IMPROVED QUALITY CONTROL PROCEDURES AND MODELS FOR SOLAR RADIATION USING A WORLD-WIDE DATABASE

SERGE YOUNES

A thesis submitted in partial fulfilment of the requirements
of Napier University for the degree of Doctor of Philosophy

March 2006

DECLARATION

I hereby declare that the work presented in this thesis was solely carried out by myself at Napier University, Edinburgh, except where due acknowledgment is made, and that it has not been submitted for any other award.

Serge Younes (Candidate)

Date

ABSTRACT

This thesis deals with various aspects of broadband horizontal solar irradiance. Quality control of measured datasets are identified and analysed. It was found that solar irradiance datasets may contain significant errors. These sources of errors were divided in two categories, the inherent instrument errors and operation related errors. Methods of assessing the quality of the datasets were evaluated and found to be unsatisfactory. A new method was hence developed to quality control the solar irradiance data.

The quality control procedure consists of two tiers of tests. The first tests are physical tests that identify and remove data points that are physical impossibilities. The second tier tests consist of the creation of a mathematical envelope of acceptance in a sky clarity index domain. This envelope of acceptance is based on multiples of standard deviations of the weighted mean of clearness index to diffuse ratio. The available datasets in this study were thus quality controlled to remove any obvious outliers.

Modelling the solar resource is an important tool for engineers and scientists. Such models have been developed since the second half of the 20th century. Some models rely on one or two meteorological parameters to estimate the solar irradiance, while other models are more complex and require a far greater number of points. Two of these models have been analysed and evaluated. The two models are all-sky, broadband solar irradiance models.

The first model analysed is the Meteorological Radiation Model, or MRM. This model is in fact a sunshine based model, with atmospheric turbidity taken into account as well. The beam irradiance component was found to be acceptable given the number of inputs required by the model. Any extra parameters would increase the complexity of the model, without noticeable improvements. The regressions were modified to take into account sunshine fraction banding. However the diffuse irradiance was identified as one which had the potential for improvement. Thus, in the present work an attempt has been made to develop improved models. The new model was found to be far superior to the older, original model, thus the name Improved Meteorological Radiation Model, IMRM.

The second type of model investigated is the cloud based radiation model. This type of model is simple to use and rely on regressions between irradiation, solar altitude angle and the cloud cover. Careful analysis of the cloud distribution reveals certain flaws in the current regressions. New regressions were formulated and the result was a model superior to all its predecessors.

Clear-sky modelling is important for maximum load calculations; however, there is no method of extracting with accuracy clear-sky broadband data. Clear-sky identification techniques were evaluated and a new method was devised. These new datasets were used on four clear-sky models, MRM, Page's Radiation Model, PRM, Yang's radiation model and Gueymard's REST2 model. It was found that using this new method of extracting extreme-clear-sky data, the models performed better than when using quasi-clear-sky data.

Solar radiation modelling is not an end by-itself, it must serve a purpose for engineers in their applications. Napier University has installed a 160m^2 photovoltaic facility in 2003. A 27-year solar radiation dataset was available for Edinburgh, to do feasibility calculations for the project; however this dataset contained gaps in the data. The cloud radiation model developed in this study was utilised to this end. In addition a complete life cycle analysis was performed on the project, and it was found that with an average efficiency at around 12%, the facility will payback its embodied energy in eight years, and based on a relatively conservative forecast of energy prices, the financial payback is set at under 100 years.

AKNOWLEDGMENTS

I would like to thank my Parents, brother and sister for their moral and financial support over the past few years. This thesis is dedicated to you. I hope I make you proud. I'd like to thank my Uncles and Aunts and all my cousins for their non-stop support and encouragement.

I would also like to thank my director of studies, Professor Tariq Muneer. Thank you for believing in me and helping me find my potential. In addition and special thank you to my supervisory and advisory team, Professor Jorge Kubie and Dr. Mohamed Asif for the support you have given me.

I can't forget the technical and administrative staff at the Faculty and the school of engineering: Bill Young, Bill Campbell, Ian Campbell and all the other technicians who were always so helpful in providing the proper equipment when needed. Also Sandra , Claire-Louise, and Alison from the school office for answering all the questions I had. Thank you to the teaching staff with whom, I shared many lectures.

For his support, and for motivating me and tutoring me on solar radiation and quality control before starting my PhD, I would like to thank Randall Claywell. I'd like to thank those who worked with me during my PhD, Stephane, Nicolas and my esteemed colleague Saima Munawwar. I wish Andrew, Celine, Haroon and Peter the best of luck with their PhDs. I'd also like to thank my friends and peers in the solar energy field, Chris Gueymard, Daryl Myers, Dany Li, Avraham Kudish and John Page.

Finally I would like to thank my friends in Edinburgh who were very supportive throughout the past few years: Alistair, Annabelle, Jahan, John, Laurent, Luisa, Mareike, Nasser, Onintza, Simon, Tim, and the Greek boys.

A special regards to my friend and partner, Annalisa for her constant support over the last year.

Contents

1 INTRODUCTION

2 LITERATURE REVIEW

3 QUALITY CONTROL OF SOLAR RADIATION DATASETS

4 THE IMPROVED METEOROLOGICAL RADIATION MODEL

5 CLOUD RADIATION MODELLING

6 CLEAR-SKY DATA SELECTION AND MODELLING

7 CONCLUSIONS

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendix D

LIST OF TABLES

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

NOMENCLATURE

DN	Julian day number	
DNI	direct normal irradiance	(W/m^2)
DEC	declination	(radians)
DBR	diffuse to beam ratio	
F(α)	function of solar altitude	
HDD	hourly/daily datasets	
HSHD	hourly/sub-hourly datasets	
I_E	extraterrestrial horizontal irradiance	(W/m^2)
$I_{reflected}$	reflected irradiance by atmosphere	(W/m^2)
$I_{absorbed}$	absorbed irradiance by atmosphere	(W/m^2)
I	measured irradiance	(W/m^2)
I'	estimated irradiance	(W/m^2)
I_B	beam horizontal irradiance	(W/m^2)
$I_{B,C}$	clear-sky beam horizontal irradiance	(W/m^2)
I_D	diffuse horizontal irradiance	(W/m^2)
$I_{D,C}$	clear-sky diffuse horizontal irradiance	(W/m^2)
I_G	global horizontal irradiance	(W/m^2)
$I_{G,C}$	clear-sky global horizontal irradiance	(W/m^2)
I_m	mean irradiance	(W/m^2)
IMRM	Improved Meteorological Radiation Model	
k	hourly or sub-hourly diffuse ratio	
k_b	hourly or sub-hourly beam clearness index	
k_t	hourly or sub-hourly clearness index	
N	cloud cover	(Okta)
n	number of points	
m	air mass	
MBE	mean bias error	(% - W/m^2)
MRM	Meteorological Radiation Model	
PRM	Page Radiation Model	
Kd	mean earth-sun distance correction factor	
RMSE	root mean square error	(% - W/m^2)
REST2	Gueymard's Radiation Model	
R^2	coefficient of determination	
SF	sunshine fraction	
T_{LK}	Linke turbidity coefficient	
YRM	Yang Radiation Model	

Greek symbols

α	solar altitude angle	(Radians)
β	Angstrom's Beta coefficient	
σ	standard deviation	
δ_r	Rayleigh optical depth	(m)
τ_a	aerosol transmittance coefficient	
τ_g	gas transmittance coefficient	
τ_{NO2}	NO_2 transmittance coefficient	
τ_o	ozone transmittance coefficient	
τ_r	Rayleigh transmittance coefficient	
τ_w	water vapour transmittance coefficient	
τ_α	Mie transmittance coefficient	

AIMS AND OBJECTIVES

This research aims to create a series of simplistic broadband solar radiation models, requiring easily available meteorological parameters for use by Engineers and designers of solar systems to simulate the potential project when measured solar radiation is non-available.

All-sky condition solar radiation models are used to simulate long-term operation of a solar facility. When cloud cover is the only parameter available at the local location, then a model that is based on this parameter could be used. The Cloud Radiation Model was thus created.

However, if more parameters are available, such as humidity, temperature and sunshine duration, then another type of models can be used, such as the Meteorological Radiation Model. In fact this model was improved in the research and hence called the Improved Meteorological Radiation Model.

Clear-sky condition solar radiation data are needed for calculating maximum or extreme solar gains. These data are important for building services engineers for example to calculate maximum heat gains for buildings and thus calculate effective sun-shading and ventilation or air-conditioning loads. However, it is very difficult to extract clear-sky data from all-sky databases as this often require the use of extra measurements and multiple meteorological parameters. To solve this problem, a simple to use method was developed to extract clear-sky data by using the global and diffuse irradiance components as well as cloud cover and atmospheric turbidity. In fact, by using this procedure, there is a possibility of selecting blue-sky data as well as quasi-clear-sky data. The effect of choosing absolute clear-sky solar radiation on clear-sky models was studied and conclusions were made.

1 Introduction

The work done for this thesis relates to broadband solar irradiance, data quality control and all-sky as well as clear-sky modelling based on readily available meteorological data.

1.1 Energy and humanity; demand and supply

Since recorded human history, humanity has relied on their own physical power as well as animal power to produce their work and ensure their livelihoods. Later they have managed to harness wind to drive mills in order to produce flour and other cereals, essential ingredients of our diets.

Ever since the industrial revolution, the great fossil fuels- coal, then petrol and natural gas- have progressively reached the status of essential commodities in our lives. The improvements in industrial products, transport have only sped the demand for energy. In the early 20th century, the worldwide commercial consumption of energy was of the order of 500 million tons of equivalent petrol, Tep to satisfy a population of 1.6 billion people. A century later, the demand has multiplied by a factor of 18 to reach 9 billion Tep, while the population has multiplied by a factor of 5.6. This equates to approximately 1.6 Tep per person per year. Yet this figure hides the true image of inequality of the energy distribution in the world. In fact the United States uses roughly 25% of the world energy demands, while its inhabitants represent only 5% of the worldwide population. In comparison, China uses 10% of the world energy while hosting 35% of the world's population. In fact the average American resident uses 8 Tep compared to 3.5 Tep for a European resident while European salaries are 75% of their American counterparts! It is estimated that 2 billion individuals do not have access to the sources of modern energy, i.e. electricity and petrol. These individuals still use wood and wood coal as their primary energy resource, thus putting more and more pressure on deforestation and desertification of world's land mass. These inequalities do correlate strongly with the distribution of wealth (Chevalier J-M, 2004).

The 20th century has seen landmark inventions such as the development of engines- the internal combustion engine, electric motors, turbines and turbo-reactors - that replaced the steam engine for large scale power demands, while small-scale electric motors have invaded the domestic sector, factories and offices as well as automobiles. In the last three decades of the last century, electricity has become a necessity for human life sustainability since it is omnipresent in our every need in life.

Access to the sources of energy has become in a few decades the source of major strategic operations to ensure the proper functioning of economies, thus playing major roles in national and international politics and war efforts.

Many economic sectors grew only because of availability of energy, such as the automobile, aerospace, rail, naval, and electricity industries. The sources of energy have become a daily battleground between countries- importers and exporters- public and private organisations. The mechanisms involved in politics, economy, financial and environmental management are colossal and are the basis of the dilemma facing the world in the next century or two.

We can observe that our primary energy consumption derives 40% from petrol, 25% from coal, 25% from natural gas, and the remaining 10% from alternative sources of energy, namely hydroelectric, nuclear, and a mix of renewable sources. Indeed 90% of our consumption comes from non-renewable sources that are of limited stock.

What the future hides from us, we do not know, yet one thing is sure, we will use up our valuable sources of fossil fuels, the demand will continue to increase, new regulations will be enforced to manage extended globalisation, an extended dependence on fossil fuels on the medium term, associated with a renewed interest in nuclear as well further development in renewable energies. There are more geopolitical uncertainties as well that will be associated with the energy situation, climatic changes will occur and will have an important influence on policy making, new technologies will emerge, a possible new World Order will arise, and in whole, fundamental changes to our culture of energy dependence. Current forecasts for 2030 by major organisations namely the *Conseil Mondial de l'Energie,* CME, *Agence Internationale de l'Energie,* AIE, and the *World Energy, technology and Climate Policy Outlook*, WETO all describe a similar picture. The increase in consumption will reach 15 to 18 billion Tep, with nuclear increasing from the current 2000 MTep to between 3000-5000 MTep; petrol will increase from 4000 MTep to 6000 MTep on average, and gas will be the major player and will increase to 5000 MTep from the current 1000 MTep share. Renewable are due to see the major increase in market share, with the most conservative estimates placing the increase at 400% and in the most optimist forecasts at 2000 MTep compared with today's 400-500 MTep (Chevalier J-M, 2004).

The increase in gas is mainly attributed to a renewed interest in this resource and the new discoveries in Siberia (Sakhalin fields co-owned by Gasprom and Shell) and the extraction from current oil fields. The extra investments in infrastructure have a very positive influence on demand. The enforcement of current greenhouse gases legislations in the developed nations is smoothing the anti-nuclear position of the public that was caused by the nuclear incidents in the late 70's and early 80's, namely the Chernobyl meltdown.

The wars in the Middle-east were caused by developed nations worrying about their security of supplies and the health of their economies. In fact there is a clear indication that the first Gulf war has improved the economies of the USA and most major European economies. mostly due to heavily subsidised and low prices of crude petrol- $30/barrel- on the commodities market. The scenario has change in the new century, terrorism and tensions in the Middle-east have tickled the sensitivity and volatility of the energy markets thus inducing price hikes of the crude petrol to a record $70 (at the time of writing this thesis) and thus increased wholesale prices to increases of up to 45% over the past two years (2004-2006). These eco-political conditions have forced developed nations to limit their dependency on fossil fuels and seek measures of securing their energy demands from more safer locations and closer to home. Some nations started creating huge emergency supplies of oil derivatives, enough to keep the markets stable for a few months. The US government emergency reserve allows it to remain unaffected for up to 160 days without rationing.

1.2 Engineering the three e's dilemma

Engineers, scientists and professionals are the providers of new technologies; they also are a source of innovation and supply of products and services. Since energy will be a major concern in the years to come, engineers hold a special role to play; they are central to architecting the future of societies by influencing decision-making and delivering solutions to the imminent crisis.

Engineers have to face a major problem in solving these problems, as engineering revolves around three main principles, the triple e: Energy, Economy and the Environment. The three elements not only act as constraints on the engineering solution but they are interconnected and interact between each other in complex ways, thus influencing the adaptability of the solutions.

The interaction between energy and economy has been demonstrated, so every energy engineering solution has to take into account this important complexity. In fact delivering the energy demands massive investments, yet markets need to be created to use this energy and thus recap the investment costs, however one might wonder how one can balance the transport and hence the security of transport of commodities without hindering the final wholesale costs.

It was also shown how energy interacts with environment, 25% of the world's population do not have access to modern energy sources therefore use the available semi-renewable forestry resource. Forestry is an important energy sources, yet leaving this resource intact has far greater benefits in medicine, oxygen and CO2 regeneration as well sustainability of life, flora and fauna diversity.

1.3 Alternative energies

Alternative energy is a term used to describe methods of conserving energy and non-fossil energy resources. Energy conservation is an important factor in reducing our dependency on energy. Proper thermal insulation, efficient lighting, hybrid cars, public transport; all these methods are the ways forward. In fact a major rule of economy is exhibited here, the higher the demand the more and cheaper the supply. Not only does this control the availability of the resource, but it also creates jobs and opportunities to innovate and produce the new engineering solutions to cope with the energy crisis.

Renewable energies are in fact solar! Indeed these are categorised differently to solar energy, which now is a term of directly harnessing the solar radiation, in fact, wind, tidal, hydro, geothermal and bio-fuel energies all are methods of harnessing the sun's energy indirectly.

Wind has gained prominence recently, as a leader of the renewable energy sector, with onshore and offshore farms being built weekly all over the world. In effect this surge in the market is caused by the influence of national green policy initiatives. Wind installations come in variable sizes, from the domestic 1kW units to larger 1MW units installed offshore.

Hydroelectric generation has played an integral role in bringing energy to remote areas in the world since the late 19th century. Big projects in the mid-west USA, such as the great Colorado projects, have created communities and spurred growth in economy where no economy ever existed. China harnesses a large part of its energy demands through hydroelectric dams on its numerous rivers. Yet this solution, even though makes perfect sense in economic and energy components of the engineering triangle, the environment suffers directly and indirectly from such solutions. Yet it is often debated that creating lakes just changes a natural eco-system by a man-made one. It is still to be seen what is the long-term damage caused by dams.

Tidal and wave energy solutions are still in their infancy, yet the potential of harnessing the potential energy is great. More research needs to be made, and more public awareness should be encouraged.

Bio-fuels, are a problematic issue to deal with. On one hand, bio-fuels are a natural solution to replace petrol in petrol hungry applications such as transport, and bio-fuels are considered renewable, since crops are replanted the following season, but the problem lies elsewhere. For every ton of bio-fuel produced, 10 tons of fresh, drinking water is consumed. Fresh water is an essential product for sustainability of life, and fresh water constitutes only 10% of all water sources, with most of it in hard to reach and remote locations. The United Nations forecast that the three major causes of worldwide conflicts and international political tensions are: The wars of Energy, the wars of Water and Ethnic and Religious wars. In fact all three are combined and inter-related in regions such as sub-Sahara Africa, Middle East, and South America. We can see its effects already all over the world. In fact compromising and substituting fresh water resource to get energy resources is a very dangerous and reckless attitude, when a lot of countries are currently having to sacrifice energy to obtain fresh water through desalination of sea water, and recycling wastewater. (Cleveland C.J. 2004).

Nuclear energy, even though it does not count as a renewable energy, is definitely a zero-greenhouse emission technology. It has its merits and demerits. It is noted though that nuclear does not have a clean record for safety represented by the accident at Three Mile Island in 1979, and Chernobyl in 1986, the fear of terrorism and the problem of storing radioactive materials is still to be solved. However countries such as France produce 77% of its electricity from 58 nuclear stations. In total 441 nuclear stations are still in operation around the world. Many experts in the energy sector agree that we have reached or are reaching peak supply of fossil fuels, and we are thus starting the decline in production. The problem is so major, that countries such as the UK, could find themselves in the next few years with massive shortfalls in electricity supply. Professionals do guarantee that European over-production could cover the energy gap, however this is only a short-term solution to an ever-increasing problem. In view of this the anti-nuclear sentiment in Europe is on the decline in certain countries and many agree that nuclear could solve our medium-term energy problems. Nuclear has to face competition from the new champion of fossil fuels- natural gas. In fact gas turbines are far more flexible than nuclear stations, as they are on-demand energy sources. Their size is also a contributor; in fact gas turbine stations come in different sizes, from 50MW to 600MW. Thus diversifying the locations of placing them. In fact both technologies can be counted in sustainable energy category, since the potential of co-generation is possible with both. This places gas turbines at an advantage as they could be placed closer to communities to which they serve heating and electricity. Nuclear comes in 4 sizes compared to six for gas turbines, these are: 600, 900, 1200 and 1400 MW. There are price considerations as well, a nuclear station costs 1500-$2500/kW while $1000-1200/kW for gas turbines, similarly it would take 6-8 years to build the former compared to 2 years for the latter. They also have their own complexities with regards to security, security of transport and supply guarantees. In addition neither is environmentally friendly, on one hand we have no carbon emissions, but have massive amounts of toxic radioactive waste to deal with, and on the other hand we have a prominent carbon emitter. (Chevalier J-M., 2004).

Using Alternative energies, especially solar energy, could be one of the most promising of the solutions with regards to Economy, Energy and the Environment thus solving the three e dilemma!

1.4 Harnessing solar energy

Solar energy can be split in two major categories, based on the method of harnessing the potential energy. Active and Passive harnessing are the two categories. Active implies that the use of mechanical or chemical processes are used to obtain the energy, while passive, as the word implies, gets the energy without any forced process. Active solar energy is then split into two categories, thermal and electrical. The differences between the two are logically attributed to how this energy is used.

Electrical energy is harnessed by using a chemical-electrical device called photovoltaic, PV, cells. PV cells are assembled together to produce modules of fixed rating and then modules can be assembled together to create strings or arrays, the backbone of a PV facility. One such facility is presented in Appendix D, a project installed at Napier University. PV cells come in different shapes and sizes, with different ratings and electricity output. They can be monocrystalline or polycrystalline modules can be fixed and bulky or can come in thin-film sheet form. In fact the differences are caused by the differences in the chemical processes to transform photons into electrons, and the associated costs and efficiency variations associated with the different technologies. On a larger scale, arrays of parabolic dish collectors can be used to concentrate the collective sun beams onto a single focus points. The heat generated, can create steam thus driving turbines to create electricity as well.

Active thermal solar energy can then be split into two categories, water and space heating. Solar water heating has been used for many centuries, in its crudest form, solar collectors are just bulky black metal water containers exposed to the sun. Although this ancient technology still remains in vogue, more effective technologies have emerged recently such as vacuum tube heaters as well as more perfected solar collectors.

Active space heating is associated with ventilation and heat recovery in such a way that hot air due to solar radiation is transported to heat the inside environment. Not that non-forced space heating comes into the category of passive solar energy. Passive heating is due to smart, energy conscious architectural design. More of this attitude is encouraged to adhere to the policy of smart use of energy and sustainability. Note that active and passive techniques can be combined. For example, a solar atrium designed to harness the sun energy during winter could have adverse effect during summer, thus what is used to heat during winter, can overheat during summer and then require means of forced ventilation and air-conditioning to cool the indoor environment thus increasing the energy demand. An aesthetic and energy conscious method is to provide shading, not any regular shading, but rather thin-film PV technology shading mechanism. This shading effect could also be useful to control glaring, an essential part of indoor environment control, a headache for most architects and engineers. In such a way both heat and electricity can be produced by harnessing the sun's radiation. (Cleveland, C.J., 2004).

1.5 Solar radiation data acquisition and engineering design

Renewable energy projects don't just get installed out of nowhere. There is often a very long and complex process of architectural, engineering and financial feasibility studies that need to be carried out before the final facility is decided, products are sourced and contracts for installation awarded.

Part of the engineering study is to assess the energy requirements of the facility and the choice of one or a combination of energy conservation procedures. When solar energy is considered, it is important to assess with relative accuracy the resource available. One method is to install measurement equipment to record the irradiance and illuminance on the site. However, in practice this is not feasible. The equipment costs are too high to allow for such a use, as well as long-term measurements are needed, this implies that the equipment needs to be put in place for long periods of time before the data is available. Note that long-term datasets are required to cancel out the seasonal variations on the radiation. The longer the datasets, the more the data follows a normal distribution, and lessens the effect of freak or extreme conditions.

Another way to access solar radiation data is to obtain long-term datasets from local and international networks of meteorological measurement stations. These datasets can either be terrestrial stations or satellite derived. Although these datasets are available for most urban areas in the developed countries, they are quasi-inexistent in rural areas of developed countries and certainly in developing countries. Thus where the resource is needed the most, the data is usually unavailable.

Not only is the data often unavailable, also the datasets are often too difficult to get hold of, or when they are widely available the cost of purchasing them is too prohibitive for the average engineering and architectural firm.

It was found that more common meteorological parameters are recorded in far more stations and do cover huge areas of the world. Parameters such as temperature, sunshine duration, rain amount, relative humidity, atmospheric pressure and often cloud cover are very common. Thus the need for solar radiation extrapolation is necessary.

Data from 15 stations from across the northern hemisphere were available for this study. These included one to all three components of the broadband solar radiation. In addition certain stations do provide extra meteorological parameters. More details are given in Table 1.1. The 15 stations used overall in this study cover two continents, Europe and Asia, then subdivided in different countries: Bahrain, Japan, India, Spain and the United Kingdom. Not only does each country represent a specific regional climate, however within the same country there are clearly defined local climates, and more discrete meso-climates. For example, the two Japanese sites listed in Table 1.1, while one site is classified as marine climate, using the Koppen classification, the other site is classified as Humid-Continental. More details about each site classification is available in Table 1.2. In addition a world climatic map is presented in Appendix D, Figure D.1. Similar, yet more subtle differences are hence observed when observing the climatic map, as well as the Table referred to above, in which Indian sites differ slightly between each other. A similar observation is made for Spanish sites. In addition, the sites utilised have human-imposed meteorological factors, such as atmospheric pollution caused by urbanisation and industrialisation. In fact these human factors influence greatly the availability of solar radiation for solar energy applications as well as agriculture and other applications by the variation of atmospheric turbidity and the effect it has on absorption and refraction of solar irradiance. Monthly-mean averages of Linke Turbidity as well as ozone depth are provided in Appendix D, Figs.D.2-D.7 and D.8-11, for Linke Turbidity and ozone depth, respectively.

1.6 Data assessment

Measurement stations vary in size and operation. Most stations do operate remotely, with very little human involvement other than the frequent maintenance, while other stations are very well staffed. This does reflect on the quality of the data. In addition certain stations only measure one meteorological parameter using only one instrument, while in other stations; batteries of instruments are available to measure every possible meteorological and atmospheric parameter possible.

There are five types of solar radiation measurement stations. The best station is located at the World Meteorological Office, WMO, in Switzerland. This station is often termed as the world reference. Instruments at this station are inter-calibrated and offer reference calibration to all other stations in the world. Every country has a reference station as well. The instruments at this station are usually calibrated at yearly interval with the WMO instruments. These nation reference instruments serve as calibration reference to other station of the network. Other stations are ranked as 1st to 3rd Class stations, and are classed based on the accuracy of the measurements taken.

Yet even in 1st class measurement stations, errors occur, and these errors need to be identified and removed, so the errors are not transmitted to the designs created by engineers and architects who rely on this data.

In Chapter 2, various methods of assessing the quality of the data will be explained. These methods vary in complexity and rely on numerous control parameters to assess whether solar radiation data collected are valid for use by professionals. A new method to identify erroneous data was developed and will be presented in Chapter 3. This method relies is based on physical tests to assess the validity of the measurement to identify extreme outliers, and a statistical process to eliminate smaller outliers.

1.7 Present work

When measured solar radiation data are unavailable, they can be extracted from other meteorological data through correlations between associated measured synoptic parameters. These regressions are the basis of estimation models

There are many different types of models, and they vary by complexity and by the availability of the input parameters required. In engineering, broadband irradiance is often sought and it can be obtained from a combination of input parameters such as temperature, humidity, atmospheric turbidity, cloud cover and sunshine duration. Other parameters could be added, and the more influencing factors are used, the more accurate the estimation, yet increasing the complexity of the model. It is important to take into account the need for simple yet accurate models for engineering use. Also the availability of the data is an important factor in selecting the models.

In effect the work done in this thesis tackles three aspects of solar energy engineering: Measurement and errors, modelling and applications.

Solar energy professionals require extreme conditions for maximum loads scenarios and all sky conditions for simulations. Getting clear-sky data has its own set of complications. For instance there are no specific instruments that record clear-sky broadband solar radiation. These data need to be extracted from all-sky datasets. An evaluation of such procedures will be available in Chapter 2 and more in depth in Chapter 6. Available and trusted sky-identification techniques were examined and new easier and accurate methods were devised. These newer techniques were compared to older methods by using four established clear-sky broadband solar irradiance models. It is noted that all these models require very accurate and often difficult to obtain atmospheric parameters. Therefore, these models have also been used using more coarse atmospheric reading, such as monthly averages, and the results of this experiment was thus recorded and compared to experiments done by peers using more accurate measurements.

For system performance simulations, designers need hourly or even sub-hourly radiation day-long data. Thus, there is a need for long-term all-sky radiation. This type of modelling is usually the combination of extreme-sky, i.e. clear- and overcast-skies as well as mixed-sky algorithms. For mixed skies, sunshine and cloud related meteorological parameters are often used. When sunshine information as well as temperature and humidity are available, models such as the Meteorological Radiation Model, MRM, can be used. One such model was developed and is based on the original MRM with multiple improvements. In Chapter 4, the newly proposed model was compared to its predecessor to assess the improvements.

Another parameter often used in modelling is cloud cover. Thus, when cloud cover data are available, professionals can use Cloud Radiation Models, CRM, to estimate the available solar resource. Cloud based radiation models, were created by various researchers since the 80's and are still used by engineers well into the 21st century. A new CRM was developed to address certain weaknesses in older established models. The result of this study will be presented in Chapter 5.

Note that error-free data needs to be available to create and validate the models, since the models can be only as good as the original data on which they are based. Therefore, broadband solar radiation measurement techniques and their associated measurement errors are investigated. To correct these erroneous measurements, methods are developed by using global and diffuse horizontal irradiance. This work will be presented in Chapter 3.

Since measurements of solar resource are not as wide-spread as other atmospheric parameters, solar irradiance could be extracted from associated atmospheric parameters. However since each of the models requires different input parameters, and not all sites contain the required parameters, the choice of sites differ for the evaluation of the models described above. More details are given in Table1.2.

Once the data is finally produced, designers can use this information to create an accurate feasibility study on the project undertaken. One such project was done at Napier University, in March 2005. A 160m^2 PV facility was launched, capable of producing up to 17kW$_p$. A sample feasibility study was undertaken using both modelled data and long-term measured data. To complete the feasibility study, a life cycle analysis was undertaken to cover the energetic, economic and environmental liabilities of the facility.

Table. 1.1 Details of available datasets used in this research.

Location	Radiation Measured G=global, D=diffuse, B=beam	Temperature DBT, WBT	Radiation measurement Frequency Hourly/ 10 minute	SF measurement Frequency Daily/Hourly/10 minute	CC measurement Frequency Three times daily/Hourly
Aldergrove	G,D	none	Hourly	none	Hourly
Bahrain	G,D	D,W	Hourly	Hourly	none
Bracknell	G,D,B	D,W	Hourly	Hourly	Hourly
Camborne	G	none	Hourly	none	Hourly
Chennai	G,D	none	Hourly	Daily	Three times daily
Edinburgh	G,D	none	Hourly	none	Hourly
Eskdalemuir	G,D	D,W	Hourly	Hourly	Hourly
Fukuoka	G,D,B	D,W	10 minutes	10 minutes	none
Gerona	G,D	none	Hourly	Daily	Three times daily
London	G,D	None	Hourly	none	Hourly
Madrid	G,D	None	Hourly	Daily	Three times daily
Mumbai	G,D	None	Hourly	Daily	Three times daily
New Delhi	G,D	None	Hourly	Daily	none
Pune	G,D	None	Hourly	Daily	Three times daily
Sapporo	G,D,B	incomplete	10 minutes	10 minutes	none

Table. 1.2 Datasets used in each Chapter.

Location Name	Country	Longitude and Latitude	Climate*	Span	Quality Control	IMRM	CRM
Aldergrove	UK	54.65N; 6.25W	Marine, Cfc	1990-1994	N	N	Y
Bahrain	Bahrain	26.22N; 50.65E	Subtropical-Desertic, BWh	2000-2002	Y	Y	N
Bracknell	UK	51.26N; 0.45E	Marine, Cfc	1990-1995	Y	Y	Y
Camborne	UK	50.22N; 5.32W	Marine, Cfc	1999	N	N	Y
Chennai	India	13.0N; 80.18E	Tropical-Savannah, Aw	1990-1994	Y	Y	Y
Edinburgh	UK	55.95N; 3.35W	Marine, Cfc	1990-1995	N	N	Y
Eskdalemuir	UK	55.32N; 3.20W	Marine, Cfc	1995-1999	Y	N	N
Fukuoka	Japan	33.52N; 130.48E	Marine, Cfb	1992-1994	Y	Y	N
Gerona	Spain	41.97N; 2.88E	Mediterranean, Csa	1995-2001	Y	Y	Y
London	UK	51.23N; 0.46W	Marine, Cfc	1993	N	N	Y
Madrid	Spain	40.45N; 3.73W	Mediterranean, Csa	1999-2001	Y	Y	Y
Mumbai	India	19.12N; 72.85E	Tropical-Savannah, Aw	1990-1994	Y	Y	Y
New Delhi	India	28.60N; 77.20E	Humid-Subtropical, Cw	1988-1998	Y	Y	N
Pune	India	18.53N; 73.85E	Tropical-Savannah, Aw	1990-1994	Y	Y	Y
Sapporo	Japan	43.05N; 141.33E	Humid-Continental, Dfb	1991-1993	Y	Y	N

* Based on the Koppen Climate Classification System.

2 LITERATURE REVIEW

2.1 Introduction

The fundamental of terrestrial solar radiation is fairly straightforward. Solar radiation leaves the surface of the sun and emanates in all directions into space. The radiation travels through space and lands on any surface it crosses. In this fashion the earth is irradiated by the sun. Note that the earth is not at a circular path around the sun, rather as with the other planets in our planetary system, the path of the planets around the sun is elliptical. Therefore the solar radiation is variable depending on the tempo-spatial coordinates of the earth; a distance correction factor therefore applies to estimate the terrestrial solar irradiation. Note that there are no regular solar radiation attenuation factors in the portion of space between the sun and earth, although solar eclipses, caused by the other planets and the moon coming between the sun and earth, do occur at regular intervals and cause in the case of a total lunar eclipse a quasi-total shadowing of the solar radiation.
The earth like some other planets in our planetary system does have a thick atmosphere composed mainly of carbon dioxide, oxygen, ozone, water vapour and various other gases. These gases have an attenuation factor on the terrestrial solar irradiance. The distance corrected solar radiation, before entering the earth's atmosphere is termed the extraterrestrial irradiation. Some components of the atmosphere absorb the solar irradiance other reflected back to space and some just scatter it through the atmosphere. The major contributors to this phenomenon are the following atmospheric components: Ozone, water vapour, aerosols and other pollutants and most importantly cloud formation, smoke, haze, smog and fog. After passing through the atmosphere the solar irradiation is termed as global horizontal irradiation. The global horizontal irradiation can be split into two components; the diffuse horizontal and the beam horizontal irradiation. These components of the terrestrial solar irradiation are used most commonly in this study.

Terrestrial solar radiation modelling has been studied rigorously since the end of World War II by various researchers in Europe, Russia, Japan, India and North America. To study the phenomena of solar radiation it is necessary to measure the radiation. To this end, instruments have been crated to measure the different components of the solar radiation. The instruments are of various precisions and built qualities so the World Meteorological Office, WMO and the *Comité Internationale d'Eclairage,* CIE, has established measurement equipment classification based on the sensitivity and error tolerances. Similarly measurement stations are also classified based on the equipment classification. There are errors in measurement of solar radiation, these are due to inherent equipment errors, due to manufacturing and other processes or the errors are due to operation related errors. To be able to use the data with confidence, it is required to identify and correct or remove the erroneous data. This is done via means of quality control.

Solar radiation modelling is used to obtain estimations of the solar radiation for locations where this is not measured. Models rely on two major elements to estimate the solar radiation. The first part is the solar geometry component of the model. This in effect simulates the positioning of the sun at any given time. The second part simulates the state of the sky coverage and atmospheric conditions. In all cases the models rely on regressions with other meteorological parameters in order to estimate the irradiation. Some models require one meteorological parameter, other multiple. Models have been also created for specialized sky conditions. The three main types of sky types are as follows: clear-sky, mixed-sky and overcast-sky. Certain models are specific for certain sky types; others are more general and are for all sky conditions. Some models use and estimate spectral irradiation, which are based on wave length bands of the solar irradiation, other models are broadband. In this research work, broadband solar irradiation was used. Therefore the models for estimating solar irradiance vary in complexity, and therefore the availability and use of the models vary. The models studied in this work are mostly simple, engineering based models that require few input parameters to estimate the broadband solar irradiance.

2.2 Measurement of solar radiation and meteorological parameters

Broadband terrestrial solar irradiation is constituted of three components. Diffuse irradiation, beam irradiation and finally the global irradiation. These components are measured at a horizontal plane, or at a pre-defined tilted plane, depending on the use. Typically measurement stations record horizontal irradiation.

The global irradiation is measured using a device called a pyranometer. The world reference pyranometer is stored at the world meteorological office 'WMO' in Switzerland, and each country has a reference pyranometer, that is calibrated yearly to the WMO pyranometer, stored at its meteorological office. Measurement stations pyranometers are calibrated to the county's reference pyranometer.

The diffuse irradiation is measured using a pyranometer that is fitted with a shade-ring. The shade-ring as its name suggests is a ring that shades the direct portion of the sun, thus only recording the scattered irradiation.

Tilted global and diffuse irradiation is measured by tilting the instruments onto the desired angle.

The direct normal irradiation 'DNI' is measured using a pyrheliometer. Similarly to the pyranometers, the measurement station equipments are calibrated to a reference pyrheliometer that is in turn calibrated to the WMO pyrheliometer. The DNI is the measurement of the direct portion of the sunbeam at the sunbeam's normal plane with the earth. Horizontal direct irradiation is a geometric translation of the sunbeam onto the horizontal plane. Often direct irradiation is termed beam irradiation by reference to the sunbeam.

The two major manufacturers of pyranometers and pyrheliometers are Kipp & Zonen and Eppley. There are other players on the market, one of which is Delta-T. The Delta-T unit not only measures the horizontal global- and diffuse-irradiation and also the sunshine duration (See below). In effect this device replaces three rather expensive devices into one. Wood et al. (2003) have compared the Delta-T irradiation measurements with a Kipp & Zonen device. Note that both units were calibrated to the reference WMO pyranometer. They noted that the accuracy was similar. See Fig.2.1 a-f for images of a Kipp & Zonen pyranometer, a shade-ring installation- external and internal view, a Kipp & Zonen pyrheliometer, the Delta-T instrument, and details of the Delta-T instrument, respectively.

There is a relationship between the three components and the extraterrestrial component as described in the following two equations:

$$I_E = I_G + I_{reflected} + I_{absorbed} \qquad (2.1)$$

and,

$$I_G = I_D + I_B \qquad (2.2)$$

With, I_E the extraterrestrial irradiation, I_G the horizontal irradiation, $I_{reflected}$ being the irradiance that is reflected back to space, $I_{absorbed}$ the irradiance that has been absorbed by the atmosphere, I_B the beam or direct horizontal irradiation and I_D the diffuse horizontal irradiation. This inter-relationship is also represented graphically in Figure 2.2, the beam normal-, diffuse and global horizontal- irradiation.

Meteorological stations measure dry bulb- and wet bulb- temperature, relative humidity, atmospheric pressure, wind speed and wind direction and precipitation. These elements are often used in solar radiation modelling, however their affect on solar irradiation is minimal. Muneer (2004) has discussed the effect of each of the elements. He has concluded that alone they do not explain the attenuation of solar irradiation and thus, on their own they are unreliable methods for solar radiation modelling.

Muneer (2004) and Page (CIBSE Guide J) have demonstrated that sunshine duration and cloud cover have the most influence on solar irradiation attenuation. Cloud cover and sunshine duration or sunshine fraction have a strong correlation.

Cloud cover data are measured at numerous locations around the world, and are often recorded in Oktas, sky division in eight parts, or in percentage. Cloud cover is measured in three main methods, each have inherent errors. Some stations measure the column of sky above the station and record the time the column has hit clouds, and therefore record in percentage the cloud cover. Alternatively, as have been recorded traditionally, a trained technician at the station would record by visually monitoring the sky and noting the cloud cover. This is often recorded in Oktas. Finally, in the UK, at night, the Alidade unit would scan the sky in the eight Oktas and record the cloud cover. Alternatively, cloud cover is often derived from satellite imagery. This information is often referred to in literature as more accurate than ground-based observations, since cloud cover data is available for a wider number of locations by use of satellite-based imagery than terrestrial station networks.

Sunshine duration is recorded via two major instruments. A simple instrument to measure the sunshine duration over the length of a day is the Campbell-Stokes instrument. The Campbell-Stokes instrument is effectively a magnifying glass that burns a sheet of paper when the sunbeam is passing through it. The length of the sunburn indicates the sunshine duration. Kipp & Zonen have produced an electronic instrument that records sunshine duration to higher precision than the Campbell-Stokes instrument.

2.3 Measurement stations

Measurements stations vary from country to country and even from network to network within the same country. Most stations belong to meteorological offices and hence measure synoptic parameters, other belong to private institutions and education organisations. Each network has their own choice about instruments they use and the each measure different parameters depending on the use of the particular station.

Meteorological stations often record temperature, humidity, wind and rain. Others also record cloud data, and other sunshine data. Less frequently solar radiation is measured. Even when solar radiation is measured only one or two of the components are measured. In the UK for example up to 600 meteorological stations are operational at any time. Less than 300 would measure cloud cover and sunshine duration, while fewer than 100 would measure horizontal broadband diffuse and global irradiation. Only a handful of stations measure direct radiation. (BADC, 2005)

Solar radiation stations are classed in three categories depending on the classification of their instruments. UK met office stations are in most cases classified as first class stations. This classification guarantees the quality of the data that is gathered at the location and thus provides the general users of the datasets with relatively error free data.

Coulson (1975) classifies pyrheliometers and pyranometers in categories based on the uncertainties and errors of the measurements. A pyrheliometer is deemed first class if the measurement errors compared to a reference pyrheliometer are in the range of ±4% while it is deemed 2nd class if the errors are in the range of ±8%. Similarly, for pyranometers, they are deemed first class if the measurement errors compared to a reference pyranometer are in the range of ±10%, 2nd class in the range of ±25% and 3rd class in the range of ±32%.

Generally broadband horizontal radiation is measured, however in some stations around the world, they measure spectral radiation.

2.4 Sources of error

Any likely sources of errors or problems related to solar radiation measurement may be categorised under the following two major categories: equipment error and uncertainty and operation related problems and errors.

With any measurement there exist errors, some of which are systematic and others inherent of the equipment employed. The most common sources of error arise from the sensors and their construction. These are broken down into the most general types of errors and described below:

- Cosine response
- Azimuth response
- Temperature response
- Spectral selectivity
- Stability
- Non-linearity
- Shade-ring misalignment
- Dark offset (nocturnal) long-wave radiation error

In addition to the above sources of equipment-related errors, care must be taken to avoid operational errors highlighted below:

- Operation related problems and errors
- Complete or partial shade-ring misalignment
- Dust, snow, dew, water-droplets, bird droppings etc.
- Incorrect sensor levelling

- Shading caused by building structures
- Electric fields in the vicinity of cables
- Mechanical loading of cables
- Orientation and/or improper screening of the vertical sensors from ground-reflected radiation
- Station shut-down

Such errors are best highlighted via cross plotting diffuse ratio ($k = I_D / I_G$) against clearness index ($k_t = I_G / I_E$), and a sample plot is shown in Figure2.3. Note that any consistent errors emanating from an operational problem, such as misaligned shade ring are easily picked up by this type of plot.

Stoffel et al. (2000) give us a good representation of the scale of errors for carefully managed irradiance sensors. For their study, Stoffel et al. (2000) found that the range of error for a pyranometer compared with a reference pyranometer was from +2.5% to -10%; while for a pyrheliometer the range was ±2.5%. Myers (2006) proposes methods of calibrating instruments to reduce equipment related errors.

There are various other errors associated with the other measured synoptic data needed for this research. Sunshine data is especially prone to errors if the instrument used is a Campbell-Stokes instrument. The readings from this instrument indicate a daily sunshine duration. This is very useful for, say, daily or lower resolution modelling; however converting daily to hourly sunshine duration or fraction is not an easy exercise. In addition, this device does not accurately represent the state of the skies; a solar beam could burn a mark on the device, while the skies are intermittently cloudy. In this situation, the instrument would record a clear-sky, while in fact it is not.

Barker (1992) states that estimation of total cloud cover by real observations is subject to perspective errors and this causes inherent errors in the available datasets. Harrison and Coombes (1986) noted that the weather observer generally overestimates clouds. Brinsfield et al. (1984) go further and remark that an observer has a general tendency to underestimate the cloud cover under low overcast conditions and overestimate the cloud cover during high overcast conditions.

Bennett (1969) states that cloud cover explains less than 50% of insolation variance while sunshine fraction per example explains between 70-85% of the insolation variance. The reason for the weakness of cloud cover is that it does not take into consideration the type and depth of the cloud. Cloud type varies immensely the amount of scattering and shading of the terrestrial solar radiation.

2.5 Computational techniques

Computational techniques were utilised to compile the solar irradiation data and their related meteorological data in order to analyse the data and create the estimation models. Programming languages were used to facilitate this process. For data analysis, Visual Basic for Applications in the MS Excel environment was predominantly used. To compile and compare the final models with other established models, Fortran was used. This latter programming language is the most commonly used programming language by peers and thus utilising it for processing the data is an efficient way to do results comparisons. For comparison purposes, qualitative and quantitative procedures were taken into consideration. Graphing software was utilised in the qualitative process, as such as scatter plots and various histograms, while statistical indicators were calculated to obtain a quantitative comparison between the models.

2.5.1 Statistical indicators

The following statistical indicators were used in the research to quantitatively evaluate the performance of each of the models that were evaluated and validated. These were either calculated via programming routines or in some cases by using statistical packages such as SPSS. The reason behind this is explained in the following Section 2.5.2.

The slope of the best-fit line between the computed and measured variable is desired to be equal to one. Slope values exceeding one indicate over-estimation; while slope values less than one indicate under-estimation of the computed variable.

The coefficient of determination, R^2 is the ratio of explained variation to the total variation. It lies between zero and one. A high value of R^2, thus indicating a lower unexplained variation, is desirable. R^2 is often used to judge the adequacy of a regression model but it should not be the sole criterion for choosing a particular model as the value of R^2 increases with the number of coefficients in the model.

$$R^2 = \frac{\sum (I' - I_m)}{\sum (I - I_m)} \tag{2.3}$$

The MBE provides an indication of the trend of the model, whether it has a tendency to under-predict or over-predict its modelled values. MBE can be expressed either as a percentage or as an absolute value. Nevertheless, within a data set overestimation of one observation can cancel underestimation of another. An MBE nearest to zero is desired. It is given by the following equation:

$$MBE = \frac{\sum (I' - I)}{n} \tag{2.4}$$

The RMSE gives a value of the level of scatter that the model produces. This is an important statistical test, as it highlights the readability and repeatability of the model. It provides a term-by-term comparison of the actual deviation between the predicted and the measured values. Since it is a measure of the absolute deviation, RMSE is always positive. A lower absolute value of RMSE indicates a better model. Mathematically it is given by the following equation:

$$RMSE = \sqrt{\frac{\sum (I'-I)^2}{n}} \qquad (2.5)$$

Note that in Eqs. 2.3-2.5, I' is the estimated-, I the measured and I_m the mean value of the given irradiance component and n the number of data points.

Skewness is defined as a measure of the lack of symmetry in a distribution. A positively skewed distribution tails off to the high end of the scale while negative skew tails off the low end of the scale. If the distribution is normal or, in other words, has no skewness, then the skewness statistic will be zero. This will indicate a robust model.

Kurtosis is defined as a measure of the degree of peakedness in the distribution, relative to its width. The kurtosis statistic will be zero (mesokurtic) for a normal distribution, positive for peaked distributions (leptokurtic) and negative for flat distributions (platykurtic). A leptokurtic distribution of the errors is highly desirable. A high positive value of kurtosis represents that there are fewer outliers in the estimation.

2.5.2 Programming

Two programming languages have been used during the course of this research, VBA in MS Excel and Fortran. The reason behind this duality of programming interfaces resides in the fact that both tools are widely used by the solar energy professionals and engineers. In addition, they tend to complement each other as the weaknesses of one are addressed in the other.

Often during the course of this research, both were used simultaneously or one after the other to achieve the desired computation effect.

The main reasons behind using the Fortran environment, is its extensive inbuilt mathematical libraries. In addition, most fellow researchers use this medium to create and validate models. Therefore to ease the use of other author's codes, Fortran was used. It is important to use the original codes used by the authors when possible, this reduces the risk of compiling errors and other language migration errors.

Excel on the other hand is user-friendlier to use than Fortran, especially by combining the tabular data acquisition of excel with the inbuilt Visual Basic for Applications, VBA, bundled with the software. In addition to this, some functions in Excel are particularly suited for engineering modelling as described by Liengme (2003) and are available in VBA (Jacobson, 2001). However Excel has two major flaws or limitations, Excel cannot tabulate more than 65'000 rows, and cannot graph more than 22'000 data points. When more than the allowed number of points are needed for validation or modelling purposes, VBA can handle external files in a similar fashion to Fortran. For graphing purposes, other packages have been used, such as, Origin, Kaleidagraph and SPSS (Pallant, J.). Excel, Origin and Kaleidagraph have been used to create scatter plots, while histograms were plotted by Kaleidagraph and SPSS.

In some instances the statistical evaluation of some models were done directly in SPSS or Kaleidagraph if the respective software were used for graphing purposes.

2.5.3 Dataset compilation

Most raw datasets used, were previously compiled by Claywell (2003). Other datasets were recently added to the large worldwide dataset available at our research facilities. These new datasets are courtesy of the UK Meteorological Office and the National Renewable Energy Laboratories, NREL as well as private donations by universities and research centres.

The radiation and other meteorological data were collated together in the correct time frame and time stamp. To this end, Excel was mostly used to compile the datasets and export the resultant arranged data into single files per location. For the turbidity and ozone data, these were extrapolated from multi-chromatic figures supplied by the SODA network. The details of the datasets used in this research will be presented in more details in later chapters.

2.6 Quality control of solar radiation datasets

As was seen in Section 2.5, there are many sources of errors in solar radiation measurements, and therefore dataset can be ridden with outliers, thus rendering the dataset obsolete. There is thus a need to quantify the errors by identifying outliers, and then either omit the erroneous data or correct it whenever this is possible. Various researchers have addressed the issue in the past and these are exposed in this section.

2.6.1 Page quality control

The Page model is based on work undertaken for production of the European Solar Radiation Atlas, ESRA, by Greif and Scharmer (2000) and the CIBSE Guide on weather and solar data (1997). Page sets upper and lower boundaries for diffuse irradiation and also sets an upper boundary for global irradiation. For the former component the overcast and clear-sky irradiance respectively set the upper and lower limits. For the latter component the upper limit is set by global clear-sky model.

The Page clear-sky model computes hourly beam and diffuse irradiances under clear-sky conditions thus,

$$I_{B,C} = 1367 K_d e^{(-0.8662 m T_{Lk} \delta_r)} \sin \alpha \qquad (2.6)$$

$$I_{D,C} = K_d T_{rd} F(\alpha) \qquad (2.7)$$

With $I_{B,C}$ and $I_{D,C}$ the beam and diffuse irradiances under clear-sky conditions respectively, and K_d is the mean earth-sun distance correction factor. The relative air mass 'm' takes account of the presence of gases, liquids and solid particulate matter in the atmosphere.

The global irradiance $I_{G,C}$ under a clear-sky is simply the sum of the beam and diffuse components.

The Linke turbidity factor T_{LK} applies throughout the electromagnetic spectrum. Values of T_L at air mass of 2 are typically used in Page's model. T_{LK} data are readily available on a monthly basis for many European locations via the SODA network (2004). The Rayleigh optical depth δ_r is an attenuation coefficient due to Rayleigh scattering.

The diffuse transmittance T_{rd} is the theoretical diffuse irradiance on a horizontal surface when the sun is the zenith. Thus,

$$T_{rd} = -21.657 + 41.752 T_{LK} + 0.51905 T_{LK}^2 \tag{2.8}$$

The solar elevation function F(α) is a polynomial function of the sine of the solar elevation,

$$F(\alpha) = x_0 + x_1 \sin\alpha + x_2 \sin^2\alpha \tag{2.9}$$

The coefficients used in Eq. 2.9 are 0.0382, 1.5458 and -0.5998 for x_o, x_1 and x_2, respectively for clear-sky condition; while for overcast conditions, the coefficients are -0.0067, 0.7860 and 0.2240 for x_o, x_1 and x_2, respectively.

Under overcast skies, global ($I_{G,OC}$) and diffuse ($I_{D,OC}$) irradiances are equal due to the absence of the beam component. Thus,

$$I_{G,OC} = I_{D,OC} = 572\alpha \tag{2.10}$$

Further details and software can be found for the Page model by Muneer (2002).

2.6.2 Helioclim quality control algorithm

Geiger et al. (2002) have described the availability of a web-based service for quality control of solar radiation data. The service is available through the web site www.helioclim.net. The quality control procedure is part of an on-going effort of the Group 'Teledetection and Modelisation' of the Ecole des Mines de Paris/Armines to provide tools and information to the solar radiation community through the worldwide web. The object of that service is not to perform a precise and fine control but to perform a likelihood control of the data and to check their plausibility. This is achieved by comparing observations with expectations based upon the extraterrestrial irradiation and a simulation of the irradiation for clear skies. It offers a means to check time series of irradiation data. Inputs are provided via an HTML page by a copy and paste procedure and the return is also via similar means. Suspicious data are flagged upon return.

The user is requested to provide information to compute the quality control procedure: Geographical co-ordinates, elevation and dates. HTML pages are available to better understand and fill the forms for each quality control procedure. Documents explaining the algorithm used in the calculation and references to articles, web sites of interest and publications on solar radiation topics are also provided.

The quality control procedure has been divided into four HTML documents:

a. Daily irradiation for a single day: Single-value examination.
b. Daily irradiation for several days: Several daily values spread over several months and years to be analysed.
c. Hourly irradiation for a single hour: Single-value examination.
d. Hourly irradiation for several hours: Several hourly values spread over several days, months and years.

The quality algorithm used in the Helioclim website is part of the Soda project (2004).

The Helioclim algorithm provides an interval of acceptance for hourly global irradiation data. The algorithm has been designed for locations with noon solar altitude angle greater than 2°.

$$\text{Upper limit} = \text{Min}\ (1.1\ I_{G,C}\ ,\ I_E) \tag{2.11}$$

$$\text{Lower limit} = 0.03\ I_E \tag{2.12}$$

2.6.3 Molineaux and Ineichen's web based procedure and tools

Molineaux and Ineichen (2003) describe the availability of yet another web-based facility for quality control of solar radiation data. Their computer programmes allow validation limits to be set on the tests so as to enable the user to increase the precision of the tests. The programme reads an input file based on the ASCII format and in turn creates an output file in the same format with AQC flags. Visualisation of the comparisons between measured and predicted values (based on well established solar radiation models) is used to trace the errors. Their programme carries out a series of coherence tests which is then followed by the creation of a number of plots based on comparisons between modelled/calculated and measured values.

2.6.4 NREL SERI QC programme

The US-based NREL (National Renewable Energy Laboratory) has developed alternate procedures and software for performing post-measurement quality assessment of solar radiation data. The assessments are also performed on the uncertainty of measured solar radiation data. In this respect a quality assessment software package SERI QC was developed by Maxwell et al. (1993) from NREL to address the above needs. SERI QC is based on the establishment of boundaries or limits within which acceptable data are expected to lie. This is similar to previous quality assessment procedures that use extraterrestrial values for the upper limit and zero for the lower limit within which solar radiation data were expected to lie. SERI QC increases the sophistication of the latter approach by establishing much more restrictive boundaries specific to each station month. SERI QC operates in a dimensionless space, i.e. solar radiation normalised to extraterrestrial values. An example of the expected limits and boundaries established by SERI QC is given in Figure 2.4. The variables that form the abscissa and ordinate in this figure: k_B the atmospheric transmission of the direct beam radiation defined as, $k_B = I_B / (I_E \sin\alpha)$.

The hourly values plotted in Figure 2.4 were the data collected by NREL for Nashville, Tennessee for the period April 1978–April 1980. Established empirical limits and boundaries of acceptable data for this station are also shown within the latter figure. The heavy dashed lines represent the expected maximum global horizontal, and direct normal values and the curved boundaries around the scatter plot of the data were empirically determined by these data. This was implemented by positioning a limited set of boundary shapes around the data. The position of the boundaries was then adjusted in k_t increments of 0.025 such that up to 5% of the data lay outside the boundaries.

This criterion was based both on the assumption that some of the data were in error and a desire to limit the acceptance of erroneous data to small percentages.

The three parts of Figure 2.4 show data, maximum–minimum limits and boundaries for three different air masses. SERI QC assigns limits and boundaries for three air mass ranges (low: 1.0–1.25, medium: 1.25–2.5 and high: 2.5–5.58). Changes in limits and boundary positions with smaller changes in air mass are not significant.

When all three of the solar radiation elements are available (global horizontal, direct normal and diffuse horizontal) redundancies may be used to further reduce the uncertainty of the data. This is accomplished by calculating the global from the direct normal and diffuse, and by comparing the calculated global with the measured global radiation. This comparison provides a direct indication of the accuracy of all three measurements.

Nevertheless, when hourly values of global horizontal, direct normal and diffuse horizontal radiation agree within a specified error limit, the lowest possible uncertainty for the data can be assigned. In addition to determining whether the solar radiation data fall within expected boundaries, SERI QC calculates the distance in k-space by which data fall outside the boundaries.

The flagging system used by SERI QC records these distances and indicates whether one-element, two-element or three-element data were involved and whether the data point was below or above expected boundaries. The SERI QC flags, therefore, permit the assignment of uncertainties that are dependent on the nature of the test performed (one, two or three components) and the distance by which the data point exceeds expected limits.

A point worth mentioning is that once the above filtering process has been completed and the erroneous data removed, there is a need to fill-in the ‘holes’ within the dataset. Unless this procedure is undertaken the time series would be incomplete. Building energy simulation programmes in particular are prone to hick-ups with such problems. Gaps identified within the dataset may either be filled by generation of irradiation data from other synoptic data such as sunshine or cloud cover, or data averaging techniques. In this respect the reader is referred to the work of Muneer and Fairooz (2002).

Furthermore, Rymes and Myers (2001) have presented a method for smoothly interpolating averaged (coarsely resolved) data into data with a finer resolution, while preserving the deterministic mean of the data. Their technique claims to preserve the component relationship between direct, diffuse and global solar radiation when values for at least two of the components are available.

2.6.5 CIE Automatic Quality Control

The Commission Internationale de l'éclairage 'CIE' (1994) proposes the following quality control tests. They note that automatic testing should not be performed when the solar elevation is less than 4° and when the global irradiance is less than 20 W/m^2.

They propose five levels of tests; the first two are related to global, beam and diffuse radiation and corresponding illuminance. The third test is related to the north, east, south and west global irradiance and illuminance. The fourth level test involves inter-comparisons between irradiance and illuminance and finally the fifth level test compares the zenith luminance with either diffuse irradiance or illuminance.

Herein we are interested with the first two levels of tests from the CIE. These are described in more details below.

The first level tests are rough absolute checks that insure that no major problems exist.

$$0 < I_G < 1.2 I_E \tag{2.13}$$

$$0 < I_D < 0.8 I_E \tag{2.14}$$

$$0 < I_B < I_E \tag{2.15}$$

The second level tests are consistency tests that utilize the redundancy existing between direct, diffuse and global components:

$$I_G = (I_B \sin\alpha + I_D) \pm 15\% \tag{2.16}$$

For stations that do not measure the direct component:

$$I_D < I_G \pm 10\% \tag{2.17}$$

Note that the 10% margin is an allowance for shade ring correction.

2.6.6 Muneer and Fairooz quality control procedure

The Muneer and Fairooz (2002) quality control procedure is in addition to other filters a combination of tests based on the CIE quality control procedure (Section 2.7.5) and Page irradiance model (Section 2.7.1).

The Muneer and Fairooz model has four test levels. The first test is the CIE quality control method given by Eqs. 2.13 to 2.17. The level two tests include consistency tests between diffuse and global irradiation, and between global and horizontal extraterrestrial irradiation. The third level tests are based on an expected diffuse ratio (I_D/I_G) - clearness index (I_G/I_E) envelope. This test involves a check that the diffuse irradiation data conforms to the limits set out by an envelope of acceptance. A further fourth level check on the quality of diffuse irradiance is undertaken by comparing its value with the diffuse irradiance under the two extreme conditions, as defined by Page's clear and overcast sky model.

As a final (fifth-level) measure of check on global and diffuse irradiance data, turbidity is calculated for the given time-series and checked for its limits. A Linke turbidity value that is less than 2.5 (obtained under exceptionally clear skies) or greater than 12 (under dust storm conditions) would demand close inspection of data.

2.7 The Meteorological Radiation Model

The meteorological radiation model (MRM) has been developed by Muneer and Gul (2000), Muneer et al. (1998) and Gul et al. (1998) as a simple broadband irradiation estimation model based on synoptic information.

The MRM utilizes only four commonly measured variables namely, atmospheric pressure (P_{atm}), dry bulb temperature (DBT), sunshine fraction/

duration (SF/SD) and wet bulb temperature (WBT). WBT can be obtained via DBT and relative humidity (RH), if direct measurements of the former elements are not recorded.

The MRM is based on regressions between the ratio of hourly diffuse horizontal irradiation (I_D) to beam horizontal irradiation (I_B) and beam clearness index. Note that the above two quantities are herein referred as DBR= I_D/ I_B and K_B= I_B/ I_E where I_E is the extraterrestrial horizontal irradiation. Muneer and Gul (2000), Muneer et al. (1998) and Gul et al. (1998) have expressed the relationship between the above two dimensionless variables in the form of a power function as shown in Fig. 2.5 and represented mathematically as:

$$DBR = a(K_B)^b \tag{2.18}$$

They validated the MRM using data from the UK and Japan; the coefficients to use in Eq. 2.18 for UK are a=0.285 and b=-1.00648

The calculated beam horizontal irradiation (I_B) is a function of the extraterrestrial horizontal irradiation attenuated by the sunshine fraction (SF) and atmospheric transmittances thus,

$$I_B = I_E.SF.\tau_r.\tau_\alpha.\tau_g.\tau_o.\tau_w \tag{2.19}$$

τ_r and τ_α are transmittances due to Rayleigh and Mie scattering; τ_g, τ_o and τ_w the transmittances due to mixed gases, ozone and water vapor scattering, respectively. These are expressed mathematically as,

$$\tau_\alpha = \exp\left\lfloor -k_\alpha^{COF(1)}(1 + k_\alpha - k_\alpha^{COF(2)})m^{COF(3)} \right\rfloor \tag{2.20}$$

$$\tau_O = 1 - \left\lfloor 0.1611x_o(1 + 139.48x_o)^{-0.3035} - 0.002715x_o(1 + 0.044x_o + 0.0003x_o^2)^{-1} \right\rfloor;$$

$$x_0 = l_o m \tag{2.21}$$

$$\tau_r = COF(4) - COF(5)m + COF(6)m^2 - COF(7)m^3 + COF(8)m^4 \tag{2.22}$$

$$\tau_w = 1 - COF(9)x_w\left[(1 + COF(10)x_w)^{COF(11)} + COF(12)x_w\right]^{-1};$$

$$x_w = l_w m \tag{2.23}$$

$$\tau_g = \exp(-COF(13)m^{COF(14)}) \tag{2.24}$$

Refer to the work by Muneer and Gul (2000), Muneer et al. (1998) and Gul et al. (1998) for the values of above used coefficients for UK sites.

Once I_B is calculated, I_D is then calculated via Eq. 2.18. The calculated beam and diffuse horizontal irradiation is then summed up to obtain the calculated global horizontal irradiation (I_G).

2.8 *Cloud Radiation Modelling*

Simple radiation estimation models using meteorologically observed input parameters are often used in the applications requiring rough estimations of solar horizontal radiation. Solar radiation modelling based on cloud cover data is still used around the globe, since cloud measurements are easily available from ground measurement stations and satellite imagery. Kasten and Czeplak (1980), Gul et al. (1998), Muneer and Gul (2000), and Lam and Li (1998), have proposed cloud-based models for the estimation of global and diffuse horizontal irradiance. New uses for these models are constantly being investigated, such as the work by Perez et al. (2005) on solar radiation forecasting based on the Kasten and Czeplak model used on cloud cover forecasting services. Other commercial tools are being developed based on the Muneer and Gul model as presented by Bing (2005).

Kasten and Czeplak (1980) formulated Eqs. 2.25-2.27 for the estimation of solar irradiance based on cloud cover information. For their research, Kasten and Czeplak used ten years (1964-1973) of continuous hourly data from Hamburg, Germany; they also validated their model using German and UK datasets, thus, the coefficients they provide are for German and UK regional climatology. Initially they proceed by calculating the global horizontal irradiance under clear-sky conditions I_{GC}, and then the global and diffuse horizontal irradiance are calculated.

$$I_{GC} = 910.\sin\alpha - 30 \tag{2.25}$$

$$I_G = I_{GC}(1 - 0.75(N/8)^{3.4}) \tag{2.26}$$

$$I_D = I_G(0.3 + 0.7(N/8)^2) \tag{2.27}$$

Gul et al. (1998), and Muneer and Gul (2000) furthered the work of Kasten and Czeplak (1980) to provide equations that can accommodate local coefficients for their datasets, as the original coefficients could not accurately estimate the irradiance in their analysis. The Kasten and Czeplak Eqs. 2.25 and 2.26 have been modified while Eq. 2.27 remains the same in both models. Eqs.2.25 and 2.26 become as follows:

$$I_{GC} = A_1 \sin\alpha - B_1 \tag{2.28}$$

$$I_G = I_{GC}(1 - C_1(N/8)^{D_1}) \tag{2.29}$$

Lam and Li (1998) have explored the incorporation of multiple linear regressions between global irradiance and cloud cover involving solar altitude. The equations are given below,

$$I_G = 217 - 485(N/8) + 696\sin\alpha \tag{2.30}$$

$$I_D = 30.5 - 62.9(N/8) + 294.7\sin\alpha \tag{2.31}$$

The coefficients given in Eqs. 2.30 and 2.31 are for use for Hong Kong datasets only.

Note that for the above-mentioned models, the beam horizontal irradiance is calculated by subtracting the diffuse component of the global horizontal irradiance.

2.9 *Clear-sky solar radiation selection procedures*

The definition of clear skies is very loosely used in solar radiation modelling. Quite often skies are described as clear when in fact the skies are cloudless, as described by CIE (1973) in their three sky condition, where clear-sky is quasi-clear-sky. However, very few researchers provide a clear identification as to whether low and high turbid skies, under zero cloud cover should qualify as clear skies or only low turbid cloudless skies should be tagged as clear.

It is important to note that clear-sky irradiance data are extracted from long all-sky irradiance datasets. To this end, different methods have been used to classify sky conditions from the available synoptic parameters.

When cloud cover information is available with the irradiance components, cloudless skies – 0 oktas – are assumed representative of clear skies, hence the irradiance is assumed clear-sky. Sunshine fraction has often been used to classify sky conditions too. A sunshine fraction close to unity indicates the associated irradiance data to belong to clear-sky conditions. Either of the two parameters has been used independently to this end; however some people have combined both to get more accurate indication of clear-skies as shown in Section 2.8.1

Some other researchers have used the clearness index and diffuse ratio dimensionless plan to identify sky conditions. High k_t and low k values are representative of clear-skies.

Low turbidity skies are often identified using coefficients that quantify aerosols in the atmosphere. As clouds count as aerosols, it is concluded therefore that very low Linke turbidity coefficients are representative of cloudless low turbid skies, often referred to as absolute clear skies.

2.9.1 Synoptic based procedures

Cloud cover data are recorded at numerous locations around the world, and are often given in oktas, sky division in eight parts, or in percentage. Cloud cover is measured in three main methods, each have inherent errors. Traditionally, a trained technician at the station would record cloud cover by visually monitoring the sky and noting the cloud cover. This is often recorded in oktas. In the UK, the Alidade unit employs lasers to scan the sky in eight oktas and records the cloud cover.

Sources of errors relative to above measurements have been discussed by Barker (1992), Harrison and Coombes (1996) and Ododo et al. (1996).

Lam and Li (2001) adopted an absolute and restrictive limit for cloud cover being nil to represent cloudless skies. However to compensate for the bias accompanied by potential errors in recording, cloudless skies are considered to be less than one okta. Babaro et al (1981) have addressed matters as well and have classified clear skies in the cloud cover range of 0 to 2 oktas, inclusive.

Sunshine based classification of clear skies has also been adopted in literature, in which case a sunshine fraction close or equal to unity would represent a cloudless and thus a clear sky. In fact, this procedure has been discussed by Littlefair (1988), Muneer (2004), and Lam and Li (2001) who commonly agree that a value greater than 0.9 is representative of the desired sky. They have also discussed the shortcomings of this method and have concluded that sunshine fraction only indicates whether the sun is blocked by the cloud and does not provide information relative to the other parts of the sky.

A combination of both cloud cover and sunshine fraction data could also be used to determine more accurately the sky conditions. It is important to note that the two parameters are interlinked statistically as was shown by Page in the CIBSE Guide J. (2002).

2.9.2 Sky clarity indices and turbidity based procedures

Solar horizontal terrestrial irradiation is often used to describe the sky type by use of dimensionless ratios or indices that are based on different components of solar irradiance. The most common of indices are the clearness index k_t, the diffuse ratio k, the clearness function k_b and the diffuse coefficient k_d. k_t is the ratio of global horizontal terrestrial irradiance I_G to extraterrestrial irradiance I_E. As its name indicates, the clearness index represents the clarity of the atmosphere from transmittance agents. The diffuse ratio indicates the amount of transmittance effect in the atmosphere. It is the ratio of diffuse horizontal irradiance I_D to I_G.

