomena, which are perceived as renewable energy resources (RER). In order to study the effects of these fluctuations, the RES need to be modeled and analyzed in detail [3]. These dynamic models have one horizontal requirement: an accurate description of weather behavior.

This thesis investigates developments for a general concept which aims to provide a synthesized description of weather phenomena behavior for RER modeling and other possible applications. Weather data as a whole have a high number of phenomena, such as irradiation, precipitation and others, that describe the RER conditions in detail. The phenomena of weather data can be separated in two different categories, driving forces and external forces. A driving force is a phenomenon of weather for which there is a technology that allows for the harness of its energy. Examples of such phenomena are solar irradiation and wind speed through PV panels and wind turbines respectively. An external force is a phenomenon which has either a positive or negative influence in RPG but is not the source of power generation. Examples of such phenomena are air temperature and precipitation. The RER which are the focus of this thesis are global horizontal irradiation, air temperature and wind speed. The purpose of the model introduced in this thesis is to provide a fundamental description of weather phenomena, therefore it mainly addresses the most widespread RPG technologies, namely photovoltaic and wind power systems. The inclusion of temperature in the concept despite being an external force is due to its significant correlation to both driving forces addressed, as well as its potential impact in many RPG technologies.

Synthetic weather data generation is just a portion of what this model offers, under-

standing the dynamics of weather phenomena is the most important objective. Nevertheless, weather data play an important part. It offers a way to test if the weather model is displaying the characteristic behaviors of the weather phenomena which it is trying to model. Typically, 8 years of measured data are necessary to capture the specific variances and frequencies of weather behavior. The sample size reflects the uncertainty in the mean and standard deviation of the data, *i.e.*, larger sample sizes in years of data result in synthetic data with higher accuracy. In general, the weather model, internally generates independent random normal distributed time series with any given time step and proceeds to shape it and introduced the modeled information of the weather data. The general concept is first introduced in [3] and uses 61 parameters to describe weather behavior. The number of parameters is the main problem to be tackled. The original model also aims to be a complete continuous description and falls short in modeling of solar irradiation. Furthermore, it aims to be a general description of weather worldwide but is still lacking the results from various locations around the world that corroborates such a statement. An additional approach presented with the analytical weather model is an analysis concept. It is an algorithm that takes in measured weather data and proceeds to capture its specific behavior through a parameter set which is applied in the weather model.

This thesis aims to reduce the complexity of the dynamic weather by reformulating the transformations and the general equations of the modeling approach. By doing so, the parameter set necessary to describe the weather behavior is reduced by a substantial amount. The main obstacle in reducing the number of parameters is the loss of accuracy. To counteract this behavior, the transformations applied need to be specifically designed with the subsequent modeling approach in mind. A set of assumptions are also introduced in the model in order to correlate previously uncorrelated parameters and reduce the need of numerical approximation processes. The new simplified version of the weather model is tested and iterated for 12 different locations around the world. Half of them are used for the development of the weather model and the other half for producing final results, *i.e.*, synthetic weather data and their respective errors.

Overall, this weather model provides a less complex approach to generate synthetic weather data than the initial concept [3] while maintaining reasonable accuracy. It uses a total of 26 numerically determined parameters and it produces an output with an average relative shape error below 3% for solar irradiation, below 5% for air temperature and below 10% for wind speed. The wind speed is a considerably chaotic weather phenomenon. The solar irradiation and air temperature have a somewhat cyclical and consistent behavior and therefore the weather model is able to model their behavior with relative ease.

2 Background

In this chapter, a brief history and overview of weather generators is provided. The main characteristics of such models are explained and a clear distinction is done between them and other weather models, both in their approach and their expected output. This chapter aims to summarize the core ideas, elaborate and detailed [4], [5] and [6], necessary to understand both the starting point of this thesis work and the developments that are introduced.

Modeling daily weather conditions and sequences is common in water engineering design, agricultural planning or even in simulations related to climate change studies. Models are used because the measured weather data available in some cases have poor quality, either in their length, by containing gaps in time or even due to insufficient spatial coverage. These models are referred to as 'Weather generators'. Typically, they either provide unlimited sequences of synthetic weather data or are used to fill in the gaps of missing data [4]. There are two main approaches to model weather data, one is numerical global climate models and the other is weather generators. The differences that distinguish these two approaches are their spatial scale and computational speed. Climate models aim to replicate the behavior of the atmosphere as a whole for an extended period in time on a global scale. Due to the enormous amount of information that these models try to include, they become limited in the number of simulations which they can provide and they do not resemble any specific location but rather a large spatial grid. On the other hand, weather generators keep their focus on much smaller spatial scales, usually just a few locations within an area of a few kilometers. Weather generators are, typically, computationally very fast in providing random simulations whose outputs show the same or very similar distributional properties as the input time series which is specially visible at daily or sub-daily scales. These main differences are the reason why weather generators are used in impact studies as computationally inexpensive tools to generate synthetic daily time series of atmospheric phenomenon at specific locations [5]. These outputs are produced by modeling important features of the measured weather data such as its daily mean, covariances and variances, maximums and minimums, its frequencies and possibly others.

Daily weather generators are the most widespread and utilized type of weather models. One reason for it is that the availability of measured data which serves as input is, on a daily timescale, much higher than continuous long sequences of measured weather data. Furthermore, there is also an abundance of impact models which use daily weather conditions as inputs. Weather generators can be described in two complementary ways. One is that they are stochastic models for diurnal (and by compilation also long periods of time) fluctuations in weather conditions. In this view, the parameters that the weather generator determines represent a summary of the weather features of a specific climate. The second way to describe weather generators is, when used for Monte-Carlo simulation (*i.e.*, weather generation), as complex generators of random numbers whose outputs have the same or very similar statistical properties as the measured weather data at a specific location.

It is imperative to stress that a weather generator is not a weather forecast algorithm.

A weather forecast algorithm is a deterministic model which works by numerically integrating the partial differential equations which describe fluid flows [4], thus each operates in completely different fashion. A crucial implication of the different ways that these two types of model operate is that weather generators (stochastic models) have an output which is not expected to have a synthetic sequence that matches or is in any way duplicated by measured weather data at any given time, past or future, despite of having a similar statistical behavior.

The majority of weather generators have been focused on precipitation. Precipitation can be the most important weather phenomenon for many applications and it can furthermore affect the statistical behavior of other weather phenomenon which need to be simulated. The other commonly modeled weather phenomenon is air temperature. In combination with precipitation, it has long been used in, for example, crop estimation models. In [7], it is presented that using extensive synthetic weather data time series the estimation of the frequency distributions of the synthetic yields for wheats is smoother than if otherwise done with short measured weather data. In addition, these synthetic weather data have also been used to test crop behavior in case of severe weather conditions.

The field of climate change impact assessment also has a long history of using weather generators. They have been used to create artificial scenarios with possibles changes in weather statistical behavior. These artificial scenarios have afterwards been used as an input in response and energy systems models. The attractiveness of weather generators in this field comes from their ability of providing extensive synthetic samples of expected future weather data which, by definition, are not available. The changes in future weather seasonal mean and the fluctuations in both diurnal and seasonal behavior can be easily achieved through manipulation of the initial parameters determined for the present condition measured weather data. Furthermore, a relationship between the parameters that model measured weather data at different spatial scales can be established in order to 'downscale' the output, *i.e.*, to go from a vast weather scale to a local one before providing it to impact simulation models [4].

In the following sub-chapters further context is given to better understand the starting point of this thesis and how the methodology is obtained. The definitions and constants for this work are presented, followed by the initial concept for the weather model. Afterwards, the main mathematical tools utilized to describe weather data are presented.

2.1 Definitions

This sub-chapter presents a number of definitions which are essential to understand the explanation of the modeling concept. It also presents the values of some constants and the reasoning behind such assumptions.

Parameter

In this thesis, the word *parameter* holds a more specific meaning. It represents a constant that is numerically determined in order to represent specific weather behaviors.

Therefore, any constant that holds the same value for any given site is called *horizontal constant* instead of *parameter*. The following definitions are *horizontal constants* present in this weather model.

Diurnal and seasonal frequency

In this thesis, a diurnal and seasonal behavior are characterized by two different frequencies. A diurnal behavior has a period of 24 hours, therefore its frequency is represented by $\omega_d = 2\pi/24\,[rad]$. A seasonal behavior includes the four seasons of the year and is characterized by a period of 8760 hours, therefore its frequency is represented by $\omega_y = 2\pi/8760\,[rad]$.

Solar Constant

The definition of the solar constant (E_0) is the overall irradiance of the sun at the average orbital distance of the Earth. The solar constant represents the total amount of solar energy that is reaching the top of the atmosphere. Despite the name, the total solar irradiance does not remain constant throughout the years, it can have relatively small fluctuations [8]. Satellite measuring of the solar constant has a value of 1.361 kW/m^2 at solar minimum and when at solar maximum it has an approximate value of 1.362 kW/m^2, 0.1% greater than the minimum [9]. For the purposes of this thesis, the solar constant value assumed is 1.362 kW/m^2.

Atmospheric proportion

The atmospheric proportion (ρ) can be interpreted as a "filter" of the maximum possible solar irradiance. For the purposes of this thesis it takes a value of 0.9 and its impact is reflected in the maximum possible GHI value ($E_g = \rho \cdot E_0$) which has the value of 1.226 kW/m^2.

2.2 Initial concept

The initial concept [3] for the investigation in this thesis consists of two distinct procedures, a dynamic weather model or 'Synthesis' and an analysis concept. The Synthesis takes as an original input an internally generated stochastic time series of time-invariant standard normal distributed $N(0, 1)$ random values. The principle is that from such a signal, weather phenomena can be described through the introduction of its specific information such as its expected value, standard deviation, rate of change and correlation to other RER. The analysis concept can be understood as a tool which tries to capture the behavior of real weather through a series of numerical minimization problems. This base model possesses most of the characteristics described in the previously chapter, however in its approach to describe weather behavior, it requires a higher amount of parameters than what had initially been envisioned for a simple model framework. Furthermore, it sets itself to be a general concept which can be applied worldwide but the test proven data available is only for the location of Garching, Germany. Here a

general and short overview of its implementation is described in order to understand the sequence of transformations applied.

2.2.1 Synthesis

This section describes the procedure for analytically modeling three RER as time series x_n, with $n \in (I, A, V)$ for global horizontal irradiation, air temperature and wind speed respectively. Typically, the generated time series x_n are not normal distributed. To be independent of all the possible types of distributions and their particular domain of definition, the model assumes a space $\chi \in \mathbb{R}$ as the domain of definition with normal distributed values. The approach of the model hinges on stochastic time series that are time-invariant standard normal distributed $\big(N(0,1)\big)$ random values χ'_n. These signals are defined as white noise and have a sampling time of T_s, hence time discrete. The white noise signals are internally generated and assembled in a vector

$$\chi' = (\chi'_I, \chi'_T, \chi'_V) \tag{1}$$

which is the elementary input for the synthesis. The first information to be introduced in the random white noise is the correlation between RER and the specific rate of change for every RER. The specific rate of change is introduced by a differential operator, which results in a differential equation. Introducing the rate of change and causality results in the correlated noise χ''_n. The differential equations can be summarized in the following form

$$\chi''_n = \mathrm{f}_n(\chi'_I, \chi'_T, \chi'_V). \tag{2}$$

The correlated noise signal χ''_n is at this point auto- and cross-correlated colored $N(0,1)$ noise. The following step is to include time-variance in the correlated noise signal. Time-variance is introduced by applying $\chi_n = \chi''_n \cdot \sigma_n(t) + \mu_n(t)$ which is equivalent to (15). The mean and standard deviation are based on diurnal and seasonal variations. Since these are periodic signals, they are described by Fourier series. A periodic signal is perfectly described by its fundamental amplitudes A_k, frequencies A_0 and phases ϕ_k, where the $k \in [1, +\infty[$ is the order of the harmonics. Here a simplification is introduced by considering only the first order harmonics ($k = 1$). Thus, the mean and the standard deviation are modeled through

$$z_n(t) = A_{0z_n}(t) + A_{1z_n}(t) \cdot \cos(\omega_d t + \phi_{1z_n}), \tag{3}$$

$$A_{0z_n}(t) = A_{0A0z_n}(t) + A_{1A0z_n}(t) \cdot \cos(\omega_y t + \phi_{1A0z_n}), \tag{4}$$

$$A_{1z_n}(t) = A_{0A1z_n}(t) + A_{1A1z_n}(t) \cdot \cos(\omega_y t + \phi_{1A1z_n}), \tag{5}$$

$$\phi_{1z_n}(t) = A_{0\phi1z_n}(t) + A_{1\phi1z_n}(t) \cdot \cos(\omega_y t + \phi_{1\phi1z_n}). \tag{6}$$

The final step is to transform the time-variant normal distributed signals χ_n into the natural distribution of their respective RER, *i.e.*, to take the dimensionless time series χ_n and transform it into appropriate magnitude and unit for each RER. In general, the final step can be described by the following expression

$$x_n = Tr_n(\chi_n). \tag{7}$$

In the this base model, the solar irradiation has no established statistical distribution. First the values of time-dependent χ_I need to be mapped into a ratio η with a domain of $\eta \in [0, 1]$ through

$$\eta := \frac{1}{\exp(-\chi_I) + 1}. \tag{8}$$

Afterwards, the synthetic solar irradiation I is obtained by multiplying the ratio η with the theoretical maximum solar irradiation curve $I_{\max}(t)$ [10]

$$I := \eta(\chi_I) \cdot I_{\max}(t) \quad \text{and} \tag{9}$$

$$I_{\max}(t) = \max\big[0, E_g \cdot (\sin(\psi)\sin(\delta) + \cos(\psi)\cos(\delta)\cos(h))\big], \tag{10}$$

where ψ is the latitude of any given location, the hour angle is $h = \frac{2\pi}{1d}t$ and the solar declination is $\delta = 0.409 \cdot \sin\left(\frac{2\pi}{1y}(t - 81d)\right)$. The air temperature is assumed to already be normal distributed and therefore only requires the correct units, $T(t) = \chi_T \cdot 1\ ^\circ C$. The wind speed is assumed to be Weibull distributed and its transformation (36) is described in the methodology chapter.

2.2.2 Analysis

This section describes how the available measured data is analyzed to obtain the necessary parameters to generate synthetic weather data. An initial time series is created by combining the measured GHI, air temperature and wind speed into $x_{n'} = (I', T', V')$. A minimum of 8 years of consecutive weather data is necessary to ensure that the model accurately captures the average values and the deviations.

After the measured data are combined, a transformation to normal distributed time series is applied. This transformation is considerable different for the three weather phenomena since all of them have different natural distributions as previously mentioned. The transformation consist in applying the inverse transformation of each respective distributions such as $\chi_{n'} = Tr_n^{-1}(x_{n'})$. The transformation of GHI into a normal distributed time series is based on (8) and (9). The equation (9) cannot however be applied directly because the actual maximum solar irradiation $I_{\max}(t)$ is zero at night. Therefore the ratio η is solved by

$$\eta' = \frac{I'[t] \cdot I_{\max}(t)}{I_{\max}^2(t) + \epsilon} \tag{11}$$

with $\epsilon \to 0^+$, which is an infinitesimal positive constant. The transformation to a normal distributed time series is concluded by using the inverse (8).

The following step consists in analyzing the time-variance of the newly determined normal distributed time series $\chi_{n'}$. There are two components to the time-variance of a normal distributed series, the mean $\mu_{n'}$ and the standard deviation $\sigma_{n'}$. After being determined, these components need to be modeled for each of the RER. At this point, a division is made between GHI and the other RER. Unlike air temperature and wind

speed, the only relevant values in GHI are daytime values, therefore a cost function is introduced in order to disregard night time values ($I' = 0$) when modeling both the mean and the standard deviation of the GHI. Here, once again a discrete function is used and therefore it is addressed in the methodology chapter. The cost function is given by

$$\text{cost}(t) = \min[I_{\max}(t), W_{\max}] \tag{12}$$

where W_{max} is a constant design parameter which sets a maximum weight for the cost function. The time-variance $(\mu_{n'}, \sigma_{n'})$ in the now normal distributed time series are modeled by fitting the parameters (p_{nz}) in (3) through a numerical optimization process like the Levenberg-Marquardt algorithm. The determination of the set of parameters (p_{nz}) of every RER, 54 in total, is by far the main contributor for the high amount of parameters necessary to generate synthetic weather data and therefore one of the main topics that this thesis investigates. After determining the parameter for μ_n and σ_n, the next step is the removal of the time-variance of the normal distributed time series through $\chi''_{n'}[t] = (\chi_{n'}[t] - \mu_n)/\sigma_n$, which is the application of (15).

The two final steps are a spectrum analysis and a casual dependency analysis. The spectrum analysis is done through a fast Fourier transform. The time constant T_n is arbitrarily determined so that the resulting amplitude spectrum of $\chi''_{n'}$ is fitted the amplitude spectrum of a designed filter. The time constant is interpreted as the rate of change of the signal χ''_n. The casual dependency analysis determines the parameter ζ_n, which represents the correlation between the RERs. This correlation parameter is determined only for air temperature and for wind speed since the initial assumption is that only the GHI is not correlated from another RER. When determining the covariance between wind speed and air temperature, all points in time are weighted equally, however the same does not apply for air temperature and GHI. Here, there is once again the need to use the cost function (12) to maintain consistency.

2.3 Characteristics for the model

Each of the weather phenomenon, global horizontal irradiation, air temperature and wind speed, requires a specific implementation due to their different behavior. This general concept includes a set of requirements which aim to ensure the quality of the generated synthetic data. These requirements are: analytical modeling, time-variance, rate of change, causality and application of simple mathematic methods [3].

(i) Analytical modeling: The requirement for the model to have an analytical description is the pillar of the methodology introduced in this general concept. This requirement allows the dynamic model to be used for the assessment of the potential of renewable energies and for potential system designs. Analytical description also allows the model to generate weather data in different time interval than the input data. This feature satisfies the demands from various power systems/models which require considerable different time step intervals in their input data. Example of such cases are energy management and grid control models which require time steps between input data of hours/minutes and seconds/milliseconds respectively.

(ii) Time-variance: The consideration of time-variance is related to the necessity of describing both seasonal and diurnal variations of all the weather phenomena. Such behavior is easily described through solar irradiation. Throughout the course of a day, there is a clear distinction between day time and night time with a relatively short transition period between them (sunrise/sunset) which corresponds to diurnal behavior. A seasonal behavior corresponds to the difference in average and absolute values in solar irradiation between seasons such as winter/summer.

(iii) Rate of change: A particular important requirement for modeling weather phenomena is its rate of change. Each RER has its own typical rate of change and it represents the typical duration of a certain weather phenomenon. It can be understood as temporal inertia. Each value in time is, until a certain extent, influenced by its previous immediate value and therefore it cannot be radically different from it. An example of such behavior is the values of air temperature during a time period of a day. It increases and decreases during the course of the day but it never manifests a sudden radical drop or peak in a short time interval of, for example, an hour.

(iv) Causality: RER have a casual relationship between them. This requirement is a fundamental assumption for the general concept and not always evident. The most evident example of causality between different RER is the correlation between global horizontal irradiation and air temperature.

(v) Simple Model: One of the main objectives in this thesis is to be able to describe the chaotic weather phenomena with the use of simple mathematics methods and equations. In this context, simplicity means that the behavior worldwide is modeled by stochastic processes in order to avoid using complex deterministic processes as applied in numerical global climate models. This requirement of simplicity is arguably what distinguishes the most between a synthetic weather generator model from a forecasting model.

2.4 Normal Distribution

In this thesis, the normal distribution is vastly used, therefore its introduction is done shortly here. The normal distribution is commonly used to describe many natural phenomena (physical, biological, and social measurement). For example, the distribution of the errors made in repeatedly measuring a 1 kilogram weight, of the wingspans of a large colony of butterflies and of the amount of sleep a normal human being gets per night are approximately normal [11]. It is a useful tool to describe a quantity which varies around a mean value with a given distribution. For example, the simplest way to characterize a random variable is through the expected value and standard deviation. In addition, it is relatively easy to mathematically work with it. In this thesis, it is used to describe a stochastic signal and introduce a modeled daily and seasonal behavior of weather phenomena. The equations and parameters of the normal distribution are manipulated to take different shapes and forms but their basis is in the same equations.

The *probability density function* $\mathrm{f_N}$ of the normal distribution is given by

$$\mathrm{f_N}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-\mu)}{\sigma^2} \right) \tag{13}$$

where the probability of observing x is $\mathrm{f_N}(x)$, the expected value μ of the distribution is and the standard deviation is σ. The *cumulative probability function* $\mathrm{F_N}$ for a generic normal distribution is given by

$$\mathrm{F_N}(x) = \Phi\left(\frac{(x-\mu)}{\sigma} \right) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}} \right) \right]. \tag{14}$$

Manipulation of a standard normal distributed time series into a different normal distributed time series can be done if the mean and standard deviation of the latter are know. Given a standard normal deviate Z ($\mu_Z = 0, \sigma_Z = 1$), then

$$X = Z \cdot \sigma + \mu, \tag{15}$$

where X has a normal distribution with μ and σ as its expected value and standard deviation respectively. Conversely, given X as a standard normal deviate with parameters μ and σ, then $Z = (X - \mu)/\sigma$ will have a standard normal distribution. This variate is called the standardized form of X. This manipulation of time series is unfortunately not directly possible if a set of data has a different natural distribution as for example a Weibull distribution.

2.5 Weibull Distribution

Wind speed is an important weather phenomenon for RES and Weibull distribution is the most commonly used statistical distribution to model its behavior. In general, prediction methods for wind power are divided into two different groups, a physical and a statistical group. The first group implies physical considerations such as local temperature, terrains, topography and pressure to more accurately estimate the wind field. The latter group, on the other hand, uses statistical models to establish the correlation between power and other variables while taking into account their forecast values and their historical data [12]. The Weibull distribution is able to better fit to the probability distributions than the Rayleigh model (which is a special case of the Weibull distribution) and analyzes the wind speed data by using statistical distributions [13]. The Weibull distribution is able to provide a reasonable accurate approximation to the probability of many weather and other phenomena in nature. It has been used to represent wind speed distributions for application in wind loads studies [12]. The *probability density function* of the Weibull distribution f_W is given by

$$\mathrm{f}_W(v) = (\lambda k)(\lambda v)^{k-1} \exp\left[- (\lambda v)^k \right] \tag{16}$$

where the Weibull shape parameter is k and the Weibull scale parameter is λ. These parameters have reference values in the same units as wind speed. The *cumulative*

probability function of the Weibull distribution F_W is given by

$$\mathrm{F}_W(v) = 1 - \exp\Big[- (\lambda v)^k \Big]. \tag{17}$$

In order to determine the parameters of the Weibull distribution, an equation is necessary to fit the existing discrete cumulative frequency distribution. By applying the natural logarithm to both sides of the equation, this equation is now given by

$$\ln\big(- \ln[1 - F(v)]\big) = k \ln(v) - k \ln\Big(\frac{1}{\lambda}\Big). \tag{18}$$

2.6 Lambert W function

The Lambert W function is a useful mathematical tool which can be used to solve equations where an unknown quantity appears inside a logarithm or a exponential function and also outside of it, for example $2x + 5 = e^x$. These types of equations are not explicitly solvable, except for particular cases. The definition for the Lambert W function is the multivalued inverse of the function $w \to we^w$ [14]. This function has a variety of uses in both pure and applied mathematics; a particular one is used in this thesis for the manipulation of an analytical expression. The Lambert W function allows for the explicit solving of the inverse transformation for the solar irradiation. The analytical description of the model is one crucial objective in this work and therefore the use of such a function becomes a necessity rather than a luxury. The form which is used in this thesis is [14]

$$W(z)e^{W(z)} = z. \tag{19}$$

3 Methodology

In this chapter, a detailed description of the specific implementations introduced in the base model and iterations to the modeling approach is done. First and foremost, the central objective of the methodology is to provide the fundamental description of weather phenomena, *i.e.*, to model realistic behavior of weather through a dynamic system with the bare minimum of mathematical complexity.

