Weather Database Analysis and Simulation for Solar Heating and Cooling System

Research Report

Submitted to the Graduate School of Asia Pacific Studies

in partial fulfillment of the requirements

for the degree of Master of Science in International Cooperation Policy

by

Edwin Purwanto

51209628

JAPAN-GERMANY DUAL MASTER'S DEGREE PROGRAM

INTERNATIONAL MATERIAL FLOW MANAGEMENT

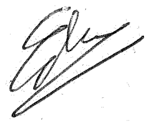
RITSUMEIKAN ASIA PACIFIC UNIVERSITY

2011

Certification Page

In light of academic, moral, and ethical principles, I hereby declare that this report represents original research efforts that I have made with the advice of my academic supervisors. I have tried my absolute best to cite each portion of this report that uses other sources of information with proper referencing procedures. And whosever work that has been used as content in or to support any part of this report is clearly detailed explicitly in the reference section.

Bolzano, 30 June 2011



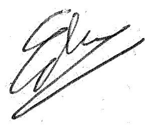
Edwin Purwanto

Author

Disclaimer

The author strives to provide accurate data from various sources. However data presented here is not guaranteed to be free from error. If there is error, the author would be very pleased to be informed.

Bolzano, 30 June 2011



Edwin Purwanto

Author

Acknowledgement

I am grateful to APU, FH Trier, MEXT and DAAD for the opportunity of pursuing further study and life experience that has been given to me. I would like to specially deliver my gratitude to Professor Doctor Michael Bottlinger for his guidance, encouragement, and support as the supervisor of this research. Without his help, the completion of this report would be impossible. I owe my thanks as well to other professors, lecturers, and staff who have made available their support academically and motivationally in a number of ways since the beginning of this program.

I am indebted to EURAC for the chance I had during the internship term to write this report with direct support from and in the atmosphere of the experts in the field. I would like to thank Doctor Wolfram Sparber who provided this thesis opportunity. I also would like to thank to Mr. Roberto Fedrizzi for this thesis idea and has unwearyingly supervised the progress of this research. I would also like to thank Mr. Matteo D'Antoni and my other EURAC colleagues for their kind help, especially the administrator, Ms. Jöchler Corrine, who underwent a lot of difficulties to manage my problem with my visa.

Unforgettable friendships and interactions with my classmates from many countries in APU and UCB, has really been an inspirational life-time experience, with which this time has been so precious to endure through. My utmost thankfulness for sure is to the great persons of my seventeen 2009 IMAT classmates.

Bolzano, 30 June 2011

Author

Abstract

It is considerably vital to have coherent weather data to help planning solar thermal (heating and cooling) systems. Currently, various types and methodologies of weather databases, each with their own strengths and weaknesses could be used for those purposes. To note some prominent examples, weather data could be primarily obtained by either ground weather stations or meteorological satellites. Data could also be artificial data generated from simulation software using formerly established models. Methodologies to process the raw weather database into a developed database could also produce divergent results. Regarding these points, it is important to have an insight to compare and analyze the databases as well. It is due to the fact that different databases exist for different purposes. Different methodologies within those databases will also bring implications towards the planning of solar thermal systems.

The research is primarily a literature review, comparing several major weather databases throughout the world. From this point, it has resulted in a comprehensive algorithm of how the weather database is made, a comprehensive table of weather databases, which serves as a tool to influence the selection of data sources. Weather database classification is thus proposed as a tool to facilitate the comparison. Secondly, a comparison between samples of weather databases in certain locations versus real time measurements with the same conditions is therefore performed to validate the database from error and uncertainty. Finally, a simplified method for quantifying the potential of solar heating and cooling systems will be expected.

Keywords: weather database, weather database classification, typical year generation, radiation chain of algorithm

Table of Contents

List of Figures

List of Tables

Appendix List

List of Abbreviation

Acronym	*Meaning*
ARMA	Autoregressive Moving Average
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
BRY	Biomass Reference Year
BSRN	Baseline Solar Radiation Network
CDF	Cumulative Distribution Function
CIBSE	Chartered Institution of Building Services Engineers
DRY	Design Reference Year
ESRA	European Solar Radiation Database
HC	Helio Climate
IGDG	Italia Dati Climati"G. De Giorgio"
IWEC	International Weather For Energy Calculation
KSI	Kolgorov Smirnov Index
MTM	Markov Transition Matrices
NASA SSE	NASA Surface Meteorology and Solar Energy
NCDC	National Climatic Data Center
NREL	National Renewable Energy Laboratory
PRT	Platinum Resistance Temperature
PVGIS	Photovoltaic Geographical Information System
SoDa	Solar Radiation Data
SRB	Surface Radiation Budget
SWERA	Solar and Wind Energy resource Assessment
TAG	Time dependent Autoregrresive Gaussian
TMY	Typical Meteorological Year
TRY	Typical Reference Year
USCRN	US Climate Reference Network

1. Introduction

1.1 Background

One of the most prominent challenges in establishing solar heating and cooling systems is the planning component. It should comprise of the estimation on how much heating or cooling is needed and how much potential lies in the particular sites of installation. Hypothetically speaking, temperature (dry bulb) should be the most significant parameter affecting the supply of heating energy and the demand of heating and cooling. Solar radiation is also a dominant factor, because it has proportional direct relation with temperature. When the sky is clear (cloudless), radiation received on the surface of the earth is assumed to be higher along with the temperature. Relative humidity is yet another parameter that contributes to the heating and cooling demand due to the fact that people are normally sensitive to humid air. The human body utilizes evaporative cooling as the primary mechanism to regulate internal temperature. Wind speed is also another influencing factor that could affect the heating and cooling demand because it contributes to temperature and humidity distribution. Within this paper, it is expected that the reader understands the relation of these parameters, the influence each parameter has on the other and how the databases affect it, whether it is from actual measurements or derived from other existing parameters, and how accurate it is supposed to be.

Currently, weather databases exist to fulfill planning and simulation needs. However, it is common practice to simply apply the existing weather databases to the simulations and calculations, therefore neglecting the accuracy of those databases, how those databases are made, and the assumptions behind it. The result in some cases is that simulations contain bugs which cause inaccuracies for various reasons such as the primary database losing some critical data while there is no method to fill those missing gaps. Other similar cases occur when simulations assume that databases come from real-time measurements. Most of the time, it does not. Databases of course are developed with different purposes in mind. For example, TMY3 specifies in its user manual that the database is not suitable for evaluating real-time energy production or efficiency in building design applications.

Another tool that easily and quickly examines heating potential and its demand is climate classification. It could be a rule of thumb to say that in equatorial or tropical areas, it is expected that the temperature will be very hot and relative humidity will also be high. So, cooling demand will likely be high throughout the year. However this approach is largely based on temperature and precipitation. In most cases it is not feasible to deduce the potential of solar thermal systems merely by climate classification alone. Thus, climate classification is usually a side approach that could be used to complement primary weather data. Sometimes in weather data files, a climate classification code (usually from Köppen climate classification) is written to further complement the data. It is adequate to quickly have an overview of a region being examined and cross check the weather data.

1.2 Purpose

Realizing the need to have comprehensive knowledge about weather databases, this paper is intended firstly to provide information over major weather databases available all over the world and compare it. It is expected also to understand how these databases work. As was mentioned, databases are produced with different processes and are aimed towards different applications. It is expected at the end of this first phase to have knowledge about how the databases are produced. Not all databases are expected to follow the same steps, nevertheless common similarities of properties are expected to be found between them.

Second, knowing the assumptions, production processes, and validation of databases, it is expected to benchmark the known databases on particular sites with real proxy measurements on the ground. The proposed sites will be in Bolzano. It is expected that NASA SSE, PVGIS, IGDG, and HC3 databases have data over these sites so it can be compared to the real measurements and produce comparisons on which one is more accurate.

Third, from examining major parameters influencing solar thermal systems such as: solar radiation, dry bulb temperature, wind speed, and relative humidity, it is expected that the reader understands the relation between them. Therefore an algorithm to deduce supply potentials of solar thermal or its heating and cooling demand could be expected.

1.3 Objective

At the end of this research, these following objectives could be expected to meet:

1. An elaborative comparison table between each database.
2. Knowledge of how weather databases are produced and verified and verification on what is claimed by weather databases is also expected from this thesis.
3. Weather database classification to facilitate quick overviews of properties of weather databases available in the market.
4. Knowledge of relations between parameters affecting the potential of solar thermal and its demand is expected.

1.4 Assumption, Scope, and Limitation

This paper has the following assumptions: scope and limitation.

1. Weather databases would be defined as databases that are ready to use and available in the market. This paper will try to check whether the databases are a derivation of another database or a primary one. The paper *will not* try to further check the primary database being used as a benchmark or primary source. As a result, several assumptions have been made which are:
    a. *Primary parameter* is a central element that will be used in a database that is being examined. It will not matter if within the other primary databases, they are being derived.
    b. Error is the difference between values on database minus value on benchmarking databases. Although databases might come with different benchmarking and validity methods, it is assumed that databases used as *benchmarks are uniformly valid and accurate*.
2. There are numerous weather databases available in the market, but this paper will focus on these databases:
    a. Satellite based:
        i. NASA SEE (NASA Surface Meteorology and Solar Energy)
        ii. PVGIS (Photovoltaic Geographical Information System)
        iii. ESRA (European Solar Radiation Atlas)
        iv. HC3 (Helio Climate 3)
    b. Ground Station based:

i. Meteonorm

ii. IWEC (International Weather for Energy Calculation)

iii. TMY (Typical Meteorological Year)

iv. Other nationals weather database: IGDG (Italia Gianni de Giorno, Italy), and CIBSE (Chartered Institution of Building Service Engineer, United Kingdom)

c. Other prominent databases such as SWERA (Solar and Wind Resource Assessment), or RETScreen are not discussed here.

d. The division between ground stations versus satellites is not entirely valid. In some cases, databases are made by data from both ground weather stations and satellites. The division is just to note the *most prominent contributors*. So a database being listed as a ground station based could have 70% of its data coming from ground station while the other 30% of it comes from satellites.

3. The validation of sites throughout the world would result in different values. Nevertheless, the *global accuracy value is a mean value* of sites all over the world. This means that if databases are validating their values from 5 stations all over the world, the end result of the global accuracy of its databases is a mean value of all 5 values.

4. Converting from shorter time increments into a longer ones (hourly to daily to monthly) *is assumed to be simple* and done without error so will not be discussed in this paper. The conversion from larger time increments into shorter time increments is another case and will be discussed in the stochastic time generation section.

5. The relation between influencing parameters towards potentials of solar thermal as well as their demand will be researched in this paper. However, the exact relation of how

much these parameters affect solar thermal potential in number is not involved, due to the complexity of the calculation and time constraints.

6. This thesis will not discuss about real time weather data or weather prediction for 24 hour upfront. There may be weather database capable of doing such things, but this is very limited. Furthermore, real time weather data and weather prediction cannot be used to predict the solar thermal potential.

1.5 Organization

The thesis will be organized into five major chapters which are: introduction, literature review, comparison and analysis, result and discussion, and conclusion. Similar subjects with different depths of explanation can be found between chapter 2: Literature review, chapter 3: comparison and analysis, and chapter 4: results and discussion. For example, a topic will be briefly explained in chapter 2, while the comparison between them can be found in chapter 3. Finally, the expected impact into solar heating and cooling systems can be found in chapter 4.

Subsequently, each chapter will be organized into different sub-chapters as shown below:

1. Chapter 1: Introduction

   Background, purpose, and the objectives of this paper are explained in this chapter. The organization of this paper plus assumptions, scopes and limitations are also briefly explained within this chapter. However, in the main body segment of this paper further elaborative assumptions and limitations should be expected.

2. Chapter 2: Literature Review

a. This chapter will provide explanations of parameters that could affect heating or cooling demand and solar thermal potentials which are:
    i. Primary parameters such as: temperature, solar radiation, relative humidity, and wind speed.
    ii. Secondary derived parameters such as: derivation of solar radiation (global, direct, diffuse, and its geometry adjustment: normal, tilted, horizontal), derivation of temperature (dew point temperature, heating degree day, cooling degree day)
b. Weather databases overview in which each database is divided into two categories while detailing the summary and its source of:
    i. Satellite based databases such as: NASA SSE, PVGIS, ESRA, HC3
    ii. Ground weather station based databases such as: Meteonorm, IWEC, TMY
c. Methodology outlook which is an overview of the methodology used to produce the weather databases.
d. Climate classification overview which briefly explains climate classification and is divided into Köppen Climate Classification and Briggs Climate Classification.

3. Chapter 3: comparison and analysis, will compare different weather databases, methodologies, climate classifications which were previously mentioned, and analyze them. Elements of chapter 3 will subsequently lead into chapter 4: results and discussion.
4. Chapter 4: results and discussion, will present the outcome of comparison in chapter 3.

5. Chapter 5: conclusion, will present the outcome of this research as well as some recommendations and barriers. Overcoming the barriers of this research is also briefly discussed.

Additionally, the appendix section is created to accommodate very large tables, figures and maps that otherwise will not fit into the main text. The appendix still cannot be constructed in a similar fashion with the excel database. So the tables are broken and separated into several tables due to space limitation. Several erroneous details were also left out due to space constraints.

2. Literature Review

2.1 Influencing Parameter

2.1.1 Primary Parameters

2.1.1.1 Dry Bulb temperature

This is a parameter mostly referred to by people when they speak about temperature. Actually there are two other kinds of temperatures that give the information about the atmospheric state aside from dry bulb temperature, which are: wet bulb temperature and dew point temperature. Dry bulb temperature is considered as the most valid parameter to represent the state of temperature because it only measures ambient air temperature. The temperature is supposedly exposed to the atmosphere, but shielded against radiation and moisture. An ordinary thermometer is able to produce this parameter.

If there is a dry bulb temperature, there is also a wet bulb temperature. Wet bulb temperature is a measurement where the thermometer's bulb is covered by a wet cloth as its name suggest. The need to know this parameter is because it could suggest the minimum temperature that could be achieved by solely evaporative cooling. Although not considered as a weather parameter (and usually not included in weather databases), this parameter helps to indicate the moisture content in its surrounding air. If the humidity in its surrounding air is high, heat will be trapped more within the wet cloth and thus the temperature is higher than dry bulb temperature measurement. Subsequently, if the humidity of its surrounding air is low, heat will be easily released towards the air by evaporative cooling and thus temperature will be lower than the dry bulb temperature measurement. The difference between wet bulb temperature and dry bulb temperature is called wet bulb depression.

There is another indirect indication of humidity in the form of temperature, which is dew point temperature. It indicates at which point water vapor starts to condense. Above this temperature, water vapor will stay in gaseous form. Dew point temperature is also considered as an indirect indicator towards relative humidity because when the moisture content in the air is high, the dew point temperature is also high, and vice versa. The dew point temperature can be observed with an ice cube within a metal container. When the vapor starts to condense in the container, it indicates the dew point temperature.

The temperature measurement for weather databases is certainly not measured using normal bulb thermometer. USCRN uses PRT (Platinum Resistance Thermometer) to measure temperature (USCRN, 2011). Other possible temperature sensors are thermocouple and thermistor. This is done because to make databases, it should produce electronic signals to be able to automatically log into databases. Manual bulb temperature that requires human force to log cannot be considered in this case.

2.1.1.2 Relative humidity

Relative humidity is an indicator of how much moisture is in the air relative to what could be held (at a maximum amount) at a given temperature (Nave, 2011). Mathematically, it can be denoted as $RH(\%) = \frac{actual\ vapor\ density}{saturated\ vapor\ density}.100\%$. When the air cannot hold the moisture, it will be condensed as dew. Therefore, this is directly related to temperature, especially to dew point temperature that indicates at which point the moisture starts to condense.

The relation between temperature, dew point temperature, and relative humidity is illustrated in figure 2.1 below.