Any of the above four indices may be used to determine clear-sky conditions. Iqbal (1983) considers clear sky to be represented by a clearness index greater than or equal to 0.7 and less than 0.9. Ianetz et al. (2005) do not restrict an upper limit in the same way as Iqbal (1983) has, therefore a k_t greater or equal to 0.7 is considered to be clear sky. In addition, Ianetz et al. (2005) discuss the fact that this limit is not static and that it varies. In their study in Israel, they noticed that during the months of January and February, the kt limit is reduced to 0.67 and further to 0.65 for the months of November and December.

Alternatively, Thevenard and Brunger (2001) have used the diffuse ratio, upper limit of 0.4 and lower limit of 0.2, to determine clear sky conditions.

Perez et al. (1990) have proposed a method to identify clear-sky by using an alternative sky clarity index, $k_t^{'}$ that is a function of k_t and air mass 'm'. Clear-sky condition may occur if,

$$k_t^{'} \geq 0.7 \qquad (2.32)$$

$$k_t^{'} = \frac{k_t}{1.031.e\left[\dfrac{-1.4}{0.9+9.4/m}\right]+0.1} \qquad (2.33)$$

Ianetz et al. (2005), Chendo and Maduekwe (1994) and Collares-Periera and Rabl (1979) refute the existence of a clear-cut k_t value to define the sky conditions as each respective researcher have tended to use their own values depending on the location and the month of the year.

There are some shortfalls when using only one sky clarity index to determine sky conditions as compared to combining two indices as shall be shown in this work. As observed in Figure 2.6a at high values of the clearness index, which can be interpreted as clear-sky condition, the associated diffuse ratio ranges between 0.3 and 0.04. In addition, from Figure 2.6a with respect to the limits set by Thevenard and Brunger (2001) for the diffuse ratio, the associated clearness index values have a large scatter. Similarly, it was noted by Lam and Li (2001) that at high k_t, there is a wide scatter of k_d as represented in Figure 2.6b. Both scatter plots indicate that a combination of clear and partly cloudy skies are present at high k_t values rather than exclusively for clear skies.

Therefore, to indentify the sky conditions by use of the sky clarity indices, it is necessary to study the combination of two of the indices mentioned above. The most common combination are k_t and k.

Battles et al. (1998) have developed a lower limit for k_t (k_{tt}) and an upper limit for k (k_k). Both limits are functions of the solar geometry.

$$k_{tt} = -0.3262 - 0.0032\alpha + 0.6843\log\alpha \quad (2.34)$$

$$k_k = 1.0827 - 0.3893\log\alpha \quad (2.35)$$

The atmospheric turbidity may also be used to evaluate and classify the sky conditions. A Linke turbidity T_{LK} value of less than 3 has been described as representative of clear skies, and a value of near unity being very clear skies. This was discussed in ESRA, by Greif and Scharmer (2000).

Lam and Li (2001) have introduced a method to determine the upper and lower limit of k and k_t, respectively, by means of analyzing the cumulative frequency of occurrence in percentage of cloud cover, sunshine fraction, k_t and k. From the cloud cover analysis, they have concluded that there was a good statistical concordance between the cloud cover and the above-mentioned indices. Figures 2.7 a-c represent the cumulative frequency distribution of the cloud cover, clearness index and diffuse ratio, respectively for Chennai, a site used in this study. In the case of Figures 2.7 a-c at 9% cloud cover less than 1, the k_t minimum limit is 0.78 and the k maximum limit is 0.21.

Ineichen (2006) proposed a completely different approach that is based on the values of either the beam or global irradiance. The following conditions needs to be met to identify clear-sky irradiance.

$$I_G > 0.9 I_{Gc} \tag{2.36}$$

Where, $I_{Gc} = I_E . e^{\left[-2m(9.4+0.9m)^{-1}\right]}$ (2.37)

And $\Delta I_G = \dfrac{\left[I_G(n-1) + I_G(n+1)\right]}{I_G}$ (2.38)

Also $\left|\Delta I_G - \Delta I_{Gc}\right| \leq 10\%$ (2.39)

Long and Ackerman (2000) have derived a new method of determining clear-sky data from I_G.

$$\frac{I_g^{\varepsilon}}{1000} \geq \theta^{\varepsilon} \tag{2.40}$$

ϵ is a site dependent coefficient and its average value is 1.3

2.10 Clear-sky modelling

Clear-sky modelling is necessary for engineers when doing extreme cases feasibility studies. Many models exist for spectral irradiance, however these models are often too complex for the average engineer to use, especially due to the difficulty in obtaining the necessary data. Four broadband models are often used by professionals, clear-sky Page Radiation Model, Meteorological Radiation Model, Yang Radiation Model and Gueymard's REST2 model. These models are explained below and will be examined and compared in more detail in Chapter 6.

2.10.1 Clear-sky Page Radiation Model

The Page Radiation Model, PRM evolved from the development of the European Solar Radiation Atlas, ESRA. The specific clear-sky model of PRM was originally developed by Page and Lebens (1986) as a key component needed for the development of ESRA. This cloudless-sky model predicts the horizontal irradiance as a function of solar altitude and Linke turbidity based on air mass 2. The cloudless-sky beam and diffuse horizontal irradiance are estimated separately and then global horizontal irradiance is obtained by summation.

The beam horizontal irradiation is thus calculated,

$$I_B = I_E K_d \exp(-0.8662 m T_{LK} \delta_r(m)) \sin(\alpha) \tag{2.41}$$

with I_E the extraterrestrial irradiance, K_d the earth-sun correction factor, m the air mass, T_{LK} the Linke turbidity, δ_r the Rayleigh optical depth and α the solar altitude.

Rayleigh's optical depth can be obtained by using Kasten (1993) formula:

$$\delta_r(m) = \left[6.6296 + 1.7513m - 0.1202m^2 + 0.0065m^3 - 0.00013m^4\right]^{1} \tag{2.42}$$

The diffuse irradiation depends on the solar altitude and the Linke turbidity at air mass 2. The estimation of the diffuse irradiance is in two stages. First the theoretical diffuse transmittance, Trd(n) is established for day n and was given in Eq.2.8 while the solar elevation function F(α) is evaluated using Eq.2.9. Thus, the diffuse irradiance is given by:

$$ID = K_d Trd(n)\sin(\alpha) \tag{2.43}$$

Based on studies by Muneer (2004) on 5 UK locations, it was found that PRM, in clear-sky conditions (clearness index, K_t>0.6) yielded an average MBE=90.6 W/m^2 and RMSE=163 W/m^2.

2.10.2 Clear-sky Meteorological Radiation Model

This model is the clear-sky part of the all-sky MRM reviewed in Section 2.7. Thus the clear-sky MRM is a broadband horizontal irradiance model as well. Beam irradiance is estimated in a similar fashion as for all-sky conditions, however clear-sky algorithms are used to estimate the diffuse horizontal irradiance. The clear-sky model is based on the work of Dave (1979), Bird and Hulstrom (1979) and Pisimanis (1987).

Diffuse horizontal irradiance is calculated using the following equation:

$$I_D = I_E \tau_{\alpha\alpha} \tau_g \tau_o \tau_w \left[\frac{0.5(1-\tau_r)}{1-m+m^{1.02}} + \frac{0.84(1-\tau_{as})}{1-m+m^{1.02}} \right] \tag{2.44}$$

with, $\tau_{\alpha\alpha} = 1 - 0.1(1-\tau_\alpha)(1-m+m^{1.06})$ (2.45)

and, $\tau_{as} = 10^{-0.045 m^{0.7}}$ (2.46)

The global horizontal irradiance is given by

$$I_G = (I_B + I_D)\left(\frac{1}{1 - r_s r'_\alpha} \right) \tag{2.47}$$

where r_S is the ground albedo and r'_α the cloudless sky albedo. $r_s' = 0.0685 + 0.17(1-\tau_\alpha')$ with τ'_α the Rayleigh scattering transmittance calculated at air mass m=1.66.

Based on the study by Muneer (2004) on 10 UK locations, in clear-sky conditions (600-1000 W/m^2) the model has yielded an average absolute mean bias error, AMBE= 35.4 W/m^2, and an average root mean square error, RMSE=10.8 W/m^2. For another study on 5 UK locations and by using different clear-sky conditions (K_t>0.6) it was found that the average MBE=-45 W/m^2, and average RMSE=77.6 W/m^2. This shows that careful attention needs to be given to clear-sky selection as they influence the results of the model validations.

2.10.3 Clear-sky Yang Radiation model

This model is given by Yang et al. (2003) and is based on a product of atmospheric transmittances. The beam and diffuse transmittances are given with m' the atmospheric pressure corrected air mass and I_E the extraterrestrial irradiance.

$$\tau_B = \tau_r \tau_o \tau_g \tau_w \tau_a - 0.013 \qquad (2.48)$$

$$\tau_D = \tau_o \tau_g \tau_w (1 - \tau_a \tau_r) + 0.013 \qquad (2.49)$$

In effect, beam and diffuse transmittances are functions of the Rayleigh, ozone, gas, water vapour and Mie transmittances. These transmittances are calculated as follows:

$$\tau_r = \exp[-0.008735 m' F(m') - 4.08] \qquad (2.50)$$

$$F(m') = 0.5474 + 0.01424 m' - 0.0003834 m'^2 + 0.00000459 m'^3 \qquad (2.51)$$

$$\tau_g = \exp(-0.0117 m'^{0.3139}) \qquad (2.52)$$

$$\tau_o = \exp\left[0.0365 (m u_o)^{-0.2864}\right] \qquad (2.53)$$

$$u_o = 0.44 - 0.16 \left\{ [(LAT - 80)/60]^2 + [(d - 120)/(263 - LAT)]^2 \right\}^{0.5} \qquad (2.54)$$

for Julian day number, DN<300, $d = DN$ (2.55)

else, $d = DN - 366$ (2.56)

$$\tau_w = \min[1.0, 0.0909 - 0.036 \ln(mw)] \qquad (2.57)$$

$$\tau_a = \exp\left\{ -m\beta \left[0.6777 = 0.1464 m\beta - 0.00626 (m\beta)^2\right]^{1.3} \right\} \qquad (2.58)$$

Thus, $I_B = I_E \tau_B$ and $I_D = I_E \tau_D$ (2.59, 2.60)

Note that β is the angstrom turbidity coefficient.

2.10.4 Gueymard's REST2 model

Gueymard (2003, 2003, 2004, 2004) has proposed the Reference Evaluation of Solar Transmittance Model, REST model. The model calculates the beam normal irradiance based on attenuating the extraterrestrial irradiance with the Rayleigh-, ozone-, uniformly mixed gases-, water vapour-, aerosol- and NO_2 atmospheric transmittances. Thus the beam horizontal irradiance is given by:

$$I_B = I_E \sin(\alpha)\tau_r \tau_g \tau_o \tau_w \tau_a \tau_{NO2} \tag{2.61}$$

The transmittances are dependant on the solar geometry, the vertical ozone column amount (atm-cm), the precipitable water in vertical column, the site atmospheric pressure, air mass and β the angstrom turbidity coefficient.

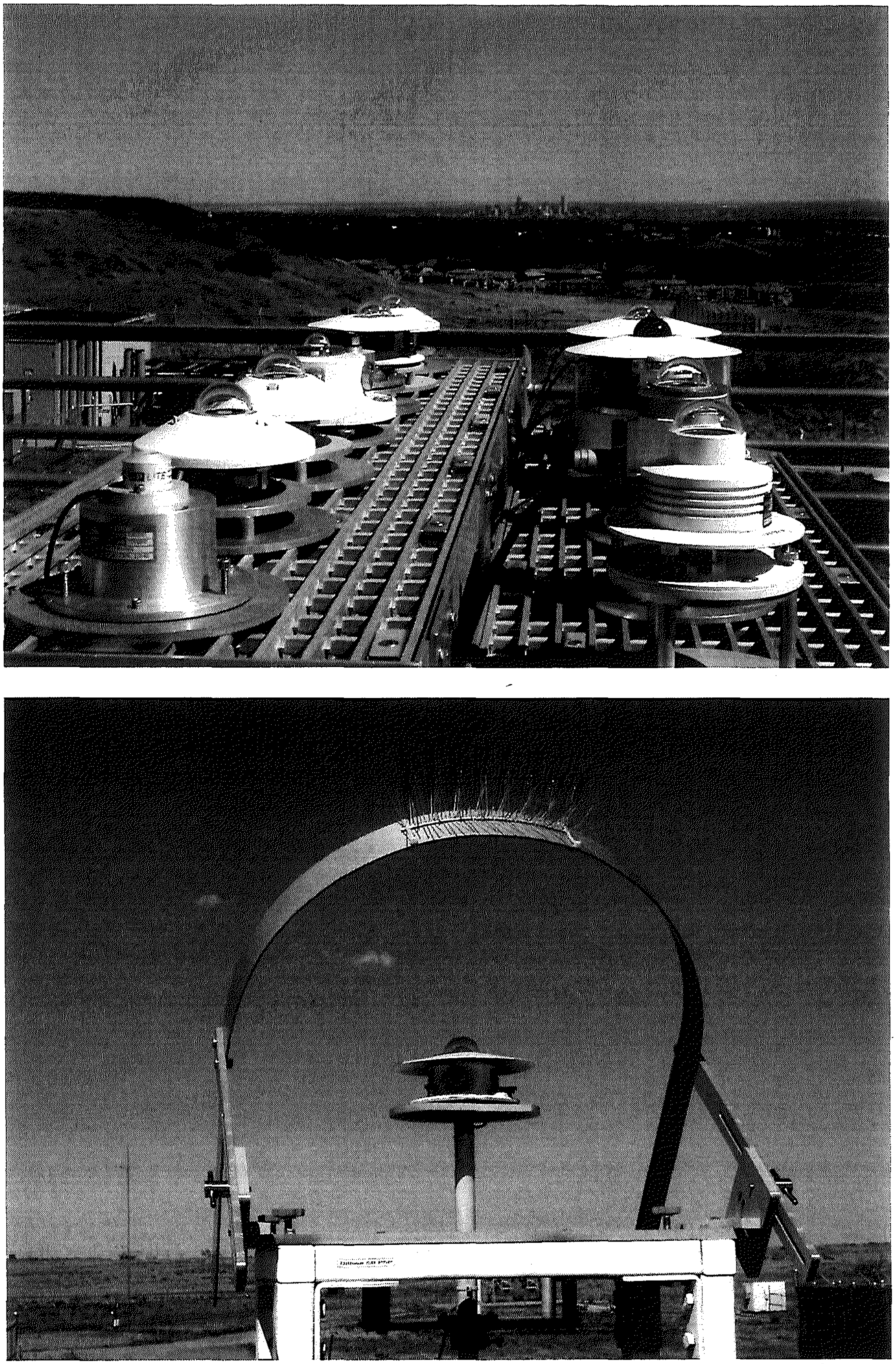


Figure 2.1 a. A battery of various Kipp & Zonen pyranometers. b. Kipp & Zonen pyranometer with shade-ring attached. (Courtesy of NREL, www.nrel.gov)

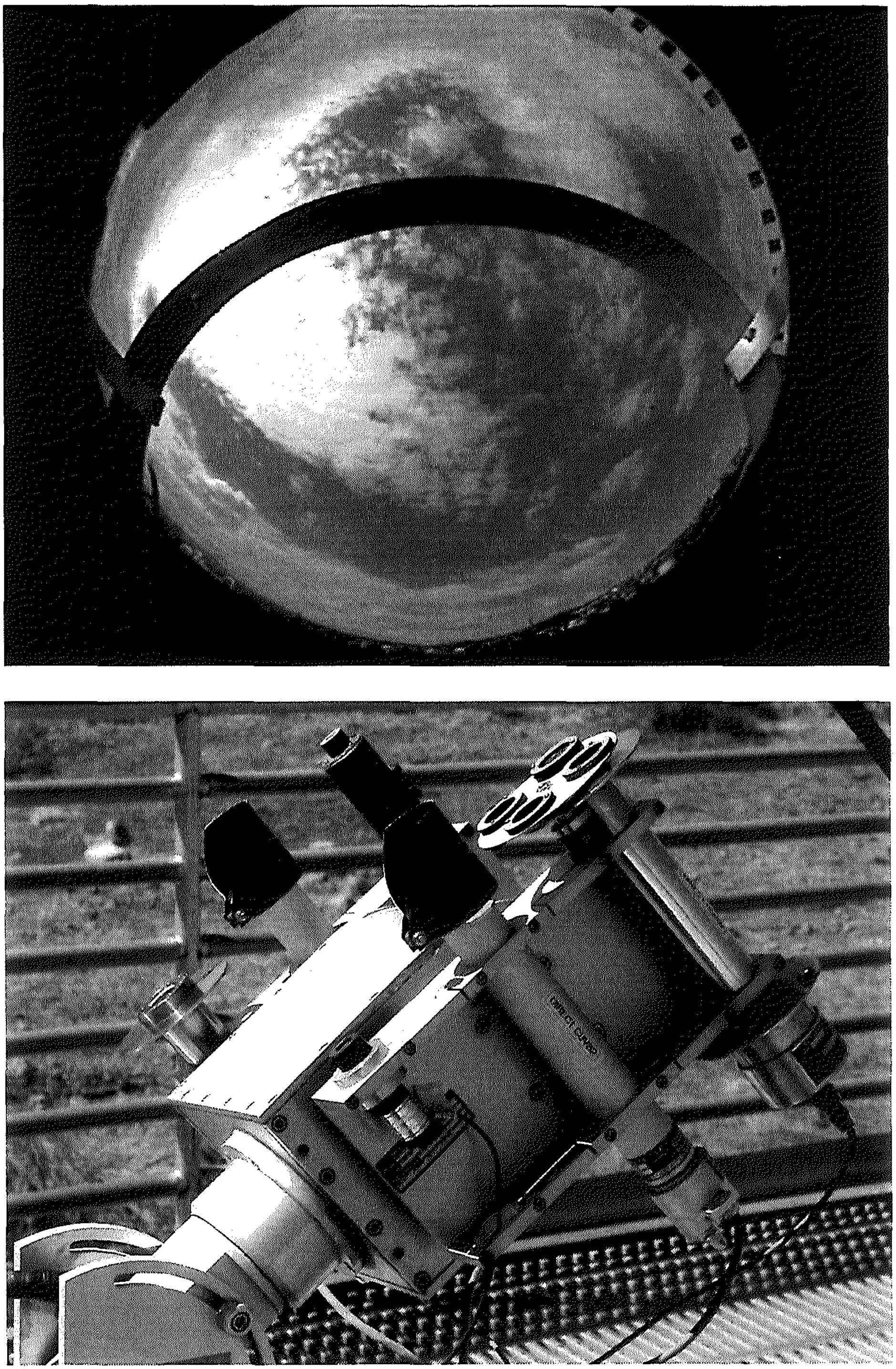


Figure 2.1 c. Kipp & Zonen pyranometer with shade-ring attached, inside-out view. d. Kipp & Zonen pyrheliometers with different lens filters attached. (Courtesy of NREL, www.nrel.gov)

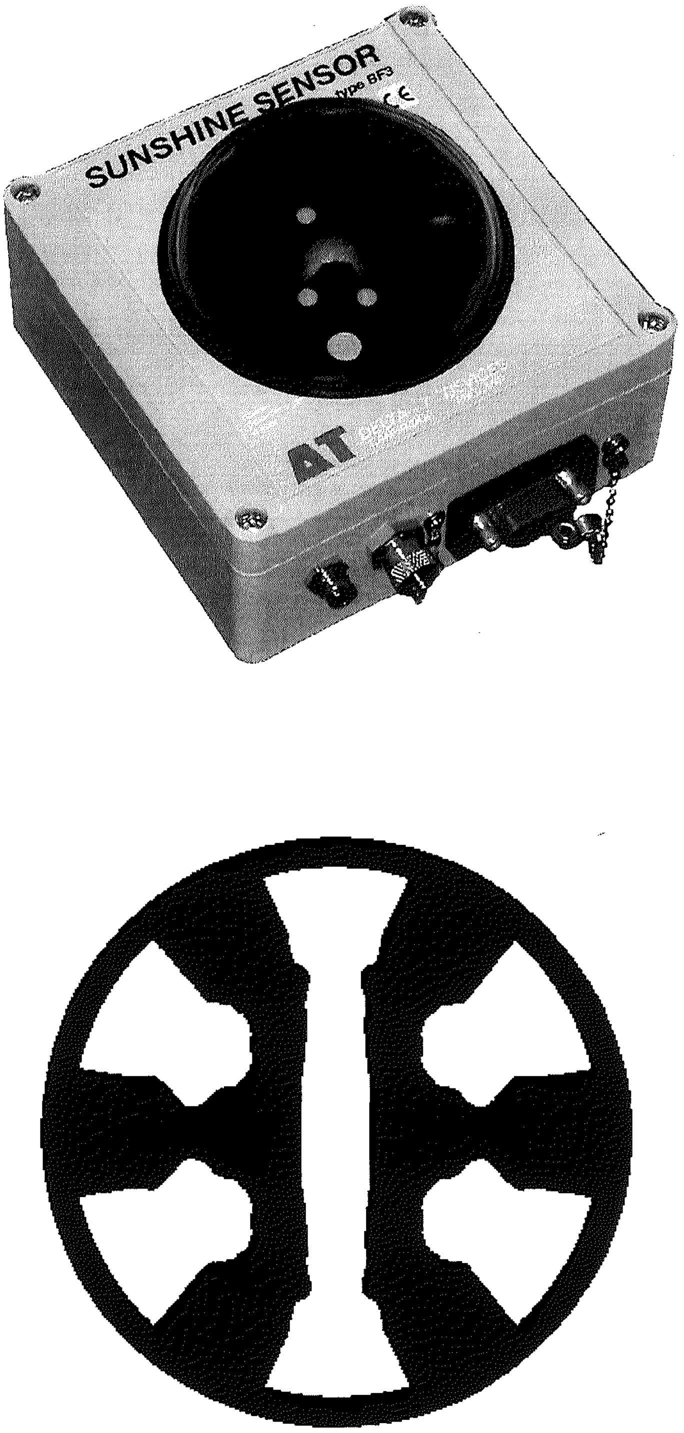


Figure 2.1 e. Delta T instrument. f. Lens details of the Delta T instrument.

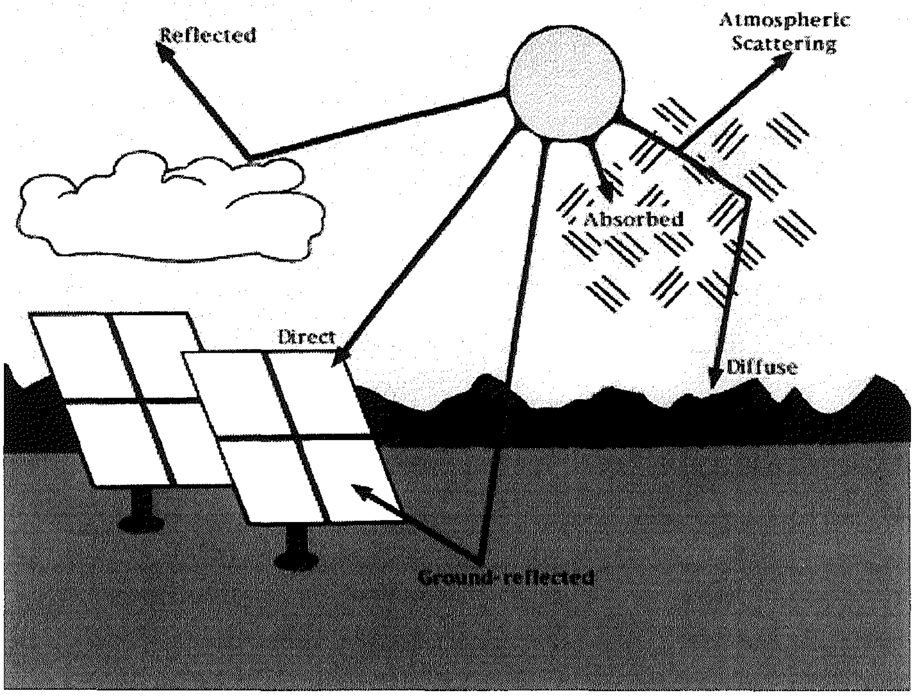


Figure 2.2 Solar irradiance components. For horizontal surfaces, global irradiance is the sum of diffuse and direct components. For tilted surfaces, ground-reflected irradiance is added. (Courtesy of NREL, www.nrel.gov)

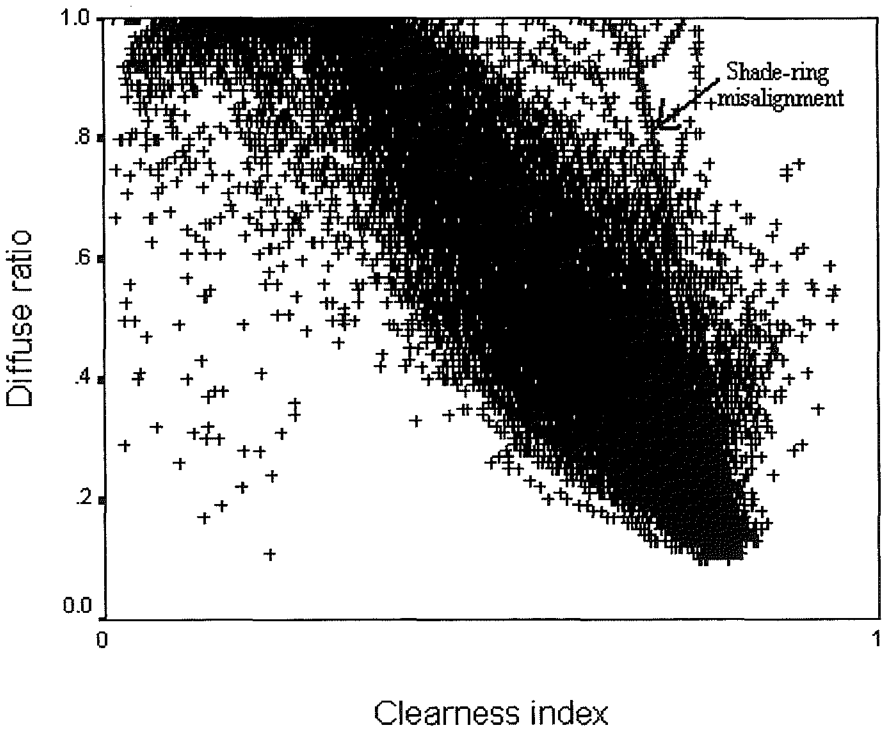


Figure 2.3 Diffuse ratio – clearness index plot for Bahrain, five-minute averaged data for 28 March - 30 September 2000

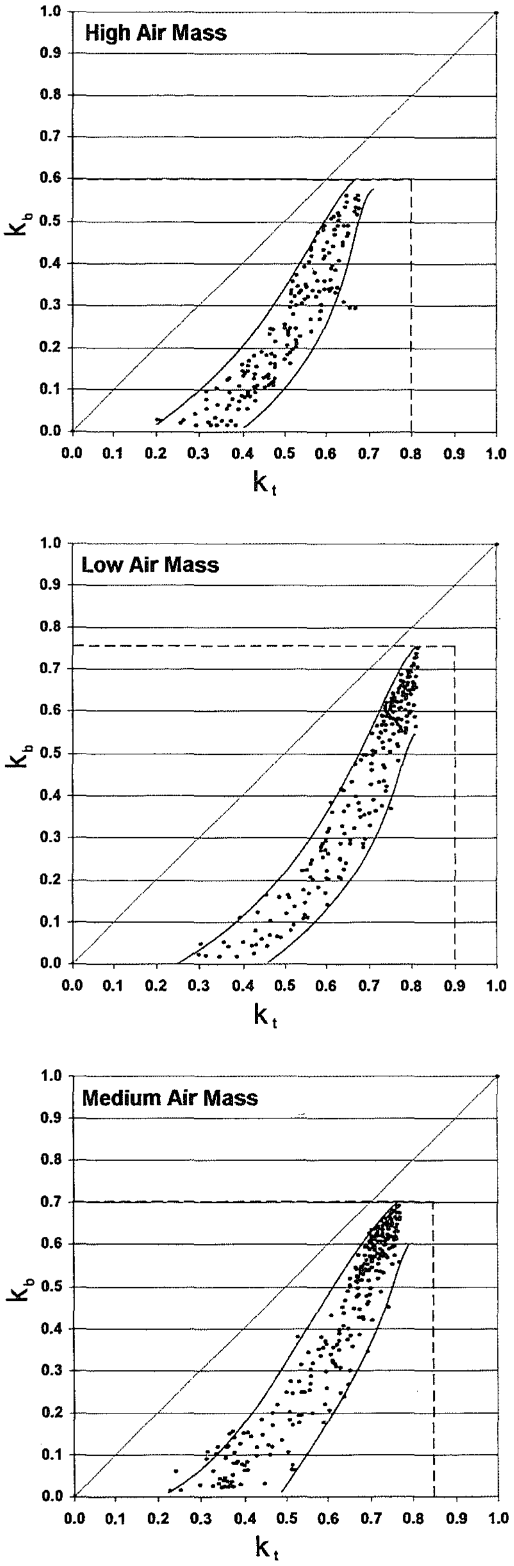


Figure 2.4 Hourly beam-to-extraterrestrial irradiance plotted against clearness index (NREL's quality control procedure).

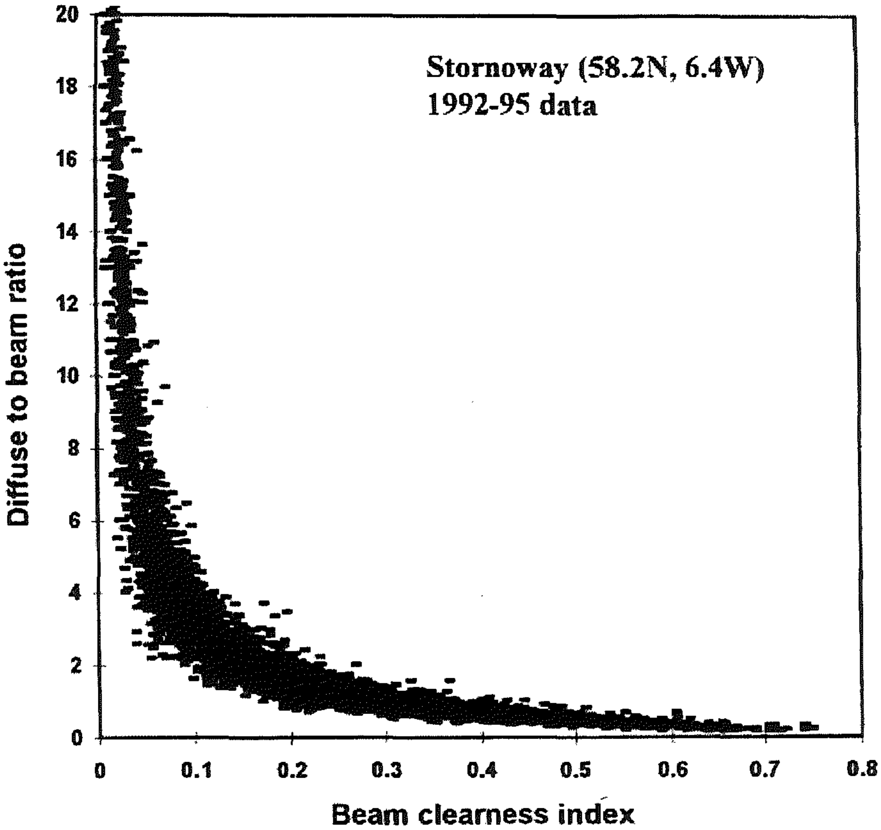


Figure 2.5 Scatter plot of Diffuse to Beam Ratio vs. Beam Clearness Index for the Stornoway site in the U.K.

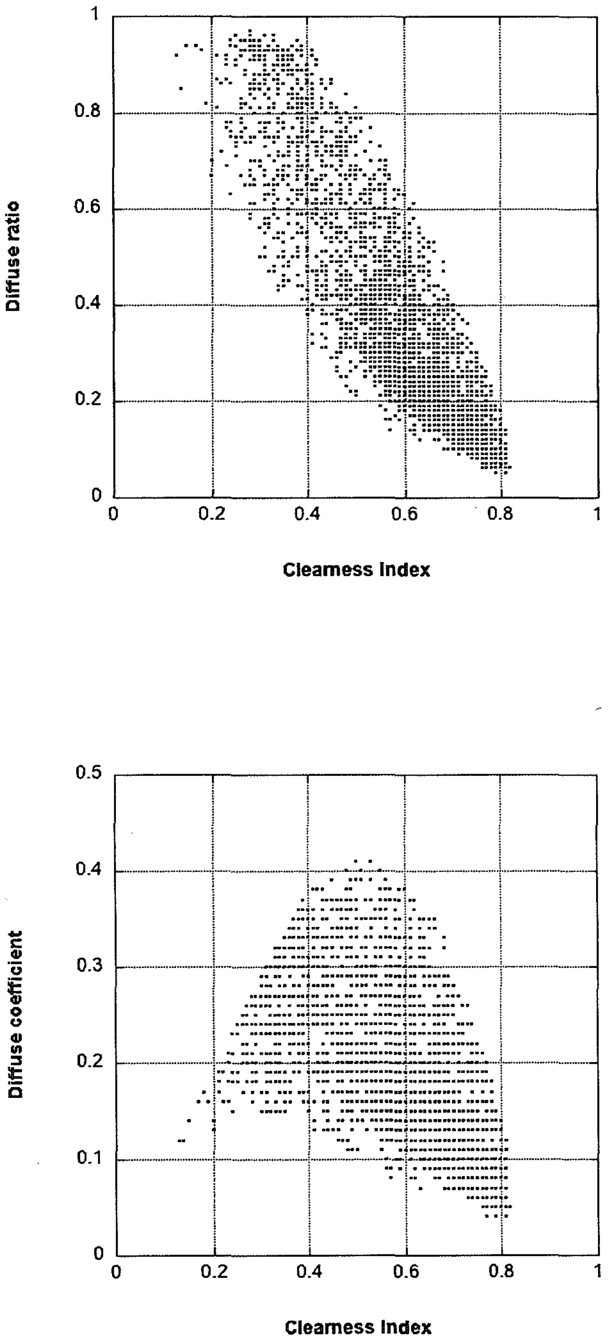


Figure 2.6 a. Clearness index - diffuse ratio scatter plot for Madrid. b. Clearness index - diffuse coefficient scatter plot for Madrid

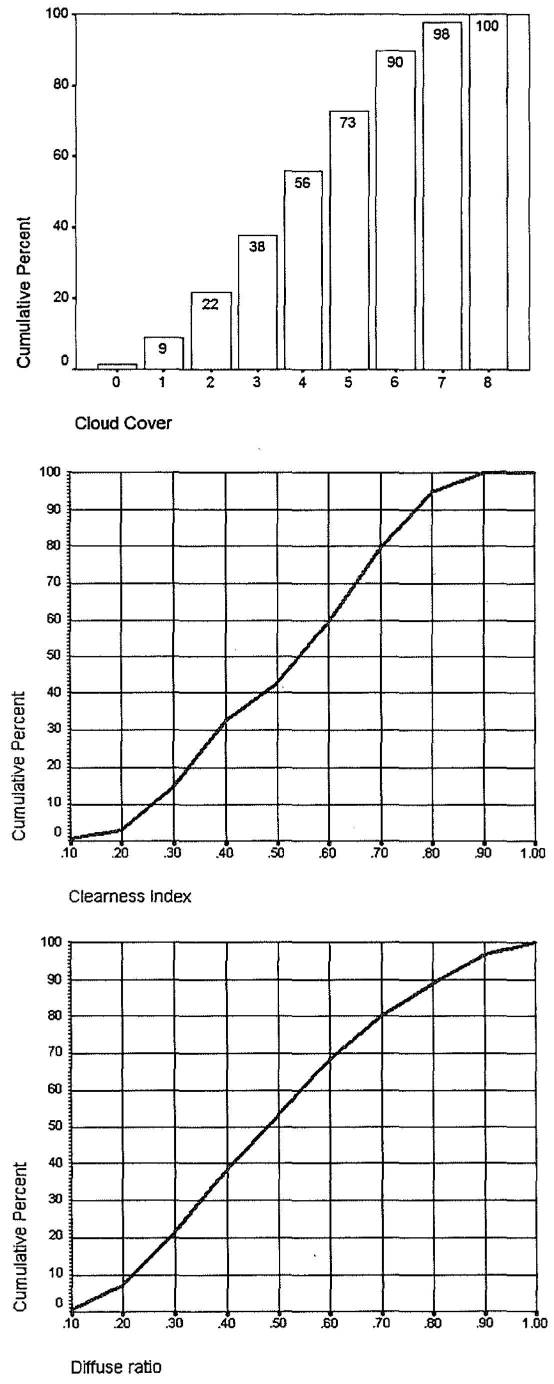


Figure 2.7 a. Cloud cover frequency distribution diagram for Chennai. b. Clearness index frequency distribution diagram for Chennai. c. Diffuse ratio frequency distribution diagram for Chennai.

3 QUALITY CONTROL OF SOLAR RADIATION DATASETS

3.1 Introduction

Architects, engineers and scientists involved in the field of solar energy systems require solar data measured at the vicinity of their application. It is important for active and passive uses of solar energy to know the global solar energy that is available and its diffuse component. The quality of measurements in solar irradiance dataset needs to be assessed by staff with minimal training and with precision.

The patterns of the availability of solar resource, in time, are important, as this dictates the design of energy storage systems. Long-term solar irradiance measurements are available from a number of meteorological measuring stations around the globe. Close examination of the data regularly reveals problems with the data often for extended periods of time. This lack of complete datasets, free from any erroneous measurements, can be due to a combination of factors, already explained in Section 2.4.

Datasets covering the northern hemisphere from Europe and Asia covering the last two decades were available to produce a quality control procedure using statistical and physical based tests.

The dataset used for this work consisted of 11 locations from the northern hemisphere and cover two continents. Out of the eleven sites, four were from Europe, two British sites: Bracknell (51.26N; 0.46W) and Eskdalemuir (55.32N; 3.20W) and two Spanish sites: Madrid (40.45N; 3.73W) and Gerona (41.97N; 2.88E). These are the most westerly locations used. We then cover Bahrain in the Middle-East (26.22N; 50.65E) before looking at the Indian sub-continent where we utilized data from four sites that are geographically and topographically different. The Indian sites are as follows: Chennai (13.0N; 80.18E), Pune (18.53N; 73.85E), Mumbai (19.12N; 72.85E) and New Delhi (28.60N; 77.20E). Finally the most easterly sites we covered are from Japan: Fukuoka (33.52N; 130.48E) and Sapporo (43.05N; 141.33E).

These sites not only cover different longitudes and latitudes from the northern hemisphere but also different climates and topographies, some have semi-arid climate others are temperate. The locations also differ by specific climatologies, Bahrain site is affected with seasonal sand storms, while Indian sites are affected by seasonal monsoons.

Note that the sites mentioned above were utilised to develop the quality control procedure, however for Chapter 5, four more British datasets were added, and quality controlled. These sites are as follows: Aldergrove (54.65N; 6.25W), Camborne (50.22N; 5.32W), Edinburgh (55.95N; 3.35W) and London (51.23N; 0.46W).

Traditionally, the solar irradiation datasets would be quality controlled by using various physical tests. In addition, some researchers have started investigating methods of quality control using statistical procedures. Certain major sources of errors will be exposed and a new hybrid physical and statistical based quality control procedure will be developed. These resultant error-free datasets will be used in following chapters as basis for model development and validation.

3.2 Outliers in solar radiation databases

Any likely sources of errors or problems related to solar radiation measurement may be categorised under the following two major categories: equipment error and uncertainty and operation related problems and errors.

With any measurement there exist errors, some of which are systematic and others inherent of the equipment employed. The most common sources of error arise from the sensors and their construction. These are broken down into the most general types of errors and described below: Cosine response, azimuth response and temperature response. Spectral selectivity, stability and non-linearity are also major causes of errors in addition to shade-ring misalignment and dark offset (nocturnal) long-wave radiation error.

Of all the above listed errors the cosine effect is the most apparent and widely recognised. This is the sensor's response to the angle at which radiation strikes the sensing area. The more acute the angle of the sun, i.e. at sunrise and sunset, the greater the error will be (at altitude angles of sun below 6 degrees). Cosine error is typically dealt with through the exclusion of the recorded data at sunrise and sunset times.

The azimuth error is a result of imperfections of the glass domes and in the case of solarimeters the angular reflection properties of the black paint. This is an inherent manufacturing error, which yields a similar percentage error as the cosine effect.

Like the azimuth error the temperature response of the sensor is an individual fault for each cell. The photometers are thermostatically controlled hence the percentage error due to fluctuations in the sensor's temperature is reduced. The pyranometers rely on their construct, i.e. a double-glass envelope to prevent large temperature swings.

The spectral selectivity of the pyranometers is dependent on the spectral absorbance of the black paint and the spectral transmission of the glass. The overall effect contributes only a small percentage error to the measurements. Each sensor possesses a high level of stability with the deterioration of the cells resulting in approximately ±1% change in the full-scale measurement per year. Finally the non-linearity of the sensors is a concern especially with photometers. It is a function of illuminance or irradiance levels. It however tends to contribute only a small percentage error towards the measured values.

The work undertaken by Stoffel et al. (2000) at NREL under the US continental climate and a desert site in Saudi Arabia has shown that zero offsets of -5 to -20 W/m^2 occur in diffuse pyranometer measurements due to thermal imbalances. This error was reported for all instruments using black sensors. The offset for a black and white detector, however, was found to be insensitive to such offset errors.

In addition to the above sources of equipment-related errors, care must be taken to avoid the following operational errors: Operation related problems and errors; Complete or partial shade-ring misalignment; Dust, snow, dew, water-droplets, bird droppings etc; Incorrect sensor levelling; Shading caused by building structures; Electric fields in the vicinity of cables; Mechanical loading of cables; Orientation and/or improper screening of the vertical sensors from ground-reflected radiation; Station shut-down.

The sources of operation relation errors itemised above are self-explanatory. It is good practise to protect cables from strong electric fields such as elevator shafts. Another source of error that may arise is from cables under mechanical load (piezoelectric effects). The piezoelectric effect is the production of electrical polarization in a material by the application of mechanical stress. Failure to protect cables from the above sources may produce ‘spikes’ in the data and these are shown as unusually high values of irradiance. Such errors are best highlighted via cross plotting diffuse ratio ($k = {}^{I_D}\!/_{I_G}$) against clearness index ($k_t = {}^{I_G}\!/_{I_E}$), and a sample plot was shown in Figure 2.3. Note that any consistent errors emanating from an operational problem, such as misaligned shade ring are easily picked up by this type of plot. Any serious departure of data from the normally expected envelope is thus identified.

Stoffel et al. (2000) give us a good representation of the scale of errors for carefully managed irradiance sensors. In fact they found that the range of error for a pyranometer compared with a reference pyranometer was from +2.5% to -10%; while for a pyrheliometer the range was ±2.5%.

3.3 *Physical tests*

The most similar quality control procedure is the one developed by Muneer and Fairooz (2002) quality control method. This quality control procedure is in addition to other filters a combination of tests based on the CIE quality control procedure (Section 2.7.5) and Page irradiance model (Section 2.7.1). A similar series of physical tests were adopted and are explained below in further details.

The code for the presently proposed procedure was written in FORTRAN to process the databases available. This code is available in Appendix C. Geographical information for the site is required such as site elevation above sea level, latitude, longitude and local time meridian. Also, logging related information is required, either solar time or local civil time is accepted.

Operation related information is also required such as desired standard deviation, interval number and envelope cut off point. Details regarding this type of information are provided later.

Before starting to test the data for its validity, solar position calculations are performed for each data entry. These are described by Muneer (2004) and are listed below as per order of occurrence. Calculation of solar hour angle, apparent solar time, declination angle, solar altitude and finally the calculation of solar azimuth is performed by the code.

Overall there are four series of tests for this model, and the first three are the physical tests.

3.3.1 First Quality Control test

The calculation then proceeds by eliminating entries that show a solar altitude less than 7°. This test is far more restrictive than what is usually used by peers, who set this limit at 10°. Researchers often agree that at low solar altitudes, i.e. at sunrise and sunset, the atmosphere scatters quite a great amount of solar irradiance, which results in erroneous readings of the solarimeters. To ensure absolute error free readings due to sunset and sunrise, any solar irradiance data that is under this limit is discarded.

For entries that have passed the first test, the day number, the horizontal extraterrestrial irradiation and finally the clearness index (k_t) and the diffuse ratio (k) are calculated.

3.3.2 Second Quality Control test

This test is a logical test, as the clearness index and the diffuse ratio are always positive and have values between zero and one. Any data that is found above this limit is obviously erroneous since global horizontal irradiance cannot exceed the extraterrestrial irradiance, and in a similar fashion, diffuse horizontal irradiance cannot exceed global horizontal irradiance.

$$0< k_t <1 \text{ and } 0< k <1 \qquad (3.1)$$

3.3.3 Third Quality Control test

At this stage global and diffuse irradiation are compared with their corresponding Page-model upper and lower boundaries (Section 2.6.1). The global horizontal irradiation aught to be less than or equal to the clear day global horizontal irradiation. Thus,

$$I_G \leq I_{G,C} \qquad (3.2)$$

With $I_{G,C} = I_{B,C} + I_{D,C}$ (3.3)

Note that $I_{B,C}$ and $I_{D,C}$ can be calculated using Eqs.2.6-2.9.

From Muneer and Fairooz (2002) quality control procedure, it is proposed to test if the diffuse horizontal irradiation is sandwiched between the clear day diffuse and the overcast day horizontal irradiation as defined by Page (Greif and Scharmer (2000) and the CIBSE Guide on weather and solar data (1997)). Thus,

$$I_{D,OC} \leq I_D \leq I_{D,C} \qquad (3.4)$$

With the global ($I_{G,OC}$) and diffuse ($I_{D,OC}$) irradiances are equal due to the absence of the beam component. Thus we refer to Eq.2.10.

The flow chart for the above three tests are presented in Fig.3.1. This is how the FORTRAN programme operates. Note that three files are used and created during the process; the first file contains the raw datasets that are the basis of the whole procedure. This contains everything that is needed for the future chapter, the time stamp, the irradiance data and the synoptic data.
The data that did not pass the physical tests are stored.

3.4 Statistical test

Hand drawn envelopes developed from visual inspection of the datasets such as those presented in Fig3.2 were used by Muneer and Fairooz (2002) to clean datasets for clearness index-diffuse ratio scatter plots. The visual type of boundary that is hand drawn is herein referred to as envelope of expectancy, or the quality envelope.

The coordinates defining the envelope are then noted thus producing constraints for data filtering, i.e. the upper and lower bounds of acceptability of k for any given k_t are used.

Another technique that is employed by statisticians to identify erroneous data is the outlier analysis. Note that an 'outlier' is a term that indicates an abnormality, and suggests that the datum is not typical of the rest of the data. As a rule, an outlier should be subjected to careful examination to identify logical explanations for its unusual behavior. Outliers may, however, be rejected if the associated errors can be traced to erroneous observations, due to any one or a number of factors described above. Statistically, a 'near-outlier' is an observation that lies outside 1.5 times the inter-quartile range. The inter-quartile is the interval from the 1st quartile to the 3rd quartile. The near-outlier limits are mathematically defined by:

Lower outlier limit=1st quartile-1.5 (3rd quartile-1st quartile) (3.5)

Upper outlier limit=3rd quartile+1.5 (3rd quartile-1st quartile) (3.6)

A high number of outliers in the dataset signify that the observations have a high degree of variability or a large set of suspect data indicating poor station operation. Note that 'far outliers' are those for which the factor of 1.5 used within Eqs.3.5 and 3.6 is replaced with a value of 3.

Outlier analysis is however very computation intensive as it involves ordering of large datasets in an ascending order. The presently proposed technique based on standard deviation is much more economical in terms of CPU time. The software that was developed under the present research programme, however, produces both, the near-outlier and standard deviation based envelopes. These are included in Fig.3.2. Note that the 'crude' hand-drawn envelope based purely on a visual observation is shown as greyed area.

The presently proposed construction of k-k_t quality control envelope is a statistical procedure that requires estimation of k_t banded mean, weighted mean, ($\bar{k}$) and standard deviations of 'k' values (σ_k). Typically the k_t range of data may be divided in, say, 10 bands of equal width. For each band the above-mentioned statistics is obtained. Other band sizes were tested, 5, 10, 20, 30 data bands, and there was no differences observed in the results. From this information an envelope may be drawn that connects those points that respectively represent the top ($\bar{k}+2\sigma_k$) and bottom ($\bar{k}-2\sigma_k$) curves.

It can be found that the standard deviation at $\pm 2\sigma_k$ envelope is more restrictive in the mid range k-k_t zone compared to a quartile envelope at the exception of the extreme zones; i.e. low k_t and high k_t.

For the datasets used, it was found that the $\bar{k}\pm 2\sigma_k$ provides the most optimum envelope. Note that $\bar{k}$ is the weighted mean of k values within a given k_t band. It also outperforms the quartile analysis method in 90% of cases, i.e. in terms of the restrictiveness of data inclusion. There was a need to use weighted means as to reduce the effect of the outliers on the banded data, and thus the envelope. The weights that are in inverse proportion to the deviation of any given datum from the mean of the population are used.

Once the envelope constituted by the upper and lower boundaries is identified it is possible to fit a polynomial for a mathematical description of the envelope of acceptance. A second degree polynomial was found to be adequate. Thus, the upper and lower boundaries are respectively represented as,

$$A(k_t) = Max(1, a_1 k_t^2 + b_1 k_t + c_1) \qquad (3.7)$$

$$B(k_t) = \min(0, a_2 k_t^2 + b_2 k_t + c_2) \qquad (3.8)$$

Note that any given polynomial may generate data that can go beyond the physical limits of k, which lie between 0 and 1. The formulation given in Eqs.3.7 and 3.8 satisfy the above constraints. Furthermore, due care has to be taken to incorporate the 'shoulder' effect caused by the respective intersection of the upper and lower polynomials with the k=1 (upper) and k=0 (lower) limits for the plot. By visual inspection of the envelope scatter plot it is possible to ascertain the latter points of intersection. The procedure of quality control can now be completed with the envelope of acceptance fully defined.

Note that presently proposed quality control procedure is therefore semi-automatic, as the user has to select the cut-off points by visual inspection of the trend of the upper and lower-bound polynomials. This is the only visual part of the process. For the locations quality controlled, the cut-off point has been below k_t =0.4 (upper bound), and between k_t =0.85, and k_t=1 (lower bound).

Figure 3.3 presents the latterly mentioned upper, C (k_t) and lower, D (k_t) lines of intersection in addition to the A(k_t) and B(k_t) sections of the envelope of acceptance. The last item to be mentioned in this context is that in certain cases there may be a need for the control of the lower-bound polynomial with respect to its upper limit. Notice that within Fig.3.3 an unconstrained 'flow' of the B (k_t) curve would exclude a small proportion of data belonging to heavy overcast regime. A cut-off shown as E (k_t) line is thus required, once again by visual inspection. There is a need for this cut-off as there could be valid data that can be tagged as outliers and removed by the envelope test as can be deduced from Fig.3.2.

This procedure was applied to the present datasets. Sample plots of raw, then filtered data are presented in Figs.3.4 and 3.5 for Fukuoka and Gerona respectively. Both envelopes are a second degree polynomial of $\pm 2\sigma_k$. Note that 2.5 and 5.4% of the raw data was flagged as outliers by the statistical test for Fukuoka and Mumbai, respectively.

Attention needs to be drawn towards Fig.3.5. The plot reveals that the shade ring correction factor has not been applied to diffuse irradiance measurements as is evident via examination of the left-hand corner of the plot. Note that the data in the corner would be expected to attain the limiting value of k=1 as $k_t \rightarrow 0$. This was later confirmed by the provider of the dataset.

Figure 3.6 also reveals that problems exist with respect to shade ring correction. Figure 3.6 also demonstrates that the optimum envelope of acceptance for this particular location is $\bar{k} \pm 2.3\sigma_k$. This is evident via a comparison of Figs.3.6c and d. Note that the former plot rejects an excessive amount of data using a tighter ($\bar{k} \pm 2\sigma_k$) envelope. Initially based on the work by Claywell (2003) and his conclusions that $\bar{k} \pm 2\sigma_k$ is the optimal envelope, all the datasets were quality controlled on the above mentioned envelope. It was found however, that for some datasets this envelope can either be too restrictive or too lenient. Following this observation, the datasets were then individually quality controlled starting from an envelope of $\bar{k} \pm 1.8\sigma_k$ and increasing by intervals of $\pm 0.1\sigma_k$ up to $\bar{k} \pm 2.4\sigma_k$. It was found that the envelope of acceptance optima lay between the ($\bar{k} \pm 2\sigma_k$) and ($\bar{k} \pm 2.3\sigma_k$) limits. This adds more responsibility on the operator or the user of the software to asses the proper variables for the best operation of the quality control procedure.

In general, a second order polynomial was deemed to be a good envelope of acceptance. However there are cases where a 2nd order polynomial $\bar{k} \pm 2\sigma_k$ envelope was unacceptable. Such a case is for the Bombay database. Further examination of the envelope in Fig.3.7b clearly denotes that the envelope is not appropriate. A third degree polynomial envelope was hence drawn as shown in Figs.3.7c and d, in which it is possible to determine that the 3rd degree polynomial $\bar{k} \pm 2\sigma_k$ envelope is more appropriate than the previous envelope. This change of envelope was due to the fact that all possible 2nd degree polynomial envelopes were inaccurate as is shown in Figs.3.7e and f where even the $\bar{k} \pm 2.2\sigma_k$ 2nd order polynomial envelope was flagging some valid data in the top right as outliers.

The results of the quality control procedure are given in Table 3.1 for the 11 original sites (excluding the four later UK sites used in Chapter 5).

Note that the FORTRAN code in Appendix C does not contain the algorithm for the 3rd order polynomial envelope. This code is as follows:

```
DEFINE POLYNOMIALS
do i = 1,NITER
if (a(i).le. PlimUP) then
        polyUp(i)=1.0
        else
        polyUp(i)=xtc(1)+xtc(2)*a(i)+xtc(3)*a(i)*a(i)+xtc(4)*a(i)*a(i)*a(i)
endif
if (a(i).le. PlimLW. and. a(i).gt. cutoff) then
        PolyLW(i)=xlc(1)+xlc(2)*a(i)+xlc(3)*a(i)*a(i)+xlc(4)*a(i)*a(i)*a(i)
        else if (a(i).le. cutoff) then
        PolyLW(i)=polcut
        else
        PolyLW(i)=0.0
endif
```

Note that *PlimUp* and *PlimLW* are the intersection points between the 3rd order polynomial and the X and Y axis respectively.

3.5 Conclusions

It is acknowledged that solar radiation datasets often contain errors. These errors are minimal in 1st class meteorological stations, but become more serious in less well maintained stations. With any measurement there exist errors, some of which are systematic and others inherent of the equipment employed. The most common sources of error arise from the sensors and their construction. In addition to the equipment-related errors, care must be taken to avoid the following operational errors: Operation related problems and errors; Complete or partial shade-ring misalignment; Dust, snow, dew, water-droplets, bird droppings etc; Incorrect sensor levelling; Shading caused by building structures; Electric fields in the vicinity of cables; Mechanical loading of cables; Orientation and/or improper screening of the vertical sensors from ground-reflected radiation; Station shut-down.

There are two schools or lines of though for quality control of solar irradiance data. The first method is to verify the physical aspects of the data. However more complex models use certain statistical procedures either separately from the physical tests or simultaneously.

The proposed quality control procedure is a hybrid method, combining both physical and statistical tests. There are three physical tests overall. The first test excludes sunrise and sunset hours. The following test checks that global horizontal irradiance is less than the horizontal extraterrestrial irradiance, and the diffuse component is less than the global horizontal irradiance. The third physical test is a semi-empirical test in which the diffuse irradiance is sandwiched between the Page's calculated absolute clear-sky and heavily over-cast sky diffuse irradiance. Similarly the global horizontal irradiance is less than or equal to the absolute clear-sky global irradiance as calculated by Page.

The statistical test used is to create an envelope of expectancy, which contains the non-erroneous data. This envelope is based on standard deviations of a weighted mean of the data. To achieve this, the data is split into bands of the clearness index and for each band the weighted mean is calculated. The weighted mean was used rather than the mean because the weights would reduce the effect of extreme outliers on the bulk of the data. For each band the $\pm 2\sigma$ of the diffuse ratio is obtained. The envelope of expectancy is thus two polynomial curves linking the upper and then lower points of the standard deviation of the bands.

It was found that ±2.0σ_k was not always accurate and in some instances exclude valid data. It was found overall that 8 sites had an envelope of ±2.0σ_k, one of ±2.2σ_k and two of ±2.3σ_k. In all instances the envelope was a second order polynomial with the exception of one dataset. It was found for Mumbai that the 2nd order polynomial envelope to be very inaccurate, and thus was replaced by a third order polynomial envelope.

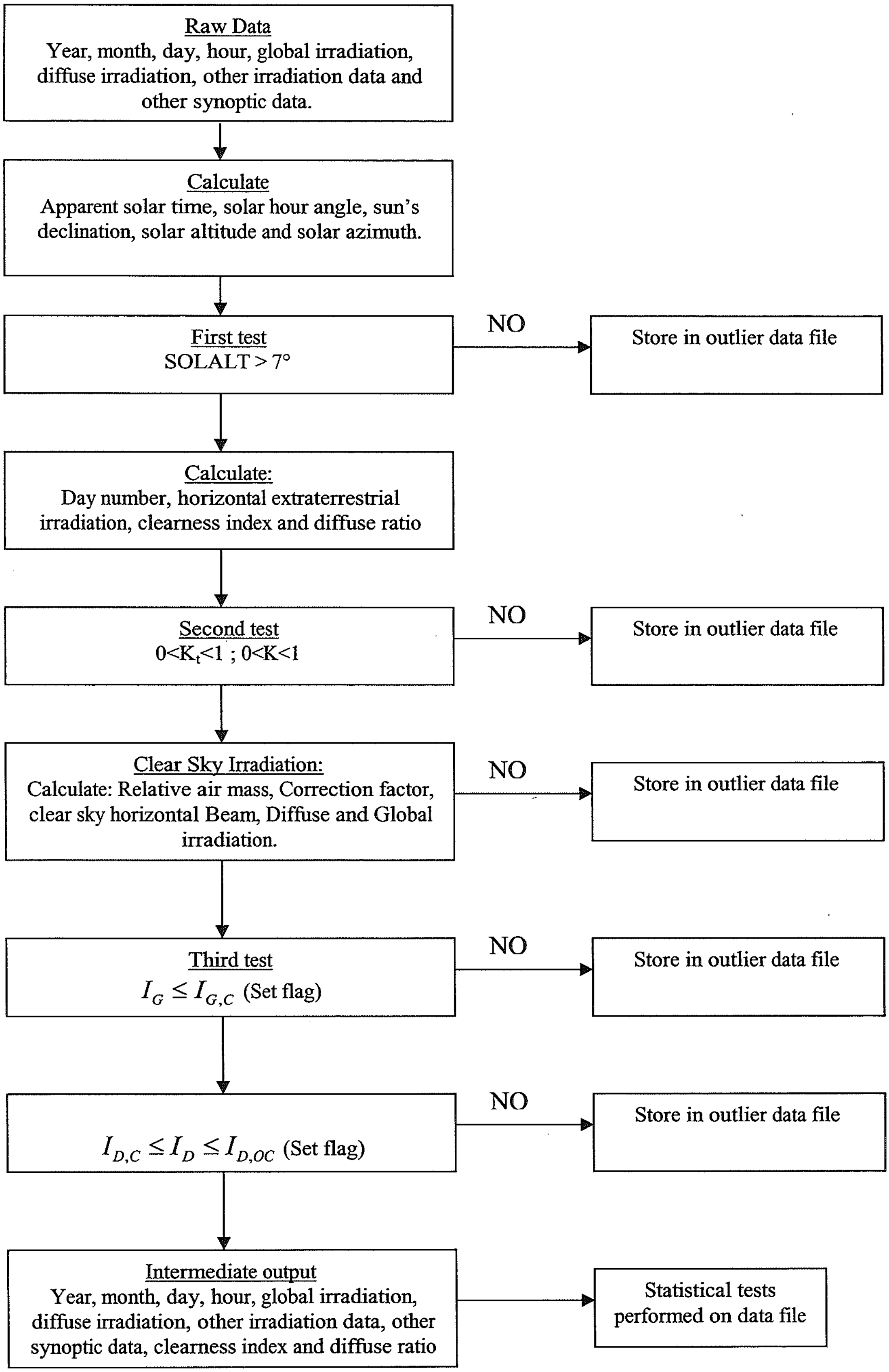


Figure 3.1 Flow chart for processing raw solar irradiation data via presently proposed procedure.

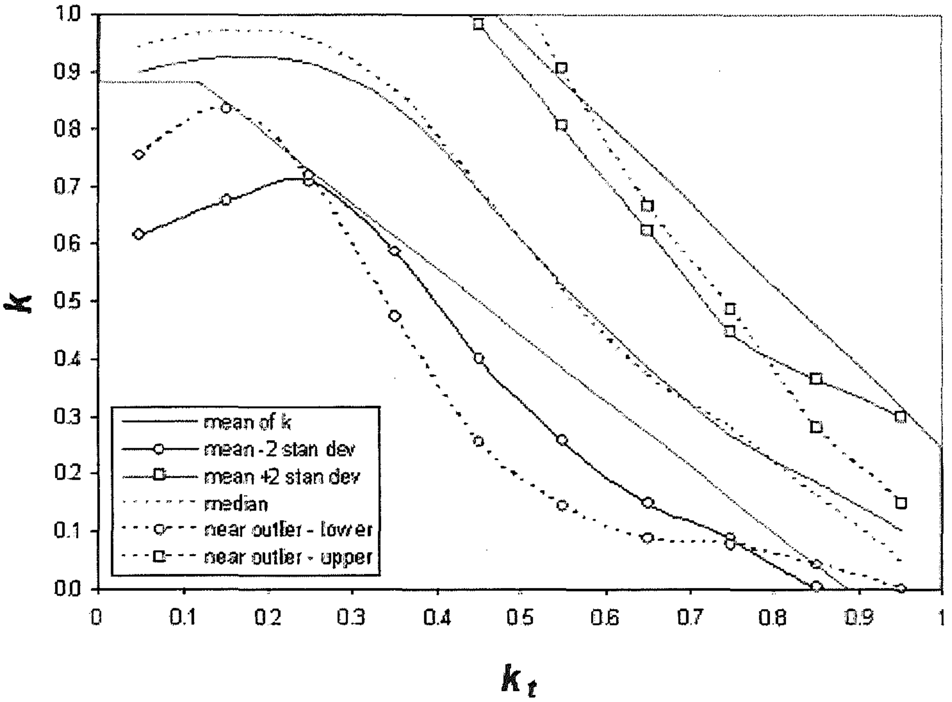


Figure 3.2 Comparison between standard deviation and quartile analysis.

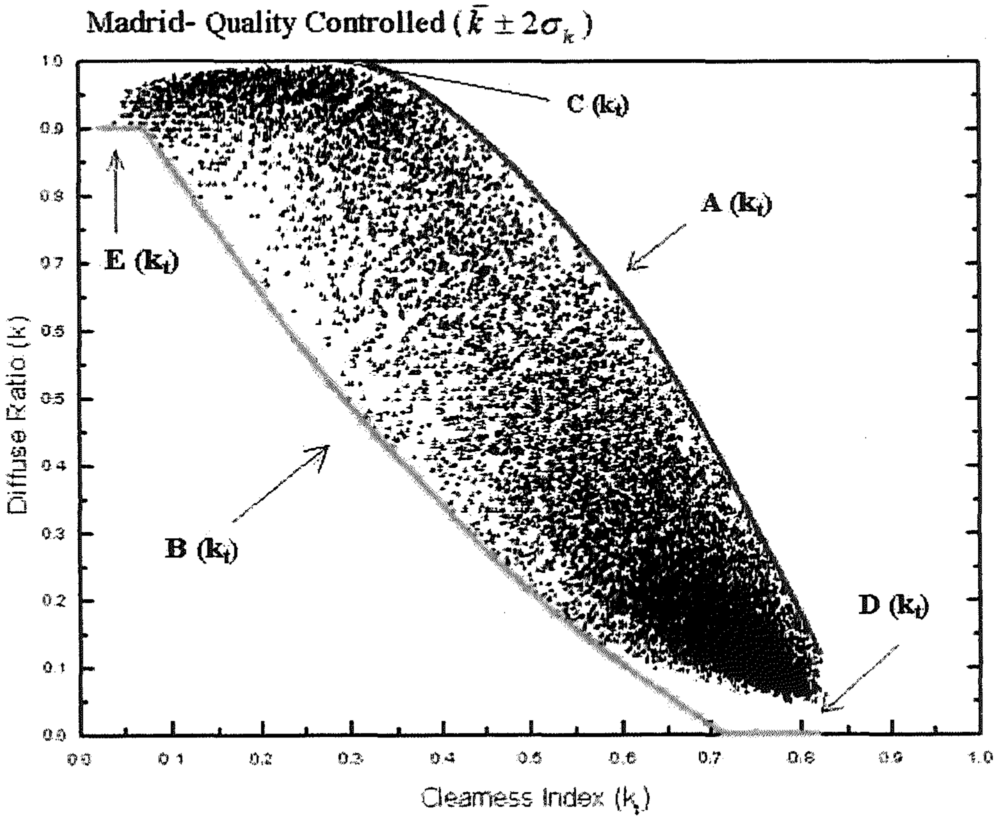


Figure 3.3 Picture shows the boundary equations and functions for a typical database analysis, in this case Madrid.

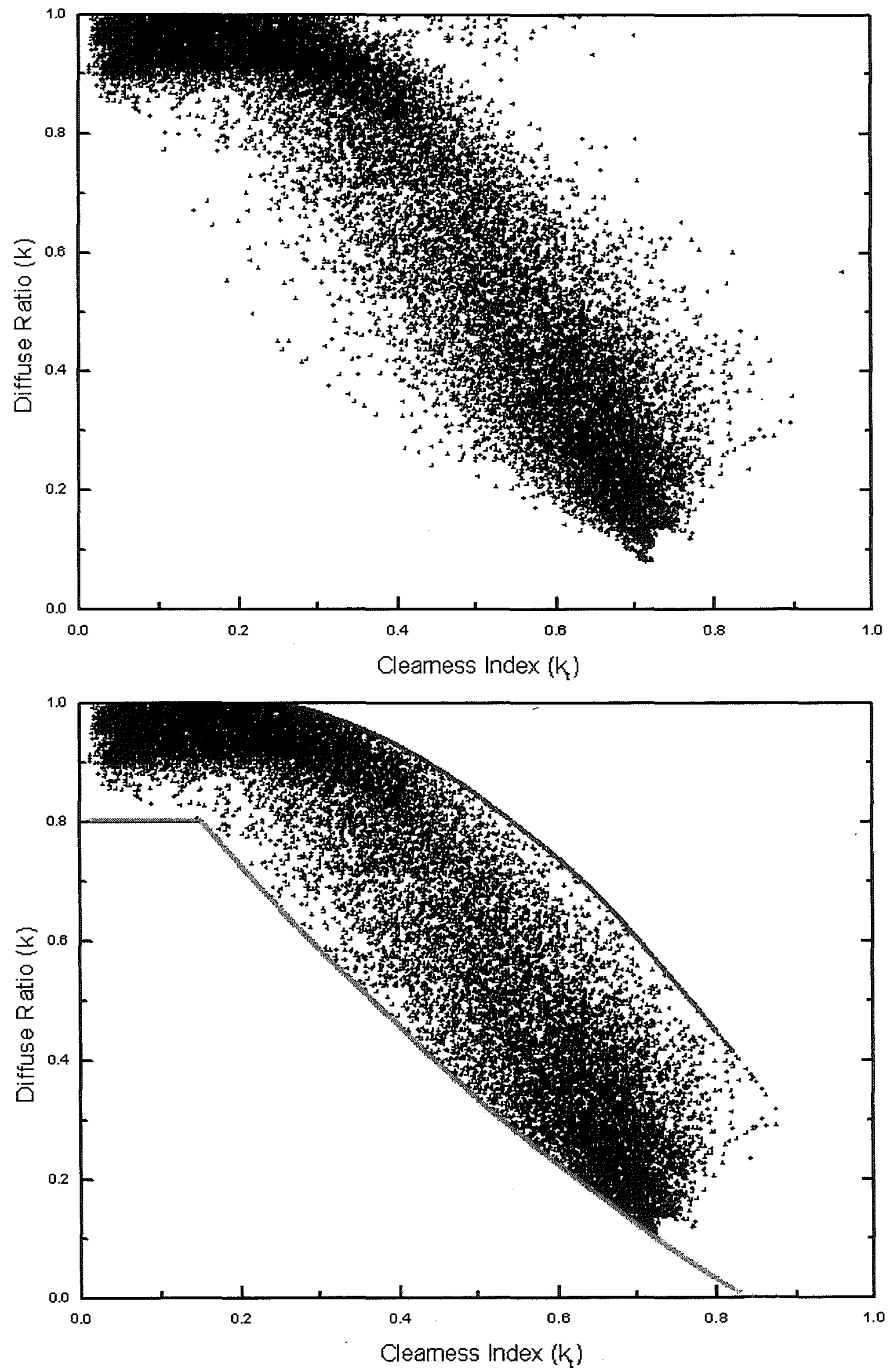


Figure 3.4 Scatter plot for Fukuoka. a. Raw data. b. Quality controlled ($\bar{k} \pm 2.0\sigma_k$). Note that $\bar{k} \pm 2.0\sigma_k$ envelope contains 97.5% of the data that passed the physical tests.