The approach for investigating developments in the base model consisted in three different phases. Firstly, developments to reduce the number of parameters and assuring the continuity of the entire model were investigated for the site location of Garching, Germany. The advantage of this initial approach is a reliable set of data and comparable results from the base model. Once ensured that the developments properly fit the expected output of Garching, the second phase is initiated.

The second phase consists in ensuring that the implementations are not over-fitted for the initial location tested. Here sets of data for a number of different sites around the world, with different climate zones, are evaluated. Similar to the procedure in the first phase, the implementations are tested and upon failure to produce reasonable results, the implementation is iterated. The continuous refinement of the implementation is an iterative process with no limit in iterations. The overall objective is minimizing errors for any given latitude of a location rather than having intervals in latitude where the model produces exceptional results and other intervals where it produces mediocre results. Once the model is deemed fit to handle any given location, the third phase of the approach is initiated.

The third phase of the approach consists in testing the model for another number of different locations from the second phase with possibly similar climate zones. In this phase, a remarkably different constraint is used from phase two. The testing and results of these new locations do not contribute in any way to improving the implementation done. The constrain is set to evaluate the results of unpredicted inputs and therefore these are used to determine the real errors and accuracy of the model. In Figure 1 this approach for the development of the weather is presented in a schematic way.

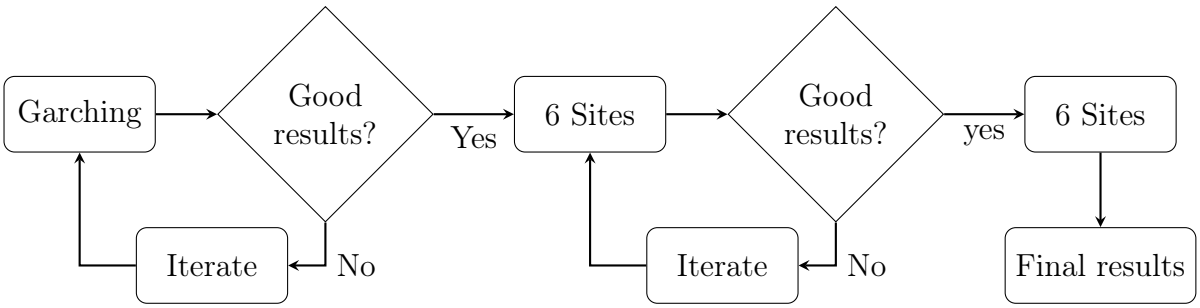


Figure 1: Flow chart of the approach taken to develop the model.

3.1 Structure

The structure of this work consists in two main parts, the dynamic model for weather phenomena and an analysis of the measured weather features as it is done in the base concept for this work. The framework of the base model shows flexibility for alteration of the transformations applied and shows no fundamental theoretical flaw, therefore it remains valid in this thesis. However, specific implementations are altered to reduce its mathematical complexity. In a practical way, the analysis of a site is done prior to the synthesis, otherwise the synthesis does not have the necessary parameters to model the weather. However, theoretically, the analysis is a function of the synthesis. The dynamic model is designed first for a number of steps, transformations and parameters. Only afterwards is the analysis designed to capture the specific features which the synthesis aims to describe in the form of parameters.

3.2 Synthesis

The synthesis is the dynamic model and where synthetic weather data are generated. It is a analytical description of the fundamental features of weather behavior. The weather phenomena are $n \in [I, T, V]$ as previously described and they necessarily follow this order, which means that the index $n + 1$ refers to the immediate following weather phenomenon. The model uses as its input a random standard normal distributed white noise signals independently generated for each weather phenomenon. These are time series with a sampling time of $T_s \in \mathbb{R}^+$. The probability density function of these signals is represented in the following Figure 2.

Their generation is done through the MATLAB function *randn*, which returns a random scalar drawn from the standard normal distribution. It is done internally to ensure that every simulation run is different and to avoid an external input requirements. This input is transformed into synthetic weather data through three step of modeling

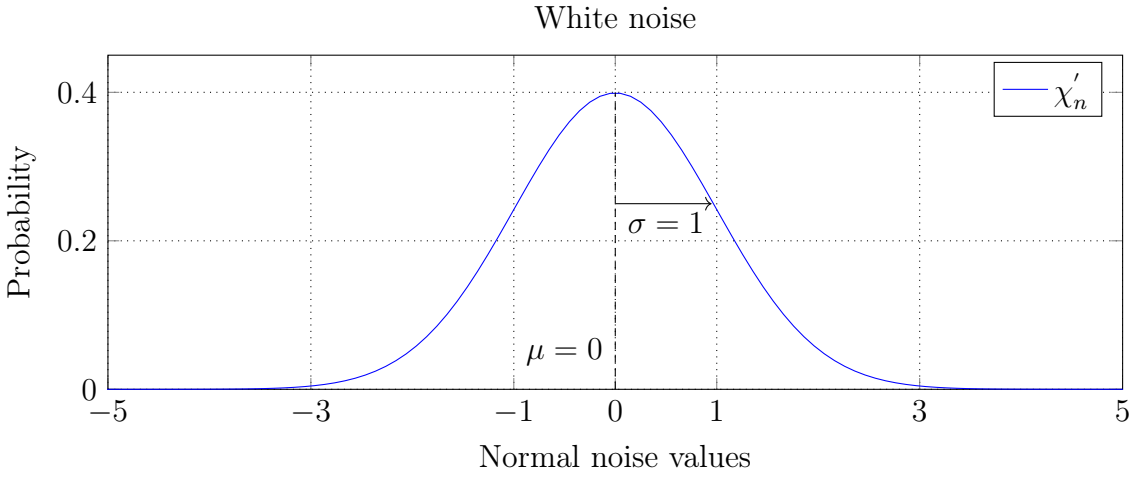


Figure 2: Probability density function for the fundamental input of the model χ_n'.

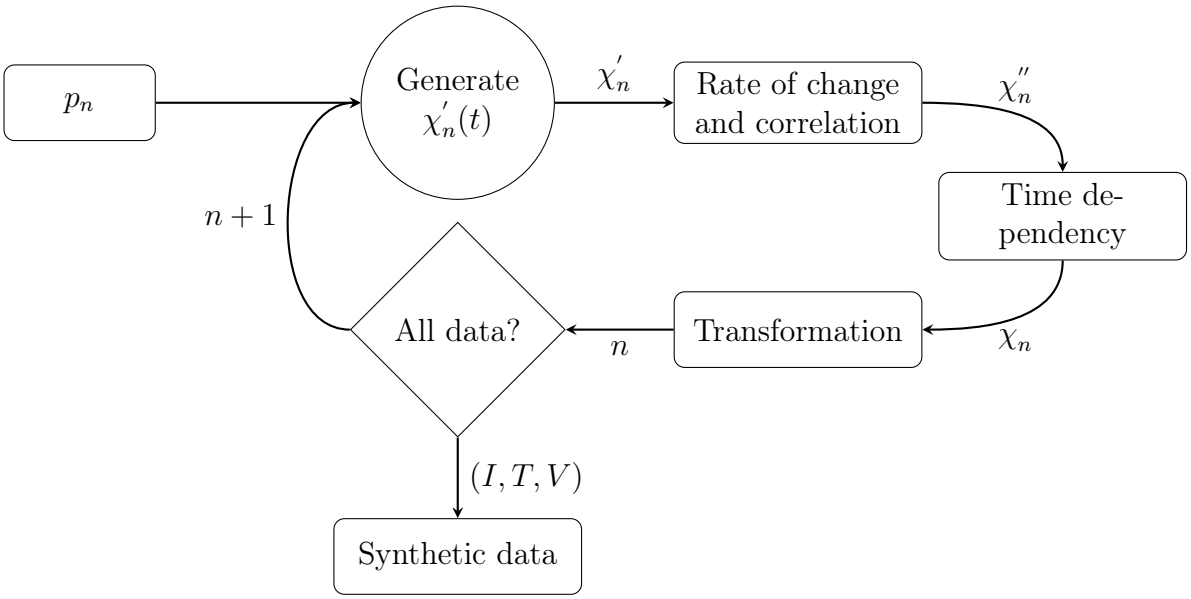


Figure 3: Flow chart of the synthesis model.

which are presented in Figure 3. The flow chart shows a cyclical behavior and it is done in such a way to maintain the order of the weather phenomena n.

3.2.1 Step 1: Introduction of rate of change and correlation between RER

The first step is to incorporate the dynamics of the weather. These dynamics are implemented through an ordinary differential equation with the objective of limiting the rate of change and introducing a causality relationship between the different weather phenomena. A rate of change means that values have a correlation in time, therefore at any given value is correlated to its immediate previous value and can only change so much from it.

A correlation between weather phenomena means that a value of one phenomenon influences to a certain extent the value of a correlated phenomenon. The modeling of the rate of change and the causality is based on the dynamics of a first order low pass system with a numerically determined gain parameter B_n and time constant T_n for each weather phenomenon. GHI is a source RER and does not depend on any other RER, therefore its dynamics result in

$$\chi_I^{''} = \frac{B_I}{1 + T_I s} \cdot \chi_I{}' \tag{20}$$

where s is the Laplacian frequency parameter and the gain parameter B_n is expressed by

$$B_n = \sqrt{\frac{2T_n}{T_s}}. \tag{21}$$

Only a time correlation is introduced in (20). To introduce such behavior, only the time constant T_n needs to be numerically determined since the time T_s is an horizontal

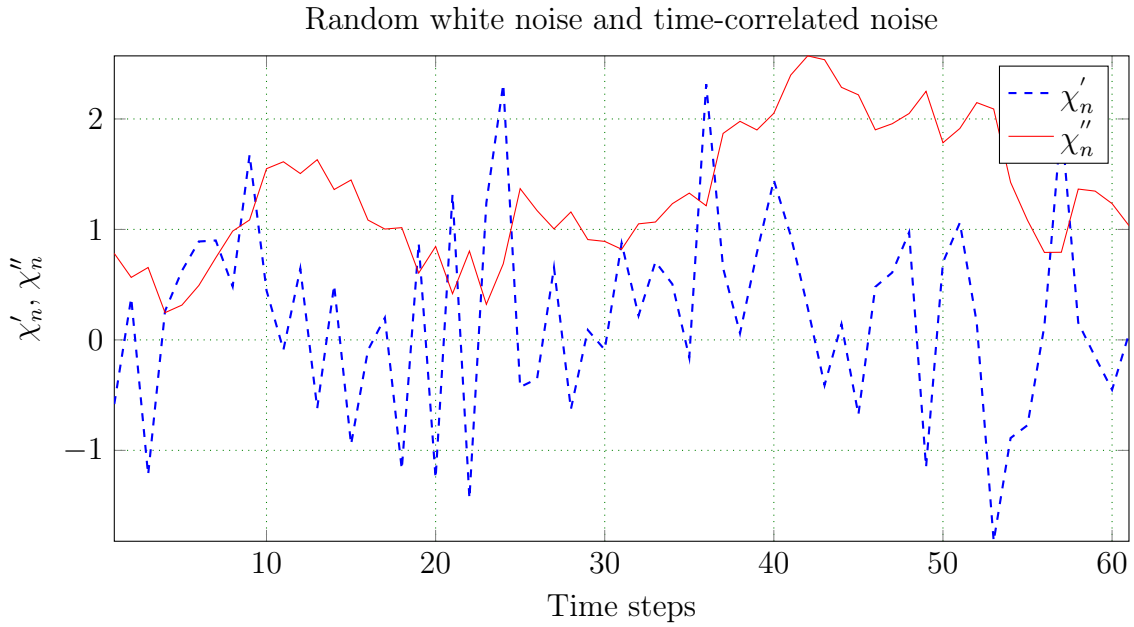


Figure 4: Visual introduction of rate of change in random white noise.

constant introduced in the model and the gain parameter is a function of these two which is used for a compact spelling of the filter.

Assumption 1. The RERs are correlated to each other.

The air temperature and wind speed are assumed to be correlated to their immediate previous RER, *i.e.*, air temperature depends on the solar irradiation values and wind speed depends on the air temperature values. Therefore a correlation parameter ζ_n is required. For $n \in (T, V)$ the time and cross correlation is introduced through

$$\chi_n{}'' = \frac{B_n}{1 + T_n s} \cdot \frac{\zeta_n \, \chi_{n-1}^{''} + \chi_n^{'}}{\zeta_n + 1}. \tag{22}$$

This standardization of the filters ensures that when random white noise $\chi_n^{'}$ is transformed to time- and cross-correlated noise $\chi_n^{''}$ by (20) and (22) it remains $N(0, 1)$-distributed. The impact of this first step in the signal is presented in Figure (4). The time series χ_n' is a very chaotic signal with abrupt changes in value. However, after applying the 1st order low pass system, it becomes a much smoother signal, *i.e.*, the fluctuations in value are now considerably smaller.

3.2.2 Step 2: Introduction of time dependency

The second step in the model is to introduce time variance, which are the diurnal and seasonal behaviors of weather. The introduction of these behaviors leads to a seasonal and diurnal dependent mean $\mu(t)$ and standard deviation $\sigma(t)$ of the time series. The $\mu(t)$ and $\sigma(t)$ are periodic signals which are described by a series of sinusoidal waves. There

are four different components necessary to completely describe $\mu(t)$ and $\sigma(t)$. The first component is their expected value, represented here by the parameter A_0. The second component is the seasonal behavior, represented by the sinusoidal wave $A_y \cdot \cos(\omega_y t - \phi_y)$ with seasonal amplitude A_y and seasonal phase ϕ_y. Here the frequency ω_y emulates the seasonal variations and seasonal phase ϕ_y sets the peak of this wave. The third component is the diurnal behavior, represented by the sinusoidal wave $A_d \cdot \cos(\omega_d t - \phi_d)$ with diurnal amplitude A_d and diurnal phase ϕ_d. Here the frequency used is ω_d which represents the diurnal variations and diurnal phase ϕ_d sets the peak of this wave to a particular hour during a day period. The most evident cycles of $\mu(t)$ and $\sigma(t)$ are described by the first three components, however their sum would generate a wave that has a constant difference between the maximum and the minimum value of a day. This invariability is solved by the fourth component, the seasonal variation of the diurnal variation. This component is represented by the sinusoidal wave $A_s \cdot \cos(\omega_d t - \phi_d) \cdot \cos(\omega_y t - \phi_y)$ with the amplitude A_s and same diurnal and seasonal phases of the second and third component. This component ensures that there is no constant difference between maximum and minimum values throughout seasons. The description of $\mu(t)$ is finally achieved by summing all four components and results in

$$z_n(t) = A_0 + A_d \cos(\omega_d t - \phi_d) + A_y \cos(\omega_y t - \phi_y) + A_s \cos(\omega_d t - \phi_d) \cos(\omega_y t - \phi_y), \quad (23)$$

where $z \in (\mu, \sigma)$. The visualization of this sinusoidal signal is done in Figure 5. This figure separates the signal into its diurnal variation ($t_{diurnal} \in [1, 24]$ h) and seasonal variation ($t_{seasonal} \in [1, 8760]$ h) in order to create a surface. The impact that the first three components have in the overall shape of the surface is explicitly pointed in the Figure 5. The fourth component is not however present because its magnitude is in general considerably lower than A_0, A_d and A_y, making it difficult to perceive from a visual point of view.

The modeling of the mean behavior $\mu(t)$ by (23) is applicable as well for the behavior of the standard deviation as previously pointed out. Despite the similarities, its influence in the outcome of the model is lower than the mean, making it the prime target for reduction of parameters. This reduction is achieved by introducing a series of assumptions that correlate the parameters of the mean with the parameters of the standard deviation. The consequence of introducing assumptions is that the model becomes stiffer, *i.e*, less capable of handling unexpected or irregular data inputs. It also results in (23) splitting into three different equations. The number of parameters necessary is downsized from 6 per RER to 1 which is the average of the input time series $\bar{\sigma}$. Despite the reduction in complexity, the model is still able to reproduce both daily and seasonal behavior with a reasonable accuracy. The modeling of the standard deviation is done individually for each RER in the general order of the model, solar irradiation, air temperature and finally wind speed.

Irradiation

Assumption 2. The mean and the standard deviation of GHI have the same daily and seasonal behavior.

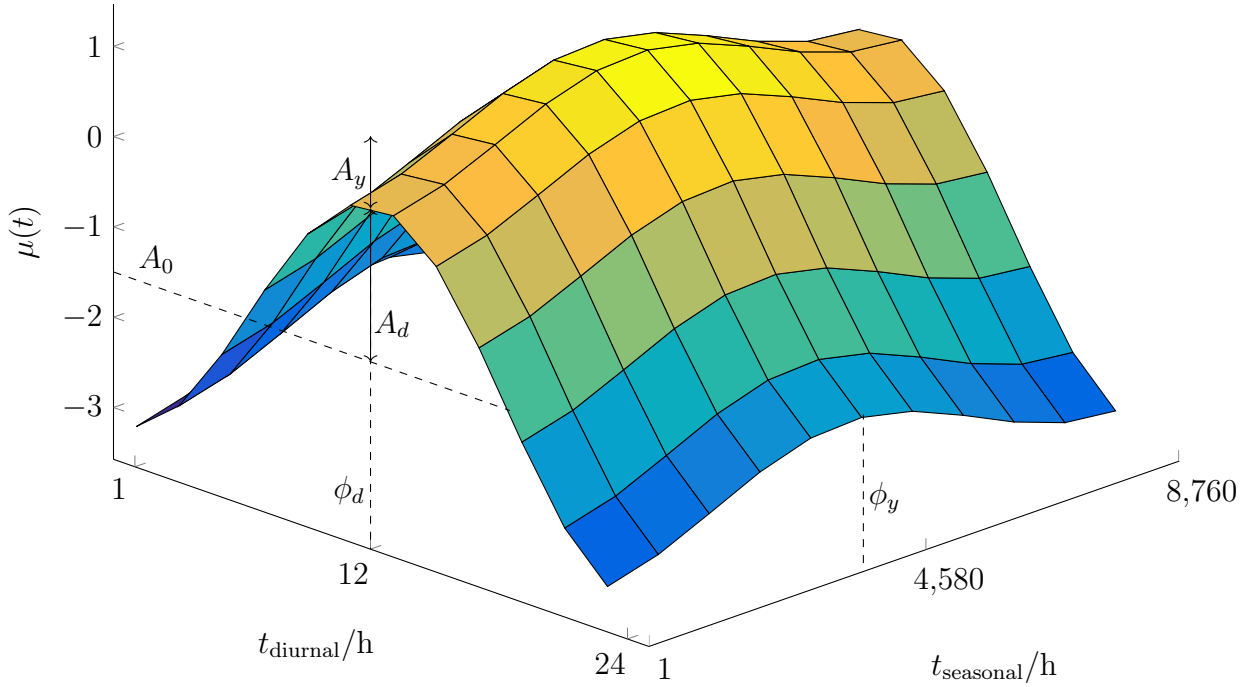


Figure 5: Illustration of the diurnal and the seasonal variation of the mean $\mu(t)$ by two distinct axes.

The assumption introduced correlating the mean and standard deviation of the GHI is that both daily and seasonal peaks have the same phase, *i.e*, they happen at the same time. The implication of this assumption is that the already numerically determined parameters ϕ_d and ϕ_y of the mean are used in the modeling of the standard deviation. The implementation of the assumption is not trivial though, the phase ϕ_y can not be implemented before assuring that the amplitude A_y of the mean $\mu(t)$ has the same sign as $\bar{\sigma}$. The problem exists because two identical sinusoidal waves can be described by

$$C \cdot \cos(\omega t + \phi) = -C \cdot \cos(\omega t + \phi + \pi) \tag{24}$$

where C, ω, ϕ are an arbitrary constant, frequency and phase respectively. When modeling the mean $\mu(t)$, the parameter A_y does not have any boundary conditions for its value ($A_y \in \mathbb{R}$). However, the average value of any standard deviation is necessarily a real positive value ($\bar{\sigma} \in \mathbb{R}_0^+$). It is necessary to ensure that regardless of the value of A_y, the respective phase ϕ_y is in the form of $C \cdot \cos(\omega t + \phi)$ when applied in the standard deviation equations and it is done by

$$\phi_{\sigma_n} = \phi_{y_n} - \left(\frac{\pi}{2} - \frac{\pi}{2} \cdot \frac{A_{y_n}}{|A_{y_n}| + \epsilon} \right) \tag{25}$$

where ϕ_{σ_n} is the correct standard deviation seasonal phase and $\epsilon \to 0^+$ is an infinitesimal positive constant. However, this approach brings an undesirable level of complexity into the methodology. Therefore, an iteration to the modeling procedure of the mean is applied, such that when A_y is numerically determined, it is forced to be positive ($A_y \in \mathbb{R}_0^+$). By doing so, it is possible to ensure that the sinusoidal wave has the appropriate form of $C \cdot \cos(\omega t + \phi)$ and $\phi_{y_\mu} = \phi_{y_\sigma}$. In the case of solar irradiation, the almost binary type of behavior from the standard deviation (0 at night, positive during the day) requires particular attention and is the reason why the simplified equation for it suffers the most alterations of all the RER. The seasonal behavior of the standard deviation in introduced by

$$H_I(t) = \bar{\sigma_I} + \frac{A_{y_I}}{|A_{0_I}|}\bar{\sigma_I} \cdot \cos(\omega_y t - \phi_{y_I}) \tag{26}$$

where A_{y_I} and A_{0_I} are numerically determined parameters from the mean $\mu_I(t)$. $H_I(t)$ is, once again, merely a variable which allows a more compact form of the final standard deviation equation. Finally, the daily behavior is introduced by

$$\sigma_I(t) = \frac{H_I(t)}{2} - \frac{H_I(t)}{2} \cdot \cos(\omega_d t - \phi_{d_I}) \tag{27}$$

where $\sigma_I(t)$ is the modeled standard deviation for GHI and ϕ_{d_I} is the daily phase for the modeled mean $\mu_I(t)$. Despite the unintuitive mathematical description of the equation, its shape is still remarkably similar to the one of the modeled mean. The alteration of its original shape of (23) simply guarantees that for any time step the value of the modeled standard deviation is positive ($\sigma_I \ \forall\, t$) which is the actual behavior of the standard deviation from measured data.

Temperature

Assumption 3. The mean and the standard deviation of air temperature have the same daily behavior and inverse seasonal behavior.

