Assesing the potential of using satellite data on radiation in the analysis of earth's climate system to complement the scarce radiation data measured from the ground stations

Collins Ochieng ONYAGA (collins@aims.ac.tz)[1], Samson Wangila Wanyonyi(swangila@aims.ac.tz)[2], and Roger Stern(r.d.stern@reading.ac.uk)[3]

[1]African Institute of Mathematical Science Tanzania Alumni.
[2]PhD student in Biostatistics, department of Mathematics and Computer Science, University of Eldoret.
[3] Department of statistics University of Readings, UK.

Abstract

High quality solar radiation data is required for the appropriate monitoring and analysis of the Earth's climate system as well as efficient planning and operation of solar energy systems. However, well maintained radiation measurements are rare in many regions of the world. Therefore, satellite-derived radiation estimates are an alternative to these scarce solar radiation measurements from the weather stations. Satellite estimates of solar radiation have an advantage over solar radiation measurements from weather stations because of their high spatial and temporal resolutions. These satellite radiation estimates at approximately 5-6 Km resolution derived from geostationary Meteosat satellites are available through the EUMETSAT Satellite Application Facilities (SAFs). CM-SAF (SAF on Climate Monitoring) provides consistent dataset of hourly, daily and monthly solar radiation from 1983 to 2013. In this study, we examined the potential of using satellite estimates of solar radiation to fill in the data gaps in records from the weather stations as well as the areas where radiation data is not available. The analysis carried out showed that the satellite data had fewer missing values than the ground data, and that they are both similar in distribution. The average correlation between the two data sets was found to be 0.71 for both monthly and daily analysis. However, the month of September showed a very low correlation of 0.21. Mean percentage error, mean bias error and mean absolute deviation were found to be 2.46, 18.84, 50.32 and 3.08, 559.87, 1135.93 for daily and monthly analysis, respectively.

The solar radiation distribution in Dodoma was found to follow Weibull distribution throughout the year.

Keywords: Weibull Distribution, CM-SAF ,EUMETSAT, geostationary Meteosat satellites.

1. INTRODUCTION

Most of the biochemical processes that occur on the surface of the earth, such as photosynthesis and plant development, derive their energy from the sun Evrendilek and Ertekin (2008). Solar radiation is also important in determining the performance and monitoring the working of solar devices such as solar furnaces, solar collectors, and photovoltaics. It is therefore important to get solar radiation data of high quality at every place where these applications are applied for appropriate monitoring.

Solar radiation data are important for many application fields such as ecology, biodiversity, hydrology, agriculture, meteorology, and climatology. For example,

1. Hydrology: Most of the hydrological models require weather variables such as precipitation, maximum and minimum temperatures and solar radiation (Richardson, 1981).

2. Agriculture: Farmers are faced with a challenge of choosing which crop to plant in a given area. This is normally influenced by a number of factors such as soil moisture of the area and plant's rate of evapotranspiration. Soil moisture depends on the rate of evaporation in the soil, which to a large extent, depends on solar radiation. To understand this basic physiology of crop growth and development, many crop models have been developed (Whisler et al., 1986). These models are of great importance in quantifying the environmental limits to specific crop production at a given area. They are also applicable in prediction of crop yields. According to (Nemani et al., 2003), solar radiation is an important factor in crop growth in Europe and other areas of the world (*Figure* 1.1). This justifies the importance of accurate measurement of radiation data.

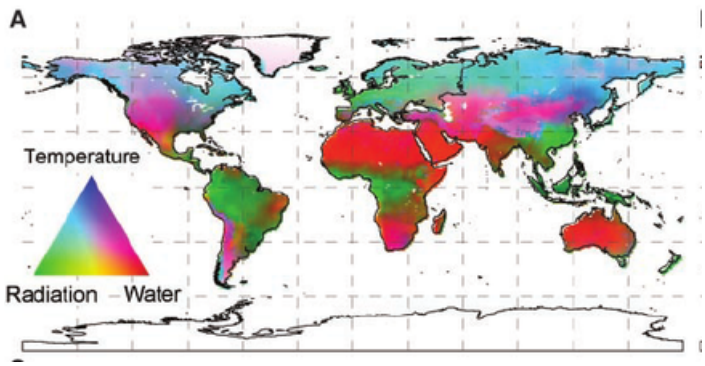

Figure 1.1: Geographic distribution of potential climatic constraints,

temperature, radiation and water, to plant growth derived from long-term climate statistics. (Nemani et al., 2003)