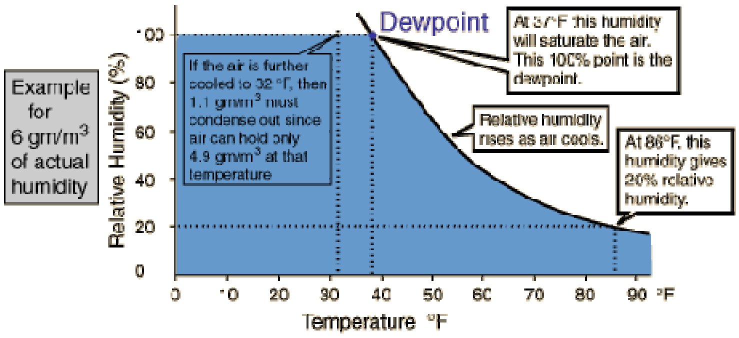


Figure 2.1: Relative humidity-temperature relation (Nave, 2011)

The conversion from relative humidity and dew point temperature is nonlinear. So it is too complex to be done without a calculator or computer assistance. However, (Lawrence, 2005) has found a simpler principle of approximation of relative humidity-dew point temperature conversion, which does not appear to be widely known by the meteorological community. He takes the linear section of the temperature-humidity chart to establish the simple linearity. More information about this matter will be further discussed in chapter 4: results and discussion.

Humidity is a prominent factor of heating or cooling demand because people tend to be sensitive to humid air. The human body uses evaporative cooling to regulate temperature. Thus, when humidity is high, evaporative cooling is slower. This means that the cooling demand is thought to be slightly higher even if the temperature stays the same.

This is also thought to be critical for an absorption cooling device that uses evaporative cooling as one of the steps to transfer heat to its surrounding area. High humidity is postulated to hinder the performance of the cooling device for the same reason mentioned above.

2.1.1.3 Solar Radiation

Basically, there are two ways to represent solar radiation, which are as an energy unit (kWh or Joule) or as a power unit (W). The parameters will need to have a time or area division to represent it in a more favorable manner. A division per time constant is needed to represent the duration of the timeframe while a division per area constant is required to know the range of energy received. In most cases, it is even more precise to represent it in a unit of energy or power per unit of area per unit of time (kWh/m^2.day). Solar irradiance refers to power (instantaneous exposure to solar radiation energy), while solar irradiation refers to energy (Renewable Training, 2010). From this point, there are several parameters that can be derived simply by varying the energy or power divided by time and/or area constant. Units can be easily converted into a suitable representative of parameter by the above method. In this case there is no agreeable primary unit of measurement.

On the other hand, solar radiation measurements are not so easily compared to temperature measurement. A Pyranometer is a device used to measure broadband (direct plus diffuse, hence global) solar radiation in a planar surface. A Pyrheliometer is a device used to measure direct solar radiation. So, a tracker is required for the Pyrheliometer to enable it to measure direct radiation.

2.1.1.4 Wind Speed

Wind speed could also be another prominent meteorological factor. Wind is a prominent factor in driving off humidity into or out of a region. (Nkemdirim, 1991) noted that his evaporation model is better when he included wind speed factor in his equation, instead of a mere temperature and relative humidity. This is an important point because evaporative cooling is a prominent factor to determine heating and cooling potential. More information regarding this can be found in chapter 4: results and discussion.

On a side note, it is already well known that wind blows from an area that has higher pressure to an area that has lower pressure. Pressure difference is created by temperature. An area that has higher temperatures will create a low pressure area and subsequently an area with low temperatures will create high pressure. This relation has been denoted by Gay Lussac's law as follows:

$\frac{P}{T} = k$, where k is a constant, P is pressure, T is Temperature. This explains why wind blows. Therefore, higher temperature differences between regions could proportionally relate to wind speed.

2.1.2 Derived Parameters

2.1.2.1 Degree Day

The foundation of this degree day estimation concept was originated from Lt. Gen Sir Richard Strachey in 1878. His original work dealt with crop growth. The basic terminologies and parameters that were used recently such as "degree-days" and "base temperature" appear to be coined from his work. Afterwards, several other publications followed and customized the concept in order to be used in the building, heating and cooling fields.

When a heating or cooling device is used, energy demand will vary according to how cold or hot the weather is. The fact that weather varies within years, months, weeks, days, or even hours makes averaging degree per time interval an inaccurate estimation of energy requirements. For example, the temperature measurements from April of this year do not produce the same mean temperature with April from last year. In some cases, when the outside air temperature is about the same with the inside temperature of a building (or slightly deviated because of internal heat gain), heating or cooling is not needed. The parameter above will be called *base temperature*. If the outside temperature is above or below the base temperature, heating and cooling is needed. The heating energy requirement (or cooling) will be in proportion with the temperature deficit in degree. If the difference between the outside temperature and base temperature is high, more energy is needed to heat up or to cool down. That cumulative deficit over a month is called *degree days.* In other words this parameter is simply a representation of temperatures (degree) multiplied by a time period. Whenever needed, this could also be improved into *degree-hours* for some more precise data. As has been discussed above, degree days show proportional relationships with energy demand. Thus a month with 240 degree-days is expected to spend two times more heating (or cooling) energy expenses than a month with 120 degree-days.

2.1.2.2 Solar Radiation Derivation

Depending on sun-plane geometry and its ray form, radiation can basically be derived into more varied parameters (SoDa, 2011). The matrix table below explains that there could be around 11 forms of radiation depending on its geometry and ray form.

	Horizontal Plane	*Inclined Plane*	*Plane normal to sun rays*
Global	A	B	C
Direct	D	E	F
Diffuse	G	H	I
Reflected	-	K	L

Table 2.1: Various form of solar radiation depend on its rays and geometry position

On a horizontal plane as shown in figure 2.2 below, the global radiation (A) could be divided into a direct (D) and diffuse part (G). On an inclined plane, the global radiation (B) is a summation of the direct (E), diffuse (H), and reflected (K) parts. On a plane normal to sun rays (this geometry is optimized for optimum direct sun rays), global radiation (C) is a summation of direct (F), diffuse (I) and reflected (L) rays. Note that in a horizontal plane, the reflected part is not expected and thus assumed to be involved in diffuse rays. It is also worth noting that the most commonly used parameters are global horizontal (A), direct normal (F), and diffuse horizontal (G). This will be further proved below in a database comparison. In addition, there could be parameters that quantify radiation on top of atmosphere/extraterrestrial radiation. However this scenario is uncommon. The reflected part is usually neglected because it is used less.

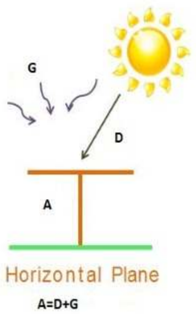


Figure 2.2: Horizontal Plane (SoDa, 2011)

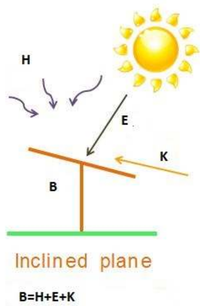


Figure 2.3: Inclined Plane (SoDa, 2011)

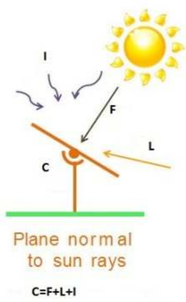


Figure 2.4: Plane normal to sun rays (SoDa, 2011)

2.1.2.2.1 Global Horizontal Radiation

Global Horizontal Radiation is usually the primary parameter if the databases use ground weather stations as their primary data gathering method. Then, by knowing global horizontal radiation, it is possible to know the other two major derived parameters, which are direct normal and diffuse horizontal radiation. Splitting these parameters will be further discussed in the methodology section below.

Global horizontal radiation is the radiation received on the surface of the earth that includes its diffuse and direct component. Due to the fact that global horizontal radiation should be measured on the surface, meteorological satellites normally cannot give direct measurements of this parameter. They can only derive the information from other meteorological parameters such as cloud cover.

2.1.2.2.2 Direct Normal Radiation

Direct normal radiation is the radiation received on the surface of the earth perpendicular to the direction of sun rays. In designing solar thermal or PV systems, this parameter is the best case and what is sought to be optimized by systems. Giving trackers to the solar collector or PV is one way to ensure a steady input of direct normal radiation.

2.1.2.2.3 Diffuse Horizontal Radiation

Diffuse horizontal radiation is the radiation reaching the earth’s surface after being scattered by molecules in the atmosphere. It is thought to significantly contribute to temperature and thus heating and cooling demand. But this is not as much as contributing factor comparing to direct radiation.

2.1.2.3 Sol-Air Heating and Cooling Temperature and Degree Days

(Erbs, Klein, & Beckman, 1984) formulated a parameter that incorporated temperature and radiation. Suppose a radiation incident falls on an opaque wall of unit area, Sol-Air temperature could be denoted as $T_{sa} = T_a + \frac{I_T \alpha_s}{h_o}$, where:

T_{sa}= Sol-Air Temperature

T_a= Ambient Temperature

I_T= Solar radiation absorbed in an opaque wall section

α_s= Solar absorptance

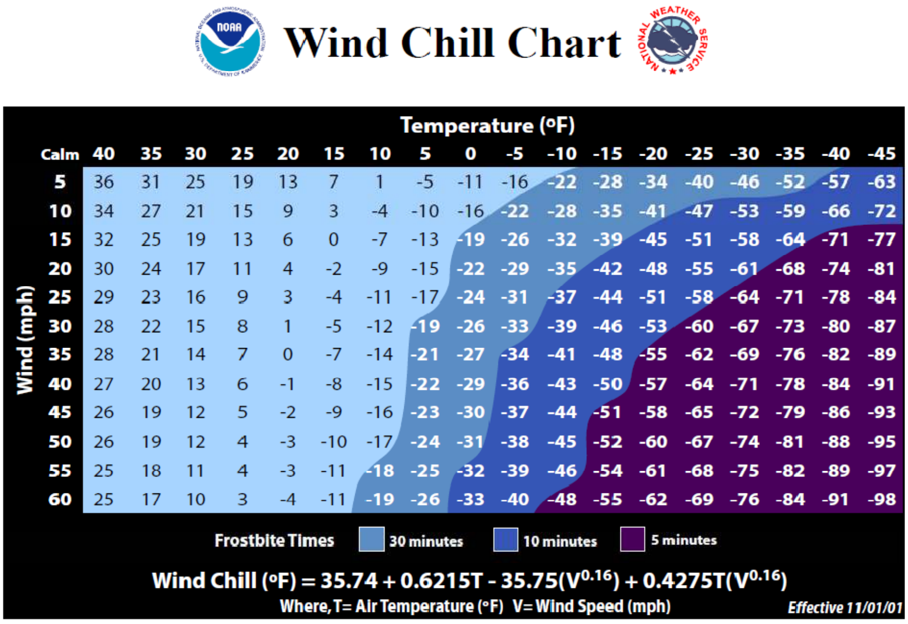
h_o= heat transfer coefficient between wall and its surrounding

This parameter is supposedly a fictitious (not pure meteorological parameter) temperature of surrounding the opaque wall that produce the same heat transfer rate without absorbing solar radiation. Sol-Air degree days are subsequently an integration of difference between Sol-Air base temperatures with Sol-Air over a time period. Sol-Air base temperature is the value of Sol-Air temperature at the point where heating or cooling is required (much about the same concept of base temperature in normal degree days in chapter 2.1.2.1 above)

The purpose and usage of this parameter is to equip building simulations with better tools to estimate heating or cooling demand. Knowledge about combined effect of temperature and solar radiation is a better approach to control net energy exchange between a building and its environment.

2.1.2.4 Wind Chill

Wind chill is the effect of wind on human skin that affects the temperature felt by people (not real measured temperature). For example, if the actual temperature measurement from a thermometer shows 10^{o}C while the wind is blowing at 5 km/h, the temperature felt in this condition might be much lower than 10^{o}C. Figure 2.5 below is a wind chill chart released by NOAA which helps to quickly determine wind chill effects.



Calm	40	35	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45
5	36	31	25	19	13	7	1	-5	-11	-16	-22	-28	-34	-40	-46	-52	-57	-63
10	34	27	21	15	9	3	-4	-10	-16	-22	-28	-35	-41	-47	-53	-59	-66	-72
15	32	25	19	13	6	0	-7	-13	-19	-26	-32	-39	-45	-51	-58	-64	-71	-77
20	30	24	17	11	4	-2	-9	-15	-22	-29	-35	-42	-48	-55	-61	-68	-74	-81
25	29	23	16	9	3	-4	-11	-17	-24	-31	-37	-44	-51	-58	-64	-71	-78	-84
30	28	22	15	8	1	-5	-12	-19	-26	-33	-39	-46	-53	-60	-67	-73	-80	-87
35	28	21	14	7	0	-7	-14	-21	-27	-34	-41	-48	-55	-62	-69	-76	-82	-89
40	27	20	13	6	-1	-8	-15	-22	-29	-36	-43	-50	-57	-64	-71	-78	-84	-91
45	26	19	12	5	-2	-9	-16	-23	-30	-37	-44	-51	-58	-65	-72	-79	-86	-93
50	26	19	12	4	-3	-10	-17	-24	-31	-38	-45	-52	-60	-67	-74	-81	-88	-95
55	25	18	11	4	-3	-11	-18	-25	-32	-39	-46	-54	-61	-68	-75	-82	-89	-97
60	25	17	10	3	-4	-11	-19	-26	-33	-40	-48	-55	-62	-69	-76	-84	-91	-98

Figure 2.5: Wind Chill Effect Chart (NOAA, 2011)

2.1.2.5 Evaporation

As was previously mentioned, (Nkemdirim, 1991) successfully created an algorithm to relate temperature, wind speed, and relative humidity as the driving force behind evaporation as follows:

$$E = 0.045\exp\left(0.35T + 0.025u - 0.133(e^{*} - e_{a})\right),$$

Where:

E = Evaporation

T = Temperature

u = wind speed

e^* = saturation vapor pressure

e_a = actual vapor pressure

Evaporation helps to reduce the cooling demand in the evaporative cooling concept. It has been explained that the human body uses the evaporative cooling concept. This concept however is not enough to model the temperature regulation of the human body. According to (Nave, 2011), other concepts such as radiation, conduction, and convection also play important roles in regulating temperatures of the human body aside from evaporative cooling. While conduction and convection rely on temperature differences to successfully regulate human body temperatures, evaporative cooling is one concept that relates wind speed and relative humidity as another regulator of human body temperature.

2.2 Weather Database Overview

The classification of weather databases and whether they are classified as satellite based or ground station based is not quite accurate. In producing databases, several instances of mixing and matching will still occur in a case where there is no ground station data, satellite data could be used, and vice versa. Classification below into ground based or satellite based only serves as a starting point of overview towards weather databases on how they are obtained. More elaborative classification could be found in chapter 4: results and discussion.

Usually, in ground weather station based databases, they will provide their own station list which has information about each station. This is done because of the difficulty in deciding uniformly the behavior and accuracy of ground weather databases.

2.2.1 Meteorological Satellite based database

2.2.1.1 NASA SSE

It is common knowledge that NASA is an Aeronautics and Space Agency of United Space of America. NASA is well known in the meteorological field. In this case, NASA issues its collected global meteorological data in the form of SSE (Surface meteorology and Solar Energy) (NASA, 2009). In 2009, NASA released version 6 which is currently the most accurate version. As a NASA product, they have several advantages. Firstly, being satellite based, the databases cover a global area. Weather data can be retrieved from virtually any site on Earth. Secondly, it is free to access.

Having global coverage and storing 22 plus years naturally creates a large amount of data archives. In this case, the dataset is not real, but an averaged value over the past 22 years. Data itself does not come directly from measurement. Global horizontal radiation data for example comes from other NASA projects called SRB (Surface Radiation Budged). In conclusion, the released database is simply an attempt to elaborate other NASA projects into one single elaborate database. Several other parameters are derived and not measured directly. This point is significant because it will affect its accuracy and will be discussed in chapter 3: comparison and analysis.

2.2.1.2 PVGIS

PVGIS, alongside with ESRA, HC3, and SoDa (not discussed here) is another set of prominent groups of satellite weather databases coming from different sources. The groups share some similarities. The similarities are mainly due to a product from joint research from the European Commission as part of the SOLAREC project. Primary data of PVGIS itself comes from ESRA for the European continent while HC3 supplies primary data for the African continent. Due to primary data differences, there are also differences in their time period span. European dataset is collected during the period 1981-1990 (9 years), while the African dataset is collected during the period of 1985-2004(19 years). Additionally, because of its reliance on European Satellite Heliosat/EUMETSAT, the coverage cannot be worldwide, unlike NASA SSE.