The behavior of the standard deviation of air temperature follows the same daily pattern as the modeled mean, *i.e.*, the daily peaks appear simultaneously. However, the peak in seasonal behavior is inverse, *i.e.*, while the peak highest air temperature values appear during the summer, the highest values of the standard deviation appear during winter time. An additional distinction from the behavior of GHI is that the value of the standard deviation of air temperature are at any given time over the year about zero ($\sigma_T(t) > 0 \ \forall\, t$). Since the boundary condition of $A_y \in \mathbb{R}_0^+$ is imposed in the modeling of the mean, no additional transformations such as (25) are performed. The mathematical description of the new simplified equation for the standard deviation of the temperature is the same as the general equation (23) but instead of having its parameters numerically determined, ratios from the parameters of the mean are used. The seasonal behavior is introduced by

$$H_T(t) = \bar{\sigma_T} - \frac{|A_{d_T}|}{|A_{0_T}|}\bar{\sigma_T} \cdot \cos(\omega_y t - \phi_{y_T}) \tag{28}$$

where $H_T(t)$ is, once again, merely a variable which allows a more compact form of the final standard deviation equation. Finally, the daily behavior is introduced by

$$\sigma_T(t) = H_T(t) + \frac{|A_{d_T}|}{|A_{0_T}|}\bar{\sigma_T} \cdot \cos(\omega_d t - \phi_{d_T}) + \frac{|A_{s_T}|}{\bar{\sigma_T}} \cos(\omega_y t - \phi_{y_T}) \cos(\omega_d t - \phi_{d_T}) \quad (29)$$

where $\sigma_T(t)$ is the modeled standard deviation for air temperature and $A_{d_T}, A_{0_T}, A_{s_T}$, ϕ_{y_T}, ϕ_{d_T} are parameters from the modeled mean $\mu_T(t)$. This form of the equation guarantees that the standard deviation only has positive values.

Wind Speed

Assumption 4. The mean and the standard deviation of wind speed have the same seasonal behavior and opposite daily behavior.

The behavior of the standard deviation of wind speed follows the same seasonal pattern as its modeled mean, *i.e.*, the seasonal peaks appear simultaneously. However, the peak in daily behavior is inverse, *i.e.*, when the peak for the highest wind speed values appear during the day, the highest values of the standard deviation appear during night time. The average and seasonal behavior is modeled by

$$H_V(t) = \bar{\sigma_V} - \frac{|A_{d_V}|}{|A_{0_V}|}\bar{\sigma_V} \cdot \cos(\omega_y t - \phi_{y_V}) \quad (30)$$

where $H_V(t)$ is, once again, merely a variable which allows a more compact form of the final standard deviation equation. Finally, the daily behavior is introduced by

$$\sigma_V(t) = H_V(t) - \frac{|A_{d_V}|}{|A_{0_V}|}\bar{\sigma_V} \cdot \cos(\omega_d t - \phi_{d_V}) + \frac{|A_{s_V}|}{\bar{\sigma_V}} \cos(\omega_y t - \phi_{y_V}) \cos(\omega_d t - \phi_{d_V}) \quad (31)$$

where $\sigma_V(t)$ is the modeled standard deviation for wind speed and $A_{d_V}, A_{0_V}, A_{s_V}, \phi_{y_V}, \phi_{d_V}$ are parameters from the modeled mean $\mu_V(t)$. This form of the equation guarantees that the standard deviation only has positive values.

The three different equations for the standard deviation have a remarkably similar shape to the one of the mean represented in Figure 5. At this point, both behaviors are modeled and their information is in a correct shape to be introduced in the time- and cross-correlated signal χ_n''. As previously mentioned, the signal χ_n'' is still standard normal distributed $N(0, 1)$. The introduction of the μ and σ is implemented through

$$\chi_n = \chi_n'' \cdot \sigma_n + \mu_n. \quad (32)$$

This is a simple mathematical transformation which attributes the most distinct behavior of the weather to a normal distributed time series. The impacts of this step are represented in Figure 6. In this figure the shift from the standard normal distributed mean μ_{SN} to μ and from their respectively standard deviations σ_{SN} to σ.

The time dependent signal χ_n has integrated all the fundamental behavior of weather. Here it is presented in Table 1 with all the parameters necessary to perform the first

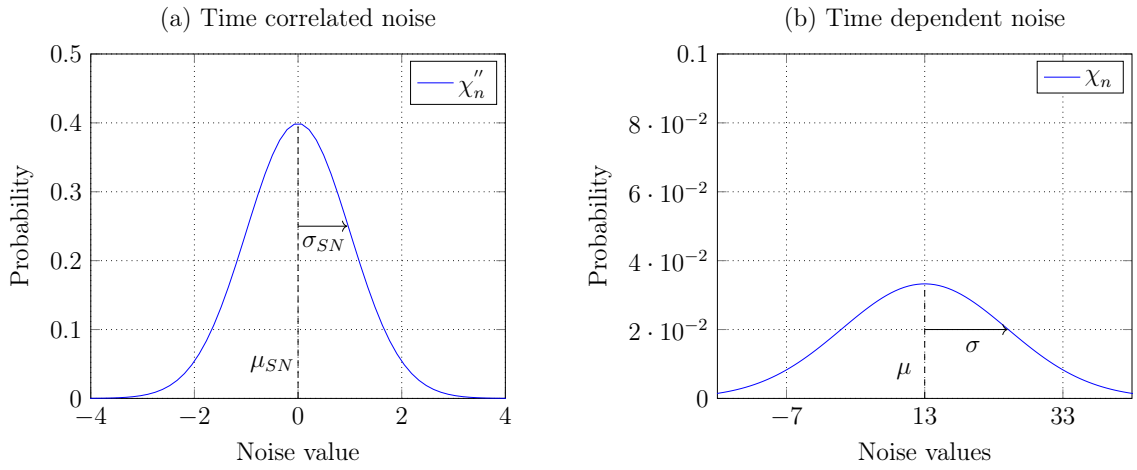


Figure 6: Visual introduction of time dependency in a time-correlated noise.

and second step separated by RER. The majority of them are dimensionless because the different signals have the same normal distribution and do not require the units of the weather phenomena into which they will be transformed. The exceptions are the phases for the diurnal and seasonal behaviors and the time constant. The phases are worked with in radians to adjust the sinusoidal waves. The basis of this model is to provide a continuous description of weather behaviors in time. The typical rate of change of every RER in this work is best described by the units of hours because it is both the most intuitive unit to understand and provides a wide but not overwhelming range of values it can take.

Table 1: Discretization of all parameters necessary to model the fundamental behaviors of weather phenomena.

Parameter Table				
Parameter type	Unit	GHI	Air Temperature	Wind Speed
Mean (μ_n)	[-]	A_{0_I}	A_{0_T}	A_{0_V}
	[-]	A_{d_I}	A_{d_T}	A_{d_V}
	$[rad]$	ϕ_{d_I}	ϕ_{d_T}	ϕ_{d_V}
	[-]	A_{y_I}	A_{y_T}	A_{y_V}
	$[rad]$	ϕ_{y_I}	ϕ_{y_T}	ϕ_{y_V}
	[-]	A_{s_I}	A_{s_T}	A_{s_V}
Standard deviation (σ_n)	[-]	$\bar{\sigma_I}$	$\bar{\sigma_T}$	$\bar{\sigma_V}$
Time constant (T_n)	[h]	T_I	T_T	T_V
Correlation (ζ_n)	[-]	-	ζ_T	ζ_W

3.2.3 Step 3: Transformation into synthetic weather data

The third and final step of the model is to transform the normal distributed signal χ_n into the natural distribution of their correspondent weather phenomena n. Each weather

phenomena has a different assumed distribution and therefore the transformation are individually designed. In this chapter, a detail overview of the different procedures used to obtain synthetic weather data is presented. Their order of transformation remains identical to the previous steps.

Irradiation

The solar irradiation has a natural distribution which is not accurately described by any established statistical distribution, *i.e.*, normal, binomial or others. For this reason, the transformation which maps the signal χ_I to realistic values of solar irradiation needs to be designed and adjusted to this modeling approach. In the base concept previously presented, a discontinuous function (10) was used. Here a continuous transformation is used to map the entire domain of $\chi_I \in \mathbb{R}$ to $I \in [0, E_g]$ where E_g is the realistic maximum possible solar irradiation determined through $E_g = E_0 \cdot \rho$. The most difficult component to integrate in such a transformation is the transition between night and day time at sunrise and sunset. This behavior is integrated by creating an asymmetric transformation focused ramping up from its minimum value within a very short domain. Furthermore, it needs to limit the values of solar irradiation to the domain of $[0, E_g]$ in an asymptotic way. The overall transformation has the following mathematical description

$$I = \frac{E_g}{\exp\Big(\exp(-\chi_I) - \chi_I\Big) + 1}, \tag{33}$$

where I is synthetic solar irradiation. The visualization of this transformation is represented in Figure 7 and offers a much more intuitive explanation for the transformation (33). First of all, the domain for the solar irradiation is respected. The focus on the transition between day and night time is pointed out in Figure 7. This transition occurs within the interval of $\chi_I \in [-2, -0.3[$ which is considerably small considering its entire domain of $\mathbb{R}$. The resulting synthetic weather time series are presented in the form of a histogram in Figure 8. This figure shows in (a) an example for what the signal χ_I could look like in its probability density function form. It is also noticeable that after incorporation the time-dependency behavior of the solar irradiation, the signal χ_I is no longer perfectly fitted by a normal distribution. In (b), there is an example of what the synthetic solar irradiation data could look like in a probability density function form. In this example, the synthetic data fit the typical distribution of measured solar irradiation. The major concentration of values is either zero or in the vicinity of zero as expected. The rest is almost evenly spread out which is typical for most regions around the world, not close to earth's poles.

Air temperature

Assumption 5. Air temperature is assumed to be normally distributed.

The air temperature shares the same statistical distribution of its time dependent signal χ_T. For this reason, the only transformation applied to the signal χ_T is the

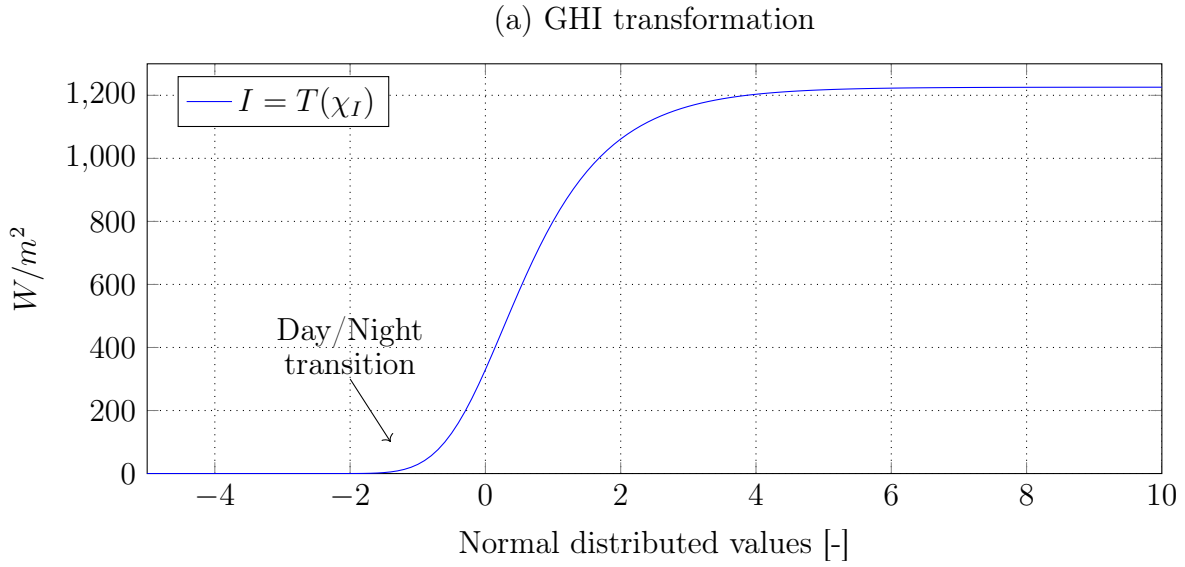


Figure 7: Visualization of the transformation applied to map a normal distributed time series into the natural distribution of solar irradiation.

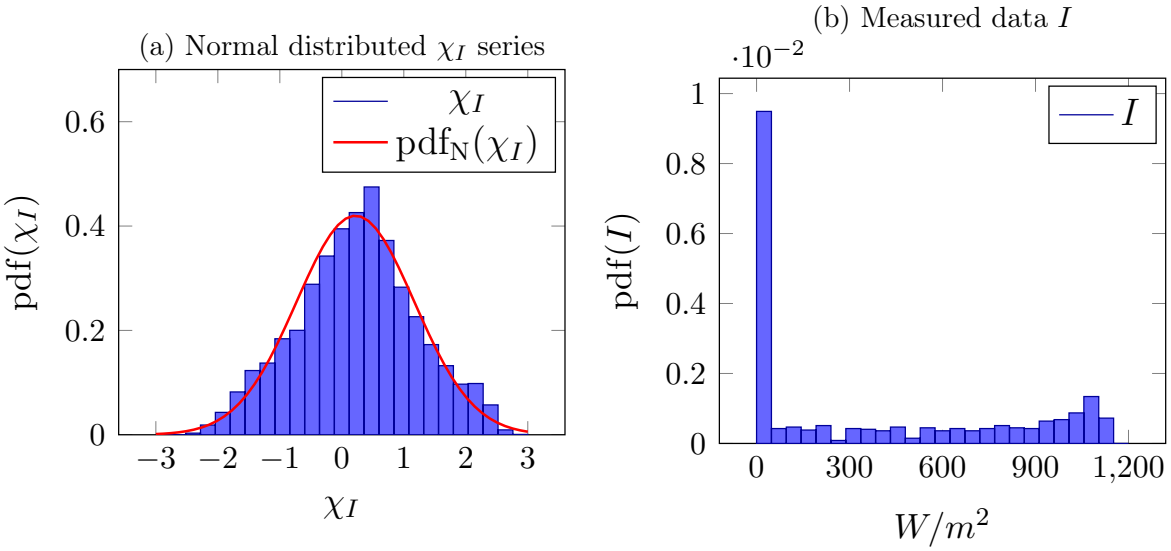


Figure 8: Histograms of GHI's natural distribution and of the normal distributed time series χ_I.

attribution of the correct units $°C$ by $T = \chi_T \cdot 1°C$. In Figure 9, an example for what the synthetic air temperature data could look like in an histogram form is represented. This figure shows that the weather behavior once again skew the normal distribution from a perfect fit but it is still a very good description.

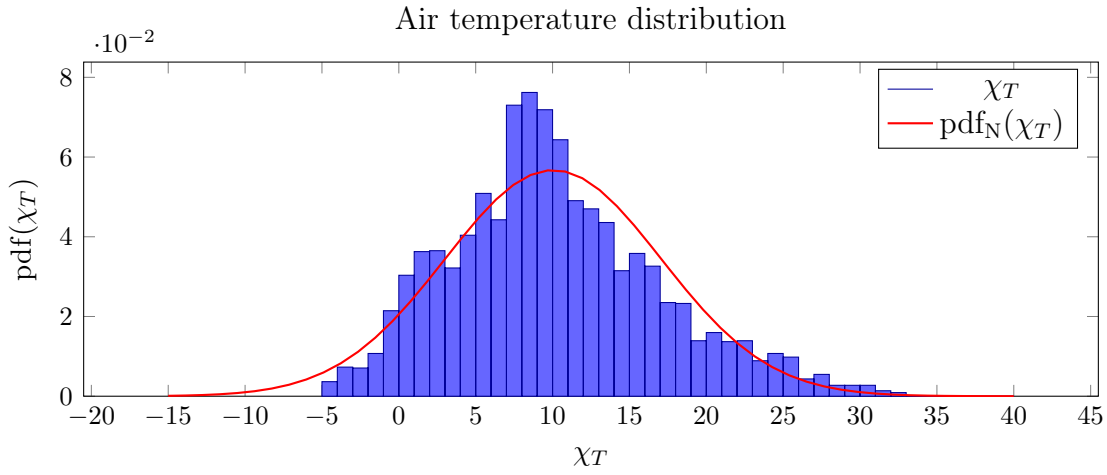


Figure 9: Histogram of air temperature with its normal probability density function $\mathrm{pdf}_{\mathrm{N}}(\chi_T)$.

Wind Speed

Assumption 6. Wind speed is assumed to be Weibull distributed.

In this work, the wind speed is modeled through a Weibull distribution. The time variant signal for the wind speed χ_V is entirely normal distributed, therefore a relationship needs to be established to transform χ_V into synthetics wind speed data V. The expected probability density function expected for the wind speed is described in (16). The signal χ_V has at this point a normal probability function of

$$\mathrm{pdf}_N(\chi_V) = \frac{1}{\sigma_N\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\left(\frac{\chi_V - \mu_N}{\sigma_N}\right)^2\right). \tag{34}$$

In order to properly apply the transformation from χ_V to V, a constraint is here introduced. This constraint guarantees that the probability for values inferior to χ_V are equal to the probability for the values inferior to $Tr_V(\chi_V)$. This condition leads to the following equality (35) and analytical transformation (36)

$$\int_{-\infty}^{\chi_V} \mathrm{pdf}_N(\chi_V)\, d\chi \overset{!}{=} \int_{T(-\infty)}^{T(\chi_V)} \mathrm{pdf}_W(V)\, dV \tag{35}$$

$$\Longrightarrow V = \frac{1}{\lambda}\sqrt[k]{-\ln\left(\frac{1}{2} - \frac{1}{2}\cdot \mathrm{erf}\left(\frac{\chi_V - \mu_N}{\sqrt{2}\sigma_N}\right)\right)} \tag{36}$$

where erf is the Gaussian error function. In this transformation, the parameters for the Weibull distribution (k, λ) are introduced. These represent two degrees of freedom for the transformation. In order to avoid further addition of parameters to the set already presented in Table 1, the two remaining degrees of freedom are solved by approximating

the Weibull parameters (k, λ). Here a correlation between the Weibull parameters and the normal distribution parameters (μ_N, σ_N) is introduced. The approximation of the Weibull's scale parameter λ is done using the expected value of the normal distribution μ_N by assuming that they are inverse correlated. The approximation of the Weibull's shape parameter k is done using the standard deviation of the normal distribution parameters by assuming that σ_N is inverse correlated to the multiplication of both Weibull's parameters. Overall, these approximations have the following description

$$\mu_V \approx \frac{1}{\lambda} \Longrightarrow \lambda : \mu_N = \frac{1}{\lambda}, \tag{37}$$

$$\sigma_V \approx \frac{1}{k\lambda} \Longrightarrow k : \sigma_N = \frac{1}{k\lambda}. \tag{38}$$

The visualization of the wind speed general transformation (36) is represented in Figure 10. In this figure, different possible specific transformations are represented mapping a normal distributed domain into wind speed values. The general transformation takes different shapes because it is a function of both the mean and the standard deviation of the normal distributed series that is trying to transform. An example for what the data sets could look like both before and after the transformation is shown in Figure 11. In (a), the normal distributed signal χ_V is represented along with its normal probability density function. In (b), there is represented the histogram for the synthetic wind speed data after the transformation (36) along with its Weibull probability density function (16). Despite some fluctuations that appear in the overall example, the Weibull distribution still offers an accurate description of the wind speed shape.

The third step for wind speed concludes the methodology for analytically modeling the fundamental behaviors of these weather phenomena. At this point, an unlimited long

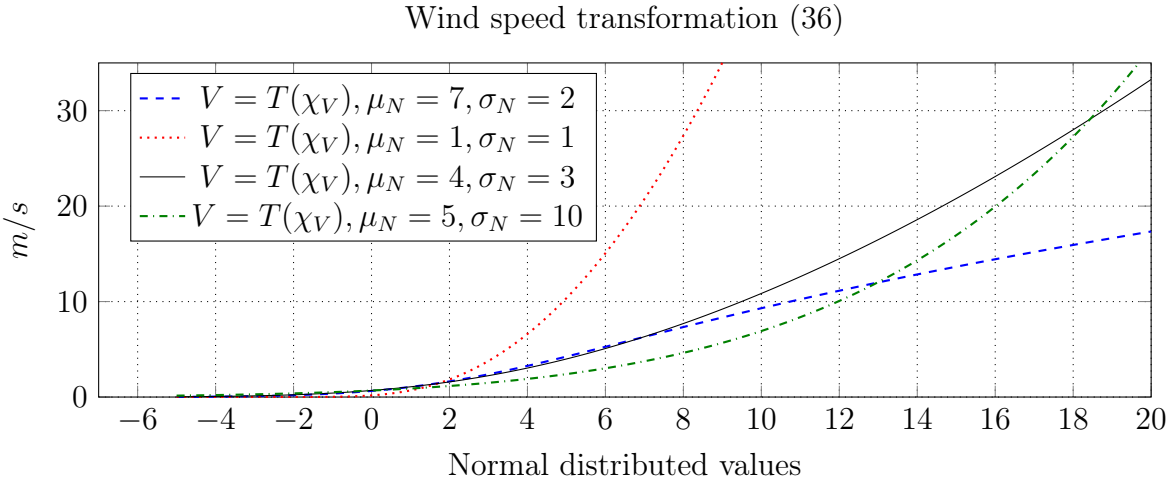


Figure 10: Visualization of the transformation applied to map a normal distributed time series into the natural distribution of wind speed with the different values of μ_N and σ_N.

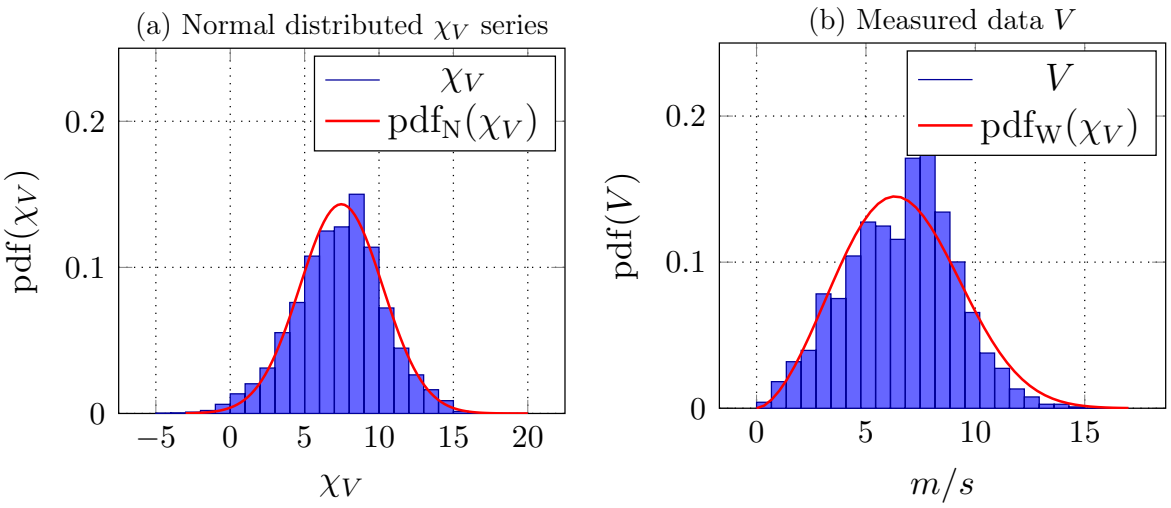


Figure 11: Histograms for wind speed with Weibull distribution and for the normal distributed time series χ_V.

time series for solar irradiation, air temperature and wind speed is able to be generated given that the necessary parameters were determined.

3.3 Analysis

The analysis section covers a complementary concept for the dynamic model. Throughout the synthesis, a series of parameters are used to describe the different behaviors of weather phenomena. This analysis concept focuses on the approach to determine the necessary parameters for the synthesis. The requirement for the analysis to be continuous does not exist, therefore it is understood as a series of numerical optimization methods. It is also not part of the dynamic model because it does not aim to describe the behaviors of the weather, merely to use a sudo-reverse methodology as the synthesis to optimize the parameters. This reverse methodology consists in applying step 3, step 2 and step 1 from the synthesis but instead of introducing weather behaviors, the analysis parametrizes them and proceeds to remove them from the input data.