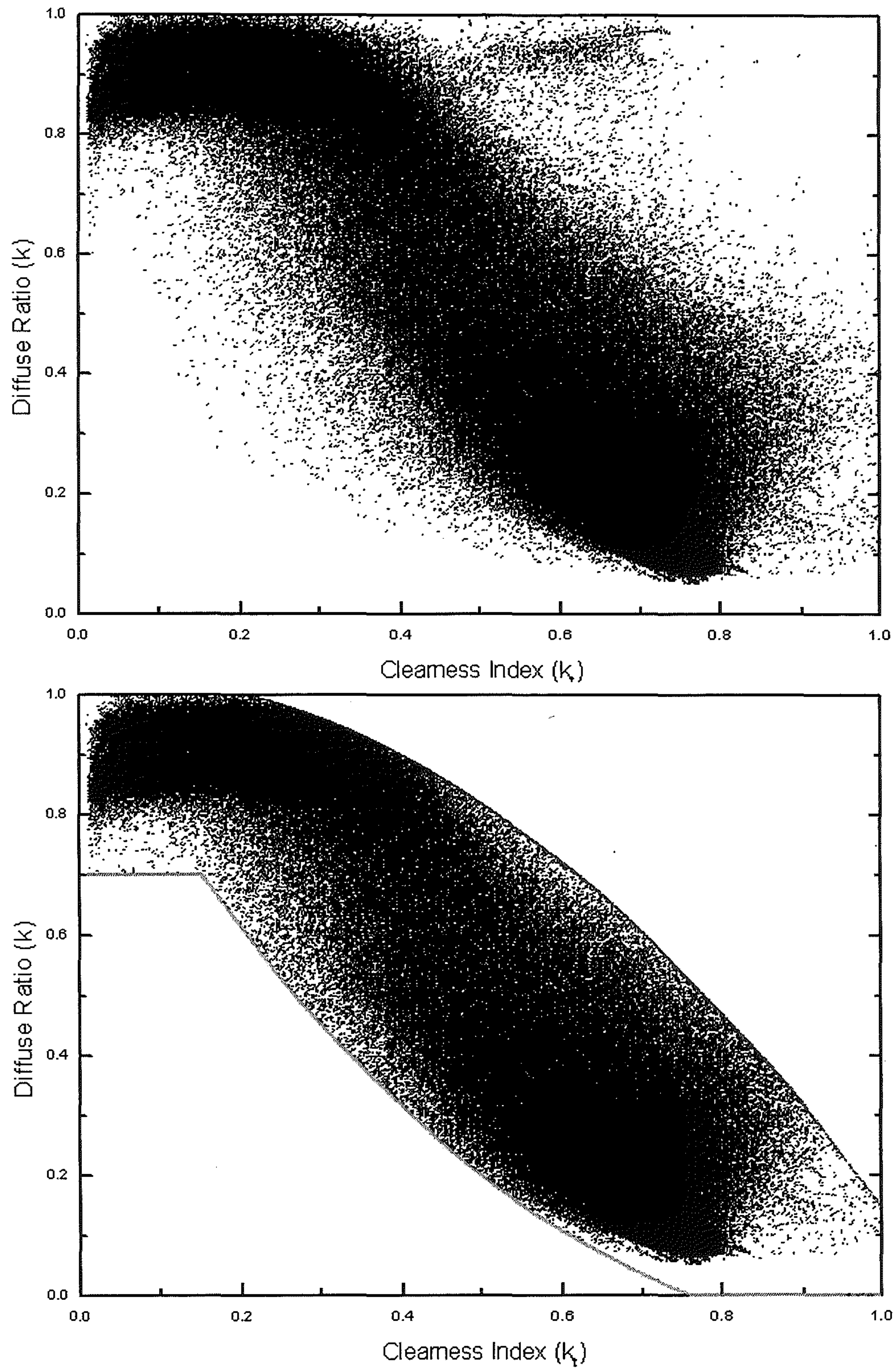


Figure 3.5 Scatter plot for Gerona. a. Raw data. b. Quality controlled ($\bar{k} \pm 2.0\sigma_k$). Note that $\bar{k} \pm 2.0\sigma_k$ envelope contains 94.62% of the data that passed the physical tests

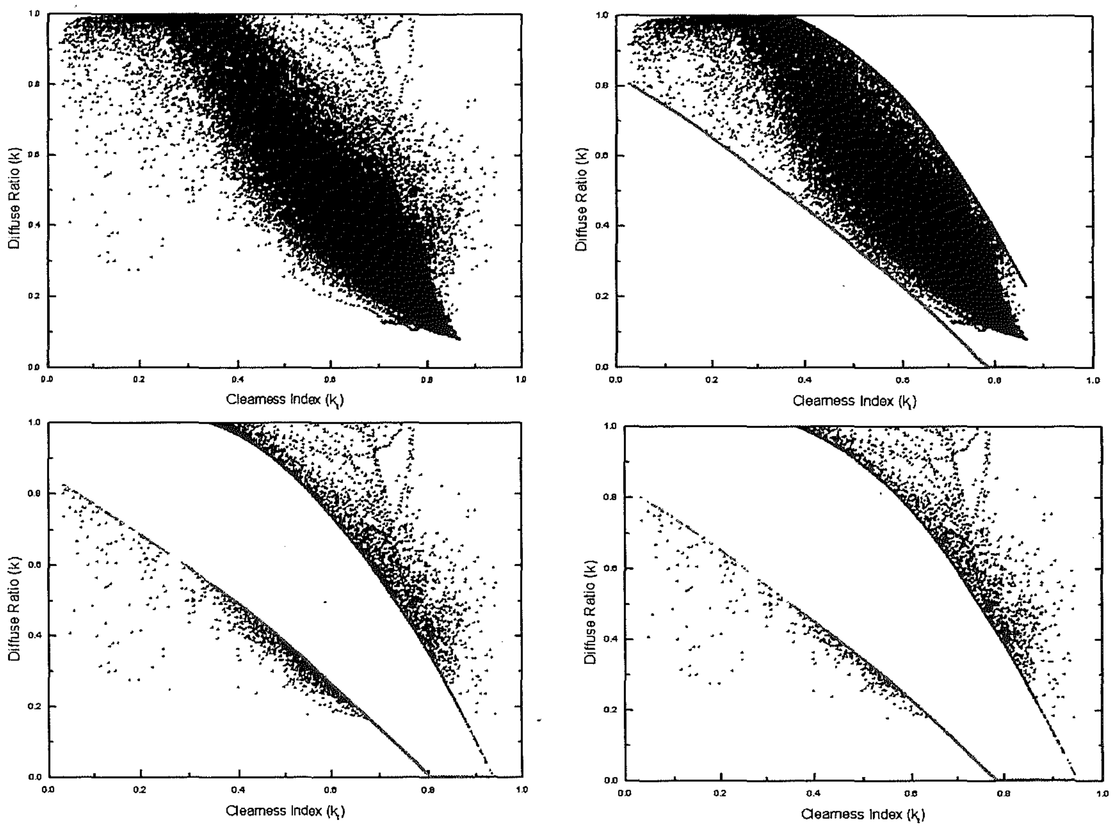


Figure 3.6 Scatter plots for Bahrain. Note that $\bar{k} \pm 2.3\sigma_k$ envelope contains 97.54% of the data that passed the physical tests. a. Raw data; b. Quality controlled ($\bar{k} \pm 2.3\sigma_k$); c. Rejected data based on ($\bar{k} \pm 2.0\sigma_k$) envelope; d. Rejected data based on ($\bar{k} \pm 2.3\sigma_k$) envelope.

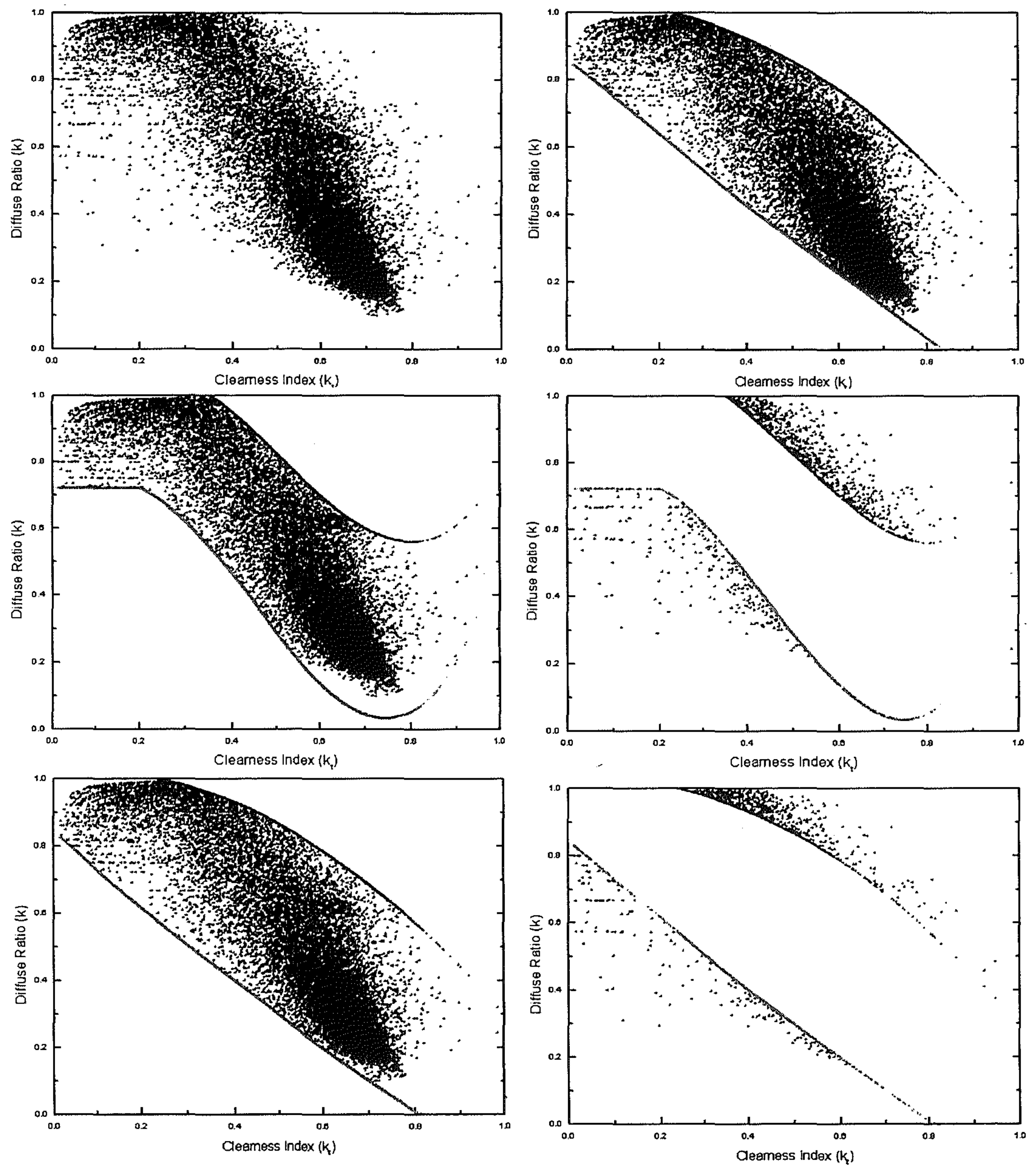


Figure 3.7 Scatter plots for Bombay. a. Raw data; b. Quality controlled ($\bar{k} \pm 2.0\sigma_k$) 2nd degree polynomial envelope; c. Quality controlled ($\bar{k} \pm 2.0\sigma_k$) 3rd degree polynomial envelope; d. Rejected data based on ($\bar{k} \pm 2.0\sigma_k$) 3rd degree polynomial envelope; e. Quality controlled ($\bar{k} \pm 2.2\sigma_k$) 2nd degree polynomial envelope; f. Rejected data based on ($\bar{k} \pm 2.2\sigma_k$) 2nd degree polynomial envelope.

Table 3.1 Quality control results of the 11 datasets.

Location	Length of dataset	Envelope			Points passed (%)
		Polynomial	Standard deviation	Cut-off point	
Bahrain	2000-2002	2	2.3	0.9	97.54
Bracknell	1990-1995	2	2.3	0.9	97.92
Chennai	1990-1994	2	2.2	0.85	96.58
New Delhi	1989-1998	2	2	1	95.26
Eskdalemuir	1995-1999	2	2	0.9	96.36
Fukuoka	1995-1999	2	2	0.8	97.54
Gerona	1995-2001	2	2	0.7	97.58
Madrid	1999-2001	2	2	0.9	95.64
Mumbai	1990-1994	3	2	0.9	94.62
Pune	1990-1194	2	2	0.8	94.32
Sapporo	1991-1993	2	2	0.85	94.93

4 THE IMPROVED METEOROLOGICAL RADIATION MODEL

4.1 Introduction

In Chapter 1, the need for solar radiation modelling, to accurately estimate the solar irradiance at remote locations where there is no equipment for measurement was discussed.

Many different models have been developed by various researchers since the mid-twentieth century, and these rely on different input parameters. The number and kind of input parameters vary and thus the more parameters, the more sophisticated the model. First generation models rely on one or two parameters, often sunshine duration, cloud cover or a combination of either with secondary parameters such as humidity or rain. More complex models also require atmospheric absorption and transmittance due to water vapour, atmospheric gases and heavy metals. One such model is the meteorological radiation model due to Muneer, Gul and Kambezidis (1998, 1998, 2000).

Following from the work of the original authors of the model, a worldwide validation of the model was performed. Usually, most models are validated using high quality datasets; however, engineers do not always have access to these precise datasets, and therefore often use monthly-averaged data. In the present work, two types of datasets were used; Fine datasets, i.e. hourly and sub-hourly datasets (HSHD) using hourly measured synoptic data and coarse datasets, i.e. hourly/daily datasets (HDD) using monthly averaged data. Note that the term synoptic refers to weather elements such as temperature, humidity, sunshine and cloud cover. The datasets used in this work have been quality controlled for errors as explained in Chapter 3.

It was found that there is room for improvement in the above mentioned model. New sunshine band based regressions were added to the original model to improve its performance. The two models, the original and the improved version were compared using the same datasets.

4.2 Radiation and synoptic datasets

The datasets used for this work are from ten locations, all from the northern hemisphere covering two continents: Europe and Asia. There are three European sites; Bracknell (51.26N; 0.46W) in UK, Gerona (41.97N; 2.88E) and Madrid (40.45N; 3.73W) in Spain. The seven Asian sites are: Bahrain (26.22N; 50.65E) in the Middle-East, Chennai (13.00N; 80.18E), Mumbai (19.12N; 72.85E), New Delhi (28.60N; 77.20E) and Pune (18.53N; 73.85E) in India, and finally Fukuoka (33.52N; 130.48E) and Sapporo (43.05N; 141.33E) in Japan.

All datasets include time stamp, global horizontal irradiance, $I_{G,}$ diffuse horizontal irradiance, I_D and sunshine fraction, SF. The datasets have been quality controlled via the quality control procedure presented in Chapter 3.

The Bracknell and Bahrain datasets included dry bulb temperature, DBT, and wet bulb temperature, WBT. Also note that none of the datasets included atmospheric pressure, P_{atm}.

The datasets were thus divided into two groups. The first group contained the hourly and sub-hourly datasets. These were based on hourly (for Bracknell and Bahrain) and ten minute averaged records (For Fukuoka and Sapporo). The second group contained hourly/daily datasets (Chennai, Gerona, Madrid, Mumbai, New Delhi and Pune).

The records that fall in this category have disparate time frequency measurements for solar radiation and sunshine fraction. The former is measured hourly while the latter is daily. Note that for the HDDs only monthly averaged maximum and minimum DBT and monthly averaged relative humidity, RH, were available.

ASHRAE's (1993) method of converting daily averaged maximum and minimum DBTs to hourly values was used to complete and append the hourly radiation and SF data. Similarly daily averaged RH was used to obtain hourly RH, following the work of Muneer (2004).

The Japanese datasets had measured beam irradiance, I_B whereas for other sites it was calculated by subtracting I_D from I_G. P_{atm} was assumed to be equal to 1.01325 bars. The daily SF value was assumed to be prevalent at each concerned hour of the day in the absence of hourly value.

Details for each site are provided in Tables 4.1 and 4.2.

The Measured beam normal and diffuse irradiance for the Japanese sites were to be checked for accuracy. It was found that the data showed some problems when I_D was plotted against I_G-I_B. Note that for the Japanese site, beam normal irradiance was provided, thus there was a need to obtain the horizontal beam irradiance. It was obvious then that I_D was not shadow band corrected. For more information on the purpose and problems of the shadow-band, refer to Section 2.2. A new approach was used to filter out the obvious erroneous data; this filter is mathematically represented as:

$$I_G - I_B = I_D \pm 10\% \tag{4.1}$$

The allowance for 10% is to take into account the circumsolar irradiance effects. Data that has passed the test described in Eq.4.1 was then used for the modelling work. Figure 4.1a represent the scatter plot between measured horizontal diffuse irradiance and its counterpart from the subtraction between global horizontal irradiance and the beam horizontal irradiance. Note the existence of the scatter; to remove all this bias, the test described in Eq.4.1

was utilised and this is represented in Fig. 4.1b. Also note in Fig. 4.1a in the low diffuse value, erroneous data that are cause by either measurement errors or most probably shadow band errors. These errors were removed by the test.

4.3 Improvements on the Meteorological Radiation Model

As was shown in Section 2.67, MRM is based on a power function relationship between DBR and K_B. This relationship is represented mathematically in Eq. 2.18.

A sample plot was shown in Fig 2.5 for Stornoway and in Fig. 4.2 for the Bracknell site. The best-fit equation for the above data was found to be $DBR=0.1894. (K_b)^{-1.1656}$ with the coefficient of determination equal to 0.95. Note that DBR is the diffuse to beam ratio and K_B the beam clearness index.

In an attempt to improve the estimation, the dataset was split into SF bands. For New Delhi, Table 4.3a shows the results of the sunshine bands regressions and their respective R^2.

It was found to be more productive to plot the above banded data in an ln(DBR) versus NLKB where NLKB= $-\ln(K_b)$ rather than DBR VS. K_b scatter plot, as shown in Fig.4.3. Not only was this more useful in identifying the seven sunshine fraction bands of data, but also allowed for easier plotting of the correlation between the two sky condition related irradiance indices. In fact in Fig.4.4a, for New Delhi India, the data was sunshine banded and as it is obvious from observing the scatter plot, the data quite heavily overlaps. The sunshine banded regressions for the same location is given in Fig.4.4b. Note that in Fig.4.4b, in the low Kb value, the regressions overlap. However when plotting the same data in a ln(DBR) – NLKB scatter, and obtaining the sunshine banded regressions, only two bands overlap as is shown in Fig.4.5.

In addition, in Table 4.3b it is clear from the improvement in R^2 that this new regression is more descriptive of the data.

It was found that the relationship between NLKB and ln(DBR) is a 4th order polynomial in the form since this regression yields the highest coefficient of determination. The regression is represented mathematically as:

$$\ln(DBR) = a_0 + a_1(NLKB) + a_2(NLKB)^2 + a_3(NLKB)^3 + a_4(NLKB)^4 \qquad (4.2)$$

However not all ten sunshine bands were used, the ones used are detailed in Table 4.4. Since the data in the range of $0<SF<0.4$ was closely populated, it was considered appropriate to classify the data into the above mentioned band rather than three consecutive bands. Any further increase in the number of sunshine fraction bands would increase the processing time of the model without any noticeable increase in accuracy. There is very little loss in accuracy since, in the case of New Delhi, the single band of $SF>0.4$ represented 90.43% of the data.

The above approach increases the coefficient of determination, R^2, of the regression. For Bracknell the power function as defined in Eq.2.18 yielded an $R^2=0.95$, while for all bands, lower SF limit=0.0 (exclusive) and upper SF limit=1.0 (inclusive) the fourth order polynomial relationship between ln(DBR) and NLKB, Eq.4.2, yielded an $R^2=0.976$. Thus, an improvement of 2.6% between the original and improved regression procedure was observed.

By plotting each data band's best fit line, it is noted that where the bulk of the data is located, the best fit lines do not overlap each other and are in a particular order. Each specific data band has an $R^2>0.8$. It is however not possible to compare the bands based on the value of R^2 as the data population in each band is different, therefore the coefficient of determination would not be valid to compare the performance of each band.

The original MRM is modified in order to estimate the diffuse component of horizontal irradiation more accurately from the regressions between ln(DBR) and NLKB; thus improving the estimation accuracy of global horizontal irradiation.

The improved meteorological radiation model, IMRM, works in a similar manner as the original MRM by Muneer et al. (1998), Gul et al. (1998) and Muneer and Gul (2000). For non-overcast skies, beam horizontal irradiation is a function of the extraterrestrial horizontal irradiation and the sunshine fraction and dampened by the transmittances due to Rayleigh, Mie mixed gases, ozone and water vapour transmittances. Figure 4.6 represents the estimation of I_B for Bracknell; In this case the slope of the calculated versus measured irradiance is of 0.99 and the coefficient of 0.82. The extraterrestrial irradiance dampening is represented mathematically as:

$$I_B = I_E.SF.\tau_r.\tau_\alpha.\tau_g.\tau_o.\tau_w \tag{2.19}$$

τ_r and τ_α are transmittances due to Rayleigh and Mie scattering; τ_g, τ_o and τ_w the transmittances due to mixed gases, ozone and water vapor scattering, respectively. These are expressed mathematically as,

$$\tau_\alpha = \exp\left[-k_\alpha^{COF(1)}(1+k_\alpha-k_\alpha^{COF(2)})m^{COF(3)}\right] \tag{2.20}$$

$$\tau_O = 1-\left[0.1611x_o(1+139.48x_o)^{-0.3035}-0.002715x_o(1+0.044x_o+0.0003x_o^2)^{-1}\right];$$

$$x_0 = l_o m \tag{2.21}$$

$$\tau_r = COF(4)-COF(5)m+COF(6)m^2-COF(7)m^3+COF(8)m^4 \tag{2.22}$$

$$\tau_w = 1-COF(9)x_w\left[(1+COF(10)x_w)^{COF(11)}+COF(12)x_w\right]^{-1};$$

$$x_w = l_w m \tag{2.23}$$

$$\tau_g = \exp(-COF(13)m^{COF(14)}) \tag{2.24}$$

With COF(i) the transmittance coefficients.

It was attempted to modify the transmittance formulas in order to improve the estimation of beam irradiance. The results proved to be weaker than the original transmittance coefficients as shown in Fig.4.7 the measured versus calculated I_B scatter plot using modified transmittance formulas. In this case, the result was a slope of 1.1, and a coefficient of determination 0.72. In comparison with the results shown in Fig.4.6, the original equations for transmittance were kept.

When processing the datasets, the improved MRM requires the same inputs as the original model, i.e. time stamp, DBT, WBT, P_{atm}, SF and the 14 coefficients for the transmittances equations. The improved model however requires 35 coefficients per site that are the result of the five coefficients of the 4th degree polynomial regression for each of the seven SF based data bands. Then by averaging each band regression coefficients for all the ten sites, a set of generalized regression coefficients we compiled, as shown in Table 4.5.

The regression lines in Fig.4.8, representing each of the bands in the range $1.0 < -\ln(K_b) < 4$ were perfectly aligned where the bulk of the data in each dataset lies. Due to scarcity of the data in the $-\ln(K_b) < 0.9$ and $-\ln(K_b) > 4.0$ bands, the correlation lines overlap and do not follow an order.

The original MRM was used to estimate the global horizontal irradiation for the ten locations used. For each of the ten sites, the model was tuned by tuning the 14-transmittance coefficients, to suit the location and maximize the accuracy of the estimation.

Since the original model has been validated, by Muneer et al. (1998) using datasets from the UK and Japan, the same transmittance coefficients were used in the current validation for the UK and Japanese datasets. However, the validation of other seven sites two Spanish, four Indian and one Middle-eastern site, needed re-tuning for the 14-transmittance coefficients for a perfect regression relationship. This was done initially for the original MRM revalidation, and the same coefficients were then used in the validation of

IMRM and inter-comparison between the two models. Note that the MRM was only processed for data with solar altitude angle greater than 10°.

The two models, original MRM and the proposed improved MRM were compared using statistical procedures/analysis. The coefficient of determination for the best fit line was calculated. Unfortunately R^2 is not a precise method of comparing the two models, as it is generally noted that the coefficient of determination increases with the increase in the amount of coefficients in the model. The original model relied on two coefficients in the regression between DBR and K_b, while the improved model relies on 35 coefficients in the regression between ln(DBR) and NLKB, thus the expected improvement in the coefficient of determination.

To avoid misjudgement of the improved R^2 for the new model two other statistical comparison methods: mean bias error, MBE and the root mean square error, RMSE, were used. The MBE provides an indication of the trend of the model i.e. whether it has a tendency to under-estimate or over-estimate its modelled values. MBE can be expressed either as a percentage or as an absolute value. The RMSE gives a value of the level of scatter that the model produces. This is an important statistical test, as it highlights the readability and repeatability of the model. R^2, MBE and RMSE are explained in more details in Section 2.51, and are mathematically expressed in Eqs.2.3-2.5.

Note that the values of MBE and RMSE are in W/m^2, same unit as the solar irradiance. Peers have often used either, the same MBE and RMSE or have expressed them in percentage. In some cases both unit and percentage MBE and RMSE have been used. Since the models, original and improved, will not be compared to other models, and will only be compared against each other, only one method of representing the statistical indicators will be used.

Thus for each location, two sets of three statistical parameters were employed, the coefficient of determination of the best fit line between computed and measured global irradiation data, mean bias error and root mean square error for both the original MRM and the improved MRM. The 14 transmittances coefficients used for each site for the improved MRM were the same as for the original MRM. For the original MRM validation, the two regression coefficients, obtained in the correlation between DBR and K_b by the work of Muneer, Gul and Kambezidis (1998, 1998, 2000) in their original validations, were used for all sites. For the improved MRM, the 35 generalized regression coefficients were used on all the sites.

4.4 Discussion of results

HSHDs and HDDs, comprising datasets from the 10 sites, were both processed by the original and the improved meteorological radiation models, by a case-by-case comparison, the improved model performed better with regards to reducing the amount of scatter and errors in the estimation process. This can be seen in Table 4.6 for HSHDs and in Table 4.7 for HDDs listing the statistical results for both the MRM and the IMRM.

The improvement can be visually observed in Fig.4.9 (a, b) for Bracknell HSHD and Fig.4.10 (a, b) for Mumbai HDD. It can be noted that in Figs.4.9b and 4.10b, the data is more symmetrical with the optimum 1:1 line, as well as the scatter is clearly reduced compared to the results of the validation of the original model.

Note that in Fig.4.10 (a, b) a considerable number of outliers are visible in the scatter plot compared with Fig. 4.9 (a, b). This is not an isolated case, as all HDD scatter plots for both the MRM and IMRM yielded considerable outliers. This is also confirmed by the results of the statistical analysis, in high MBE and RMSE values. This amount of scatter is caused by the nature of the HDDs. Single daily values of SF are assigned to hourly radiation values. If for a certain day, the SF value is given as 0.5; it does not necessarily indicate that every hour has an SF value of 0.5. In reality a daily SF value of 0.5 might indicate that the morning hours could have been in clear skies, while in the afternoon, sky was fully overcast. The outlier points, in the upper half of the scatter plots indicate that the daily SF values are low, sign of overcast sky, while the high global horizontal irradiation values indicate clear sky at the measured hours of the otherwise overcast days and vice-versa for outliers in the lower half of the scatter plots.

Further examination of the results of the evaluation of the two models, shows that on average, for the HSHD sites, there is an increase in R^2 of 7.2% and a decrease of 83.9% and 60.7% in the absolute mean bias error, AMBE and root mean square error, respectively. The same was also observed for HDD sites with an increase in R^2 of 24.2% and a decrease of 72.2% and 63.3% in AMBE and RMSE, respectively. Overall, an increase in R^2 of 0.13, and a decrease of 63- and 76 W/m^2 AMBE and RMSE, respectively, was observed. More details for each site can be found in Table 4.8.

Error histograms have also been plotted for each site from both model estimations, as can be seen in Fig.4.11 for HSHDs and Fig.4.12 for HDDs. It is another good indication of reduction in estimation error in the improved model as compared to the original model.

For further validation of the model, a residual examination was required, as discussed by Muneer (2004). This is done by a procedure that produces a graph of the residuals or errors d (the difference between the observed I_G and calculated I_G) plotted against an independent variable n (the number of data points). In an ideal scenario, the plot of error has a horizontal shape. Further inspection of the errors plots suggests the model to be adequate enough, as shown in Fig.4.13 plot of residuals for the Bracknell HSHD.

Muneer (2004) performed an evaluation of different models for estimating global horizontal irradiance. The cloud radiation model, CRM, explained in Chapter 2, and used again in Chapter 5; the meteorological radiation model, MRM; and the Page radiation model, PRM, also featured in Chapter 2 and discussed in Chapter 6. To perform the evaluation, Muneer (2004) employed a scoring procedure, based on two statistical indicators, MBE and RMSE in

W/m^2. This scoring procedure is a crude precursor of the accuracy scoring procedure developed in Chapter 5 and used as well in Chapter 6. In this scoring procedure, the model with the smallest score is the best performing. The results of that evaluation showed that for overcast skies, PRM performed best with a score of 337 W/m^2 compared to 596- and 510 W/m^2 for MRM and CRM, respectively. In mixed skies, CRM performed best, followed by MRM and PRM, with scores of 392-, 418- and 597 W/m^2, respectively. In quasi-clear skies, it was found that MRM performed best, followed by CRM and PRM with respective scores of 613-, 972- and 1268 W/m^2. Thus overall, CRM scored 1874 W/m^2, MRM 1627 W/m^2 and PRM 2202 W/m^2, placing MRM at the lead. Note that MRM and PRM will be re-evaluated in Section 6.7 under different clear-sky datasets.

Using the same scoring procedure, it was found that for both HSHD sites and HDD sites, the newer improved model outperformed the older model. The Overall scores averaged at 303- and 163 W/m^2 for MRM and IMRM, respectively. More details can be found in Table 4.9 for all the sites used. Note that, in his study Muneer (2004) has used a common site as in this study. However, the length of the dataset was different, thus the discrepancies in results between the two studies. In Muneer's study, for all sky conditions of Bracknell, MRM scored 105 W/m^2 compared to 136 W/m^2. It was found that, in the dataset used in this study, MRM scored 272 W/m^2 while IMRM scored 87.7 W/m^2. Based on this, it would logically follow that IMRM would also outperform PRM for all sky conditions. For specific clear-sky conditions, the comparison between the models will be presented in more details in Chapter 6.

It can also be extracted from Table 4.9, that for the HSHD sites, the MRM underestimates the global horizontal irradiance, while IMRM slightly overestimates it. Meanwhile, for the HDD sites, both models underestimate the global horizontal irradiance, with much improvement in the IMRM evaluation. On average, for both site classifications, the IMRM has a MBE of nearly 0 W/m^2.

4.5 Conclusions

The meteorological radiation model was developed and evaluated by Muneer, Gul and Kambezidis (1998, 1998, 2000). However it was observed and noted that there were weaknesses in the model. Various methods of improving the model were evaluated.

The MRM estimates the beam irradiation by attenuating the extraterrestrial irradiation by estimating the atmospheric transmittance coefficients. New formulae were developed to improve the estimation of the transmittance coefficients; however it was found that using the new formulae did not achieve this goal.

A different approach was undertaken to improve the model. The original MRM estimated the diffuse irradiance based on a correlation between the beam clearness index and the diffuse to beam ratio. Note as well that the global irradiance is simply the sum of the two irradiance components, beam and diffuse. The above mentioned correlation was to be sunshine banded, thus increasing the precision in the regression. It was found however that it was more straightforward to create a new sunshine banded regression based on the correlation between the negative natural logarithm of the beam clearness index and the natural logarithm of the diffuse to beam ratio. Rather than use ten sunshine fraction bands, further analysis of the data showed that only 7 sunshine bands would be enough since the overcast sky conditions bands could be amalgamated. The regression between the above indices was found to be a 4th order polynomial. Based on the 10 datasets available for this study, it was possible to create a generalised set of coefficients.

The newly developed model was evaluated against its predecessor in order to verify for any improvements. In total 10 sites were used for the evaluation process. Note that the original MRM is a site sensitive model, thus local tuning must be allowed to provide accurate estimations. In this respect the original model was locally tuned, and the same transmittance coefficients were used in the modified model.

It was found, after evaluation of both models that the modified MRM increased the coefficient of estimation by 0.13, and reduced the mean bias error and the root mean square error by 63- and 76 W/m^2, respectively. In addition, by using the same scoring procedure previously employed by Muneer (2004) to compare models, which consists of summing up the absolute MBE and the RMSE value, and the lowest scoring model is the best performing, the new modified and improved MRM, IMRM scored on average 163-, compared to 303 W/m^2 for the original MRM.

It is therefore concluded that the IMRM is a more accurate model than its predecessor.

Since the improved meteorological radiation model does demand multiple inputs to estimate the solar irradiance, and that sometimes the users do not have easy access to these parameters, other more simplistic models can be used that only rely on one or two parameters. In Chapter 5, one such breed of models is investigated.

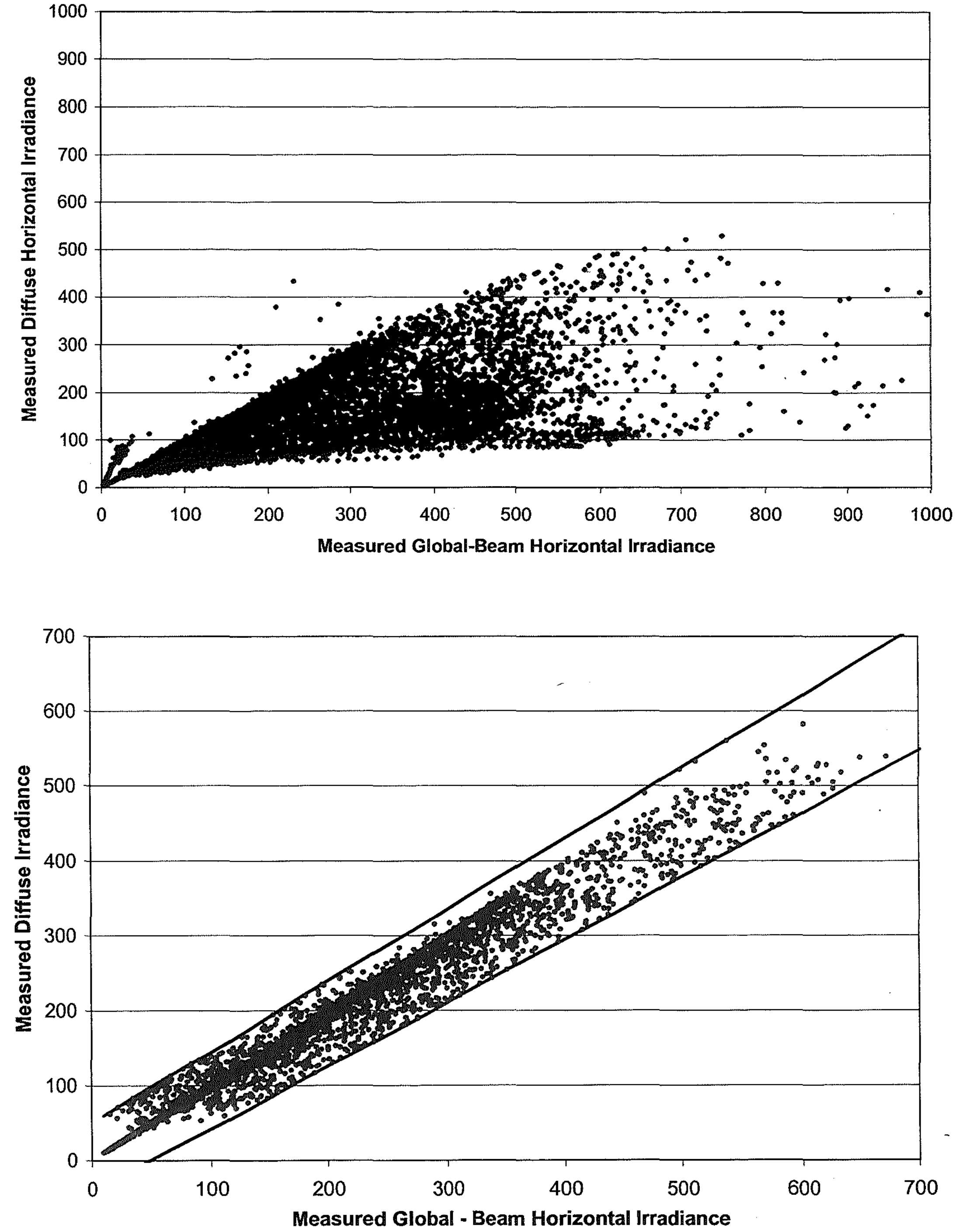


Figure 4.1 Scatter plot of I_D versus I_G-I_B for Sapporo. a. Before correction of horizontal diffuse Irradiance. b. After correction, and showing upper and lower limits.

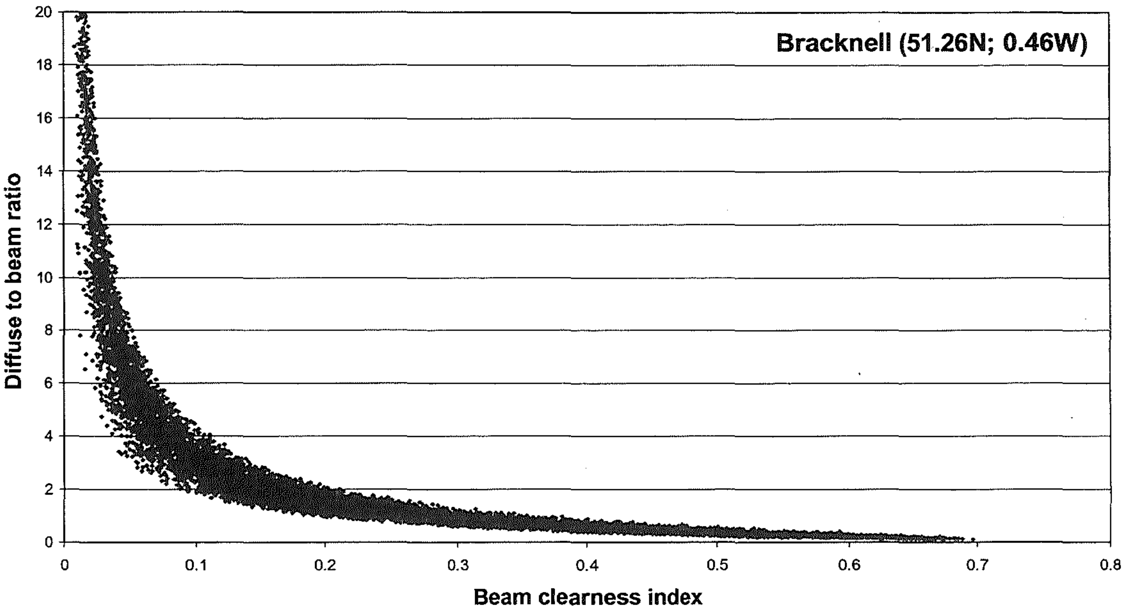


Figure 4.2 Scatter plot of DBR vs. K_b for the Bracknell site in the U.K. Note that DBR and K_b are dimensionless.

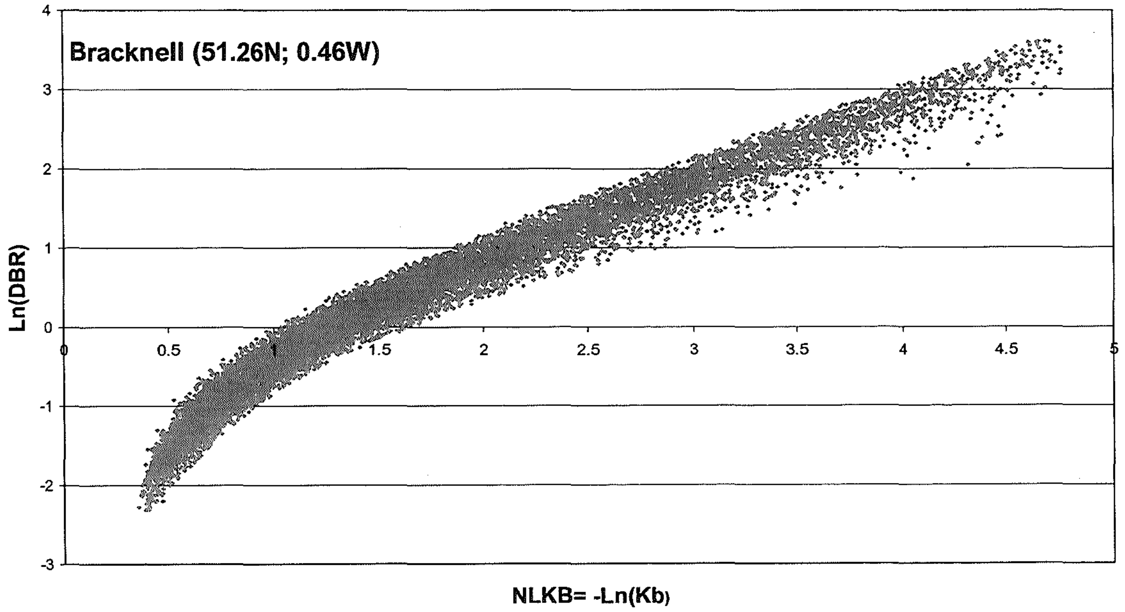


Figure 4.3 Scatter plot of ln(DBR) vs. NLKB for the Bracknell site in the U.K. Note that ln(DBR) and NLKB are dimensionless.

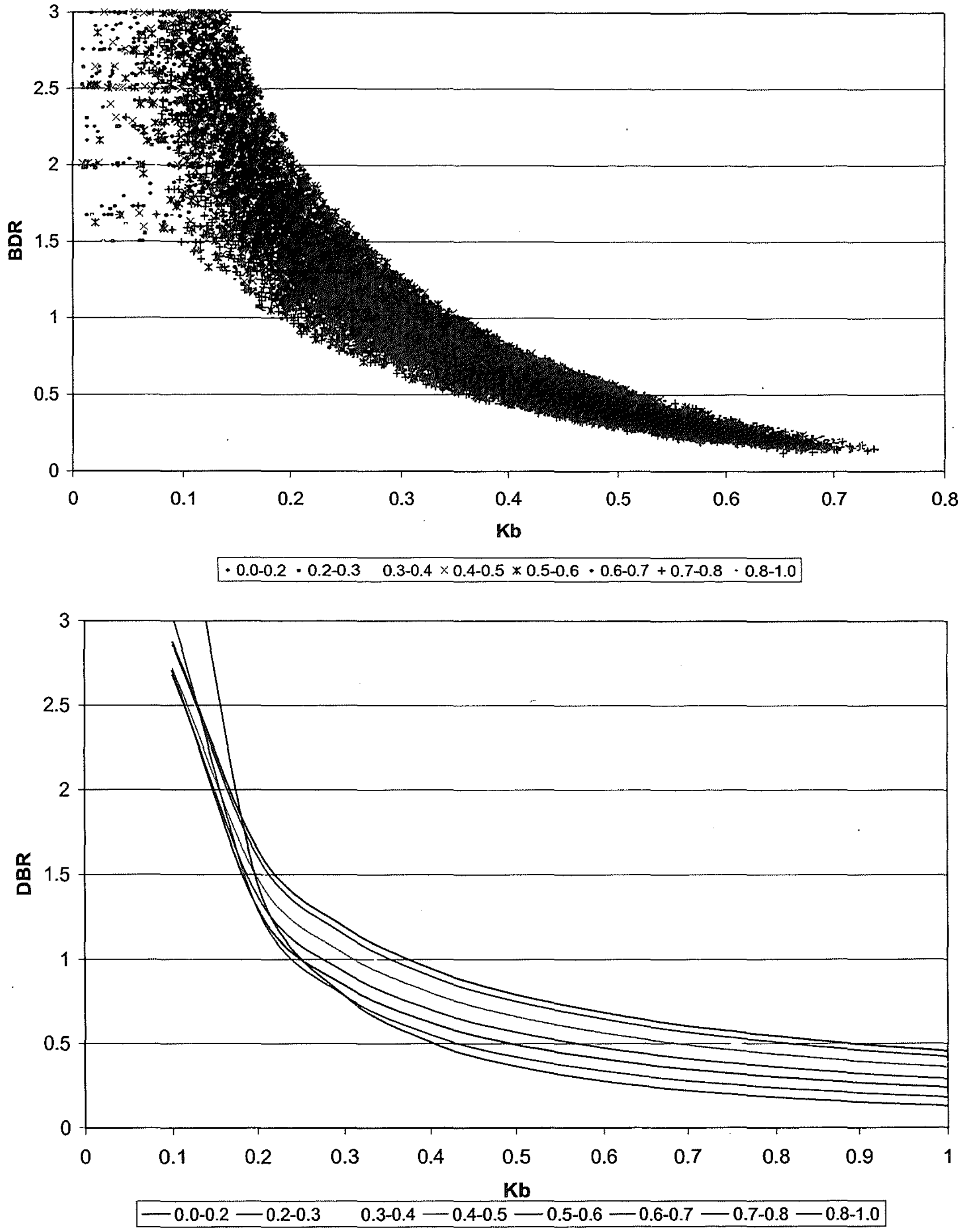


Figure 4.4 Scatter plots of a-sunshine banded DBR vs. K_b and b- sunshine band regressions for the New Delhi site in India. Note that DBR and K_b are dimensionless.

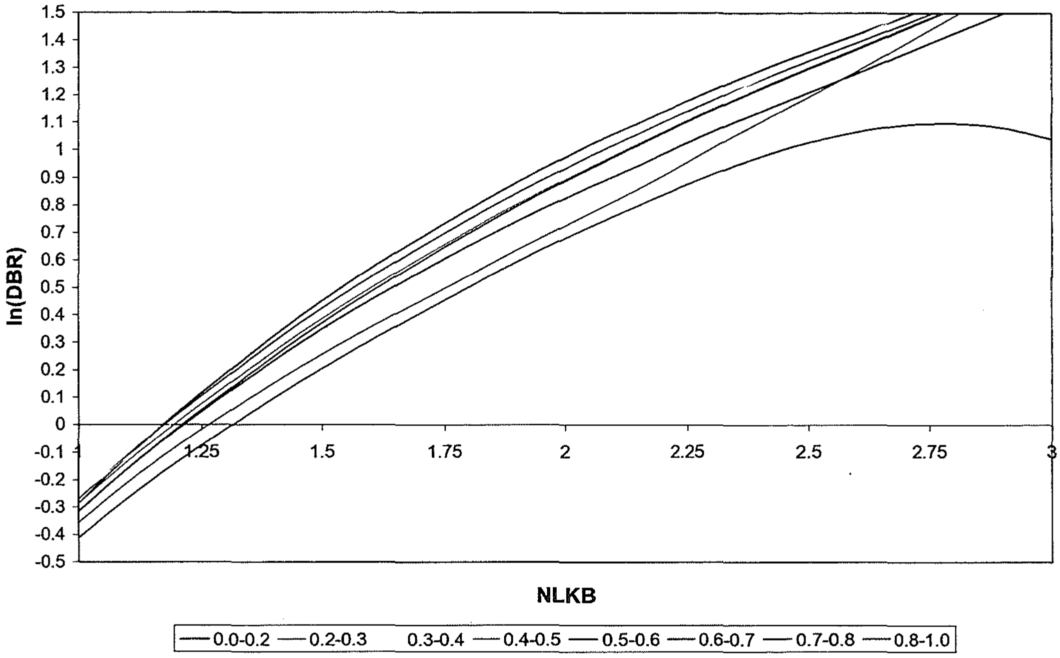


Figure 4.5 Scatter plots of sunshine band ln(DBR) – NLKB regressions for the New Delhi site in India.

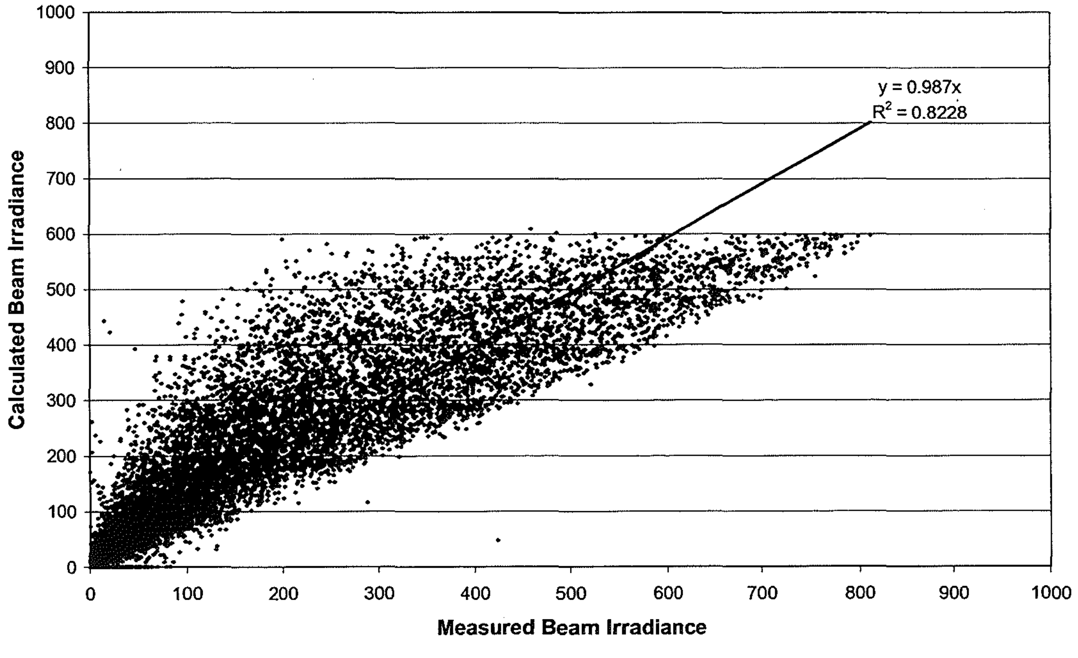


Figure 4.6 Scatter plot of measured vs. calculated I_B for the original MRM validation for the Bracknell HSHD site in the UK.

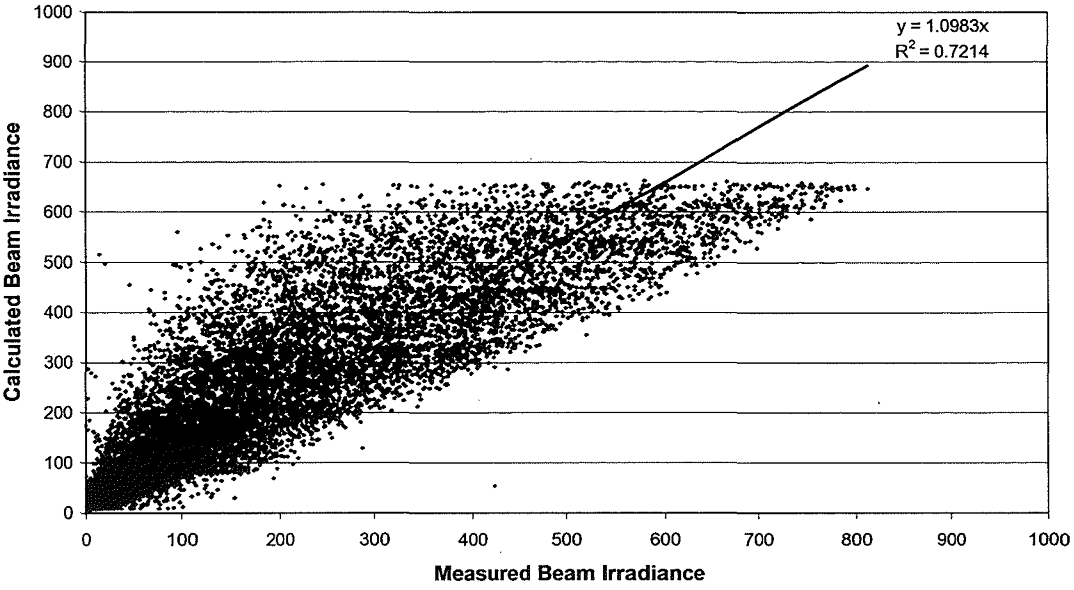


Figure 4.7 Scatter plot of measured vs. calculated I_B for the modified MRM for the Bracknell HSHD site in the UK

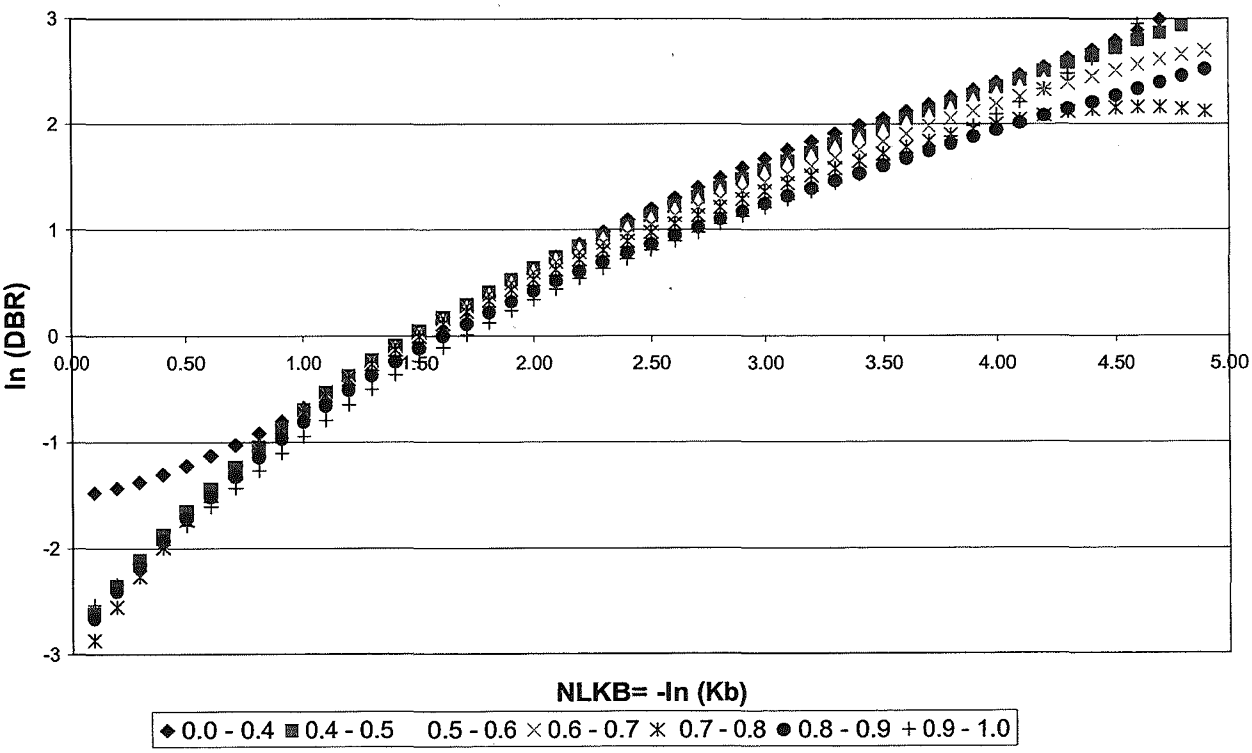
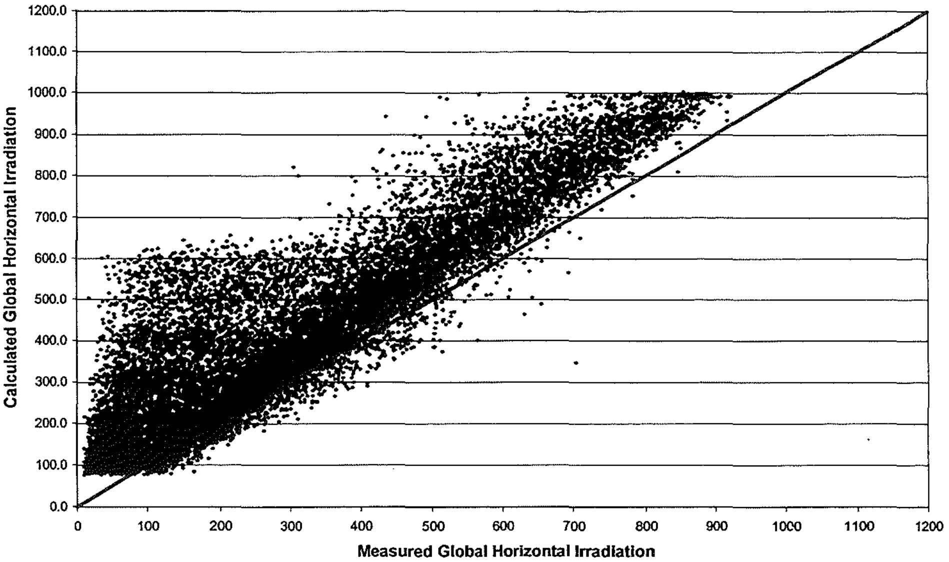


Figure 4.8 Regression between ln (DBR) and NLKB for all 7 SF bands. Note that ln(DBR) and NLKB are dimensionless.



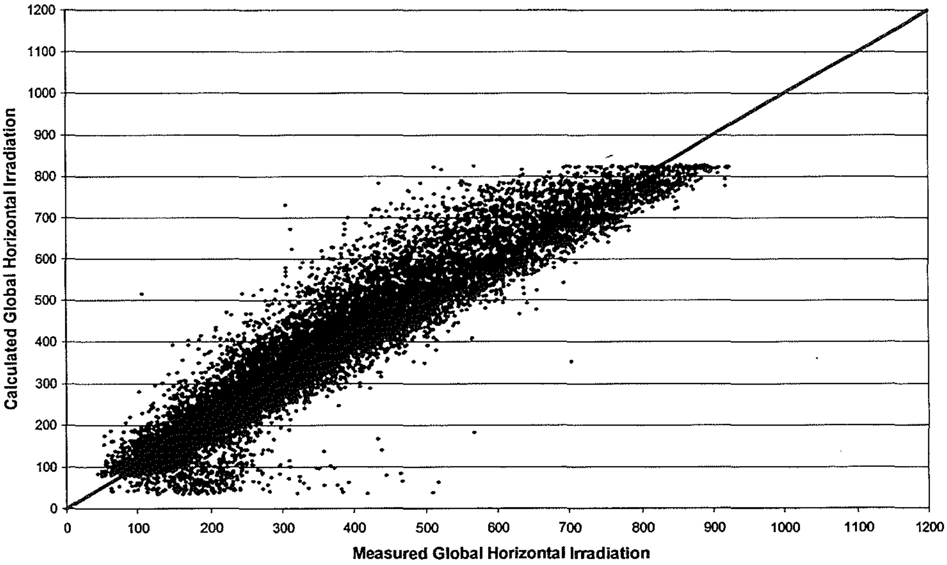
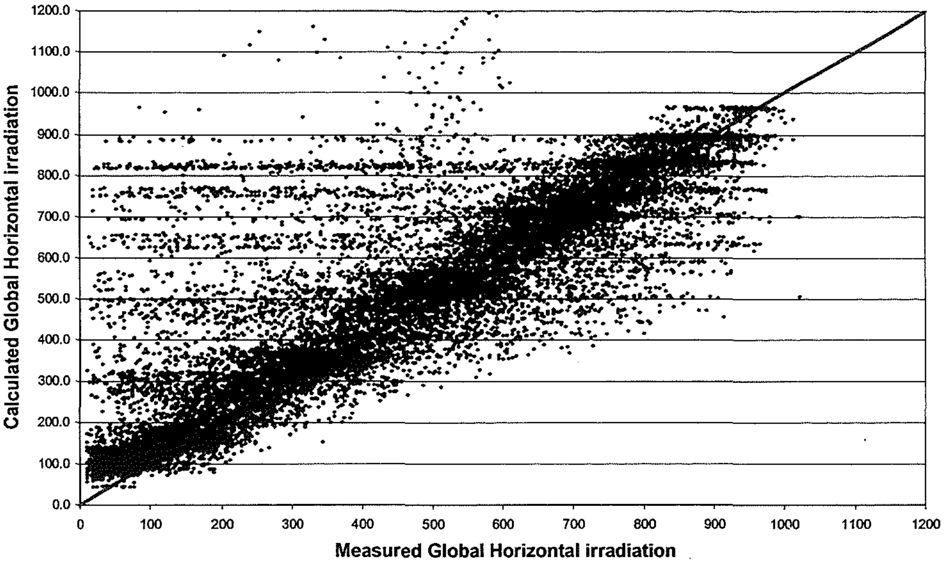


Figure 4.9 a. Scatter plot of measured vs. calculated I_G for the original MRM validation for the Bracknell HSHD site in the UK. b. Scatter plot of measured vs. calculated I_G for the proposed IMRM validation for the Bracknell HSHD site in the UK. Note that measured and calculated I_G are measured in W/m^2.



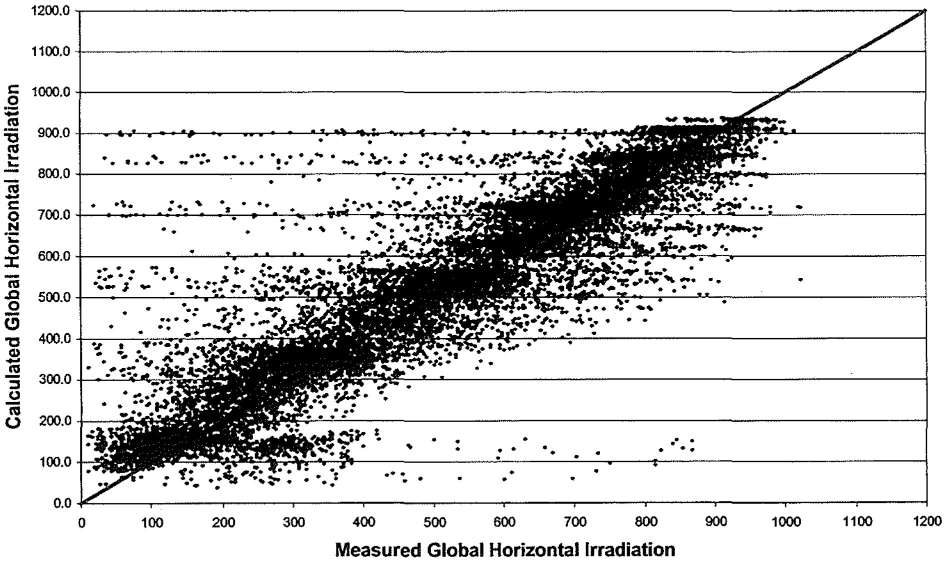


Figure 4.10 a. Scatter plot of measured vs. calculated I_G for the original MRM validation for the Mumbai HDD site in India. b. Scatter plot of measured vs. calculated I_G for the proposed IMRM validation for the Mumbai HDD site in India. Note: The outliers in the above Figures are caused by the daily SF values assigned to hourly solar irradiation data. The measured and calculated I_G are measured in W/m^2.

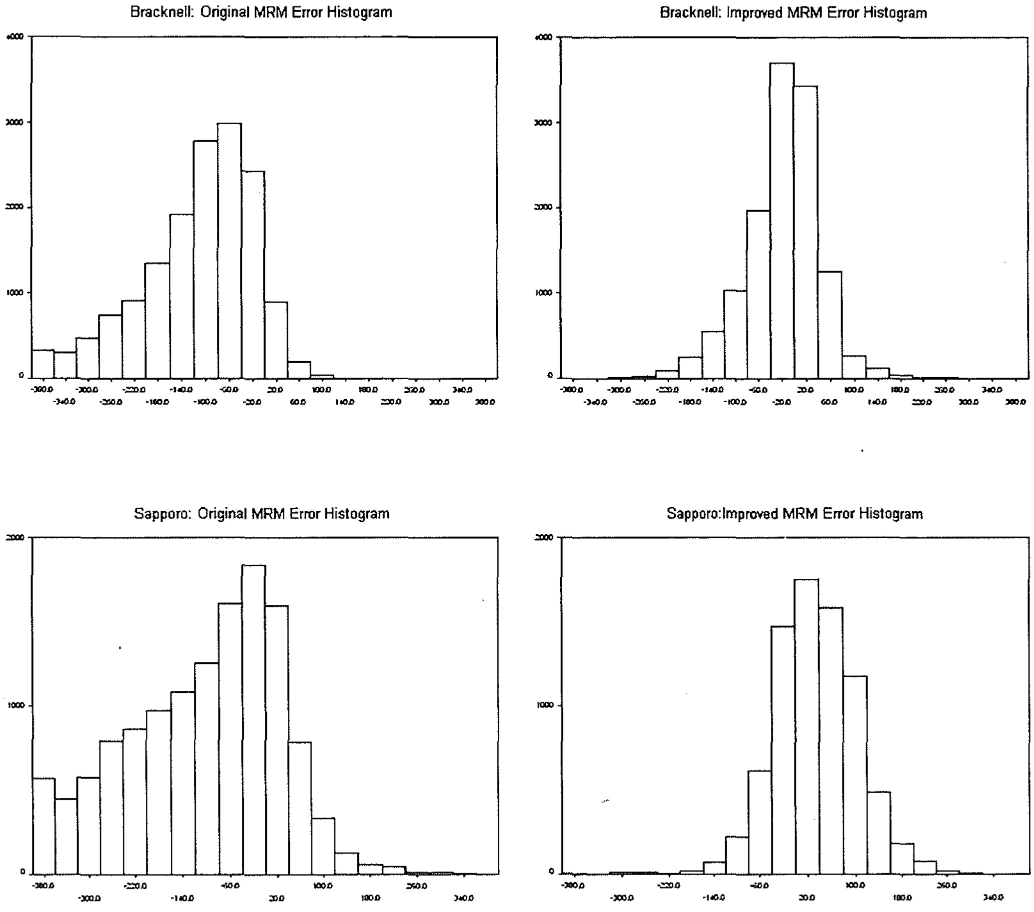


Figure 4.11 The original and proposed improved MRM error histograms for the Bracknell and Sapporo HSHDs. Note that x axis represents radiation error in W/m^2 and y axis is the number of occurrences.

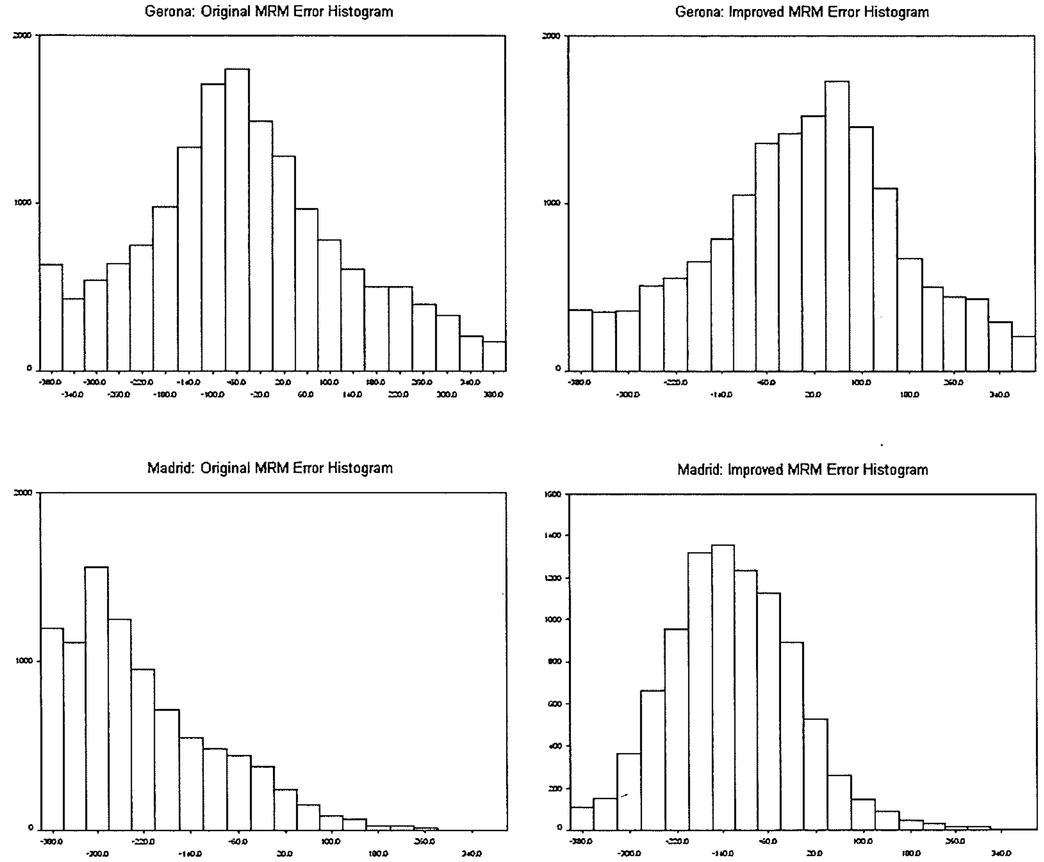


Figure 4.12 The original and proposed improved MRM error histograms for the Gerona and Madrid HDDs. Note that x axis represents radiation error in W/m^2 and y axis is the number of occurrences.