As the name describes, PVGIS is aimed toward GIS (Geographical Information System) application. Databases only contain information about temperature and solar radiation, while other less important parameters such as relative humidity and wind speed are being left out. PVGIS, like NASA SSE, is freely available in the internet as a web application.

2.2.1.3 ESRA

As was discussed above, ESRA shares some similar traits to PVGIS and HC3 mainly because the primary data itself comes from HC3 while ESRA data is being taken for PVGIS database. ESRA is also a joint European project. However this project is under the framework of JOULE II Programme (Scharmer & Greif, 2000). The institution working in this ESRA project, PVGIS, HC3, and SoDa is mostly the same. Unlike NASA and PVGIS, it comes in the form of a CD-ROM and Atlas which needs to be purchased. Coverage is only in Europe. However, ESRA complements their primary satellite data with around 600 ground weather stations around Europe. This case brings the consequence of methodological disparities of each ground station. Several ground stations could provide daily data while others provide monthly data. Several ground stations have TRY (Typical Reference Year) as their typical year generation method (Typical year generation method will be further discussed in section 2.3.2 Typical Year Generation below). Other stations have DRY (Design Reference Year). Unlike Meteonorm, the difference is kept and the user can select their desired adjustment in the software distributed in their CD-ROM.

2.2.1.4 HC3

HC3 (HelioClimate 3) is another weather database developed by The Center for Energy and Processes (www-cep.cma.fr), a joint research laboratory of the French school of engineers MINES ParisTech (www.mines-paristech.eu) and framework association for school of engineers for research activities directed to the industry, ARMINES, (www.armines.net). HC3 is based on meteorological satellite EUMETSAT. Although using satellites, the data is also benchmarked against 29 ground stations located mainly in Europe with several stations in Africa. There is only one station in Middle East (Israel). Similar to TMY, a numeric "3" behind HC signifies that it is the third update of the database. HelioClimate is a database that is most often referred by others. HC1 (first version of it) is used as a primary database for PVGIS for the Mediterranean basin, Africa, and South-West Asia. It is also referred by ESRA. This signifies its validity as a database that is often referred by other databases.

2.2.2 Ground Weather Station based database

2.2.2.1 TMY3

TMY3 (Typical Meteorological Year) is a weather database based on ground weather stations developed by NCDC (National Climatic Data Center), USA (Wilcox, 2008). It can cover the USA including Puerto Rico, Guam, and the Virgin Islands. The number "3" behind "TMY" signifies that this is the third update. There are previously TMY, and TMY2 database. Previous data sets (TMY and TMY2) claim that this will not work interchangeably because of several differences. The notable differences lie in time (solar versus local), formats, elements, and units. Furthermore, NCDC also puts a disclaimer on its user manual that TMY should not be used to predict weather for a particular period of time nor is it an appropriate basis for evaluating real-time energy production or efficiencies for building design applications or solar conversion systems.

2.2.2.2 IWEC

IWEC is basically TMY for international location. Unfortunately, it cannot cover the entire globe unlike NASA SSE. More specific information about country coverage can be seen in appendix C1. Development procedure and typical year generation is mostly the same with TMY. More information about this topic will be discussed in chapter 3: comparison and analysis. ASHRAE (American Society of Heating, Refrigeration, and Air Condition Engineering) is responsible for the development of the database (Thenevard & Brunger, 2001). A rather different aspect compared to TMY is that IWEC usually uses derivation parameters, instead of direct measurements which come primarily from their own solar radiation data and is derived from various methodologies. The methodology used will be discussed in this chapter, section 2.3 below.

2.2.2.3 Meteonorm

Meteonorm is an unusual database, compared to another database mentioned above. It was produced by METEOTEST, which is like a corporation. Other databases like those mentioned above are commonly from government agencies, educational institution, or common nonprofit institutions (Meteonorm, 2010). As such, Meteonorm is designed with better properties rather than other databases mentioned. Even though Meteonorm is ground station based, a spatial interpolation method is employed to give the database larger coverage worldwide. Meteonorm takes significant portions of regional and national level meteorological data and compiles it into a single, elaborate database from all over the world. Meteorological parameters included are also quite complete, making it on par with TMY, IWEC, and NASA SSE.

One drawback is that they receive data from all over the world which implies it comes with different assumptions and methodologies. Meteonorm only takes monthly data and then interprets it according to the need. More elaborative explanations on this subject can be found in section 3: comparison and analysis.

2.2.2.4 National level databases

On a national level, there are databases issued by local governments and scientific or educational institutes to fulfill the need of weather databases on a national level. Examples ranging from CWEC (Canadian Weather for Energy Calculation, Canada), CIBSE (Chartered Institution of Building Services Engineer, UK), CSWD (Chinese Standard Weather Data), CTWY (Chinese Typical Year Weather), ETMY (Egyptian Typical Meteorological Year), IGDG (Italia Gianni de Giorgio, Italy), IMS (Israel Weather Data), and so on. The advantage is it can be tailor made to the procedure in order to suit their current national location.

2.3 Methodology

2.3.1 Post Process Method

2.3.1.1 Splitting Global Radiation into Direct and Diffuse Component

The algorithm developed by (Liu & Jordan, 1960) is the first and widely used methodology to split global horizontal into their direct and diffuse components. HC3 employs the method to find the direct and diffuse component of their radiation. IWEC also uses this method since the Kasten model they used to produce radiation under cloudy skies (will be discussed below) only calculates global solar radiation. Thus, the direct and diffuse components are not available.

The basic principle of (Liu & Jordan, 1960) is to introduce a dimensionless coefficient, namely transmission coefficient for global horizontal radiation (T_t), direct solar radiation (T_D), and transmission coefficient for diffuse solar radiation (T_d).

$T_D = \frac{I_{Dh}}{I_{on}}$, where I_{Dh}= Direct radiation at horizontal plane; and I_{on}= Top of Atmosphere horizontal radiation,

$T_d = \frac{I_{dh}}{I_{on}}$, where I_{dh}= Diffuse radiation at horizontal plane; and I_{on}= Top of Atmosphere horizontal radiation,

$T_t = \frac{I_{Th}}{I_{on}}$, where I_{Th}= Global Horizontal Radiation; and I_{on}= Top of Atmosphere horizontal radiation,

Linear regression is applied to the T_d and T_D coefficients which are based on real measurements to obtain: firstly a relation between T_d and T_D, secondly another linear regression is performed to obtain a relation between T_d and T_t. After the two steps, a relation between T_d, T_D, and T_t can be achieved.

However, the above equations only work for clear skies. For cloudy skyies, there should be additional parameters to adjust the relations of radiation and their direct and diffuse components.

2.3.1.2 Stochastic Time Generation

In one irregular case, Meteonorm stores monthly data as their primary dataset. Storing monthly data in the case of Meteonorm is understandable since it is a database that incorporates weather data all over the world. Storing smaller time increments such as daily or hourly increments supposedly is space consuming. Storing monthly data is considered to be simpler. However, it is quite common of software modeling to request smaller time increments. So, Meteonorm is equipped with stochastic time generation to fulfill this need. When it is a conversion from a smaller time increment (hourly, daily) into a larger one (monthly), it implies that those are simple procedures and will thusly not to be discussed.

(Aguiar, Collares-Perreira, & Conde, 1988), has developed a procedure to generate a daily radiation value from monthly values using a library of Markov Transition Matrices. The

Markov Transition Matrices (MTM) was already generated from 300 months from five locations throughout the world. The procedure is subsequently adopted by Meteonorm to convert their monthly data into a synthetic daily series.

However, the same method does not work for the conversion of daily increments into hourly increments. The MTM was deemed unsuitable for this case by (Aguiar & Collares-Perreira, 1992), as their inspection of MTM in different places showed that a similar corresponding transition probability value could not be determined by this method. So, the author mentioned above (Aguiar & Collares-Perreira, 1992) a proposed solution involving the Auto Regressive Moving Average Model (ARMA) with inverse Gaussian mapping and time dependence, abbreviated as TAG (Time dependant, Autoregressive, Gaussian) model. This procedure is also subsequently adopted by Meteonorm to convert their already converted daily value into hourly values.

While the advantage of this method is already explained above, it permits smaller storage data capacity. The disadvantage is clear due to its lack of accuracy. More information about this will be discussed more in chapter 3: comparison and analysis regarding its accuracy and time increment effect.

2.3.1.3 Spatial Interpolation

This method is important in a ground based weather database to further extend their coverage, especially Meteonorm. It has been explained in (METEONORM, 2010) that one ground weather station has 50 km of coverage radius. With PVGIS' temperature measurement, they also use spatial interpolation because their data comes from ground weather measurements in this case. Basically, space interpolation is a method to determine a value in one site that is not measured directly by sensors. Meteonorm incorporates (Wald & Lefevre, 2001) method and (Zelenka, 1992) method. PVGIS uses (Huld, Süri, Dunlop, & Micale, 2006) the method to interpolate their temperature value.

For the Meteonorm case, this following algorithm is the basic operational principle of space interpolation. A meteorological parameter (for example: global radiation) at point x ($G_h(x)$) is a factor of several nearby measured values ($G_h(x_i)$, altitude of site x, and weight. Weight is a factor of several parameters, including search radius (max 2000 km), horizontal distance, and vertical scale factor. The procedure makes Meteonorm able to determine a parameter without having real measurements at certain points if several weather stations exist around 2000 km distance from its point.

2.3.1.4 Clear Sky Radiation and Cloudy Sky Radiation

The measurement of radiation directly by equipment such as pyranometers or pyrheliometers is difficult. As a result, many weather databases choose to generate radiation models. Measuring is possible in clear sky conditions. Weather databases generate radiation models because of the amount of radiation received in the top of the atmosphere is predictable. In order to predict the radiation in the atmosphere, latitude and longitude factors are considered, thus forming solar altitude and Julian day as a time factor. Julian day (j) is a time system designed to compute the geometric position of the sun and the Earth.

So, global horizontal radiation on top of atmosphere, G_o is expressed as:

$$G_o = \varepsilon . 1367 . \sin \gamma_s \, {}^{W}/_{m^2} ,$$

Where:

γ_S is a solar altitude as shown in figure 2.7 below and a factor of latitude, longitude, and time in hourly format.

ε is a correction factor of solar distance that follows this formula:

$$\varepsilon = 1 + 0.0334\cos((j - 2.8^o)),$$

Where j is Julian day.

Additionally, the constant of 1,367 shown in global horizontal equation above is solar constants to estimate that Earth receives 1,367 W/m^2 constantly, as depicted in this following figure 2.6 below:

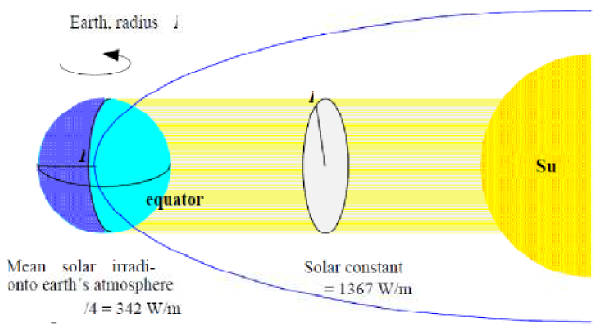


Figure 2.6: Available extraterrestrial mean of solar irradiance on top of atmosphere (Scharmer & Greif, 2000)

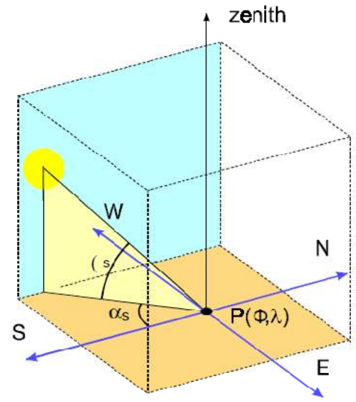


Figure 2.7: Solar Altitude angle and Azimuth angle seen from and observer point of view P (Scharmer & Greif, 2000)

That following algorithm explained above refers to the ESRA procedure (Scharmer & Greif, 2000). However, other databases such as Meteonorm, IWEC, NASA SSE, and HC3 also follow the same structure.

Obtaining the top of atmosphere data alone is not adequate to determine radiation on a surface level. Clear sky radiation has to be determined by first having the top of atmosphere radiation data. In a clear sky model, usually (Rigollier, Olivier, & Lucien, 2000) is used by most of major database such as ESRA, Meteonorm, SoDa (Remund, Wald, & Page, Chain of Algorithms to Calculate Advanced Radiation Parameters, 2003). IWEC implements another method, namely METSTAT (Maxwell, 1998) method to calculate its clear sky and subsequently adopts the Kasten method in its cloudy sky model.

It could be summarized that to convert top of atmosphere radiation into surface radiation (which is a more usable parameter) depends on atmospheric attenuation (scattering, absorption) given by (Šúri & Hofierka, 2004):

1. Gases (air molecules, CO_2,O_2, and ozone)

2. Solid and liquid particles (aerosols, including non-condensed water)
3. Clouds (condensed water)

The first factor above influences the Rayleigh optical thickness and relative optical air mass variables, while the second factor mentioned above influences the Linke Turbidity variable. Clear sky radiation could be derived knowing the first and second factors above. Knowing the third factor (cloud cover, cloud amount, opacity, etc) will result in cloudy sky radiation.

2.3.2 Typical Year Generation Method

Typical year generation is an attempt to generate a "typical" year series from a set of databases. The synthetic series is usually required for solar energy system simulation and is more preferable than the real series due to the elimination of extreme conditions. In practice, all January during observational the time span are examined and judged which one is the most typical according to the category. This is done with other months as well, creating a year that is unreal but typical.

The earliest attempt to produce a typical year resulted in TRY (Typical Reference Year) at 1976 by NCDC (Crawley, 1998). A weakness of TRY is due to the fact that the method still resulted in a historical year. Within this method, months containing extreme values are progressively eliminated until one particular year that is considered as a "mild year" remains. As well, its parameter neglected solar radiation as a main criterion. This results in a problem when TRY is inputted to simulation software that requires solar radiation as an input. The software cannot generate radiation values which naturally will result in error. TRY is still evolving as noted by (Bilbao, Miguel, Franco, & Ayuso, 2003), there are TRY4 (Pissimanis method), TRY5

(Lund method, in some other cases this is called DRY and will be discussed much below), TRY6 (Argiriou Method). CIBSE, a weather database from United Kingdom, still uses this method to generate their typical year. The difference is that they complement it with the DSY (Design Summer Year) method. Additionally, several ESRA ground stations still use TRY methods, thus part of ESRA also contains this method.

Realizing the limitations of TRY in several areas, NCDC created new procedures and datasets called TMY (Typical Meteorological Year). The procedure is used to generate the typical year and is called Sandia method. While several databases already use TMY with their own customized procedure, the basic procedure on how to generate TMY as defined by (ASHRAE Technical Committee 4.2, 2011) is described:

1. For each month in the climate record, calculate the daily means for each index. Indices generally include temperature and solar radiation, and (with lower weights) humidity and wind speed. Weight may vary between methods.
2. For each calendar month, determine the CDF (Cumulative Distribution Function) of the daily means, sorting the values in rank order.
3. For all the years in the data set, calculate CDF of the daily means.
4. Calculate the F-S for each month and select five months using a weighted sum of the F-S statistics.
5. Rank the candidate months with respect to the closeness of the month to the long term mean and median.
6. Use persistence criteria to exclude months with the longest run of temperature, the month with the most runs, and the month with zero runs.

7. Concatenate the 12 selected months by smoothing the six hours on each side of the transition between months to eliminate discontinuities.

TMY as a database itself has already reached version three, denoted earlier in this paper as TMY3. TMY 3, TMY2, and the original TMY differ slightly on their weighting and procedure. Another database using the TMY procedure is IWEC. More information about this slight difference procedure will be discussed in chapter 3: comparison and analysis.

While (Bilbao, Miguel, Franco, & Ayuso, 2003) note a particular procedure as TRY5, a study on the real paper (Lund, 1995) reckon that this procedure is called DRY (Design Reference Year). Procedure of generating DRY is generally almost the same with TMY and TRY, which are:

1. Selection process
2. Adjustment process
3. Derivation of missing parameter

Selection process is what differs much with TMY, because DRY is more qualitative procedure rather than quantitative. While retain TRY method of criteria, DRY propose rather different method of selection. Each month being observed is given a qualification of whether they are “Qualified,” “Acceptable,” “Poor,” or “Impossible.” This division is based on numbers of flags and parameters.