First and foremost, the input data are provided and read by the model. This set of data has several requirements, it needs to provide over a period of minimum eight year continuous information about GHI, air temperature and wind speed and it requires the values to be provided in the appropriate unit of $W/m^2, C, m/s$ respectively. It is also necessary that the time zone corresponds to the cost function (40) used in the model which is in Coordinated Universal Time (UTC) and it corresponds to a GHI daily peak between 12pm and 13pm. These requirements lead to the first assumption of the implementation.

Assumption 7. The input data set is provided in the correct format and without discontinuities. Therefore the model does not possess tools to manipulate input data.

The cost function in this model has the objective of, when modeling and minimizing the error, disregarding night time values of the GHI, which are near absolute 0 W/m^2 and therefore considered zero. The determination of day time and night time is represented by [10]

$$\Psi = \sin(\psi)\sin(\delta) + \cos(\psi)\cos(\delta)\cos(h) \tag{39}$$

where Ψ is the ratio between the maximum possible GHI for a given time ($I_{max}(t)$) and the maximum possible GHI value ($E_g = E_0 \cdot \rho$). The latitude of any given site in radians is ψ, the solar declination is $\delta = 0.409 \cdot \sin\big(\omega_y(t - 81d)\big)$ and the hour angle is given by $h = (\omega_d t - \pi)$ radians. In this equation, day time values are represented by positive values and night time values are represented by negative values. To properly implement the cost function, positive day time values need to be converted to the value of '1' and negative night time values need to be converted to the value of '0'. The conversion from a sinusoidal wave to an 'rectangular' wave is achieved by

$$w_I = \frac{1}{2} + \frac{1}{2}\left(\frac{\Psi}{\sqrt{\Psi^2 + \epsilon}}\right) \tag{40}$$

where $\epsilon \to 0^+$. In this equation, Ψ is divided by its equivalent positive value and therefore has a ratio of '1' for day time values and '-1' for night time values. The introduction of ϵ (an infinitesimal positive value) allows for the ratio to be defined for the cases of $\Psi = 0$ while only introducing an infinitesimal error. Finally, by summing $\frac{1}{2}$ and $\frac{1}{2}$ multiplied by the ratio, a wave similar to a rectangular wave between the values of '0' and '1' is obtained.

A draw back from a rectangular shaped cost function is that during the optimization processes errors which are only noticeable after the transformation (step 3) can be generated. In a normal distribution, the mean value divides the probability density function in exactly two symmetric areas. The same does not happen for a Weibull distribution or for the transformation applied to the solar irradiation (33). The result is that when mapping an interval from a normal distribution to one of these distributions, its domain changes. Taking an interval from the Weibull distribution as an example, if its domain consists of values lower than the expected value, then it is mapped in a larger domain in the normal distribution. If its domain consists in values greater than the expected value, then it is mapped in a smaller domain in the normal distribution. To avoid such potential errors, the transformations from step 3 can be differentiated and used as the cost function,

$$w_n = \frac{\partial Tr_n(\chi_n)}{\partial \chi_n}. \tag{41}$$

3.3.1 Step 1 : Transformation of input weather data into normally distributed time series

Irradiation

The GHI data time series needs to be transformed, similar to the base concept, into a normally distributed time series. More specifically, values of GHI between zero and the

maximum possible irradiance ($I' \in [0, E_g]$) are mapped into values of $\chi'_I \in \mathbb{R}$. Two issues are presented in the previous GHI transformation (10). Firstly, the transformation is done with a discontinuous function and because of it, any value of χ'_I during night time is transformed into '0' regardless of its otherwise value. The transformation implemented in the model solves this issue but presents an obstacle for its analytical inversion $\chi'_I = Tr_I^{-1}(I')$.

The inverse transformation of (33) is not trivial. There are three possible ways to apply the inverse transformation function. The first method is to create a lookup table which, depending on the time-step necessary between each input, is either the fastest or slowest method for implementation. The second method is to use an approximation expression through $e^x = \sum_{k=1}^{\infty}(x^k/k!)$. However, this method requires a considerable length before it yields acceptable errors for the domain of χ'_I. The third method is to analytically invert the transformation function. Here it is required the use of the Lambert W function (19) described in chapter 2. There is a recurring attempt throughout the analysis concept to maintain analytical expressions as much as possible. Therefore, the third method is applied and the inverse transformation is given by

$$\chi_{I'} = W\left(\frac{E_g}{I'} - 1\right) - \ln\left(\frac{E_g}{I'} - 1\right) \tag{42}$$

where W represent the Lambert W function.

Temperature

Air temperature is reasonably well described by a normal distribution as previously presented. Unlike the other two RER, the air temperature does not require an inverse transformation in order to be parametrized by the same procedure. While working with normal distributions, all magnitudes and amplitudes are dimensionless, so only the unit °C is removed from the measured time series I'.

Wind speed

Wind speed is a RER which is reasonably well described by a Weibull distribution with probability density function described by (16). The input time series for the wind speed has naturally only real positive values, $V' \in \mathbb{R}_0^+$. The input data requires a transformation to a normally distributed time series, resulting in the time variant wind speed noise signal χ'_V. The wind speed input data is transformed into a normal distributed time series by applying the inverse function of (36). The inversion of this function is not trivial and for implementation purposes two different methods are considered. The first is to create a lookup table for the transformation. The second method is to solve (36) analytically for χ_V. The latter is the choice used in this work since it is the most coherent with the overall concept. The inverse transformation takes the following form

$$\chi'_V = \frac{1}{\lambda} + \frac{\sqrt{2}}{\lambda k} \cdot \mathrm{erf}^{-1}\left(1 - 2 \cdot \exp\Big(-(\lambda V')^k\Big)\right). \tag{43}$$

3.3.2 Step 2: Parameterizing time dependency

The second step in the analysis concept consists of parameterizing the mean (μ_n) and the standard deviation (σ_n) for each of the normal distributed time series (χ_n). As previously mentioned, these two behaviors are responsible for the time dependency of the time series. In the base concept, this step is responsible for the vast majority of parameters in the concept, since each mean and standard deviation requires 9 parameters, amounting to a total of 54. It is then of most importance to drop any over fitting parameters and only maintain the absolute minimum amount that allows a good description of the time dependent signals χ_n. Both of these behaviors require different approaches to determine the parameters and are in this chapter presented, first for the mean and afterwards for the standard deviation.

Mean

The mean is what defines the overall shape of the noise signal χ_n, therefore its fitting demands high finesse. After the inverse transformations of the weather data $Tr_n^{-1}(n'[t])$, $\chi_{n'}[t]$ still have the same length as the input data, which is at least 8 years. $\chi_{n'}[t]$ is a periodic signal with the longest period being 1 year, therefore every value of $\chi_{n'}[t]$ within a 1 year period is averaged across the number of year in the input data. The result $\hat{\mu}_{n'}$ is defined as

$$\hat{\mu}_n[t] = \frac{1}{\Delta T} \sum_{i=0}^{\Delta T-1} \chi_{n'}[t + i \cdot [1]y]. \tag{44}$$

The measured mean behavior of the weather $\hat{\mu}_n$ needs to be fitted the designed equation (23). 'Fitting' the equation means that there is a numerical optimization method that is applied in order to determine the parameters $A_0, A_d, A_y, A_s, \phi_d$, and ϕ_y that minimize the error to $\hat{\mu}_n$. In this concept, the general mathematical representation for this fitting procedure is

$$p_{\mu_n} = \arg_{p'_{\mu_n}} \min \left\{ \sum_t w_n(t) \cdot \left(\hat{\mu}_n[t] - \mu_n\big(t; p'_{\mu_n}\big)\right)^2 \right\}. \tag{45}$$

This numerical optimization method is similar to the Levenberg-Marquardt algorithm which is also used for solving non-linear least squares problems. The only considerable variation of (45) for the different RER is the cost function $w_I(t)$ applied. The cost function applied for the solar irradiation is (40) to ignore night time values. The air temperature and wind speed need, however, to have the entirety of their measured mean ($\hat{\mu}_{T,V}$) considered for the parameters to be properly determined. Therefore, the cost function is considered to be an ineffective constant such as $w_{T,V} = 1,\ \forall t$.

Once this procedure is finished, all the 18 parameters necessary to model the different mean behaviors (23) are determined. The following procedure consists in determining the parameters for the standard deviation.

Standard Deviation

The standard deviation defines how much a particular value can stray from its expected value. As the mean, it also exhibits a time variant behavior throughout the day and seasons. To parametrize such behavior, the standard deviation from the measured data is determined first. The result is $\hat{\sigma}_{n'}$ and it is defined as

$$\hat{\sigma}_n[t] = \sqrt{\frac{\sum_{i=0}^{\Delta T-1}(\chi_{n'}[t + i \cdot [1]y] - \hat{\mu}_n[t])^2}{\Delta T - 1}}. \tag{46}$$

The approach for modeling the standard deviations of each RER has suffered considerable transformations from the base model. Instead of applying the same numerical optimization method (45) as the mean, only the average value of $\hat{\sigma}_n[t]$ is determined. As previously mentioned in step 2 of the model, relationships between the different parameters of μ_n are established to describe the diurnal, seasonal and seasonal variation of diurnal variation components. However, the relationships are merely ratios, they do not represent the base amplitude that describes these behaviors. The base amplitude is considered to be the numerically determined average value of $\bar{\sigma}_n$. This calculation is defined as

$$\bar{\sigma_n} = \frac{1}{N}\sum_{t=1}^{N}\hat{\sigma}_n[t]. \tag{47}$$

The parameters used to create the correlation between the mean and the standard deviation for the different time variant components are chosen empirically. Typically, each amplitude A_0, A_d, A_y, A_s has a different range of dimensionless values which they can take. The same is true for the standard deviation. So ratios between these amplitudes are created and multiplied by the base amplitude $\left(\frac{|A_i|}{|A_j|} \cdot \bar{\sigma_n}\right)$ to obtain the typical range of values for the standard deviation, described in (27), (29) and (31).

At this point, the time dependent behaviors are parametrized and account for 21 of the total amount of parameters. The following step is to determine the specific rate of change of each RER and the correlation between them.

3.3.3 Step 3: Parametrize the rate of change and correlation between RER

The third and final step of the analysis concept consists in determining the time constant T_n and the correlation constant ζ_n. Before determining these parameters, it is necessary to remove the time dependency from the the signals $\chi_{n'}$ and obtain a standard normal distributed signal $\chi''_{n'}$. Removing these behaviors from $\chi_{n'}$ is done through

$$\chi''_{n'}[t] = \frac{\chi_{n'}[t] - \mu_n(t)}{\sigma_n(t)}. \tag{48}$$

The resulting signal $\chi''_{n'}$ is now only time- and cross-correlated. The time constant is determined for the dynamic equation (20) in the case of solar irradiation and for (22) in the case of air temperature and wind speed. The correlation parameter is only

determined for (22) and its correspondent RER. To determine T_n, first a spectrum analysis is done based on a discrete fast Fourier transform (FFT). This FFT obtains the t frequency response $A_{n'}(\omega)$. The frequency range of interest for the modeling of the rate of change is $(\log_{10}(\omega) \in [-0.7, -1.0])$ which is equivalent to an hour range of $T \in [31.49, 62, 83]$. This range is deemed to have a sufficient number of samples for the fitting of the time constant parameter T_n. This parameter is assumed to be greater than 20 hours ($T_n > 20[hr]$) and the frequency response is expected to obtain an affine relation between $\log_{10}(\omega)$ and the frequency response

$$\log_{10}(A_n^{'}(\omega)) \approx -\log_{10}(\omega) + b. \tag{49}$$

The cut off frequency of (20) applied to the approximation of (49) leads to $\sqrt{\frac{2T_n}{T_s}} \stackrel{!}{=} -\log_{10}(T_n^{-1}) + b$. Once again, by applying (49) it is possible to compute (50).

$$b = \text{mean}(\log_{10}(A_n^{'}(\omega)) + \log_{10}(\omega)) \tag{50}$$

Finally, the time constant is fitted by numerically solving

$$T_n = 10^{-\omega_T} : \sqrt{\frac{2 \cdot 10^{-\omega_T}}{T_s}} + \log_{10}(10^{\omega_T}) - b = 0. \tag{51}$$

The visualization of this process is represented in Figure 12. In this figure the majority of the frequency response $A_{n'}(\omega)$ from the FFT is plotted along with equation 51. The objective is to minimize the error through which 51 fits the values of within interval limited by dashed lines. This correspondent hour value is presented for clarification. Any value higher than 63 hours ($\log_{10}(\omega) < -1$) is deemed to be too slow for an accurate parametrization of T_n. The reverse is applied for values lower than 31 hours ($\log_{10}(\omega) > -0.7$) which are deemed to be too fast for weather phenomena in general.

The last parameter to be determined is the correlation ζ_n between each RER. The methodology for its determination remains unchanged from the base concept but for coherence purposes, it is described here. For the air temperature and the wind speed, the correlation parameter analyzed is $\zeta_n \; \forall \; n \in (T, V)$. The frequency spectrum is shifted from the comparative series $n-1 \; \in (I, T)$ to the frequency spectrum of n. The specific frequency spectrums results form the filter part of the dynamics, which can be expressed by the transfer

$$G_n = \frac{\sqrt{\frac{2T_n}{T_s}}}{1 + T_n s}. \tag{52}$$

The frequency spectrum of the comparative series $n-1$ is implemented by a combination of the transfer functions G_n and G_{n-1}, resulting in the shifted comparative series

$$\begin{aligned} \chi^*_{n'-1}[t] &= \frac{G_n}{G_{n-1}} \chi''_{n'-1}[t] \\ \chi''_{n'-1}[t] &= \frac{\chi_{n'-1}[t] - \hat{\mu}_{n-1}[t]}{\hat{\sigma}_{n-1}[t]}. \end{aligned} \tag{53}$$

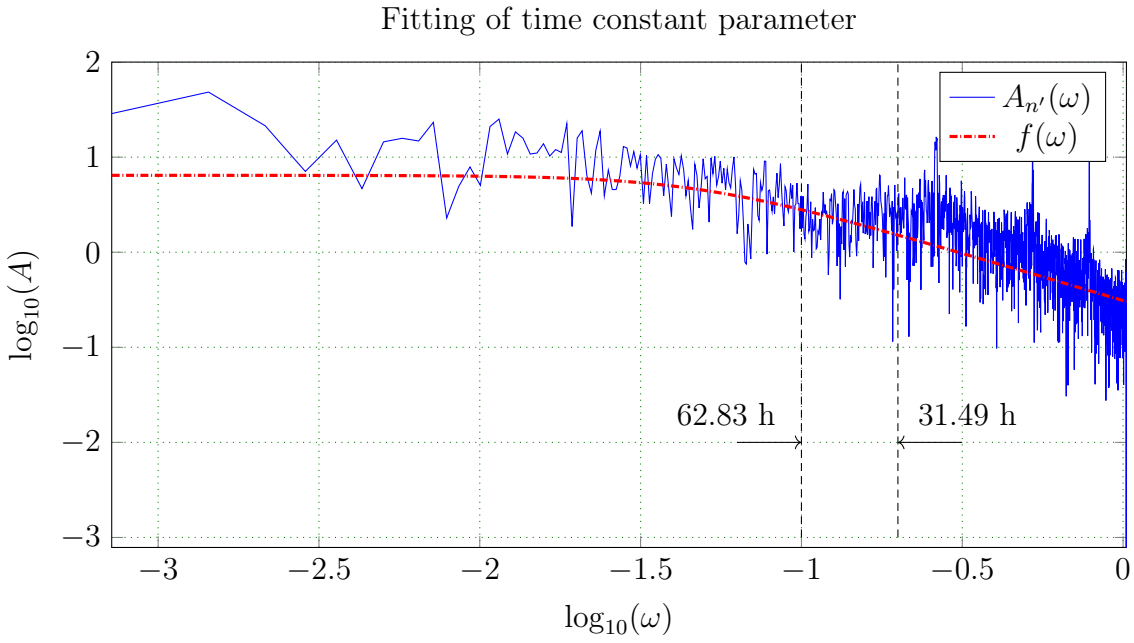


Figure 12: Example of equation (51) fitting rate of change of a signal.

The correlation $\tilde{\mathrm{E}}$ weights each time sample by its influence w_n on the final result n by

$$\tilde{\mathrm{E}}(\chi''_{n'}[t], \chi^*_{n'-1}[t]) = \frac{\sum_t w_n \cdot \chi''_{n'}[t] \cdot \chi^{*''}_{n'-1}[t]}{\sum_t w_n} \tag{54}$$

with $\chi''_{n'}[t] = \frac{\chi_{n'}[t]-\hat{\mu}_n[t]}{\hat{\sigma}_n[t]}$. The value of the weight function is, however, only changing for the solar irradiation according to (40). For the air temperature and wind speed all points are considered with the same weight of '1'. The correlation parameter results according to [3] in

$$\zeta_n = \frac{\tilde{\mathrm{E}}(\chi''_{n'}[t], \chi^*_{n'-1}[t])}{1 - \tilde{\mathrm{E}}(\chi''_{n'}[t], \chi^*_{n'-1}[t])}. \tag{55}$$

At this point, all parameters necessary for the dynamic model are determined and the analysis concept is concluded. In Figure 13, the flow chart for the analysis concept is presented. Every single transition described in this chapter is pointed out for each RER along with their respective input and output. The parameter set is saved at the end of the analysis and can be used in the synthesis to generate unlimited variations of the weather with the parametrized same behaviors.

3.3.4 Available data

The data sets used in the development of the model consisted of satellite data from the MERRA database of NASA [15]. The precise data extracted for all the sites are the GHI, air temperature and wind speed at 50 meters between a period of 2005-2012. The

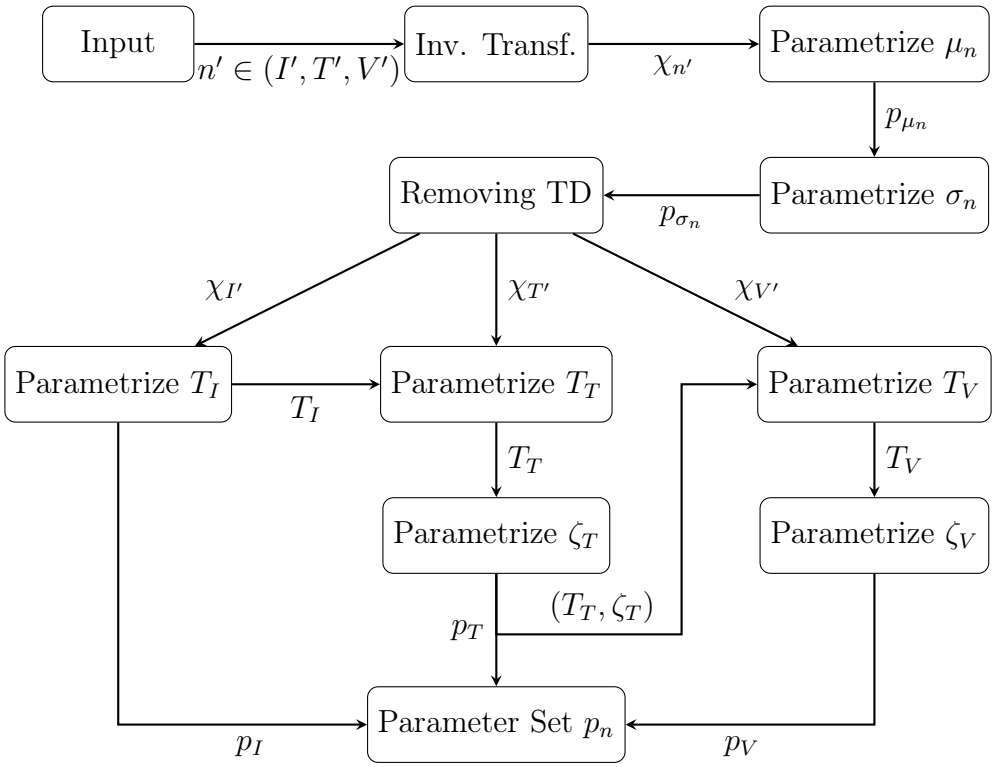


Figure 13: Flow chart of the analysis concept.

data set has a time step of 1 h, which results in 70128 total amount of points. The choice of site consisted in parameters, one is the latitude of the site and other is their climate zone [16]. Only the latitude is used to determine daytime hours and peaks in GHI. The optimization of the model is done with data which is from both the most largest climate zones and spread out from south to north hemisphere. The coordinates of a location determine the center of data images from the MERRA data base. The images are generated from satellite data and therefore have a standardized size of 0.625°x 0.5°. The list of sites and their respective coordinates used in the first and second stage of development of the model are presented in the following table

3.3.5 Determination of errors

The determination of errors is crucial for an accurate evaluation of the model. It is the deciding factor when assessing whether the model is producing reasonable results or whether it requires further iterations in development. The numerous stages that the model goes through before producing an output are measured for accuracy in order to better identifying possible unexpected abnormalities in the output. In order to obtain the errors of model as accurate as possible, a set of 8 years of synthetic and measured data is used. The relative shape error measure e_{pdf} indicates deviations of the histogram of a measured time series $n'(t)$ and the synthetic time series $n(t)$. Both time series are

Table 2: Discretization of all the sites and coordinates for the center of their respective data sets used in the development of the model.

List of sites			
X Site	Coordinates (Lat.,Long.)	Y Sites	Coordinates (Lat.,Long.)
Fortaleza, Brazil	(-3,76°, -38,54°)	Perth, Australia	(-31,92°, 115,8°)
Namibe, Angola	(-15,19°, 12,16°)	Tehran, Iran	(35,68°, 51,38°)
Seoul, South Korea	(37,54°, 126,9°)	Beijing, China	(39,9°, 116,4°)
Galicia, Spain	(42,57°, -8,13°)	Madrid, Spain	(18,56°, -68,39°)
Toronto, Canada	(43,65°, -79,38°)	Bogota, Colombia	(4,62°, -74,06°)
Punta cana, Dominique Rep.	(18,56°, -68,39°)	Capetown, South Africa	(-33,92°, 18,42°)

sorted by their values resulting in $n'(i)$ and $n(i)$ with i indicating the position of the value. The relative shape error is defined as

$$e_{\mathrm{pdf_n}} = \sqrt{\frac{\left(\sum_i n'[i] - \sum_i n[i]\right)^2}{\sum_i n'^2[i]}}, \tag{56}$$

indicating deviations of the absolute values as a percentage error value. This error is calculated for 10 independent simulations in order to obtain its confidence interval. Its representation takes the form of $e_{\mathrm{pdf_n}} = A \pm B\%$, where 'A' is the average value of the relative shape error for the 10 simulations and 'B' represents the boundaries of the confidence interval. The relative mean and standard deviation error measure e_μ and e_σ indicate deviations of the overall mean and standard deviation. These measures are defined as

$$e_{\mu_n} = \frac{E(n[t]) - E(n'[t])}{E(n'[t])}, \tag{57}$$

$$e_{\sigma_n} = \frac{\sum_t \left(x[t] - \frac{\sum_t x[t]}{\sum_t 1}\right)^2 - \sum_t \left(\hat{x}[t] - \frac{\sum_t \hat{x}[t]}{\sum_t 1}\right)^2}{\sum_t \left(\hat{x}[t] - \frac{\sum_t \hat{x}[t]}{\sum_t 1}\right)^2}, \tag{58}$$

where $E[n(t)] = \frac{1}{N}\sum_t^N n(t)$. These errors, unlike the relative shape error, can have a negative value. The interpretation of their signal is that a negative relative error indicates an underestimation of the behavior when compared to the measured data. A positive relative error represents an overestimation of these behaviors instead.