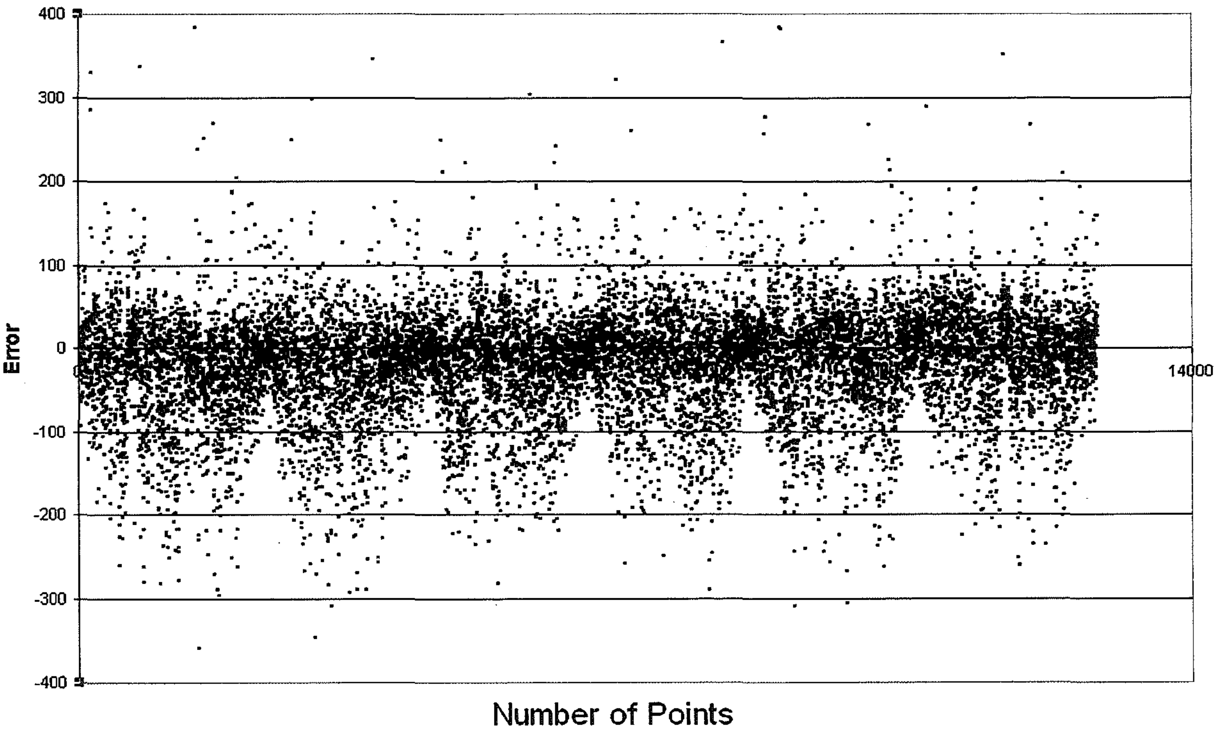


Figure 4.13 Plot of residuals of the proposed improved MRM validation for the Bracknell HSHD site in the UK. Note that y axis is the radiation error in W/m^2.

Table 4.1 Table shows the classification and length of the datasets used.

Country	Location	Latitude; Longitude	Year	Description of dataset
Bahrain	Manama	26.22N; 50.65E	2000-2002	HSHD
India	Chennai	13.00N; 80.18E	1990-1994	HDD
India	Mumbai	19.12N; 72.85E	1990-1994	HDD
India	New Delhi	28.60N; 77.20E	1988-1998	HDD
India	Pune	18.53N; 73.85E	1990-1994	HDD
Japan	Fukuoka	33.52N; 130.48E	1992-1994	HSHD
Japan	Sapporo	43.05N; 141.33E	1991-1993	HSHD
Spain	Gerona	41.97N; 2.88E	1995-2001	HDD
Spain	Madrid	40.45N; 3.73W	1999-2001	HDD
UK	Bracknell	51.26N; 0.46W	1990-1995	HSHD

* HSHD refers to hourly or sub hourly radiation and sunshine duration dataset.

** HDD refers to hourly radiation to daily sunshine duration dataset.

Table 4.2 Table shows the synoptic and radiation details of the ten datasets used in this study.

Location	4.5.1 Radiation Measured G=global, D=diffuse, B=beam	Temperature DBT, WBT	Radiation measurement Frequency Hourly/ 10 minute	SF measurement Frequency Daily/Hourly/10 minute
Manama	G,D	D,W	Hourly	Hourly
Chennai	G,D	none	Hourly	Daily
Mumbai	G,D	none	Hourly	Daily
New Delhi	G,D	none	Hourly	Daily
Pune	G,D	none	Hourly	Daily
Fukuoka	G,D,B	D,W	10 minutes	10 minutes
Sapporo	G,D,B	incomplete	10 minutes	10 minutes
Gerona	G,D	none	Hourly	Daily
Madrid	G,D	none	Hourly	Daily
Bracknell	G,D,B	D,W	Hourly	Hourly

Table 4.3 Sunshine banded a-K_b and DBR regression coefficients and their respective R^2; b- NLKB and ln(DBR) regression coefficients and their respective R^2; for MRM using New Delhi dataset.

SF	a	b	R^2
0.0-0.2	0.797	-0.777	0.85
0.2-0.3	0.829	-0.859	0.89
0.3-0.4	0.863	-0.942	0.93
0.4-0.5	0.880	-1.026	0.92
0.5-0.6	0.965	-1.237	0.92
0.6-0.7	1.056	-1.438	0.92
0.7-0.8	1.230	-1.723	0.90
0.8-1.0	1.510	-2.061	0.88

SF	a_0	a_1	a_2	a_3	a_4	R^2
0.0-0.2	-2.803	3.432	-1.077	0.171	-0.009	0.86
0.2-0.3	-2.545	3.052	-0.913	0.143	-0.008	0.90
0.3-0.4	-2.159	2.420	-0.609	0.092	-0.005	0.94
0.4-0.5	-2.454	2.900	-0.867	0.143	-0.008	0.94
0.5-0.6	-2.505	2.905	-0.840	0.131	-0.007	0.95
0.6-0.7	-2.682	3.346	-1.191	0.227	-0.015	0.95
0.7-0.8	-3.228	4.826	-2.652	0.783	-0.084	0.95
0.8-1.0	-3.268	4.977	-3.039	1.067	-0.149	0.92

Table 4.4 The 7 SF bands used to split the datasets to improve the accuracy of the regressions.

Band Number	Sunshine Fraction
1	$0 \leq SF < 0.4$
2	$0.4 \leq SF < 0.5$
3	$0.5 \leq SF < 0.6$
4	$0.6 \leq SF < 0.7$
5	$0.7 \leq SF < 0.8$
6	$0.8 \leq SF < 0.9$
7	$0.9 \leq SF \leq 1.0$

Table 4.5 The 35 generalized coefficients needed for the IMRM based on Eq. 4.2 for the specified range of the data.

SF Band	-ln (K_b),min	-ln (K_b),max	Regression Coefficients for Eq. (8)					
			a_0	a_1	a_2	a_3	a_4	R^2
1	0.000	5.000	-1.4786	0.3528	0.8191	-0.2744	0.0273	0.94
2	0.000	5.000	-2.6147	2.6782	-0.7084	0.1201	-0.0079	0.97
3	0.000	5.000	-2.8485	3.0903	-0.9961	0.2044	-0.0168	0.97
4	0.000	5.000	-2.8717	3.2533	-1.1234	0.2305	-0.0182	0.97
5	0.000	5.000	-2.8661	3.2627	-1.1995	0.2691	-0.0241	0.96
6	0.000	5.000	-2.6631	2.6301	-0.7453	0.1302	-0.0088	0.93
7	0.000	5.000	-2.53370	1.9348	-0.1349	-0.0751	0.0155	0.87

* Note that the coefficient of determination for each of the 7 bands is above 0.85.

Table 4.6 Table shows validation results of HSHDs.

Location	Model	R^2	MBE (W/m^2)	RMSE (W/m^2)
Bahrain	MRM	0.78	-47.1	170.5
	IMRM	0.75	12.3	151.9
Bracknell	MRM	0.81	-117.0	155.3
	IMRM	0.90	-19.4	68.3
Fukuoka	MRM	0.93	-67.4	96.9
	IMRM	0.92	30.5	76.2
Sapporo	MRM	0.71	-123.2	199.2
	IMRM	0.91	33.8	80.9

* The MRM is the original model.
**The IMRM is the proposed improved model.

Table 4.7 Table shows validation results of HDDs.

Location	Model	R^2	MBE (W/m^2)	RMSE (W/m^2)
Chennai	MRM	0.41	134.9	254.5
	IMRM	0.58	-25.7	163.2
Gerona	MRM	0.37	-59.3	226.0
	IMRM	0.37	2.7	201.5
Madrid	MRM	0.63	-242.8	309.0
	IMRM	0.76	-121.9	171.1
Mumbai	MRM	0.26	-92.2	346.9
	IMRM	0.79	-15.1	114.8
New Delhi	MRM	0.78	61.1	119.3
	IMRM	0.91	74.8	94.0
Pune	MRM	0.68	-42.4	163.9
	IMRM	0.72	18.5	155.2

* The MRM is the original model.
**The IMRM is the proposed improved model.

Table 4.8 Comparison between MRM and IMRM for global horizontal irradiance.

Location	Classification	Δ R^2	Δ AMBE (W/m^2)	Δ RMSE (W/m^2)
Bahrain	HSHD	-0.03	-34.8	-18.6
Bracknell	HSHD	0.09	-97.6	-87.0
Chennai	HDD	0.17	-109.2	-91.3
Fukuoka	HSHD	-0.01	-36.9	-20.7
Gerona	HDD	0.00	-56.6	-24.5
Madrid	HDD	0.13	-120.9	-137.9
Mumbai	HDD	0.53	-77.1	-232.1
New Delhi	HDD	0.13	13.7	-25.3
Pune	HDD	0.04	-23.9	-8.7
Sapporo	HSHD	0.20	-89.4	-118.3
Average HSHD		0.06	-64.7	-61.2
Average HDD		0.17	-62.3	-86.6
Average Total		0.13	-63.3	-76.4

* Note that positive values indicate an increase, and negative values a decrease.

Table 4.9 Evaluation of MRM and IMRM for global horizontal irradiance.

		MBE (W/m^2)		RMSE (W/m^2)		Score (W/m^2)	
Location	Classification	MRM	IMRM	MRM	IMRM	MRM	IMRM
Bahrain	HSHD	-47.1	13.3	170.5	151.9	217.6	165.2
Bracknell	HSHD	-117.0	-19.4	155.3	68.3	272.3	87.7
Chennai	HDD	134.9	-25.7	254.5	163.2	389.4	188.9
Fukuoka	HSHD	-67.4	30.5	96.9	76.2	164.3	106.7
Gerona	HDD	-59.3	2.7	226.0	201.5	285.3	204.2
Madrid	HDD	-242.8	-121.9	309.0	171.1	551.8	293.0
Mumbai	HDD	-92.2	-15.1	346.9	114.8	439.1	129.9
New Delhi	HDD	61.1	74.8	119.3	94.0	180.4	168.8
Pune	HDD	-42.4	18.5	163.9	155.2	206.3	173.7
Sapporo	HSHD	-123.2	33.8	199.2	80.9	322.4	114.7
Average HSHD		-88.7	14.6	155.5	94.3	244.2	118.6
Average HDD		-40.1	-11.1	236.6	150.0	342.1	193.1
Average Total		-59.5	-0.9	204.2	127.7	302.9	163.3

5 CLOUD RADIATION MODELLING

5.1 Introduction

Solar radiation is being measured at many sites around the globe. These sites however are mainly concentrated in the developed countries, and are scarcer within the developing world. Even where measured data are available for a given region, the data becomes less accurate the further away the location, of implementation for use of solar energy, is from the measuring station. To an extent, that for a given location that is beyond 50 km from the measurement station the use of the respective measurement station's data is obsolete in the assessment of solar energy applications. Thus, simple and easy to use solar radiation estimation models that are based on more commonly available data are needed. Note that the number of surface weather stations recording solar radiation is very small compared to the number of meteorological stations that record cloud cover and other synoptic information.

We have seen in Chapter 4, a multivariate model, the meteorological radiation model. In some instances it could be very difficult to obtain all the required parameters, therefore models that are simplistic and that are based on one or two parameters are necessary. Such models are based on either one or a combination of the following parameters: cloud amount, relative humidity, rain amount, sunshine duration, temperature.

Since the 1960's researchers have been developing solar radiation estimation models based on cloud information. In this chapter, we shall compare three established models to newly proposed, improved models. The datasets used in this work have been quality controlled based on the work developed in Chapter 3.

5.2 *Radiation and synoptic datasets*

The datasets available for the present analysis cover three countries, each with specific meteorological and geographical characteristics. Hubbard (1994) demonstrated that the length of data series should be more than one year to characterize the seasonal pattern in special variability. Gueymard (1999) recommends that in order to validate radiation estimation models, a minimum three-year dataset is needed. Thus, the following seven datasets presently used are based on a span of more than three years and are used for the comparison and conception of the models.

Initially, only two datasets came from UK locations, Aldergrove (54.65N; 6.25W) and Bracknell (51.26N; 0.45E). The Aldergrove dataset contained a five-year period (1990 to 1994), and Bracknell set spanned six years (1990 to 1995). The two UK locations are an example of temperate, maritime weather found · in northern Europe. The two, radiation and meteorological measurement, stations are classed as first class stations, and are part of the UK Meteorological Office network. The radiation is recorded hourly from minute-by-minute averaged data. The cloud cover is recorded hourly using the Alidade unit. This unit allows the recording of the cloud cover and the height of the cloud base.

Data from two Southern European sites were also utilized, both sites being from Spain, i.e. Gerona (41.97N; 2.88E) and Madrid (40.45N; 3.73W). The Gerona dataset spanned seven years, covering the period from 1995 to 2001 and was provided by the University of Gerona, while the Madrid dataset spanned three years, covering 1999-2001 years. For both Spanish sites cloud cover data are recorded at 0700, 1300 and 1800 hours local clock time. The Instituto National De Meteorologia based in Madrid provided these data.

Data from three Indian sites were also used in the present work. All of these datasets were provided by the Indian Meteorological Department based in Pune. Data from Indian sites - Chennai (13.0N; 80.18E), Mumbai (19.12N;

72.85E) and Pune (18.53N; 73.85E) spanned five years covering 1990 to 1994. As with the Spanish sites, the Indian cloud cover data were also not recorded hourly. Rather this information was noted at 0830, 1130, 1430 and 1730 hours local time.

For both the Indian and Spanish sites, the cloud cover was observed and recorded by trained personnel.

For validation purposes, three additional UK Meteorological Office sites were used, Camborne (50.22N; 5.32W), Edinburgh (55.95N; 3.35W) and London (51.23N; 0.46W). Camborne and London datasets are one year long, 1999 and 1993, respectively. While the Edinburgh dataset is five year long, spanning 1990 to 1995. Excluding Camborne, these datasets include the same parameters and follow the same format as the Aldergrove and Bracknell datasets. Camborne on the other hand did not measure diffuse and beam irradiance, therefore only global horizontal irradiance was available for validation purposes.

The hourly irradiance data were quality controlled using the Younes et al. (2005) algorithms, also demonstrated in Chapter 3. The mathematical envelope of acceptance can be observed in Figure 5.1, for the Aldergrove dataset.

5.3 Cloud Radiation Models under review

As seen in Chapter 2, Kasten and Czeplak (1980) formulated Eqs.2.25- 2.27 for the estimation of solar irradiance based on cloud cover information.

$$I_{GC} = 910.\sin\alpha - 30 \tag{2.25}$$

$$I_G = I_{GC}(1 - 0.75(N/8)^{3.4}) \tag{2.26}$$

$$I_D = I_G(0.3 + 0.7(N/8)^2) \tag{2.27}$$

Gul et al. (1998), and Muneer and Gul (2000) furthered the work of Kasten and Czeplak (1980) to provide equations that can accommodate local coefficients for their datasets, as the original coefficients could not accurately estimate the irradiance in their analysis.

$$I_{GC} = A_1 \sin\alpha - B_1 \qquad (2.28)$$

$$I_G = I_{GC}(1 - C_1 (N/8)^{D_1}) \qquad (2.29)$$

To calculate the diffuse horizontal irradiance, they used the same formulation of Kasten and Czeplak (1980), Eq.2.27.

Lam and Li (1998) have explored the incorporation of multiple linear regressions between global irradiance and cloud cover involving solar altitude. The equations are given below,

$$I_G = 217 - 485(N/8) + 696\sin\alpha \qquad (2.30)$$

$$I_D = 30.5 - 62.9(N/8) + 294.7\sin\alpha \qquad (2.31)$$

The coefficients given in Eqs. 2.30 and 2.31 are for use for Hong Kong datasets only. Therefore to accommodate for local coefficients for the datasets used in the analysis, the Lam and Li equations were modified as follows,

$$I_G = A_2 - B_2(N/8) + C_2 \sin\alpha \qquad (5.1)$$

$$I_D = D_2 - E_2(N/8) + F_2 \sin\alpha \qquad (5.2)$$

Note that for all the above-mentioned models, the beam horizontal irradiance is calculated by subtracting the diffuse component of the global horizontal irradiance.

To facilitate the model comparison, each model has been renamed. M1, M2 and M3 are the Kasten and Czeplak model, Gul et al. model and Lam and Li model, respectively. While, M4 is the modified Lam and Li model.

5.4 Proposed modifications to the Kasten-Czeplak based formulae

Based on the work from Kasten and Czeplak (1980), Gul et al. (1998), and Muneer and Gul (2000) three new models are presently proposed. The new models have in common with the Gul et al. (1998) model, the equation to calculate the clear sky global horizontal radiation. The global and diffuse horizontal irradiation, are estimated using empirical equations based on constant, linear and quadratic power functions, referred to as M1, M2 and M3 respectively. The equations are as follows:

M5:

$$I_G = I_{GC}(a_0 + a_1\varphi + a_2\varphi^2)^{b_0} \quad (5.3)$$

$$I_D = I_G(c_0 + c_1\varphi + c_2\varphi^2)^{d_0} \quad (5.4)$$

M6:

$$I_G = I_{GC}(a_0 + a_1\varphi + a_2\varphi^2)^{(b_0+b_1\varphi)} \quad (5.5)$$

$$I_D = I_G(c_0 + c_1\varphi + c_2\varphi^2)^{(d_0+d_1\varphi)} \quad (5.6)$$

And M7:

$$I_G = I_{GC}(a_0 + a_1\varphi + a_2\varphi^2)^{(b_0+b_1\varphi+b_2\varphi^2)} \quad (5.7)$$

$$I_D = I_G(c_0 + c_1\varphi + c_2\varphi^2)^{(d_0+d_1\varphi+d_2\varphi^2)} \quad (5.8)$$

Note that φ=N/8; with N the cloud cover in oktas.

Initially the seven models were compared for their accuracy using the Aldergrove dataset. For the Kasten and Czeplak model and the Lam and Li model, the dataset was processed using the authors' original coefficients. However for the other five models, the coefficients were fitted against the local dataset for Aldergrove. The regressed Aldergrove coefficients were:

M2: A_1=1046.5, B_1=81, C_1=0.73, D_1=3.17, E_1=0.28, F_1=0.7, G_1=2.34

M4: A_2=210.7, B_2=-346.3, C_2=630.5, D_2=9.1, E_2=16.1, F_2=376.4

M5: a_0=1.001, a_1= -0.012, a_2=0.021,
b_0= -102.567,
c_0=70.870, c_1= -154.435, c_2=86.117,
d_0= -0.321

M6: a_0=1.01, a_1= -0.04, a_2=0.04,
b_0= -10.804, b_1= -102.911
c_0=0.867, c_1= -0.175, c_2= 0.189,
d_0= 8.709, d_1= -8.244

M7: a_0=5.926, a_1= -13.428, a_2=10.189
b_0= -0.001, b_1=0.410, b_2= -1.730
c_0=277.173, c_1= -574.280, c_2= 298.360
d_0= -0.252, d_1=0.137, d_2= -0.068

Figures 5.2 to 5.8 enable a visual evaluation of all the seven models. One may observe that the M2 and M4 perform better than the models on which they are based. This indicates that using local coefficients significantly improves the performance of the models. It was also noted that the Kasten and Czeplak based models performed better than the Lam and Li based models (both original and modified adaptations).

However, it is difficult to compare the accuracy gains of M2 against those of M5, M6 and M7 from the scatter plots alone, as this is only a qualitative method. Therefore, to quantitatively evaluate the performance of all models under discussion a multiple set of statistical indicators were employed. These indicators were, the slope and coefficient and determination, R^2, of the best-fit line between the calculated and measured global and diffuse irradiance. Furthermore, the mean bias error, MBE, and root mean square error, RMSE along with kurtosis and skewness of the error histograms related to the estimation of global and diffuse irradiation were used. The results for all seven models for Aldergrove are shown in Table 5.1.

The statistical indicators presented in Table 5.1 show that M2 and M7 perform better than the other five models, and that they perform with almost equal effectiveness. The two models, M2 and M7 yielded the same slope, R^2, MBE and RMSE. Comparison between the measured and estimated radiation for the presented models yielded low MBEs. However, MBE on its own does not give an insight into the performance of the individual models as stated by Gueymard (1999). Negative values imply an overestimation of the solar radiation, and vice-versa for positive values of MBE.

The two models, M2 and M7, clearly reduced the scatter as indicated by lower RMSEs. This is true for both I_G and I_D. To more closely examine the scatter in the estimation of I_G and I_D, error histograms have been generated and these are shown in Figs. 5.9 and 5.10 for I_G and I_D, respectively.

5.5 Accuracy scoring system

The Accuracy Score procedure, referred as AS, was developed with the aim to facilitate a discrete comparison between all the models. Numerically, it is expressed as:

$$AS = \frac{R_i^2}{R_{i,\max}^2} + [1 - \frac{|MBE|_i}{|MBE|_{i,\max}}] + [1 - \frac{RMSE_i}{RMSE_{i,\max}}] + [1 - \frac{|Skewness|_i}{|Skewness|_{i,\max}}]$$

$$+ \frac{|Kurtosis|_i}{|Kurtosis|_{i,\max}} + [1 - \frac{\Delta Slope_i}{\Delta Slope_{i,\max}}] \qquad (5.9)$$

$$\Delta Slope_i = |1 - Slope_i| \tag{5.10}$$

$$\Delta Slope_{i,\max} = |1 - Slope_{i,\max}| \tag{5.11}$$

$Slope_i$ is the slope of the best-fit line between the computed and measured irradiance component. The subscript $_{i,\ max}$ indicates the largest value of the given parameter for all of the models. AS is a convenient accuracy index, by means of which it is possible to compare the performance of any suite of models. Note that for the above mix of statistical parameters it is not necessary that a clear 'best' model may emerge owing to the fact that a gain in accuracy indicated by, say a low RMSE may be offset by a high value of kurtosis. The AS scoring system is thus of good value when such a contradictory picture emerges. The model that yields the highest values of R^2 and kurtosis would have the highest score. For MBE, RMSE and skewness, the model that yields the values closest to 0 would score the highest. Similarly, for the slope of the best-fit line, the model that yields a score closest to 1 would score the highest. The maximum obtainable score per statistical indicator is 1, therefore for any given solar irradiance component; a model would have a maximum obtainable score of 6. Since the models, described in Section 5, estimate all three components of the horizontal solar irradiance, I_G, I_B and I_D then the maximum obtainable score for any model is 18. Note that this scoring procedure gives the same weight to each of the statistical indicators.

However obtaining the maximum score does not necessarily imply that the model is accurate and best performing; it only indicates that the model yields better results compared to the other models in the evaluation. It is therefore essential to take into consideration the statistical indicators.

The statistical comparison method will be used in this chapter as well as in Chapter 6 in clear-sky modeling.

5.6 Results of modifications

Based on the AS results, M7 is the superior model for estimating I_G while M5 is superior for estimating I_D as can be observed from Table 5.2. However by summing up the AS scores for both I_G and I_D, as can be seen in Table 5.2, M7 obtains the highest score, followed by M2.

To further validate the models, the same procedure shown above for Aldergrove was performed on the remaining six datasets: Bracknell, Chennai, Gerona, Madrid, Mumbai and Pune. However in this latter analysis, only M1, M2 and M7 were compared, as they clearly performed better than the other models, as noted for Aldergrove. M7 yielded the closest result to M2 between the newly proposed models, M5, M6 and M7, and thus M7 is deemed more accurate than M5 and M6. Each of the above sites had its own specific local coefficients tuned for the M2 and M7 models, while for M1 Kasten and Czeplak original coefficients were used.

The models, M2 and M7 were in turn fitted with local coefficients for each location. Once again, with the coarse and fine datasets, M2 and M7 outperform M1. This is easily observed visually from the measured versus calculated global and diffuse irradiance scatter plots respectively shown in Figures. 5.11 a-c and Figures. 5.12 a-c for the Gerona dataset.

Tables 5.3 and 5.4 show the statistical indicators obtained for each of the six locations for M2 and M7. There is a slight reduction in MBEs for M7 compared to M2 for most sites, and the RMSEs are also reduced. The greatest reduction in RMSE was observed for Madrid I_D estimation, the reduction was 23W/m^2 and the smallest change was noted for Bracknell. The average reduction in RMSE in the estimation of I_G is in the range of 2-3W/m^2.

It is noted that although the results for I_G for both M2 and M7 models have very similar performance, in the estimation of I_D, M7 performs slightly better than M2. This observation is also observed by looking at the error histograms for both I_G and I_D, in Figures 5.13 and Figures 5.14 respectively for comparison between both M2 and M7 models. The number of occurrences of errors in the central bands of the estimation error histograms is relatively the same between the two models for each site.

5.7 Discussion of results due to modifications

Barker (1992) states that estimation of total cloud cover by real observations is subject to perspective errors and this causes inherent errors in the available datasets and is partly the reason for higher MBEs related to the estimation of I_G and I_D. Harrison and Coombes (1986) noted that the weather observer generally overestimates clouds. In this regard Myers (2005) remarks that the estimation models can only be proven as good as the data. In respect of his studies with uncertainties and error in solar radiation measurement instruments, Myers (2005) states that absolute measurement uncertainties are of the order of 25-100 W/m^2 in hemispherical measured data. Coulson (1975) classifies pyrheliometers and pyranometers in categories based on the uncertainties and errors of the measurements. A pyrheliometer is deemed 1st class if the measurement errors compared to a reference pyrheliometer are in the range of ±4% while it is deemed 2nd class if the errors are in the range of ±8%. Similarly for pyranometers, they are deemed 1st class if the measurement errors compared to a reference pyranometer are in the range of ±10%, 2nd class in the range of ±25% and 3rd class in the range of ±32%.

The percentage errors between measured and calculated I_G and I_D using the M2 and M7 models have been calculated, and the data has been split into error bands. The first band contains data in the ±10% range, while the second band included data in the ±25% range. Most sites had more than 20% of their data points in the ±10%, in the case of Mumbai 49% of I_G estimation errors are in the 1st band. Similarly for most sites, more than 50% of their data points were in the 2nd band, i.e. the ±25% range. For Mumbai, 77% of I_G estimation errors are in the 2nd band. Table 5.5 provides the above information in further detail for all sites under discussion.

Brinsfield et al (1984) remark that an observer has a general tendency to underestimate the cloud cover under low overcast conditions and overestimate the cloud cover during high overcast conditions. With respect to this observation and referring to the work conducted by Muneer and Gul (2000), the estimation of I_G and I_D datasets were divided in k_t bands to observe the patterns of estimation under overcast-, part overcast- and clear-skies. The data was divided in three k_t bands of $k_t \leq 0.2$, $0.2<k_t<0.8$ and $k_t \geq 0.8$ to respectively represent the above sky conditions. MBE and RMSE were obtained to analyze the performance of the models. The results of this statistical evaluation are given in Table 5.6. Both M2 and M7 models performed poorly under overcast skies. This is due to the fact that k_t attains a value of 0.2 under overcast conditions, irrespective of the type of cloud. The models performed satisfactorily under the remainder of the sky conditions.

A similar conclusion was reached by dividing the data into the above three sky conditions using the cloud cover information. The three categories that were chosen in this case were $N \leq 1$, $1<N<7$ and $N \geq 7$, to represent the above mentioned sky conditions. Once again, both models performed badly under overcast-sky. This is due to the fact that once the cloud cover attains a value of 8 oktas the model is unable to differentiate between thin and heavy overcast. However, there will be sliding scale of radiation receipt as the sky condition changes from thin to a heavy overcast. The results of this latter, cloud cover based analysis, is given in Table 5.7.

Bennett (1969) states that cloud cover explains less than 50% of insolation variance while sunshine fraction for example explains between 70-85% of the insolation variance. The reason for the weakness of cloud cover is that it does not take into consideration the type and depth of the cloud. Cloud type varies immensely the amount of scattering and shading of the terrestrial solar radiation.

5.8 Further examination of the cloud cover databases

It is possible to identify weaknesses in the Kasten and Czeplak (1980) and Gul et al. (1998) models. Any errors in estimation in the first calculated component, I_{GC}, are transmitted to the second component calculated, I_G, and thereafter to I_D and I_B. This is due to the dependence of the component algorithms to each other. Therefore, if the user seeks I_B and/or I_D, then the errors found in the estimation are to be greater than for a user who seeks the estimation of I_G. Lam and Li (1998) have identified this problem and therefore proposed two independent algorithms to estimate I_G and I_D. However, validation of models revealed flaws in them. The Lam and Li models often result in negative estimation values. Moreover, the dependency problem applies in the estimation of I_B when using the Lam and Li derived models.

Cloud information is often used as an indicator of the sky conditions. Skies with no cloud are represented with a cloud cover index of 0 Oktas, and skies that are fully covered by cloud are therefore represented by a cloud cover index of 8 Oktas. These represent the clear and overcast skies respectively. Sky conditions in between 1 and 7 Oktas are considered as mixed sky types.

A good indication of the sky conditions can be obtained by analyzing cloud cover datasets. This is done by plotting a cumulative frequency of occurrence diagram. One such diagram is shown in Fig. 5.15 for Madrid. This plot provides an indication on the typical type of the local skies. As an example, we can conclude from Fig. 5.15 that up 29% of the year, the skies above Madrid are not covered by clouds (0 Oktas), while only for 3% of the year the sky is heavily covered with cloud (8 Oktas).

A relationship between the solar altitude, the cloud cover and the horizontal components of solar irradiance has often been used in the formulation of solar irradiance models that are based on cloud cover, as can be seen in the models described in Section 5.3, Eqs. 2.25-2.31, 5.1 and 5.2. The relationship could be visually observed in Fig. 5.16 a, b and c, which presents a three-dimensional scatter plot of cloud cover (N), sine of solar altitude and a horizontal solar irradiance component for all sky conditions, then clear to mixed skies, and finally mixed to overcast skies.

It was therefore concluded that the solar irradiance components are dependent on the cloud cover and sin α,

$$I_i = f(\sin\alpha).f(N) \tag{5.12}$$

It was observed that there is a linear regression between the solar irradiance components and the sin α for each cloud cover band. This can be clearly observed from Fig. 5.16. We thus conclude that Eq. 5.12 may be expanded,

$$f(\sin\alpha) = A\sin\alpha + B \tag{5.13}$$

For clear to mixed skies (0-5 Oktas), in Figure 5.16b, linearity can be observed between cloud cover and global horizontal irradiance. However, for more overcast skies (6-8 Oktas), in Figure 5.16c, there is no evidence of linearity. From cloud analysis, it was found that for the first mentioned cloud band, a linear formula could result in coefficients of determination greater than 0.9, while for the latter mentioned cloud band the linear formula yielded R^2 of less than 0.7. For all band a linear formula as presented in the work of Lam-Li yielded an R^2 lower than 0.8 for all the datasets used in this study. This low coefficient of determination explains certain weaknesses in the models by Lam-Li as will be mentioned in latter sections. No linearity was found when analyzing for beam or diffuse horizontal irradiances.

It was found that by applying a non-linear regression for the data, the R^2 for all sky conditions is greater than 0.9. Thus to complete Eq. 5.13, Eq. 5.14 is as follows:

$$f(N) = C + D(N/8)^E \tag{5.14}$$

Similar results were found when analyzing for the diffuse and beam horizontal irradiances.

5.9 Proposed multivariate models

Two new empirical models are proposed to address the findings of the cloud cover analysis. The first proposed model is simplistic and calculates the global and beam horizontal irradiance based on the formulations given in Eq. 5.12.

$$I_G = (A_3 \sin\alpha - B_3).(C_3 + D_3(N/8)^{E_3}) \quad (5.15)$$

$$I_B = (F_3 \sin\alpha - G_3).(H_3 + I_3(N/8)^{J_3}) \quad (5.16)$$

$$I_D = I_G - I_B \quad (5.17)$$

As observed in Figure 5.16a, the cloud cover function is linear for cloud cover values under five Oktas for I_G. It was observed that there was no apparent linearity for I_B and I_D. Therefore a hybrid bi-variate model has been developed. For cloud cover above five Oktas, the model is the same as the previously proposed model. It is mathematically represented the same way as the previously proposed model in Eqs. 5.15-5.17, however in Eq. 5.15 the coefficient E_3 is equal to unity for $N \leq 5$, for higher Oktas, it remains variable.

5.10 Discussion of results and validation of multivariate models

Similarly to the analysis of results in Section 5.6, a quantitative and qualitative procedure was used to examine the models. The same statistical indicators were used to obtain the accuracy score to compare the models. It is important to note that obtaining the maximum score does not necessarily imply that the model is accurate and best performing; it only indicates that the model yields better results compared to the other models in the evaluation. It is therefore essential to take into consideration the statistical indicators. The range of the statistical indicators can be observed in Table 5.8 for Aldergrove.

For all the sites that were processed in this frame of work, it was found that the models that allow the use of locally tuned coefficients yielded better results than the models that use the original coefficients supplied by the authors. This was also observed visually by inspecting both the scatter plots and the error histograms in Figures 5.17 and 5.18, respectively. The statistical indicators and hence the AS indicate the superiority of the local coefficient variants of all models.

The Lam-Li models, original and modified, have a specific weakness; the algorithms used in the models cannot estimate the horizontal solar irradiance components of the available datasets to an acceptable degree of accuracy. It was noted however that the modified variant does accurately estimate the global and diffuse horizontal irradiance for the clear-sky data. However, in all other sky conditions the result was found to be unsatisfactory. In all cases, the Lam-Li models yielded negative estimations, thus making the estimated data obsolete since the irradiance cannot be negative.

It was noted that the Kasten-Czeplak model performed exceptionally well in estimating I_G and I_D for UK datasets. This is due to the fact that Kasten and Czeplak had created then validated the model using UK and German datasets.

The Katen-Czeplak, Muneer-Gul and the proposed models all estimated I_G and I_D with good accuracy; however, the proposed models were superior in the estimation of I_B. It was found that for the proposed models having separate equations to estimate I_D yielded little improvement on the procedure described in Section 5.9. Therefore, I_D could simply be extracted by estimating I_G and I_B.

Based on the AS results for the estimation of all three components of the horizontal solar irradiance, It was found that the proposed models performed better than the older models. This can be observed in AS results for Bracknell, presented in Table 5.9.

The second proposed model being a more complex version of the first proposed model does yield improvements, however due to the increase in the number of coefficients used; the improvement is of questionable value.

For all the datasets that were utilized in this paper, the first proposed model is the most accurate of the models compared. This is mainly due to the independence of the horizontal solar irradiance components. The proposed model shares the same basic structure with the Kasten-Czeplak variant models for the estimation of I_G, therefore no particular improvement was observed, however it was observed that the proposed model estimated I_B and hence I_D with greater accuracy than the other models in this study. This can be observed visually in Figure 5.19 a-i, scatter plots of measured versus calculated I_G, I_D and I_B for the Muneer-Gul, modified Lam-Li and the proposed models. Similarly, the improvement can also be observed in Figure 5.20 a-i, I_G, I_D and I_B estimation error histograms for the same three models. Both Figures 5.19 and 5.20 are for the Madrid data. For the proposed model, the coefficients used for each of the datasets that was used in this research are given in Table 5.10. It was observed that for Aldergrove, Bracknell, Gerona, Madrid and Mumbai the coefficients were very similar and had very little variations. However, it was noted that Chennai and Pune had different coefficients than the rest of the sites.

Further research was done to evaluate the use of regional coefficients compared to using local coefficients. Based on the coefficients for Aldergrove and Bracknell for the UK, the first proposed model in this study is validated for three other UK locations. These locations are Camborne, Edinburgh and London. The model, using the regional coefficients for the UK, performed as well as it would when using locally tuned coefficients. For the Edinburgh dataset, the MBE of I_G estimation for regional coefficients is -8.45 W/m^2 while it is of 4.45 W/m^2 when using local coefficients. There is also an increase in RMSE of 2.18 W/m^2. Similar results were found for the other validation datasets. The details of the results are found in Table 5.11. The generalized coefficients adopted are: A= 1046.44, B=75.24, C=1.03, D=-0.72, E=3.31, F=995.11, G=121.87, H=-0.83, I=1.73

5.11 Discussion on model scoring

When multiple models are compared, there is a specific need for quantitative and qualitative methods to compare them. Traditionally, peers have used a combination of statistical indicators to compare quantitatively the models under evaluation. The most common have been to use the mean bias error and root mean square error in either units or as a percentage.

It was found by other researchers that when multiple models are compared, and especially for a large pool of datasets, it was rather confusing to make an opinion on the efficacy of the models under evaluation. Therefore, Muneer (2004), developed a scoring procedure for his work on the evaluation of the meteorological radiation model, MRM, the Page radiation model, PRM, and the cloud radiation model by Gul, Muneer and Kambezidis, previously discussed in this chapter. His method was fairly simple and straightforward and is expressed mathematically as:

$$Score = |MBE| + RMSE \tag{5.18}$$

In this method, the lower the score, the better the model. This scoring method was used in Chapter 4 to evaluate the improved meteorological radiation model.

This model has evolved from then, into a similar Accuracy Score as the one described above in Section 5.5. This model was first used my Munawwar and Muneer (2005) and is described mathematically as follows:

$$AS = \frac{R_i^2}{R_{i,\max}^2} + [1 - \frac{|MBE|_i}{|MBE|_{i,\max}}] + [1 - \frac{RMSE_i}{RMSE_{i,\max}}]$$
$$+ [1 - \frac{|Skewness|_i}{|Skewness|_{i,\max}}] + \frac{|Kurtosis|_i}{|Kurtosis|_{i,\max}} \qquad (5.19)$$

Note that the Accuracy Score described in Eqs.5.9-5.11 is an evolution of this method described above, with the addition of the Slope as a statistical component.

From Personal communication, Dr. Avraham Kudish from Ben Gurion University, Beer Shiva, Israel, has come up with a scoring procedure as well. This new scoring method is expressed mathematically as:

$$Score = 0.5(1 - Slope.R^2) + 0.25(RMSE/100) + 0.25(MBE/100) \qquad (5.20)$$

To compare these scoring procedures, the Bracknell CRM evaluation was used. The Kasten and Czeplak's, as well as Lam and Li's original and modified models that were studied in this chapter, were compared to the best performing newly developed model, by means of using the four scoring method described above. The results of this analysis can be observed in Table 5.12 (a, d). It was found that each scoring method yielded different results. For example the original Kasten-Czeplak CRM was found to have ranked 3rd based on the current AS, 2nd based on Munawwar and Muneer AS, 4th based on the Muneer procedure, and 3rd based on Kudish's formula. The ranks of the other models are represented in Fig.5.19.

It is worth noting that all procedures were capable of identifying the weakest model, in addition all except the Kudish procedure, identified the proposed CRM as the most accurate.

The procedures, in which Muneer has been involved in developing, have shown that the proposed bi-variate model shows between 15- and 21% improvement on the Muneer et al. CRM. However, based on Kudish procedure the Muneer et al. CRM is still 3.2% more accurate than the proposed CRM. It is therefore important to note that the way the models are scored or ranked depends on the statistical indicators used. Note that some researchers have included standard deviations and student test in the mix of quantitative methods of evaluating the various models. More details can be found in Table 5.13.

5.12 The Napier University PV Facility

The previous chapters and sections have discussed a purely theoretical side of solar radiation - data manipulation and modelling. For practical uses of solar radiation, models could be used to estimate the resource available or direct measurements at the site.

Napier University's School of Engineering has been involved in education and research in renewable energy for the past 35 years. With the aim of demonstration of the viability of production of solar electricity at a high latitude location, such as Edinburgh (56 degree north), the School undertook to commission a medium-sized PV electricity generation project. The installation of 32 rows of BP Solar silicon panels covering a total nominal area of 160 square metres ensures generation of 13.7 kWp (kW peak) AC power. This value of peak power is the highest obtained to date.

DC power is produced from the BP Solar high efficiency mono-crystalline panels, each of which produces 90 W of power at 22 V, which is then converted by four inverters to a stable AC supply. The PV facility is fully instrumented with both input (incident solar energy) and PV electrical energy output, recorded at a frequency of 15 minutes.

The solar irradiation is locally measured by use of two pyranometers installed on the roof of the building. One pyranometer is installed horizontally and thus records the global horizontal irradiance and second pyranometer records the inclined irradiance at the same tilt and orientation as the photovoltaic facade. In addition to local measurement, global irradiance and cloud data was also obtained from local meteorological station. This allows for the comparison of the local measurement with long-term averaged data and cloud modelled data.

5.12.1 The measurement facility

For the purpose of this study, energy inputs and outputs were required for the analysis of the installation. To measure the incident irradiation, pyranometers were used. Photovoltaic DC and AC outputs were measured via data logging equipment installed within the inverter housing. Other synoptic information was obtained from nearby Meteorological Office measurement station at Turnhouse (Edinburgh Airport).

Two types of Kipp and Zonen pyranometers, CM11 and CM6b, were installed on the roof of the Merchiston Campus building belonging to Napier University. The CM11 unit is more precise, but is also more expensive than the CM6b unit. An error of ±10 W/m^2 at 1000 W/m^2 of incident radiation for CM11 is quoted against a ±20 W/m^2 error for CM6b. Both types of pyranometers have low response-time of 18 seconds at 95 % precision. This information was obtained by the manufacturer, Kipp & Zonen (2005).

The photovoltaic (PV) modules installed are BP-Solar BP790S units (2004). A PV module generates DC power from incident solar irradiance. The PV modules incorporate mono-crystalline silicone semiconductor technology. Metallic grids collect electrons from the semiconductor and generate a DC electrical current transferred via an electric circuit to the inverters, which convert the produced DC output into AC power. The system contains four Fronius inverters (2005). Two large (IG60) inverter are connected to four strings of 18 modules wired in series; each of these strings is wired in parallel with the others. Two small (IG20) inverters are connected to two strings of 12 modules wired in the same way as the IG60 inverters. Note that the input voltage rises quickly as soon as the PV modules are irradiated. The maximum input voltage is 500 Volts DC. DC current generated is proportional to the intensity of light falling on the modules.

Napier University's PV system is supplied with monitoring and data-logging equipment, which permit storage of the data for the whole system over long periods (up to 6 weeks at 15-minute frequency). The data can be viewed on a computer screen using "Fronius-IG Access" software. This program enables automatic long-term storage of measured data.

Meteorological data, when not measured at the experimental station, can be obtained from a local station belonging to the UK Meteorological Office network via the British Atmospheric Data Center website (2005). For the purpose of this study long term, inclined solar irradiation was required. In the absence of such measurements, a 27-year dataset of horizontal global and diffuse irradiation based on the local Meteorological Office records at Turnhouse was used to produce estimates of slope irradiation for the given PV aspect of 37° East of South and a tilt of 75° from horizontal irradiation using the work by Muneer (1990). In addition, the cloud cover data was obtained for the Edinburgh measurement station for the period of operation of the facility as well as the local measurement period. Edinburgh Airport measures total cloud cover up from 1976 to 2000, and solar global irradiance from 1976 to 2002. Beyond 2000, cloud cover at the lower four layers of the atmosphere are measured.

Figure 5.20 shows the long-term, hourly-averaged horizontal global and diffuse irradiation profile, which is based on the above dataset. It is interesting to note that the AM and PM irradiation trace is not symmetrical. A slightly higher transmission is observed for the afternoon period. This is a common observation in Edinburgh, particularly so during summer months when the morning sun gradually 'burns' any prevailing cloud-cover. It is therefore to be expected that solar energy collection modules that have aspects of west of south will gain more energy than those that are east of south.

Although the installation has been under operation since the 6th of April 2005, and the PV-AC output recorded, the solar power input and PV AC power production data were simultaneously recorded only over a five-month period, from beginning of June to end of October 2005. The latter information was used to obtain an estimate of the average PV conversion efficiency.

5.12.2 PV output calculations

Table 5.14 provides the calculated and measured cumulative energy quantities for the PV facility under discussion. An overall efficiency, $\eta_{PV,AC}$, of 11.5 % was obtained for the irradiation-AC power conversion. Note that the differences between the calculated and measured cumulative energy is due to seasonal variations that lead to differing frequencies of cloud-cover and hence the colour of the sky.

The AC electrical output, $E_{PV,AC}$, can be obtained from the tilted global irradiance I_{TLT}, the useful surface area of the PV array A_{PV}, (A_{PV} = 125.5 m^2) and the overall efficiency $\eta_{PV,AC}$ as,

$$E_{PV,AC} = I_{TLT} A_{PV} \eta_{PV,AC} \tag{5.21}$$

Using Eq.5.21 an analysis may be undertaken to obtain the optimum PV module aspect and tilt. For this purpose a local, 27-year solar radiation database with hourly records of global and diffuse components was employed.

The incident irradiance was calculated using Muneer (1995) slope irradiance model.

Muneer's model treats the shaded and sunlit surfaces separately and further distinguishes between overcast and non-overcast conditions of the sunlit surface. In this model, the slope diffuse irradiation for surfaces in shade and sunlit surfaces under overcast sky is computed as:

$$I_{D,TLT} = I_D[TF] \tag{5.22}$$

A sunlit surface under non-overcast sky is modelled as:

$$I_{D,TLT} = I_D[TF(1-F)+Fr_B] \tag{5.23}$$

TF is calculated via the following equation:

$$\frac{I_{DB}}{I_D} = \cos^2\frac{TLT}{2} + \frac{2b}{\pi(3+2b)}\left[\sin TLT - TLT\cos TLT - \pi\sin^2\frac{TLT}{2}\right] \tag{5.24}$$

For worldwide locations an average b=2.5 can be used, and for Northern Europe, F is calculated via the following equation:

$$2b\{\pi(3+2b)\}^{-1} = 0.0033 - 0.415F - 0.6987F^2 \tag{5.26}$$

To find the tilted global irradiance, $I_{TLT} = I_{D,TLT} + I_{B,TLT}$ (5.27)

$$I_{B,TLT} = I_B r_b \tag{5.28}$$

$$r_B = \max(0, \cos INC/\sin\alpha) \tag{5.29}$$

The above-mentioned model was used to calculate the incident irradiance on the PV facade. In addition to calculating theoretical inclined irradiance of the current installation, the following combinations of tilt and orientation were calculated in order to find the optimum parameters for an installation in Edinburgh, UK. The calculated orientations are as follows: East, South-East, South, South-West and West. For each of the orientations the followings tilts were used in degrees: 0° (vertical), 10°, 20°, 30°, 35°, 40°, 45°, 50°, 60°, 75° and 90° (horizontal). The results of the matrix of orientation and tilt was compared to the Napier University installation at 37° South-East and a tilt of 75°. It was however found that for Edinburgh, the optimum orientation and tilts are South, at 35°-40°. The annual-total energy yield was hence calculated on the combination of cardinal aspects and tilt.

It was found that for Edinburgh, the optimum aspect is South with the tilt range between 35°-40° producing the highest yield. The above yields are also compared to the Napier University installation with an aspect of 37° East of South and a tilt of 75°. There is also a considerable influence of the preponderance of morning clouds with an aspect of south-west producing much higher energy yield than the corresponding south-east tilts and likewise 'western' aspects performing better than their 'eastern' counterparts.

Due to architectural and planning restrictions, the only possible location to install the photovoltaic array was in its current location with an aspect of 37° East of South and a tilt of 75°. It was found that for the presently installed aspect and tilt, an average of 10.84 MWh AC output would be generated yearly.

5.12.3 Solar radiation estimation based on developed models.

For Edinburgh, the UK Met Office records the global irradiance and cloud cover at the Turnhouse Edinburgh Airport. Therefore of the models developed in this thesis, only the cloud based model from Chapter 5 can be used.

The chosen model estimates the global horizontal irradiance based on the solar altitude angle and the total cloud cover in Oktas. Diffuse horizontal irradiance is then obtained by the difference between the global horizontal irradiance and the direct horizontal irradiance, which is calculated similarly to the global horizontal irradiance.

A one-month dataset for June 2005 of layered cloud cover was compiled. For modelling purposes, total cloud amount was required. By analysing a year's data of Edinburgh, a simple method was devised to obtain total cloud amount from layered cloud amounts. For each hour, the highest reading of the four cloud layers is assumed representative of the total cloud amount. When compared to total cloud amount measurements for the same period, an accuracy of 94% was observed.

For the chosen period, the solar global horizontal irradiance is calculated using CRM based on the cloud data obtained from the UK meteorological Office. The same coefficients were used for this one-month dataset as the ones used for the Edinburgh five-year dataset. The average irradiation for that period was compared to long term monthly averages for June based on the 27-year dataset and from work by Muneer (2001). More details are given in Table 5.15. It is clear from the table that the month of June 2005 yielded less incident energy than the same month in 1999 and less than a typical June taken from average of 27-year data for Edinburgh.

Discrepancies between calculated and measured parameters in Tables 5.15 and 5.16 are due to a series of cumulative errors, although seasonal variability plays a major role to explain this discrepancy. Cloud cover information was needed to perform the irradiance calculations. It should be noted that solar irradiance is greatly affected by scattering caused by the various cloud types and their depth. The measurement station at Turnhouse Edinburgh airport is approximately 10 miles from the PV installation, the presence of minute differences, between the two locations, are due to microclimates and mesoclimates variations.

The cloud cover that could be recorded at the measurement station might not be that same if it were recorded at Merchiston campus; unfortunately, there is no method to quantify the associated error. Since the cloud information, is not measured as total cloud amount, but rather in cloud amount at different levels, the associated error in compiling the total cloud amount is 6% as described earlier. As was seen in Table 5.15, the cloud radiation model performs accurately, with low estimation errors, however, Myers (2005) remarks that the estimation models can only be proven as good as the data.

There are also estimation errors in the slope irradiation model. The cumulative errors are also transmitted to the calculated PV output. The measured irradiance values also present operational and equipment errors and uncertainties as described in Chapter 3.

Perez et al (2005), based on their study in the USA, report that ground measured cloud cover is generally over-estimated at airport locations compared to regular meteorological stations. This is often due to air traffic safety concerns that tend to make the cloud observation reports more conservative. This could explain the results found in Table 5.15 in which cloud based estimation of the June 2005 incident energy is less than long term averages and the June 1999 record. Note that this minute variation could also be explained by seasonal variation and more long-term comparisons between records and estimations should be attempted in the future to better assess the found discrepancies. However to undertake this work at present is not possible due to none availability of 2005-2006 cloud data.

5.13 Conclusions

Cloud cover based solar radiation models, have been developed and used as of the mid 20th century. In the early eighties, Kasten and Czeplak (1980) published their first model based on two parameters only, the solar altitude angle and the cloud cover in Oktas, to obtain the global and diffuse horizontal irradiance. Note that the original model was generalized by tuning it to a multitude of datasets from the UK and Germany.

Many models evolved from the Kasten and Czeplak model, by the work of Gul et al. (1998) and derived models are still used in research such as the work by Perez et al. (2005), as well as major work by Thevenard and Brunger (2001) for ASHRAE and for commercial purposes such as the work by Biggs (2005). Most of these modifications were done by fine tuning the coefficients of the original Kasten-Czeplak model for the locations were they are needed. A huge database of these regional and national coefficients is available in literature.

Lam and Li (1998) adopted a different approach for their model. They stuck with the same twin-parameter approach to cloud cover based modelling, by examining the correlation between solar horizontal irradiance, solar altitude angle and cloud cover. The model was published with local coefficients for the Hong Kong area. For the purpose of this study, the Lam and Li model has been modified to allow for local tuning since the Honk Kong coefficients are not representative of all climatologies.

By comparing these models, it was found that the Kasten-Czeplak based models performed better than the Lam-Li based models. In the same instance, the locally tuned models performed better than their original/generalized coefficient counterparts.

New models were proposed based on the Kasten and Czeplak approach, by means of modifying the equations themselves. Thus, the global and diffuse horizontal irradiation, are estimated using empirical equations based on constant, linear and quadratic power functions. The new models performed better than the Lam-Li based models, yet the improvement compared to the Gul et al. (1998) model was minute to compensate for the increase in the number of coefficients and complexity of the model.

Due to the large amount of models involved in this comparison and the multitude of statistical indicators used for this task, a comprehensive scoring procedure was adopted, the Accuracy Score, to streamline this task. However it was noted that different scoring methods yield different results. In fact in three out of the four scoring procedures evaluated, the newly proposed model was ranked best model. This is due on the weighting and the quantity of statistical indicators used.

More critical analysis of the correlation between the cloud cover and the solar irradiance was needed. Based on this analysis, a new model was developed, using the same two parameters as the previous models. The new model yielded higher accuracy than any of the previous models presented in this study. Yet this model still had a weakness, it requires local coefficients.

Further analysis of the coefficients, allowed for the obtaining of generalized coefficients. It was noted that the Indian sites had coefficients that differ greatly from the others, as was noted in previous chapters, this difference is either due to local weather patterns that differ from the other locations in the database used in the study or that the data provided is of low quality. The generalized coefficients were validated using three extra datasets from the UK, and it was noted that the drop in precision in the model when used with the generalized coefficients is minimal and acceptable compared to the locally tuned model.

Further work on the scoring procedure should be undertaken, and is the basis of future development and cooperation with peers for creating a standardised method of evaluating solar radiation models.

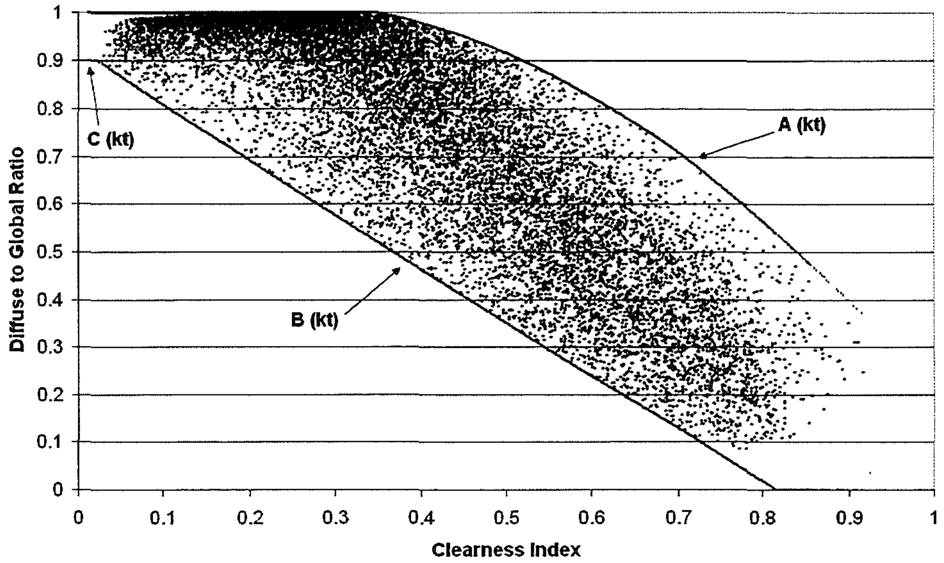


Figure 5.1 Quality Control boundaries in a k_t-k plot for Aldergrove.

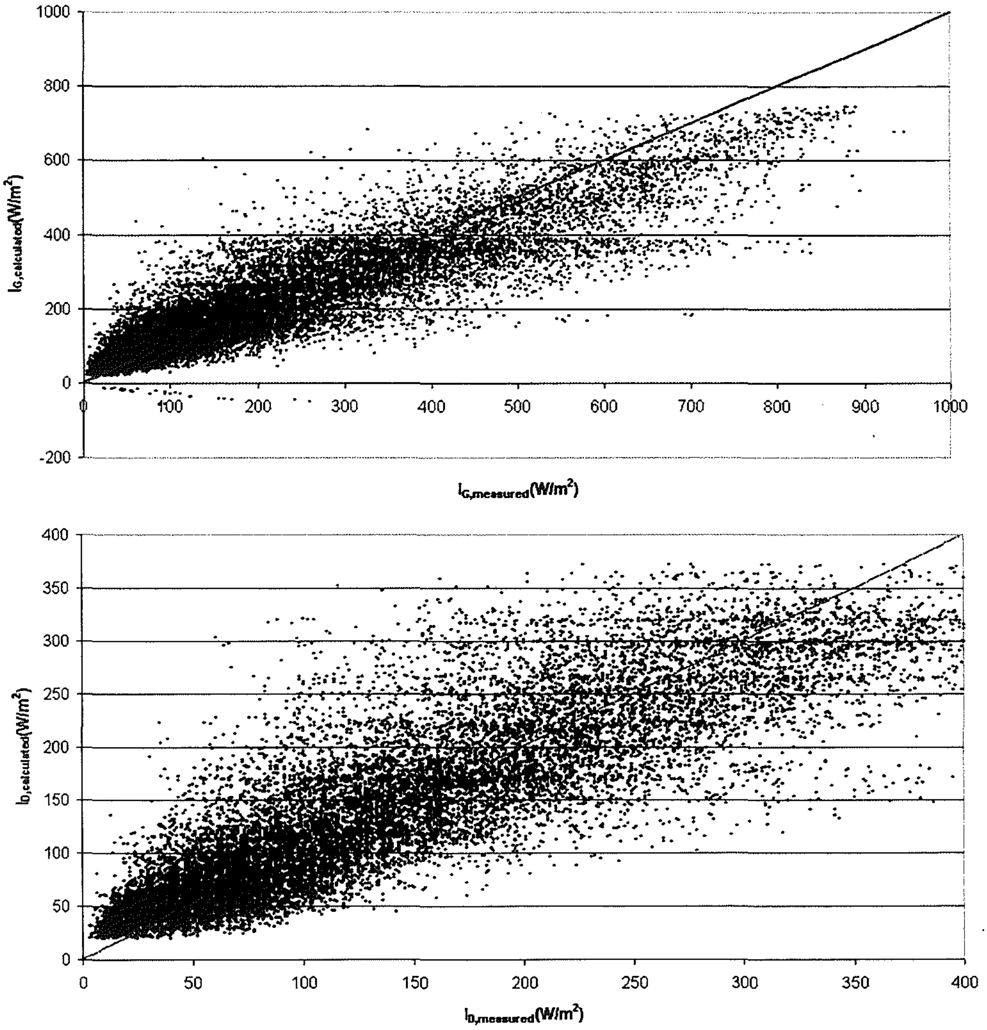


Figure 5.2 Scatter plot of measured versus calculated global and diffuse irradiance for Aldergrove using the Kasten and Czeplak, M1 model.

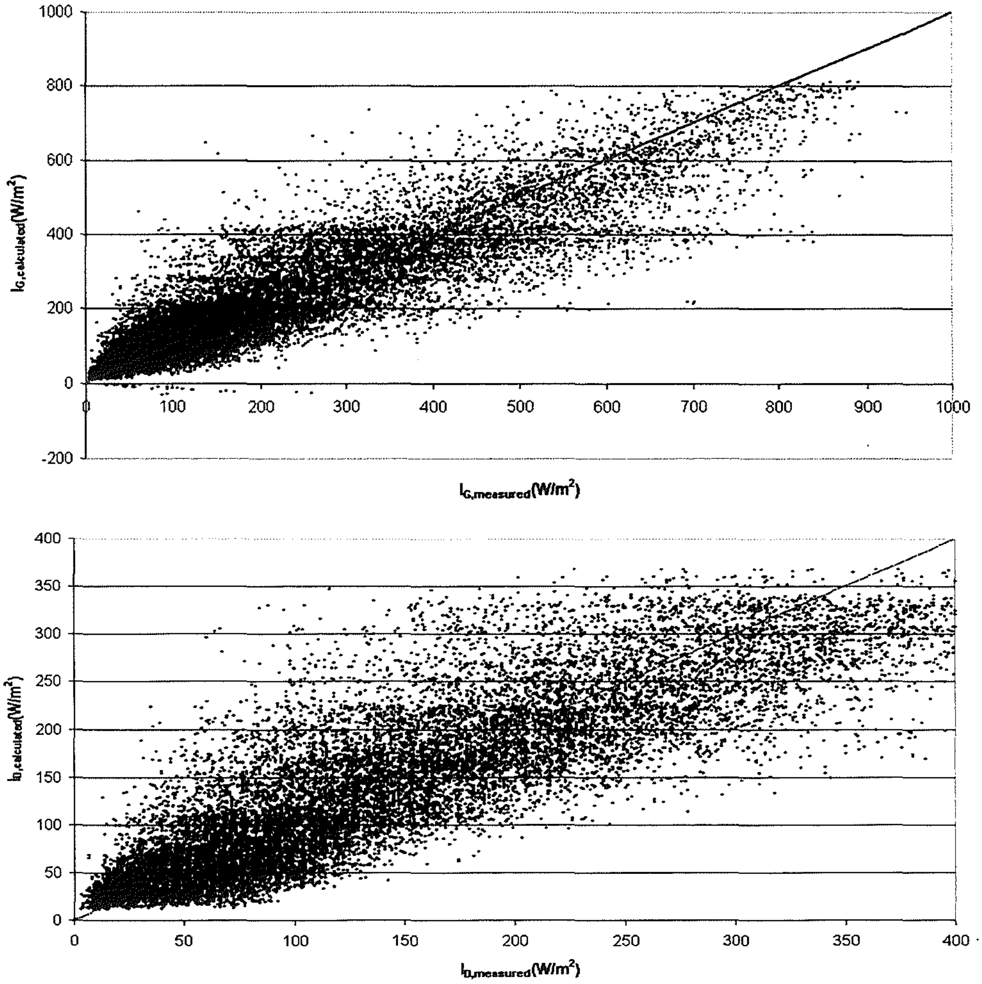


Figure 5.3 Scatter plot of measured versus calculated global and diffuse irradiance for Aldergrove using model M2.

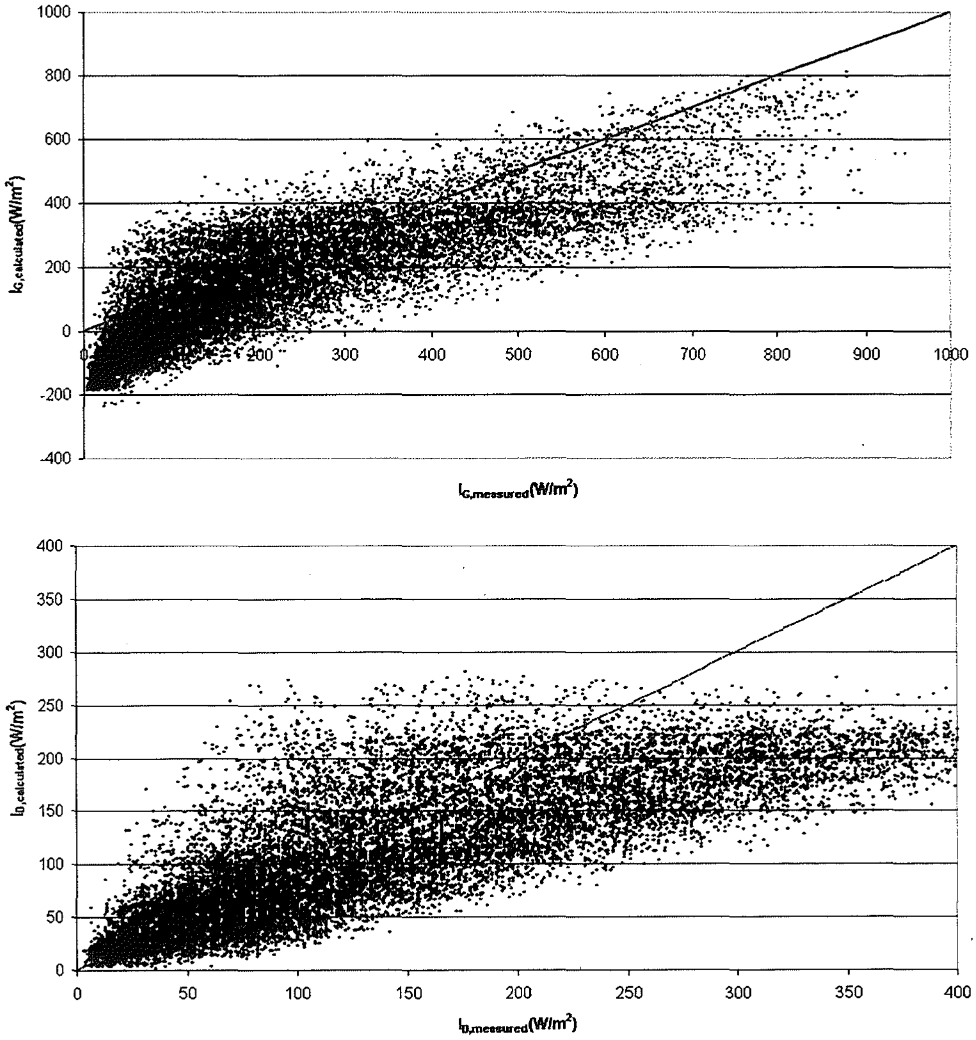


Figure 5.4 Scatter plot of measured versus calculated global and diffuse irradiance for Aldergrove using the Lam and Li, M3 model.

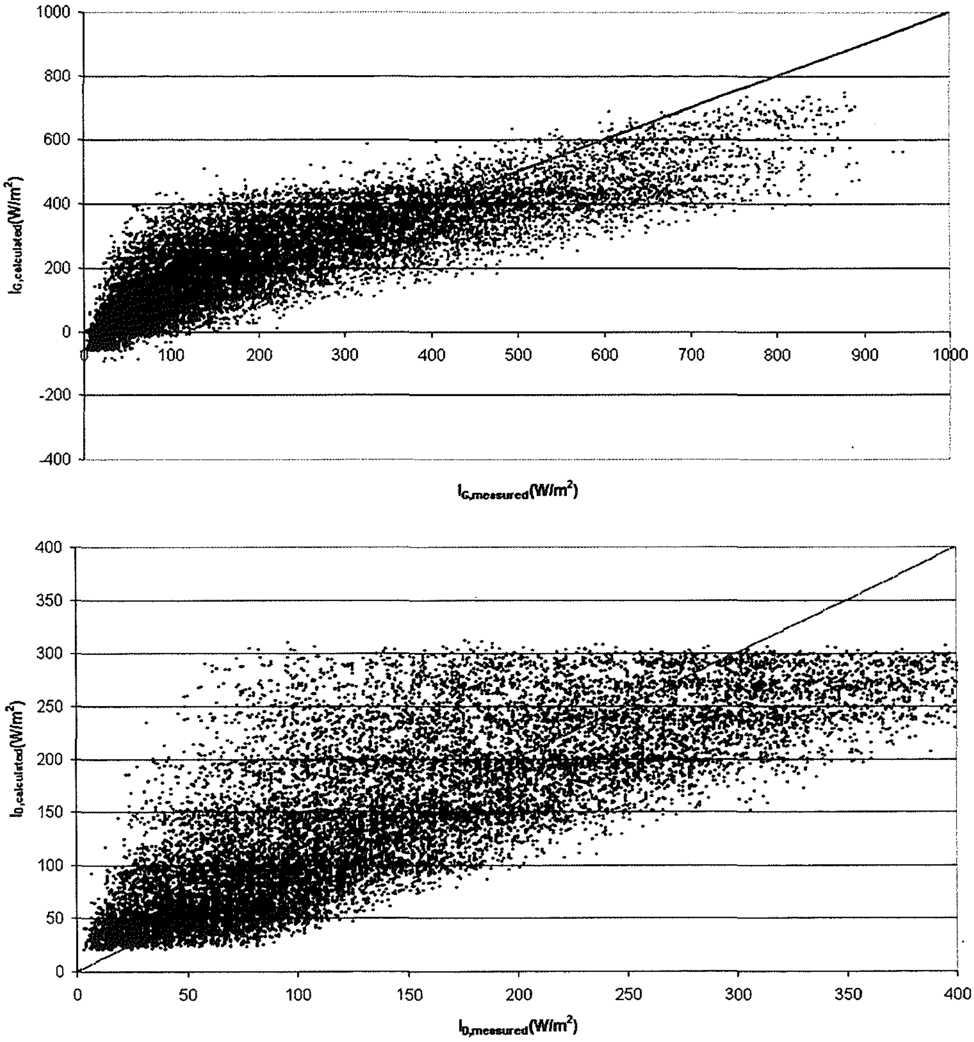


Figure 5.5 Scatter plot of measured versus calculated global and diffuse irradiance for Aldergrove using model M4.