2.4 Climate Classification Method

Two climate classifications being discussed in this paper is Köppen (Kottek, 2006) climate classification, and Briggs (Briggs, 2002) climate classification. Köppen is used by IWEC, TMY, and IGDG, while NASA SSE prefers to use the Briggs model. Meteonorm uses theTroll and Paffen model, which is not discussed here.

2.4.1 Köppen Climate Classification

The most well known and widely used climate classification system is Vladimir Köppen's work from 1900. It has been updated several times, while the most prominent version used now is the latest version from Rudolf Geiger in 1961. Köppen was trained as a plant physiologist and realized that plants are indicators for many climatic elements (Kottek, 2006). Therefore, the classification is mainly distinguished by vegetation group. As the classification is developed further, it is realized that the main factor towards vegetation group is precipitation and temperature. Therefore, the classification is basically done from those two main parameters.

2.4.2 Briggs Climate Classification

In Briggs's paper (Briggs, 2002), the prime mover is not vegetation, but energy efficiency measures in building is now the main concern. By this classification, the primary concern is no longer temperature, but HDD (Heating Degree Days) and CDD (Cooling Degree Days), which relate in direct proportion to the heating or cooling demand of a building. Certainly HDD and CDD are derived values from temperature which will vary from one region to another based on their base temperature preference. Precipitation is still a main differentiated parameter as it is in Köppen's work. However, it is not differentiated under dry and humid. Rather, Briggs adds one more type which is marine. Classification is simpler in Briggs compared to Köppen because it is only a two-letter designation of degree days in first letter (1, 2, 3, …) and precipitation (A, B, C) as the second letter. The comparison between Köppen and Briggs is displayed in table 2.2 below.

Zone No.	Climate Zone Name and Type	Thermal Criteria[1,8]	Representative U.S. City*	Köppen Class.	Köppen Classification Description
1A	Very Hot – Humid	5000 < CDD10°C	Miami, FL	Aw	Tropical Wet-and-Dry
1B[7]	Very Hot – Dry	5000 < CDD10°C	---	BWh	Tropical Desert
2A	Hot – Humid	3500 < CDD10°C ≤ 5000	Houston, TX	Caf	Humid Subtropical (Warm Summer)
2B	Hot – Dry	3500 < CDD10°C ≤ 5000	Phoenix, AZ	BWh	Arid Subtropical
3A	Warm – Humid	2500 < CDD10°C ≤ 3500	Memphis, TN	Caf	Humid Subtropical (Warm Summer)
3B	Warm – Dry	2500 < CDD10°C ≤ 3500	El Paso, TX	BSk/BWh/H	Semiarid Middle Latitude/Arid Subtropical/Highlands
3C	Warm – Marine	HDD18°C ≤ 2000	San Francisco, CA	Cs	Dry Summer Subtropical (Mediterranean)
4A	Mixed – Humid	CDD10°C ≤ 2500 AND HDD18°C ≤ 3000	Baltimore, MD	Caf/Daf	Humid Subtropical/Humid Continental (Warm Summer)
4B	Mixed – Dry	CDD10°C ≤ 2500 AND HDD18°C ≤ 3000	Albuquerque, NM	BSk/BWh/H	Semiarid Middle Latitude/Arid Subtropical/Highlands
4C	Mixed – Marine	2000 < HDD18°C ≤ 3000	Salem, OR	Cb	Marine (Cool Summer)
5A	Cool – Humid	3000 < HDD18°C ≤ 4000	Chicago, IL	Daf	Humid Continental (Warm Summer)
5B	Cool – Dry	3000 < HDD18°C ≤ 4000	Boise, ID	BSk/H	Semiarid Middle Latitude/Highlands
5C[7]	Cool – Marine	3000 < HDD18°C ≤ 4000	---	Cfb	Marine (Cool Summer)
6A	Cold – Humid	4000 < HDD18°C ≤ 5000	Burlington, VT	Daf/Dbf	Humid Continental (Warm Summer/Cool Summer)
6B	Cold – Dry	4000 < HDD18°C ≤ 5000	Helena, MT	BSk/H	Semiarid Middle Latitude/Highlands
7	Very Cold	5000 < HDD18°C ≤ 7000	Duluth, MN	Dbf	Humid Continental (Cool Summer)
8	Subarctic	7000 < HDD18°C	Fairbanks, AK	Dcf	Subarctic

Table 2.2: Comparison of Briggs Climate Classification against its Köppen counterparts (Briggs, 2002)

3. Comparison and Analysis

One of main objectives of this research is to compare weather databases. Therefore, this chapter is central to the argumentation of the thesis. Comparisons are done by comparing user manuals from databases. Tables and figures are generated from self-made databases if there are no references attached. The table, in some cases is small enough and presentable in the main text. **A larger table is attached to the appendices because if presented within the main text the table will severely disrupt paper layout.*

3.1 Effect on Different Time Increment

The two major parameters that will be discussed in this paper differ greatly. Temperature estimation usually uses degree day, while solar radiation estimation still uses energy or power received per area per time scale such as W/m^2 year. While the latter is a considerably more appropriate parameter (it is easily comparable and easily observed), it also contributes to complexities in estimation. On the other hand, the degree day approach will provide better estimations since it already incorporates a time constant on its value. It must be noted that the degree day approach is not an appropriate parameter to be displayed in a report. As mentioned above, degree day is regarded as proportional with the energy consumption or requirement of a heating or cooling system (with a certain extent).

This difference might be caused by differences in the utilization of parameters. Temperature is a parameter that is proportional to how much energy is required /consumed to heat up or cool down a building. Conversely, solar irradiation is a parameter to measure how much energy is received. This explains why the amount of energy or power received per square meter per given time frame must be provided.

(Aguiar, Collares-Perreira, & Conde, 1988) have developed a simpler mathematical method to estimate daily future radiation received in a given area and time frame using the Markov Transition Matrix. Meanwhile, previously common methods using ARMA (Auto Regressive Moving Average) model is basically regarded as more complex with the requirement of a more precise dataset (daily or hourly data).

However, (Ransome & Funtan, 2005) have indicated that the approximate attempt might not be enough and in some cases misleading to estimate the energy received. An example is shown in figure 3.1 below. It is one point that could be used to argue against such an estimation method.

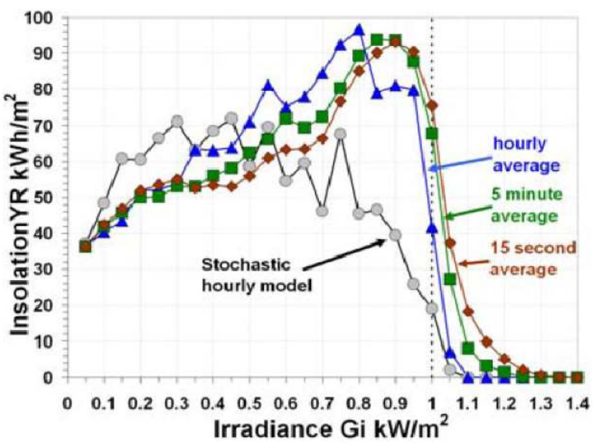


Figure 3.1: Plane of array insolation vs irradiance comparing stochastic hourly model to measured data (Ransome & Funtan, 2005)

It can be seen that the approximate model suggests that most of the radiation will occur below 0.6 kW/m^2. However, real measurements show that even at 0.8 kW/m^2 point, energy from radiation still can be well received. As the time interval of observation is increased, it is even shown that a larger proportion of energy occur at higher light level.

More frequent radiation measurements bring several consequences. On a positive note, it reveals better information as shown by (Ransome & Funtan, 2005) above. It is particularly important that the data is available to consider the inverter power rating. However, in some cases, the shorter timeframe might be considered to be redundant in planning and scaling a solar thermal system. In short, it is generally agreed that having lower scale measurements (more conservative number) will benefit the financial analysis. Furthermore, there is a serious problem in terms of solar thermal software modeling (TRNSYS, TSOL, etc). The requirement for data in time intervals within minutes is nearly non-existent. Most of them only require data in hourly format. So, a shorter time frame could produce a redundant dataset, of which several data is not needed and will be truncated by the modeling software. Mostly, monthly data is what is needed in estimating solar radiation. This part will be discussed further in section 3.3.4 typical year generation below. The effects of different time increments will be most apparent in the Meteonorm database because it employs stochastic time generation (from monthly, to daily and hourly).

On the other hand, it is not compulsory to have the same timeframe for a dataset in degree day method. It could be done in hourly, daily, or even monthly averages. Adding up all temperature variations over longer periods still represents the variation of temperature over the whole periods. In contrast, if the variations are averaged over a time period, it will only result in one parameter and will also indicate nothing about the variation. (CIBSE, 2006) has indicated that there are basically four major methods to calculate degree days. The methods are as follows:

1. Mean degree-hours
2. Meteorological equation, Daily outdoor maximum and minimum temperatures
3. Mean daily temperature
4. Mean monthly temperature, e.g. Hitchin's Formula

(CIBSE, 2006) also indicates that a smaller timeframe other than hour (possibly every minute) might be used, but little could be gained in terms of accuracy. This indicates a contrasting point to the solar radiance parameter displayed in (Ransome & Funtan, 2005), where timeframe difference matters.

3.2 Methodology used to Benchmark and Validate Database

There are several ways to evaluate models and thus express the validity of data. The simplest way is by comparing the measured dataset with its own mean value, thus forming a *standard deviation*. However, the most acceptable ways is usually done by comparing the dataset from measurement with another dataset that have better accuracy as a benchmark. Some validation goes even further as to invoke the distribution, whether it is fit to the benchmark dataset or not by *Kolgorov-Smirnov Test*.

Suppose the data from measurement is $x = \{x_1, x_2, \dots, x_n\}$ Mean value of x could be denoted as $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. Thus, a *standard deviation* of our measured dataset could be denoted as $s_n = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$. Meanwhile, if our benchmark dataset is supposedly denoted as $\hat{x} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n\}$, a *bias* of the estimator could be denoted as $E = [\hat{x} - x]$. The *Mean Bias*

Error (MBE) is thus could be denoted as $MBE = \frac{1}{N}\sum_{i=1}^{N} E_i$. The *Root Mean Square Error (RMSE)* is thus denoted as $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2}$. This indicates that the root mean square is still retaining its original unit of parameter being compared. However, a model with a root-mean-squared error of 0.25 is not much better than one with an RMSE of 0.5 for instance. It is because of the width of the confidence intervals that is proportional to the RMSE. Thus, to make the comparison more valid *NRMSE (Normalized Root Mean Square Error)* is often used. *NRMSE* is done by dividing *RMSE* by its maximum and minimum value. The unit of NRMSE is thus a percentage. *NRMSE* denotes as $NRMSE = \frac{RMSE}{x_{max} - x_{min}}$. Other means to evaluate is rMAE (Relative Mean Absolute Error) which is defined as $rMAE = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{x}_i - x_i|}{\hat{x}}$. As shown above, these validation methods require other datasets that are more reliable as a benchmark.

(METEONORM, 2010) explain this Kolgorov-Smirnov Test to review several of their measured dataset versus BSRN Network dataset, particularly for their stochastic time generation procedure. The methodology could be described as follows. Suppose a cumulative distribution function (CDF) of that x dataset and its subsequent benchmark dataset $\hat{x}$ could be denoted as $F(x_i)$ and $R(x_i)$, the distance between the two CDF could be denoted as $D_n = max|F(x_i) - R(x_i)|, x_i \in [x_i + (n+1)p, x_{min} + np]$, where the interval distance is $p = \frac{x_{max} - x_{min}}{100}$. Example of CDF difference result could be seen in figure 3.2 below.

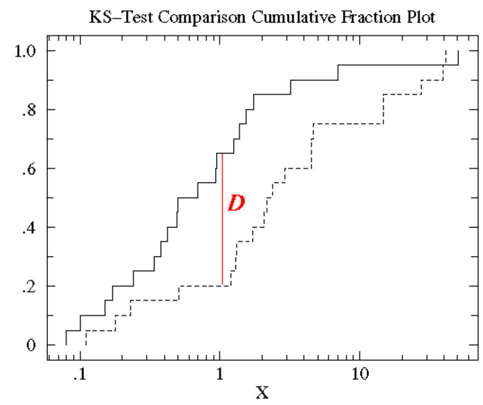


Figure 3.2: Result example of KS Test indicating point where CDF differs between database and its benchmarking value (College of St. Benedict and St. John University, 2011)

Then, if any of the intervals are considered, the distance is above the critical value, V_c (which depends on the population size, N), the null hypothesis that the sets are statistically the same must be rejected.

$$V_c = \frac{1.63}{\sqrt{N}}, N \geq 35$$

$$aux = \begin{cases} D_n - V_c, if\ D_n > V_c \\ 0, \quad if\ D_n \leq V_c \end{cases}$$

The KSI over %, which is the indicator of how fit the dataset towards the benchmark, is thus defined as:

$$KSI\ over\ \% = \frac{\int aux\ dx}{a_{critical} \cdot 100}$$

Where, $a_{critical} = V_c \cdot (x_{max} - x_{min})$

The databases commonly used in benchmarking are the BSRN (Baseline Solar Radiation Network, [global scope] which is only concerned with solar radiation), the USCRN (US Climate

Reference Network, USA only scope), and the NCDC (National Climatic Data Center, USA only scope).

According to the theory above, RMSE should retain the unit of parameter while NRMSE should use percentage. Nevertheless, it is found in several publications that RMSE does not use its actual unit, rather it uses percentages. The disparity needs to be verified with the producer of the weather database because it creates confusion.

It is clear that there is no general agreement on which methodology should be used. As a result, databases published with a validity of NRMSE 5% against BSRN cannot be considered to be more accurate than a database with rMAE of 7% against USCRN. Methodologies exist to suit the need of each database validation. For example, the KSI method is useful in measuring how far CDF of databases goes against benchmark CDF. This is particularly done in Meteonorm due to its generation of stochastic time sequences. Naturally, it is of their interest to know whether their generated sequence is accurate or not. Another problem found is the lack of uniformity throughout all locations. All data provided in this paper here has already been averaged over all reporting locations. If the second objective of this paper (measured error of database against proxy ground value) is attainable, it is still not yet valid to decide the accuracy of databases. It may be true within certain regions but not to other regions.

In general however, it could be inferred that benchmarking databases in ground stations, will have uniform procedures that produce consistent values so that the accuracies are most likely to be equally valid throughout all sites. RMSE (Root Mean Square Error) is the most used validation method in all databases. This will be used in chapter 3.4 as a main validation method to compare databases, despite its range of weaknesses.

3.3 Basic Production process of weather databases

The production process of a weather database is usually following these major steps showed in figure 3.3.

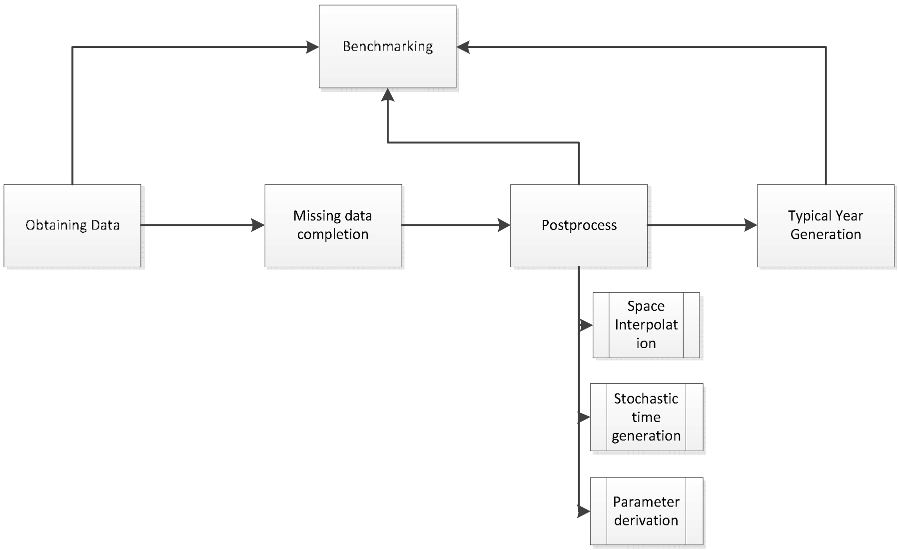


Figure 3.3: Basic Production Process of Weather Database

The first step is data collection. In this step, databases are divided into two major categories, which are meteorological satellite derived, and ground weather station derived database. Note that this category does not imply that all the data is obtained by satellite or ground stations. Mixing and matching will still occur. Meteonorm for example, whose primary data obtaining method is from ground stations, still need satellite data to obtain measurements where no weather station is around by a distance of 300 km. As a general rule, it is usually assumed that ground weather station data has better accuracy rather than satellite data (NASA, 2009). Specifically, this is highly debatable and depends on conditions and parameters of which the measurements are conducted. There is also a limitation for ground weather stations regarding their area coverage. At most, ground weather stations can only cover areas of 50 km within its radius (METEONORM, 2010).