The two remaining modeled behaviors of the weather data are the rate of change and correlation between RER. The relative error for these two behaviors is determined by

$$e_{T_n} = \frac{R\big(n[t], n[t-1]\big) - R\big(n'[t], n'[t-1]\big)}{R\big(n'[t], n'[t-1]\big)}, \tag{59}$$

$$e_{\zeta_n} = \frac{R\big(n[t], n_{-1}[t-1]\big) - R\big(n'[t], n'_{-1}[t-1]\big)}{R\big(n'_{-1}[t], n'[t-1]\big)}, \tag{60}$$

where $R[s_1(t), s_2(t)] = E[s_1(t) \cdot s_2(t)]$ and n_{-1} represents the previously determined RER to which n is correlated, for example $(n = T, n_{-1} = I)$.

The quality of the output synthetic weather data is an amalgamation of all the errors presented, however, the relative shape error is the most general description for the synthetic data as a whole. Therefore it is used as the highest weighting factor when determining the quality. Since the model is a less complex version of its original base model and explores the implementation of new assumptions for each RER, the output data are considered to have 'good' quality if the relative shape error is inferior to 5%. If the synthetic data have a relative shape error greater than 10% then it is considered 'poor' quality data. The data with a relative shape error between these two intervals is considered 'reasonable' quality data.

4 Results

This chapter illustrates the final results obtained for three sites described in Table 2. Each location is analyzed individually and presented as a case study. Afterwards, a general assessment is done to the weather model with all the information extracted from these data sets. The general approach to present each site of study starts by a brief introduction to the location, followed by the discretization of the parameters and errors in the analysis part of the model. Afterwards, the synthesis part of the model is presented with special detail and finally, a discussion is conducted to express how well the weather model works.

4.1 Perth, Australia

The city of Perth in Australia is one of the most southern sites in the series of data sets evaluated in the results chapter. In this work, it is used as the most representative point for a site in the south hemisphere that is not close to the equator, which is considered to be within a range of [-20°; 20°]. Perth is located in the western coastal line of Australia and due to the resolution of the satellite MERRA data there is a considerable amount of sea within it. This characteristic of the data may distort in a small extent the average values for Perth but it also allows for the assessment of the impact that presence of sea has in the analysis and synthetic weather data.

4.1.1 Analysis

The inputs given to the weather model are the latitude of the site (-31°57' North) and its data set composed by the GHI, air temperature and wind speed at 50 meters altitude. The time zone in Perth is leading Coordinated Universal Time (UTC) by 8 hours, therefore, in order to provide the data set with the correct time zone, it needs to be shifted by 8 hours. Following the presented methodology, the first step applied is the transformation of all the RER to normal distributed time series through (33) and (43) respectively. From this normal distributed measured data, the mean and standard deviation over a year long period are determined. The mean and standard deviation are the data's time dependency information and are modeled through (23), (27), (29) and (31) respectively. Afterwards, the rate of change and correlation parameters are determined through (51) and (55) resulting in the parameter Table 3. These are the parameters which are saved and will serve as the only necessary external input for the synthesis part of the model.

4.1.2 Synthesis

The synthesis algorithm starts by loading the parameter set in Table 3 and by creating the random standard normal distributed time series χ'_n (2). Afterwards, the rate of change, correlation and time dependency information are introduced. Finally, the respective transformation for each RER is applied and the output of synthetic weather

Table 3: Discretization of all numerically determined parameters for Perth's data set.

Table of numerically determined parameters			
Parameter	GHI	Air Temperature	Wind Speed
$A_{0_n}[-]$	-1.1267	19.8459	7.4655
$A_{d_n}[-]$	-2.2756	5.6154	0.9039
$\phi_{d_n}[rad]$	0.0846	-2.5532	-0.0242
$A_{y_n}[-]$	0.7002	6.7125	0.5051
$\phi_{y_n}[rad]$	0.0013	0.4805	-0.0373
$A_{s_n}[-]$	-0.0291	2.6185	0.4415
$\bar{\sigma_n}[-]$	0.1684	3.1904	2.5625
$T_n[hr]$	8.5007	30.1639	197.2352
$\zeta_n[-]$	-	0.0533	-0.4606

data is obtained. The following subsections present the synthetic weather data in both a macro- and micro-perspective, *i.e.*, from a seasonal scale and from a daily scale. A comparison between the expected result and actual result is conducted mainly from a visual approach. Lastly, the errors of the output weather data are presented and discussed.

Irradiation

The output GHI data (I) from both a macro- and micro-perspective is presented in Figure 14, (a) and (b) respectively. For comparison purposes, the Figure 14 also includes measured data (I') and the theoretical maximum GHI at any given time ($\Psi \cdot E_g$). Representing every single synthetic and measured point does not allow for an appropriate visualization because the density of points is too high for a plot of this size. To visualize the seasonal behavior of the GHI (a), only the maximums of each day are represented. The minimums are neglected in (a) because their value is zero for each day (night time). The first important characteristic that is observed in (a) is that the seasonal fluctuation occurs in phase with the measured data, which is an indicator that the parameter ϕ_{y_I} is modeled correctly. The difference between the average synthetic maximum in the summer and in the winter is also remarkably similar to the one of the measured data, which indicated that the parameter A_{y_I} is correctly modeled. Also noticeable is that the synthetic maximums always remain below the theoretical maximum GHI curve. The behavior that distinguishes the synthetic data from the measured data is for the sporadic occurrences of maximums, which are much lower than the expected average. This inability to reproduce extreme variances is mainly related to the standard deviation (σ_I) and the time constant (T_I) which, for the majority of the time period, does fit the measured data.

To visualize the daily behavior of the synthetic air temperature, a plot of every single point generated over a time period of a week is deemed appropriate (b). The phase for the daily behavior of GHI is the same for both curves, their peaks occur with little

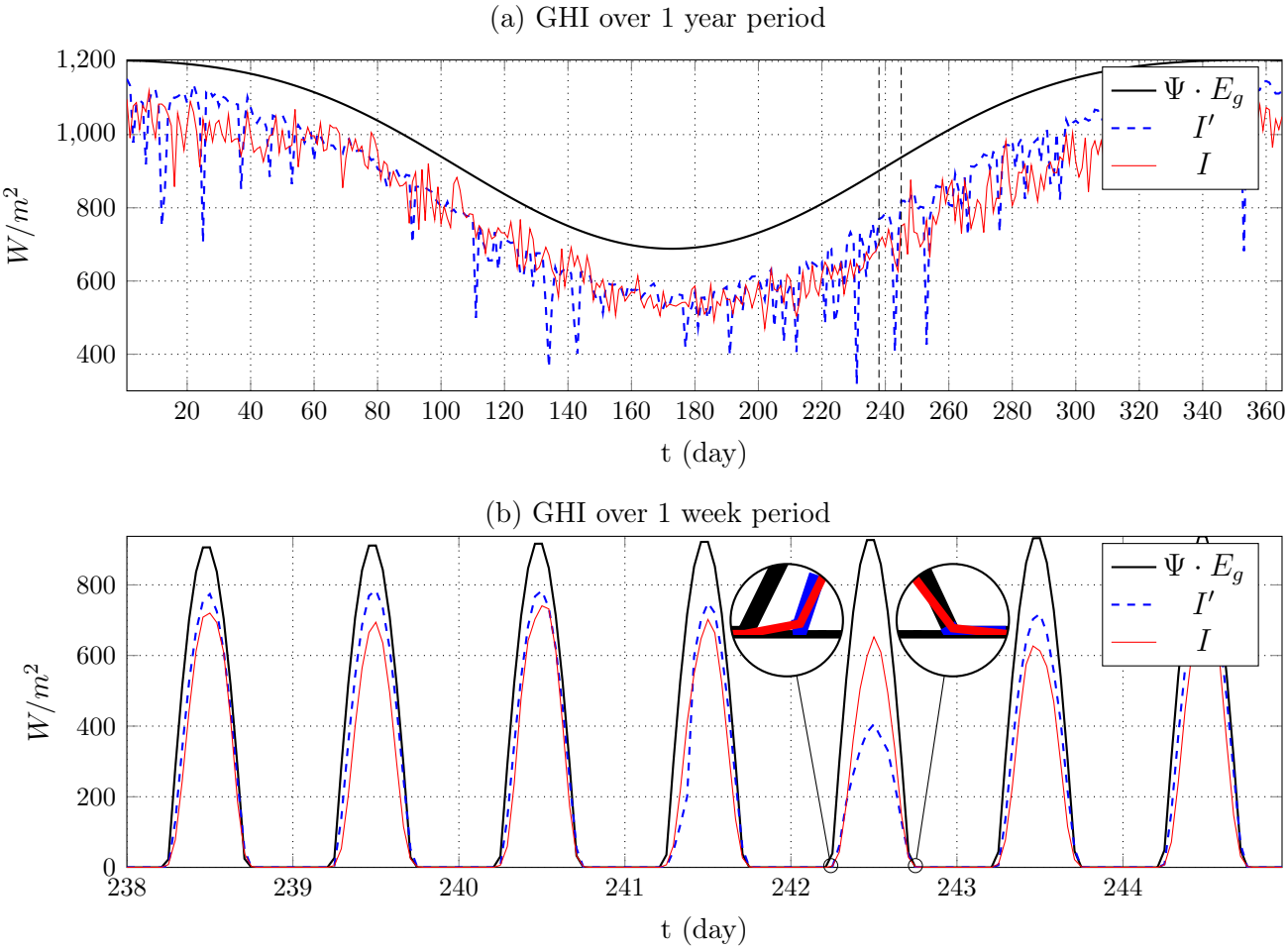


Figure 14: Synthetic and measured weather for Perth's data set discretized by day (a) over a year long period with its respective maximum values, and (b) over a week long period with all of its respective values.

to no lag between each other. Here it is now possible to visualize another important characteristic of the daily phase, the ramping up and down of GHI at sunrise and sunset, which is magnified in (b). The synthetic data match the behavior of the measured data at these two specific time steps almost perfectly. This particular behavior assures that the parameter for daily phase ϕ_{d_I} is modeled correctly and it shows the importance of transformation (33) for not mapping the values of I' symmetrically to the vertical axis ($\chi_{I'} = 0$). Nevertheless, it is not possible to fully judge the modeling of the daily amplitude parameter A_{d_I} from only a week long period. However, subfigure (b) still provides a quick way to confirm that there are no fundamental problems with its modeling. The time constant parameter T_I is also adequately modeled for the daily behavior of GHI since there are no signs of noise, *i.e.*, unexpected and unreasonable changes in value of the synthetic GHI.

Air Temperature

The output air temperature data T and the measure air temperature data T' are presented in Figure 15 from both a macro- and micro-perspective, (a) and (b) respectively. In (a), the maximums and minimums for each day are represented. The first important observable characteristic in (a) is that the seasonal fluctuation occurs in phase with the measured data, which is an indicator that the parameter ϕ_{y_T} is modeled correctly. The difference between the average synthetic maximum in the summer and in the winter are also similar to the one of the measured data, which indicates that the seasonal amplitude parameter A_{y_T} is also correctly modeled. The amplitude for the daily behavior A_{d_T} is assessed by comparing the difference between the maximum and minimum values of each day. Sub-figure (a) provides a general good estimation for the quality of A_{d_T} and it is further corroborated by a more detailed view represented in (b). During winter time, more specifically between the period of $t \in [180, 230]$, there are occurrences of sporadic

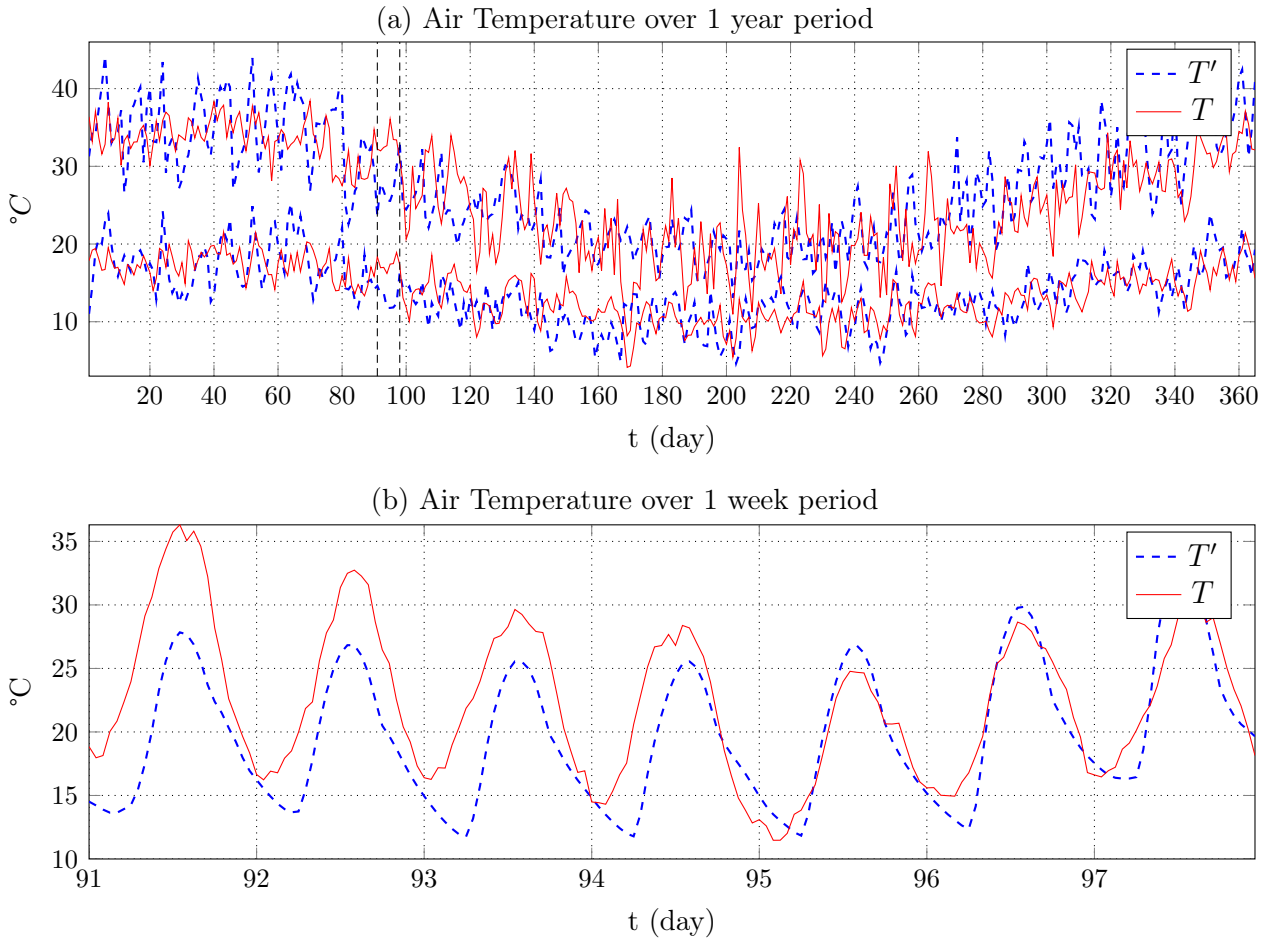


Figure 15: Synthetic and measured air temperature for Perth's data set discretized by day (a) over a year long period with its respective maximum and minimum values, and (b) over a week long period with all of its respective values.

'spikes' in the maximum values of the day for the synthetic data which are not observed in the measured data. The standard deviation is responsible for the interval from which the synthetic data can vary, therefore the modeled standard deviation $\sigma_T(t)$ has greater values during winter time than the corresponding measured standard deviation σ'_T. To visualize the daily behavior of the synthetic air temperature, a plot of every single point generated over a time period of a week is deemed appropriate (b). The phase for the daily behavior of the air temperature is the same for both curves, their peaks occur with little to no lag between each other which leads to the preliminary conclusion that the daily phase parameter ϕ_{d_T} is modeled correctly. The influence of the time constant parameter T_T is also observable in (b). The fact that there is no accentuated noise in the synthetic data indicates that the T_T is modeled reasonably well. The small variances that appear during a day long period, specially during peak hours when the standard deviation is at its theoretical maximum value, are possibly smoothened out with a slight increase in the T_T. From a visual perspective, the correlation parameter ζ_T is indistinguishable from the seasonal variations. For some insight on how well it is modeled, its numerical error needs to be determined.

Wind Speed

The output wind speed data V and the measure wind speed data V' are represented in Figure 15 with the same approach and reasoning previously mentioned. The first most noticeable characteristic of the wind speed data in (a) is that it appears to have much more chaotic behavior than either the GHI or the air temperature. It is also particularly difficult to quickly judge if the seasonal parameters A_{y_V} and ϕ_{y_V} are correctly modeled. The main reason for it is that there are no outstanding differences in the measured wind speed data between seasons. However both data sets fluctuate throughout the entire year around an average value $(\bar{\mu}'_V, \bar{\mu}_V)$ which is remarkably similar. An important difference presented in (a) is the magnitude of maximum values in both data sets. The synthetic data shows several short spikes in value that are greater than any maximum in the measured data. This occurrence is noticeable during the summer and it is an indicator that the modeled standard deviation $\sigma_T(t)$ is overestimated for this season. Due to the reduced amount of information that can be extracted from (a), the relevance of a detail view (b) is even higher than for other RER. The most evident behavior that the synthetic wind speed shows in (b) is the noisy pattern, *i.e.*, very frequent small variances in the signal that are not as noticeable in the measured wind speed data. This behavior is mainly attributed to the time constant (T_V) and the transformation to Weibull distributed data which reflects either a poor modeling of its behavior or an inadequate mapping of the normal distributed values χ_V. Despite the noisy behavior, it is still possible to identify that the synthetic data's daily phase parameter ϕ_{d_V} appears to be correctly modeled. Unfortunately, the case is not the same for the daily amplitude parameter A_{d_V}. There are some instances where it appears to have a reasonable value but too often the presence of noisy behavior does not allow for an accurate statement. Similarly to the air temperature, the correlation parameter ζ_V is not addressed from a visual perspective, only by its numerical error.

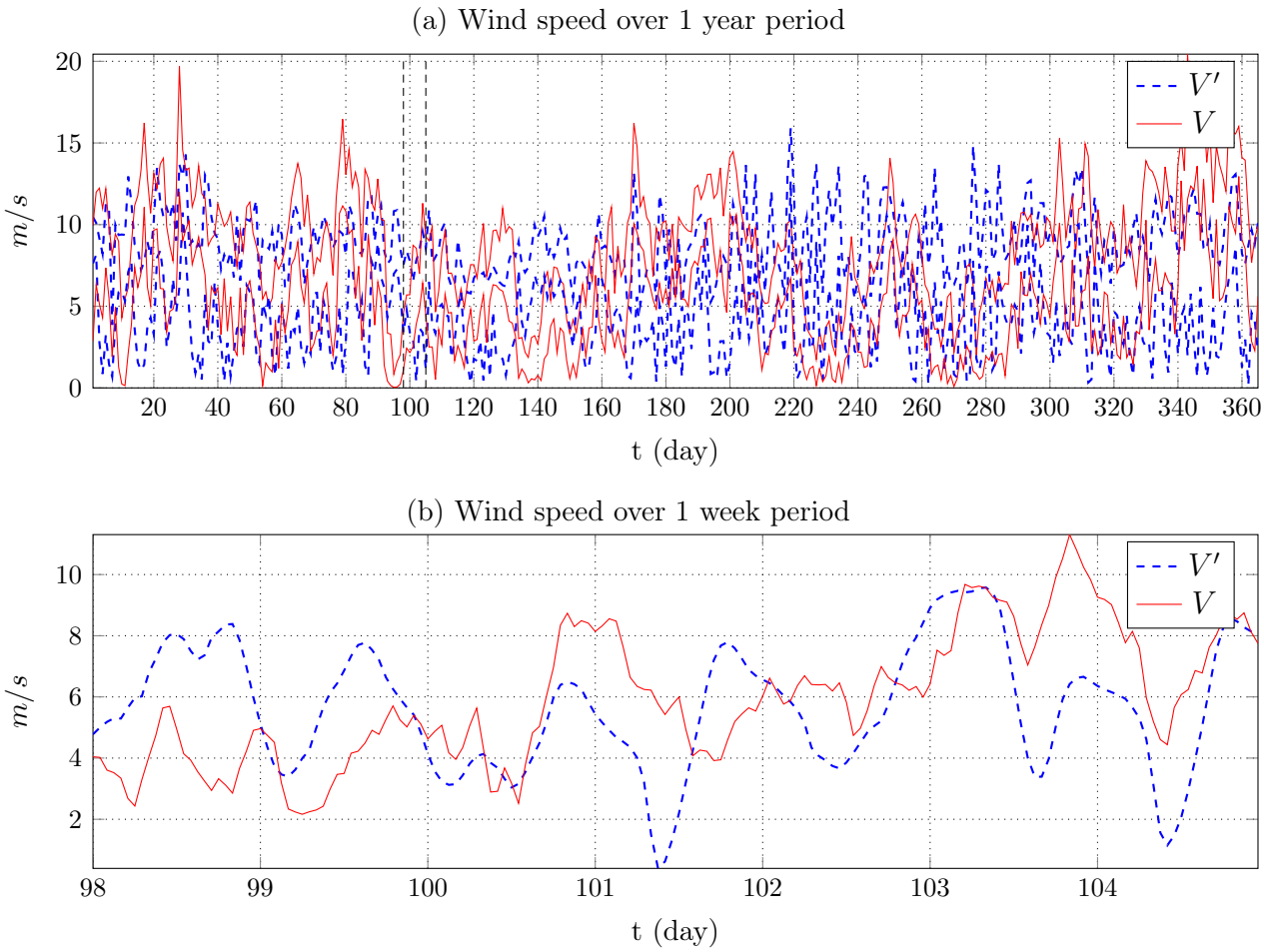


Figure 16: Synthetic and measured wind speed for Perth's data set discretized by day (a) over a year long period with its respective maximum and minimum values, and (b) over a week long period with all of its respective values.