In the recent past, there has been a lot of work on examining the feasibility of solar cooking and solar water purification. Development of these technologies requires accurate information on hourly solar radiation as this will help in knowing when to use these appliance. The knowledge of the hourly solar radiation is also vital in determining how much of the energy can be stored for use when the sun goes down. According to (Cuce and Cuce, 2013), "it is anticipated that solar cooking technology will be demanded by a huge group of people in the near future because of its outstanding features". Therefore, accurate measurement of solar radiation in most parts of the world will be of great importance to help in designing the solar cooker and solar water purifiers that can be can be used even in areas that receive low solar irradiance.

However, location-specific and accurate solar radiation data is still a great challenge for most of the African countries, even though the problem has been discussed at many local and global meetings including the World Meteorological Organization (WMO) Regional Association (Africa) sessions that

are held every four years Jarraud (2008). This is mainly because there are few weather stations in Africa. On 6th Nov, 2006, Michel Jarraud said in a press conference that was held in Nairobi, Kenya, that Africa still needs approximately 200 automatic weather stations as a step towards rescuing historical data and also improving capacity building on climate and weather reporting. According to Medany et al. (2006), there are 1152 World Weather Watch (WWW) stations in Africa, giving a station density of 1 station per $26,000Km^2$ which is still 8 times below the minimum recommended level by WMO. In the working paper 3 of United Nations Economic Commission for African Climate Policy Centre, ACPC (2011), it is documented that the shortage of data in Africa is exacerbated by the uneven distribution of available weather stations leaving large areas unmonitored.

As a way to reduce the problem of insufficient climatic data in Africa, United Nations Development Programme rolled out a number of support programs that encourages climate-resilient economic development in Africa $(CIRDA)$ UNDP (2016). The main component of these programs is to provide credible and appropriate information on both weather and climate to help in decision making in most of the sectors that require these data UNDP (2016). To achieve this, Trans-African Hydro-meteorological Observatory (TAHMO) has been commissioned with the task of installing 20,000 on-the-ground automatic stations across Africa. These stations are to provide rainfall, temperature, and other critical data at high temporal resolution, which TAHMO plans to make freely available in their website. TAHMO is currently piloting this in several parts of Chad, Kenya, DRC, Uganda, among others.

Figure 1.2: Figure showing a weather Station installed at Homabay- Kenya

(TAHMO) Trans-African Hydro-meteorological Observatory (TAHMO) has the task of installing 20,000 stations in Africa to help solve the problem of insufficient climatic data in the continent.

In the absence of solar radiation measurements, a number of methods have been developed to derive estimates of solar radiation for places where it is not measured directly. The simplest of these methods is assigning the measured values from the nearest station to these places (Hunt et al., 1998) or using spatial interpolation methods (Bechini et al., 2000). Solar radiation, at times is also estimated using other meteorological observations that are available, such as sunshine duration, cloud cover and air temperature using empirical models Mavromatis (2008). The estimated radiation calculated from these models varies in accuracy depending on variables used Bojanowski et al. (2013). The model that uses the sunshine duration has been found to be more accurate, since it gives a measure of the time per day of direct solar radiation.

But since the number of weather stations is low, solar radiation data recorded or estimated may not be enough to create a time series of solar radiation that has a high spatial resolution Bojanowski (2014). Thus, estimated solar radiation data from sources that provide spatially - continuous information at high resolution such as data provided by the satellite have to be considered to supplement the station-specific estimates.

Satellite data on radiation provides continuous estimates of these radiation data with a very high spatial and temporal resolution of $0.05^\circ \times 0.05^\circ$ and 30 min, respectively, for over 30 years. These estimates from the satellite data are therefore worth investigating to see if they correlate well with ground truth, and hence can be used to fill the large data gaps.

The objective of this work was To determine the potential of using satellite-derived radiation data as a substitute of scarce radiation data measured from the ground stations.

2. Solar Radiation Data

Data as well as preparation of the data used in this work is presented here.