In the second step there are occasionally missing data. So the second step deals with how the database fulfills the missing data. This is an important issue if the database is aiming to generate a typical year at the end of the steps (for example: IWEC, TMY). For databases that do not generate typical year and instead choose a mean value along the assessment period, this is not critical and can be left alone (for example: NASA SSE).

The third step is post processing data already received. This step includes space interpolation, time conversion (and stochastic time generation in some cases), and parameter derivation. Given that the data obtained in step one is valid, it could be used to derive other parameters as well. For example, if global horizontal radiation is obtained, it is usually possible to derive direct normal and diffuse horizontal data as well. Space interpolation is also needed at this point to obtain data outside its designated location for the ground weather station because of the above reason that a weather station could only cover 50 km within its radius. Time conversion is also matter in this case. Granted that the initial database value is usually a small time increment (minute or hour) while larger time increments are usually needed (day, or month), it is a not a complex task to convert up designated time increments. A peculiar one is Meteonorm database, because it contains monthly initial values. It argues that most users do not need smaller time increments, which is generally true. In such a case, Meteonorm is already equipped with stochastic time generation to convert down a longer time increment (month) into shorter one (day, hour). However, this method is at the expenses of its accuracy.

The fourth step is typical year generation. In some cases this part is omitted (NASA SSE, HelioClim3) because simpler methods exist. Simply averaging it over an observational time span also yields monthly values (mean, max, min), but this approach generates less accurate results when compared to typical year generation. On the other hand, some databases (IWEC, CIBSE, TMY3) put higher emphases on this because they realize that accuracy matters due to this data will usually be fed into simulation software. The methodology used in this step will vary between TMY, TRY, DRY, etc. Lastly, Meteonorm already has monthly values as its initial data. So in this case, this step might happen at a much earlier stage. For the Meteonorm cases, it is hard to determine the methodology that was used to generate its typical year because it is not mentioned in its manual; however, they do exist (see discrepancies between weather database and methodology). For convenient use, it is touted as “Meteonorm” method as compared to other established method such as TMY, DRY, TRY, etc. The Meteonorm method is probably a mix and match between those methods, simply because it receives monthly data directly from the weather station. The typical year method will depend on what the procedure is of that weather station. The event happens in ESRA as well because they already receive data in a monthly format from the weather station. Their typical year methodology varies between TRY, DRY, and BRY.

The next step is benchmarking. It cannot be placed as the last step because it must be done continuously within each step. For each step, it is usually imperative to benchmark its result with a more reliable database. It is done usually by subtracting the obtained value and the “real” value yielding a bias, error, or other similar parameter. The “real” value is obtained from other databases with higher accuracy, for example BSRN (Baseline Solar Radiation Network). This ground station solar measurement network guarantees 0,01% accuracy for solar radiation.

However, this is not true for the second step (missing data completion), because if there is no data to be compared, it makes no sense to compare. In some cases, databases only indicate its uncertainty due to the procedure mentioned in step two. This indicates why in the above diagram, there is no arrow pointing at the benchmark, unlike other steps.

3.3.1 Step 1: Obtaining Data

Obtaining data method can be divided into two major categories, which are satellite-based or ground weather station-based. Certain parameters are unable to be fulfilled by satellite and other parameters cannot be measured by the ground station. Subsequently, these two methods are often used jointly to create weather databases. Table 3.1 below shows which parameters cannot be measured for certain methods. However, it must be noted that ground station measurement value is simpler in handling, in a sense that less conversion and derivation is needed. Ground station measures radiation on surface directly, which is needed in several cases, as opposed to the top of atmosphere measurement of a satellite. (Mendelsohn & al., 2007) completed a comparison between ground station and satellite data. He concluded that to measure temperature, satellite data is more accurate. In contrast, a ground weather station is more accurate to measure precipitation.

Unmeasurable parameter	***Satellite***	***Ground station***
Radiation	surface radiation	top of atmosphere
		cloud condition
temperature	wet bulb temperature	

Table 3.1: Comparison between ground station and satellite based measurement parameter

Table 3.2 below show comparison of weather databases during step 1:

Parameter	***Benchmark***		***Result***	
	Name	*Characteristic*	*Satellite*	*Ground Weather station*
Monthly Global Radiation	BSRN (Baseline Solar Radiation Network), Global scale	Uncertainty: 2%	10,28% (2)	
	USCRN (US Climate Reference Network), US Scale	Accuracy: 70 W/m^2	6% (3)	5% (3)
Daily Global Radiation	USCRN (US Climate Reference Network), US Scale	Accuracy: 70 W/m^2	27% (3)	9% (3)
Monthly Temperature	NCDC (National Climatic Data Center		2,13 ^{0}C (2)	
Hourly Temperature	NSRDB (National Solar Radiation Database)			0,6^{0}C(2)
Wind Speed	RETScreen		1,3 m/s (2)	
Relative Humidity	NCDC (National Climatic Data Center		9,4% (2)	

Validation:

(1)	NRMSE	Normalized Root Mean Square Error
(2)	RMSE	Root mean square error
(3)	RMAE	Root mean absolute error
(4)	KS Test	Kolgorov-Smirnov Test
(5)	MBE	Mean Bias Error

Table 3.2: Comparison of accuracy between satellite and ground station

As can be seen, different validation methods, different benchmarking datasets, and procedures make it difficult to estimate which database is more precise and accurate, even if the scope is lowered down until only step one.

4.3.2 Step 2: Missing Data Completion

Table 3.3 below compares of procedures of each database in handling missing data. As shown below, only several databases are equipped with such procedures, namely HC3, TMY, and IWEC. Other databases such as NASA SSE, Meteonorm, and ESRA obtain their data from other projects. This shows why they are not equipped with such procedures.

Weather Database	*Short Term filling (depends on its smallest time increment)*	*Medium term filling (1 day missing)*	*Long Term filling (up to 1 year missing)*	*Last Ditch Filling*
HC3	15 minutes missing: intelligent oversampling	the radiation is interpolated along the day	the average daily irradiation is computed for all available days in the year and multiplied by the number of days in the year	no procedure
TMY	5 hour missing: linear interpolation	substitution of data from the same hours of adjacent days	substitution of data from the same calendar days from another similar year	substitution of data from same calender day from random year
IWEC	6 hour missing: linear interpolation	linear interpolation with the day before and after	not filled	extrapolation

Table 3.3: Missing Data Filling Methodology

Intelligent oversampling is an interesting method employed by HC3 to fill its missing data of short terms by using only maximum, minimum, or mean values of larger time increments (Lockhart, 1997). So the focus itself is not the waveform of a signal, but some more critical component of the waveforms represented by its maximum, minimum, or mean value. The method naturally permits some slight short time errors while the bigger picture still remains.

Other procedure like linear interpolation and extrapolation is a normal procedure. Last ditch filling, described in the latest column of table four above, is a last attempt procedure if everything else fails. This procedure is usually not benchmarked, as mentioned above, because it has no corresponding data. If they do have similar data to be benchmarked, the method usually generates several errors.

3.3.3 Step 3: Post-process Method

What is being defined as a post process method in this paper is any attempt to modify the database parameters and time increments into a much desirable outcome. This will usually involve:

1. Time conversion; as has been discussed above, time conversion at this point refers to stochastic time generation from higher time increment (monthly) into smaller (daily, hourly).
2. Parameter derivation; in several cases parameter is not measured, but generated from other available parameters that relate to them.
3. Space interpolation; is usually done in ground station based database because it will extend their coverage.

A complete comparison of weather databases can be obtained in the Appendix A, because the table will be too big to be presented in main text. Comparing and pointing out several major differences of post-process methodologies are the main objectives of this chapter.

As has been said above, Meteonorm is an uncommon example of this case. This figure 3.3 below (Remund, Quality of METEONORM v 6.0) shows how the algorithm of Meteonorm works.

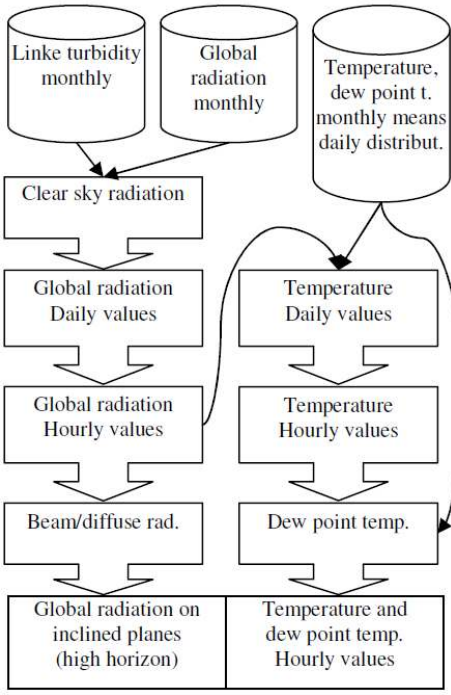


Figure 3.4: Meteonorm chain of algorithm (Remund, Quality of METEONORM v 6.0)

There is also another interesting behavior for Meteonorm, despite its uncommon time conversion. It is shown in figure 7 above that daily temperature values depend in global hourly values. Daily temperatures also depend on other parameters such as: dew point and relative humidity which are found to be related each other. The risk Meteonorm with greater error compared towards other databases, even when Meteonorm is based on ground weather station data that commonly shows more superior data accuracy compared to its satellite counterparts.

NASA SSE, HC3, TMY, and ESRA behave rather commonly. The four systems are similar because their time conversions start from smaller values into bigger values. Its algorithm also shares common behavior such as starting from a primary parameter of global horizontal, then splitting it into direct and diffuse, then looking for radiation on a tilted plane, and lastly finding an illuminance parameter. However, it is already a known issue that satellites cannot measure global horizontal radiation on a surface directly. The NASA SSE and HC3 input primary parameter of global horizontal was determined to have come from other sources.

As a matter of primary data, NASA SSE radiation primary data is obtained from the SRB (Surface Radiation Budget) Project. Other NASA projects also provide substantial data to be agglomerated into one single database which is now presented as NASA SSE version 6.0. HC3 uses the Heliosat-2 method to obtain its primary surface radiation data. As has been explained before, the ESRA and the PVGIS use the HC3 dataset as their primary data. Meteonorm also agglomerates all weather stations datasets all over the world, especially if they are listed under WMO (World Meteorological Organization).

3.3.4 Step 4: Typical Year Generation

The difference between various methodologies to generate a "typical" year could vary within weight. Another notable difference is how to treat "special" cases months such as mountain eruptions, leap years, etc. TMY omit those special cases months as their occurrences are too low. However, TRY still includes special cases in their database. Nevertheless, the special months are expected to be eliminated in the selection phase since their occurrences are low. Although TRY still includes those special months into calculation, the result is not too different from TMY.

Originally, TMY adopted the Sandia Method, which has been developed by Sandia National Laboratories. As TMY was developed into TMY2 and TMY3, several notable differences, particularly in weighting, took place. Table 3.4 below shows weight difference between TMY (original Sandia Method) and TMY2/TMY3.

Index	Sandia Method	NSRDB TMY2s
Max Dry Bulb Temp	1/24	1/20
Min Dry Bulb Temp	1/24	1/20
Mean Dry Bulb Temp	2/24	2/20
Max Dew Point Temp	1/24	1/20
Min Dew Point Temp	1/24	1/20
Mean Dew Point Temp	2/24	2/20
Max Wind Velocity	2/24	1/20
Mean Wind Velocity	2/24	1/20
Global Radiation	12/24	5/20
Direct Radiation	Not Used	5/20

Table 3.4: Difference between Sandia Method (TMY) and NSRDB (TMY2/3) (Marion & Urban, 1995)

While the usual TMY and TRY uses Finkelstein-Schafer method (Wilcox, 2008) to select month, DRY, on the other hand, uses its standard deviation to select typical months. So a month that deviates too much from its mean value will be eliminated. Complete comparisons between methodologies (TRY, TMY, DRY) can be seen in the appendix.

The general methodology of typical year generation could be summarized as shown in figure 3.4 below.

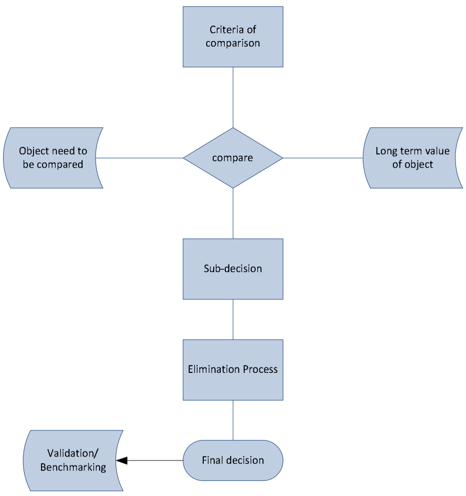


Figure 3.4: Summary of Typical Year Generation Process

Criteria of comparison as explained above could be the Finkelstein-Schafer method for TMY and TRY, but standard deviation for DRY. An object of comparison could be a CDF (Cumulative Distribution Function) of each designated parameter for TRY and TMY, and mean value for DRY. This is also the same for its long term value of object. After the comparison, the dataset is usually reduced into three to five candidate months (TMY, DRY) or years (TRY). Then, several elimination procedures occur to further reduce candidate months or years into a single most typical series. As usual, at the end of a procedure, validation is done to determine the validity of result from this process. All of these procedures, comparisons, and validity measures are presented in table 3.5 below. TMY 2 and TMY 3 have the same procedure as shown in the table below. They only differ in their observational time span, which are 15 years for TMY3 and 30 years for TMY2.

In conclusion, the typical year generation methodology exists to provide a series of datasets in a year that is typical during the observational time span. So, in some cases, merely averaging the value over observational time span is not enough because some extremes are still taken into account. In that sense, the best typical year generation method is the one that can ensure less variation and avoid extreme values on simulation. (Crawley, 1998) has indicated that for this reason, TRY should be avoided. He also noted that in most cases, TMY works well.

		TMY3/2	***TMY***	***TRY***	***DRY***
Criteria of Comparison	*Dry Bulb Temperature*	monthly max, min, mean	monthly max, min, mean	monthly max, min, mean	monthly max, mean
	Dew Point Temperature	monthly max, min, mean	monthly max, min, mean		
	Wind Velocity	monthly max, mean	monthly max, mean	monthly max, mean	monthly max, mean
	Radiation	monthly global, direct	monthly global		monthly global
	Relative Humidity			max, min, mean	monthly max, mean
Compare	*object*	candidate monthly CDF	candidate monthly CDF	candidate monthly CDF	candidate monthly value
	benchmark	Long Term CDF	Long Term CDF	Long Term CDF	long term mean value
	method	Finkelstein-Schafer Stat	Finkelstein-Schafer Stat	Finkelstein-Schafer Stat	Standard deviation
Elimination Process		Persistence criteria of temperature and radiation	Persistence criteria of temperature and radiation	Persistence criteria of radiation	Seasonal variation
		does not include volcanic eruption	will not include volcanic eruption		
		does not include leap year			

Table 3.5: Comparison between Typical Year Generation Method with regards to figure 3.4

3.3.4.1 Ambiguity between weather database and methodology

In some cases, the name of a methodology to generate a typical file weather database corresponds with the database itself. This is a special case in TMY (Test Meteorological Year). As a methodology, TMY is referred by a lot of other databases including IWEC. However, TMY itself is a valid database issued by NCDC (National Climatic Data Center), USA. As a weather database, it satisfies the requirements to be defined as a database above that it contains meteorological data of a certain location within a certain timeframe.