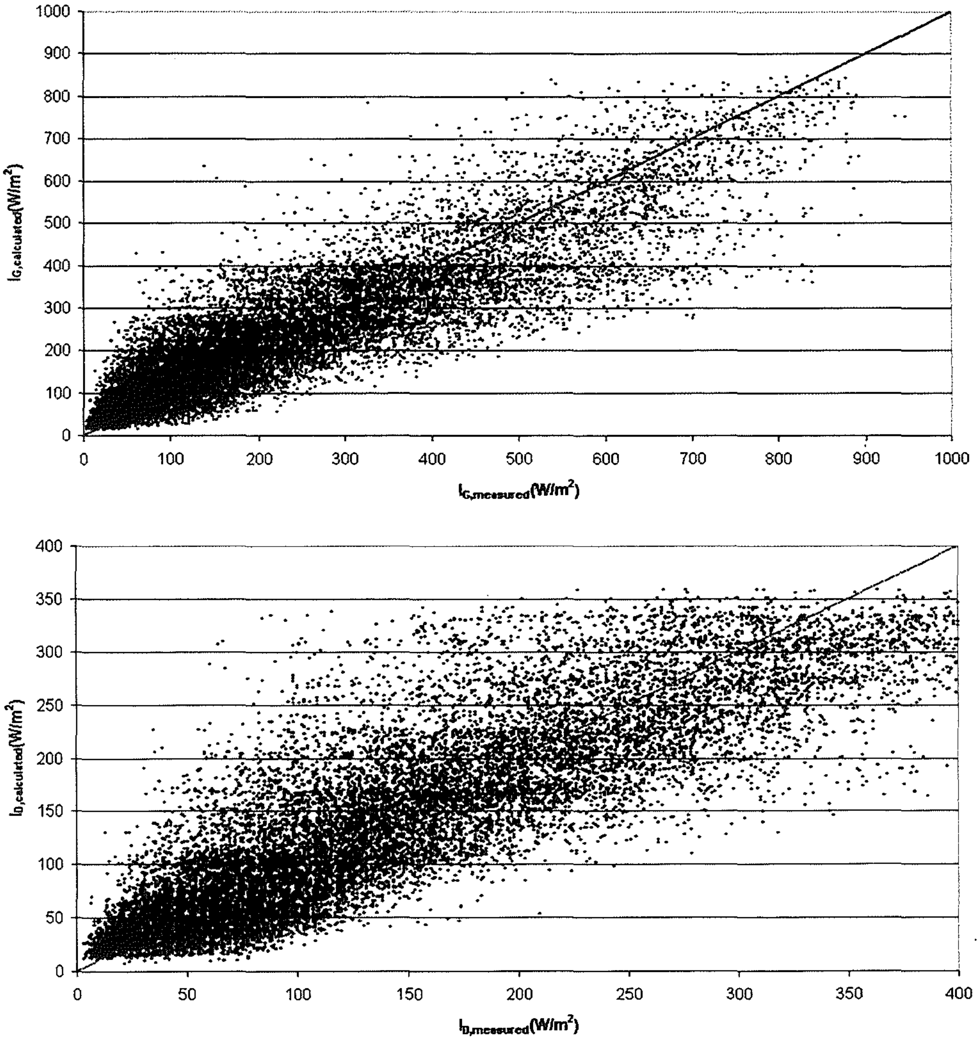


Figure 5.6 Scatter plot of measured versus calculated global and diffuse irradiance for Aldergrove using model M5.

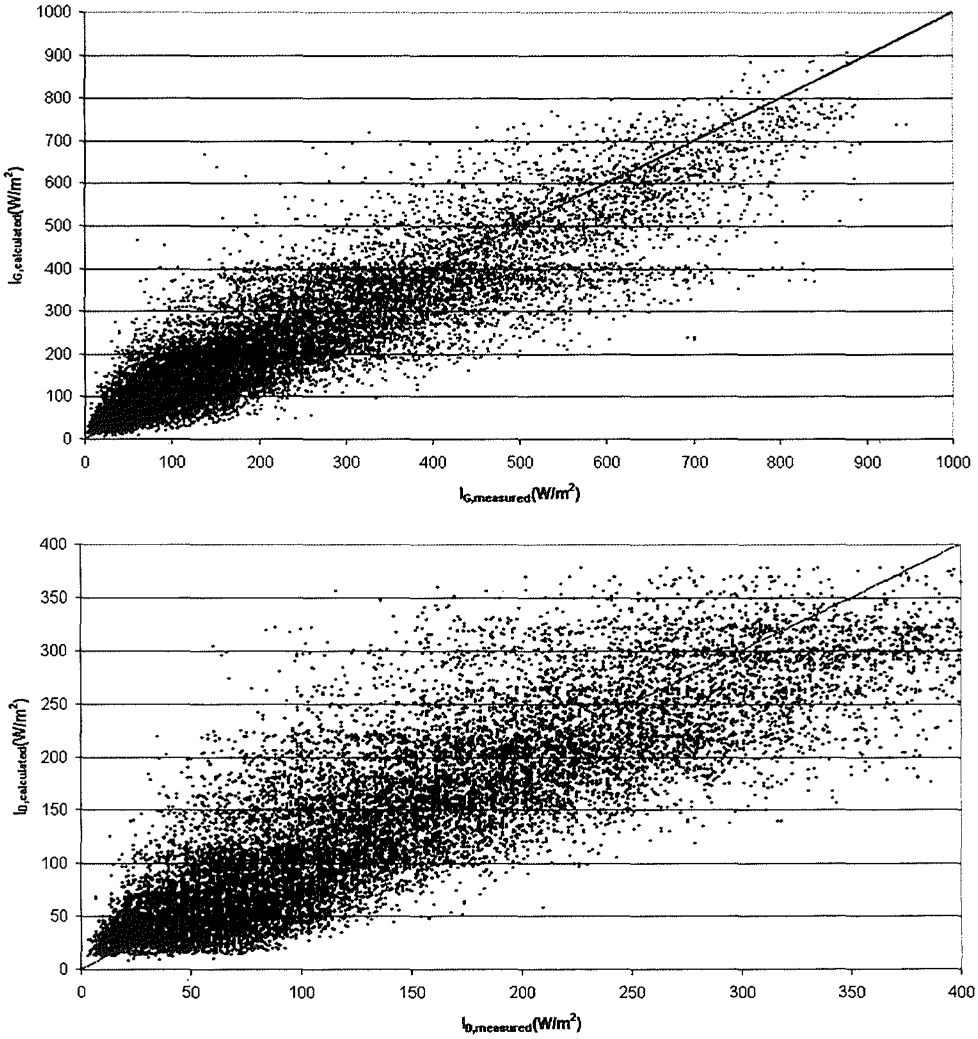


Figure 5.7 Scatter plot of measured versus calculated global and diffuse irradiance for Aldergrove using model M6.

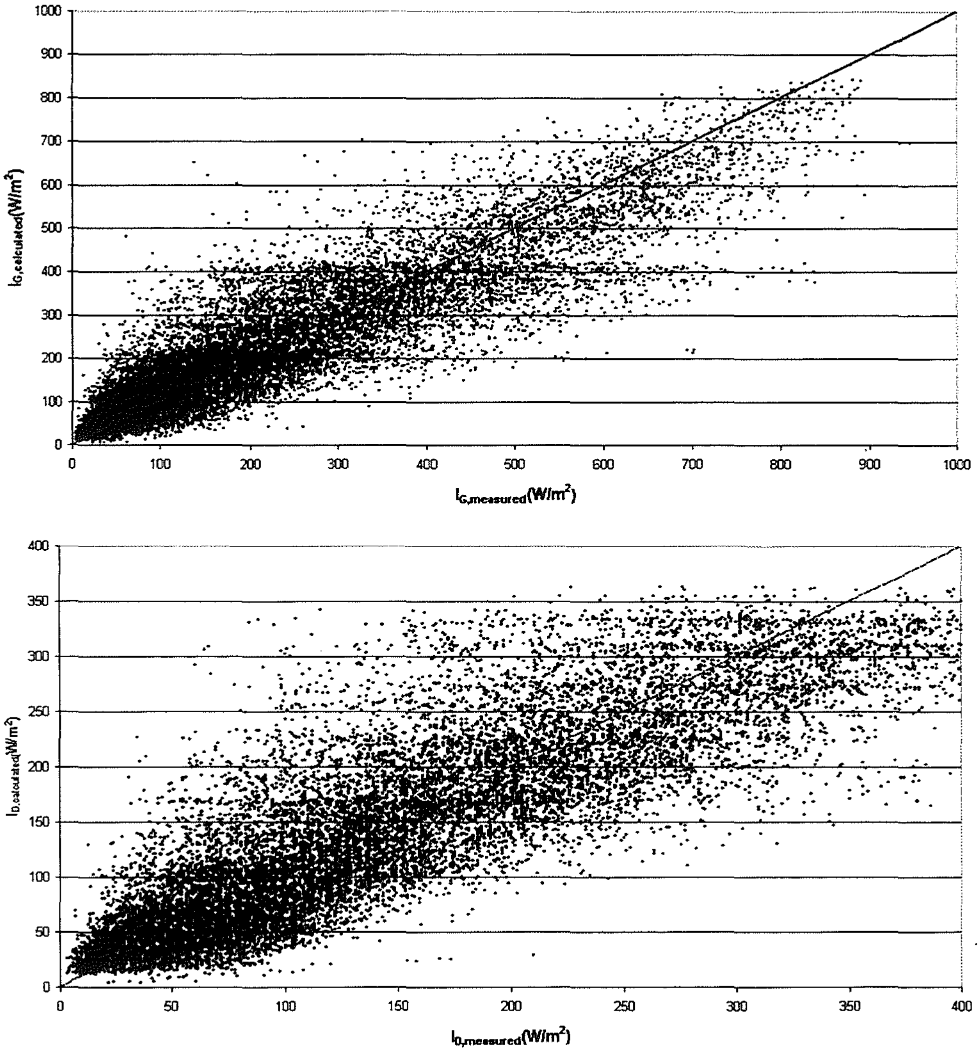


Figure 5.8 Scatter plot of measured versus calculated global and diffuse irradiance for Aldergrove using model M7.

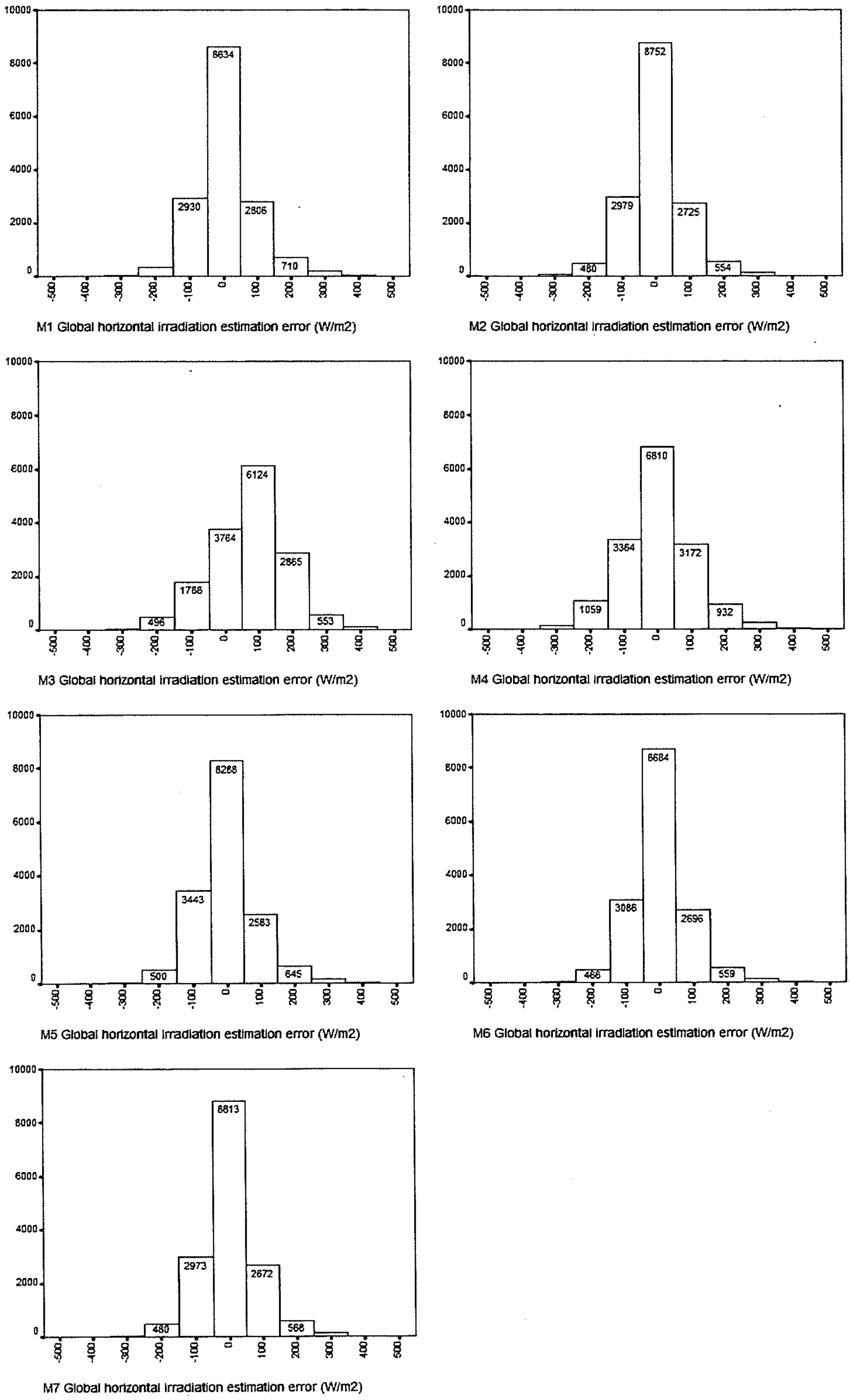
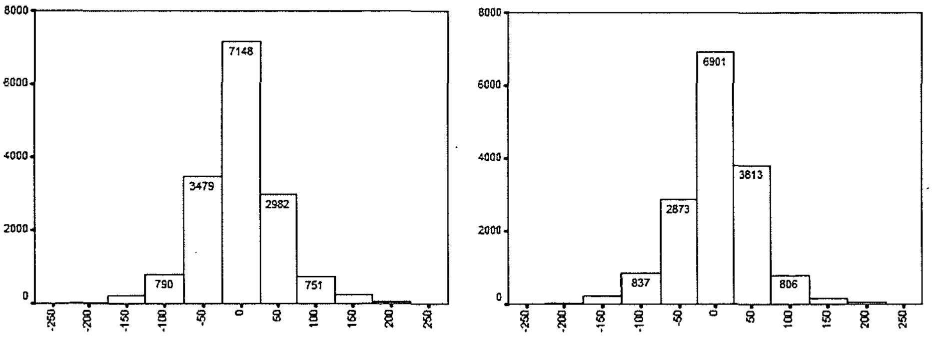
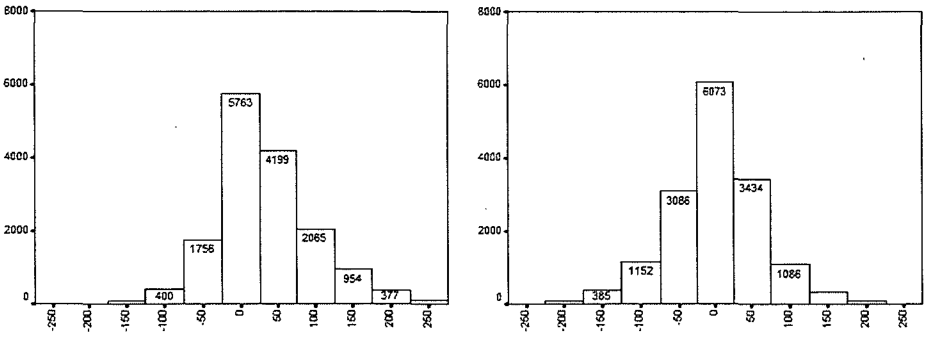


Figure 5.9 Estimation error histograms for global horizontal irradiance for Aldergrove using models M1 to M7.



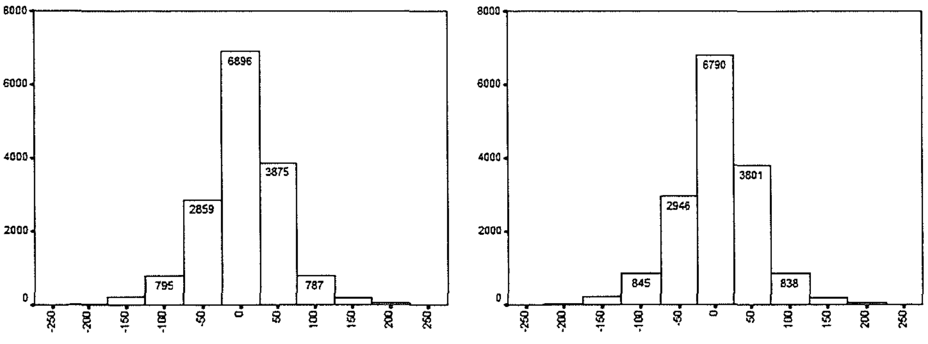
M1 Diffuse horizontal irradiation estimation error (W/m2)

M2 Diffuse horizontal irradiation estimation error (W/m2)



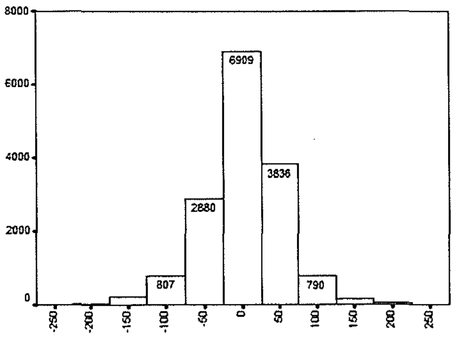
M3 Diffuse horizontal irradiation estimation error (W/m2)

M4 Diffuse horizontal irradiation estimation error (W/m2)



M5 Diffuse horizontal irradiation estimation error (W/m2)

M6 Diffuse horizontal irradiation estimation error (W/m2)



M7 Diffuse horizontal irradiation estimation error (W/m2)

Figure 5.10 Estimation error histograms for diffuse horizontal irradiance for Aldergrove using models M1 to M7.

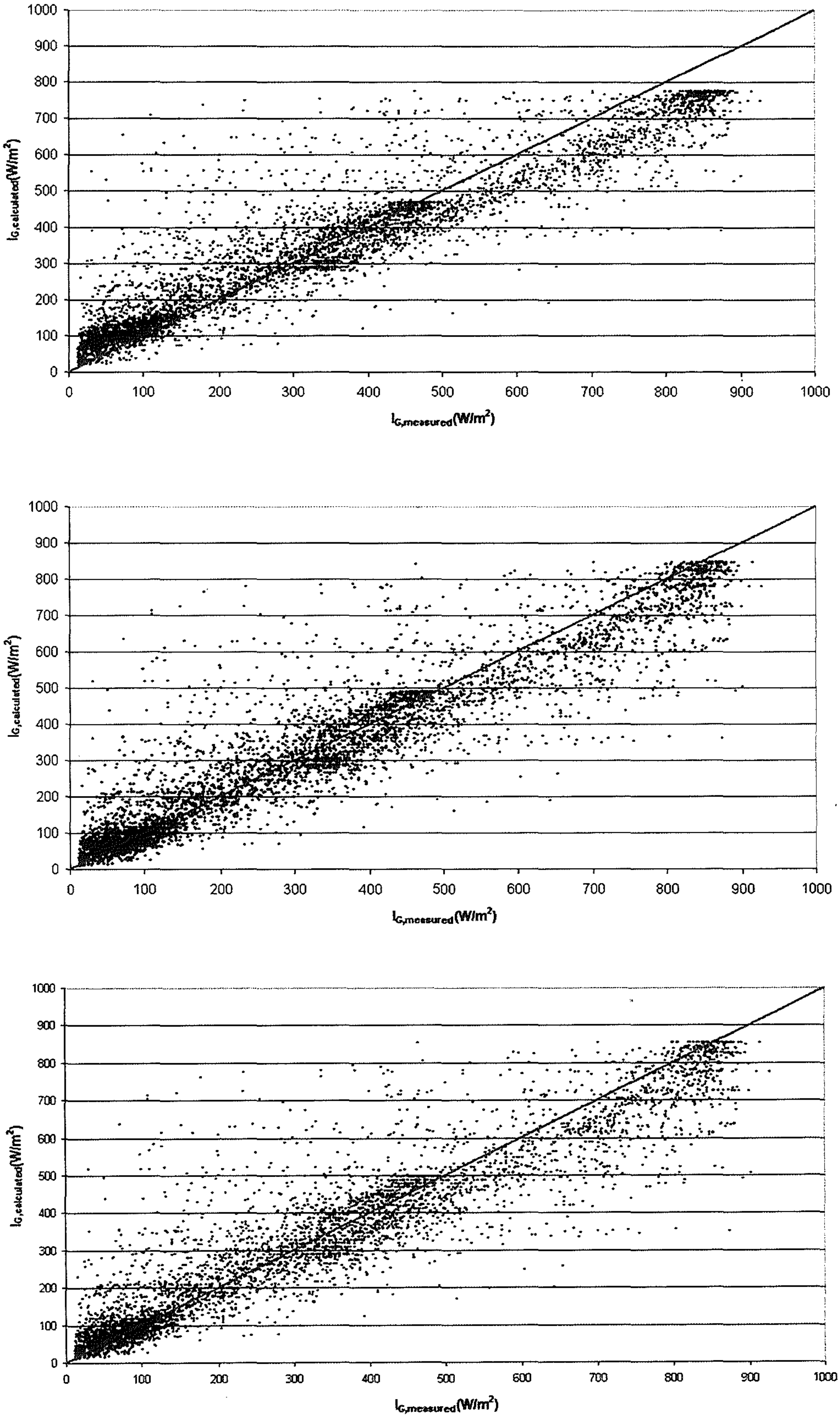


Figure 5.11 Scatter plots of measured versus calculated global irradiation for Gerona. a. M1 validation; b. M2 validation; c. M7 validation.

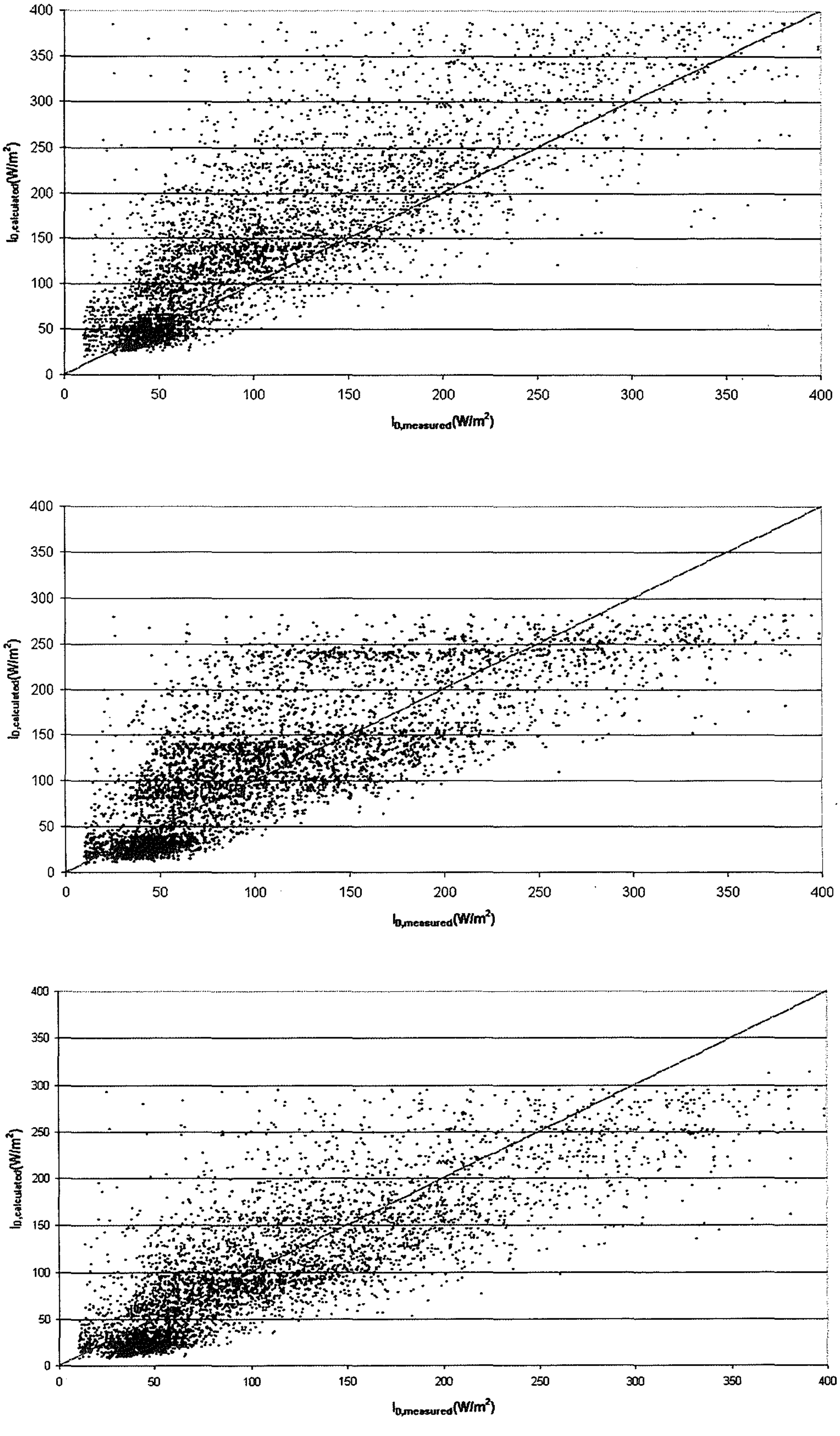


Figure 5.12 Scatter plots of measured versus calculated diffuse irradiation for Gerona. a. M1 validation; b. M2 validation; c M7 validation.

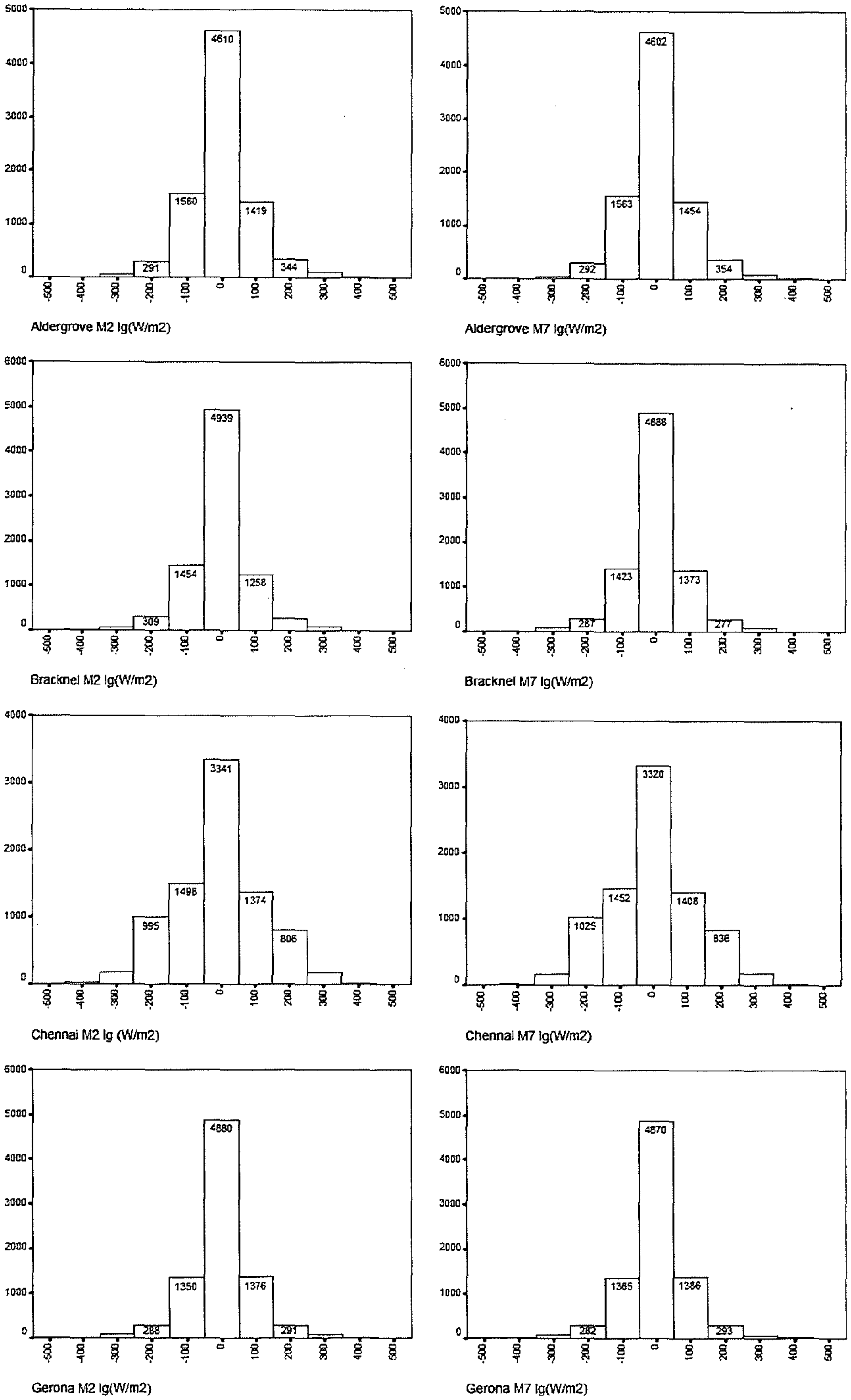


Figure 5.13 Global horizontal irradiation estimation error histograms of M2 and M7 for the seven datasets. Note that x-axis represents radiation error in W/m^2 and y-axis is the number of occurrences. Plots for 4 of the 7 sites.

Fig. 5.13-continued

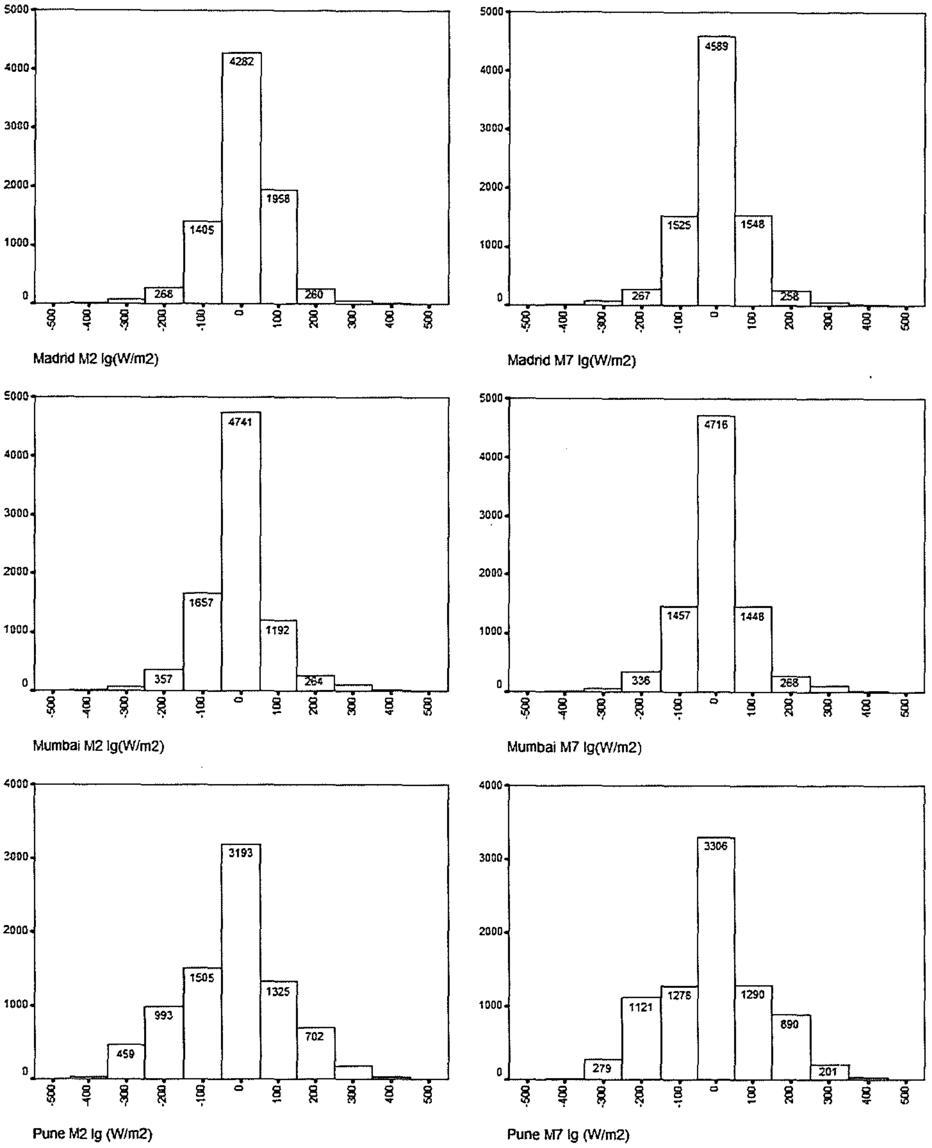


Figure 5.13 Global horizontal irradiation estimation error histograms of M2 and M7 for the seven datasets. Note that x-axis represents radiation error in W/m^2 and y-axis is the number of occurrences. Plots for 3 of the 7 sites.

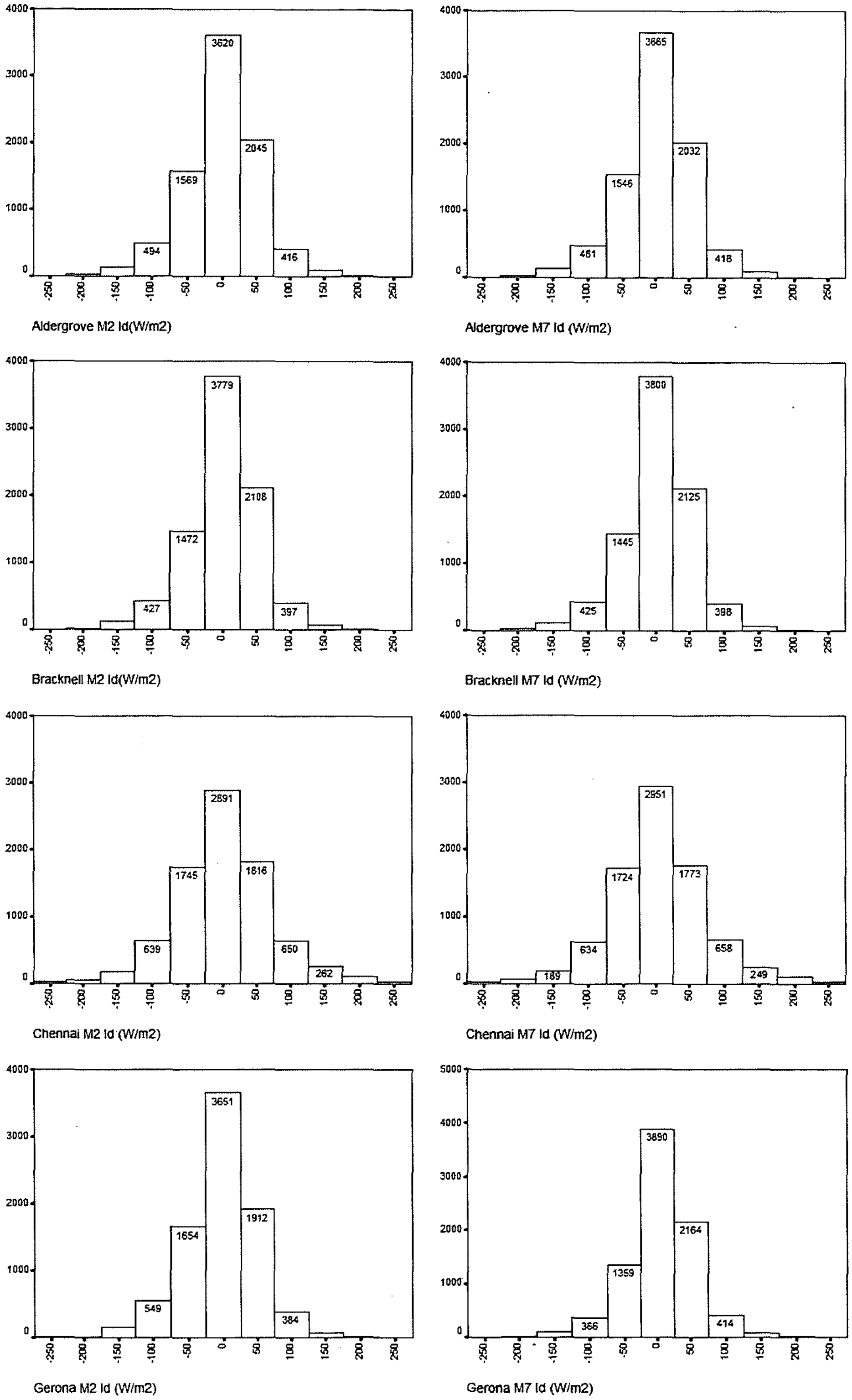


Figure 5.14: Diffuse horizontal irradiation estimation error histograms of M2 and M7 for the seven datasets. Note that x-axis represents radiation error in W/m^2 and y-axis is the number of occurrences. Plots for 4 of the 7 sites.

Figure 5.14-continued

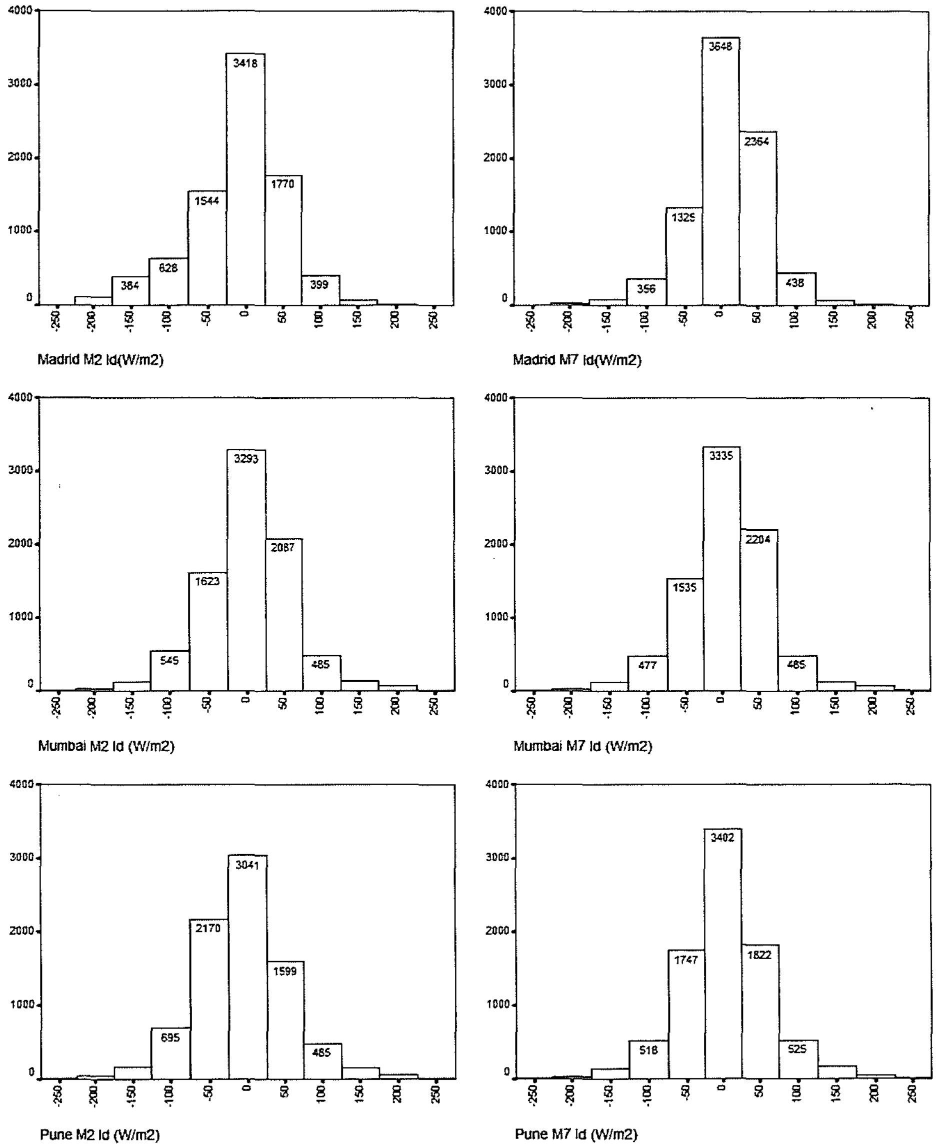


Figure 5.14: Diffuse horizontal irradiation estimation error histograms of M2 and M7 for the seven datasets. Note that x-axis represents radiation error in W/m^2 and y-axis is the number of occurrences. Plots for 3 of the 7 sites.

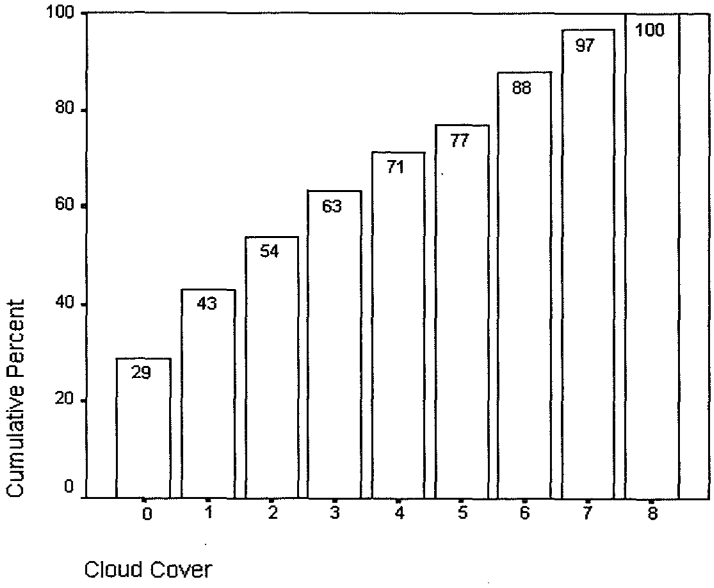


Figure 5.15 Cumulative percentage frequency diagram of cloud cover for Madrid.

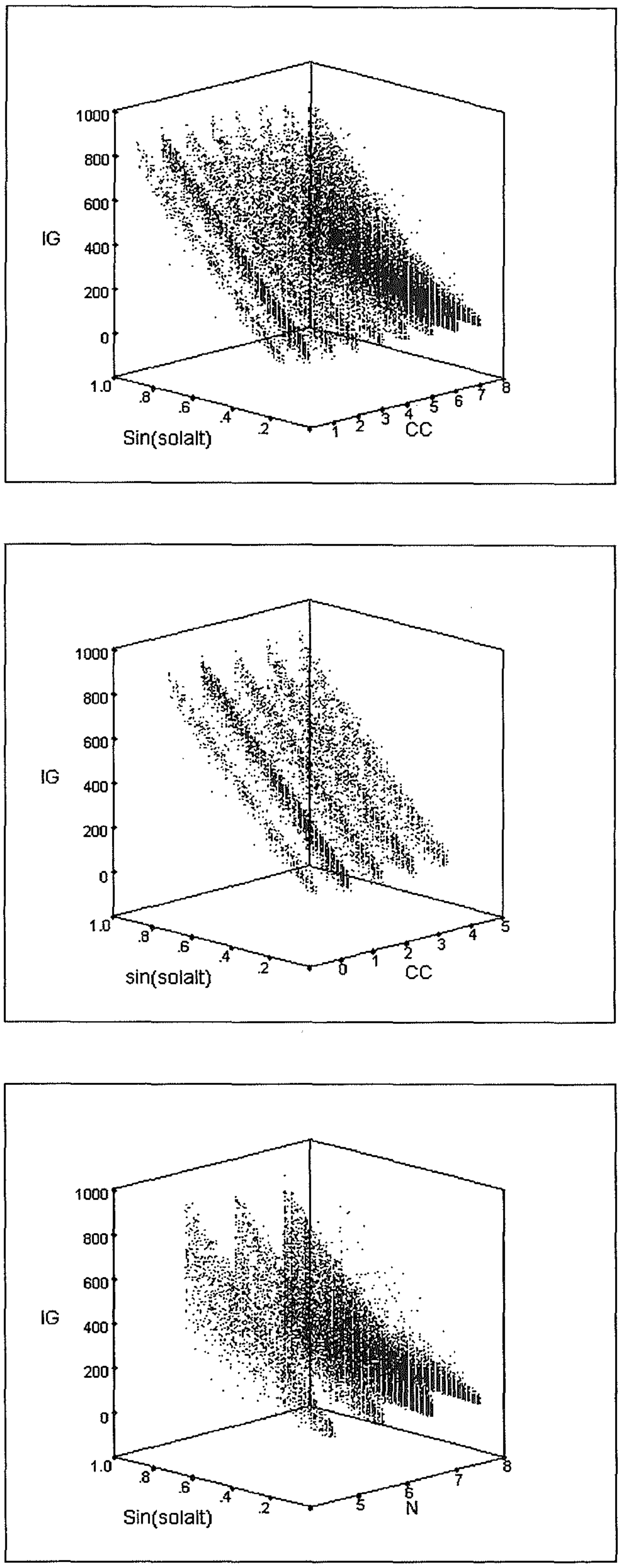


Figure 5.16 Three-dimensional scatter plots of the sine of the solar altitude, cloud cover and I_G for Bracknell. Note that N in Oktas and I_G in W/m^2.

a-c.

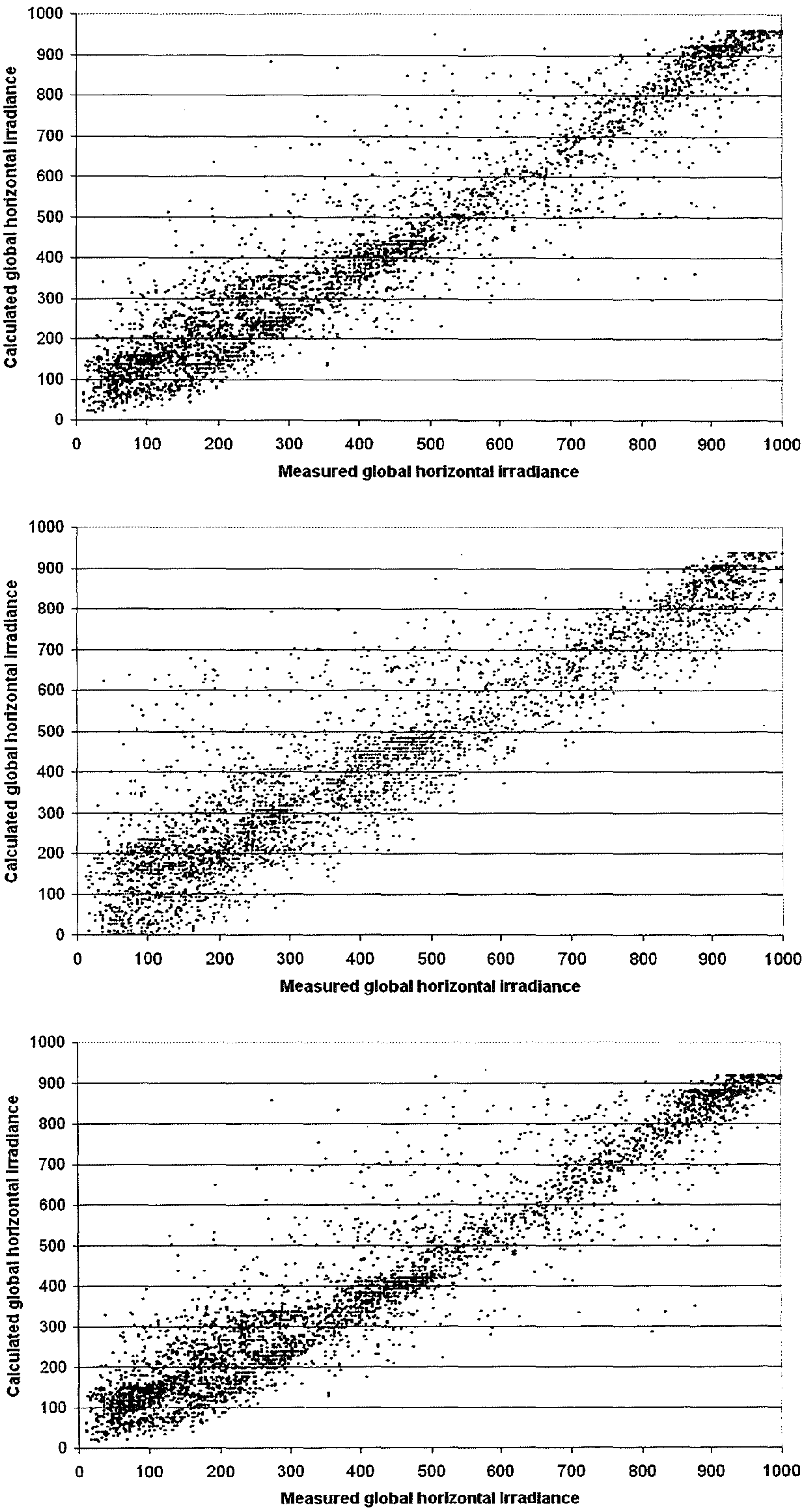


Figure 5.17 Scatter plot of measured versus calculated solar horizontal irradiance components, a-c. I_G; d-f. I_D and g-i. I_B for Madrid using Muneer-Gul, modified Lam-Li and the proposed bi-variate models, respectively. Global irradiance.

Fig. 5.17-continued
d-f

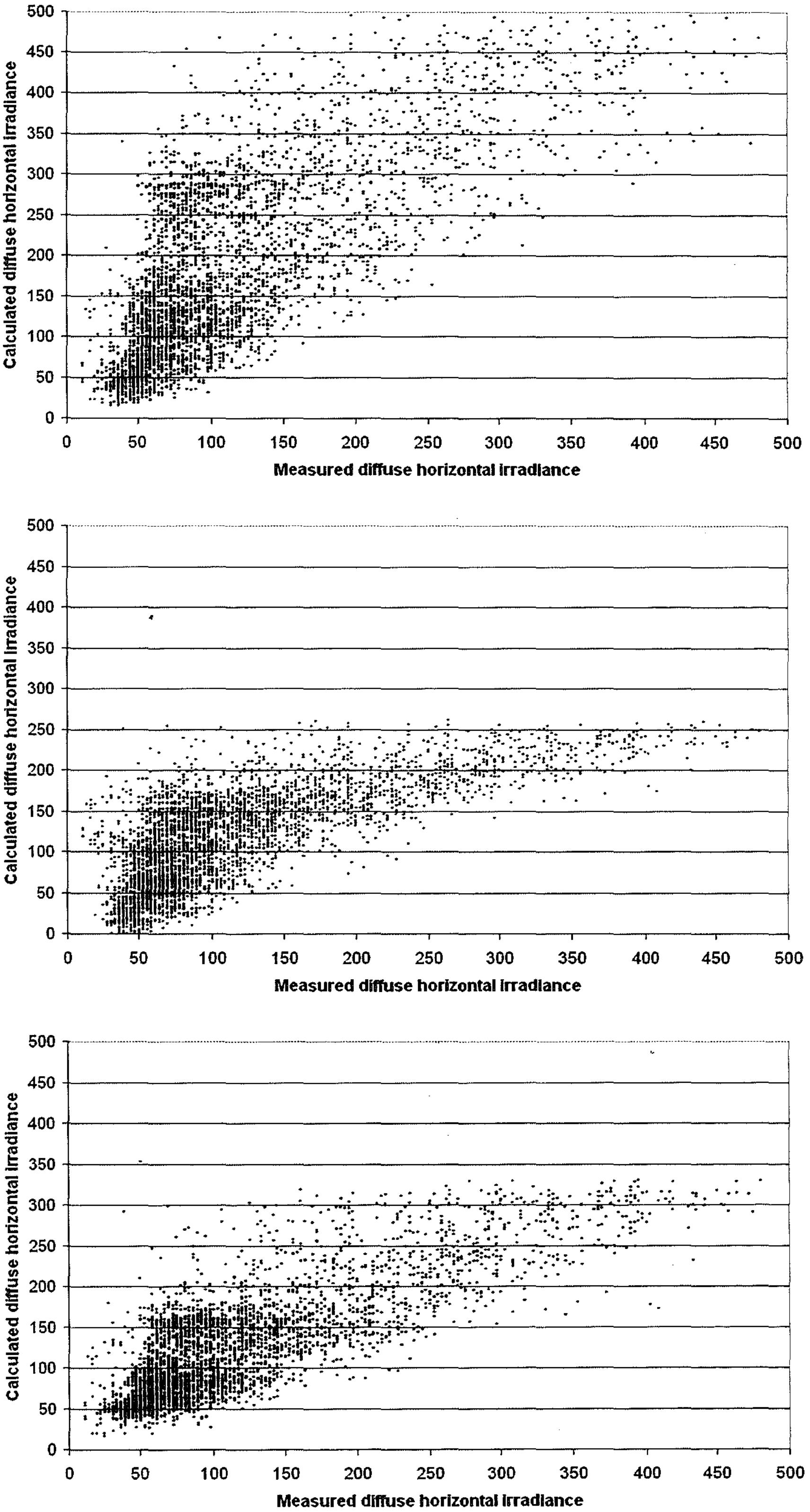


Figure 5.17 Scatter plot of measured versus calculated solar horizontal irradiance components, a-c. I_G; d-f. I_D and g-i. I_B for Madrid using Muneer-Gul, modified Lam-Li and the proposed bi-variate models, respectively. Diffuse irradiance.

Fig. 5.17-continued
h-j.

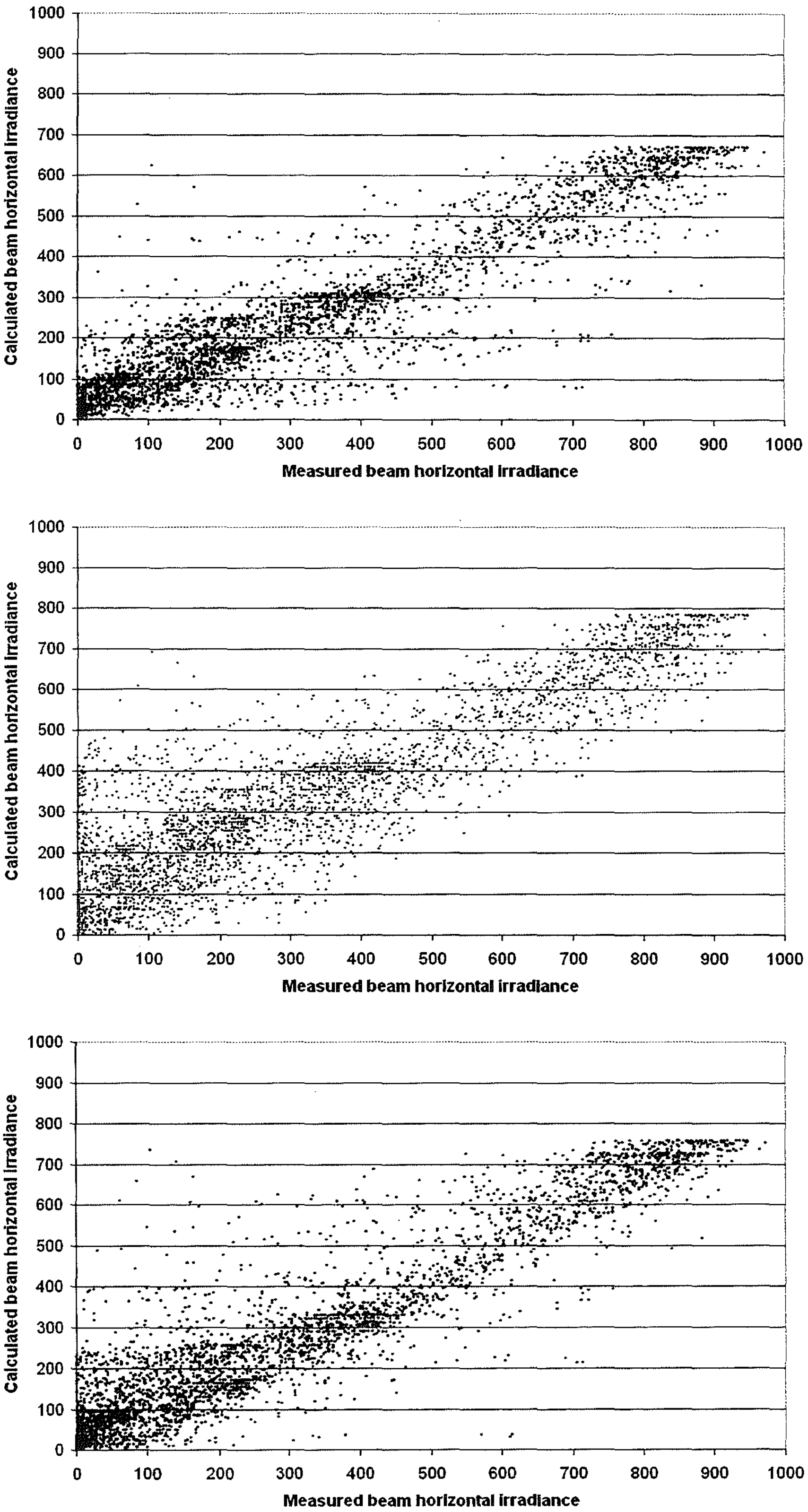


Figure 5.17 Scatter plot of measured versus calculated solar horizontal irradiance components, a-c. I_G; d-f. I_D and g-i. I_B for Madrid using Muneer-Gul, modified Lam-Li and the proposed bi-variate models, respectively. Beam irradiance.

a-c.

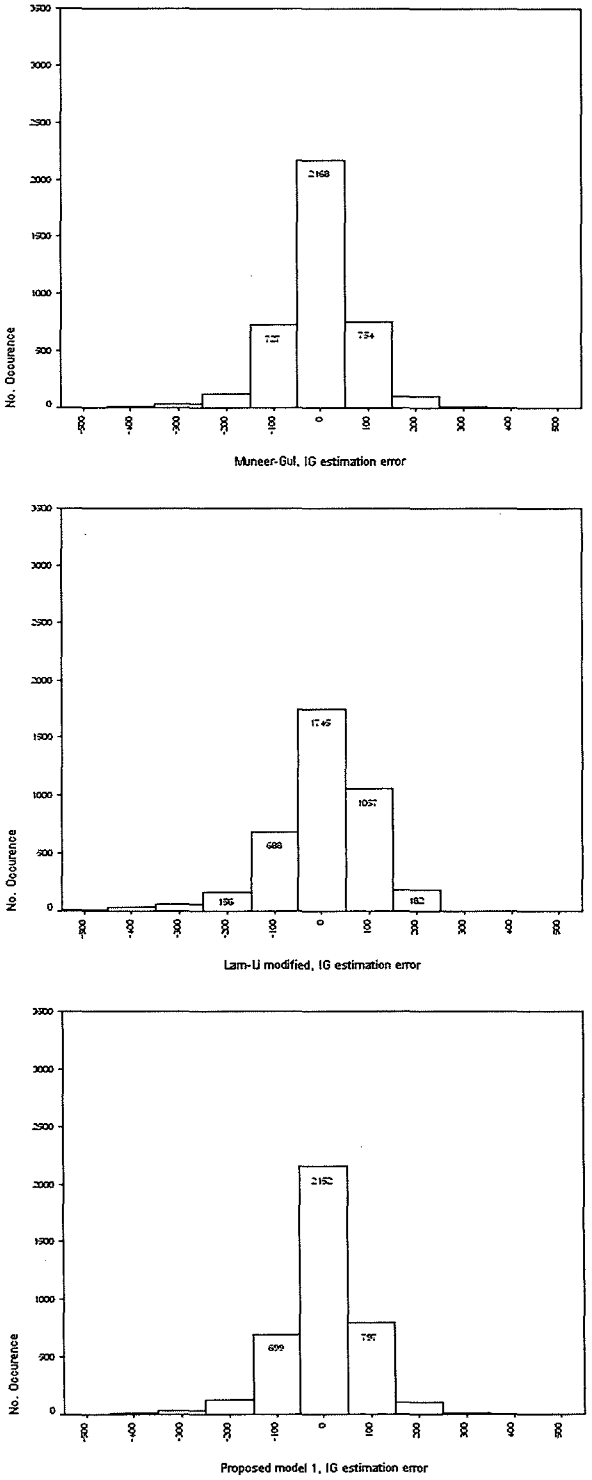


Figure 5.18 Error histograms for solar horizontal irradiance components estimation; a-c. I_G; d-f. I_D and g-i. I_B for Madrid using Muneer-Gul, modified Lam-Li and the proposed bi-variate models, respectively. Global irradiance.

a-c.

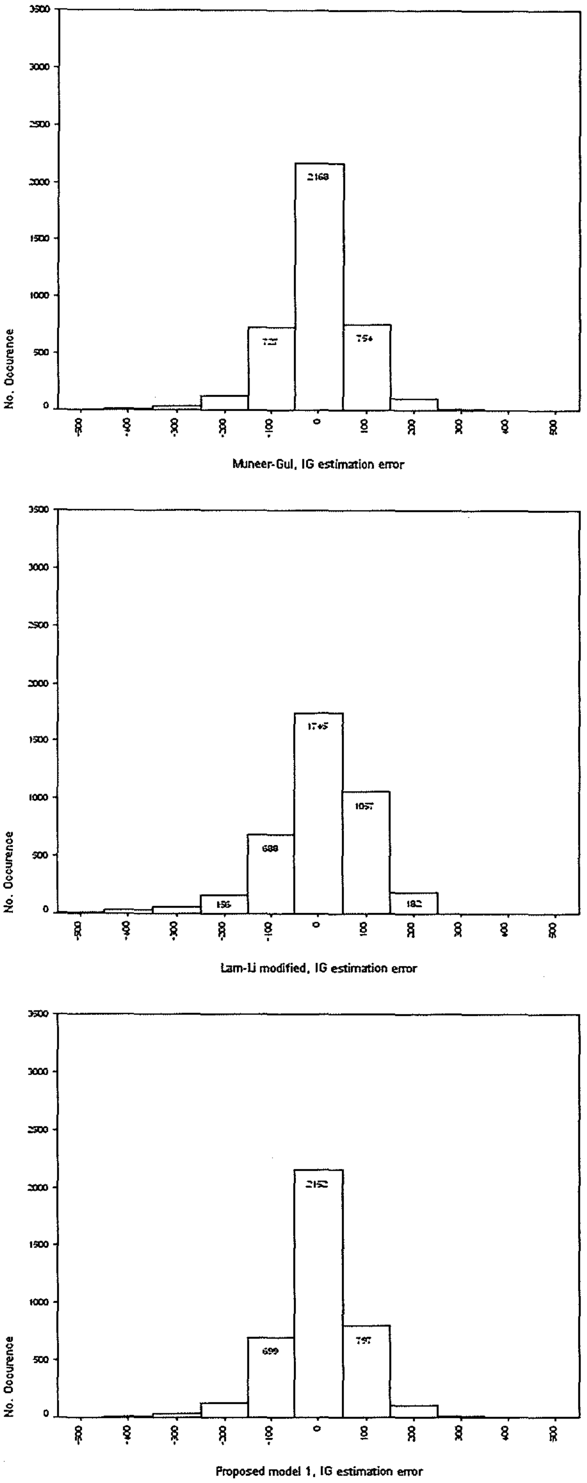


Figure 5.18 Error histograms for solar horizontal irradiance components estimation; a-c. I_G; d-f. I_D and g-i. I_B for Madrid using Muneer-Gul, modified Lam-Li and the proposed bi-variate models, respectively. Global irradiance.

Fig. 5.18-continued
d-f.

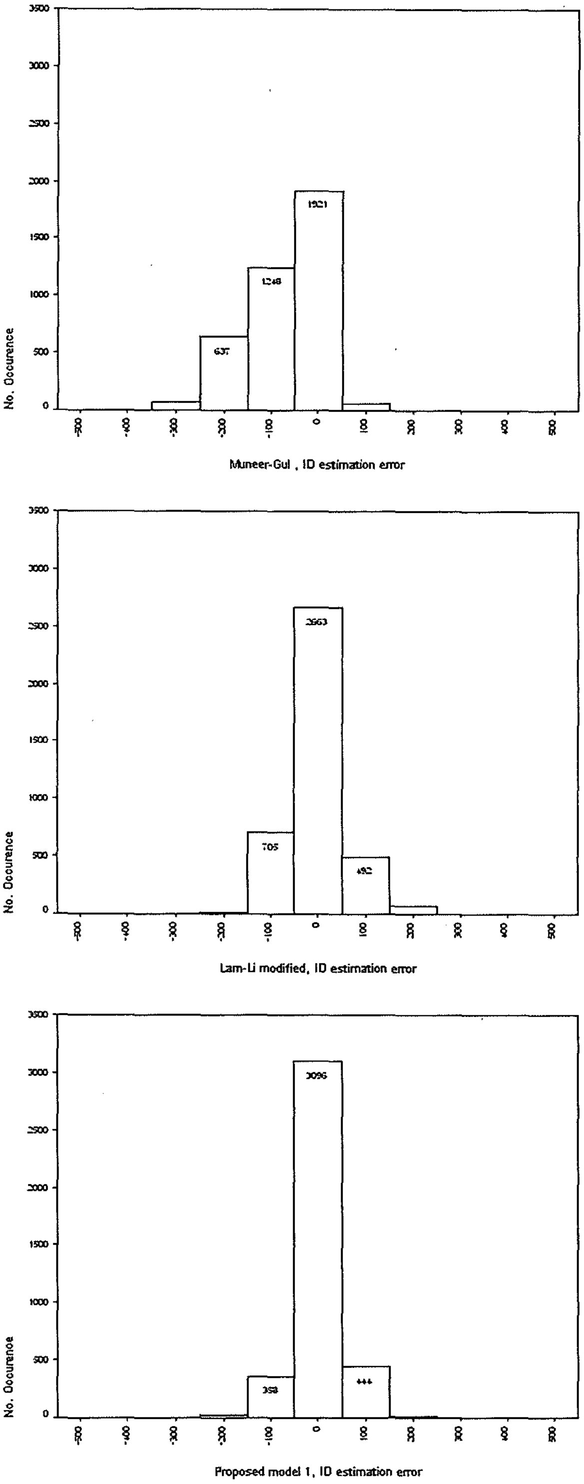


Figure 5.18 Error histograms for solar horizontal irradiance components estimation; a-c. I_G; d-f. I_D and g-i. I_B for Madrid using Muneer-Gul, modified Lam-Li and the proposed bi-variate models, respectively. Diffuse irradiance.

Fig. 5.18-continued
g-i-

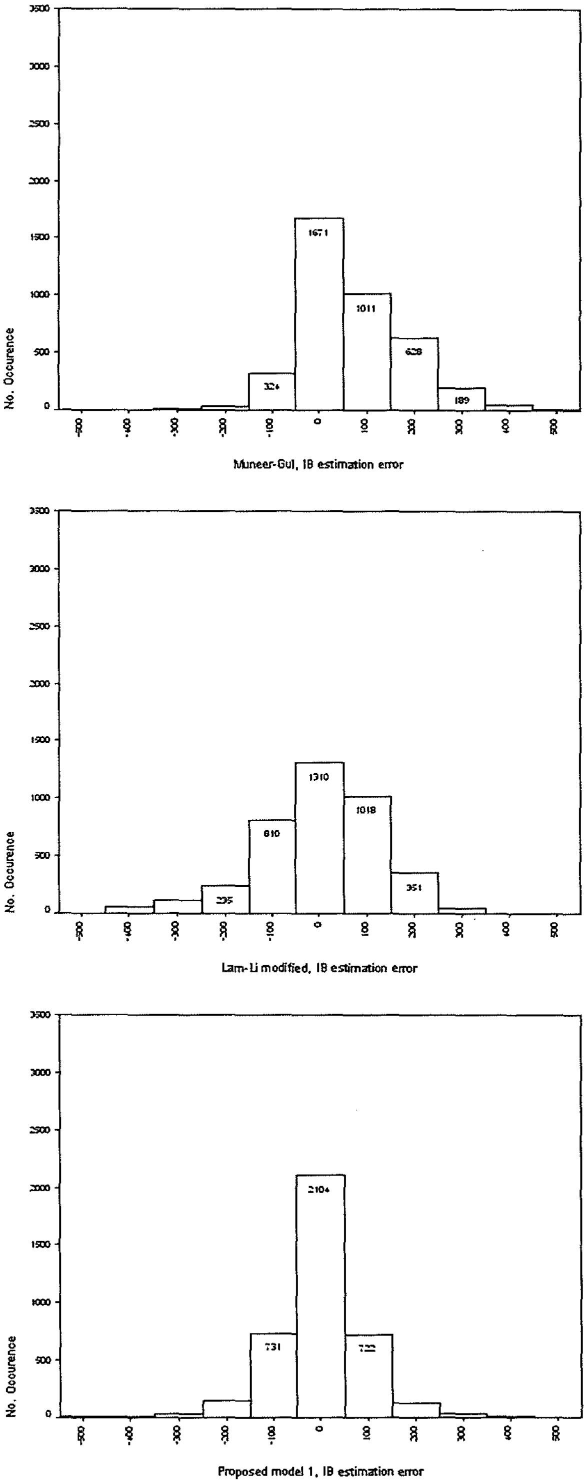


Figure 5.18 Error histograms for solar horizontal irradiance components estimation; a-c. I_G; d-f. I_D and g-i. I_B for Madrid using Muneer-Gul, modified Lam-Li and the proposed bi-variate models, respectively. Beam irradiance.