Errors

The visualization of the synthetic weather data allows for a general assessment of the output of the model. The errors relative to the visualized data offer a detailed perspective of their behavior. The first and most relevant error measurement determined is the relative shape error e_{pdf_n}. This error is represented both graphically in Figure 17 and numerically in the overall accuracy Table 4 for all phenomena, in this case study. The GHI has a relative shape error of $e_{\text{pdf}_I} = 3.90 \pm 0.08\%(< 5\%)$ which puts it in the category of good quality data. In Figure 17 (a), the sorted values of i are represented for both synthetic and measured data of GHI. An expected feature of the data is that $I'[i]$ is stiffer in shape than $I[i]$, *i.e.*, the sorted measured data take a smoother and almost linear behavior for positive values of GHI and the synthetic sorted data fluctuate very slightly around this 'line'. Another important feature of the model is represented in the magnified detail of (a). Here is shown that the transition of the measured data

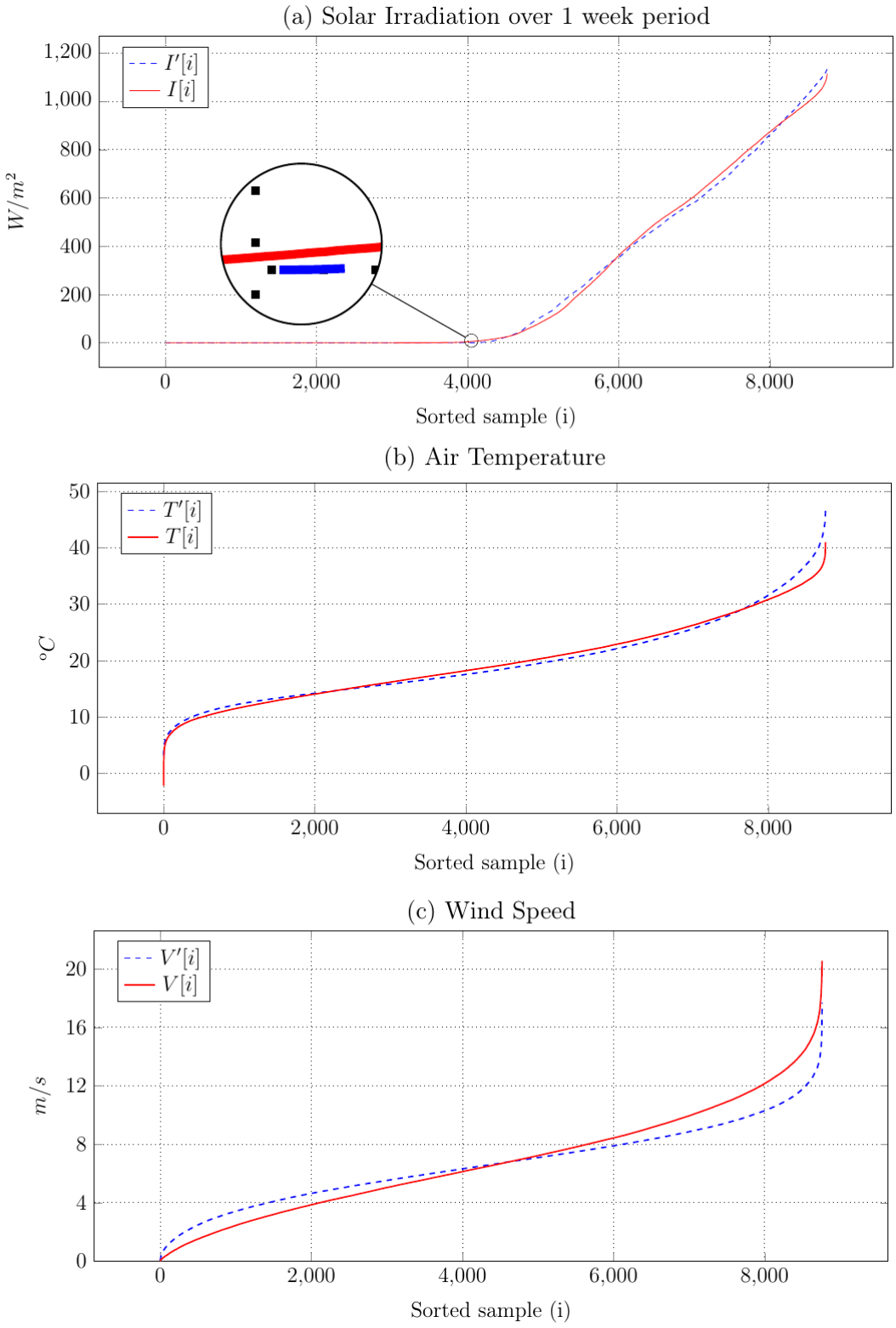


Figure 17: Sorted time series of measured and synthetic weather data for each RER

Table 4: Overall accuracy table for the synthetic RER of Perth's data set.

Overall accuracy table			
Parameter	GHI	Air Temperature	Wind Speed
$e_{\text{pdf}_n}[\%]$	3.90 ± 0.08	4.94 ± 0.14	12.50 ± 1.78
$\bar{\mu}'_n \ [W/m^2, ^\circ C, m/s]$	251.6032	19.8459	6.6545
$\bar{\mu}_n \ [W/m^2, ^\circ C, m/s]$	254.6961	19.9060	6.7561
$e_{\mu_n}[\%]$	0.93	-0.14	3.64
$\bar{\sigma}'_n \ [W/m^2, ^\circ C, m/s]$	337.5043	7.3314	2.6330
$\bar{\sigma}_n \ [W/m^2, ^\circ C, m/s]$	341.8805	6.7560	3.5369
$e_{\sigma_n}[\%]$	1.45	-23.14	49.35
$e_{T_n}[\%]$	1.91	-0.96	19.41
$e_{\zeta_n}[\%]$	–	-1.37	1.96

$I'[i]$ from values approximately $0W/m^2$ to values greater than $1W/m^2$ in an almost discontinuous fashion. The model has a continuous behavior by design and cannot emulate this discontinuous output of data with 1 hour or greater time step. Instead of generating data with sharp angles at sunrise and sunset time it smooths this day/night transitions as much as possible, resulting in the spacing/relative error between the two curves. It is also worth mentioning that this characteristic only amounts to a very small portion of the total relative shape error e_{pdf_I}.

In Table 4, is possible to further observe the numerically determined errors from the model. The expected values of the measured $\bar{\mu}'_I$ and synthetic $\bar{\mu}_I$ data are very close together, with an approximated absolute difference of $3W/m^2$. The relative error of the expected value e_{μ_I} shows that this difference is less than 1% and that it is a negligible overestimation of the model. The standard deviation of the measured $\bar{\sigma}'_I$ and synthetic $\bar{\sigma}_I$ data are notably close to each other, with only an approximated absolute difference of only $4.4W/m^2$. The corresponding relative error is $e_{\sigma_I} = 1.45\%$ which represents an overestimation. Despite the small overestimation, Figure 17 (a) shows that the synthetic data cover the entire range of values in the measure and no substantial difference exists. The time constant parameter T_I is also slightly overestimated by $e_{T_I} = 1.91\%$ which means that the model describes a stronger correlation in time for GHI. The GHI is considered a source RER with no previous correlation and therefore the parameter ζ_I does not exist.

The air temperature has a relative shape error of $e_{\text{pdf}_T} = 4.94 \pm 0.14\% (< 5\%)$ which puts it in the category of good quality data, although borderline. In Figure 17 (b), the overall shapes of both curves are remarkably similar in the majority of the sorted values i. The vast majority of the error obtained is attributed to values in the synthetic air temperature interval $T(i) \in [30, 50]^\circ C$. The model shows some reluctance in generating values for the peaks in temperature. Another interval of interest is the one which the values of measured data are not replicated by the model, specifically $T' \in [42, 48]^\circ C$. This reluctance is traced down to the standard deviation behavior. From (b), is possible

to infer that the peaks of values in the standard deviation signal during summer season are lower than optimal.

In Table 4, it is possible to further observe the numerically determined errors for the air temperature. The expected values of the measured $\bar{\mu}'_T$ and synthetic $\bar{\mu}_T$ data are notably close to each other, with only an absolute difference of approximately 0.06°C. The relative error of the expected value is $e_{\mu_T} = -0.14\%$, which is a negligible underestimation from the model. The standard deviation of the measured $\bar{\sigma}'_T$ and synthetic $\bar{\sigma}_T$ data are not as close together as the expected values. Their absolute difference is approximately $0.6W/m^2$. The corresponding relative error is $e_{\sigma_T} = -23.14\%$ which represents a distinguishable underestimation. This relative error corroborates the interpretation of Figure 17 (b) previously done. The time constant parameter T_T is slightly underestimated by $e_{T_T} = -0.96\%$ which means that the model describes a weaker correlation in time. The correlation between the synthetic GHI and air temperature is represented by ζ_T and it has a relative error of $e_{\zeta_V} = -1.37\%$. This error means that the correlation between the measured GHI and air temperature is slightly stronger than the one in the synthetic data.

The wind speed has a relative shape error of $e_{\mathrm{pdf}_V} = 12.50 \pm 1.78\% (> 5\%)$ which puts it in the category of poor quality data. In Figure 17 (c), it is represented the sorted values i for both synthetic and measured data of wind speed. The overall shapes of both curves are similar, *i.e.*, they still represent the same type probability density function but it is immediately possible to observe some differences in their parameters. The majority of the relative shape error of wind speed is attributed to two intervals almost identical in the number samples $V \in [0, \bar{\mu}_V[\cup]\bar{\mu}_V, 20]\ m/s$. This interval excludes only the mean value of the synthetic wind speed data and reflects that the standard deviation is by far the main contributor for the shape error e_{pdf_V}. The density of synthetic values in the vicinity of $\bar{\mu}_V$ is lower than the one of measured values which reveals that the modeled standard deviation is greater than the measured one.

The values for the mean are indeed very close to each other with an absolute error of approximately 0.1 m/s. The corresponding relative error is $e_{\mu_V} = 3.64\%$ which reveals an overestimation of the model. The standard deviation has, as previously mentioned, poorer representation of reality. Its absolute error is 0.9 m/s which corresponds to a relative error of $e_{\sigma_V} = 49.35\%$. This is a significant overestimation of the model in this particular behavior. The wind speed's time constant T_V also experiences a considerable overestimation of $e_{T_V} = 19.41\%$, showing that the measured wind speed varies values in time with greater freedom. Finally, the correlation between air temperature and wind speed shows a realistic behavior from the model. The relative error of ζ_V is $e_{\zeta_V} = 1.96\%$. This is a minor overestimation of the correlation between the measured air temperature and wind speed.

4.2 Beijing, China

The second case study is done for the city of Beijing in China. It is used to determine how the model responds to an input weather data set from the north hemisphere. An additional reason for choosing this location is its time zone. Perth and Beijing share the same time zone so these results help understand what differences, if any, appear in the output of the model when the hemisphere of the data set changes. Beijing is located in the northeast of China and with the resolution of the satellite MERRA data there is no considerable amount of sea within it. This characteristic of the satellite data give a realistic perspective of the weather in land and nearly unaffected by the presence of sea.

4.2.1 Analysis

The inputs given to the weather model are the latitude of the site (39°54′ North) and the its data set, composed by the GHI, air temperature and wind speed at 50 meters altitude. The time zone in Beijing is also leading Coordinated Universal Time (UTC) by 8 hours, therefore, in order to provide the data set with the correct time zone, it needs to be shifted by 8 hours. The first step is to input the weather data in the analysis concept to determine its specific parameters. These parameters are presented in the following Table 5.

Table 5: Discretization of all numerically determined parameters for Beijing's data set.

Table of numerically determined parameters			
Parameter	GHI	Air Temperature	Wind Speed
$A_{0_n}[-]$	-1.0827	13.1884	5.0980
$A_{d_n}[-]$	-1.7315	5.6810	0.9090
$\phi_{d_n}[rad]$	0.0602	-2.4822	-0.6043
$A_{y_n}[-]$	0.4708	16.4607	0.8740
$\phi_{y_n}[rad]$	2.8541	-2.9268	0.9637
$A_{s_n}[-]$	-0.0371	0.9397	-0.0915
$\bar{\sigma_n}[-]$	0.1815	3.1475	2.1864
$T_n[hr]$	18.1072	23.5604	66.2755
$\zeta_n[-]$	–	0.1882	-0.0846

4.2.2 Synthesis

The following subsections present the synthetic weather data in both a macro- and micro-perspective, *i.e.*, from a seasonal scale and from a daily scale. A comparison between the expected result and actual result is conducted mainly from a visual approach. Lastly, the errors of the output weather data are presented and discussed.

Irradiation

The output GHI data I from both a macro- and micro-perspective are presented in Figure 18, (a) and (b) respectively. For comparison purposes, the Figure 18 also includes measured data I' and the theoretical maximum GHI at any given time ($\Psi \cdot E_g$). To visualize the seasonal behavior of the GHI (a), only the maximums of each day are represented. In general, the characteristics of the synthetic GHI generated for Beijing are very close to Perth's synthetic GHI. The most noticeable change is that the seasons are inversed, as expected. Here, the comparison between the synthetic data of Perth and Beijing is done for their respective seasons and not in their regular time line, *i.e.*, summers are compared to each other. The seasonal parameters A_{y_I} and ϕ_{y_I} appear to be correctly modeled as previously presented. The sporadic occurrence of maximums which are considerably lower than the expected average is present here as well. However,

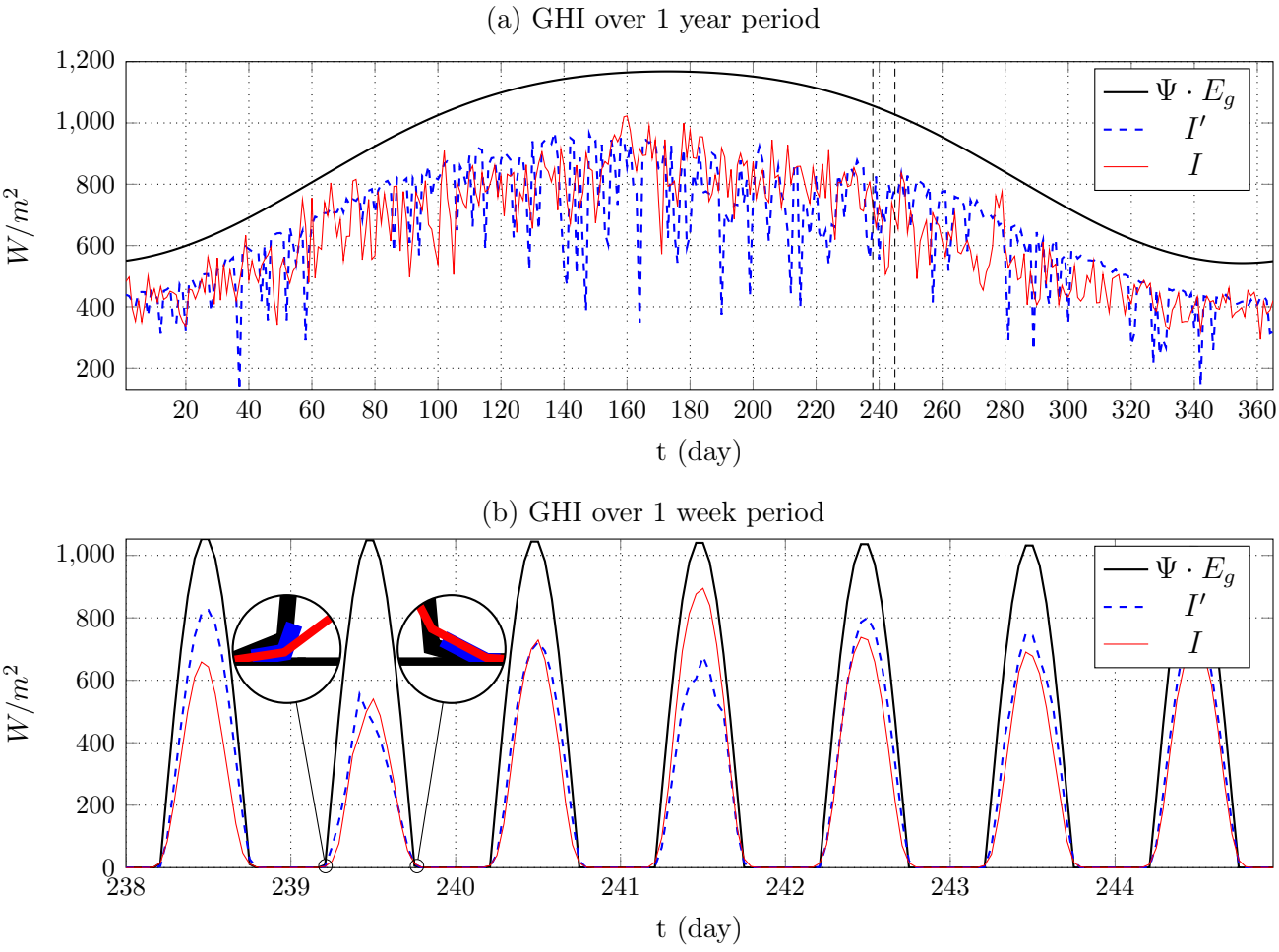


Figure 18: Synthetic and measured weather of Beijing's data set discretized by day (a) over a year long period with its respective maximum values, and (b) over a week long period with all of its respective values.

it is more concentrated during the summer season.

The daily behavior of the solar irradiation is represented in Figure 18 (b). The daily behavior of Beijing's synthetic GHI shows remarkably similarity to that from Perth. This implies a correct modeling of the daily parameters ϕ_{d_I} and A_{d_I}. The time constant parameter T_I is also adequately modeled for the daily behavior of GHI since there are no signs of noise, *i.e.*, unexpected and unreasonable changes in value of the synthetic GHI.

Air Temperature

The output air temperature T and the measure air temperature data T' are presented in Figure 19 from both a macro and micro perspective, (a) and (b) respectively. In (a), the maximums and minimums for each day are represented. The general characteristics for the synthetic weather data are once again similar to those from Perth. The seasonal

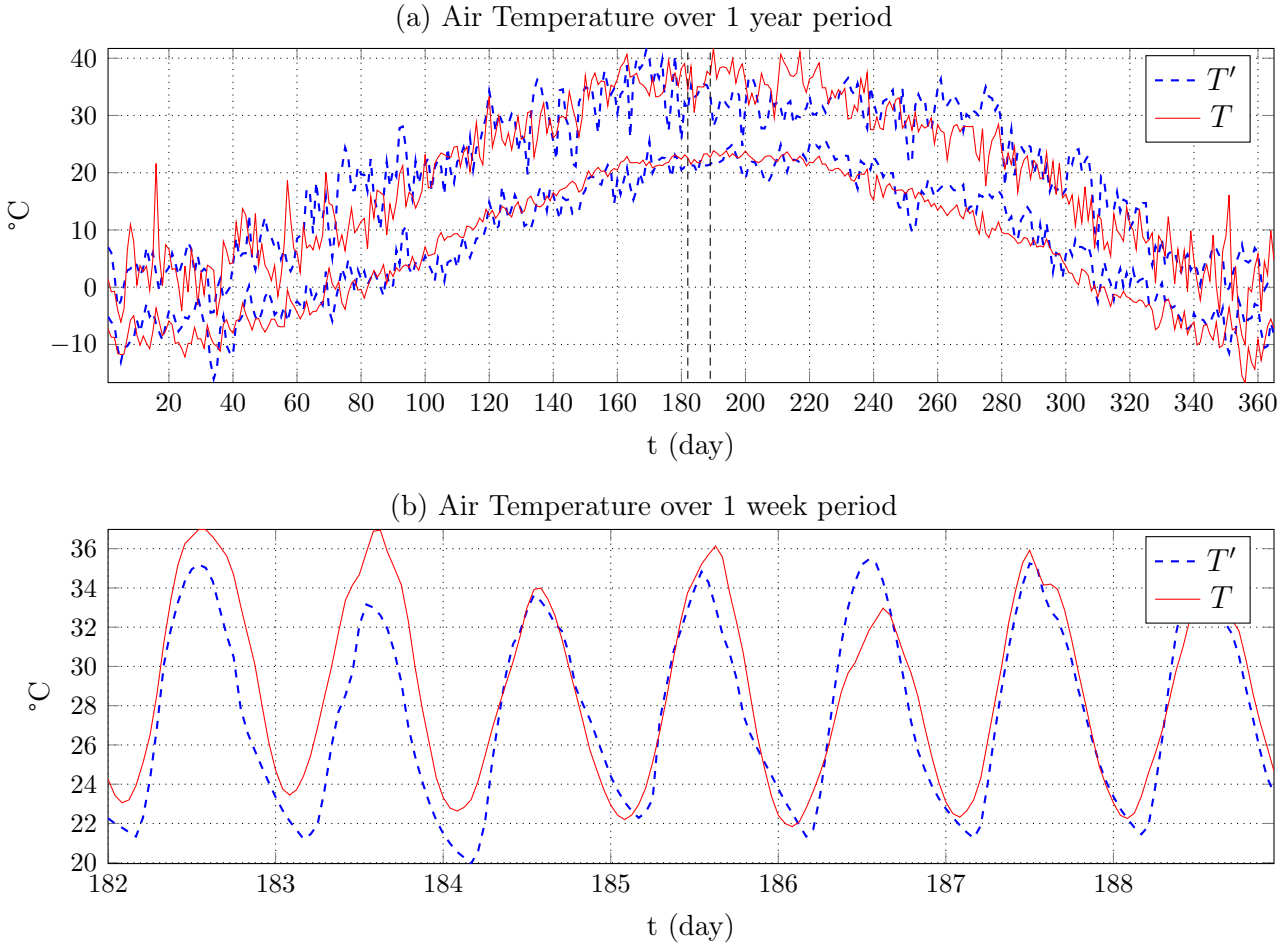


Figure 19: Synthetic Air Temperature data for Beijing's data set discretized by day (a) over a year long period with its respective maximum and minimum values, and (b) over a week long period with all of its respective values.

parameters A_{y_T} and ϕ_{y_T} show good signs of being correctly modeled. During winter time, more specifically between the period of $t \in [1, 20] \cup [340, 365]$, there are occurrences of sporadic 'spikes' in the maximum values of the day for the synthetic data which are not observed in the measured data. The modeled standard deviation $\sigma_T(t)$ has therefore greater values during winter time than the corresponding measured standard deviation $\hat{\sigma}_T[t]$. To visualize the daily behavior of the synthetic air temperature, a plot of every single point generated over a time period of a week. Again, similar to Perth's synthetic data the daily behavior of the air temperature appears to be correctly modeled. The fact that there is no accentuated noise in the synthetic data indicate that the T_T is reasonably well fitted. The correlation parameter ζ_T is indistinguishable from the seasonal variations. For some insight on how well it is modeled, its numerical error needs to be examined.

Wind Speed

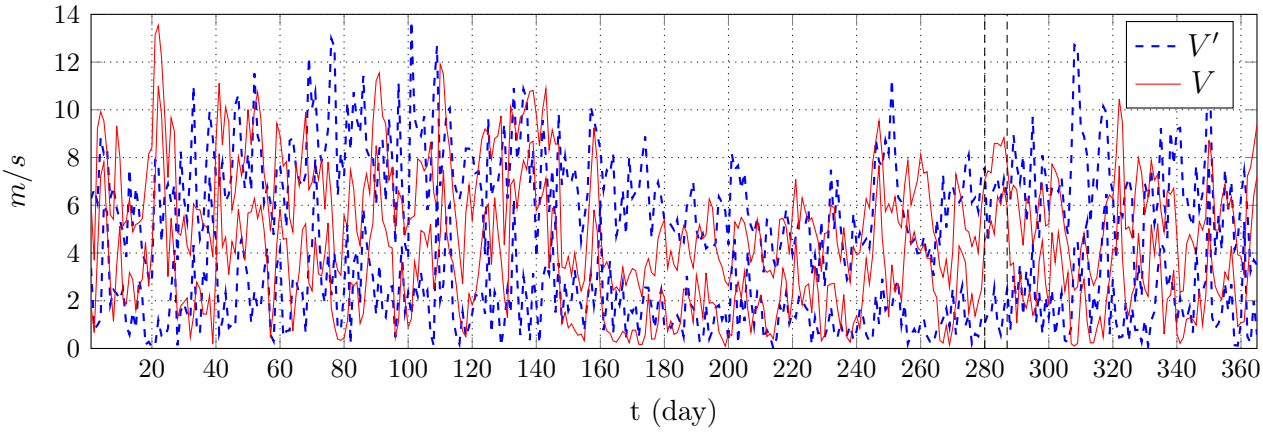
The output wind speed data V and the measure wind speed data V' are represented in Figure 20 with the same previous approach. The first most noticeable characteristic of the wind speed data set in (a) is, again, that they have a much more chaotic behavior than either the GHI or the air temperature. It is also particularly difficult to quickly judge if the seasonal parameters A_{y_V} and ϕ_{y_V} are correctly modeled. The main reason for it is that there are no outstanding differences in the measured wind speed data between seasons. However both data sets fluctuate throughout the entire year around an average value $(\bar{\mu}'_V, \bar{\mu}_V)$ which is remarkably similar. Unlike the synthetic wind speed from Perth, the magnitude of the maximum values in both data sets is negligible. The maximums reached by the measured data are replicated by the synthetic data within the same season but at different time steps. In this case, the standard deviation looks to be well modeled.