Another special case is Meteonorm. Meteonorm itself is proprietary or commercial software that can be classified as a weather database. Meteonorm itself on its user manual (METEONORM, 2010) specifies its purposes as follows:

1. A meteorological **database** containing comprehensive climatological data for solar engineering applications at every location of the globe.
2. A **computer program** for climatological calculations.
3. **Data source for engineering design programs** in the passive, active and photovoltaic application of solar energy with comprehensive data interfaces.
4. A **standardization tool** permitting developers and users of engineering design programs access to a comprehensive, uniform data basis.
5. **Meteorological reference** for environmental research, agriculture, forestry and anyone else interested in meteorology and solar energy.

However, several publication (Müller, 2001), (David M. , Adelard, Lauret, & Garde, 2010), (David M. , Adelard, Garde, & Boyer, 2005), (Ebrahimpour & Maerefat, 2009), has regarded Meteonorm with the same capabilities as other methodologies like TMY, TRY, or DRY. (Ebrahimpour & Maerefat, 2009) made a comparison between Meteonorm vs Weathergenerator (another software tool to generate typical year) and Sandia method, that is generally employed in TMY. Each parameter then plotted (temperature, solar radiation, wind speed, etc) of each RMS (Root Mean Square) value of each three method as shown in example below:

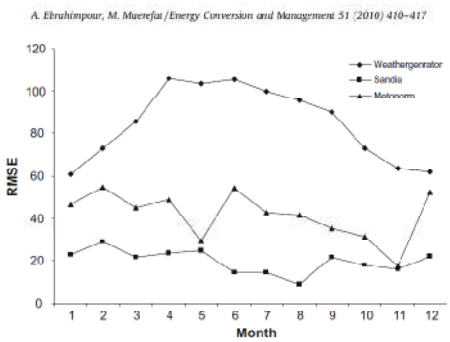


Figure 3.5: Monthly RMSE of global solar radiation of several methods (Ebrahimpour & Maerefat, 2009)

Finally (Ebrahimpour & Maerefat, 2009) concluded this following table, suggesting what method best fits for each month.

Month	Method
1	Metonorm
2	Weathergenrator
3	Metonorm
4	Metonorm
5	Metonorm
6	Metonorm
7	Metonorm
8	Weathergenrator
9	Sandia
10	Sandia
11	Sandia
12	Sandia

Table 3.6: Best method for every month (Ebrahimpour & Maerefat, 2009)

It can be concluded that Meteonorm itself as a software is quite versatile, with capabilities of generating a typical year. Nevertheless, after carefully looking at the Meteonorm documentation and help files, the theory how do they generate typical file is still unidentified.

The disparity in the cases shown above confirms that it could be misleading in some cases to have a lack of understanding between the methodologies and databases.

3.4 Comparing Database Accuracy

It has been explained above that database accuracy depends on quite a lot of factors. Weather database producers themselves are normally aware of accuracy matters and usually have performed validity measures of their own dataset. However, there is no uniform rule on how that should be done. As a result, validity measures differ (see chapter 3.2: Methodology used to benchmark) from one database to another database as well as with their benchmarking status.

In this chapter, an attempt to compare databases regardless of whether their parameters are primary or derived, or whether it is from satellites or ground stations is presented in table 3.7.

	NASA SSE	**ESRA**	**PVGIS**	**TMY**	**HC3**	**Meteonorm**
Monthly Global Horizontal	0,018 kWh/m2	0,027 kWh/m2	0,024 kWh/m2	0,026 kWh/m2	0,069 kWh/m2	0,015 kWh/m2
Monthly Direct Normal	1,3887 kWh/m2			0,057 kWh/m2	14,8 kWh/m2	0,01 kWh/m2
Monthly Dry Bulb Temperature	2,13ºC		0,7ºC			1,4ºC
Monthly Wind Speed	1,3 m/s					1.1m/s
Monthly Relative Humidity	9,4%					40%

Table 3.7: Summary of Database Accuracy Comparison

The main validity method here in this comparison is RMSE (Root Mean Square Error). It is being regarded as a main validation method by most databases, since many weather databases provide data using this same method. It also still retains its unit of measurement, so that it will be easier to convert into a more uniform parameter and compare. However, since accuracy is usually denoted by percentage this display may not be well understood. The main difficulty to convert this RMSE value into a percentage (NRMSE, Normalized Root Mean Square Error) is due to the fact that its range of measurement for each dataset is unknown.

Benchmarking datasets (BSRN, USCRN, NCDC, etc) may differ between each other. In this case, it is regarded as being equally valid and accurate.

As can be seen above, no database could really claim to have superior accuracy over another. Meteonorm and NASA SSE are two databases that can fulfill the need of complete meteorological parameters demanded in this paper, including radiation, temperature, humidity, and wind speed. As predicted before, Meteonorm as a ground weather station based, fare slightly better than NASA SSE.

3.5 Effect on Global Warming

As most weather databases are issued with a focus towards finding the mildest dataset series throughout the observational time span, global warming can disrupt conditions. To address the problem, several databases have been equipped with distinctive procedures.

The most common way to deal with global warming is by periodically issuing new versions of datasets. This is done for example in NASA SSE (which is currently version six), TMY3 and HC3 (which is both currently in version three). By renewing the procedures and observational time span on its new release of dataset, it is expected to cope with climate change.

Meteonorm has a Hadley CM3 model to cope with climate change (Meteonorm, 2010). The complete procedure is unable to be obtained in the Meteonorm user manual and has to be retrieved from the UK Meteorological Office. In short, the Meteonorm user manual assumes that carbon dioxide will double in the 21st century. The implications of this model are that the temperature model from Meteonorm is expected to be slightly adjusted per year.

Climate classification is also expected to be severely affected from global warming. Köppen climate classification, as the oldest and first attempt to classify climate, has undergone several modifications. In 1954 and 1961, it was updated by Rudolf Geiger. Thus, in several cases it is referred to as Köppen-Geiger classification because of the update. Please note that Geiger update is not connected as an attempt to adjust to global warming. (Rubel & Kottek, 2010) is the most recent paper concerning adjusting Köppen-Geiger climate classification into the climate change phenomenon. It has presented the possibility of climate change adjustment to Köppen Geiger climate classification, from 1900-2100. It has been noted that the most visible climate change may happen in climate type B, C, D, E that shift successively to the north. On the other hand, the Briggs climate classification procedure adjusted to climate change cannot be found. Perhaps this is due to the fact that climate classification has recently been issued in 2002. So, the adjustment is omitted.

4. Results and Discussion

4.1 Definition and categorizing of Database

As one of the objectives of this paper is to gain knowledge about weather databases, it is necessary to postulate the definition of weather database and a simplified categorization of weather database.

After the research, a weather database can be defined as a database that contains meteorological information such as temperature, radiation, etc. of a certain location and within a certain timeframe (this is a subject of different techniques and methodologies and thus cannot be held true that a weather database must have a distinctive timeframe). It does not necessarily need to be a real number while real time data in the above description has already proven further that synthetic databases perform much better in certain application areas rather than real data. It is also common practice to have weather databases to feed inputs into simulation software for solar thermal like TRNSYS or TRANSOL. It is expected that weather databases are able to synergize the software. In other words, a database might qualify as a weather database if the software refers to them as a source (either primary or complementary). However, this brings some consequences, in a sense that a borderline between databases and simulation software are rather blurred. In some cases, software can have their own databases.

Basically, by this paper, it is proposed to categorize various weather databases into major segments, which are:

I. From the standpoint of primary data obtaining methodology:

a. Satellite based weather database

b. Ground station based weather database

II. From the standpoint of time increment

1. Performing typical year generation methodology
2. Doing averaging over observational period
3. Giving real data from real time sequence
4. Obtaining typical year from other sources.

So a weather database could be classified as "a1" if it is satellite-based plus performing typical year generation, "a2" if satellite-based plus doing averaging over observational period, and so forth. An example of a proposed categorization is displayed in the column "category" in the table below.

The table summarizes the weather data being observed. Being ground station based is usually applying typical year generation. Satellite-based weather databases cannot be decided by which tendency time increment is processed. It is variable from type two (performing typical year generation methodology) to type four (obtaining typical year from other sources). However, satellite based weather databases are most likely not doing typical year generation by itself, as shown in table below; there is no "a1" classification.

No	Name	Coverage	Observational Year Span	Category	Created by	Expected Usage	Form	Cost	Main method
1	PVGIS (Photovoltaic Geographical Information System)	Europe	9	a2	European Commission, Joint Research Centre, EU	PV, GIS	time series, web based application, interactive map	free	Meteorological Satellite
		Africa	19						
2	IWEC (International Weather for Energy Calculation)	See map in appendix	30	b1	ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers), USA	PV, Solar Thermal	time series database	pay	ground weather station
3	TMY3 (Test Meteorological Year, version 3)	USA	15	b1	NCDC(National Climatic Data Center), USA	PV, Solar Thermal	time series database	pay	ground weather station
4	TMY 2 (Test Meteorological Year, version 2)	USA	30	b1	NCDC(National Climatic Data Center), USA	PV, Solar Thermal	time series database	pay	ground weather station
5	IGDG (Italia Dati Climati "G. de Giorgio")	Italy	13	b1	Politecnico di Milano, Italy		time series database		ground weather station
6	HC3(Helio Clim, version 3)	See map in appendix	7(current)	a3	MINES ParisTech, France	PV	time series databaseweb based application	pay	Meteorological Satellite
7	NASA SSE(Surface meteorology and Solar Energy)	Global	22	a2	NASA, USA	PV, Solar Thermal, Solar Cooker	web based application	free	Meteorological Satellite
8	Meteonorm	Global	20	b4	METEOTEST, Switzerland	PV, Solar Thermal	application	pay	ground weather station
9	ESRA (European Solar Radiation Atlas)	Europe	10	a4	Integrated group of scientist based on JOULE II programme, EU	PV, Solar Thermal, Biomass, Building	CD ROM, Book		Meteorological Satellite
10	CIBSE (The Chartered Institution of Building Services Engineers)	UK	21	b1	CIBSE, UK	PV, Solar Thermal	time series database	pay	ground weather station

Table 4.1: Summary and Categorization of Weather Database

4.2 Determining Parameter that Influence Heating and Cooling Demand and Supply Potentials of Solar Thermal System

This chapter attempts to integrate all knowledge gained from the process of making weather databases into an algorithm to determine potentials and demands of solar thermal systems.

4.2.1 Demand Side Analysis

It has been discussed in the literature review that HDD (Heating Degree Day) or CDD (Cooling Degree Day) directly influences heating and cooling demand. Temperatures (in this case HDD and CDD is a temperature derived parameter) unarguably plays the most important rule. Other parameters also play important roles in determining demand of heating and cooling.

(Erbs, Klein, & Beckman, 1984) postulate that the combined effect between temperature and radiation will yield better estimations of energy demand. The concept is denoted as Sol-Air temperature. By this concept, solar radiation can now be assuredly considered to influence the heating and cooling demand proportionally as well. The factor will depend on the opaqueness of a wall section (surfaces such as windows) where radiation incidentally passes through. This explains that on sunny days, temperatures are usually higher than the mean temperature.

Additionally, (Meteonorm, 2010) also provides an algorithm to derive temperature from radiation. It has been shown that the ratio between extraterrestrial radiation to surface radiation (kx) is proportional to temperature. When kx increases, temperatures also increase. At night, this effect is replaced by a cooling effect. (Meteonorm, 2010) also provides an algorithm to derive relative humidity from temperature (dew point temperature, and dry bulb temperature).

Wind could slightly alter heating demand if the condition is met (building structure allows it). Wind chill can normally decrease the temperature felt by human skin. This is a factor that will likely decrease the cooling demand and increase heating demand. However, wind chill is only valid at wind speeds above 4.8 km/h and temperatures below $10^{o}C$. In some cases, for example in tropical areas, the factor could be somewhat achievable by permitting larger ventilation in building structures. Subsequently, the building design will allow faster wind to blow through and thus wind chill factor will contribute to lower cooling demand. Wind chill is usually left out of weather databases. Wind chill can be easily determined by figure 2.5 above.

Another prominent factor of heating and cooling demand is evaporative cooling. The human body uses evaporative cooling to regulate their body temperature. (Nkemdirim, 1991) has presented that evaporation depends on temperature, wind speed, saturation vapor pressure, and actual vapor pressure. The ratio between saturation vapor pressure and actual vapor pressure is denoted as relative humidity. In this case, it is also implied that relative humidity affects the demand side of heating and cooling. For example, a room with the same temperature (for example $30^{o}C$) but different relative humidity creates a different cooling demand. A room with higher humidity will need more cooling because evaporative cooling, which should occur naturally in the human body, is being somewhat negated by that high humidity.

A description about the interrelation of parameters and concept above can be summarized by figure 4.1 below.

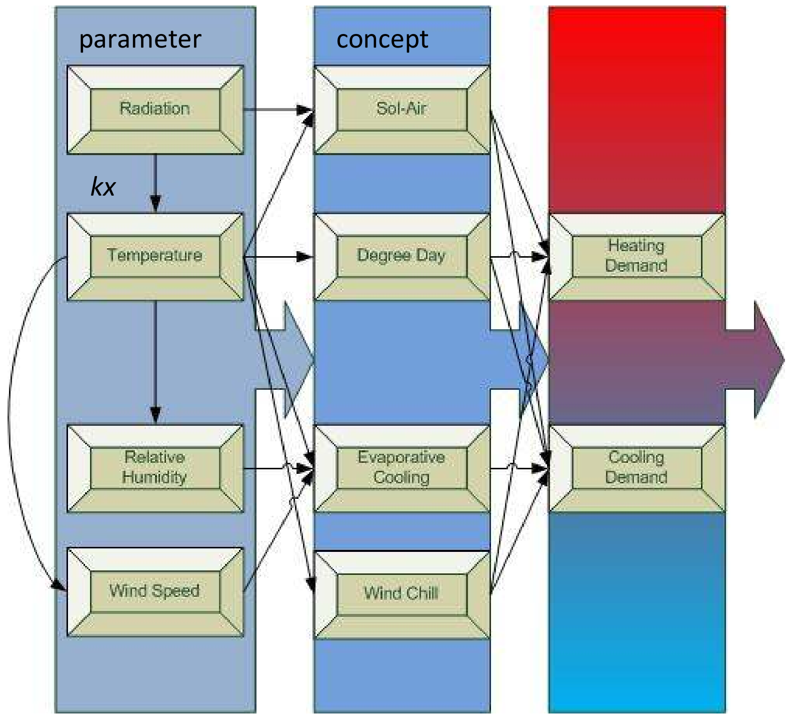


Figure 4.1: Parameter influencing demand side

Four concepts: Sol-Air, degree days, wind chill, and evaporative cooling, are put in the center layer column in the above diagram. This denotes a concept that could affect the heating and cooling demand directly and proportionally. The "concept" denotation is coined because all the real parameters are directly measurable. They are a fictitiously derived parameter used to facilitate the calculation and understanding of a system.

On the top left column, four parameters: temperature, radiation, wind speed, and relative humidity are put included. This denotes parameters. Temperature, having the most branches (four arrows going out) has the most significant influences on the demand side. Additionally, temperature is thought to have the most influential parameter because it also gains control upon wind speed and relative humidity. However, temperature is influenced by radiation as is seen in by the arrow with *kx* denoted above.

Heating demand relates proportionally to the degree days and the Sol-Air concept as is shown in (CIBSE, 2006) and (Erbs, Klein, & Beckman, 1984). Wind chill may occasionally increase heating demand in the case that the building is designed to allow wind exposure. Evaporative cooling does not have any connection with heating demand as it helps cooling rather than heating.

Cooling demand relates proportionally to all four concepts mentioned above. The degree days and Sol-Air concepts relate proportionally with cooling demand. Meanwhile, evaporative cooling might increase cooling demand much further than already indicated by degree days and Sol-Air.

4.2.2 Supply Side Analysis

As is nearly the same with the demand side analysis, the supply of solar thermal energy naturally depends on temperature. Other influencing parameters are solar radiation, which is still in line with the Sol-Air concept as explained above. The effects of evaporative cooling and wind chill will be nonexistent because those are the temperatures felt by people, which is not exactly the real temperature.