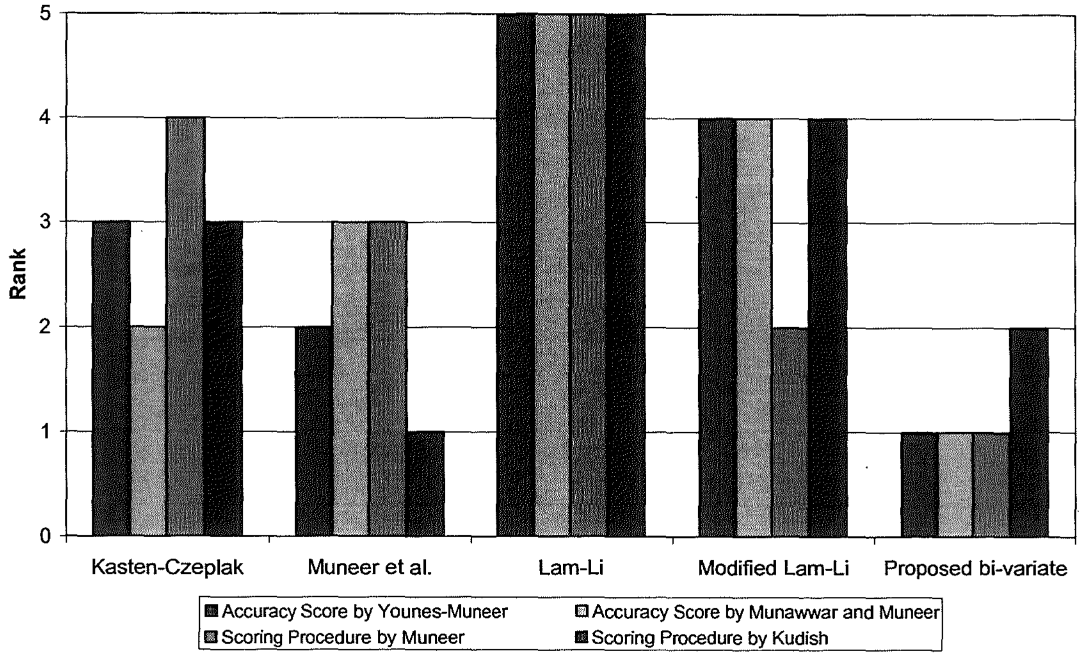


Figure 5.19 Ranking of the CRM models based on the four difference scoring procedures.

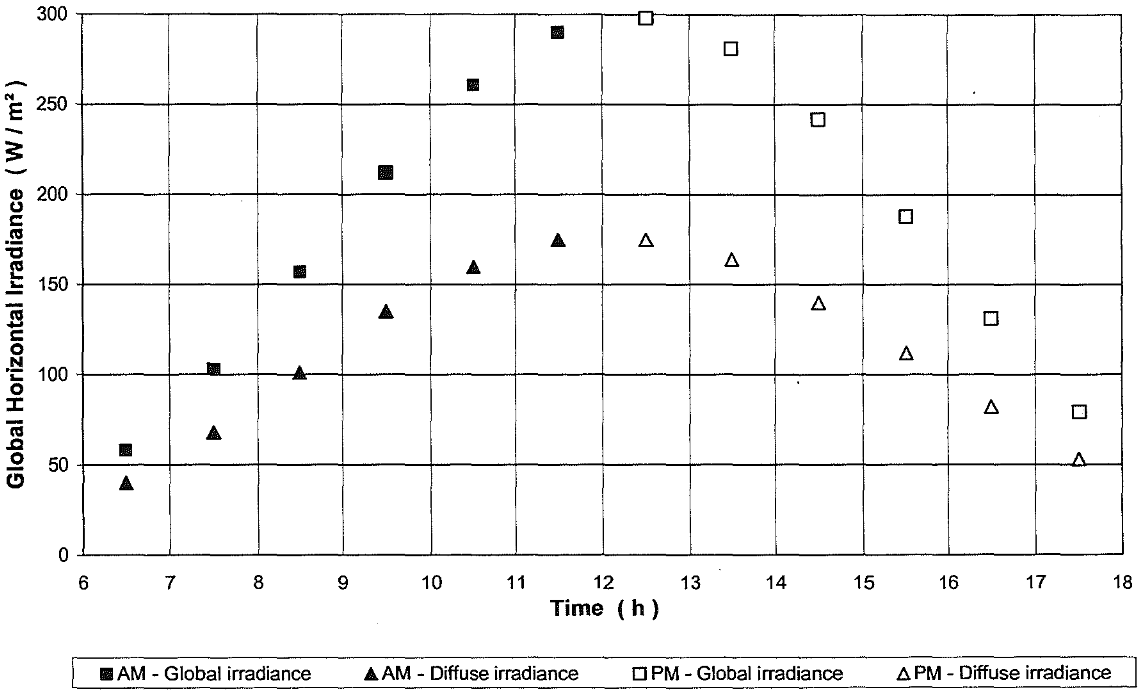


Figure 5.20: Scatter plot of morning and afternoon, global and diffuse horizontal irradiance based on 27-year, hourly data.

Table 5.1 Results of the seven cloud radiation models for Aldergrove.

Model Number	I_G						I_D					
	Slope	R^2	MBE*	RMSE*	Kurtosis	Skewness	Slope	R^2	MBE*	RMSE*	Kurtosis	Skewness
M1	0.73	0.8	6	86	2.41	0.77	0.74	0.7	-1	53	2.15	0.29
M2	0.80	0.8	1	84	2.45	0.39	0.77	0.7	3	53	1.61	-0.07
M3	0.84	0.7	70	131	0.21	-0.22	0.50	0.6	34	73	0.98	0.63
M4	0.71	0.7	0	103	0.63	0.15	0.64	0.6	0	61	0.91	-0.03
M5	0.78	0.8	-2	88	2.17	0.56	0.77	0.7	3	52	1.68	-0.11
M6	0.80	0.8	-1	84	2.50	0.44	0.77	0.7	2	53	1.42	-0.06
M7	0.80	0.8	-1	84	2.52	0.40	0.77	0.7	3	52	1.66	-0.10

* Units in W/m^2

Table 5.2 AS results of the seven cloud radiation models for Aldergrove.

Model Number	I_G AS	I_D AS	AS $_{Total}$
M1	3.3	4.3	7.6
M2	4.1	4.6	8.7
M3	2.7	1.3	4.0
M4	3.1	3.8	6.9
M5	3.7	4.7	8.4
M6	4.1	4.5	8.6
M7	4.2	4.7	8.8

*. Note that the maximum attainable score is 12.

Table 5.3 Global horizontal irradiation M2 and M7 results for the seven sites.

Location	Model No.	Slope	R^2	MBE	RMSE	Kurtosis	Skewness	AS I_G
Aldergrove	M2	0.80	0.81	-1.00	84.00	2.63	0.38	3.56
	M7	0.80	0.81	-1.00	84.00	2.55	0.38	3.55
Bracknell	M2	0.84	0.81	-3.00	92.00	3.39	0.02	4.67
	M7	0.83	0.81	0.00	91.00	2.94	0.03	4.59
Chennai	M2	0.53	0.68	-16.00	158.00	0.14	-0.02	2.53
	M7	0.53	0.68	-14.00	158.00	0.02	0.00	2.50
Gerona	M2	0.85	0.86	1.00	94.00	4.44	-0.24	5.75
	M7	0.86	0.86	-1.00	94.00	4.46	-0.26	5.78
Madrid	M2	0.87	0.91	12.00	87.00	3.66	-0.25	6.01
	M7	0.91	0.91	-1.00	85.00	3.60	-0.09	5.36
Mumbai	M2	0.90	0.87	-13.00	92.00	3.01	0.35	3.63
	M7	0.87	0.88	-2.00	91.00	2.89	0.23	4.17
Pune	M2	0.53	0.64	-36.00	173.00	0.07	-0.03	1.84
	M7	0.51	0.65	-15.00	170.00	-0.03	0.01	2.29

Table 5.4 Diffuse horizontal irradiation M2 and M7 results for the seven sites.

Location	Model No.	Slope	R^2	MBE	RMSE	Kurtosis	Skewness	AS I_D
Aldergrove	M2	0.77	0.74	3.00	53.00	1.46	-0.20	4.25
	M7	0.77	0.74	3.00	52.00	1.49	-0.20	4.26
Bracknell	M2	0.77	0.71	4.00	51.00	1.72	-0.29	4.24
	M7	0.83	0.72	4.00	51.00	1.81	-0.29	4.43
Chennai	M2	0.55	0.56	0.00	87.00	2.06	-0.09	3.38
	M7	0.57	0.57	0.00	86.00	1.68	0.03	3.50
Gerona	M2	0.74	0.60	-7.00	53.00	1.32	-0.37	3.28
	M7	0.76	0.67	6.00	47.00	2.16	-0.42	4.28
Madrid	M2	0.69	0.46	-25.00	72.00	1.42	-0.67	1.65
	M7	0.82	0.70	9.00	48.00	2.66	-0.46	4.67
Mumbai	M2	0.70	0.61	2.00	66.00	1.73	0.01	4.13
	M7	0.72	0.62	6.00	66.00	2.10	0.07	4.60
Pune	M2	0.59	0.69	-18.00	69.00	1.31	0.12	3.18
	M7	0.66	0.70	-2.00	65.00	1.54	0.10	4.09

Table 5.5 Estimation percentage error occurrences in the upper ±10% band and the ±25% band based on M2 and M7 for Bracknell, Chennai, Gerona, Madrid, Mumbai and Pune.

	Bracknell				Chennai				Gerona			
	M2		M7		M2		M7		M2		M7	
	I_G	I_D	I_G	I_D	I_G	I_D	I_G	I_D	I_G	I_D	I_G	I_D
±10%	6018	4786	5760	4814	1105	1096	1113	1100	1725	726	1729	829
±25%	11647	11418	11914	11486	2373	2507	2339	2507	3233	1870	3237	2071
Total no. Points	22111	22111	22111	22111	4184	4184	4184	4184	4863	4863	4863	4863

	Madrid				Mumbai				Pune			
	M2		M7		M2		M7		M2		M7	
	I_G	I_D	I_G	I_D	I_G	I_D	I_G	I_D	I_G	I_D	I_G	I_D
±10%	1345	662	1558	631	1917	926	1966	971	967	909	1105	1090
±25%	2649	1573	2637	1592	3058	2196	3083	2253	2180	2105	2020	2415
Total no. Points	3936	3936	3936	3936	4011	4011	4011	4011	4008	4008	4008	4008

Table 5.6 M1, M2 and M7 results, based on clearness index for all seven sites.

Site	Model	Band	I_G MBE	I_D MBE	I_G RMSE	I_D RMSE
Aldergrove	M1	$k_t \leq 0.2$	-43	-33	67	50
		$0.2 < k_t < 0.8$	26	13	92	54
		$k_t \geq 0.8$	147	-24	165	60
	M2	$k_t \leq 0.2$	-46	-33	72	51
		$0.2 < k_t < 0.8$	17	19	88	53
		$k_t \geq 0.8$	136	-14	150	58
	M7	$k_t \leq 0.2$	-46	-32	71	51
		$0.2 < k_t < 0.8$	17	19	87	53
		$k_t \geq 0.8$	137	-15	152	59
Bracknell	M1	$k_t \leq 0.2$	-43	-33	70	52
		$0.2 < k_t < 0.8$	31	2	101	52
		$k_t \geq 0.8$	211	-83	215	102
	M2	$k_t \leq 0.2$	-51	-31	81	51
		$0.2 < k_t < 0.8$	14	16	95	51
		$k_t \geq 0.8$	164	-41	168	75
	M7	$k_t \leq 0.2$	-52	-30	83	51
		$0.2 < k_t < 0.8$	18	16	93	51
		$k_t \geq 0.8$	221	-52	222	77
Chennai	M1	$k_t \leq 0.2$	-146	-86	178	99
		$0.2 < k_t < 0.8$	20	17	169	101
		$k_t \geq 0.8$	244	12	253	78
	M2	$k_t \leq 0.2$	-280	-177	296	206
		$0.2 < k_t < 0.8$	-45	2	141	83
		$k_t \geq 0.8$	206	13	212	76
	M7	$k_t \leq 0.2$	-56	-34	92	58
		$0.2 < k_t < 0.8$	6	14	106	59
		$k_t \geq 0.8$	211	13	216	76
Gerona	M1	$k_t \leq 0.2$	-92	-63	141	91
		$0.2 < k_t < 0.8$	7	-35	89	62
		$k_t \geq 0.8$	188	-146	189	150
	M2	$k_t \leq 0.2$	-73	-33	127	58
		$0.2 < k_t < 0.8$	13	-2	87	52
		$k_t \geq 0.8$	155	-114	160	121
	M7	$k_t \leq 0.2$	-74	-30	127	57
		$0.2 < k_t < 0.8$	12	13	87	45
		$k_t \geq 0.8$	152	-78	158	84

Table 5.6 M1, M2 and M7 results, based on clearness index for all seven sites (continued).

Site	Model	Band	I_G MBE	I_D MBE	I_G RMSE	I_D RMSE
Madrid	M1	$k_t \leq 0.2$	-47	-34	76	53
		$0.2<k_t<0.8$	53	-43	110	84
		$k_t \geq 0.8$	114	-60	123	99
	M2	$k_t \leq 0.2$	-92	-50	118	69
		$0.2<k_t<0.8$	19	-22	83	71
		$k_t \geq 0.8$	102	-45	104	89
	M7	$k_t \leq 0.2$	-94	-44	120	64
		$0.2<k_t<0.8$	6	14	81	47
		$k_t \geq 0.8$	86	1	90	43
Mumbai	M1	$k_t \leq 0.2$	-62	-56	97	82
		$0.2<k_t<0.8$	39	-5	103	68
		$k_t \geq 0.8$	229	22	271	49
	M2	$k_t \leq 0.2$	-96	-73	134	98
		$0.2<k_t<0.8$	-6	8	87	63
		$k_t \geq 0.8$	204	33	234	63
	M7	$k_t \leq 0.2$	-98	-66	135	94
		$0.2<k_t<0.8$	6	12	86	63
		$k_t \geq 0.8$	209	35	237	64
Pune	M1	$k_t \leq 0.2$	-179	-70	202	79
		$0.2<k_t<0.8$	14	-23	170	82
		$k_t \geq 0.8$	256	-20	270	59
	M2	$k_t \leq 0.2$	-232	-81	241	97
		$0.2<k_t<0.8$	-68	-15	157	69
		$k_t \geq 0.8$	195	-3	205	51
	M7	$k_t \leq 0.2$	-219	-65	228	80
		$0.2<k_t<0.8$	-47	0	148	65
		$k_t \geq 0.8$	220	15	226	52

Table 5.7 M1, M2 and M7 results, based on cloud cover information for all seven sites.

Site	Model	Band	I_G MBE	I_D MBE	I_G RMSE	I_D RMSE
Aldergrove	M1	N≤1	22	4	67	41
		1<N<7	9	-6	90	48
		N≥7	4	0	86	56
	M2	N≤1	13	9	57	44
		1<N<7	3	7	88	48
		N≥7	-4	0	85	55
	M7	N≤1	1	10	57	45
		1<N<7	3	8	87	48
		N≥7	-3	-1	84	55
Bracknell	M1	N≤1	11	43	5	46
		1<N<7	6	-17	89	54
		N≥7	15	-2	99	51
	M2	N≤1	-2	33	7	48
		1<N<7	-11	5	88	50
		N≥7	3	3	93	50
	M7	N≤1	15	39	36	58
		1<N<7	0	6	84	49
		N≥7	-1	2	95	51
Chennai	M1	N≤1	N/A	N/A	N/A	N/A
		1<N<7	29	1	170	83
		N≥7	179	119	260	182
	M2	N≤1	N/A	N/A	N/A	N/A
		1<N<7	-15	1	157	81
		N≥7	-21	-7	168	127
	M7	N≤1	N/A	N/A	N/A	N/A
		1<N<7	-14	1	157	81
		N≥7	-10	-6	165	122
Gerona	M1	N≤1	13	-32	56	53
		1<N<7	-11	-44	108	71
		N≥7	-19	-32	98	68
	M2	N≤1	4	-29	44	55
		1<N<7	0	-1	104	51
		N≥7	-2	-3	96	57
	M7	N≤1	0	8	44	34
		1<N<7	-1	6	104	47
		N≥7	-2	3	96	57

Table 5.7 M1, M2 and M7 results, based on cloud cover information for all seven sites (continued).

Site	Model	Band	I_G MBE	I_D MBE	I_G RMSE	I_D RMSE
Madrid	M1	N≤1	52	-55	80	81
		1<N<7	38	-60	105	89
		N≥7	60	14	137	65
	M2	N≤1	34	-54	56	84
		1<N<7	6	-22	87	69
		N≥7	0	3	114	62
	M7	N≤1	7	15	45	33
		1<N<7	-3	8	86	48
		N≥7	-1	3	114	61
Mumbai	M1	N≤1	18	0	55	57
		1<N<7	32	-25	97	70
		N≥7	47	12	151	83
	M2	N≤1	-16	-2	52	59
		1<N<7	-16	5	84	62
		N≥7	-2	-1	138	82
	M7	N≤1	2	8	48	58
		1<N<7	-4	8	82	62
		N≥7	-2	-1	138	81
Pune	M1	N≤1	-19	-27	164	50
		1<N<7	52	-48	197	85
		N≥7	144	41	238	108
	M2	N≤1	-62	-30	165	52
		1<N<7	-28	-11	179	69
		N≥7	-1	-8	175	98
	M7	N≤1	-20	1	159	41
		1<N<7	-15	-3	178	68
		N≥7	-4	-8	173	94

Table 5.8 The ranges of the statistical indicators for the Aldergrove data evaluation

Statistical Indicator	I_G	I_D	I_B
Slope	0.71 to 0.84	0.50 to 0.83	0.55 to 0.76
R^2	0.70 to 0.81	0.61 to 0.74	0.55 to 0.73
MBE	-0.78 to 70.13	-7.00 to 33.56	-1.24 to 36.56
RMSE	84.12 to 130.57	52.69 to 72.93	70.32 to 104.55
Kurtosis	0.21 to 2.48	0.91 to 2.15	0.27 to 7.41
Skewness	-0.22 to 0.77	-0.28 to 0.63	0.12 to 1.92

Table 5.9 The accuracy score results of the CRM comparisons on Bracknell.

Model Name	Accuracy Score			
	IG (/6)	ID (/6)	IB (/6)	Total (/18)
Kasten-Czeplak	4.00	4.51	3.47	11.98
Muneer et al.	4.98	3.24	3.88	12.09
Lam-Li	2.65	2.40	2.77	7.83
Modified Lam-Li	4.06	3.70	3.45	11.21
Proposed bi-variate	4.96	4.86	4.48	14.29
Proposed hybrid bi-variate	4.89	4.94	4.46	14.30

Table 5.10 Values of coefficients used in the proposed bi-variate model.

Coefficient	Aldergrove	Bracknell	Chennai	Gerona	Madrid	Mumbai	Pune
A	1046	1046	1295	1045	1046	1046	1024
B	81	75	326	81	81	81	260
C	1.00	0.94	1.05	1.01	1.05	0.96	1.30
D	-0.73	-0.68	-0.38	-0.78	-0.65	-0.68	-0.74
E	3.17	4.02	2.09	2.37	3.06	4.55	5.25
F	997	995	1405	997	997	995	965
G	111	121	492	111	111	126	289
H	0.88	0.84	0.89	0.94	1.04	0.80	1.14
I	-0.90	-0.83	-0.81	-0.94	-0.96	-0.80	-1.09
J	1.47	1.73	1.52	1.53	1.81	2.45	2.70

Table 5.11 Results of validation of the bi-variate model.

Location	Coefficient	I_G MBE	I_G RMSE	I_D MBE	I_D RMSE	I_B MBE	I_B RMSE
Edinburgh	Local	-8.45	224.39	-6.30	108.48	-2.15	183.10
	Generalized	-31.34	240.71	-6.61	104.46	-24.72	201.41
London	Local	-0.13	98.63	0.89	61.60	-1.02	82.80
	Generalized	-11.28	99.54	-1.48	62.42	-9.80	83.74
Camborne	Local	-0.61	101.01	N/A	N/A	N/A	N/A
	Generalized	13.60	103.09	N/A	N/A	N/A	N/A

Table 5.12 Results of the evaluation of the leading CRM by using a- the proposed AS; b- Munawwar and Muneer AS; c- Muneer scoring procedure; d- Kudish scoring procedure; for Bracknell.

a-

Model Name	Global	Diffuse	Beam	Total
Kasten-Czeplak	4.00	4.52	3.47	11.99
Muneer et al.	4.98	3.25	3.88	12.11
Lam-Li	2.65	2.41	2.77	7.84
Modified Lam-Li	4.06	3.71	3.45	11.22
Proposed bi-variate	4.96	4.87	4.48	**14.31**

Note that the model that scores the closest to 18 is the best performing.

b-

Model Name	Global	Diffuse	Beam	Total
Kasten-Czeplak	3.11	3.66	2.62	9.39
Muneer et al.	3.99	2.25	2.96	9.20
Lam-Li	1.65	1.83	1.77	5.25
Modified Lam-Li	3.21	3.04	2.67	8.92
Proposed bi-variate	4.00	4.07	3.51	**11.58**

Note that the model that scores the closest to 15 is the best performing.

c-

Name	Global	Diffuse	Beam	Total
Kasten-Czeplak	106.09	59.67	115.20	280.97
Muneer et al.	94.63	75.48	108.02	278.12
Lam-Li	204.85	90.40	162.13	457.39
Modified Lam-Li	111.74	58.51	107.54	277.79
Proposed bi-variate	91.71	51.21	89.82	**232.74**

Note that the model that has the lowest score is the best performing.

d-

Model Name	Global	Diffuse	Beam	Total
Kasten-Czeplak	0.46	0.33	0.58	1.38
Muneer et al.	0.39	0.28	0.54	**1.21**
Lam-Li	0.72	0.58	0.70	2.00
Modified Lam-Li	0.52	0.46	0.61	1.59
Proposed bi-variate	0.40	0.37	0.48	1.24

Note that the model that has the lowest score is the best performing.

Table 5.13 Performance evaluation of the proposed bi-variate model compared to the other CRMs using the various scoring procedures.

Model Name	New Accuracy Score	Accuracy Score	Scoring Procedure by Muneer	Scoring Procedure by Kudish
Kasten-Czeplak	16.2%	19.0%	17.2%	9.7%
Muneer et al.	15.4%	20.6%	16.3%	-3.2%
Lam-Li	45.2%	54.7%	49.1%	37.9%
Modified Lam-Li	21.6%	23.0%	16.2%	21.9%

Note that a negative result implies that the proposed bi-variate model did not improve on the performance of the estimation of the terrestrial horizontal irradiance.

Table. 5.14 Cumulative energy assessment for the given PV installation

Month	Calculated Global Tilted Irradiation MWh	Measured Global Tilted Irradiation MWh	PV AC Output MWh	PV System Conversion Efficiency %
April*	10.44	Not measured	1.59	Not Calculated
May	13.00	Not measured	1.47	Not Calculated
June	12.53	10.27	1.14	11.1
July	12.56	11.98	1.34	11.2
August	11.58	11.20	1.33	11.9
September	8.77	9.72	1.19	12.3
October	5.97	5.39	0.57	10.6
Totals	**51.41****	**48.56**	**5.57****	**11.5*****

* Measurement started on 6th April

** Only June to October data were used in Totals

*** Only June to October data were used to calculate the overall system efficiency

Table 5.15: June 2005 Incident Energy Based on Calculations Compared to Typical June Month Using Various Long Term Records.

Incident Energy (kWh/m^2)	Method
4.1	June 2005 from Cloud Radiation Model calculations
4.99	27 year average from Meteorological Office measured dataset
4.5	1999 average irradiation estimates based on sunshine records

Table 5.16: Calculated Versus Measured Inclined Irradiance and PV and AC Output

*Measurements started at 15:30. **Measurements ended at 10:30

Totals of I/O between 25/05/05* and 28/06/05**	
Inclined Global Irradiation, calculated (MWh)	13.1
Inclined Global Irradiation, measured. (MWh)	10.3
PV AC output, calculated (MWh)	1.3
PV AC output, measured. (MWh)	1.1

6 CLEAR-SKY DATA SELECTION AND MODELLING

6.1 *Introduction*

Energy engineering and building services applications require maximum load calculations. Terrestrial horizontal global irradiance is at its maximum in clear skies where the atmospheric transmittance is at its minimum.

The acquisition of clear-sky data is not done by direct measurement but rather by extraction of clear-sky data from all-sky datasets. Examining indicators of sky conditions such as clearness index and diffuse ratios are often used to determine data of particular sky condition. Finding a universal method to determine clear-sky data is difficult especially when relying on irradiation based indices. Often the upper and lower limits are location linked and differ due to meteorological and geographical variations between sites.

The scientific community has not yet established a proper, universal, classification of clear-skies for use in solar radiation modelling. Often, cloud-free skies are classified as clear-skies, while others classify low turbid skies as clear-skies. Cloud-free skies are determined by either measuring sunshine or by recording the cloud cover. Cloudless skies are therefore flagged when sunshine fraction is near unity and cloud cover nil.

The details of this lack of uniformity are observed in Section 2.8. Certain established procedures were reviewed and analyzed. Some new procedures were developed and compared to the established procedures; these will be put forward and discussed in this chapter.

As it was emphasized in earlier chapters, energy engineering and building services applications require maximum load calculations. Terrestrial horizontal global irradiance is at its maximum in clear skies where the atmospheric transmittance is at its minimum. When measured data is unavailable, designers and engineers require estimates of the clear-sky irradiance. Models that are based on geographic parameters have been proposed since the mid-20th century, yet these simple models are very inaccurate. In the past decade some more complex models have been created to estimate the absorption and scattering of solar radiation in the atmosphere in order to estimate, more accurately, the terrestrial solar irradiance. The two types of models are accurate for all-sky conditions; however this is not the case for clear-sky conditions. In cloudless skies, the contributors to the solar radiation extinction are ozone, water vapour, aerosols and other absorption and scattering elements in the atmosphere.

Four models were evaluated using the datasets extracted via the clear-sky procedure adopted in this chapter. The models under evaluation are the following: the Meteorological Radiation Model 'MRM' by Muneer et al. (1998), Gul et al. (1998) as well as Muneer and Gul (2000), the Page Radiation Model 'PRM', ESRA (2000), Yang's model by Yang et al. (2001) and REST2 by Guemard (2003, 2003, 2004, 2004).

6.2 Data acquisition

The datasets available for the present analysis cover three countries, each with specific meteorological and geographical characteristics. Hubbard (1994) demonstrated that the length of data series should be more than one year to characterize the seasonal pattern in special variability. Gueymard (1999) recommends that in order to validate radiation estimation models, a minimum three-year dataset is needed. Thus, the following six datasets presently used are based on a span of more than three years and are used for the comparison of the models.

One datasets came from a UK location, Bracknell (51.26N; 0.45E). The Bracknell dataset spanned six years (1990 to 1995). The UK location is an example of temperate, maritime weather found in northern Europe. The radiation and meteorological measurement station is classed as first class station, and is part of the UK Meteorological Office network. The radiation is recorded hourly from minute-by-minute averaged data. The cloud cover is recorded hourly using the Alidade unit. This unit allows the recording of the cloud cover and the height of the cloud base. Hourly sunshine fraction is also provided.

Data from two Southern European sites were also utilized, both sites being from Spain, i.e. Gerona (41.97N; 2.88E) and Madrid (40.45N; 3.73W). The Gerona dataset spanned seven years, covering the period from 1995 to 2001 and was provided by the University of Gerona, while the Madrid dataset spanned three years, covering 1999-2001 years. For both Spanish sites cloud cover data are recorded at 0700, 1300 and 1800 hours local clock time. The Instituto National De Meteorologia based in Madrid provided these data.

Data from three Indian sites were also used in the present work. All of these datasets were provided by the Indian Meteorological Department based in Pune. Data from Indian sites - Chennai (13.0N; 80.18E), Mumbai (19.12N; 72.85E) and Pune (18.53N; 73.85E) spanned five years covering 1990 to 1994. As with the Spanish sites, the Indian cloud cover data were also not recorded hourly. Rather this information was noted at 0830, 1130, 1430 and 1730 hours local time.

For both the Indian and Spanish sites, the cloud cover was observed and recorded by trained personnel. The hourly sunshine fraction was extracted from daily sunshine duration records for these sites.

The dry and wet bulb temperature, atmospheric pressure, and relative humidity are measure at the respective meteorological station of the sites under study. If the wet bulb temperature and the water vapour are not provided, the ASHRAE's (1993) method of estimating the wet bulb temperature and water vapor from dry bulb temperature and relative humidity.

The monthly averaged ozone optical depth and Linke turbidity were obtained from the SODA network (2002). The hourly Linke turbidity was calculated using the ESRA method at air mass=2 as described by Rigolier et al. (2000). The Angstrom Beta coefficient (β) was obtained from the Linke turbidity using Grenier et al. (1994) method.

All the datasets used in this study have been quality controlled using the Younes et al. (2005) procedure, also described in Chapter 3. This procedure relies on multiple physical tests to remove outliers that are caused by operation errors and then a statistical test to remove unacceptable errors due to instrumentation errors as discussed by Myers (2005).

6.3 Clear-sky identification procedures.

The definition of clear skies is very loosely used in solar radiation modelling. Quite often, skies are described, as clear when in fact the skies are cloudless, thus clear-sky is in reality quasi-clear-sky. However, very few researchers provide a clear identification as to whether low and high turbid skies, under zero cloud cover should qualify as clear skies or only low turbid cloudless skies should be tagged as clear.

It is important to note that clear-sky irradiance data are extracted from long all-sky irradiance datasets. To this end, different methods have been used to classify sky conditions from the available synoptic parameters.

When cloud cover information is available with the irradiance components, cloudless skies – 0 Okta – are assumed representative of clear skies, hence the irradiance is assumed clear-sky. Sunshine fraction has often been used to classify sky conditions too. A sunshine fraction close to unity indicates the associated irradiance data to belong to clear-sky conditions. Either of the two parameters has been used independently to this end; however some people have combined both to get more accurate indication of clear-skies. Some other researchers have used the clearness index and diffuse ratio dimensionless plan to identify sky conditions. High k_t and low k values are representative of clear-skies. Low turbidity skies are often identified using coefficients that quantify aerosols in the atmosphere. As clouds count as aerosols, it is concluded therefore that very low Linke turbidity coefficients are representative of cloudless low turbid skies, often referred to as absolute clear skies.

Lam and Li (2001) adopted an absolute and restrictive limit for cloud cover being nil to represent cloudless skies. However to compensate for the bias accompanied by potential errors in recording, cloudless skies are considered to be less than one Okta. Babaro et al (1981) have addressed matters as well and have classified clear skies in the cloud cover range of 0 to 2 Oktas, inclusive. This proposed limit of cloud cover less than 1 Okta is considered as the first clear-sky test, presently in this study.

Sunshine based classification of clear skies has also been adopted in literature, in which case a sunshine fraction close or equal to unity would represent a cloudless and thus a clear sky. In fact, this procedure has been discussed by Littlefair (1988), Muneer (2004), and Lam and Li (2001) who commonly agree that a value greater than 0.9 is representative of the desired sky. They have also discussed the shortcomings of this method and have concluded that sunshine fraction only indicates whether the sun is blocked by the cloud and does not provide information relative to the other parts of the sky. In this study, sunshine fraction as a determinant of sky condition will be evaluated namely as clear-sky test 2.

A combination of both cloud cover and sunshine fraction data could also be used to determine more accurately the sky conditions. It is important to note that the two parameters are interlinked statistically as was shown by Page in CIBSE Guide J (2002). This procedure will be referred to in the present study as clear-sky test 3.

The typical methods of clear-sky identification using sky clarity indexes have been discussed in Section 2.8. It was concluded from literature that using one index is not enough to classify the sky conditions thus the use of two indexes is necessary.

Battles et al. (1998) have developed a lower limit for k_t (k_{tt}) and an upper limit for k (k_k). Both limits are functions of the solar geometry. The Battles method will be referred to as clear-sky test 4 in this study. It is mathematically represented in Eqs. 2.34, 2.35.

The atmospheric turbidity may also be used to evaluate and classify the sky conditions. A Linke turbidity T_{LK} value of less than 3 has been described as representative of clear skies, and a value of near unity being very clear skies . For the clear-sky test 5, a T_{LK} value of less than 2.5 was used to describe clear sky conditions. However, for some locations it was found that the monthly averaged T_{LK} value was less than 2.5, in such cases the monthly-averaged value was used as the upper limit. This was discussed in ESRA, by Greif and Scharmer (2000).

Lam and Li (2001) have introduced a method to determine the upper and lower limit of k and k_t, respectively, by means of analyzing the cumulative frequency of occurrence in percentage of cloud cover, sunshine fraction, k_t and k. From the cloud cover analysis, they have concluded that there was a good statistical concordance between the cloud cover and the above-mentioned indices. Figures 2.7 a-c represented the cumulative frequency distribution of the cloud cover, clearness index and diffuse ratio, respectively for Chennai, a site used in this study. In the case of Figures 2.7 a-c at 9% cloud cover less than 1 Okta, the k_t minimum limit is 0.78 and the k maximum limit is 0.21. The relevant details for other sites are given in Table 6.1.

Three clear-sky tests have been derived from the procedure described by Li and Lam (2001) to address certain irregularities in the procedure. Clear-sky test 6 applies the k_t and k limits as found in Table 6.1 with the addition of a cloud cover limit as in the clear-sky test 1. The cloud cover limit is dropped and replaced with a Linke turbidity coefficient limit, as in clear-sky test 5, in clear-sky test 7. Finally, both the cloud cover limit and Linke turbidity coefficient were added to the k_t and k limits in clear-sky test 8.

Long and Ackerman (2000) have derived a new method of determining clear-sky data from I_G (See Eq.2.40).

Thus, in total, there are nine clear-sky tests that are under evaluation in this study. These tests are presented in Table 6.2.

6.4 Discussion of clear-sky tests

All six datasets have been independently used in the evaluation of the clear-sky tests mentioned in Table 6.2. The results of the tests are given in Table 6.3 in percentage of points that passed the clear-sky filtering process. The results are dependent on the respective climatology of the locations. Generally sunnier, less cloudy and less polluted sites produced more clear-sky hours than their cloudier, heavy polluted counterparts did.

It was observed that the first four tests were not very restrictive as shown in Figs.6.1a-e, k_t - k scatter plots for Mumbai. In Fig.6.1b it was noted that eliminating cloudless skies based on cloud cover alone is not a good method to identify clear-skies as there is a wide scatter in the low k_t and high k zone. Using the sunshine fraction, to define clear-skies was more accurate than solely using cloud cover, except in the case of Bracknell. However, for the same reason as in clear-sky test 2, sunshine fraction cannot be used as shown in Fig.6.1c. Combining the cloud cover and the sunshine fraction resulted in even more precision, but as observed in Fig.6.1d, for the high k_t values, there is a large scatter in the high values of k. The Battles et al. (1998) test was found least accurate to identify clear-sky hours. This concludes that using synoptic data by itself is not an accurate method to determine clear-sky conditions. In this same respect, one can conclude as well from observing

Fig.6.1a that using the diffuse ratio or the clearness index alone cannot yield a true sky classification method as discussed by various researchers and referenced in Chapter 2, Section 2.8. It was therefore safe to exclude the use of a single sky index as a clear-sky test.

The Long and Ackerman (2000) test yielded very mixed results. This is due to the low resolution of the datasets used. From personal communication, it was found that the above-mentioned procedure works well for sub-15 minutes resolution datasets and is unreliable for hourly resolutions. The results of clear-sky test 9 are given in Figs.6.2a-d for Pune. Notice that the points that passed this test are in the high k_t zone as shown in Fig.6.2a, and most data points are located under the monthly averaged Linke turbidity values in Fig.6.2c. This procedure worked only for the Indian sub-continent locations, with mixed results. For all other sites the test identified no clear-sky points, as can be observed in Table 6.3.

It was also found that by selecting only low turbid skies as a means of identifying clear-skies does not equate to cloudless skies. Thus it was concluded that using the Linke Turbidity as means of sky classification is not efficient.

Figures 6.3a-e represent the evaluation of tests 6-8 for Bracknell. In test 6, the clear-sky data points that were identified are all in the correct- high k_t, low k – zone as observed in Fig.6.3a. However, very few of the data points identified are in the low turbidity region as shown in Fig.6.3b. This was partially addressed in test 7 Fig.6.3c, however the clear-sky hours identified do not belong to cloudless skies as can be seen in the cloud distribution histogram in Fig.6.3d. Cloudless skies and low turbid atmosphere clear-sky hours were identified in test 8 as shown in Fig6.3e.

Test 8 was found to be the most appropriate to identify cloudless and low turbid skies, i.e. clear sky conditions.

6.5 Clear-Sky models reviewed

The meteorological radiation model, by Muneer et al. (1998), Gul et al. (1998) as well as Muneer and Gul (2000), has been developed to estimate all-sky beam, global and diffuse irradiance. In this study, only the clear-sky section of the model is evaluated.

Similarly, Page's radiation model, by ESRA (2000) has been developed to estimate all-sky irradiation components. The model's clear-sky section was evaluated in this study.

Yang's model developed by Yang et al. (2001) is a hybrid model and is composed of two parts. The first part deals with clear-sky conditions and is based on Yang's own work, while the second part, is an adaptation of Angstrom's model (1924). Only the clear-sky model is evaluated herein.

REST2 is a clear-sky model developed by Gueymard (2003, 2003, 2004, 2004) to estimate broadband irradiance.

MRM is the only site dependent model used in this study. The model had to be locally tuned for each location. The coefficients used are presented in Table 6.4. In a study of all-sky MRM conducted by the Muneer and Younes (2005) and described in more details in Chapter 4, it was found that coefficients differed significantly for Chennai and Pune (Indian locations). The sites used in the current study are derived from those used in the all-sky evaluation of MRM by Muneer et al. (2005) and in Chapter 4, therefore a similar observation was made in the clear-sky analysis.

Further details about the models under evaluation are found in Section 2.9.

6.6 Evaluation of clear-sky models

To evaluate the models, the same statistical indicators used in Chapter 5 were utilized. Thus, the slope and coefficient of determination of the best fit line, between the measured and calculated irradiance, were obtained. As well as the root mean square error and mean bias error in W/m^2 rather than percentage. In addition the kurtosis and skewness of the estimation error histograms were calculated. In order to streamline the comparison process, the Accuracy Score procedure adopted in Chapter 5, and explained in Section 5.5, was used.

MRM was found to have performed with good accuracy for the estimation of I_G, I_B and I_D as can be observed in Table 6.5 giving the ranges of R^2, Δ Slope, AMBE, RMSE, ASkewness and kurtosis. The average MBE and RMSE for I_G was -1.6 and 85 W/m^2. Calculated versus measured beam, diffuse and global horizontal irradiances scatter plot can be observed in Figs.6.4a-c for Bracknell.

The Page radiation model was less efficient than MRM in the estimation of the solar terrestrial irradiance. The model yielded very high residuals and considerable underestimation of all three irradiation components. The statistical indicators averages are shown in Table 6.6. Similarly to MRM, the calculated versus measured beam, global and diffuse horizontal irradiances scatter plot can be observed in Figs.6.5a-c for Bracknell.

It was found that the Yang radiation model performed well in the estimation of beam and global irradiance as can be observed in Fig.6.6 for Gerona. The evaluation result was poor for the diffuse irradiance as is observed from the averages of the statistical indicators, shown in Table 6.7.

The REST 2 model performed satisfactorily for all six datasets considering the low accuracy of certain input parameters supplied that are required for the model, this can be observed in Fig.6.7 for Gerona. The averages of the statistical indicators are shown in Table 6.8.

The four models were compared using the AS procedure for the three irradiance components. The results are given in Tables 6.9 a-c for I_B, I_G and I_D, respectively. Based on those AS results, it was observed that the MRM performed best in estimating I_B and I_G and REST2 for I_D. The overall results are shown in Fig.6.8, where REST2 performs best, followed by MRM, YRM and PRM for Bracknell, and MRM performs best followed by REST2, YRM and PRM for the other five sites. The overall AS results for all sites and all three irradiance components are as follows, 77.7, 70.9, 59.2 and 27.6 for MRM, REST2, YRM and PRM respectively. However note that MRM was locally tuned for the sites, while the other models do not require local tuning. In this respect REST2 and YRM performed very well.

6.7 Effect of improper clear-sky data selection on modelling

Alam (2005) had performed a clear-sky evaluation of Gueymard's (2003, 2003, 2004, 2004) CPCR2 and REST models in addition to Yang's model for Indian climatology at four locations. Note that the REST model is the predecessor of the REST2 model under evaluation in this study. In his study, the predicted DNI was compared with reference DNI that was estimated from measured global and diffuse radiation. The REST model showed maximum RMSE of 6.5% in the prediction of DNI as compare to more than 10% errors in Yang and CPCR2 models. The predicted global radiation showed a maximum RMSE of 7% in REST model, 13.4% in Yang model and 25.9% in CPCR2 model. Alam concluded that the REST model was the most appropriate for the Indian climatology. It was also concluded, given the input parameters used in this study, that REST2 is the most appropriate model for the Indian Sub-Continent.

Previously, Muneer (2004, 1998) compared the estimation of I_G by using the cloud radiation model 'CRM' also referred to as the Gul et al. (1998) model in Chapter 5, MRM and PRM for 5 UK locations: Aughton, Bracknell, Finningley, London and Stornoway. Muneer evaluated the three models for three sky conditions: quasi-overcast-, mixed- and quasi-clear-skies. The quasi-clear-sky zone was selected to be for $k_t \geq 0.6$. On average, the models performed satisfactorily with MBEs and RMSEs averaging at 78.6-, 45- and 90.6 W/m^2 and 115.8-, 77.6- and 163 W/m^2 for CRM, MRM and PRM, respectively. Muneer concluded from the statistical indicators of his study that for global irradiance the meteorological radiation model performed satisfactorily for UK climatology. He also concluded, based on a crude scoring procedure, described in Section 4.4, that MRM outperformed PRM under his pre-defined clear-sky conditions. The scores are as follows, 613 W/m^2 for MRM and 1268 W/m^2 for PRM.

The use of the clear-sky datasets for the MRM yielded results that were found to be better than the evaluation by Muneer (2004) in which, clear-sky data was selected at $k_t \geq 0.6$. This is due to the selection of very clear-skies conditions for the evaluation rather than quasi-clear-sky conditions. Note that using very clear-skies, i.e. cloudless and low turbid sky conditions, PRM did not perform as well, when quasi-clear-sky data were utilized.

6.8 Conclusions

Various researchers, for modelling purposes, adopted different procedures to identify clear-sky conditions from large all-sky databases. Some procedures relied on cloud or sunshine data to extract the clear-sky irradiance, while others relied on sky clarity indexes.

The synoptic procedures were found inaccurate in identifying the clear-sky solar irradiance. Even when associating the two tests of cloudless skies and complete sunshine hours, the associated clearness index to diffuse ratio plots showed scatter in the mixed- to quasi-clear- skies zones.

Various literatures have shown that using one sky clarity index is not enough to categorize the sky conditions; therefore an association of two indexes can be used to determine clear-sky data. Yet, the information, found in literatures, seldom agrees on a fixed classification using the sky indexes. Recently researchers have agrees that the diffuse ratio and clearness indexes when used to define clear-skies are variable depending on seasons and on geographic location.

The Lam and Li (2001) approach of setting clear-sky limits to the diffuse ratio and clearness index based on the frequency of occurrence of cloudless skies, was investigated. It was concluded that on it's own it was unable to achieve clear-skies. By adding a cloudless sky condition to the above method, cloudless skies can be achieved. However further analysis of the data show that the resultant data after the clear-sky test lays in a mixed turbid region.

Here is the main problem facing solar radiation modellers, should one use cloudless skies with mixed turbidity as clear-skies, or should one adopt a more restrictive cloudless skies associated with low turbid condition to associate with clear sky conditions? Note that cloudless low turbid skies are often referred to as blue skies or absolute clear-skies.

Turbidity as a clear-sky classifier was attempted; however it was found that low turbid skies are not always associated to cloudless skies. Therefore using turbidity solely was not accurate enough to be used as a clear-sky identifier.

Combining the clearness index and diffuse ratio, cloudless sky limits with a filter to assure cloudless skies, as associate a filter to select only low turbid skies, assures that cloudless, and low turbid conditions are obtained, thus very selective/restrictive clear-sky conditions achieved.

Using the clear-sky datasets, obtained, four established clear-sky solar radiation models were compared. The models evaluated are as follows: the Meteorological Radiation Model 'MRM' by Muneer et al. (1998), Gul et al. (1998) as well as Muneer and Gul (2000), the Page Radiation Model 'PRM', ESRA (2000), Yang's model 'YRM' by Yang et al. (2001) and REST2 by Guemard (2003, 2003, 2004, 2004).

By use of a previously developed scoring system, the Accuracy Score, the four models were ranked as follows: First MRM, second REST2, third YRM and finally PRM. Note that MRM, in this study, has been locally tuned. For non-tuned models, REST2 would be the most accurate model that may be used to estimate clear-sky broadband horizontal irradiance.

The effects of the use of different clear-sky databases are observed by comparing the current study with the study by Muneer et al. (1998). It was found that for the same locations, the use of extreme clear-sky datasets improved the estimation of solar irradiation of the MRM. It, however, had an adverse effect on the PRM. Thus it can be deduced that the MRM performs very well in extreme sky clarity, and PRM performs well in quasi-clear skies.

Even though the inputs used for the REST2 and Yang model are not as accurate as the models require, it was found that by using extreme clear-skies data, the models performed exceptionally well.

a-c

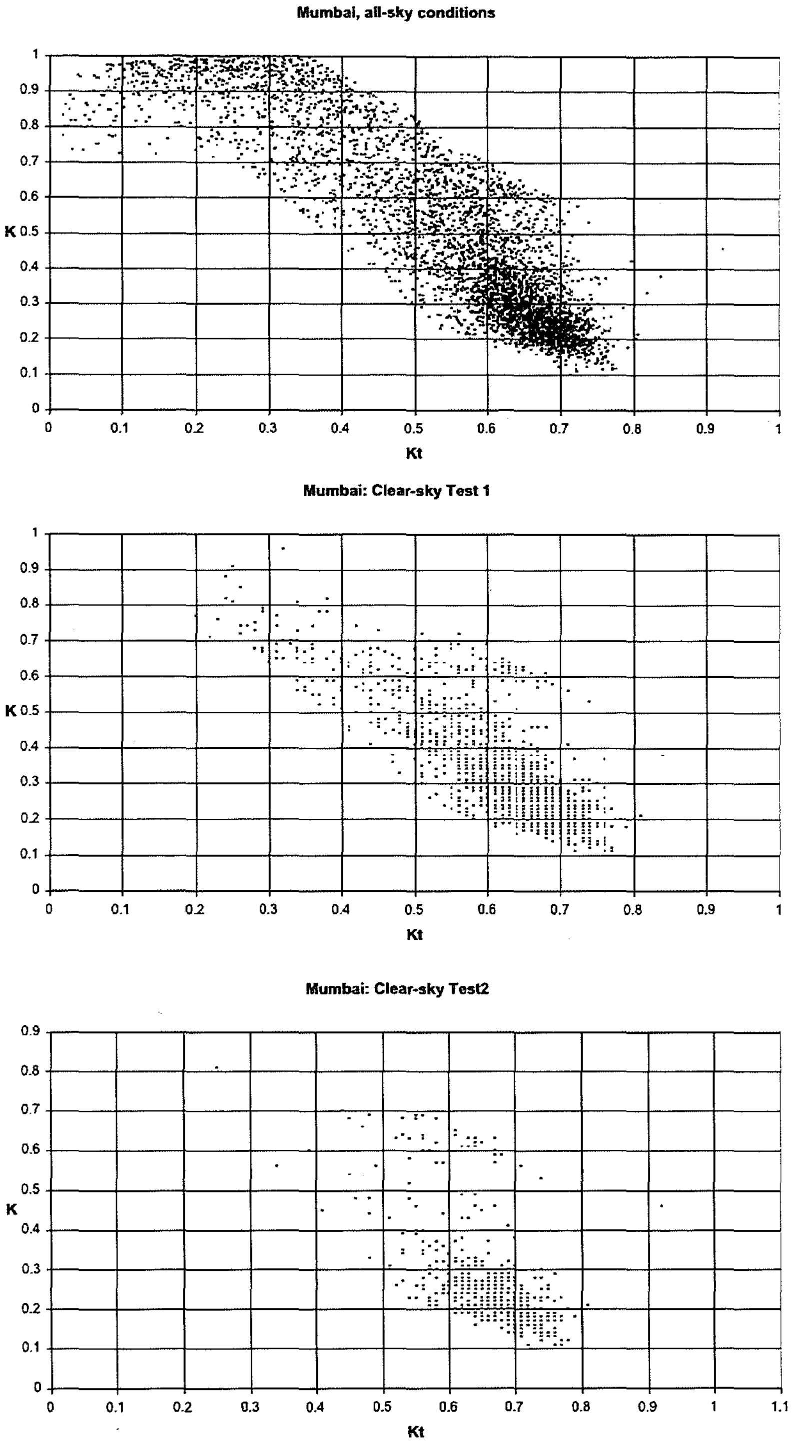


Figure 6.1 Clearness index - diffuse ratio scatter for Mumbai. a. for all sky conditions; b. after clear-sky test 1; c. after clear-sky test 2;

Fig. 6.1- Continued

d-e

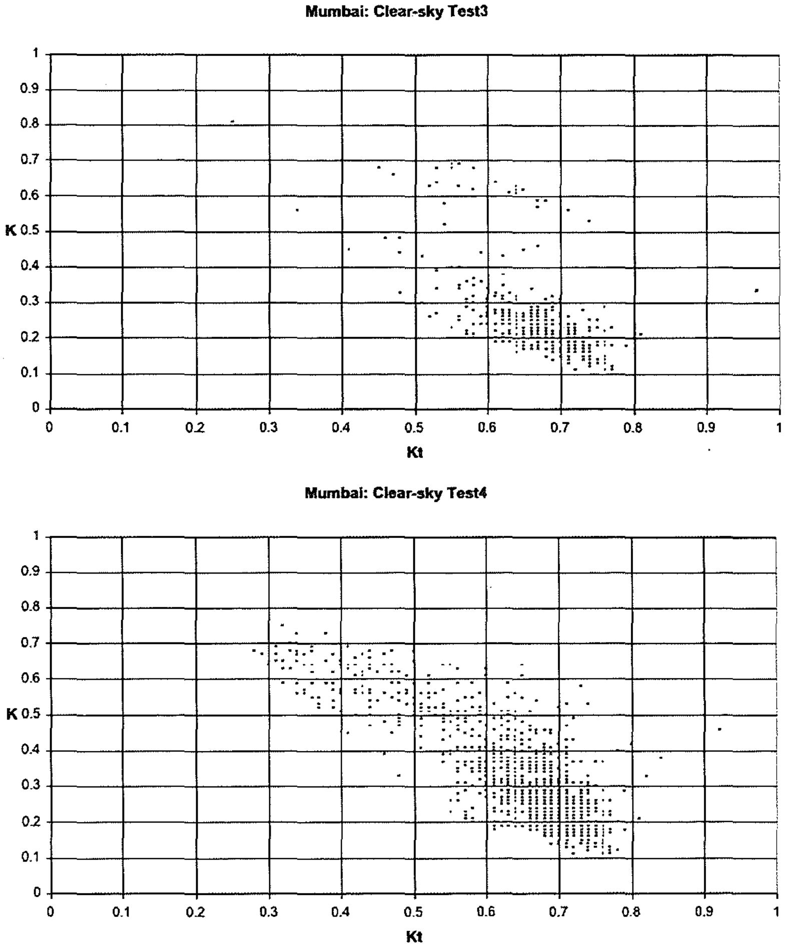


Figure 6.1 Clearness index - diffuse ratio scatter for Mumbai. d. after clear-sky test 3; e. after clear-sky test 4.

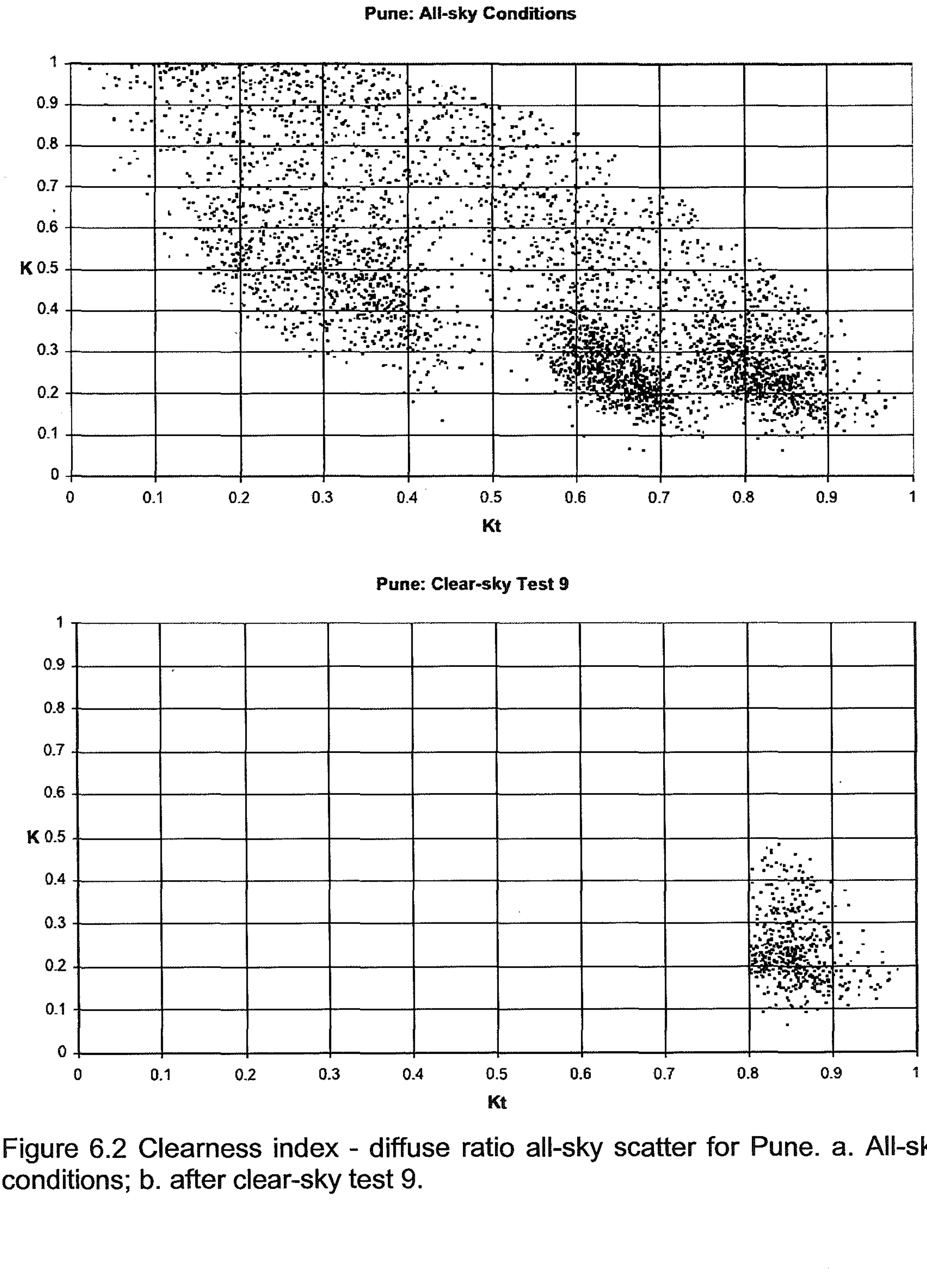


Figure 6.2 Clearness index - diffuse ratio all-sky scatter for Pune. a. All-sky conditions; b. after clear-sky test 9.

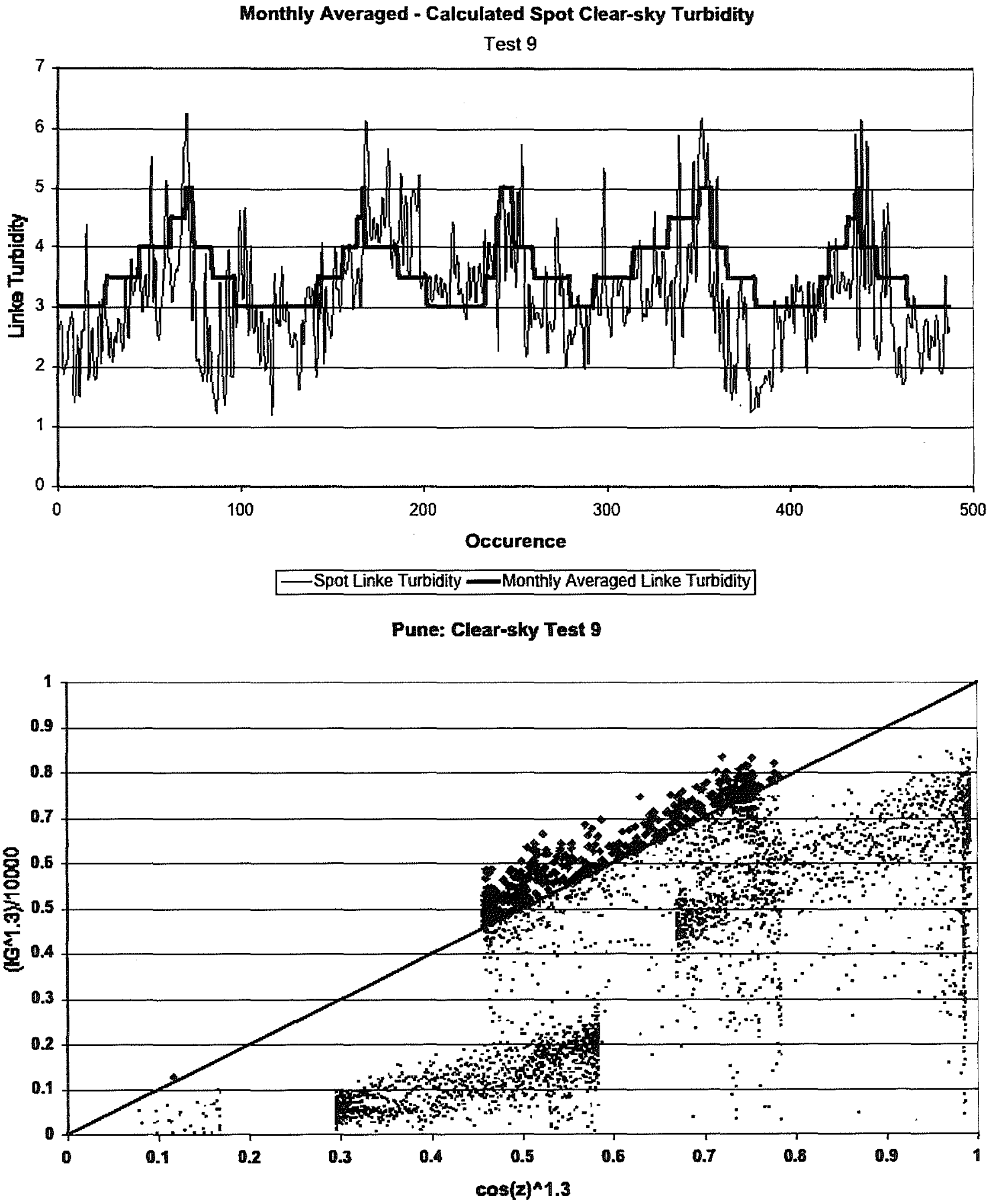


Figure 6.2 c. Linke turbidity profile after clear-sky test 9 for Pune; d. scatter plot of function represented by Eq. 2.40. Points in larger bold format are considered to be clear-sky.

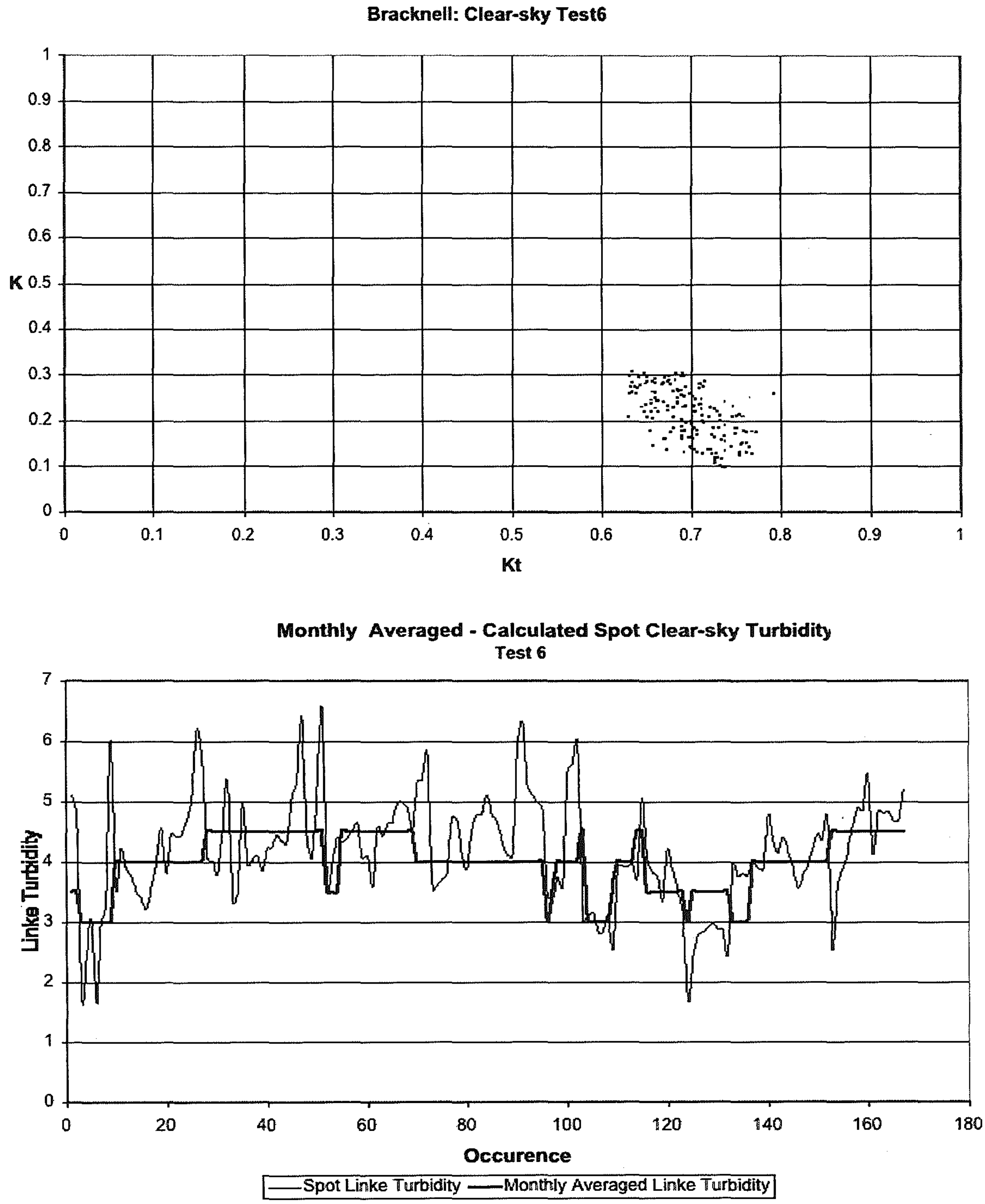


Figure 6.3 After clear-sky test 6 for Bracknell a. Clearness index - diffuse ratio scatter; b. Linke turbidity profile.

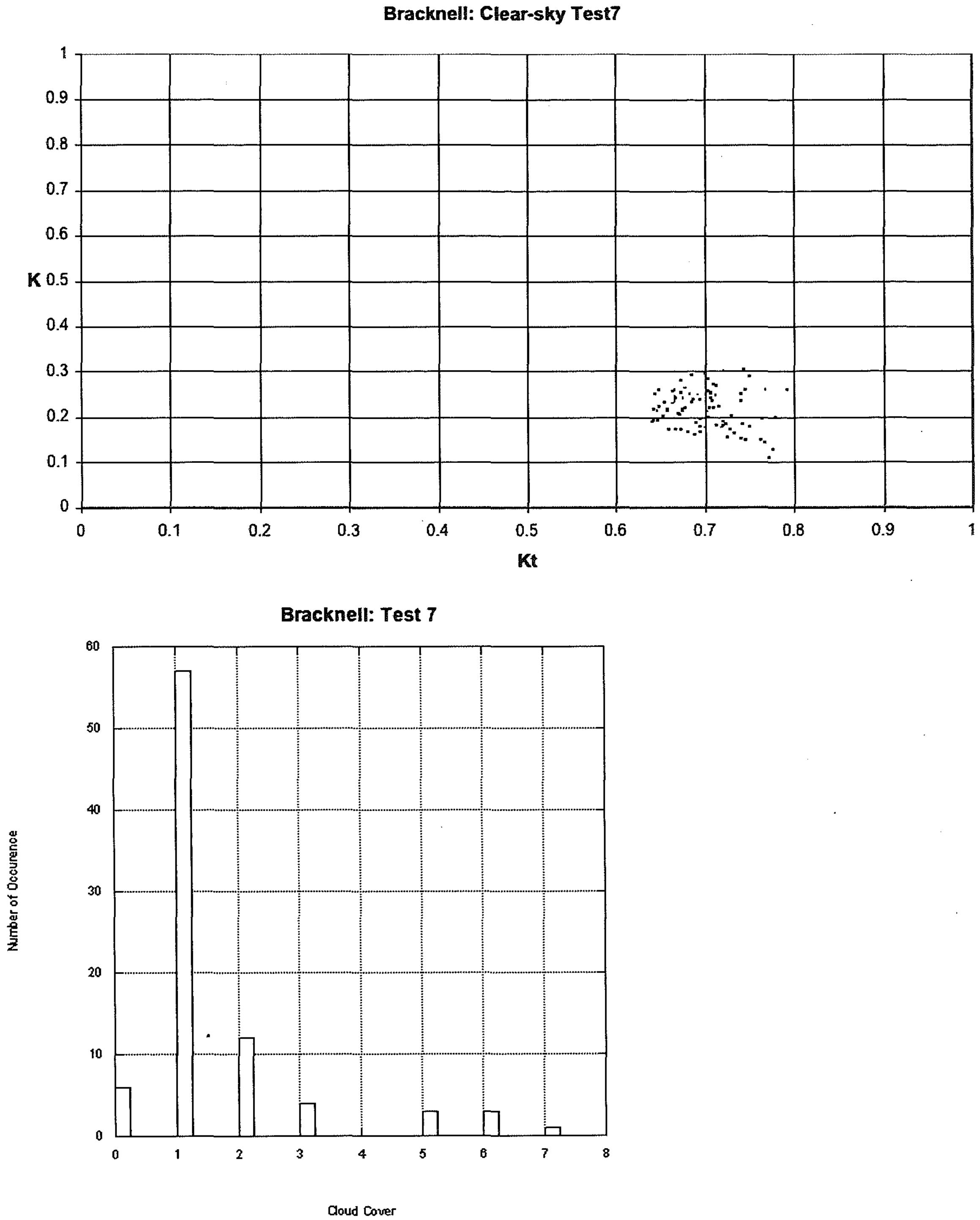


Figure 6.3 After clear-sky test 7 for Bracknell; c. Clearness index - diffuse ratio scatter; d. Cloud cover histogram of occurrence after clear-sky test 7 for Bracknell.