Figure 20 (b) shows the daily behavior of the wind speed. The most evident behavior that the synthetic wind speed shows in (b) is the noisy pattern, *i.e.*, very frequent small variances in the signal that are not as noticeable in the measured wind speed data. Despite the noisy behavior, it is still possible to identify that the daily phase parameter of the synthetic data ϕ_{d_V} appears to be correctly modeled. Unfortunately the case is not the same for the daily amplitude parameter A_{d_V}. There are some instances where it appears to have a reasonable value but too often the presence of noisy behavior does not allow for an accurate statement. Similarly to the air temperature, the correlation parameter ζ_V is not addressed from a visual perspective, only by its numerical error.

Errors

The relative shape error for the synthetic of Beijing is represented graphically in Figure 21. The numerically errors are shown in the overall accuracy Table 6. The GHI has a relative shape error of $e_{\text{pdf}_I} = 1.83\%(< 5\%)$ which puts it in the category of good quality data. In Figure 21 (a), the sorted values i for both the synthetic and the measured data of GHI are represented. In general, the probability density function of GHI in this

(a) Wind speed over 1 year period



(b) Wind speed over 1 week period

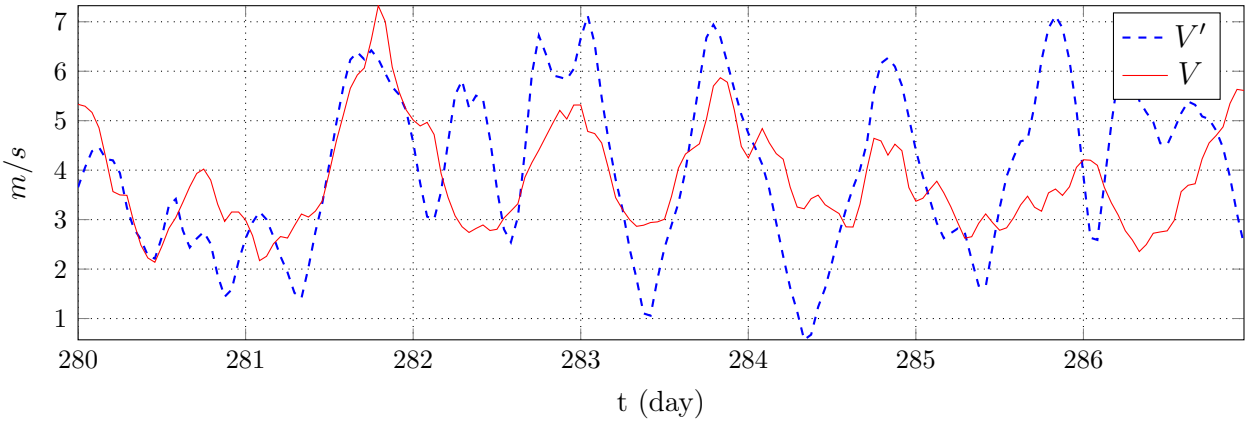


Figure 20: Synthetic wind speed data for Beijing's data set discretized by day (a) over a year long period with its respective maximum and minimum values, and (b) over a week long period with all of its respective values.

case study is extremely similar to that of Perth. The shape of the curves exhibit the same patterns as those in the previous case study, including the day/night transitions magnified in the detail. Here, the difference is the accuracy of the model. For the GHI of Beijing, the model generates a synthetic data with less noticeable fluctuation from the measure data than for Perth and therefore achieves a smaller relative shape error e_{pdf_I}.

The expected values of the measured $\bar{\mu}'_I$ and synthetic $\bar{\mu}_I$ data are very close to each other, with an absolute difference of approximately $1.4W/m^2$. The relative error of the expected value is $e_{\mu_I} = -0.57\%$, which is a negligible underestimation from the model. The standard deviation of the measured $\bar{\sigma}'_I$ and synthetic $\bar{\sigma}_I$ data are very close together, with only an absolute difference of approximately $2.2W/m^2$. The corresponding relative error is $e_{\sigma_I} = -1.65\%$ which represents a minor underestimation. Despite the small difference, Figure 17 (a) shows that the synthetic data covers the entire range of values

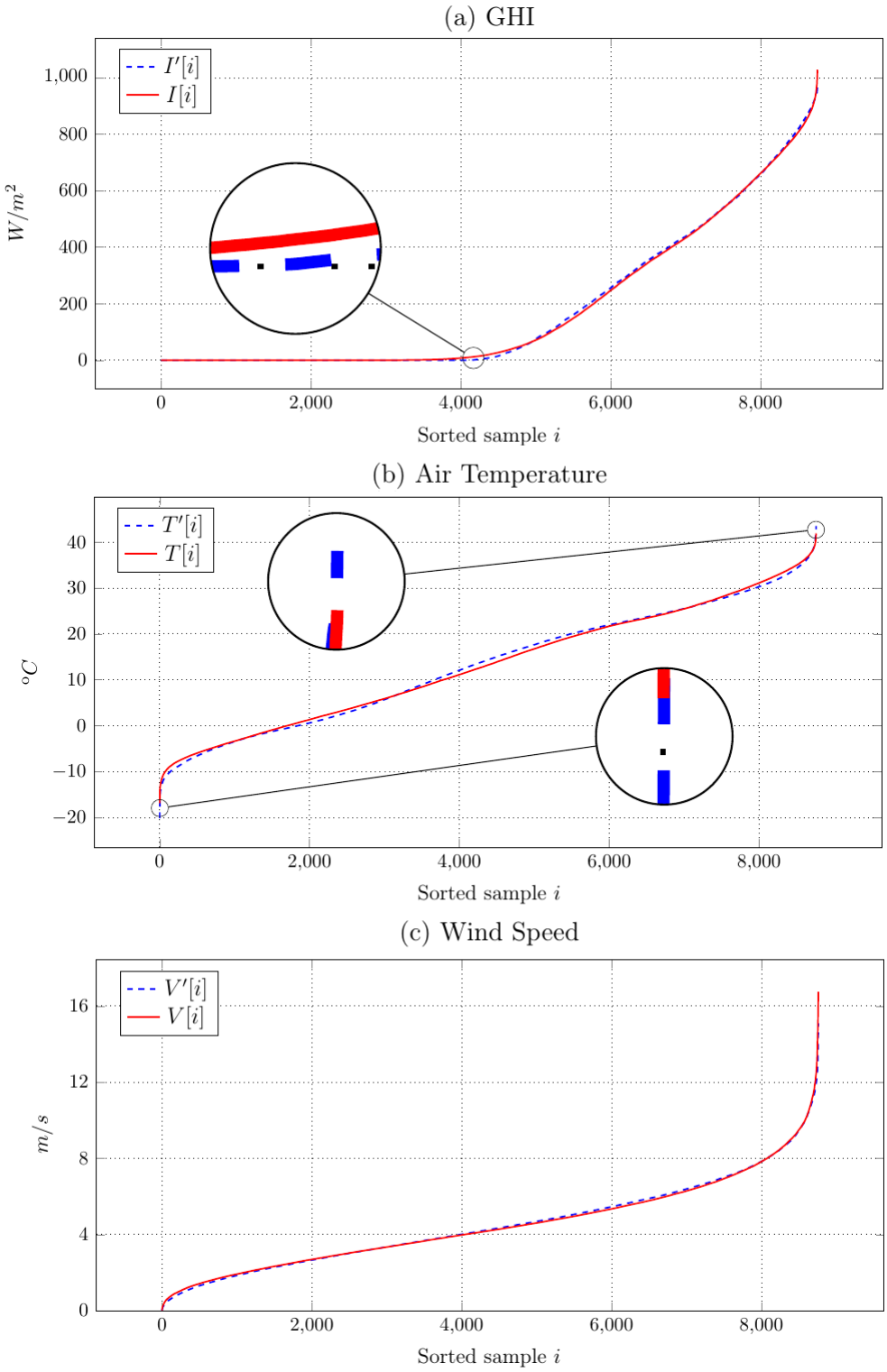


Figure 21: Sorted time series of measured and synthetic weather data for each RER

Table 6: Overall accuracy table for the synthetic RER of Beijing's data set.

Overall accuracy table			
	GHI	Air Temperature	Wind Speed
$e_{\text{pdf}_n}[\%]$	1.83 ± 0.25	3.63 ± 0.32	2.22 ± 0.97
$\bar{\mu}'_n \ [W/m^2, ^\circ C, m/s]$	190.9157	13.1884	4.5153
$\bar{\mu}_n \ [W/m^2, ^\circ C, m/s]$	189.5639	13.2092	4.5035
$e_{\mu_n}[\%]$	-0.57	-1.28	0.43
$\bar{\sigma}'_n \ [W/m^2, ^\circ C, m/s]$	262.7283	12.8948	2.3009
$\bar{\sigma}_n \ [W/m^2, ^\circ C, m/s]$	260.4302	12.7573	2.2869
$e_{\sigma_n}[\%]$	-1.65	-5.29	3.82
$e_{T_n}[\%]$	-1.72	-1.93	7.35
$e_{\zeta_n}[\%]$	$-$	-4.15	0.49

in the measure and no substantial difference exists. The time constant parameter T_I is also slightly underestimated by $e_{T_I} = -1.72\%$ which means that the model describes a weaker correlation in time for GHI.

The air temperature has a relative shape error of $e_{\text{pdf}_T} = 3.63\% (< 5\%)$ which puts it in the category of good quality data. In Figure 21 (b), the overall shape of both curves are remarkably similar for all their domain. The synthetic sorted curve appears to have a stiffer behavior than the measured data, *i.e.*, it does not fluctuate so much across its domain. The other results responsible for the error are the minimum and maximum values. The detail shows that in both cases the measured data has extremes with greater magnitude. The model does, however, come close to generating synthetic data with the same magnitudes which implies that its standard deviation is minorly underestimated.

In Table 6, the numerically determined errors for the air temperature are presented. The overall errors show a very close similarity to those of Perth. For air temperature, all modeled behaviors are once again very close to the correspondent measured behaviors and all of them show a very slight underestimation. Both the expected value and the standard deviation have low absolute errors of 0.02°C and 0.14°C, which correspond to the relative errors of $e_{\mu_T} = -1.28\%$ and $e_{\sigma_T} = -5.29\%$ respectively. The time- and cross-correlation of the model reveals to be superficially weaker than the measured data by $e_{T_T} = -1.93\%$ and $e_{\zeta_T} = -4.15\%$ respectively.

The wind speed has a relative shape error of $e_{\text{pdf}_V} = 2.22 \pm 0.97\% (< 5\%)$ which puts it in the category of good quality data. In Figure 21 (c), it is represented the sorted values i for both synthetic and measured data of wind speed. The overall shape of both curves is extremely similar. The majority of its relative error is attributed to the highest values in the synthetic data set. These have greater magnitude than the measured data and a slightly higher frequency. It evidently clear that the model was able to generate considerably better quality synthetic wind speed data for Beijing than for Perth. This statement is also corroborated by the rest of the relative errors presented in Table 6. The overall wind speed errors show some similarity to the errors of Perth

since all behaviors are modeled with an overestimation. Both the expected value and the standard deviation have low absolute errors of approximately 0.01 m/s and 0.02 m/s which correspond to the relative errors of $e_{\mu_V} = 0.43\%$ and $e_{\sigma_V} = 3.82\%$ respectively. The time-correlation is modeled with an overestimation of $e_{T_V} = 7.35\%$ which is still reasonably low. The cross-correlation of the wind speed is modeled with a negligible error of $e_{\zeta_V} = 0.49\%$, which means that the correlation between air temperature and wind speed for the Beijing is realistically modeled.

4.3 Madrid, Spain

The third case study is done for the city of Madrid in Spain. This case study gives an additional insight on how the model responds to another input weather data from the north hemisphere. Two additional reasons for choosing this location are its time zone and its climate zone. Madrid and Beijing have a similar latitude to each other but a considerably different climate and time zone. These results help to understand what differences, if any, appear in the output of the model when the input weather data remains in the same latitude but changes climate and time zone. Madrid is located in the middle of Spain and with the resolution of the satellite MERRA data there is no sea within it. This characteristic of the satellite data also gives a realistic perspective of the weather in land.

4.3.1 Analysis

The inputs given to the weather model are the latitude of the site (40°25′ North) and its respective data set. The time zone in Madrid is leading Coordinated Universal Time (UTC) by 2 hours, therefore, in order to provide the data set with the correct time zone, it needs to be shifted by 2 hours. The first step is to input the weather data in the analysis concept to determine its specific parameters. These parameters are presented in the following Table 7.

Table 7: Discretization of all numerically determined parameters for Madrid's data set.

Table of numerically determined parameters			
Parameter	GHI	Air Temperature	Wind Speed
A_{0_n}	-1.0895	13.8381	5.7140
A_{d_n}	-1.8649	5.7165	0.6886
ϕ_{d_n}	0.0800	-2.5035	-0.4258
A_{y_n}	0.6722	11.3785	0.5318
ϕ_{y_n}	3.0476	-2.7805	0.5193
A_{s_n}	-0.0140	2.2024	-0.1937
$\bar{\sigma_n}$	0.1919	3.2902	2.3734
T_n	8.8047	45.6151	70.9228
ζ_n	–	0.2247	-0.3051

4.3.2 Synthesis

The following subsections present, in the same fashion, the synthetic weather data in both a macro- and micro-perspective, *i.e.*, from a seasonal scale and from a daily scale. A visual comparison is conducted and afterwards, the correspondent numerical errors are presented and discussed.

Irradiation

Figure 22 shows the model output data I, the measured data I' and the theoretical maximum values ($\Psi \cdot E_g$) of GHI. In general, the characteristics of the generated synthetic data of Madrid are similar to those of Beijing and Perth. The seasonal behavior is represented in Figure 22 (a) and its correspondent parameters A_{y_I} and ϕ_{y_I} are well fitted. The argument supporting this statement is already presented in the previous case studies. The major difference from the previous case study of Beijing is the seasonal concentration of low maximum daily values in the measured data. In the case of Madrid, this concentration is predominant during winter. For all cases, the model shows a reluctance in replicating such sporadic behavior.

The daily scale of the GHI is presented in Figure 22 (b). At this scale, the output of the model is, again, remarkably similarity to those from Beijing and Perth. Since no

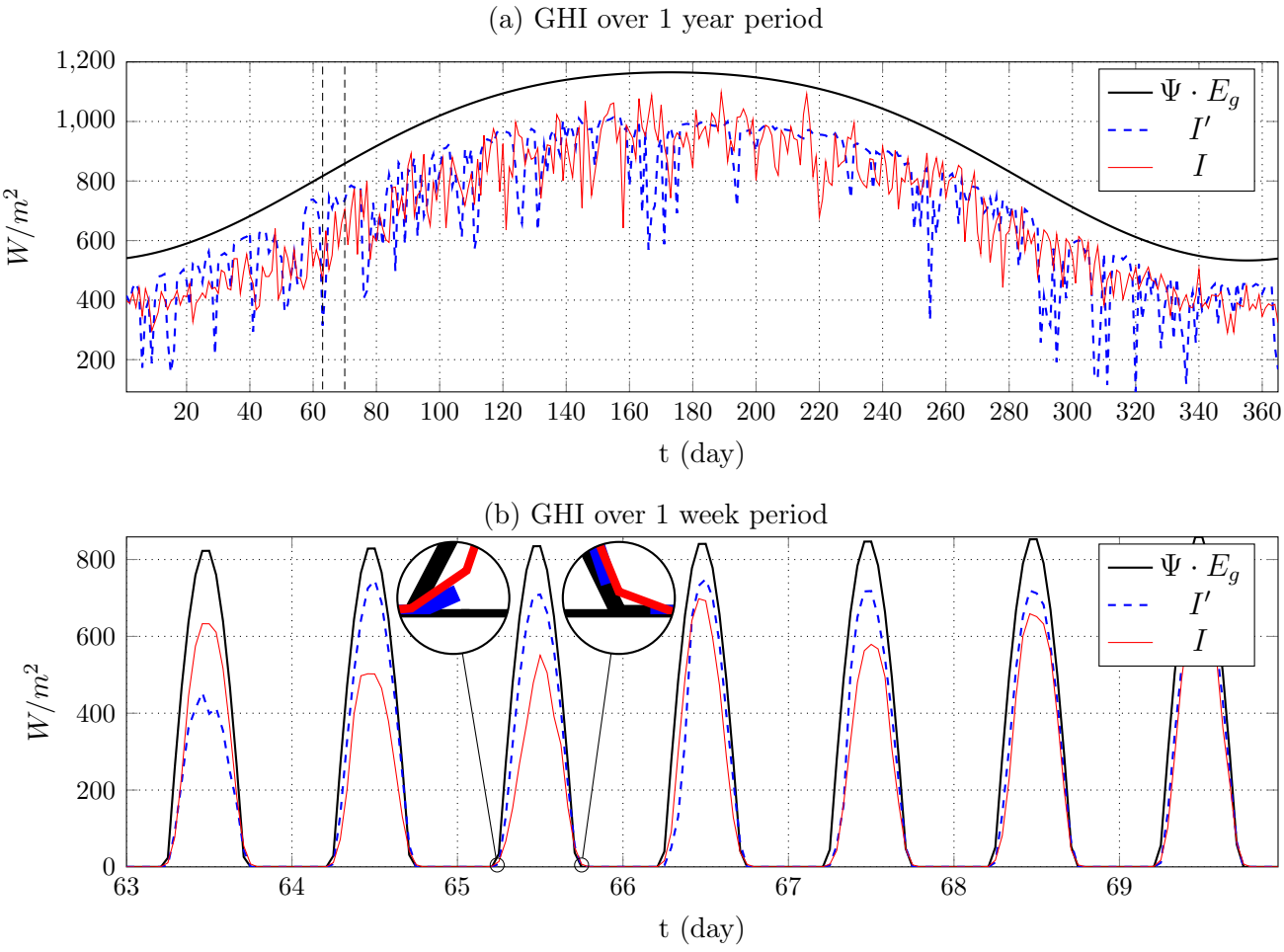


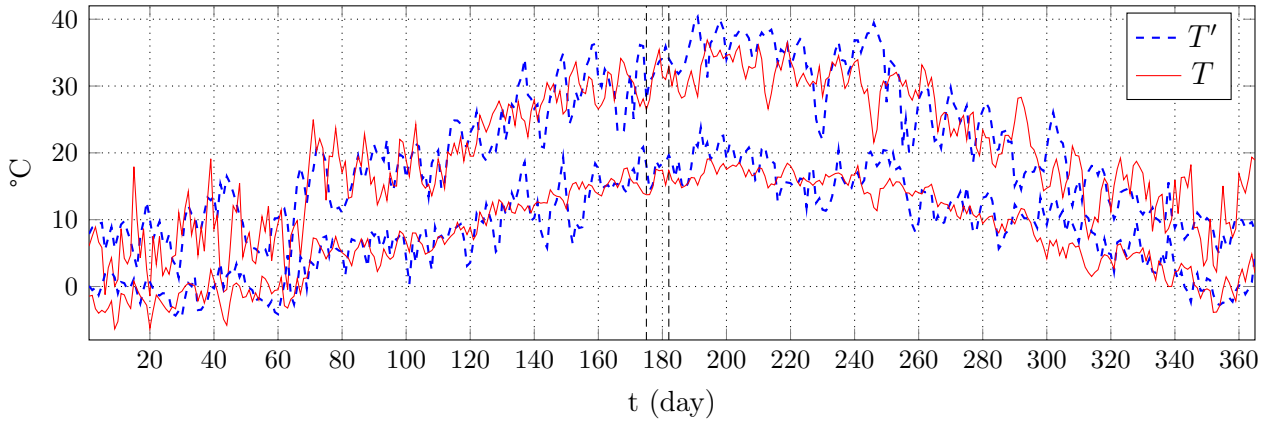
Figure 22: Synthetic weather data for Madrid's data set discretized by day (a) over a year long period with its respective maximum values, and (b) over a week long period with all of its respective values.

substantial changes appear in the GHI output of this case study, the description and arguments of Figure 18 (b) are still valid here.

Air Temperature

Figure 23 shows the model output data T, the measured data T' of air temperature. In (a), the general characteristics for the synthetic weather data are once again similar to those from Beijing and Perth. The seasonal parameters A_{y_T} and ϕ_{y_T} show good signs of being correctly modeled. During winter time, more specifically between the period of $t \in [1, 40] \cup [340, 365]$, there are occurrences of sporadic 'spikes' in the maximum values of the day for the synthetic data which are not observed in the measured data, similarly to Beijing. In (b), the air temperature for a week long period is represented. Again, similar to previous synthetic, data the daily behavior of the air temperature are

(a) Air Temperature over 1 year period



(b) Air Temperature over 1 week period

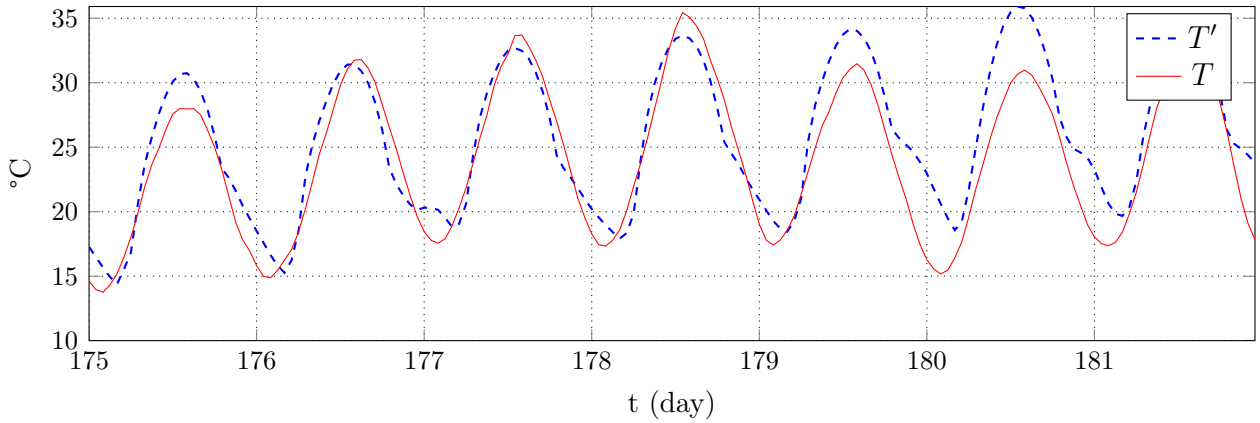


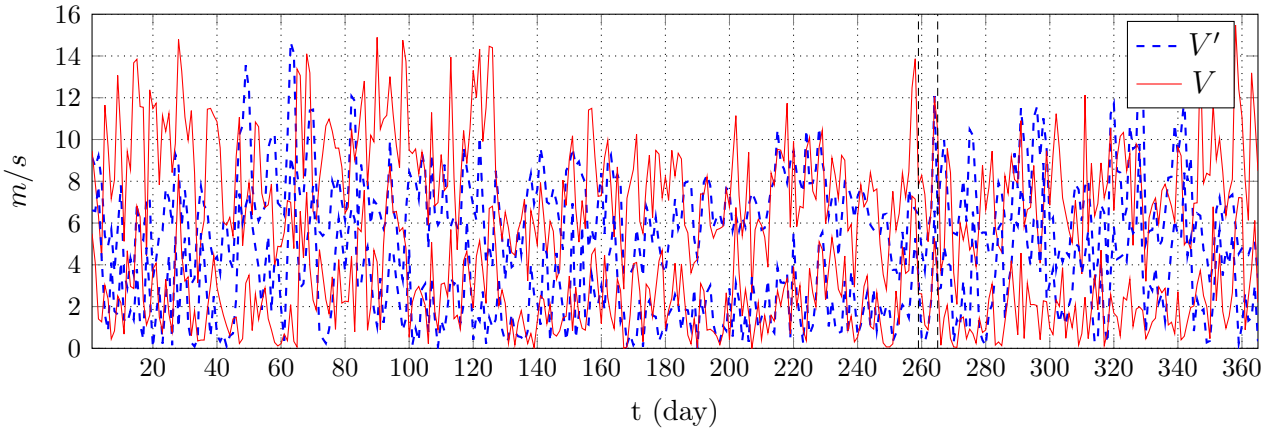
Figure 23: Synthetic Air Temperature data for Madrid's data set discretized by day (a) over a year long period with its respective maximum and minimum values, and (b) over a week long period with all of its respective values.

expected to be very well modeled. The time constant T_T impact in the output exhibits a realistic behavior, confirming that its parametrization is appropriate.