Evaporative cooling considerably influences the efficiency of absorption chillers. Evaporative cooling supposedly happens in the cooling tower of a chiller unit. The effect theoretically reduces the overall efficiency of the chiller.

5. Conclusion

From this study, it is shown that there are several possibilities of quickly determining the potential of solar thermal systems and their demand with available weather databases. Several weather databases are freely distributed and surprisingly contain quite comprehensive meteorological parameters that could influence cooling or heating demands and solar thermal potentials. It should be noted that in several weather databases, meteorological parameters could be incomplete while several other parameters are derivatives and thus selecting databases according to specific needs is crucial.

It should be recognized that this research paper is not a quantitative research but more of a qualitative research. A plan to have more quantitative research to cross check the validity of weather database against proxy ground measurement in Bolzano was undermined due to time and equipment constraints.

Because of the qualitative nature and several barriers from this research, it could not be blatantly decided, which parameters have how much influence towards heating and cooling demand and its potential. It is also difficult to determine exactly which weather database is the most accurate for each usage.

Subsequent research is being planned in Indonesia for the next stage of solar thermal system research in order to know exactly how much humidity affects heating and cooling demands and solar thermal potentials. Indonesia is known as a hot and humid region with tropical climates ("A" class according to Köppen classification).

References

Aguiar, R., & Collares-Perreira, M. (1992). TAG: A Time Dependent, Autoregressive, Gaussian Model for Generating Synthetic Hourly Radiation. *Solar Energy: Vol. 49, No, 3* , 167-174.

Aguilar, R., Collares-Perreira, M., & Conde, J. (1988). Simple Procedure for Generating Sequences of Daily Radiation Values Using a Library of Markov Transition Matrices . *Solar Energy Vol. 40, No, 3.* , 269-279.

ASHRAE Technical Committee 4.2. (2011, 6 15). *Climatic Information-FAQ*. Retrieved 6 15, 2011, from ASHRAE Technical Committee 4.2: http://tc42.ashraetcs.org/faq.html#6

Bilbao, J., Miguel, A., Franco, J. A., & Ayuso, A. (2003). Test Reference Year Generation and Evaluation Methods in the Continental Mediterranean Area. *Journal of Applied Meteorology Vol. 43* , 390-400.

Briggs, R. S. (2002). *Climate Classification for Building Energy Codes and Standards.* Pacific NW National laboratory.

CIBSE. (2006). *Degree Days: Theory and Application.* London: CIBSE.

College of St. Benedict and St. John University. (2011, 6 14). *Kolmogorov-Smirnov Test*. Retrieved 6 14, 2011, from http://www.physics.csbsju.edu/stats/KS-test.html

Crawley, D. B. (1998). *Which Weather Data Should You Use for Energy Simulation of Commercial Building?* ASHRAE.

D. Steward, G. S. (November 2009). Lifecycle Cost Analysis of Hydrogen Versus Other Technologies for Electrical Energy Storage. *National Renewable Energy Laboratory-Technical Report* .

David, M., Adelard, L., Garde, F., & Boyer, H. (2005). Weather Data Analysis Based on Typical Weather Sequences. Application: Energy Building Simulation. *Building Simulation 9th International Conference*, (pp. 239-246). Montreal, Canada.

David, M., Adelard, L., Lauret, P., & Garde, F. (2010). A method to generate Typical Meteorological Years from raw hourly Climatic Database. *Building and Environment 45* , 1722–1732.

Ebrahimpour, A., & Maerefat, M. (2009). A method for generation of typical meteorological year. *Energy Conversion and Management 51* , 410-417.

Energetic. (2011, 3 13). *Pumped-Storage Method Hydroelectricity*. Retrieved 3 13, 2011, from Hvdroelectric Energy: http://hydroelectric-energy.blogspot.com/2010/12/pumped-storage-methods-hydroelectricity.html

Erbs, D., Klein, S., & Beckman, W. (1984). Sol-Air Heating and Cooling Degree Days. *Solar Energy vol. 33* , 605-612.

Huld, T. A., Süri, M., Dunlop, E. D., & Micale, F. (2006). Estimating average daytime and daily temperature profiles within Europe. *Environmental Modelling & Software Vol 21, Issue 12* , 1650-1661.

Kottek, M. e. (2006, April 10). World Map of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift,* , pp. 253-269.

L.Gaines, J. S. (October 1, 2010). *A Review of Battery Life-Cycle Analysis:State of Knowledge and Critical Needs.* Argonne: Argonne National Laboratory.

Lawrence, M. G. (2005). Humidity and Dew Point Temperature in Moist Air: A Simple Conversion and Application. *American Meteorological Society* , 225-233.

Liu, B. Y., & Jordan, R. C. (1960). The Interrelationship and Characteristic Distribution of Direct, Diffuse, and Total SolarRadiation. *Solar Energy, Vol. 4, No. 3* , 1-19.

Lockhart, R. (1997). Enhancing Data Acquisition with Intelligent Oversampling. *SENSORS* .

Lund, H. (1995). *The Design Reference Year User Manual.* Lyngby: TU Denmark.

Marion, W., & Urban, K. (1995). *User Manual for TMY2.* Colorado: NREL.

Maxwell, E. (1998). METSTAT-The Solar Radiation Model Used in the Production of the National Solar Radiation Database (NSRDB). *Solar Enera_v. Vol. 62. No. 4* , 263-279.

Mendelsohn, R., & al., e. (2007). Climate Analysis with Satellite versus Weather Station Data. *Climate Change 81* , 71-83.

Meteonorm. (2010). *Meteonorm Version 6.0 Handbook Part 1: Software.* Bern: METEOTEST.

Meteonorm. (2010). *Meteonorm Version 6.0 Handbook Part II: Theory.* Bern: METEOTEST.

Müller, E. (2001). Development of a Test Reference Year on a Limited Data Base for Simulations on Passive Heating and Cooling in Chile. *Building Simulation 7th International Conference*, (pp. 135-142). Rio de Janeiro, Brazil.

NASA. (2009). *Surface meteorology and Solar Energy (SSE) Release 6.0 Methodology.* NASA.

Nave, R. (2011, 06 11). *relative humidity*. Retrieved 06 11, 2011, from hyperphysics: http://hyperphysics.phy-astr.gsu.edu/hbase/kinetic/relhum.html

Nave, R. (2011, 6 18). *Temperature Regulation of the Human Body* . Retrieved 6 18, 2011, from hyperphysic: http://hyperphysics.phy-astr.gsu.edu/hbase/thermo/heatreg.html#c1

Nkemdirim, L. (1991). An Empirical Relationship Between Temperature, Vapour Pressure Deficit and Wind Speed and Evaporation During a Winter Chinook. *Theoritical and Applied Climatology. 43* , 123-128.

NOAA. (2011, 1 1). *wind chill.* Retrieved 6 15, 2011, from NOAA National Weather Service: http://www.weather.gov/os/windchill/index.shtml

Peltier, R. (2010, March 10). *Can Utility-Scale Batteries Rescue Intermittent Renewables? (Improvement, market shakeout, but no 'silver bullet').* Retrieved February 21, 2011, from MasterResource: http://www.masterresource.org/2010/03/can-utility-scale-batteries-rescue-intermittent-renewables/

Ramsden, T., Kroposki, B., & Levene, J. (2008). Opportunities for Hydrogen-Based Energy Storage for Electric Utilities. *Golden, CO: National Renewable Energy Laboratory* .

Ransome, S., & Funtan, P. (2005). Why Hourly Averaged Measurement Data is Insufficient to Model PV System Performance Accurately. *PVSEC.*

Rastler, D. (September/October 2008). New Demand for Energy Storage. *Electric Perspective* .

Remund, J. *Quality of METEONORM v 6.0.* Bern.

Remund, J., Wald, L., & Page, J. (2003). Chain of Algorithms to Calculate Advanced Radiation Parameters. *ISES Solar World Congress.* Göteborg.

Renewable Training. (2010, 8 30). *Solar Radiation 101 – Irradiance and Irradiation*. Retrieved 6 15, 2011, from Renewable Training: http://renewable-training.com/?p=170

Rigollier, C., Olivier, B., & Lucien, W. (2000). On Clear Sky Model of the ESRA-European Solar Radiation Atlas with Respect to Heliosat Method. *Solar Energy 68, 1* , 33-48.

Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–2100 depicted by World Maps of the Köppen-Geiger Climate Classification. *Meteorologische Zeitschrift, Vol. 19, No. 2* , 135-141.

S.M Schoenung, W. H. (2003). *Long- vs. Short-Term Energy Storage Technologies Analysis: A Life-Cycle Cost Study.* Livermore, CA: Sandia National Laboratories: A Study For The DOE Energy Storage Systems Program.

Scharmer, K., & Greif, J. (2000). *The European Solar Radiation Atlas Vol. 1: Fundamentals and Maps.* Paris: Les Presses de l'École des Mines.

SoDa. (2011, 4 26). *Plane Orientation and Radiation Components*. Retrieved 6 15, 2011, from SoDa (Solar Radiation Data): http://www.soda-is.com/eng/education/plane_orientations.html

Šúri, M., & Hofierka, J. (2004). A New GIS-based Solar RadiationModel and Its Application to Photovoltaic Assessments. *Transactions in GIS, 8, 2* , 175-190.

Synwoldt, C. (WS 2010/2011). 13.Storage-General. *Economic Aspect of Sustainable Energy, Technology and Strategy.* Birkenfeld, Rheinland Pfalz: Ing.-Büro Synwoldt.

Thenevard, D., & Brunger, A. (2001). *Typical Weather Year for International Location.* Atlanta: ASHRAE Research Project 1015-RP.

USCRN. (2011, 6 19). *NCDC.* Retrieved 6 19, 2011, from USCRN: http://www.ncdc.noaa.gov/crn/elements.html

Wald, L., & Lefevre, M. (2001). *Interpolation schemes - "Profile Method" (a process-based distance for Interpolation Schemes).* SoDa.

Wilcox, S. a. (2008). *Users Manual for TMY3 Data Sets.* Colorado: NREL, Technical Report.

Zelenka, A. G. (1992). *Techniques for supplementing solar radiation network data.* IEA Report No.IEA-SHCP-9D-1.

Appendix

Appendix A: Complete Database Table Observed

Steps	Parameter	Method/ Description	Validation
primary parameter	monthly global horizontal	Space interpolation (Zelenka et al. 1992;Wald and Lefèvre, 2001)	8,9%(2)
	monthly linke turbidity	ESRA, 2000	
first derivative	monthly clear sky radiation	ESRA, 2000	
second derivative	daily global horizontal	Stochastic time generation	1,7%(4)
third derivative	hourly global horizontal	Stochastic time generation (Aguiar and Collares-	11,6%(4)
fourth derivative	hourly direct normal	Perez et al, 1991	7,4%(2)
	hourly diffuse horizontal		7,1% (2)
fifth derivative	hourly global on inclined plane	Perez et al. 1986	8 W/m^2(2)
sixth derivative	hourly global illuminance	Perez et al. 1990	0,55 klux (2)

Validation		
(1)	NRMSE	Normalized Root Mean Square Error
(2)	RMSE	Root mean square error
(3)	RMAE	root mean absolute error
(4)	KS Test	Kolgorov-Smirnov Test
(5)	MBE	Mean Bias error

Appendix A1: Tables of Meteonorm Accuracy and Methodology for Solar Radiation

Steps	Parameter	Method/ Description	Validation
primary parameter	monthly dry bulb temperature	Space interpolation (Zelenka et al. 1992;Wald and Lefèvre, 2001)	1,4%(2)
	daily global horizontal		
first derivative	daily temperature	stochastic auto regression process	
second derivative	daily min, max temperature		
third derivative	hourly temperature		

Appendix A2: Tables of Meteonorm Accuracy and Methodology for temperature

Steps	Parameter	Method/ Description	Validation
primary parameter	monthly dew point temperature	Space interpolation (Zelenka et al. 1992;Wald and Lefèvre, 2001)	1,5%(2)
	monthly dry bulb temperature		
first derivative	daily dew point temperature	stochastic time generation	
second derivative	hourly dew point temperature	stochastic time generation	
	hourly relative humidity		

Appendix A3: Tables of Meteonorm Accuracy and Methodology for Relative Humidity

Steps	Parameter	Method/ Description	Validation
primary parameter	monthly wind speed	Space interpolation (Zelenka et al. 1992;Wald and Lefèvre, 2001)	1,1%(2)
	daily global horizontal		
first derivative	hourly wind speed	stochastic auto regression process	

Appendix A 4: Tables of Meteonorm Accuracy and Methodology for Wind Speed

Steps	***Parameter***	***Method/ Description***	***Validation***
primary parameter	3-hourly global horizontal (19)		
first derivative	daily global horizontal		
second derivative	monthly global horizontal		
third derivative	monthly diffuse		0,39 $kWh/d.m^2$ (2)
	monthly direct normal		1,39 $kWh/d.m^2$ (2)
fourth derivative	monthly clear sky global		12,9 W/m^2 (2)
	monthly clear sky direct normal		1,16 $kWh/d.m^2$ (2)
	monthly clear sky diffuse		0,17 $kWh/d.m^2$ (2)
fifth derivative	monthly global on inclined plane		0.19%(2)

Appendix A5: Tables of NASA SSE Accuracy and Methodology for Solar Radiation

Steps	***Parameter***	***Method/ Description***	***Validation***
primary parameter	3-hourly temperature		
first derivative	daily mean, max, min temperature	space, bilinear interpolation	2,75^{0}C(2)
second derivative	daily mean, max, min temperature	Elevation correction factor applied	2,47^{0}C(2)
third derivative	Heating degree day	setpoint 18^{0}C	
	cooling degree day		

Appendix A6: Tables of NASA SSE Accuracy and Methodology for Temperature

Steps	Parameter	Method/ Description	Validation
primary parameter	hourly global horizontal		
first derivative	hourly direct normal		
	hourly diffuse horizontal		
second derivative	hourly global illuminance	Perez et al 1990	1,2%(5)
	hourly diffuse illuminance		1,6%(5)
	hourly direct illuminance		2,3%(5)
	hourly zenith illuminance		1,2%(5)
third derivative	monthly global horizontal		0.2kWh/m^2/day(2)
	monthly direct normal		0.5kWh/m^2/day(2)

Appendix A7: Tables of TMY Accuracy and Methodology for Solar Radiation

Steps	***Parameter***	***Method/ Description***	***Validation***
primary parameter	hourly dry bulb temperature		
first derivative	monthly heating degree day	setpoint 18^{0}C	45,6(2)
	monthly cooling degree day		28,2(2)

Appendix A8: Tables of TMY Accuracy and Methodology for Temperature

Steps	***Parameter***	***Method/ Description***	***Validation***
primary parameter	Hourly Top of atmosphere Radiation		
first derivative	Hourly clear sky radiation	METSTAT	14,04%(1)
	Hourly clear sky direct radiation		32.95%(1)
	Hourly clear sky diffuse radiation		45,9%(1)
second derivative	Hourly Global Radiation	Kasten	55,3%(1)
	Hourly Direct Normal		92,45%(1)
	Hourly Diffuse Horizontal		88,475%(1)
third derivative	Hourly Global Illuminance		
	Hourly Diffuse Illuminance		
	Hourly Direct Illuminance		

Appendix A9: Tables of IWEC Accuracy and Methodology for Solar Radiation

Steps	Parameter	Method/ Description	Validation
primary parameter	Clear Sky index-meteo station		
	Monthly Linke Turbidity	reinterpolate from SoDa(HC)	0.7(2)
first derivative	Clear Sky global		
	Clear sky direct		
	clear sky diffuse		
second derivative	monthly global horizontal		0,024(2)
third derivative	Diffuse Horizontal		
	Direct Normal		
fourth derivative	Direct on inclined plane		
	Diffuse on inclined plane		