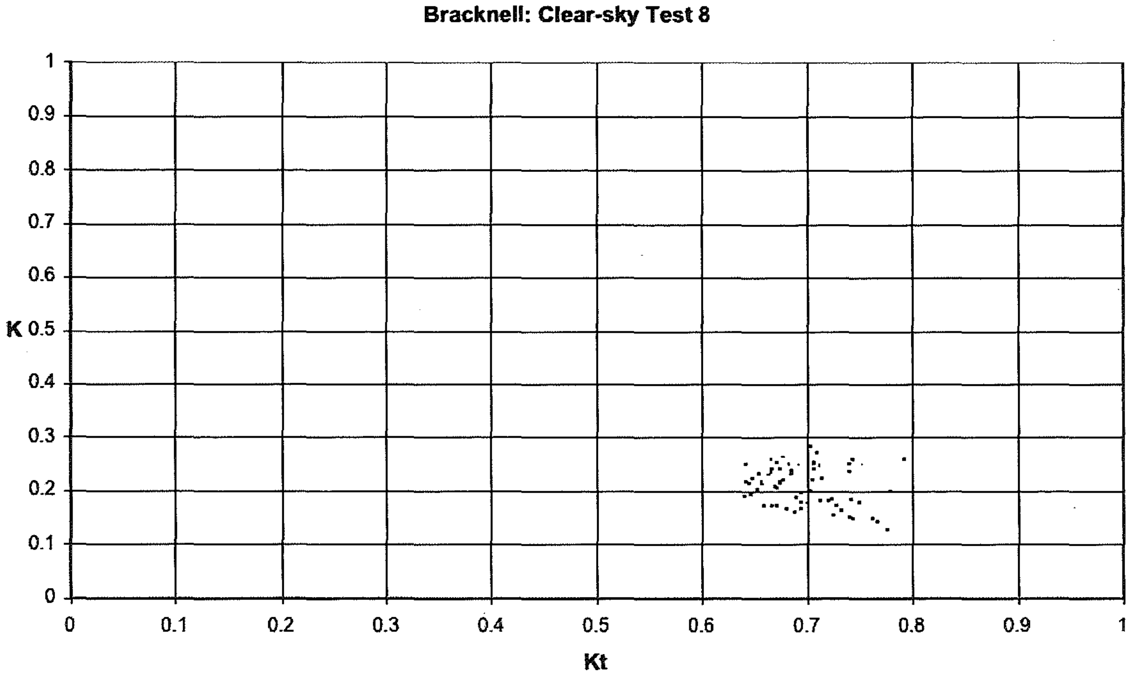


Figure 6.3e: Clearness index - diffuse ratio scatter after clear-sky test 8 for Bracknell.

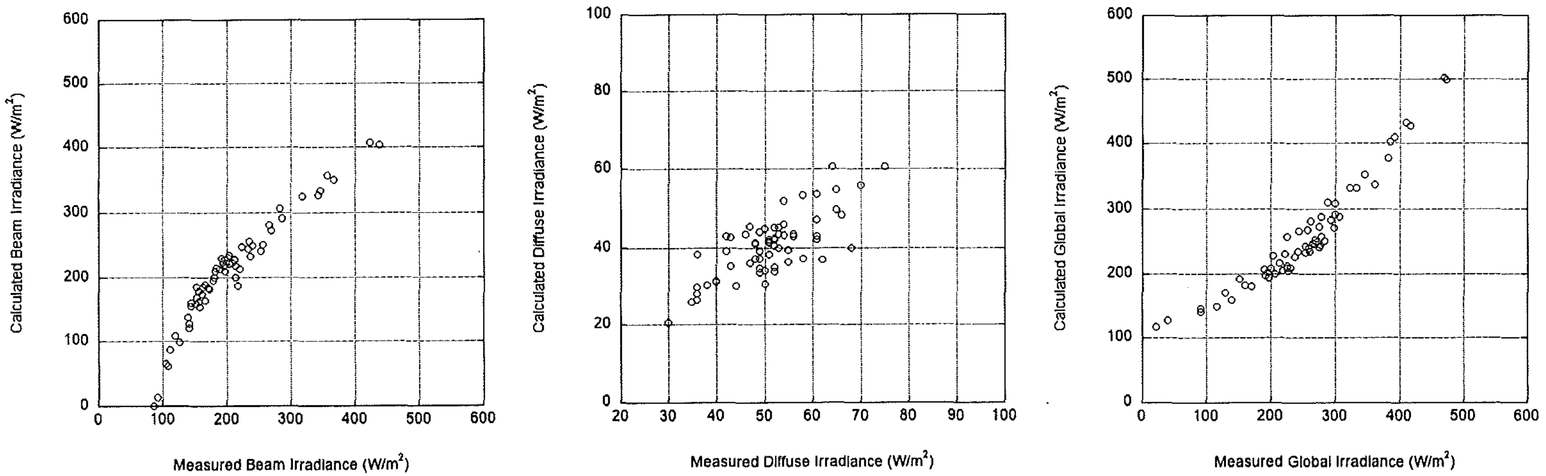


Figure 6.4 Evaluation of MRM estimation for Bracknell. a. Beam irradiance; b. diffuse irradiance; c. global irradiance.

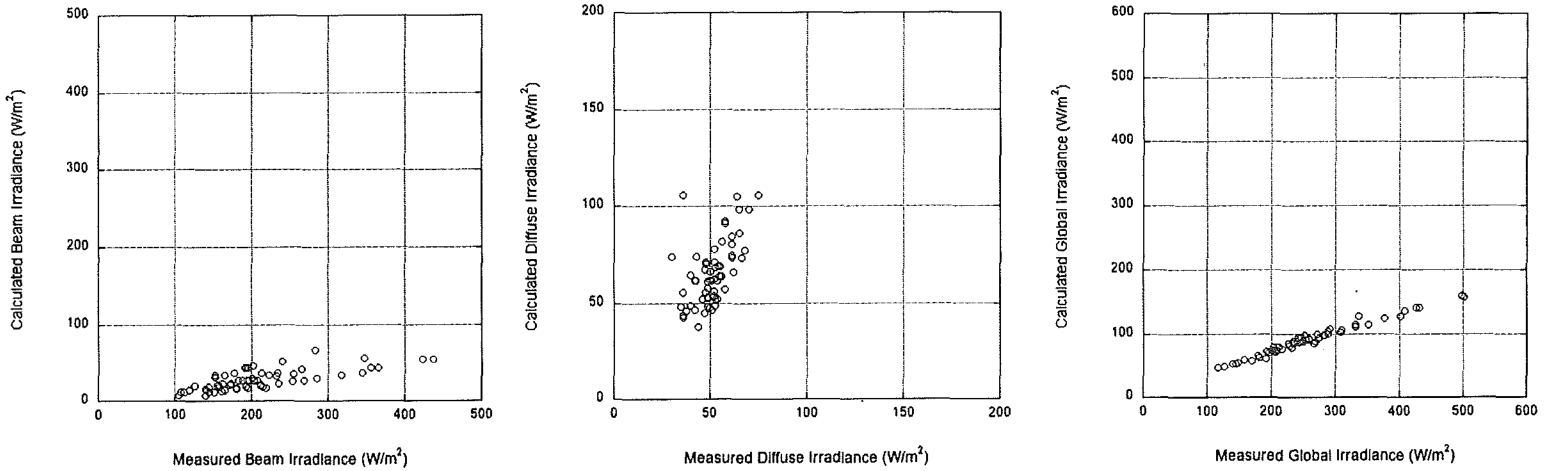


Figure 6.5 Evaluation of PRM estimation for Bracknell. a. beam irradiance; b. diffuse irradiance; c. global irradiance.

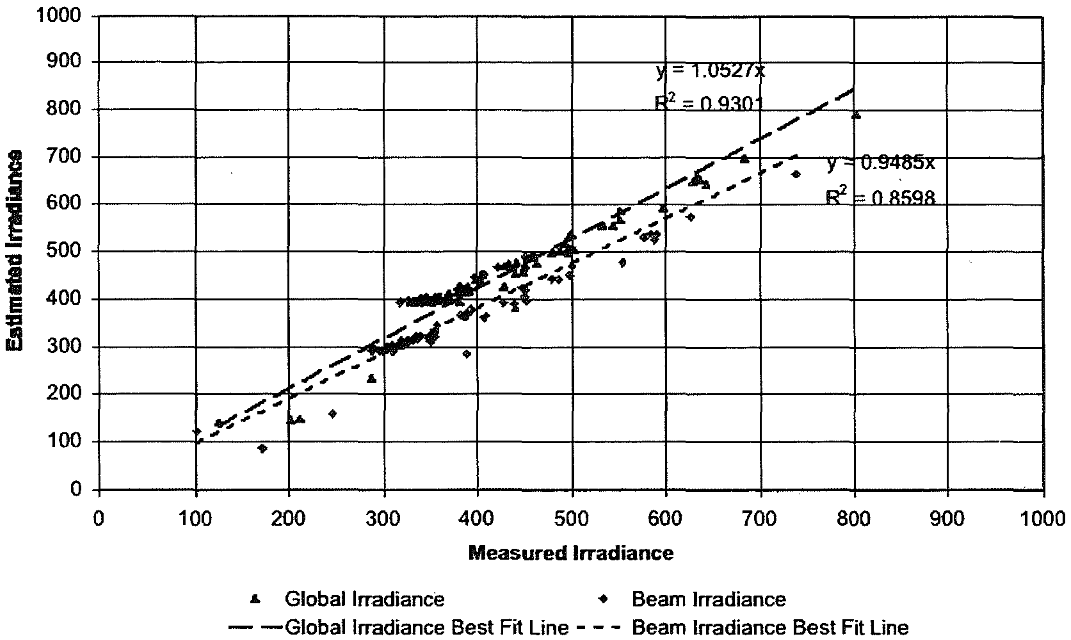


Figure 6.6 Evaluation of the Yang radiation model estimation of beam and global irradiance for Gerona.

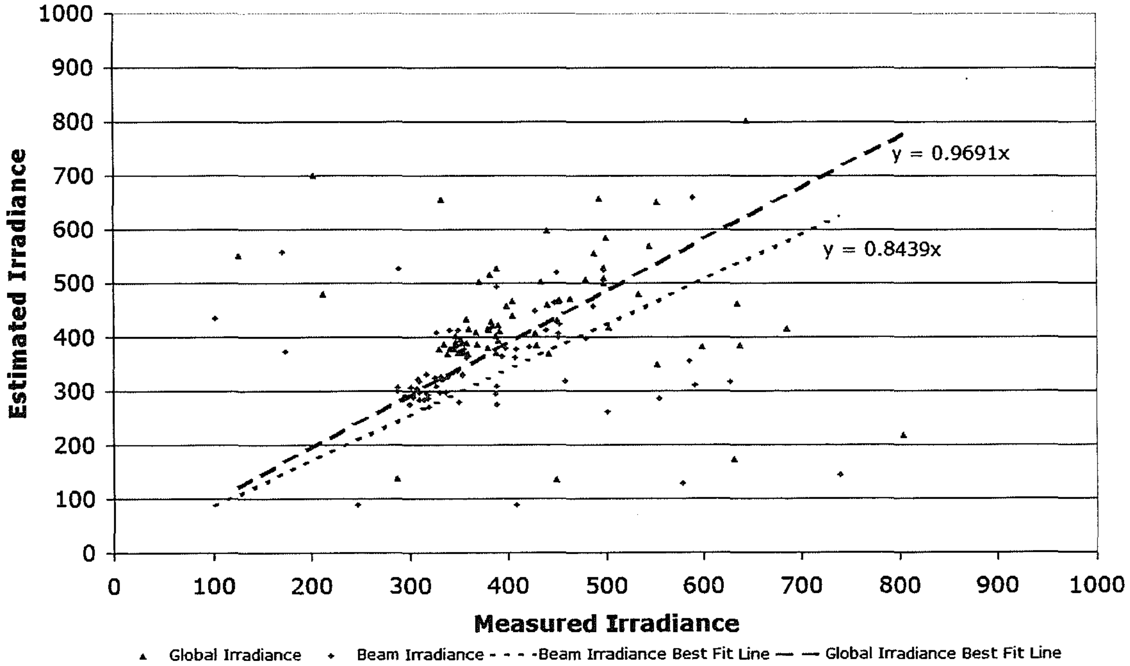


Figure 6.7 Evaluation of REST2 estimation of beam and global irradiance for Gerona.

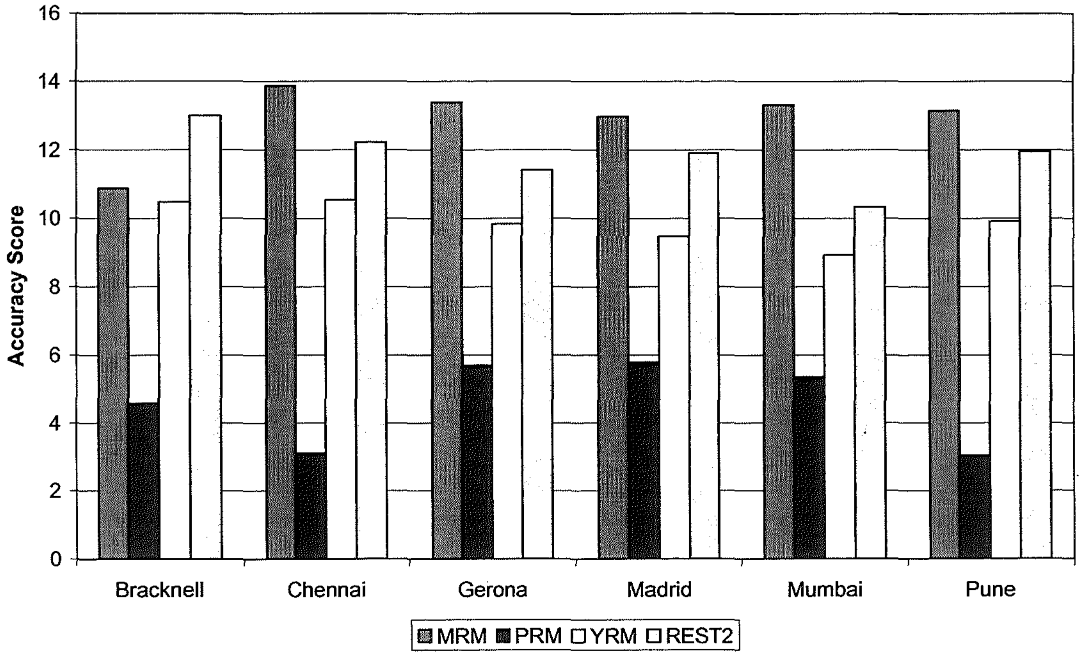


Figure 6.8 Accuracy Score results for the four models presently evaluated.

Table 6.1 kt & k limits based on cloud cover frequency diagrams for clear-sky tests 6, 7 and 8.

	Cumulative frequency plot results		
Site name	Cloud Cover ≤1	Kt, min	K, max
Bracknell	12%	0.61	0.26
Chennai	9%	0.78	0.21
Gerona	32%	0.60	0.21
Madrid	43%	0.62	0.19
Mumbai	39%	0.58	0.31
Pune	40%	0.59	0.28

Table 6.2 Description of clear-sky tests.

Test Number	Description
1	Cloud cover <1
2	Sunshine fraction ≥0.9
3	Cloud cover <1 and Sunshine fraction ≥0.9
4	Battles et al. (1998) clear sky procedure (See Eqs. 2.34 and 2.35)
5	$T_{LK} < 2.5$
6	k_t and k limits and Cloud cover <1 (See Table 6.1)
7	k_t and k limits and T_{LK} <2.5 (See Table 6.1)
8	k_t and k limits and Cloud cover <1 and T_{LK} <2.5 (See Table 6.1)
9	Long and Ackerman (2000) clear-sky procedure (See Eq. 2.40)

Table 6.3 Percentage of points that passed each clear sky test.

Clear-sky tests	Bracknell	Chennai	Gerona	Madrid	Mumbai	Pune
Total no. of points*	22111	4203	3931	3439	4011	4008
Test 1	11.70%	9.14%	32.26%	43.09%	39.04%	40.39%
Test 2	28.95%	4.71%	1.40%	38.88%	12.76%	6.39%
Test 3	10.05%	1.40%	1.07%	29.63%	10.00%	5.14%
Test 4	0.00%	18.46%	55.99%	71.56%	29.00%	25.67%
Test 5	0.57%	0.40%	3.08%	4.22%	0.30%	2.69%
Test 6	0.76%	0.14%	10.07%	16.02%	17.63%	17.84%
Test 7	0.39%	0.21%	2.62%	3.72%	0.15%	2.67%
Test 8	0.28%	0.02%	1.86%	2.38%	0.15%	1.75%
Test 9	0.00%	0.45%	0.00%	0.00%	0.02%	12.13%

*Total number of points with co-incidence solar radiation and cloud cover/ sunshine data.

Table 6.4 Coefficients used for clear-sky Meteorological Radiation Model.

Coefficients	Bracknell	Chennai	Gerona	Madrid	Mumbai	Pune
COF(1)	2.11	1.97	2.12	2.12	2.12	2.16
COF(2)	0.35	0.09	0.37	0.37	0.37	0.44
COF(3)	1.06	1.27	1.07	1.06	1.06	1.01
COF(4)	0.93	0.89	0.88	0.88	0.88	0.95
COF(5)	-0.12	-0.22	-0.07	-0.07	-0.07	-0.13
COF(6)	-0.08	0.11	0.05	0.06	0.06	0.09
COF(7)	-0.02	0.26	0.04	0.04	0.04	0.20
COF(8)	0.00	0.04	0.00	0.00	0.00	0.03
COF(9)	3.43	3.43	3.43	3.42	3.43	3.41
COF(10)	77.02	77.02	77.03	77.03	77.03	77.03
COF(11)	0.55	0.98	0.64	0.64	0.60	0.71
COF(12)	3.37	3.38	3.37	3.37	3.37	3.38
COF(13)	0.00	0.00	0.00	0.00	0.00	0.00
COF(14)	-0.05	-0.04	0.11	0.25	0.00	-0.01
COF(15)	0.06	-0.19	-0.02	0.08	-0.08	-0.23
COF(16)	0.91	2.17	1.31	1.17	2.46	2.07
COF(17)	0.06	45.33	0.55	1.00	0.19	-55.93
COF(18)	0.17	-141.53	0.17	0.17	0.17	225.62

Table 6.5 Ranges of statistical indicators for the MRM evaluation for all six sites.

Irradiance		R^2	Δ Slope	A-MBE*	RMSE	A-Skewness**	Kurtosis
I_B	Min	0.63	0.02	1.6	25.3	0.03	-0.50
	Max	0.92	0.26	97.8	141.3	1.39	2.54
I_G	Min	0.72	0.00	0.2	26.7	0.07	-0.78
	Max	0.94	0.03	5.6	135.9	1.16	2.01
I_D	Min	0.18	0.02	0.0	12.1	0.07	0.19
	Max	0.76	0.21	10.5	53.1	1.98	6.11

* A-MBE is the absolute mean bias error.

** A-Skewness is the absolute skewness of errors.

Table 6.6 Averages of statistical indicators for the PRM evaluation for all six sites.

Averages	I_B	I_G	I_D
R^2	0.53	0.79	0.45
Slope	0.22	0.41	0.37
MBE	313.9	289.7	119.5
RMSE	362.4	336.4	128.6
Skewness	0.25	0.12	0.74
Kurtosis	-0.69	-0.69	1.09

Table 6.7 Averages of statistical indicators for the Yang radiation model evaluation for all six sites.

Averages	I_B	I_G	I_D
R^2	0.78	0.88	0.17
Slope	0.88	0.94	1.29
MBE	51.7	52.4	-9.8
RMSE	68.4	71.1	51.6
Skewness	-0.18	0.19	0.96
Kurtosis	0.59	0.32	0.87

Table 6.8: Averages of statistical indicators for the REST2 model evaluation for all six sites.

Averages	I_B	I_G	I_D
R^2	0.93	0.94	0.95
Slope	1.03	0.97	0.72
MBE	16.8	-8.7	-25.5
RMSE	168.7	183.7	39.8
Skewness	0.49	0.42	0.34
Kurtosis	2.04	2.07	0.33

Table 6.9 Accuracy score for a. I_B; b. I_G; c. I_D.

	MRM	PRM	YRM	REST2
Bracknell	3.5	0.8	4.0	5.1
Chennai	4.8	1.2	3.6	3.3
Gerona	4.6	1.9	4.4	3.4
Madrid	4.1	1.8	3.6	3.9
Mumbai	5.1	2.1	2.8	3.5
Pune	5.0	0.8	3.5	3.6
Total	27.0	8.6	21.9	22.8

	MRM	PRM	YRM	REST2
Bracknell	3.5	1.7	4.4	4.1
Chennai	4.3	1.1	3.9	3.8
Gerona	5.4	2.4	3.5	3.9
Madrid	5.1	2.2	4.2	3.9
Mumbai	5.1	2.0	3.2	2.5
Pune	4.8	0.8	4.3	3.6
Total	28.0	10.1	23.5	21.7

	MRM	PRM	YRM	REST2
Bracknell	4.0	2.1	2.1	3.8
Chennai	4.9	0.8	3.1	5.1
Gerona	3.5	1.4	1.9	4.2
Madrid	3.8	1.8	1.7	4.1
Mumbai	3.1	1.2	3.0	4.3
Pune	3.4	1.5	2.1	4.8
Total	22.7	8.8	13.9	26.4

7 Conclusions

Solar radiation resource is a pre-requisite for various engineering and architectural disciplines, from agricultural engineering to building services engineering to environmental architectural design.

Often, only broadband solar radiation is required in engineering design calculations. Broadband solar radiation is measured at various meteorological and research stations. These locations are still scarce in developed countries and often only cover major population centres. In the developing areas of the world this resource assessment effort is all too often inexistent.

Engineers and designers need year-long data to assess the potential of their applications. They thus need long-term solar radiation databases for this purpose. It is important to remind that solar radiation datasets are only valid for locations under 50km from the measurement stations. Even then, the data has questionable usefulness when micro- and meso-climates are predominant. When the solar radiation data is unavailable, other means of extrapolating this data is required, thus the importance of modelling.

When plotted on a worldwide map, the areas where solar radiation is recorded are relatively insignificant compared to areas that are covered by synoptic parameters such as cloud cover, sunshine duration, relative humidity and temperatures. In the United Kingdom for example, less than 100 stations record solar radiation, compared to 300 measuring cloud and sunshine information, and 600 measuring temperature, relative humidity and other parameters. (BADC, 2006).

Solar radiation measurement stations guarantee error free data. Outliers in solar radiation datasets are caused by numerous inherent errors in the measurement instruments themselves and by operational errors. Stations are ranked in three classes, ranging from 1st class for stations that record data with the least errors present, compared to the World Meteorological Office reference instruments, and 3rd class for the least reliable stations.

In this research, 15 worldwide stations were used. These stations are all from the northern hemisphere, and represent four different macro-climates. Six UK sites and two Japanese sites represent Islands climates; yet these two countries have their particular differences, in a way that the UK represent Northern Atlantic climates, heavily influenced by the Gulf Stream, while Japan is affected by the Northern Pacific climatology. Each of the UK and Japanese sites has their own specific micro-climatology and often differs from the other sites. Two of the sites represent a desert climate, Madrid and Bahrain. Even though these two sites are geographically distant, in meteorological term, they are similar. The Indian sites represent a particular semi-tropical, semi-desert climate, which is heavily influenced by monsoons. Finally, Gerona in Spain represents a Mediterranean climate. These sites were selected for their availability and because they are representative of the main macro-climates. No sites for tropical climates and from the southern hemisphere were available for this research.

Some of the chosen sites are deemed to be first class stations, however some other stations have doubtful designation. Therefore, the quality of the data had to be assessed. After looking at the various procedures used by researchers to quality control the solar radiation datasets, these methods were found not enough to insure error-free datasets. A new procedure was thus developed.

Once the datasets were quality controlled for outliers, the modelling effort was started. Many all-sky and clear-sky models were investigated and weaknesses were found in many of them. For all-sky broadband radiation models, the Meteorological Radiation Model, as well as the Cloud Radiation Model were both found to be the most promising for improvements. However for clear-sky models the problem that was found was not in the models themselves but in the method to extract the clear-sky radiation from all-sky broadband datasets.

Comparing models for best performance is usually done by using certain statistical indicators, yet seldom do the indicators used show the complete image of the models under scrutiny. A new scoring procedure was thus developed to compare models using an array of six statistical indicators.

7.1 Data acquisition and quality control

The major sources of errors were investigated and are categorised in either instrument uncertainties and errors or operation related errors. These errors can be visually inspected in a clarity index plane. This plane is a clearness-index to diffuse ratio scatter plot. This plane of work allows the user to compare diffuse irradiance to global irradiance and thus have an indication of the clarity of the skies in the particular location and for the specific time frame inspected.

It was important to flag the data points that indicated obvious physical abnormalities. To achieve this, the tests developed by the *Commission Internationale d'Eclairage*, CIE, and the Chartered Institute of Building Services Engineers, CIBSE, have been used in this regard. Thus, global irradiance cannot exceed extraterrestrial solar irradiance, and diffuse irradiance cannot exceed global irradiance. In addition, extreme limits of particular components of the irradiance were set by the Page extreme-clear-sky irradiance components and the heavy-overcast-sky irradiance components. These tests are the basis of the physical part of the developed quality control procedure.

By inspecting the data in the clearness index to diffuse ratio plane, a certain envelope of the data could be visually observed. This is often referred to in literature as the envelope of acceptance. Previous researchers have tried to describe this envelope mathematically, in earlier attempts this was done manually, while in latter attempts this was done by statistical means, through quartile analysis. It was found that quartile analysis is a computational hungry procedure, and can be easily and with little error replaced by a standard deviation procedure.

The proposed standard deviation based procedure is achieved in the following manner. The data is split into multiple clearness index bands. For each band the positive and negative standard deviations of the weighted mean are calculated. The weighted mean is used rather than the normal mean to reduce the effect of extreme outliers on each of the data bands. The value of the standard deviation is variable, and is user defined. It was found that this value varies between 1.8 standard deviations to 2.3 standard deviations depending on the sites. The points of the positive standard deviation of each of the bands are connected and form a mathematical polynomial top envelope of acceptance. Similarly the bottom points of the negative standard deviation form the bottom section envelope of acceptance. It was found that in certain instances the bottom limit of the envelope of acceptance would remove valid points in the lower clearness index, high diffuse ratio zone. Thus a cut-off point was introduced to limit this excessive exclusion process.

The end result proved useful in removing excessive outliers from the datasets, thus insuring a better platform for modelling.

7.2 Model comparison methods

When multiple models are compared to each other for estimation performance, researchers have often used, in addition to visual means, multiple statistical indicators. The most used indicators recently have been the mean bias error and the root mean square error of the estimation. These two statistics are very appropriate indicators of the performance of the data. The lower the two values the better the model. It is recently accepted that these two parameters are not enough, and researchers have used one or more statistics to improve the image given of the models. Standard deviation, student test, coefficient of determination have been predominantly used.

When many models are compared, or many locations are used, huge tables of statistics are often presented in scientific literature, thus making model ranking very difficult.

In this study a combination of statistical indicators were calculated. These are as follows: Mean bias error, root mean square error, slope and coefficient of determination of the best fit line between calculated and measured data and the skewness and kurtosis of the estimation error histograms. With this wealth of information, there was a need to develop a scoring procedure to make sense of all this information collected on the performance of the models.

A scoring procedure was further developed and used intensively in this research to compare the models. This scoring procedure is a comprehensive method of rationalizing the statistical indicators, thus the model that yields the most desired value for each statistic would yield the best score for this statistic. Thus a model that would yield the best result for all six statistics would have achieved the highest score. If all six statistics are used, and the three components of the solar irradiance are compared, the maximum achievable score for the models is 18.

7.3 All-sky synoptic-based solar radiation modelling

For main engineering applications, long-term solar irradiance datasets for all sky conditions are necessary for the design efforts. The solar radiation can be modelled based on available synoptic parameters. If many parameters are available, more complex models can be used, however, when only few parameters are available, more simple models can be used. It is noted that the more complex the model, the more precise and accurate the estimation.

When, sunshine duration, relative humidity, dry and wet bulb temperature and atmospheric pressure are available, the Meteorological Radiation Model, MRM, can be used. The original model was found to have some inherent weaknesses in the way the all-sky radiation is calculated, while it was deemed to be accurate for clear-sky conditions. Thus, improvements were attempted to increase the overall accuracy of the model.

When only the cloud cover information is available, the Cloud Radiation Model could be used to estimate the solar irradiance components. It was found that some improvements could be made by analysing the cloud distribution.

7.3.1 Improved meteorological radiation model

The MRM works on estimating the beam irradiance by attenuating the extraterrestrial irradiance by various atmospheric transmittances. Diffuse irradiance is then estimated by a regression between the beam clearness index and the diffuse to beam ratio. Once the diffuse irradiance is estimated, the global irradiance is found by summing up the beam and diffuse components.

It was found that there are no improvements that can be made to the original method of estimating beam irradiance by using only the available input parameters and thus the current transmittance coefficients. Any improvements that can be made would require the use of extra input parameters, that are far less common than the ones currently used, thus reducing the effectiveness of the current model with regards to potential of use.

The main drawbacks of the original model is the estimation of diffuse irradiance, and thus since all three components are linked, the global irradiance.

The main method found to deal with these drawbacks is to refine the regression in the sky clarity plane. The data was banded based on sunshine fraction, and thus creating more specific regressions in the sky clarity plane, to improve the diffuse ratio estimation, hence the global irradiance. In addition to this, it was found that the original regression function which is a power function of the sky indices does not accurately represent the data. In a log plot of the sky indices, it was found that a fourth order polynomial of the logarithm of the sky indices was more indicative of the data trends based on the sunshine fractions.

The result of these modifications had a positive and clear improvement on the model for estimating diffuse and global irradiance. An improvement of 50% and more was observed for certain locations.

This model could be further improved by making atmospheric data more easily available for engineers and architects so that the transmittance equations could be changed and improved, thus improving the beam irradiance estimation, thus increasing the whole accuracy of the model.

7.3.2 Improved cloud radiation models

Cloud radiation modelling relies on very simple modelling techniques. Initially the clear-sky global radiation is estimated then the all-sky global radiation is estimated. Then the diffuse and beam radiation is estimated from the all-sky irradiation estimation. On all stages the models require regional parameters to operate, thus the cloud radiation models are location sensitive. This is due to the nature and tempo-spatial distribution of the clouds.

Weaknesses were identified in the original models. The most important weakness is the transmission of errors from component to the other. In fact any errors in the estimation of clear-sky irradiance are transmitted to all-sky global irradiance and are compounded with the errors associated with this latter estimation, and similarly for diffuse irradiance and then beam irradiance.

In addition, the original regressions between the irradiance and the cloud cover were found less than accurate when analysing the cloud distribution. New regressions were produced and thus new models were proposed.

It was found that the new model proposed did not improve on the estimation of global irradiance, however dissociating the estimation of the two other components has improved the estimation of the diffuse and beam irradiance. Thus overall the new model is a far superior model to any of the other cloud radiation models available.

A further understanding of the cloud types and their optical properties could improve the formulation of the models. However these data are often unavailable. In addition, the proposed model and its predecessors use either ground recorded or satellite derived cloud data. The problem resides in the fact that in ground measurements, only the bottom clouds are observed, while in satellite measurement, the top clouds are observed. This lack of information is obvious and observed in the cloud cover distributions where in mid-to-heavy overcast conditions, more outliers are present. Combining satellite and ground cloud cover data can improve significantly cloud based models, however until equipment are invented to assess all the cloud type distribution with thickness and cloud optical depth, no further improvements could be attempted to such models.

7.4 Clear-sky synoptic-based solar radiation modelling

Clear-sky data is required by engineers to estimate extreme case scenarios when designing their systems. In the case of HVAC engineers and architects clear-sky days are synonymous with days where the building has the most heat gains, thus influencing the design of peak air-conditioning, ventilating and shading systems. Maximum loads are thus important parts of the design of such systems, thus there is a necessity to obtain the appropriate data.

There is no method of measuring clear-sky data only. Clear-sky data points are extracted from all-sky datasets. In addition there is no defined method to extract these data points from broadband datasets. Various methods have been investigated and a hybrid methodology was developed.

Four major clear-sky models have been compared to each other to assess their performance. The performance of these models have been compared to previous assessments of these models using less stringent clear-sky data selection methods.

7.4.1 Clear-sky data acquisition

Clear-sky data was often extracted by researchers by means of using the sky clarity indices planes. Data that fitted a certain criteria based on the researchers assumptions were thus accepted and used. However there are problems with these procedures, since the clear-sky limits of these sky-indices are subject to the particular decisions of the researchers extracting those data. It is often commented in literature that these limits are not static and change depending on seasonal variations of climates.

In other cases, researchers have used cloud, or sunshine as indication of clear-skies. Yet it was found that when cloudless skies are taken and or full sunshine data points are selected, the sky clarity indices do not reflect this properly thus indicating that the synoptic and radiation parameters could often be controversial. Yet, cloud cover and clearness index give a fairly accurate image of the skies, except for some minor circumstances.

A new procedure for finding the sky clarity indices limits for clear-sky was obtained by analysing the cloud cover distribution. By combining this test and the sunshine test or the cloud test, cloudless skies were obtained with great accuracy.

However it was found that there are various degrees of sky clarity, ranging from cloudless yet heavily turbid skies to what is called blue skies. This latter occurs when the skies are cloudless and have very low aerosol occurrences.

Therefore for engineering, requiring extreme scenarios, the blue skies data points are required. This was achieved by combining the previously developed cloudless sky test with an already accepted method of finding low turbid skies, as proposed in the European Solar Atlas, ESRA, that the Linke Turbidity has to be less than 3.

7.4.2 Comparative study on clear-sky models

Four models were selected to test the new clear-sky selection procedure, and to compare their performance. These models are as follows: The MRM with only the clear-sky module of the code, the Page Radiation Model, PRM, the clear-sky module of the Yang Radiation Model, YRM and finally Gueymard's REST2 model for broadband clear-sky irradiance.

It was found that MRM performed well for all the available locations; in fact it was the best performing model over-all. PRM performed particularly poorly, while YRM and REST2 performed exceptionally well considering the use of the low-resolution parameters available. In fact the YRM and REST2 require parameters that are not easily available, and when available they are given in very-coarse form, such as worldwide maps with very low resolution of in or macro-climate tables. In this respect, these models proved to be sturdy and acceptable for use given low resolution data.

It was found that compared to a cruder method of clear-sky identification, both MRM and PRM have performed better when using extreme-clear sky datasets. Yet PRM fails to impress in both assessments.

7.5 Modelling and solar applications

In April 2006, Napier University launched its 160m^2 solar photovoltaic facility at the Merchiston campus. The mono-crystalline photovoltaic modules are manufactured and supplied by BP Solar, and have a 17kW$_p$ theoretical capability. They are installed in 32 rows on an East of South facade of the campus. The modules generate DC current that is fed to four inverters, two of 6kW$_p$ and two of 2kW$_p$ capacity each. The inverters and manufactured and supplied by Fronious. The AC current is then distributed to the university grid.

It was estimated, based on long term radiation datasets that on average the facility should generate yearly 10.81 MWh. This figure could have been obtained from modelling as well. Per example for the period of June 2005, modelling using the newly developed CRM would have resulted in an estimated PV output of 1.3 MWh compared to an actual output of 1.1 MWh.

Based on an exhaustive life cycle analysis on the facility, it is estimated that the facility has 228.8 MWh$_{th}$ and 5.17 MWh$_e$ of embodied energy and this is equivalent to 14.4 metric tons of CO_2 emission. The facility cost £155'000 and is forecasted to incur further maintenance costs estimated at £10'000-5'000 over the whole life of the facility.

It was thus found that the facility has an energetic payback time of 8 years, which compared well to other recent installations of similar size. It is very difficult to estimate how and when the facility will reach economic maturity, however with ever increasing wholesale energy prices across the world in the past two years, due to the volatility of the crude petroleum markets, the financial payback time could be anywhere from 100 years downwards. The actual cost of the facility is at £9.5/W$_p$ far higher than the \$10/W$_p$ figures quoted in more mature solar energy markets, yet PV technologies are still far from achieving market viability costs.

7.6 Future work

The quality control procedure was found to be fairly satisfactory; in fact more work can be done to produce a fully automated program. This requires a strong analysis of datasets and the associated climatology. In effect if there is a strong correlation between the quality envelope number of standard deviation and the climatology of the location, then the program could be changed to reflect this above observation.

The Meteorological Radiation Model was improved and the newer model performs up to 200% better than the older version. However both models are strongly location sensitive and regression coefficients are required for each location before the models can be accurately used. By changing the atmospheric transmittance equations this site independency could be achieved.

The Cloud Radiation Model developed in this study is the best model to be used given the current state of cloud cover reporting. If more details about cloud types, cloud thicknesses and opacity as well as location of cloud are to be available in the future, then more complex models could be developed to achieve higher accuracies in the estimation of broadband horizontal solar irradiance.

It is envisaged to create a hybrid model using in the first instance the Yang Radiation Model to estimate beam, global and diffuse clear-sky irradiance; and in the second instance the developed Cloud Radiation Model to estimate the mixed- and overcast-sky solar irradiance components. This work will feature in the following month in journal and conference publications.

REFERENCES

Alam, S. Prediction of direct and global solar irradiance using broadband models: validation of REST model. Energy Conversion Management, 2005, 31, 8. pp 1253-1263.

Angstrom, A. Solar and terrestrial radiation. Quarterly Journal Royal Meteorological Society, 1924, 50, pp.121-125.

ASHRAE. Handbook of fundamentals. American Society of Heating, refrigeration and Air-conditioning Engineers, Atlanta, USA 1993.

Barker H. Solar radiative transfer through clouds possessing isotropic variable extinction coefficient. *Q. J. R. Meteoro. Soc*, 1992; 118: 1145-1162.

Barbaro, S., Cannata, G., Coppolino, S., Leone, C. and Sinagra, E. Correlation between relative sunshine and state of the sky", Solar Energy, 1981, 26, 6, pp.537-550.

Battles, F.J., Barbero, J, Lopez, G., Perez, M., Rodrigo, F. and Rubio, M.A. Fundamentos de la radiacion solar y aspectos climatologicos de Alemria, 1990-1996. Servicio de Publicationes de la Universidad de Almeria, 1998, Spain.

Bird, R.E and Hulstrom, R.L. Application of Monte Carlo Technique to Insolation Charaterisation and Prediction. US SERI Tech. Report 1979, TR-642-761,38.

Beggs, C. A method for estimating the time-of-day carbon dioxide emissions per kWh of delivered electrical energy in England and Wales. BSER&T, 1996, 127-134.

Bennett. I. Correlation of daily insolation with daily total sky cover, opaque sky cover and percentage of possible sunshine. *Solar Energy*, 1969; 12: 391-393.

Berg, B. The ecology of building materials. Architectural press. Butterworth-Heinemann, Oxford, UK 2001.

Bing, J. A new business model for grid-connected solar generation in restructured electricity markets. Solar World Congress, ISES/ASES-05, Orlando, Florida, USA, August 2005.

BP Solar Technical Brochure. Photovoltaic Module BP790-Saturn Technology - 3012E - 05. 2004.

Brinsfield R, Yaramanoglu M, Wheaton F. Ground level solar radiation prediction model including cloud cover effects. *Solar Energy*, 1984; 33: 493-499.

British Atmospheric Data Center website. http://www.badc.ac.uk. 2005.

Chendo, MAC. and Madueke AAL. Hourly global and diffuse radiation of Lagos, Nigeria - correlation with some atmospheric parameters. Solar Energy, 1994, 52, 3, pp.247-251.

Chevalier, J-M. Les grandes batailles de l'energie. Edition Gallimard, paris, France, 2004.

Cleveland C.J. Encyclopaedia of Energy. Elsevier, Oxford, UK, 2004.

CIBSE Guide J. Chartered Institute of Building Services Engineers, London, UK, 2002.

CIE. Standardization of luminance distribution on clear skies, CIE Publication No.22, TC-4.2, 1973.

Claywell R.G. Quality assurance of solar radiation datasets. MPhil Thesis, Napier University, 2003.

Collares-Pereira, M. and Rabl, A. The average distribution of solar radiation – correlation between diffuse and hemispherical and daily and between daily and hourly insolation values. Solar Energy, 1979 22, 2, pp.155-164.

Coulson K. Solar and terrestrial radiation, methods and measurements. Academic Press, New York 1975, pp. 60,100.

Culp, A. Principle of Energy Conversion. McGraw-Hill, Singapore 1991.

Dave, J.V. Extensive data sets of the diffuse radiation in realistic atmospheric models with aerosols and common absorbing gases. Solar Energy, 1979, 21, pp. 361-9.

ESRA. European Solar Radiation Atlas. Ecole des Mines, Paris, 2000.

FRONIUS IG Technical Data Brochure. 40,000 6,2366 - 01. 2005.

Geiger M, Diabate L, Menard L, Wald L. A web service for controlling the quality of measurements of global solar irradiation. Solar Energy 2002, 73, 6, pp. 475-480.

GEMIS. Global emission model for integrated systems, GEMIS 4.1 Database (September 2002). O¨ ko-Institut Darmstadt, Germany, 2002.

Greif J, Scharmer K. European Solar radiation Atlas. 4th edition. France: Presses de l'Ecole, Ecole des mines de Paris, 2000.

Grenier, J., de la casiniere, A. and Cabot, T. A spectral model of Linke's turbidity factor and its experimental implications. Solar Energy, 1994, 52, 4, pp. 303-313.

Gueymard, C. Prediction and performance assessment of mean hourly global radiation. *Solar Energy,* 1999, 68, pp. 285-301.

Gueymard, C. Direct solar transmittance and irradiance predictions with broadband models. Part I: detailed theoretical performance assessment. Solar Energy, 2003, 74, 5, pp.355–79.

Gueymard, C. Direct solar transmittance and irradiance predictions with broadband models. Part II: validation with high quality measurements. Solar Energy, 2003, 74, 5, pp.381–95.

Gueymard, C. Corrigendum to Direct solar transmittance and irradiance predictions with broadband models. Part I: detailed theoretical performance assessment. Solar Energy, 2004, 76, 4, pp. 513.

Gueymard, C. Corrigendum to Direct solar transmittance and irradiance predictions with broadband models. Part II: validation with high quality measurements. Solar Energy, 2004, 76, 4, pp. 515.

Gul M, Muneer T and Kambezidis H. Models for obtaining solar radiation from other meteorological data. Solar Energy, 1998, 68, pp.99-108.

Hagedorn, G. Hidden energy in solar cells and photovoltaic power station. Proceeding of the 9th PV specialist conference, Germany 1989, pp. 542-546.

Institute of Civil Engineers. The State of the nation. An assessment of the UK's infrastructure by the Institution of Civil Engineers. Institute of Civil Engineers, London. 2005.

Harrison C, Coombes C. Empirical relationship of cloud shade to point cloudiness (Canada). *Solar Energy*, 1986; 37: 417-421.

Hubbard K.G. Spatial variability of daily weather variables in high plains of the USA. *Agric. For. Meteorol. 1994*; 68: 29-41.

Iqbal, M. An introduction to solar radiation. Academic Press, Canada, 1983.

Ianetz, A., Lyubansky, V., Setter, I., Evseev, E. and Kudish, A. Inter-comparison of different methods for estimating clear sky global radiation for the negev region of Israel", Proceeding of the 2nd SOLARIS conference, Athens, Greece. 2005.

Jacobson, R. Microsoft Excel 2002 Visual Basic for Applications STEP BY STEP. Microsoft Press, Redmond, WA, USA, 2001.

Jungblunth, N. Life Cycle Assessment for Crystalline Photovoltaics in the Swiss ecoinvent Database - Progress in Photovoltaics: Research and Applications. John Wiley & Sons, Ltd. 2004.

Jungblunth, N., Bauer, C., Dons, R. and Frischknecht, R. Life cycle assessment for emerging technologies: Case studies for Photovoltaic and Wind Power. Int. J. LCA, 2004, 11.

Kannan, R., Leong, K.C., Osman, R., Ho, H.K. and Tso, C.P. Life cycle assessment study of solar PV systems: An example of a 2.7 kWp distributed solar PV system in Singapore. Solar Energy. 2005. (In Press).

Karl, E.K. and Theresa, L.J.. Initial empirical results for the energy payback time of photovoltaic modules. Siemens Solar Industries, Camarillo. 2002.

Kasten F, Czeplak G. Solar and terrestrial radiation dependant on the amount of type of cloud. *Solar Energy*, 1980; 24: 177-189.

Kasten, F. Discussion on the relative air mass. Light. Res. Tech. 1993, 25, pp. 129.

Kato, K., Murato, A. and Sakuta, K. An evaluation on the life cycle of Photovoltaic energy systems considering production energy of off-grade silicon. Solar Energy Materials and Solar Cells. 1997, 47, pp. 95-100.

Kipp and Zonen. Pyranometers for atmospheric research and industry. Brochure 4414-470-W41. 2005.

Kendrick D, et al. Guide to recommended practice of daylight measurement. Report no. CIE-108. Wein, Austria: International Commission on Illumination (CIE), 1994.

Lam J, Li D. Correlation analysis of solar radiation and cloud cover. *International Journal of Ambient Energy*, 1998; 19, pp. 187-198.

Li, D. and Lam, J. An analysis of climatic parameters and sky condition classification. Building and environment, 2001, 36; pp.435-445.

Littlefair PJ., Measurements of the luminous efficacy of daylight. Lighting Res. and Tech., 1998, 20, pp.177-188.

Long, C. N., and Ackerman, T. P. Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects", J .Geophys. Res. 2000, 105, 15, pp. 609-15, 626.

Maxwell E, Wilcox S, Rymes M. Users manual for SERI QC software, assessing the quality of solar radiation data. Report no. NREL-TP-463-5608. 1617 Cole Boulevard, Golden, Colorado: National Renewable Energy Laboratoty, 1993 .

Molineaux B, Ineichen P. Automatic quality control of daylight measurement: software for IDMP stations. Vaulx en Velin, France, 2003, International Daylight Measurement Programme, Ecole National des Travaux Publics de l'Etat. See also: http://idmp.ente.fr

Muneer, T. Solar radiation model for Europe. BSER&T, 1990, 11, 153.

Muneer T. Solar irradiance and illuminance models for Japan I: sloped surfaces. *Light. Res. Tech.,1995, 27, 209.*

Muneer T. Solar radiation & daylight models. Oxford: Elsevier, 2004

Muneer T, Fairooz F. Quality control of solar radiation and sunshine measurements - lessons learnt from processing worldwide databases. Building Services Engineering Research & Technology 2002. 23, 3, pp. 151-166.

Muneer T and Gul M. Evaluation of sunshine and cloud cover based models for generating solar radiation data. Energy Conversion & Management, 2000; 41, pp. 461-482.

Muneer T, Gul M and Kambezidis H. Evaluation of an all-sky meteorological radiation model based against long-term measured hourly data. Energy Conversion & Management, 1998, 39, pp. 303-317.

Muneer, T. and Younes, S. The all-sky meteorological radiation model: proposed. Applied Energy, 2006, 83, pp. 436-450.

Muneer, T., Younes, S. and Munawwar, S. Discourses on solar radiation modelling. Renewable and Sustainable Energy Reviews, (In Press).

Muneer T, Zhang X. A new method for correcting shadow band diffuse irradiance data. Journal of Solar Energy Engineering, 2002; 124, pp. 34-43.

Myers, D. Solar radiation modeling and measurements for renewable energy applications: data and model quality, Energy, 2005, 30(9), pp. 1517-1531.

Myers, D. A method of improving global pyranometer measurements by modeling responsivity functions. Solar Energy, 2006, 80, pp. 322-331.

Ododo, JC., Agbakwuru, JA. and Ogbu, FA. Correlation of solar radiation with cloud cover and relative sunshine duration", Energy Conversion Management, 1996, 37, 5, pp. 1555-155.

Page J. Proposed quality control procedures for the Meteorological Office data tapes relating to global solar radiation, diffuse solar radiation, sunshine and cloud in the UK. Report submitted to CIBSE Guide Solar Data Task Group. Chartered Institute of Building Services Engineers, 222 Balham High Road, London, UK, 1997.

Page, J.K and Lebens, R. Climate in the United Kingdom. HMSO, London. 1986.

Pallant, J. SPSS surviving manual: a step by step guide to data analysis using SPSS for windows (version 12). Open University Press, maidenhead, UK, 2005.

Perez, R., Ineichen, P. and Seals, R. Modelling daylight availability and irradiance components from direct and global irradiance, Solar Energy, 1990, 44, 5, pp. 271-289.

Perez, R., George, R., Renné, D. Cloud cover reporting bias at major airports. ASES-04, Portland, Oregon, USA. June 2004.

Pisimanis, D., Notaridou, V. and Lalas, D.P. Estimating direct, diffuse and global radiation on an arbitrarily inclined plane in Greece. Solar Energy, 1987, 39, 3, pp. 159.

Rigolier, C., Bauer, O. and Wald, L. On the clear-sky model of the ESRA-European Solar Radiation Atlas- with respect to the Heliosat method. Solar Energy, 2000, 68, 1, pp. 33-48.

Rymes M, Myers D. Mean preserving algorithm for smoothly interpolating averaged data. Solar Energy, 2001; 71, 4, pp. 225-231

SODA. Integration and exploitation of networked solar radiation databases for environmental monitoring. 2002. http://www.soda-is.com

Stoffel T, Redo I, Myers D, Renne D, Wilcox S, Treadwell J. Current issues in terrestrial solar radiation instrumentation for energy, climate and space applications. Metroloogia, 2000; 37, 5, pp. 399-402.

Thevenard, D. and Brunger, A. ASHRAE research project 1015-RP. Typical Weather Years for International Locations. 2001.

Wald, L. The Project SoDa for Solar Energy and Radiation, 2004. Sophia Antipolis, France, 2004, Ecole des Mines de Paris, Centre d'Energetique, Groupe Télédétection et Modélisation. See also: http://www.soda-is.org

Wood, J., Muneer, T. and Kubie, J. Evaluation of a novel sensor for measuring global and diffuse irradiance and sunshine duration. ASME J. Solar Energy Eng. 2003, 125, 43.

Younes S, Claywell R and Muneer T. Improved and automated methods for the quality control of solar irradiance data. Energy, 2005; 30, pp.1533-1549.

Yang, K., Huang, GW. and Tamai, N. A hybrid model for estimating global solar radiation", Solar Energy, 2001, 70, 1, pp. 13-22.

BIBLIOGRAPHY

Alados, I., Alados-Arboledas, L. Direct and diffuse photosynthetically active radiation: measurements and modelling. Agricultural and Forest Meteorology, 1999, 93, 1, pp: 27-38.

Boer, K., Duffie, J., Prince, M., Goswami, Y. Advances in solar energy : an annual review of research and development. Earthscan, London, UK, 2005.

Duffie, J., Beckman, W. Solar engineering of thermal processes. Wiley, Chichester, 1991.

Gueymard, C., Vignola, F. Determination of atmospheric turbidity from the diffuse-beam broadband irradiance ratio. Solar Energy, 1998, 63, 3, pp.135-146.

Herzog, T., Kaiser, N., Volz, M. Solar energy in architecture and urban planning. Prestel, Munich, Germany, 1998.

Jacovides, C., Timvios, F., Papaioannou, G., Asimakopoulos, D., Theofilou, C. Ratio of PAR to broadband solar radiation measured in Cyprus. Agricultural and Forest Meteorology, 2004, 121, 3-4, pp: 135-140.

Kambezidis, H., Katevatis, E., Petrakis, M. Lykoudis S., Asimakopoulos, D. Estimation of the linke and unsworth-monteith turbidity factors in the visible spectrum: application for Athens, Greece. Solar Energy, 1998, 62, 1, pp.39-50.

Krivova, N., Solanki, S. Solar variability and global warming: a statistical comparison since 1850. Advances in Space Research, 2004, 34, 2, pp.361-364.

Li, D., Lam, J., Lau, C. A new approach for predicting vertical global solar irradiance. Renewable Energy, 2002, 25, 4, pp.591-606.

López, G., Muneer, T., Claywell, R. Assessment of four shadow band correction models using beam normal irradiance data from the United Kingdom and Israel. Energy Conversion and Management, 2004, 45, 13-14, pp.1963-1979.

Louche, A., Maurel, M., Simonnot, G., Peri, G., Iqbal, M. Determination of Ångström's turbidity coefficient from direct total solar irradiance measurements. Solar Energy, 1987, 38, 2, pp.89-96.

Markvart, T. Solar electricity. Wiley, New York, USA, 2000.

Mendoza, B. Total solar irradiance and climate. Advances in Space Research, 2005, 35, 5, pp.882-890.

Meltzer, M. Passive and active solar heating technology. Prentice-Hall, London, UK, 1985.

Miyake, J., Igarashi, Y., Rogner, M. Biohydrogen III : renewable energy system by biological solar energy conversion. Elsevier, Oxford, UK, 2004.

Muneer, T., Abodahab, N., Weir, G., Kubie, J. Windows in buildings. Architectural Press, Elsevier, Oxford, UK, 2000.

Muneer, T., Munawwar, S. Potential for improvement in estimation of solar diffuse irradiance. Energy Conversion and Management, 2006, 47, 1, pp.68-86.

Niele, F. Energy- Engine of evolution. Elsevier, Amsterdam, The Netherlands, 2005.

Norton, B. Solar energy thermal technology. Springer-Verlag, London, UK, 1992.

Ogunjobi, K., Kim, Y., He, Z. Influence of the total atmospheric optical depth and cloud cover on solar irradiance components. Atmospheric Research, 2004, 70, 3-4, pp: 209-227.

Perez, R., Hoff, T., Herig, C., Shah, J. Maximizing PV peak shaving with solar load control: validation of a web-based economic evaluation tool. Solar Energy, 2003, 74, 5, pp.409-415.

Psiloglou, B., Santamouris, M., Asimakopoulos, D. On broadband Rayleigh scattering in the atmosphere for solar radiation modelling. Renewable Energy, 1995, 6, 4, pp:429-433.

Scheer, H. A solar manifesto : the need for a total solar energy supply .and how to achieve it. James & James, London, UK, 1994.

Sheldom, M. Introduction to probability and statistics for engineers and scientists. Elsevier Academic Press, Oxford, UK, 2004.

Varo, M., Pedrós, G., Martínez-Jiménez, P., Aguilera, M.J. Global solar irradiance in Cordoba: Clearness index distributions conditioned to the optical air mass. Renewable Energy, 2006, 31, 9, pp.1321-1332.

APPENDIX A: LIST OF PUBLICATIONS

Journal articles

Younes, S., Claywell, R. and Muneer, T., 2005. Quality control of solar radiation data: Present status and proposed new approaches. Energy, 30, pp. 1533-1549.

Younes, S. and Muneer, T., 2006. Improvements in Solar Radiation Models Based on Cloud Data. Building Services Engineering Research & Technology, 27 (1), pp 41-54.

Muneer, T. and Younes, S., 2006. The all-sky meteorological radiation model: proposed. Applied Energy, 83, pp.436-450.

Muneer, T., Younes, S., Lambert, N. and Kubie, J. Life cycle assessment of a medium sized photovoltaic facility at a high latitude location. IMechE Journal of Power and Energy (In Press).

Younes, S., Muneer, T. Evaluation of Clear Sky-Classification Procedures and Models Using a World-Wide Database. Applied Energy (Communicated).

T. Muneer, S. Younes and S. Munawwar. Discourses on solar radiation modelling. Renewable and Sustainable Energy Reviews. (In Press).

S. Younes and T. Muneer. Comparison between Solar Radiation Models Based on Cloud Information. International Journal of Solar Energy. (In Review).

S. Younes and T. Muneer; Solar Horizontal Irradiation: Data Quality Control and Modelling. Solar Energy. (In Review).

Conference articles

Muneer, T., Munawwar, S. and Younes, S. Performance of Meteorological Parameters Based Solar Radiation Models. SOLARIS-05, Athens, Greece, May 2005.

Younes, S and Muneer, T. Solar Horizontal Irradiation: Data Quality Control and Modelling. Solar World Congress, ISES/ASES-05, Orlando, Florida, USA, August 2005.

Younes, S and Muneer, T. Improvements in Broadband All-Sky Solar Radiation Modelling Based on Cloud Cover Information for the UK. Eurosun 2006, Glasgow, UK, June 2006.

Younes, S. and Muneer, T. Alternative methods for clear-sky conditions identification for solar irradiance modelling and their effect on the validation of models. Eurosun 2006, Glasgow, UK, June 2006.

Younes, S. Muneer, T. and Kubie, J. Life cycle assessment of a medium sized PV facility in Edinburgh. Eurosun 2006, Glasgow, UK, June 2006.

APPENDIX B: GLOSSARY

Appendix B: GLOSSARY

Absorption - when the substance of interest is captured by another substance, reducing the amount available. For example, solar energy is absorbed by some atmospheric molecules, solar collectors, and the ocean.

Aerosol - excluding weather and clouds, any small particle that tends to stay in the air, such as smoke, dust, salt, and pollen.

Aerosol Optical Depth - (technically known as the relative aerosol optical depth) usually considered to be synonymous with the air mass, is the approximate number of aerosols in a path through the atmosphere relative to the standard number of aerosols in a vertical path through a clean, dry atmosphere at sea level.

Airmass - the relative path length of the direct solar beam radiance through the atmosphere. When the sun is directly above a sea-level location the path length is defined as airmass 1 (AM 1.0). AM 1.0 is not synonymous with solar noon because the sun is usually not directly overhead at solar noon in most seasons and locations. When the angle of the sun from zenith (directly overhead) increases, the airmass increases approximately by the secant of the zenith angle.
The figure below illustrates the concept of airmass.

Albedo - the fraction of solar radiation that is reflected. The solar energy community defines albedo as the fraction of solar radiation that is reflected from the ground, ground cover, and bodies of water on the surface of the earth. Astronomers and meteorologists include reflectance by clouds and air. To reduce confusion, some solar researchers use the term ground reflectance.

Atmospheric Pressure - the pressure (force per area) created by the weight of the atmosphere. At higher elevations, the atmospheric pressure is lower because there is less air.

Atmospheric Turbidity - haziness in the atmosphere due to aerosols such as dust (particles ranging from 0.1 to 1+ microns in diameter). If turbidity is zero, the sky has no dust. A sun photometer is used to measure atmospheric turbidity.

Attenuation - loss of a substance as it is deflected, fragmented, or absorbed. For example, solar irradiance attenuates as it passes through the atmosphere to the surface of the earth.

Azimuth Angle - the angle between the horizontal direction (of the sun, for example) and a reference direction (usually North, although some solar scientists measure the solar azimuth angle from due South).

BADC - British Atmospheric Data Center.

Beam Radiation - synonym for direct normal irradiance, the amount of solar radiation from the direction of the sun.

Bright Sunshine - when the sun casts an obvious shadow or when a Campbell-Stokes sunshine recorder is recording. The lower limit for bright sunshine (based on a Campbell-Stokes recorder) is between 70 W/m2 (very dry air) and 280 W/m2 (very humid air).

Broadband Solar Irradiance - theoretically the solar radiation arriving at the earth from all frequencies or wavelengths, in practice limited to the spectral range of radiometers, typically from 300 nm to 3000 nm wavelength. Meteorologists refer to this band as short-wave radiation.

BSRN - the worldwide Baseline Surface Radiation Network, or the program that manages it.

Calibration - the process of comparing an instrument's output signal with reality. Instruments that measure solar energy tend to "drift", that is, their output signals do not mean the same thing from one time period to another. Because of this, they are periodically (annually or semi-annually) re-calibrated against more reliable instruments. The picture below illustrates instruments being calibrated at SRRL. The radiometers on the calibration table are calibrated to a reference instrument traceable to the World Radiometric Reference (WRR).

Campbell-Stokes Sunshine Recorder - a clear glass sphere that focuses the sun's rays onto a special strip chart, producing a charred path when there is bright sunshine. The length of the path determines the bright sunshine duration. The lower limit for bright sunshine (based on a Campbell-Stokes recorder) is between 70 W/m2 (very dry air) and 280 W/m2 (very humid air).

Circumsolar Radiation - the amount of solar radiation coming from a circle in the sky centered on the sun's disk and having a radius of between 2.5 and 3.5 degrees, depending on the type of instrument being used to measure beam radiation (direct normal irradiance).

Cloud Amount - the fraction of the sky dome covered by clouds. This fraction is typically expressed either as tenths (1/10, ..., 10/10) or eighths (1/8, ..., 8/8).

Cloud Cover - the fraction of the sky dome covered by clouds. This fraction is typically expressed either as tenths (1/10, ..., 10/10) or eighths (1/8, ..., 8/8). Some researchers refer to this as cloud amount, to clarify the distinction from cloud type, which is the nature of the cloud cover.

Cosine Response - the effects of radiance incidence angle on pyranometer measurement performance. If a pyranometer is rotated while a beam of light is shined upon it, it will record the maximum energy when it is directly facing the beam, and the energy will fall to zero when it is sideways to (or facing away from) the beam. A graph of the energy reported by the pyranometer as the angle it makes with the beam of light should look like the cosine of the angle, if the instrument were perfect.

Dewpoint - the temperature at which the water in the atmosphere will condense as drops on a surface.

Diffuse Horizontal Irradiance - synonym for diffuse sky radiation.

Diffuse Sky Radiation - the radiation component that strikes a point from the sky, excluding circumsolar radiation. In the absence of atmosphere, there should be almost no diffuse sky radiation. High values are produced by an unclear atmosphere or reflections from clouds.

Direct Normal Irradiance - synonym for beam radiation, the amount of solar radiation from the direction of the sun.

Dry-bulb Temperature - air temperature measured with a thermometer, similar to ambient temperature. The term "dry-bulb" distinguishes it from the wet-bulb temperature measured by a psychrometer to determine relative humidity.

Equation of Time - the annual East-West swing of the location of the Sun which can be detected by noting the position of the Sun at the same time (such as noon) each day. This motion is caused by the libration (wobble) of the Earth

Equinox - literally "equal night", a day when the number of hours of daylight equals the number of hours of night. The vernal equinox, usually March 21, signals the onset of Spring, while the autumnal equinox, usually September 21, signals the onset of Autumn.

Extraterrestrial Radiation - abbreviated ETR, also known as "top-of-atmosphere" (TOA) irradiance, is the amount of global horizontal radiation that a location on Earth would receive if there was no atmosphere or clouds (i.e., in outer space). This number is used as the reference amount against which actual solar energy measurements are compared.

Global Horizontal Radiation - total solar radiation; the sum of direct, diffuse, and ground-reflected radiation; however, because ground reflected radiation is usually insignificant compared to direct and diffuse, for all practical purposes global radiation is said to be the sum of direct and diffuse radiation only.

Greenhouse Effect - the warming of the Earth by the atmosphere because of water vapor and gases such as carbon dioxide, which absorb and emit infrared radiation, or heat. Thus, the high-energy photons such as light and ultraviolet radiation are passed through the atmosphere to the Earth, which tends to absorb them and emit lower-energy photons which are then captured in the atmosphere and partially sent back to Earth. As the presence of infrared absorbers rises in the atmosphere, the more solar energy is retained at heat in the atmosphere and on the surface of the Earth. Because glass also passes light and tends to absorb and reflect heat, this effect is compared to that of a greenhouse.

Ground-Reflected Radiation - the radiation from the sun which is reflected back into the atmosphere after striking the Earth.

Humidity - the amount of water vapor in the air. Because the common measure of water vapor is the ratio between the measured amount and the maximum possible amount (the saturation point at which water condenses as dew), humidity and relative humidity are generally used interchangeably.

Incident Angle - the angle that a ray (of solar energy, for example) makes with a line perpendicular to the surface. For example, a surface that directly faces the sun has a solar angle of incidence of zero, but if the surface is parallel to the sun (for example, sunrise striking a horizontal rooftop), the angle of incidence is 90°. The figure accompanying the description of airmass illustrates a solar angle of incidence of 48.2° to a horizontal surface.
Incident Radiation - incoming radiation; i. e., radiation that strikes a surface.

Insolation - solar radiation on the surface of the Earth. This term has been generally replaced by solar irradiance because of the confusion of the word with insulation.

Irradiance - the rate at which radiant energy arrives at a specific area of surface during a specific time interval. This is known as radiant flux density. A typical unit is W/m^2.

Local Apparent Time - The time of day based strictly on the longitude of the locality and not on "blocky" time zones. For example, when it is 12:00 Pacific Standard Time (USA) (assumed to be 120° West Longitude), it is 11:51 Local Apparent Time in Seattle, Washington (USA), at 122° 18' West Longitude.

Local Standard Time - The time of day based on the longitude of the zone meridian associated with a locality.

Macroclimate - the general climate of a large region such as the Baltics, the British isles…etc.

Measurement Uncertainty - the bounds that should be placed on a measured value because of uncertainties in the measurement. If there are several factors pertaining to the measurement, such as voltage bias and temperature bias and precision of measurement scale, the total measurement uncertainty can be difficult to calculate and may be larger than the largest individual uncertainty of any one factor depending on the sensitivity of the measurement to the significant factors. There is no such thing as a perfect measurement, although some measurements are so precise that errors are negligible. Solar irradiance measurements are notoriously unreliable with the best methods (1% to 3% uncertainty, which means that an "excellent" method can produce results that may differ as much as 50 W/m^2), and can become worthless (10% to 30% uncertainty) with careless methods.

Mesoclimate - the climate that is peculiar to a small natural feature such as a hill or a small lake. This climate tends to be different from the general climate of the region in predictable ways. Statements such as "it snows more at the airport than downtown" are statements about mesoclimates.

Microclimate - the local climate near the ground that is peculiar to a small area (usually, the radius is less than a kilometer, and can be as small as a centimeter). A microclimate region is defined by changes in behavior of the atmosphere's surface boundary layer and not by obvious physical features.

Mie Scattering - the scattering of solar radiation by (mathematically spherical) particles in the atmosphere which have an approximate size of the wavelength of light, analyzed by Gustav Mie. While Rayleigh scattering explains the blue sky, Mie scattering explains why wet, coastal skies are whiter than dry, mountainous skies.

Minutes of Sunshine - a specific instance of bright sunshine duration, the number of minutes per hour during which the sun casts an obvious shadow or when a Campbell-Stokes sunshine recorder is recording, usually above 210 W/m2.

NIP - a Normal Incident Pyrheliometer, used to determine the amount of solar irradiance emitted from the direction of the sun. Below is a picture of two NIPs (silver) mounted on a sun-following tracker (white) at the Solar Radiation Research Laboratory. These thermopile-based radiometers have a uniform spectral response from 280 nm to 2800 nm and a 5.7° field of view.

Normal Radiation - radiation striking a surface that is facing the sun. Mathematically, the word normal is the vector that is perpendicular to a surface, and the direction of a normal radiation source is perpendicular to a radiation source. Global (total) normal solar irradiance is all radiation that strikes a flat surface that faces the sun, while direct normal solar irradiance excludes all radiation that does not come from the direction of the sun in the sky.

NSRDB - the 1961-1990 National Solar Radiation Data Base, which supplies hourly solar and meteorological data from 239 locations in the United States and its territories.

Optical Depth - (technically known as the relative aerosol optical depth) usually considered to be synonymous with the airmass, is the approximate number of aerosols in a path through the atmosphere relative to the standard number of aerosols in a vertical path through a clean, dry atmosphere at sea level.

Orientation - the direction that a solar energy collector faces. The two components of orientation are the tilt angle (the angle the collector makes from the horizontal) and the aspect angle (the angle the collector makes from North).

Ozone Layer - the layer in the atmosphere with the most ozone, usually at an altitude of 25 km. Ozone is created from oxygen by ultraviolet radiation bombardment. Because ozone tends to absorb and block ultraviolet radiation, a substantial ozone layer reduces the risk of skin cancer.

Passive Solar - technology for using sunlight to light and heat buildings directly, with no circulating fluid or energy conversion system.

Percent Possible Sunshine - the ratio of measured bright sunshine to the total possible bright sunshine in a given time period such as an hour or a day, expressed as a percent.

Photovoltaic - technology for converting sunlight directly into electricity, usually with photovoltaic cells.

Photovoltaic Cell - a single semiconducting element of small size (for example, 1 cm2) that absorbs light or other bands of the electromagnetic spectrum and emits electricity.

Photovoltaic Module - a unit comprised of several photovoltaic cells that is the principal unit of photovoltaic array. A photovoltaic module's size is on the order of 1 m2, although its size is governed by convenience and application.

Photovoltaic Array - a photovoltaic module or set of modules used for converting solar radiation to energy.

PMOD/WRC - the Physikalisch-Meteorologisches Observatorium Davos / World Radiation Center, at Davos, Switzerland. PMOD/WRC determines and maintains world-wide standards for measurement of solar radiation, including the World Radiometric Reference (WRR), for the World Meteorological Organization.

Point-focusing Concentrator - a solar power generator which uses a series of tracking mirrors (heliostats), Fresnel lenses, or a paraboloid (3-dimensional parabola, or dish) of mirrors to focus solar energy onto a single central receiver such as a boiler, engine, or photovoltaic array.

Power - the amount of work or energy expended in a given amount of time. For example, the watt is a unit of power, which is defined as a joule per second.