Wind Speed

The output wind speed data (V) and the measure wind speed data (V') are represented in Figure 20 with the same previous approach. The general characteristics of (a) are the same as Beijing. The noticeable difference in (a) from Beijing is that the maximums of synthetic wind speed during winter and spring ($t \in [1, 130] \cup [340, 365]$) are much more frequent. In Figure 20 (b), the synthetic wind speed point generated over a time period of a week. The most evident behavior that the synthetic wind speed shows in (b) is the very frequent small variances in the signal which are not as noticeable in the measured wind speed data. Despite the noisy behavior, it is still possible to identify the synthetic

(a) Wind speed over 1 year period



(b) Wind speed over 1 week period

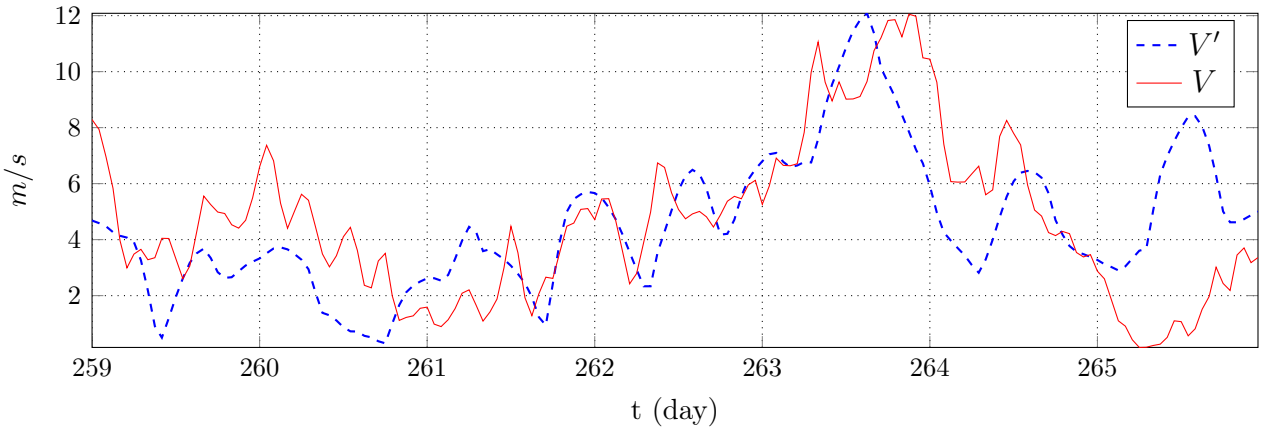


Figure 24: Synthetic wind speed data for Madrid's data set discretized by day (a) over a year long period with its respective maximum and minimum values, and (b) over a week long period with all of its respective values.

data's daily phase parameter ϕ_{d_V} appears to be correctly modeled. Unfortunately the case is not the same for the daily amplitude parameter A_{d_V}. Similarly to Beijing and Perth, the time- and cross-correlation parameter T_V and ζ_V are mainly addressed by their numerical error.

Errors

The relative shape error for the synthetic of Madrid is represented graphically in Figure 25. The numerically errors are shown in the overall accuracy Table 8. The GHI has a relative shape error of $e_{\mathrm{pdf}_I} = 1.81 \pm 0.12\% (< 5\%)$ which puts it in the category of good quality data. In Figure 21 (a), it is represented the sorted values i for both synthetic and measured data of GHI. The general characteristics of this curve are similar to the one of Beijing, including the detail. The maximum synthetic values are here once again noticeable. Only a few amount of data points surpass the measured maximums but not with enough frequency to have a considerable impact in the relative error shape.

The expected values of the measured $\bar{\mu}'_I$ and synthetic $\bar{\mu}_I$ data are notably close to each other, with only an absolute difference of approximately $0.7 W/m^2$. The relative error of the expected value shows that this difference is less than -1% and that it is a negligible underestimation from the model. The standard deviation of the measured $\bar{\sigma}'_I$ and synthetic $\bar{\sigma}_I$ data are also close together, with only an absolute difference of approximately $0.86 W/m^2$. Unlike the error from Beijing, here the standard deviation is modeled with an overestimation of $e_{\sigma_I} = 2.56\%$. The time constant parameter T_I has a negligible underestimation of $e_{T_I} = -0.55\%$ which means that the model describes a realistic correlation in time for GHI.

The air temperature has a relative shape error of $e_{\mathrm{pdf}_T} = 4.59 \pm 0.33\% (< 5\%)$ which puts it in the category of good quality data. In Figure 25 (b), the overall shape of both curves are similar for almost all their domain. The synthetic curve show similarities to the curve from Perth, especially the range of values for highest air temperatures is

Table 8: Overall accuracy table for the synthetic RER of Madrid's data set.

Overall accuracy table			
Parameter	GHI	Air Temperature	Wind Speed
$e_{\mathrm{pdf}_n}[\%]$	1.81 ± 0.12	4.59 ± 0.33	9.22 ± 0.66
$\bar{\mu}'_n \ [W/m^2, ^\circ C, m/s]$	210.7447	13.8381	5.0654
$\bar{\mu}_n \ [W/m^2, ^\circ C, m/s]$	210.0206	13.9104	5.2373
$e_{\mu_n}[\%]$	-0.34	0.52	3.39
$\bar{\sigma}'_n \ [W/m^2, ^\circ C, m/s]$	292.4902	9.8318	2.3871
$\bar{\sigma}_n \ [W/m^2, ^\circ C, m/s]$	291.6343	9.4893	2.9069
$e_{\sigma_n}[\%]$	2.56	-5.11	16.68
$e_{T_n}[\%]$	-0.55	-1.51	12.42
$e_{\zeta_n}[\%]$	$-$	-3.14	3.44

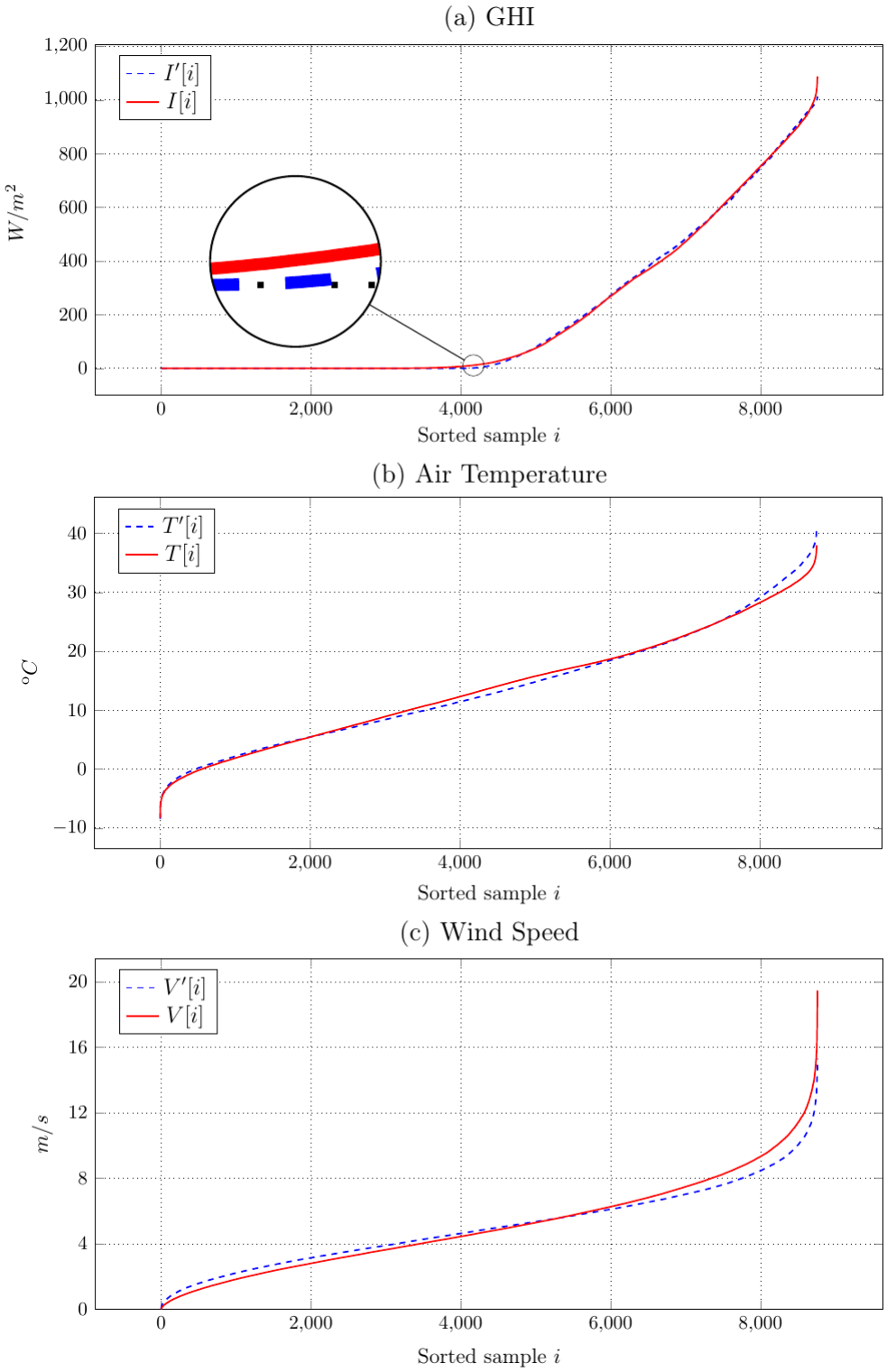


Figure 25: Sorted time series of measured and synthetic weather data for each RER

not a perfect match. The reason for such is also the underestimation of the modeled standard deviation $\sigma_T(t)$. The overall errors show in Table 8 are close to the errors of Beijing. The only substantial difference being the relative error in the modeled mean which here is overestimated by $\mu_T = 0.52\%$ and an absolute error of 0.08°C. All the other behavior relative to the air temperature are modeled with a superficial underestimation. The standard deviation has a absolute error of $0.34W/m^2$ which corresponds to a relative error of $e_{\sigma_T} = -5.11\%$. The time- and cross-correlation have a relative error of $e_{T_T} = -1.51\%$ and $e_{\zeta_T} = -3.14\%$ respectively.

The wind speed has a relative shape error of $e_{\mathrm{pdf}_V} = 9.22 \pm 0.66\% (> 5\%)$ which puts it in the category of reasonable quality data. In Figure 25 (c), it is represented the sorted values i for both synthetic and measured data of wind speed. The overall shape of both curves is similar. The majority of the relative error is attributed to the range of values for the highest wind speeds $i \in [6000, 8760]$. This behavior is a consequence of an overestimation in the standard deviation $\sigma_V(t)$. The overall wind speed errors shown in Table 8 have the same general characteristics as the errors for Perth and Beijing. All the modeled behaviors are overestimated to a certain degree. The modeled mean shows low absolute and relative error values of $0.17m/s$ and $e_{\mu_V} = 3.39\%$ respectively. The modeled standard deviation and time-correlation reveal a considerable increase in their relative error values when compared to better quality synthetic data such as Beijing. It means that the extreme values appears more often and are less prone to change their value than the measured wind speed data. The cross-correlation shows, however, a reasonable approximation of the influence that the air temperature has on the wind speed. Its corresponding error being $e_{\zeta_V} = 3.44\%$.

4.4 Discussion

The case studies in the previous chapters show the different outputs that the weather model can generate when the latitude, the time zone and the climate zone of a data set change. They are representative of all data sets in Table 2. This chapter discusses the overall results for all data sets and elaborates on the average output of the model. In general, the weather model reveals no significant issue in adapting to any specific location. This result corroborates the initial statement that this work introduces a general and analytical modeling concept for weather phenomena.

The solar irradiation shows, without doubt, the most accurate results obtained in the model. Taking the relative shape errors as the main measurement of quality, the results from the solar irradiation have a consistent e_{pdf_I} below 3%. Given the simplicity with which these weather phenomena are modeled, obtaining such realistic results is exceptionally optimistic. One of the biggest challenges that this work overcomes is the analytical modeling of an entire diurnal variation of the solar irradiation. The transition of night to day time and reverse at sunrise and sunset is accomplished by the transformation introduced (33). Across all locations tested, only a small error was obtained from these transitions which is generally visible in the magnified detail of Figures 17, 21 and 25 (a). The most significant difference in error values is related to the modeling of maximums. The behaviors responsible for it are the expected value μ_I and the standard deviation σ_I. The modeled expected value has, in every single case study, displayed a lower error than the standard deviation. Therefore, the standard deviation is the most accountable behavior for the slight lack of accuracy. This is actually an expected result. The majority of parameters removed from the original model are the ones responsible for independently modeling the standard deviation. The comparison of the overall accuracy tables for the different sites reveal no pattern in the behavior modeling for solar irradiation, i.e., there is no systematic overestimation or underestimation for any of the modeled behaviors. The relative error of all behaviors is consistently within a 3% range ($e_{\mu_I}, e_{\sigma_I}, e_{T_I} \in [-3, 3]\%$). These two previous observations imply that the weather model adapts very well to different data sets without significant fluctuations.

The air temperature comes second to solar irradiation in terms of overall data quality. For all tested locations, the relative shape error for the air temperature is consistently below 5%. Despite the air temperature having a higher average relative shape error than the solar irradiation, it is still considered to be of good quality. The error relative to its probability density function is mainly generated by the highest air temperature values. It represents a skewing of the synthetic data. The normal distribution assumed for the air temperature also shows a minor skewed fitting as illustrated in Figure 9. This issue can be addressed by assuming a Weibull distribution for the air temperature instead. In this work, such is not done because it represents an increase in complexity for the model and the concept aims to have the simplest description possible. This represents a practical trade-off between mathematical complexity and accuracy. The tables of overall accuracy for the air temperature reveal a pattern of behavior modeling, unlike the solar irradiation. The standard deviation $\sigma_T(t)$, time constant T_T and cross-correlation ζ_T show a consistent slight underestimation of the model, for all data sets. The modeled

expected value does not show the same pattern. It is consistently modeled with a negligible error and therefore any fluctuation of underestimation or overestimation are considered redundant.

The wind speed is the modeled weather phenomenon with the least positive results. Its relative shape error value is on average below 10%. There are, however, some instances when this error can be higher; such as the presented case study of Perth, but the tendency is never to stretch too far away from this value. The overall accuracy tables for wind speed also reveal a pattern in behavior modeling, however, reverse to the one of air temperature. In this case, all the modeled behavior show a consistent overestimation of different magnitudes. Throughout the different data sets, the expected value $\mu_V(t)$ and the cross-correlation ζ_V display only a slight relative error. Unfortunately, the standard deviation $\sigma_V(t)$ and the time-correlation T_V have a unpredictable magnitude for their error. The standard deviation of the wind speed is by far the behavior which suffers the greatest loss in accuracy. Its impacts on the probability density function of the synthetic data are noticeable but do not invalidate it. Unlike the other modeled standard deviation $\sigma_I(t)$ and $\sigma_T(t)$, this one displays the highest necessity for further iteration and reformulation. The wind speed also shows some abnormal fluctuations in time. The visual comparison of the generated synthetic data with measured data is particularly difficult to make because of its inherent chaotic behavior. This is a further indicator that the wind speed requires a more complex modeling approach than the other weather phenomena. In this model the only correlation assumed for the wind speed is with the air temperature; however, other weather phenomena, such as precipitation, could have a strong impact in the general behavior of the wind speed.

5 Summary and conclusion

In this thesis, a simplified analytical description for weather phenomena as a renewable energy resource has been presented. This dynamic model provides an insight regarding the fundamental behavior that describes the weather and how it can be translated into mathematical quantities. This model is designed with an approach to easily enable stochastic analysis of renewable energy systems. The model is not meant to be used in an extremely realistic investigation of a specific site but rather to allow the research of basic principles and properties of renewable systems. The most widespread renewable power generation units are photovoltaic cells and wind turbines. The weather phenomena focused in this thesis are solar irradiation, air temperature and wind speed since they are the major driving forces of renewable power generation.

The core concept of the model is that from a stochastic random process, an accurate synthetic weather description can be achieved. The modeling focus is on the most evident and noticeable behaviors of the weather while leaving other effects to be introduced by a random signal. The simplest way to describe a random signal is by its expected value and standard deviation. A normal distribution is therefore assumed. It serves as the statistical base for the introduction of all fundamental behaviors of weather. These are the specific rate of change in time, time dependency and the correlation to other weather phenomena. Each weather phenomenon also has a distinct natural statistical distribution. For that reason, it is also necessary to establish a relationship between their natural distribution and then normal distribution.

The framework for the modeling approach remains valid from the original base concept. However, the mathematical complexity, number of parameters and analytical description needed to be addressed. This work introduces an alternative transformation for the solar irradiation which removes the discontinuities presented in the base concept. The modeling of the time dependency in the base model accounted 89% of the total number of parameters. The approach taken in this work reduces the necessary amount of parameters for an accurate description of the time dependency by 61%. The general implementation for both time- and cross-correlation through the dynamics of a first order low pass system proved itself to be well adjusted to the level of complexity and number of parameters initially aspired.

This model has the fundamental objective of being a general description of weather. The emphasis in 'general' relates to its applicability worldwide. During the process of iterating the model, several locations around the world are tested in order to both achieve a general concept and avoid an over-fitting for any data set. The model proves itself to be capable of adapting to different climate zones and both hemispheres. This characteristic gives the model an additional attractiveness for the analysis of renewable energy systems.

The overall results generated by the model show a good accuracy. Both the probability density functions and the individual modeled behaviors are tested for errors. The model produces synthetic weather data with remarkably similar statistical description to the measured which leads to low relative shape errors. Furthermore, the individual modeled behaviors resemble those of measured weather data both visually and numer-

ically, through their errors. The model shows an increase of accuracy for the relative error shape reversely to the methodology applied, *i.e.*, the model consistently generates synthetic solar irradiation with the least overall errors, followed by air temperature and wind speed.

Future work could be focused on improvements for the analysis concept and further generalization of the model. Since the analysis concept is not a part of the model, the numerical optimization methods can be altered to suffer an increase in mathematical complexity in order to better parametrize the behaviors of weather. The model can also be further generalized in its approach. For example, behaviors such as the standard deviation require specific assumptions for different weather phenomena. The determination of errors is not trivial and should also be subjected to revision in order to better describe the impact that each parameter has in the output of the model. The weather data sets for future iterations of the model should also include measured data from weather stations to determine its accuracy in considerably smaller spatial scales.

References

[1] *Zeitreihen zur Entwicklung der erneuerbaren Energien in Deutschland*, http://www.erneuerbare-energien.de/EE/Navigation/DE/Service/Erneuerbare_Energien_in_Zahlen/Zeitreihen/zeitreihen.html, Accessed: 2017-10-01, 2017 (cit. on p. 1).

[2] *Flexible Resources to Help Renewables*, http://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf, Accessed: 2017-10-3, 2016 (cit. on p. 1).

[3] F. Christange and T. Hamacher, "Analytical modeling concept for weather phenomena as renewable energy resources", *IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 273–278, Birmingham, 2016 (cit. on pp. 1, 2, 5, 8, 31).

[4] D. Wilks and R. Wilby, "The weather generation game: a review of stochastic weather models", *Progress in Physical Geography*, vol. 23, no. 3, pp. 329–357, 1999 (cit. on pp. 3, 4).

[5] P. Ailliot, D. Allard, P. Naveau, and et. al., "Stochastic weather generators: an overview of weather type models", *Journal de la Societe Francaise de Statistique*, vol. 156, no. 1, pp. 1–14, 2015 (cit. on p. 3).

[6] R. Srikanthan and T. McMahon, "Stochastic generation of annual, monthly and daily climate data: A review", *Hydrology and Earth System Sciences*, vol. 5, no. 4, pp. 653–670, (cit. on p. 3).

[7] C. Richardson, "Weather simulation for crop management models", *Transactions of the ASAE*, vol. 28, pp. 1602–1606, 1985 (cit. on p. 4).

[8] J. Shirley, "Solar Constant", in *Encyclopedia of World Climatology*, J. E. Oliver, Ed., Springer Netherlands, 2005, pp. 666–667 (cit. on p. 5).

[9] G. Kopp and J. Lean, "A new, lower value of total solar irradiance: Evidence and climate significance", *Geophysical Research Letters. 38*, (cit. on p. 5).

[10] J. Hay, "Calculating solar radiation for inclined surfaces: Practical approaches", *Renewable Energy*, vol. 3, no. 4-5, pp. 373–380, 1993 (cit. on pp. 7, 26).

[11] S. Gordon, *The Normal Distribution*, http://sydney.edu.au/stuserv/documents/maths_learning_centre/normal2010web.pdf, Accessed: 2017-10-17, 2006 (cit. on p. 9).

[12] A. Azad, M. Rasul, M. Alam, and et. al., "Analysis of wind energy conversion system using Weibull distribution", *Procedia Engineering*, vol. 90, pp. 725–732, 2014 (cit. on p. 10).

[13] E. Akpinar and S. Akpinar, "Statistical analysis of wind energy potential on the basis of the Weibull and Rayleigh distributions for Agin-Elazig, Turkey", *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 218, no. 8, pp. 557–565, 2004 (cit. on p. 10).

[14] R. Corless, G. Gonnet, D. Hare, and et. al., "On the Lambert W function", *Advances in Computational Mathematics. Berlin, New York: Springer-Verlag.*, vol. 5, pp. 329–359, 1996 (cit. on p. 11).

[15] M. Rienecker, M. Suarez, R. Gelaro, and et. al., "MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications," *Journal of Climate*, vol. 24, no. 14, pp. 3624–3648, July 2011 (cit. on p. 31).

[16] M. Kottek, J. Grieser, C. Beck, and et. al., "World Map of the Köppen-Geiger climate classification updated", *Meteorologische Zeitschrift*, vol. 15, no. 3, pp. 259–263, June 2006 (cit. on p. 32).

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen, als die angegebenen Quellen und Hilfsmittel benutzt, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet habe.

Ort, den Datum