Appendix A10: Tables of PVGIS Accuracy and Methodology for Solar Radiation

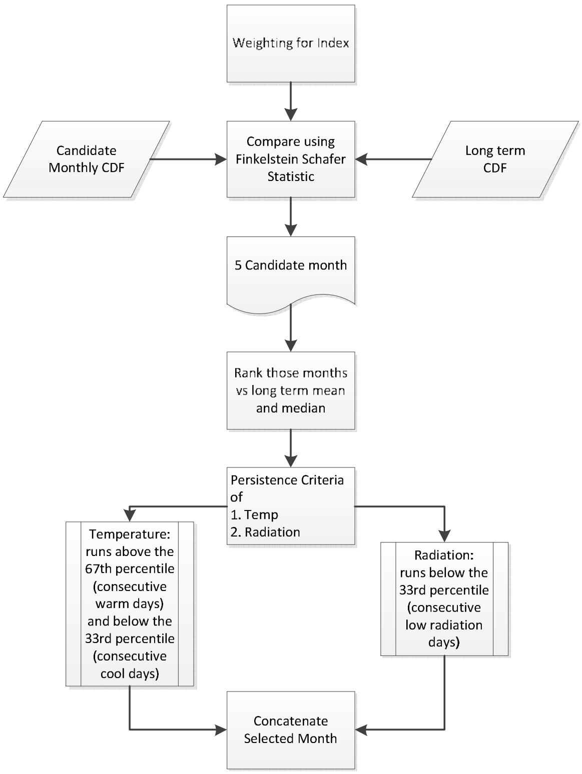
Steps	***Parameter***	***Method/ Description***	***Validation***
primary parameter	Daily temperature		1-1.2^{0}C
first derivative	Monthly average	spatial interpolation	0.5-0.7^{0}C
second derivative	Heating Degree Days		

Appendix A11: Tables of PVGIS Accuracy and Methodology for Temperature

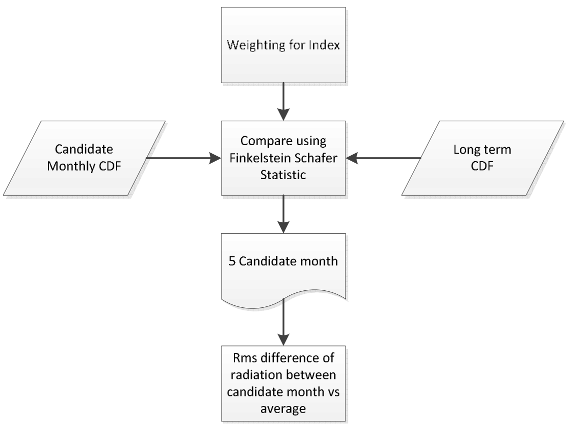
Parameter	NASA SSE	HC3	ESRA	Meteonorm	IWEC	TMY	PVGIS
Lattitude (φ)	constant (user input)	constant	constant (user input)	constant (user input)			constant (user input)
Longitude (λ)	constant (user input)	constant	constant (user input)	constant (user input)			constant (user input)
time (t)		constant (user input)	constant (user input)	constant (user input)			constant (user input)
Elevation (z)	constant						
Global Radiation on Top of Atmosphere (G_{ba})	measured primary	measured primary	constant (λ,φ,t)	constant (λ,φ,t)	constant	measured primary	constant (λ,φ,t)
sunset hour angle (SSHA)	constant	constant	constant	constant			constant
noon solar angle from the horizon (NHSA)	constant						
Solar Zenith angle (THMT)	constant						
Cloud amount (I)					constant		
Solar Azimuth angle(							
Solar altitude angle (γ)		constant	constant	constant			constant
Solar Declination Angle (δ)		constant	constant	constant			constant
Linke Turbidity Factor (T_{LK})		constant	constant	constant			constant
Relative Optical Air Mass (m)		constant (γ)	constant	constant			constant(NHSA, z)
Rayleigh Optical Thickness ($δ_m$)		constant	constant	constant	constant		constant(m)
Global Horizontal Radiation (GH)	measured primary(1)	Measured primary(1)	measured primary(2)	measured primary	derived	measured primary	derived (KT, GH_{clear})
Direct Horizontal (BH)		derived	derived	derived			derived
Diffuse Horizontal (DH)	derived (GH, BN, TMHT)	derived (GH,KT)	derived (GH,KT)	derived (GH,KT)	derived	measured primary	derived (GH,KT)
Direct Normal (BN)	derived (KT, SSHA, NHSA)	derived	derived	derived	derived	measured primary	derived
Clearness Index (KT)	derived (G_{ba},GH)	derived (G_{ba},GH)	derived (G_{ba},GH)	derived (G_{ba},GH)			constant
Clear Sky Global Horizontal (GH_{clear})							derived (T_{LK})
Clear Sky Diffuse Horizontal (DH_{clear})	derived (KT)						
Clear Sky Direct Normal (BN_{clear})	derived (Dh_{clear}, TMHT)	derived (T_{LK},m,$δ_m$)	derived (T_{LK},m,$δ_m$)	derived (T_{LK},m,$δ_m$)			derived (T_{LK},m,$δ_m$)
Clear Sky Direct Horizontal (BH_{clear})		derived(T_{LK},m,γ)	derived(T_{LK},m,γ)	derived(T_{LK},m,γ)			derived (T_{LK},m,$δ_m$)
Global on tilted surface							
Direct on tilted surface							
Diffuse on tilted surface				derived (TMHT,DH,BN)			
Global illumination				derived	derived	derived	
Direct illumination				derived	derived	derived	
Diffuse Illumination				derived	derived	derived	
Zenith Illumination				derived	derived	derived	
Dry Bulb Temperature (T)	measured primary			both primary and derived (GH)	measured primary	measured primary	both primary and derived (space interpol)
Heating Degree Days	derived (T)				derived (T)	derived (T)	derived (T)
Cooling Degree Days	derived (T)				derived (T)	derived (T)	derived (T)
Dew Point Temperature (Td)	measured primary			derived (Linear Interpol)			
Relative humidity	measured primary			derived (T,Td)			
Wind speed	measured primary			derived (Stochastic model)	measured primary	measured primary	

Appendix A12: Tables of Algorithm Derivation and Parameter Each Database

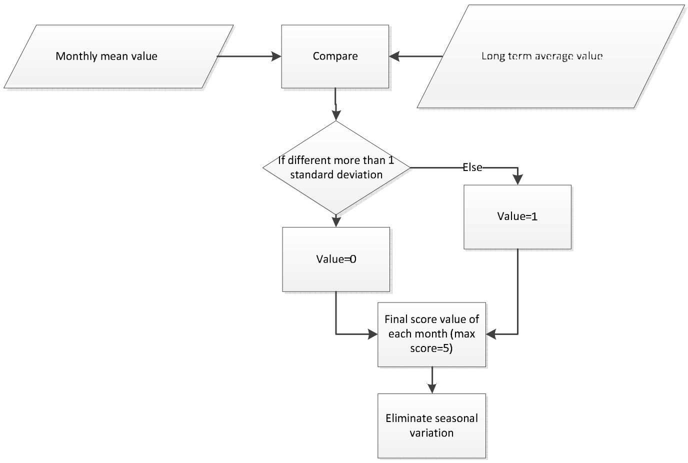
Appendix B: Flowchart of Typical Year Generation Method



Appendix B13: Flowchart of TMY



Appendix B14: Flowchart of TRY



Appendix B15; Flowchart of DRY

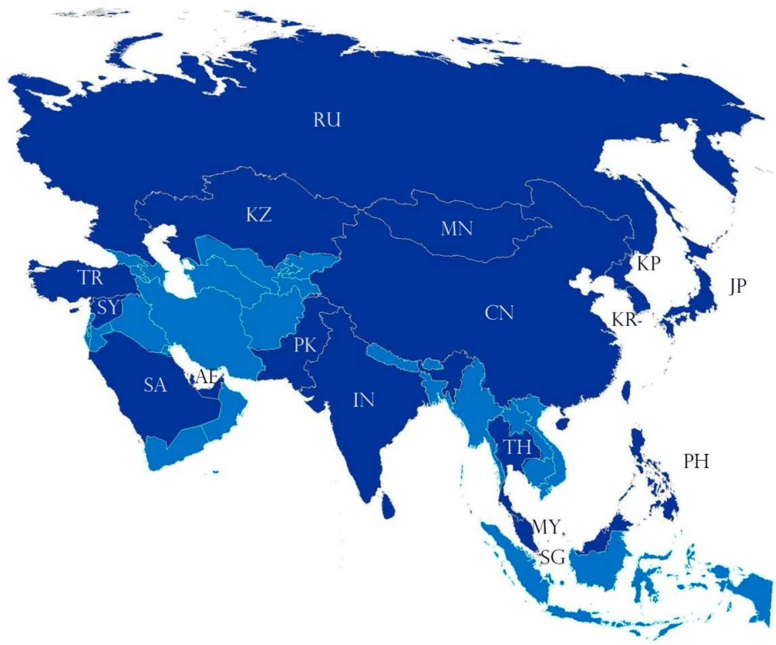
Appendix C: Coverage map of Databases

Appendix C1: IWEC Coverage

Asia

Description of Country Code

1. *AE=United Arab Emirates*
2. *CN=China*
3. *IN=India*
4. *JP=Japan*
5. *KP=North Korea*
6. *KR=South Korea*
7. *KZ=Kazakhtan*
8. *MN=Mongolia*
9. *MY=Malaysia*
10. *PH=Phillipines*
11. *PK=Pakistan*
12. *RU=Russia*
13. *SA=Saudi Arabia*
14. *SG=Singapore*
15. *SY=Syria*
16. *TH=Thailand*
17. *TR=Turkey*

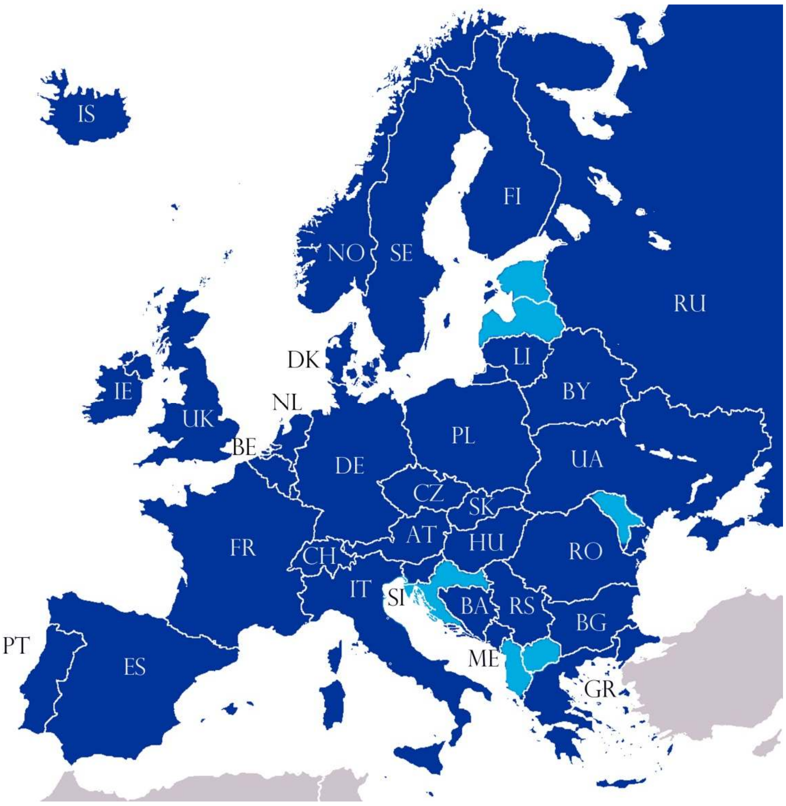


Appendix C16: IWEC Coverage Map of Asia

Europe

Description of Country Code

1. *AT=Austria*
2. *BA=Bosnia Herzegovina*
3. *BE=Belgium*
4. *BG=Bulgaria*
5. *BY=Belarus*
6. *CH=Switzerland*
7. *CZ=Czech Republic*
8. *DE=Germany*
9. *DK=Denmark*
10. *ES=Spain*
11. *FR=France*
12. *FI=Finland*
13. *GR=Greece*
14. *HU=Hungary*
15. *IE=Ireland*
16. *IS=Iceland*
17. *IT=Italy*
18. *LI=Lithuania*
19. *ME=Montenegro*
20. *NL=Netherland*
21. *NO=Norway*



Appendix C17: IWEC Coverage Map of Europe

America

Description of Country Code

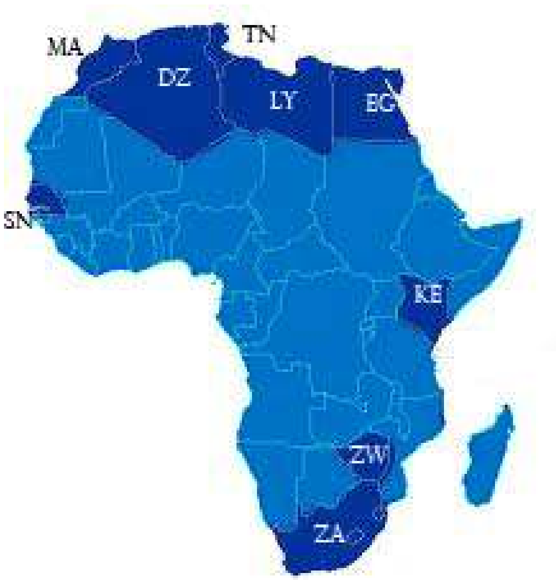
1. *AR=Argentina*
2. *BO=Bolivia*
3. *BR=Brazil*
4. *CL=Chile*
5. *CO=Colombia*
6. *EC=Ecuador*
7. *MX=Mexico*
8. *MQ=Martinique (France)*
9. *PE=Peru*
10. *PY=Paraguay*
11. *UY=Uruguay*
12. *VE=Venezuela*

Appendix C18: IWEC Coverage Map of America

Africa

Description of Country Code

1. *DZ=Algeria*
2. *EG=Egypt*
3. *LY=Libya*
4. *MA=Morocco*
5. *SN=Senegal*
6. *TN=Tunisia*
7. *KE=Kenya*
8. *ZA=South Africa*
9. *ZW=Zimbabwe*

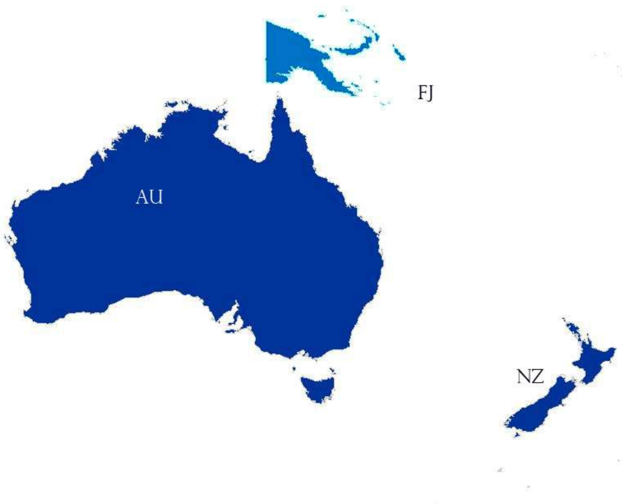


Appendix C19: IWEC Coverage Map of Africa

Australia

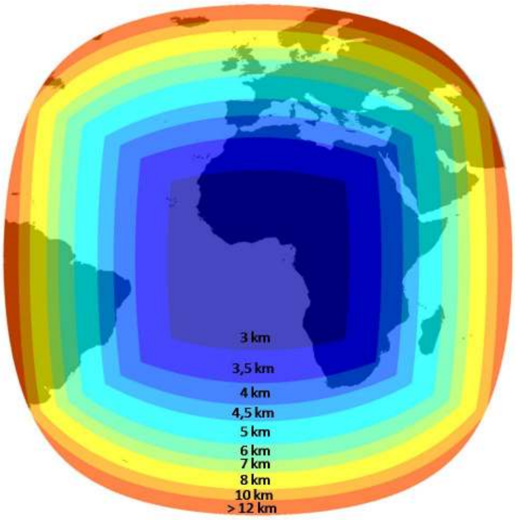
Description of Country Code

1. *AU=Australia*
2. *FJ=Fiji*
3. *NZ=New Zealand*



Appendix C20: IWEC Coverage Map of Australia

Appendix C2: HC3 Coverage



Appendix C21: HC3 Coverage Map (SoDa, 2011)

Appendix C3: ESRA Coverage

Appendix C22: ESRA Coverage Map (Scharmer & Greif, 2000)