Precipitable Water - the amount of water in a vertical column of atmosphere. The unit of measure is typically the depth to which the water would fill the vertical column if it were condensed to a liquid. For example, 6 centimeters of precipitable water (in the absence of clouds) indicates a very moist atmosphere. Precipitable water is often used as a synonym for water vapor.

Pyranometer - an instrument with a hemispherical field of view, used for measuring total or global solar radiation, specifically global horizontal radiation; a pyranometer with a shadow band or shading disk blocking the direct beam measures the diffuse sky radiation.

Pyrheliometer - instrument with a narrow (circumsolar) field of view which measures direct normal irradiance. Pyrheliometers are mounted on sun-following trackers so that the instrument is always aimed at the sun. Below is a picture of two (silver) pyrheliometers mounted on a (white) tracker.

Rayleigh Scattering - the scattering of solar radiation by (mathematically spherical) particles in the atmosphere which are much smaller than the wavelength of light, analyzed by Lord Rayleigh. Rayleigh scattering explains the blue sky.

RCC - Radiometer Calibration & Characterization (RCC) software is used to automate the BORCAL process. The RCC controls all data acquisition from the reference radiometers and those under calbration, displays several color-coded fields representing the present sky condition and instrument performance, builds an instrument calibration database, and generates the final calibration report.

Refraction - the bending of electromagnetic radiation by its passage through a medium of a high refractive index. Light is refracted by passing through a lens, water, or the atmosphere.

Relative Humidity - the amount of water vapor in the air expressed as the ratio between the measured amount and the maximum possible amount (the saturation point at which water condenses as dew).

Remote Sensing - the determination of a quantity by detecting it from a distance. A common application of remote sensing is the use of satellite-borne instruments to determine the location and amount of resources on the surface of the Earth.

Rotating Shadow Band Radiometer - an instrument that determines total solar radiation and diffuse sky radiation by periodically shading the total sky sensor from the sun with a rotating shadow band.

Saturated Air - air that has the maximum amount of water vapour; any increase in water vapour will cause condensation.

Scattered Radiation - radiation that has been reflected from particles, disrupting the original direction of the beam.

SERI - the Solar Energy Research Institute, which became the National Renewable Energy Laboratory in 1991.

Shading Disk - a disk on a tracking arm which blocks the direct normal irradiance so as to allow a pyranometer to measure only the diffuse sky radiation.

Shadow Band - a metal strip which blocks the direct normal radiation so as to allow a pyranometer to measure only the diffuse sky radiation.

Solar Constant - although not strictly constant, this number is the amount of solar power flux that passes through the mean Earth orbit. The currently accepted value is 1367 W/m2. Note that Earth-based instruments record lower values of solar power flux because of atmospheric attenuation.

Solar Irradiance - the amount of solar energy that arrives at a specific area of a surface during a specific time interval (radiant flux density). A typical unit is W/m2.

Solar Noon - the time at which the position of the sun is at its highest elevation in the sky. At this time, the Sun is either due South (typically in the Northern Hemisphere) or due North (typically in the Southern Hemisphere). This time can be quite different from noon according to local standard time.

Spectral Irradiance - the amount of radiant energy flux expressed in terms of the solar spectrum. NREL's Solar Spectral Radiation Data Base contains thousands of irradiance spectra.

Sunshine - used interchangeably with the more precise term bright sunshine, when the sun casts an obvious shadow or when a Campbell-Stokes sunshine recorder is recording, usually above 210 W/m2.

Sunshine Duration - the length of time for which the sun casts an obvious shadow or when a Campbell-Stokes sunshine recorder is recording. The lower limit for bright sunshine (based on a Campbell-Stokes recorder) is between 70 W/m2 (very dry air) and 280 W/m2 (very humid air).

Sun Position - the location of the sun in the sky, expressed in terms of azimuth angle and zenith angle.

Temporal - pertaining to time, such as temporal variation (variation over time).

TMY - Typical Meteorological Year, a "typical" year of hourly solar and meteorological values which is designed to produce the expected climate of a location throughout a year.

Total Solar Radiation - solar radiation that is the sum of direct, diffuse, and ground-reflected radiation; however, because ground reflected radiation is usually insignificant compared to direct and diffuse, for all practical purposes global radiation is said to be the sum of direct and diffuse radiation only.

Transient Response - the short-term response of an instrument caused by a change of status of the instrument's environment. For example, the switching of a power supply on and off will send very short-term power spikes that can be detected by a volt meter with sufficiently rapid response time.

Transmittance - the fraction or percent of a particular frequency or wavelength of electromagnetic radiation that passes through a substance without being absorbed or reflected.

Turbidity - a measure of the opacity of the atmosphere. A perfectly clear sky has a turbidity of 0, and a perfectly opaque sky has a turbidity of 1. Turbidity is affected by air molecules and aerosols.

Temporal - pertaining to time, such as temporal variation (variation over time).

Uncertainty - the expression of the amount of doubt that remains after a result is obtained. Although uncertainty may be subjective and without foundation ("We are 80% certain that nuclear fusion will be a power source in the 21st century"), many uncertainties are determined by statistical procedures ("Sampling polls indicate that 63% agree, with a possible 3% uncertainty") or measurement uncertainty (e.g., 3.204 ± 0.005 °C).

Visible Radiation - the range of radiation wavelengths which is visible to the human eye; goes from red to violet and then enters the ultraviolet part of the spectrum

Water Vapour - gaseous water (individual water molecules) in the atmosphere.

Wet-bulb Temperature - temperature that is measured by a wet-bulb thermometer, which has a wet cloth sleeve that covers its bulb. Wet-bulb temperature and dry-bulb temperature are used to compute relative humidity.

WRR - the World Radiometric Reference, which provides the basis for all measurements by radiometers in the world. Every five years, many of the best absolute cavity radiometers undergo an intercomparison at PMOD/WRC (Davos, Switzerland). The most stable, accurate, and precise instruments provide the World Radiometric Reference for the coming years. Any credible radiometer measurement must be traceable to the WRR.

WSG - the World Standard Group (WSG) of absolute cavity radiometers is maintained by the World Meteorological Organization's (WMO) World Radiation Center (WRC) (Davos, Switzerland). The WSG is a group of seven well-characterized absolute cavity radiometers used to define the World Radiometric Reference (WRR). International intercomparisons of national standard pyrheliometers with the WSG are held every five years at the WRC to transfer the WRR to national centers. Having participated in such comparisons since 1980, NREL has three absolute cavity radiometers directly traceable to the WRR. The WRR has an uncertainty of less than +/- 0.3%. This means that the best possible measurements of direct normal solar irradiance have at least this uncertainty.

Zenith Angle - the angle between the direction of interest (of the sun, for example) and the zenith (directly overhead).

Appendix C: QUALITY CONTROL PROGRAM

Appendix C: QUALITY CONTROL PROGRAM

```
PROGRAM FOR OUTLIER ANALYSIS:
        Parameter (kpts=9,ncoef=3)
        real am(kpts),bm(kpts),coef(ncoef)
        Dimension xlc(3),xtc(3)
        Dimension NYR(200000),NMT(200000),NDY(200000),GHR(200000),
   *XGRAD(200000),XDRAD(200000),XSf(200000),
   *IXCC1(200000),IXCC2(200000),IXCC3(200000),IXCC4(200000),a(200000),
   *b(200000),polyUp(200000), polyLw(200000)
        Dimension bsum(20),nsum(20),savrg(20),
   *xdr(50000),wavrg(20),wsum1(20),wsum2(20),wsum3(20),wsum4(20),
   *wsd(20)
        real intlmt(20),intwid
        Character*1 Anum
  1 format(A1)
   PI=3.14159
   DTOR=3.14159/180.0

DATA INPUT BY USER
    WRITE(*,*) 'SELECT THE SYSTEM USED FOR TIME LOG'
    WRITE(*,*) 'INPUT "1" FOR SOLAR, OR "2" FOR CLOCK TIME'
    READ(*,*) NTIMES
        write(*,*) 'Provide station height'
        read (*,*) htmasl
    write(*,*) 'Provide factor of standard deviation & Intervals'
        read(*,*) SDfactor, nintval
        write(*,*) 'Provide DR value for envelope cut-off'
    write(*,*) 'Use 0.2 as default and change after visual inspection'
        read(*,*) PolCut
    IF(NTIMES.EQ.2) THEN
    WRITE(*,*) 'INPUT LAT, LONG, STD. TIME MERIDIAN (real values)'
    WRITE(*,*) 'NORTH = +,  WEST = + '
    READ(*,*) YLAT,YLONG,YRLONG
    ELSE
    WRITE(*,*) 'INPUT LATITUDE (real value)'
    WRITE(*,*) 'NORTH = +'
    READ(*,*) YLAT
    YRLONG=0.0
    YLONG=0.0
    ENDIF

INPUT AND OUTPUT DATAFILES
        Open(unit=1,file='Solrad1.Prn')
        Open(unit=2,file='Envelope.dat')
        Open(unit=3,file='Outlier.dat')
        Open(unit=4,file='Gooddata.dat')
        Open(unit=5,file='crude.dat')
 801   Format(2F7.3,I6,4F7.3)
 901    FORMAT(3F7.3)
1001    Format(3I6,f5.1,2f7.1,f5.2,4I3,4f6.3)

CALCULATION SETUP
        nintval=nintval+1
        NITER=0
        miter=0
        Ngood=0
        Nbad=0
```

```
    do i=1,200000

READ RAW DATA INPUT FILE
    READ(1,*,END=90321) IYR,IMT,IDY,HR,GRAD,DRAD,SF,ICC1,ICC2,
   *ICC3,ICC4
        IF(GRAD.GT.10.0.AND.DRAD.GT.10.0) THEN

CALCULATE GLOBAL HOUR ANGLE AND DECLINATION
    XLCT=HR
    UT=XLCT+(YRLONG/15.0)
    IF (IMT.GT.2) THEN
    IYR1=IYR
    IMT1=IMT-3
    ELSE
    IYR1=IYR-1
    IMT1=IMT+9
    ENDIF
    INTT1=INT(30.6*IMT1+0.5)
    INTT2=INT(365.25*(IYR1-1976))
    SMLT=((UT/24.0)+IDY+INTT1+INTT2-8707.5)/36525.0
    EPSILN=23.4393-0.013*SMLT
    CAPG=357.528+35999.050*SMLT
    IF(CAPG.GT.360.0) THEN
    G360=CAPG-INT(CAPG/360.0)*360.0
    ELSE
    G360=CAPG
    ENDIF
    CAPC=1.915*SIN(G360*DTOR)+0.020*SIN(2*G360*DTOR)
    CAPL=280.460+36000.770*SMLT+CAPC
    IF(CAPL.GT.360.0) THEN
    XL360=CAPL-INT(CAPL/360.0)*360.0
    ELSE
    XL360=CAPL
    ENDIF
    ALPHA=XL360-2.466*SIN(2*XL360*DTOR)+0.053*SIN(4*XL360*DTOR)
    GHA=15.0*UT-180.0-CAPC+XL360-ALPHA
    IF(GHA.GT.360.0) THEN
    GHA360=GHA-INT(GHA/360.0)*360.0
    ELSE
    GHA360=GHA
    ENDIF
    DEC=ATAN(TAN(EPSILN*DTOR)*SIN(ALPHA*DTOR))/DTOR

CALCULATE SOLAR HOUR ANGLE
    IF(NTIMES.EQ.2) THEN
    SHA=GHA360-(YLONG)
    ELSE
    SHA=GHA360
    ENDIF

CALCULATE APPARENT SOLAR TIME
    IF(NTIMES.EQ.2) THEN
    AST=12.0+(SHA/15.0)
    ELSE
    AST=XLCT
    ENDIF

CALCULATE SOLAR ALTITUDE
    TRM111=SIN(YLAT*DTOR)*SIN(DEC*DTOR)
    TRM112=COS(YLAT*DTOR)*COS(DEC*DTOR)*COS((SHA+180.0)*DTOR)
```

```
     TRM11=TRM111-TRM112
     SOLALT=ASIN(TRM11)/DTOR
     IF(SOLALT.GT.7.0) THEN

CALCULATE DAY NUMBER
     DN1=(IDY+INTT1+INTT2)
     IMT9=1
     IYR1=IYR-1
     IMT1=IMT9+9
     INTT1=INT(30.6*IMT1+0.5)
     INTT2=INT(365.25*(IYR1-1976))
     DN2=(INTT1+INTT2)
     DN=DN1-DN2

CALCULATE HORIZONTAL EXTRATERRESTRIAL IRRADIANCE
     ERAD=1367.0*(1.+0.033*COS(0.0172024*DN))*SIN(SOLALT*DTOR)
     XMAXI0=1367.0*(1.+0.033*COS(0.0172024*DN))

CALCULATE kt AND k PAIRS
          xa=GRAD/ERAD
          xb=DRAD/GRAD
          if(xa.gt.0.0.and.xa.le.1.0) then
          if(xb.gt.0.0.and.xb.le.1.0) then
          NITER=NITER+1

!---------CLEAR-SKY IRRADIANCE - JOHN PAGE-----------------------
          XTLK=2.5
CALCULATE RELATIVE AIR MASS
     XAM=1./(SIN(SOLALT*DTOR)+0.50572*(SOLALT+6.07995)**(-1.6364))

CORRECTION FACTOR Kd
     Xjang=1.0*dn/365.25
     Xjangr=Xjang*dtor
     xkd=1.0+0.03344*COS(Xjangr-.048869)
     IF(xam.le.20.0) THEN
     Rd=1.0/(6.6296+1.7513*xam-.1202*(xam**2)+.0065*(xam**3)
    1-.00013*(xam**4))
     ELSE
     Rd=1.0/(10.4+.718*xam)
     ENDIF
     xLM=XTLK*.8662

BEAM HORIZONTAL (BCH) IRRADIANCE
     bch=1367.0*xkd*EXP(-xam*xLM*Rd)*sin(solalt*dtor)
     IF(bch.lt.0.0) THEN
     bch=0.0
     endif

CLEAR SKY DIFFUSE MODEL
     Trdiff=(-21.657 +41.752*XTLK+.51905*(XTLK**2))
     C0=0.26463-0.061581*XTLK+0.0031408*(XTLK**2)
     IF(C0*Trdiff.lt.3.0) THEN
     C0=3.0/Trdiff
     endif
     C1=2.0402+0.018945*XTLK-0.011161*(XTLK**2)
     C2=-1.3025+0.039231*XTLK+0.0085079*(XTLK**2)

CALCULATE CLEAR DAY DIFFUSE IRRADIANCE
     dch=xkd*Trdiff*(C0+C1*SIN(solalt*dtor)+C2*
    1(SIN(solalt*dtor)**2))
```

```
    IF(dch.lt.0.0) THEN
    dch=0.0
    endif

CALCULATE CLEAR DAY GLOBAL IRRADIANCE
    gch=bch+dch

!-----END OF CLEAR-SKY IRRADIANCE MODULE---------------

        if(Grad.le.gch) then
        icodgc=2
        endif
        if(Drad.ge.dch.and.Drad.le.572*trm11) then
        icoddc=2
        endif
        if(icodgc.eq.2.and.icoddc.eq.2) then
        miter=miter+1
        NYR(NITER)=IYR
        NMT(NITER)=IMT
        NDY(NITER)=IDY
        GHR(NITER)=HR
        XGRAD(NITER)=GRAD
        XDRAD(NITER)=DRAD
        XSF(NITER)=SF
        IXCC1(NITER)=ICC1
        IXCC2(NITER)=ICC2
        IXCC3(NITER)=ICC3
        IXCC4(NITER)=ICC4
    a(NITER)=xa
        b(NITER)=xb
        write(5,1001)NYR(niter),NMT(niter),NDY(niter),GHR(niter),
   *XGRAD(niter),XDRAD(niter),XSF(niter),IXCC1(NITER),
   *IXCC2(NITER),IXCC3(NITER),IXCC4(NITER),a(niter),b(niter)
    ENDIF
        ENDIF
        ENDIF
        ENDIF
        endif
        end do
90321 CONTINUE

QUALITY CONTROL OF GLOBAL & DIFFUSE SOLAR RADIATION DATA

MAXIMISATION / MINIMISATION OF THE ENTIRE DATASET
        amx = a(1)
        amn = a(1)
        do j = 2,NITER
        If (a(j) .gt. amx) Then
        amx = a(j)
        End If
        If (a(j) .lt. amn) Then
        amn = a(j)
        End If
        end do

DEVELOP BANDWIDTHS
        intwid = (amx - amn) * 1 / (1.0*(nintval-1))
        do k=1,nintval
        intlmt(k)=amn+intwid*(k-1)
        end do
```

QUARTILE ANALYSIS

```
      do k = 2,nintval
      npts=0
      do i = 1,NITER
      If (a(i).ge.intlmt(k-1).And.a(i).lt.intlmt(k)) Then
      npts=npts+1
      xdr(npts)=b(i)
      end if
      end do
      call Sortit(xdr,npts)
      if(npts.gt.0) then
      write(2,801) intlmt(k - 1),intlmt(k),npts,xdr(1),
     *xdr(int(0.25*(npts-1))+1),xdr(int(0.50*(npts-1))+1),
     *xdr(int(0.75*(npts-1))+1),xdr(npts)
      endif
      end do
```

MEAN, WEIGHTED-MEAN & STANDARD DEVIATION

```
      do k = 1,10
      bsum(k) = 0.0
      nsum(k) = 0
      wsum1(k) = 0.0
      wsum2(k) = 0.0
      wsum3(k) = 0.0
      wsum4(k) = 0.0
      end do
      do i = 1,NITER
      do k = 2,nintval
      If (a(i).ge.intlmt(k-1).And.a(i).lt.intlmt(k)) Then
      bsum(k) = bsum(k) + b(i)
      nsum(k) = nsum(k) + 1
      If (nsum(k).gt.0) Then
      savrg(k) = bsum(k) / (1.0*nsum(k))
      Else
      savrg(k) = -99999.9
      End If
      End If
      end do
      end do
      do i = 1,NITER
      do k = 2,nintval
      If (a(i) .ge. intlmt(k - 1) .And. a(i) .lt. intlmt(k)) Then
      bdep = Abs(b(i) - savrg(k))
      If (bdep .ne. 0.0) Then
      wt = 1.0 / bdep
      Else
      bdep = 0.0
      wt = 1.0
      End If
      wsum1(k) = wsum1(k) + wt * b(i)
      wsum2(k) = wsum2(k) + wt

      If (nsum(k) .gt. 0) Then
      wavrg(k) = wsum1(k) / wsum2(k)
      Else
      wavrg(k) = -99999.9
      End If
      End If
      end do
```

```
        end do
        do i = 1,NITER
        do k = 2,nintval
        If (a(i) .ge. intlmt(k - 1) .And. a(i) .lt. intlmt(k)) Then
        bdep2 = (b(i) - wavrg(k)) ** 2
        wsum3(k) = wsum3(k) + bdep2
        wsum4(k) = wsum4(k) + 1.0
        If (nsum(k) .gt. 0) Then
        wsd(k) = Sqrt(wsum3(k) / wsum4(k))
        Else
        wsd(k) = -99999.9
        End If
        End If
        end do
        end do
        do k = 2,nintval
        If (nsum(k) .gt. 0) Then
        write(2,901) savrg(k),wavrg(k),wsd(k)
        endif
        end do

OUTLIER IDENTIFICATION

LOWER CURVE OF ENVELOPE
        do i=1,kpts
        am(i)=0.5*(intlmt(i)+intlmt(i+1))
        bm(i)=Max(0.0,(wavrg(i+1)-SDfactor*wsd(i+1)))
        end do
        Call polfit(am,bm,kpts,ncoef,coef)
        do j=1,ncoef
        xlc(j)=coef(j)
        end do

UPPER CURVE OF ENVELOPE
        do i=1,kpts
        am(i)=0.5*(intlmt(i)+intlmt(i+1))
        bm(i)=Min(1.0,(wavrg(i+1)+SDfactor*wsd(i+1)))
        end do
        Call polfit(am,bm,kpts,ncoef,coef)
        do j=1,ncoef
        xtc(j)=coef(j)
        end do

FIND UPPER AND LOWER POLYNOMIAL kt LIMIT
    deltaUP=xtc(2)*xtc(2)-4*xtc(3)*(xtc(1)-1.0)
        deltaLW=xlc(2)*xlc(2)-4*xlc(3)*xlc(1)
        deltaCUT=xlc(2)*xlc(2)-4*xlc(3)*(xlc(1)-PolCut)
        if (deltaUP.ge.0.0 .and. deltaLW.ge.0.0) then
        PlimUP=(-xtc(2)-sqrt(deltaUP))/(2*xtc(3))
        PlimLW=(-xlc(2)-sqrt(deltaLW))/(2*xlc(3))
        Cutoff=(-xlc(2)-sqrt(deltaCUT))/(2*xlc(3))
        else
        PlimUP=0.3
        PlimLW=0.7
        CUTOFF=0.1
        end if

DEFINE POLYNOMIALS
        do i = 1,NITER
        if(a(i).le. PlimUP) then
```

```
      polyUp(i)=1.0
      else
      polyUp(i)=xtc(1)+xtc(2)*a(i)+xtc(3)*a(i)*a(i)
      endif
      if(a(i).le. PlimLW. and. a(i).gt. cutoff) then
      PolyLW(i)=xlc(1)+xlc(2)*a(i)+xlc(3)*a(i)*a(i)
      else if (a(i).le. cutoff) then
      PolyLW(i)=polcut
      else
      PolyLW(i)=0.0
      endif
      do k = 2,nintval
      If (a(i) .ge. intlmt(k - 1) .And. a(i) .lt. intlmt(k)) Then

CHECK FOR kt
      if(a(i).gt.PlimUP) then
      if(b(i).lt.Max(0.0,(PolyLW(i))).or.b(i)
      *.gt.Min(1.00,(PolyUP(i))))then
      Nbad=Nbad+1
      write(3,1001)NYR(I),NMT(I),NDY(I),GHR(I),XGRAD(I),XDRAD(I),
      *XSF(I),IXCC1(I),IXCC2(I),IXCC3(I),IXCC4(I),a(i),b(i),PolyUP(i),
   *PolyLW(i)
      endif
      elseif (b(i).lt.Max(0.0,(PolyLW(i))))
   *  then
      Nbad=Nbad+1
      write(3,1001)NYR(I),NMT(I),NDY(I),GHR(I),XGRAD(I),XDRAD(I),
      *XSF(I),IXCC1(I),IXCC2(I),IXCC3(I),IXCC4(I),a(i),b(i),PolyUP(i),
   *PolyLW(i)
      endif
      if(a(i).gt.PlimUP) then
      if(b(i).ge.Max(0.0,(PolyLW(i))).and.b(i)
      *.le.Min(1.0,(PolyUP(i))))
   *  then
      Ngood=Ngood+1
      write(4,1001)NYR(I),NMT(I),NDY(I),GHR(I),XGRAD(I),XDRAD(I),
      *XSF(I),IXCC1(I),IXCC2(I),IXCC3(I),IXCC4(I),a(i),b(i),PolyUP(i),
   *PolyLW(i)
      endif
      endif
      if(a(i).le.PlimUP) then
      if(b(i).ge.Max(0.0,(PolyLW(i)))) then
      Ngood=Ngood+1
      write(4,1001)NYR(I),NMT(I),NDY(I),GHR(I),XGRAD(I),XDRAD(I),
      *XSF(I),IXCC1(I),IXCC2(I),IXCC3(I),IXCC4(I),a(i),b(i),PolyUP(i),
   *PolyLW(i)
      endif
      endif
      endif
      end do
      end do
      write(*,*)'Total data points=',NITER
      write(*,*)'No. of points passing Page tests=',miter
      write(*,*)'No. of points that cleared all tests=',Ngood
      write(*,*)'Rejected no. of points=',Nbad
      write(*,*) 'Key-in any alphanumeric key to exit'
      read(*,1) Anum
      end
```

POLYNOMIAL FIT

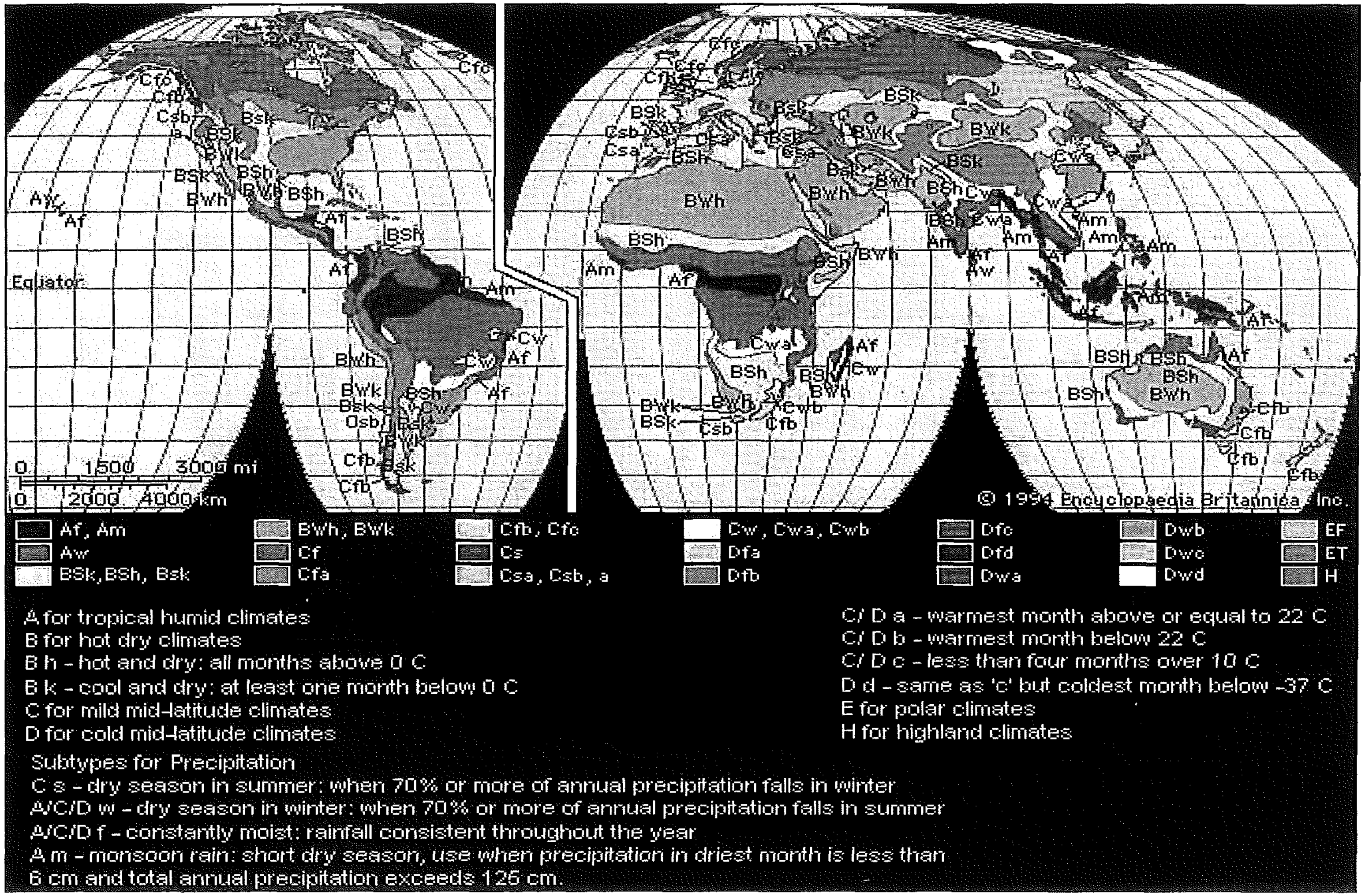
```
      Subroutine polfit(a,b,mpts,mcoef,coef)
      real a(mpts),b(mpts),coef(mcoef),P(2*(mcoef-1)),S(10,10),
  *countx(2*(mcoef-1)),county(2*(mcoef-1))
      np=mcoef-1
      np2=2*np
      do j=1,np2
      countx(j)=0.0
      County(j)=0.0
      end do
      do i=1,mpts
      do j=1,np2
      countx(j) = countx(j) + a(i)**j
      county(j) = county(j) + b(i)*(a(i)**(j-1))
      end do
      end do

MATRIX INVERSION ROUTINE
      n = mcoef
      N1 = n + 1
      N2 = n + n
      do i = 1,10
      do j = 1,10
      S(i, j) = 0.0
      end do
      end do
      do k=1,mcoef
      P(k)=county(k)
      end do
      do nfirst=1,mcoef
      do nsecnd=1,mcoef
      if(nfirst.eq.1.and.nsecnd.eq.1) then
      S(nfirst,nsecnd)=mpts
      else
      S(nfirst,nsecnd)=countx(nsecnd-1+nfirst-1)
      endif
      end do
      end do
      i = 1
      do j = N1,N2
      S(i, j) = 1.0
      i = i + 1
      end do
      do i = 1,n
      X = S(i, i)
      do j = 1,N2
      S(i, j) = S(i, j) / X
      end do
      do k = 1,n
      If (i.eq.k) Then
    GoTo 251
      End If
      ax = S(k, i)
      do j = 1, N2
      S(k, j) = S(k, j) - S(i, j) * ax
      end do
 251  continue
      end do
      end do
      do i = 1,n
      Sum = 0.0
```

```
          do k = 1,n
          Sum = Sum + S(i, k + n) * P(k)
          end do
          coef(i)=Sum
          end do
          return
          end
!Return p,q in ascending order
          Subroutine Order(p,q)
          real p,q,temp
           if (p.gt.q) then
            temp=p
           p=q
           q=temp
          end if
          return
          end
!Sorting of array A
          Subroutine Sortit(A, n)
          real A(1:n)
           do i=1, n
            do j=n, i+1, -1
call Order(A(j-1), A(j))
            end do
           end do
           return
          end
```

APPENDIX D

CLIMATOLOGY MAPS



A for tropical humid climates
B for hot dry climates
B h - hot and dry: all months above 0 C
B k - cool and dry: at least one month below 0 C
C for mild mid-latitude climates
D for cold mid-latitude climates

C/ D a - warmest month above or equal to 22 C
C/ D b - warmest month below 22 C
C/ D c - less than four months over 10 C
D d - same as 'c' but coldest month below -37 C
E for polar climates
H for highland climates

Subtypes for Precipitation
C s - dry season in summer: when 70% or more of annual precipitation falls in winter
A/C/D w - dry season in winter: when 70% or more of annual precipitation falls in summer
A/C/D f - constantly moist: rainfall consistent throughout the year
A m - monsoon rain: short dry season, use when precipitation in driest month is less than 6 cm and total annual precipitation exceeds 125 cm.

Figure D.1. World climatic map based on the Koppen classification.

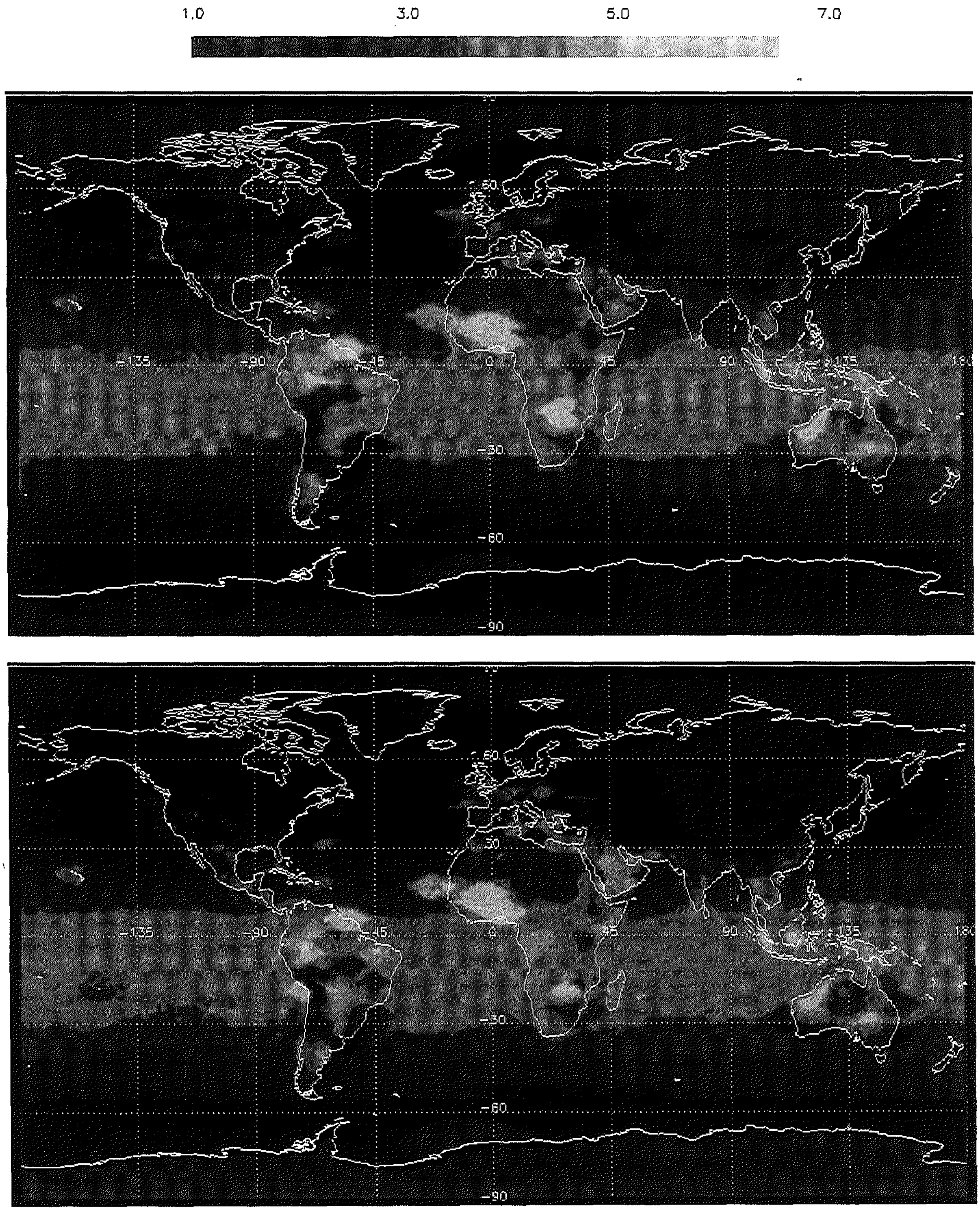


Figure D.2 . Linke Turbidity at air mass 2 monthly mean world map for January and February. (Courtesy of SODA, www.soda-is.org).

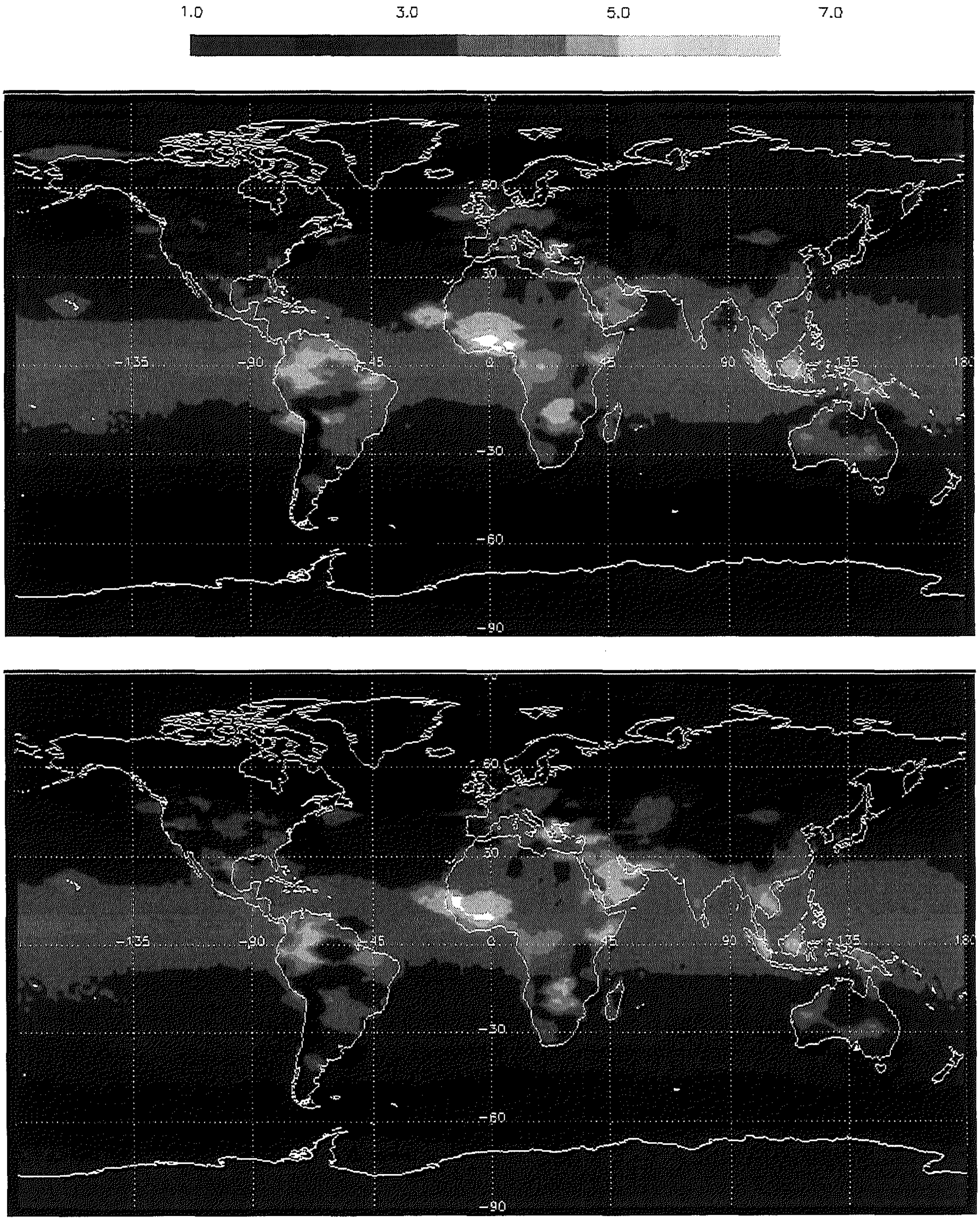


Figure D.3. Linke Turbidity at air mass 2 monthly mean world map for March and April. (Courtesy of SODA, www.soda-is.org).

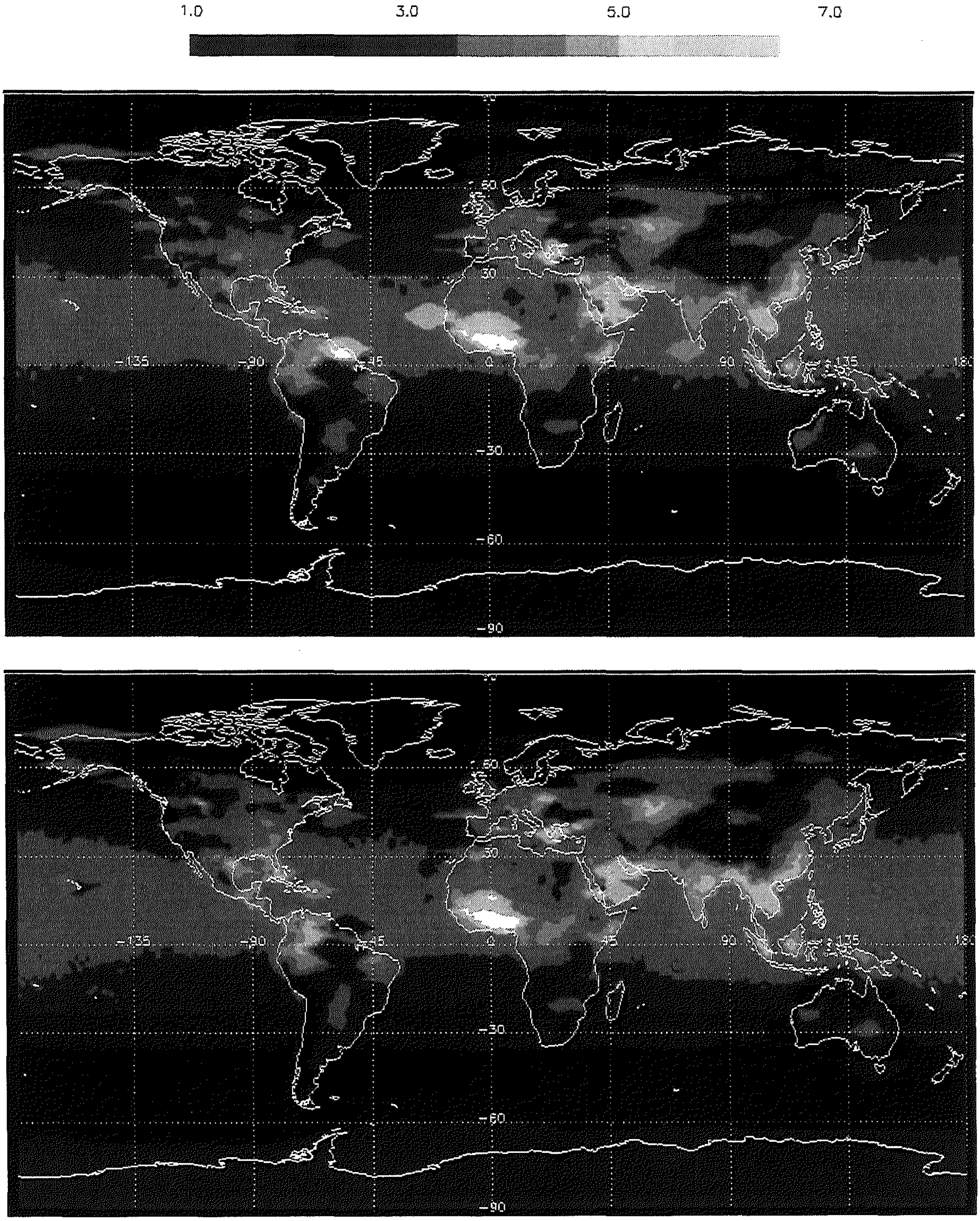


Figure D.4. Linke Turbidity at air mass 2 monthly mean world map for May and June. (Courtesy of SODA, www.soda-is.org).

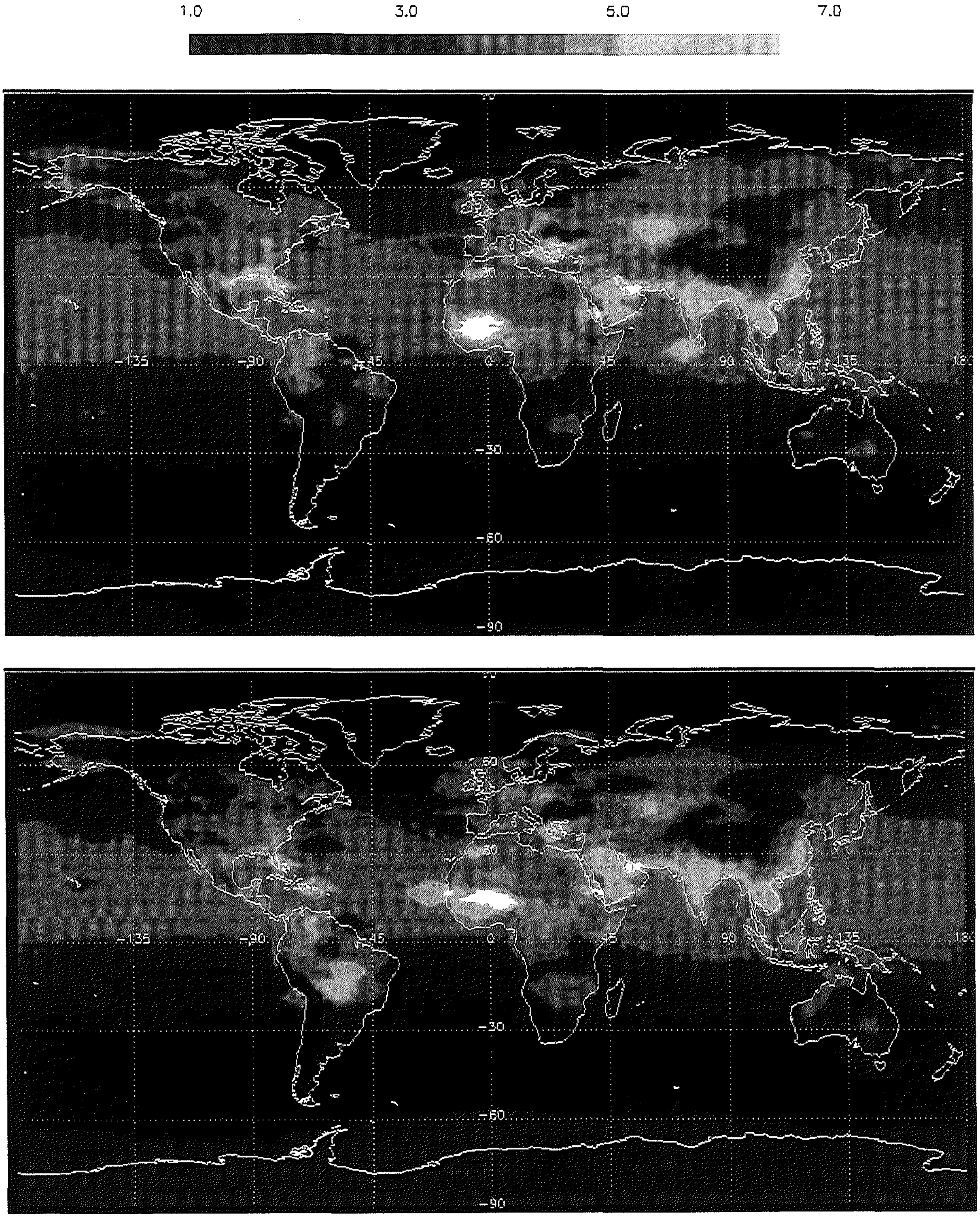


Figure D.5. Linke Turbidity at air mass 2 monthly mean world map for July and August. (Courtesy of SODA, www.soda-is.org).

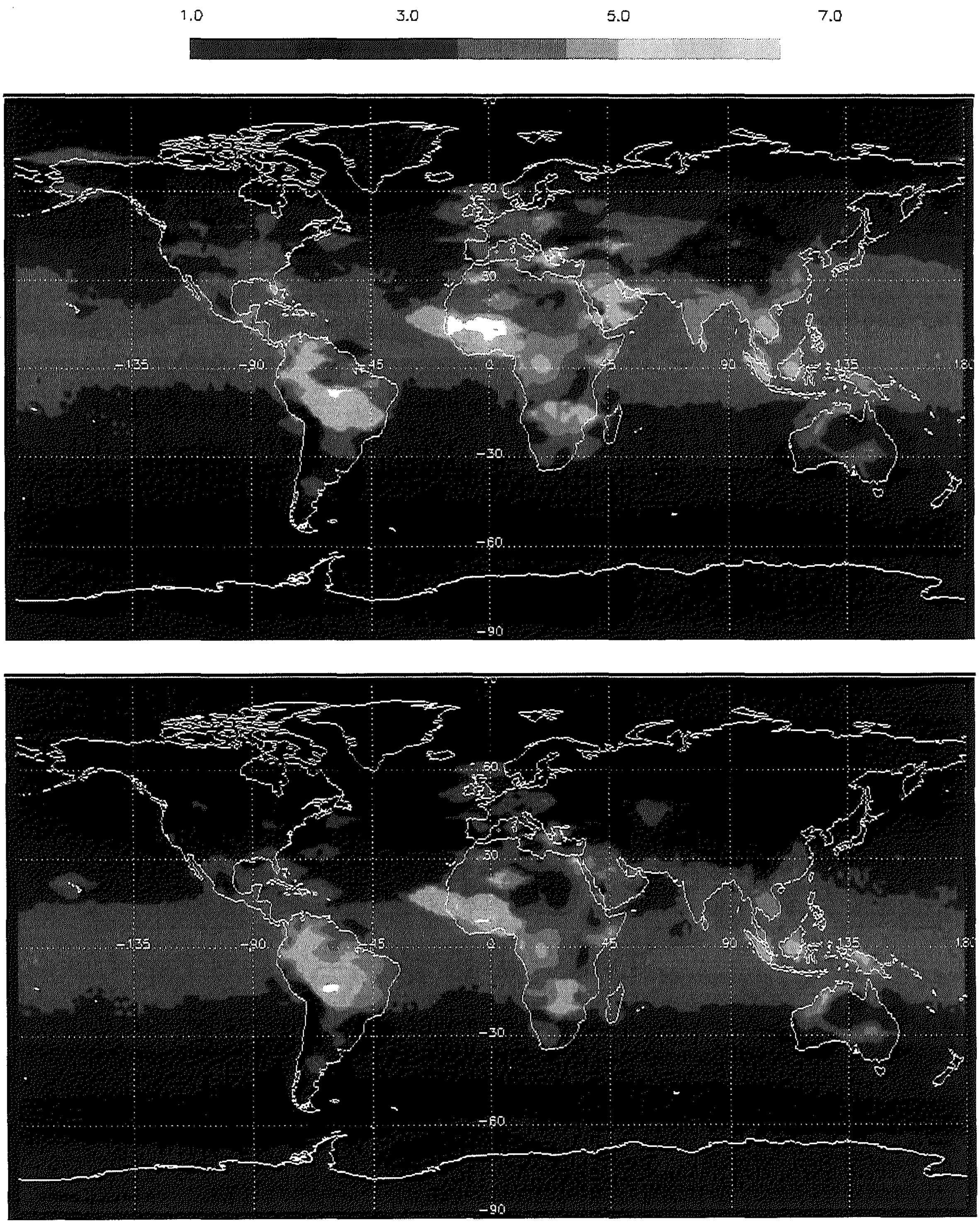


Figure D.6. Linke Turbidity at air mass 2 monthly mean world map for September and October. (Courtesy of SODA, www.soda-is.org).

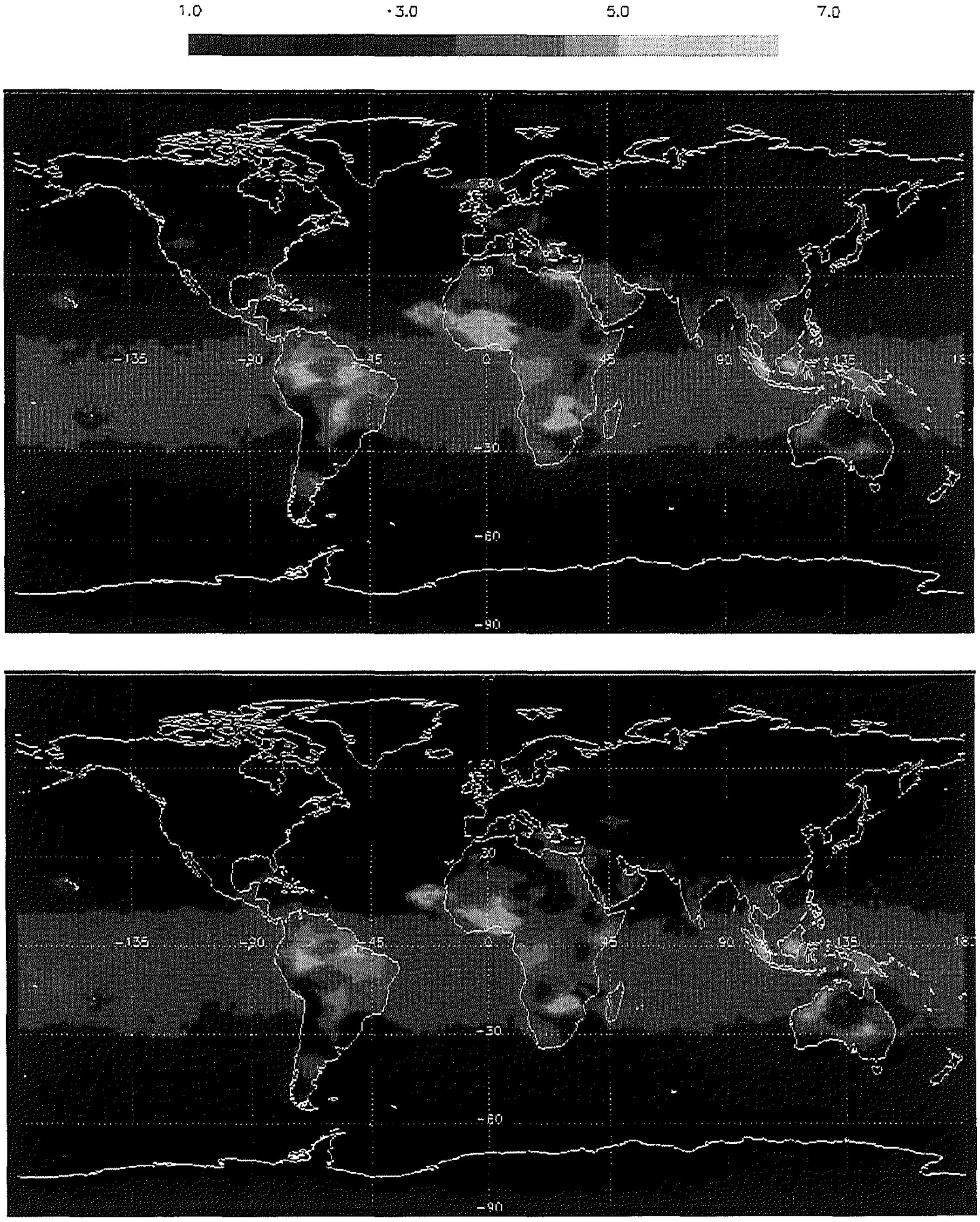


Figure D.7. Linke Turbidity at air mass 2 monthly mean world map for November and December. (Courtesy of SODA, www.soda-is.org).

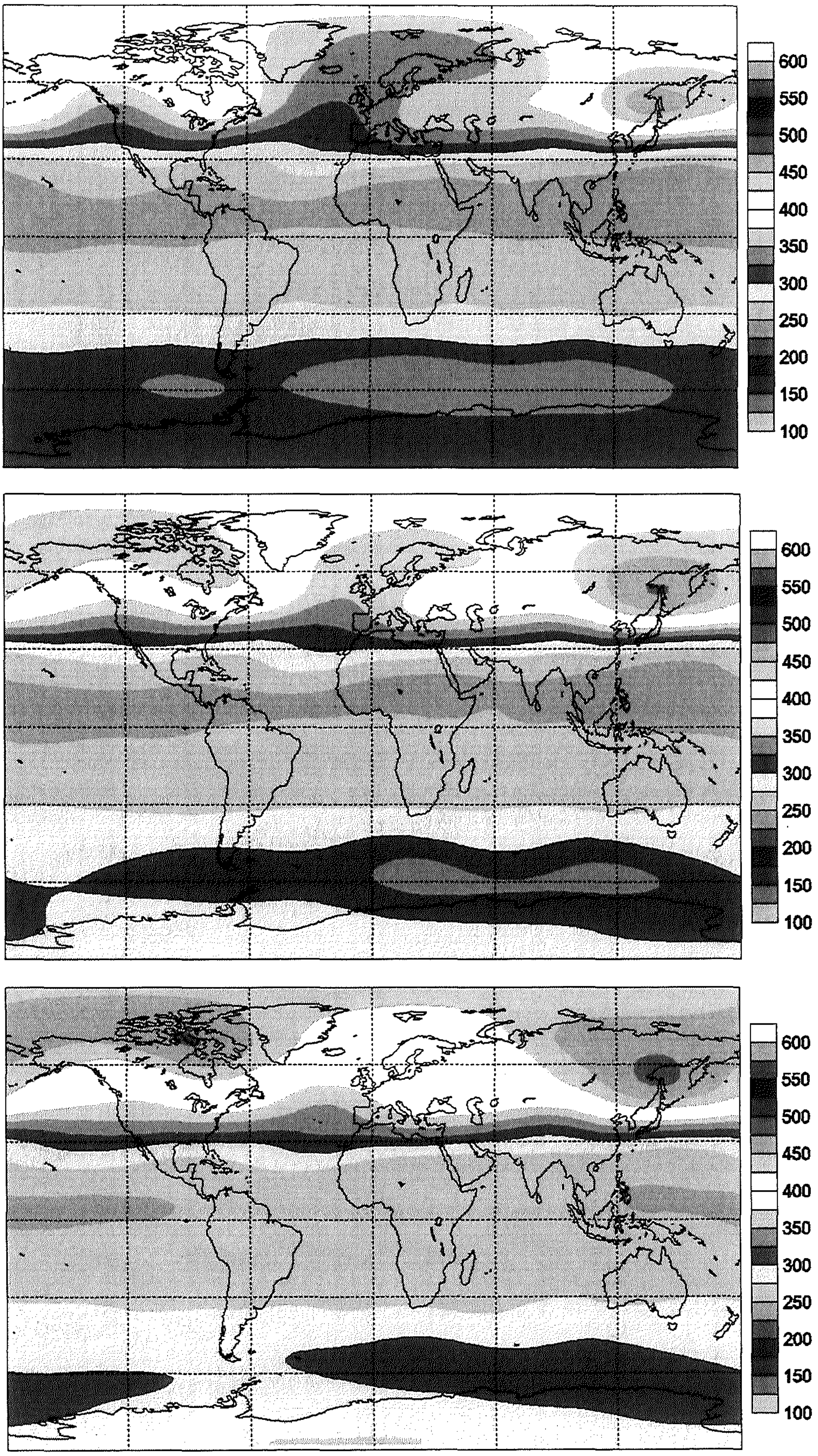


Figure D.8. Ozone depth monthly mean world map for January, February and March. (Courtesy of SODA, www.soda-is.org).

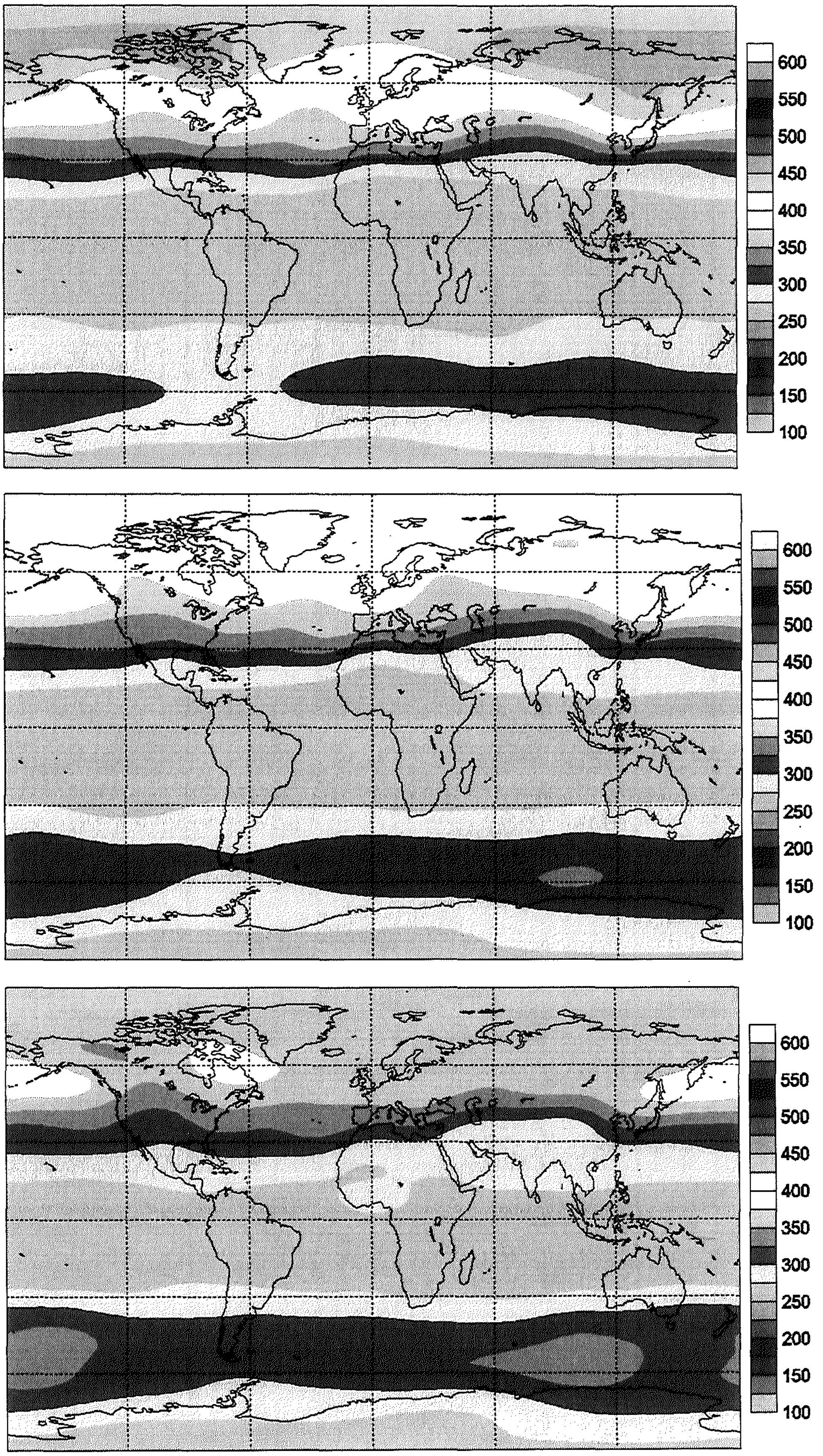


Figure D.9. Ozone depth monthly mean world map for April, May and June. (Courtesy of SODA, **www.soda-is.org**).

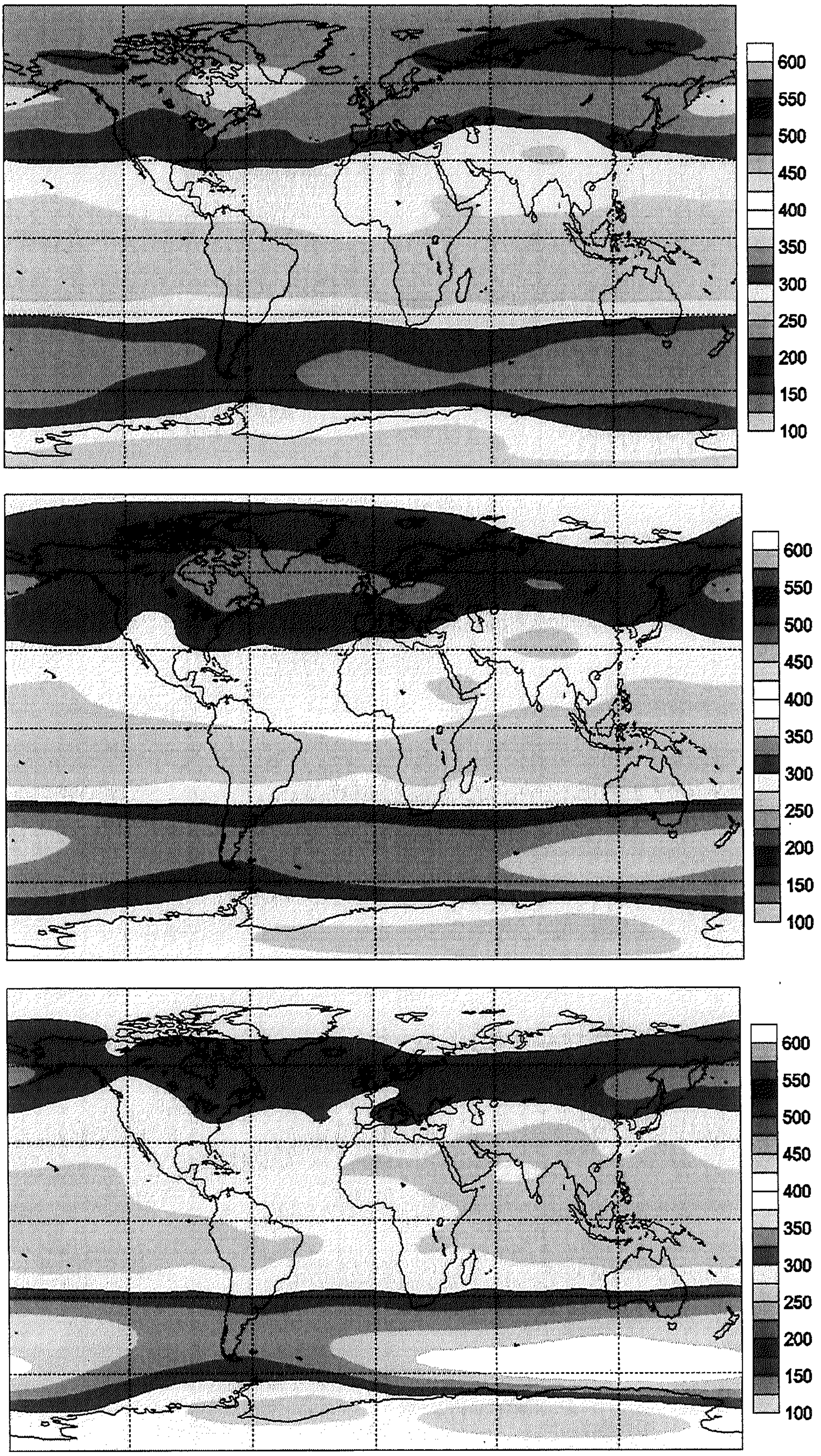


Figure D.10. Ozone depth monthly mean world map for July, August and September. (Courtesy of SODA, www.soda-is.org).

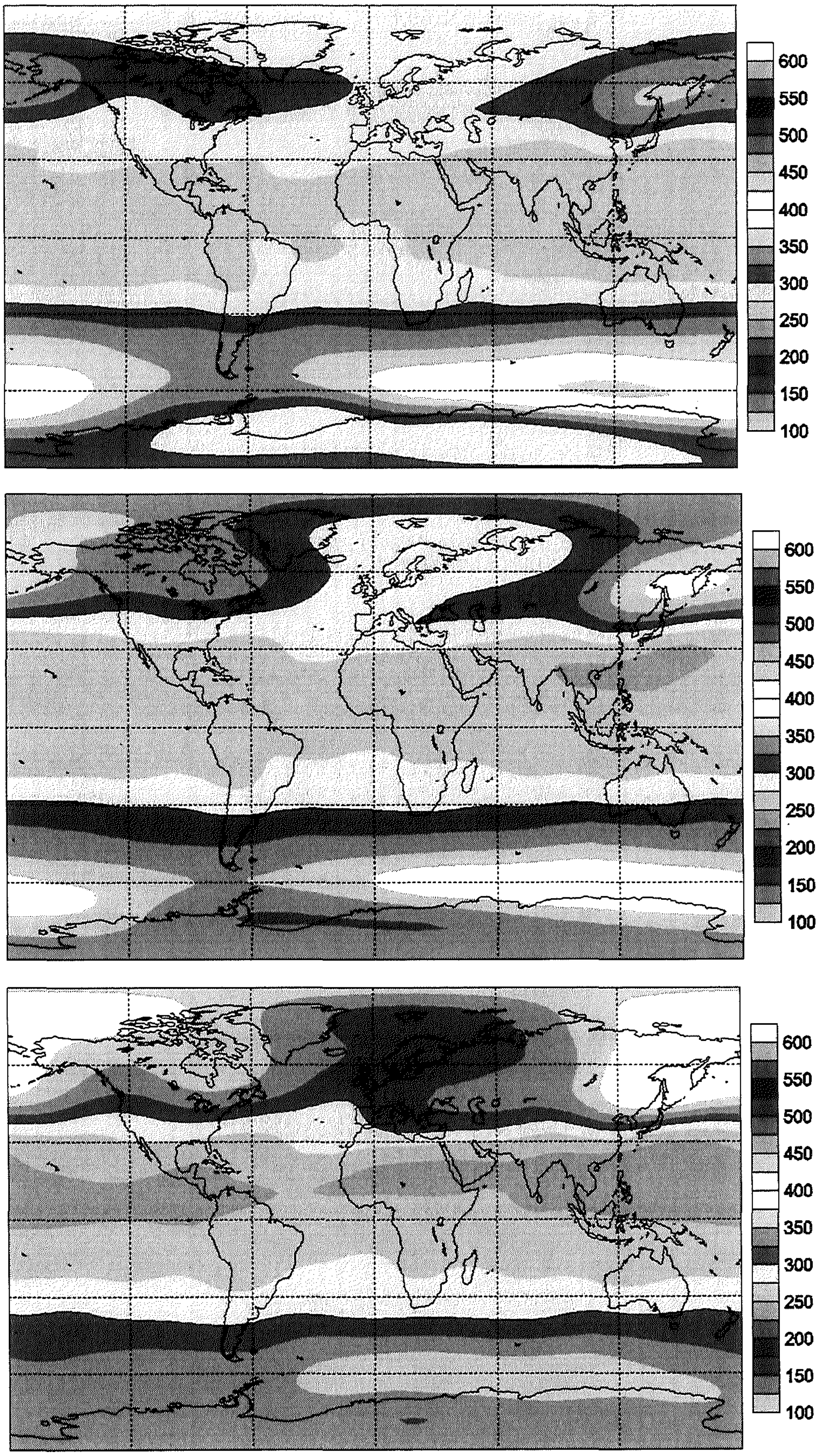


Figure D.11. Ozone depth monthly mean world map for October, November and December. (Courtesy of SODA, www.soda-is.org).