2.1 Description of the Data

2.1.1 Satellite Data on Radiation.

The data set that was used is called the Surface Solar Radiation Data Set-Heliosat (SARAH). It is a collection of satellite records of the solar surface irradiance (SIS), the surface direct normalized irradiance (DNI) and the effective cloud albedo (CAL) (Müller et al., 2015a). These records are available as hourly, monthly and yearly averages from $1983-01$ to $2013-12$, which have started to change as more recent data are being added to the data base, with a spatial resolution of $0.05° \times 0.05°$ and covering a region of $\pm 65°$ longitude and $\pm 65°$ latitude (Müller et al., 2015a) .

The SARAH data are derived from the Meteosat Visible and Infrared Imager (MVIRI) and Spanning Enhanced Visible and Infrared Imager (SEVIRI) instruments on the geostationary Meteosat satellites (Hammer et al., 2003). This is called the Heliosat Method.

The first step in this method is to retrieve CAL by using the normalized relation between all sky and clear sky reflection in the visible channel of the Meteosat instruments. CAL is then used to derive the cloud index (a measure for the impact of clouds on the clear-sky irradiance. The clear - sky irradiance is calculated using the all sky model SPECMAGIC (Mueller et al., 2012)). The combination of cloud index and clear sky irradiance gives SIS. The Surface Direct Iradiance (SID) is then derived using the diffuse model of (Skartveit et al., 1998) and the cloud index. The direct normalized irradiance (DNI) is derived from normalization of SID with the cosine of the solar zenith angle (SZA):

According to Mueller et al. (2012),

$$DNI = \frac{SID}{\cos(SZA)}. \tag{2.1.1}$$

Effective Cloud Albedo (CAL) is the normalized difference between the reflected irradiance for all sky and reflected irradiance for clear sky. It is dimensionless. CAL is a measure of the effect of cloud on the Earth's solar radiation budget, since it measures the cooling effect in the solar spectrum caused by clouds.

According to Müller et al. (2015b),

$$CAL = \frac{\rho - \rho_{cs}}{\rho_{max} - \rho_{cs}}, \tag{2.1.2}$$

where, ρ is the observed reflection for each pixel and time, ρ_{cs} is the clear sky reflection and ρ_{max} is the maximum reflection determined by the 95% percentile of all reflection values at local noon in a target region. From Posselt et al. (2011), the monthly, daily and hourly averages of CAL are calculated as

$$CAL_{mean} = \frac{1}{n}\left(\sum_{i=1}^{n} CAL_i\right), \tag{2.1.3}$$

where i is the index of the slots per hour for hourly average, index of hourly means for daily means, index for daily means for monthly mean calculations.

Surface Incoming Solar Radiation (SIS). This is the solar radiation flux on a horizontal Earth surface, expressed in Wm^{-2}. According to Mueller et al. (2011), it is calculated from CAL as;

$$SIS = SIS_{CLS}(1 - CAL), \tag{2.1.4}$$

where SIS is the solar surface irradiance, and SIS_{CLS} is clear sky irradiance. The SIS daily mean is

$$SIS_{DA} = SIS_{CLSDA}\left(\frac{\sum_{i=1}^{n} SIS_i}{\sum_{i=1}^{n} SIS_{CLS_i}}\right), \tag{2.1.5}$$

where SIS_{DA} is daily average of SIS, SIS_{CLSDA} is the daily averaged SIS for clear sky, SIS_i is calculated SIS for satellite image i, SIS_{CLS_i} is the calculated clear sky SIS for image i and n is the total number of images available during the day.

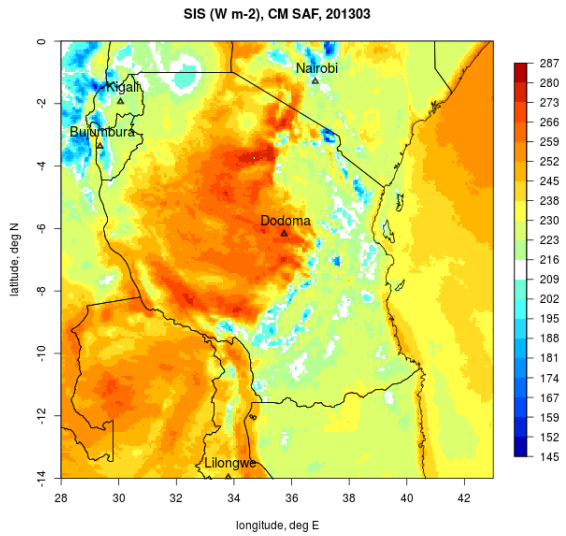

Figure 2.1: SIS plot for satellite data

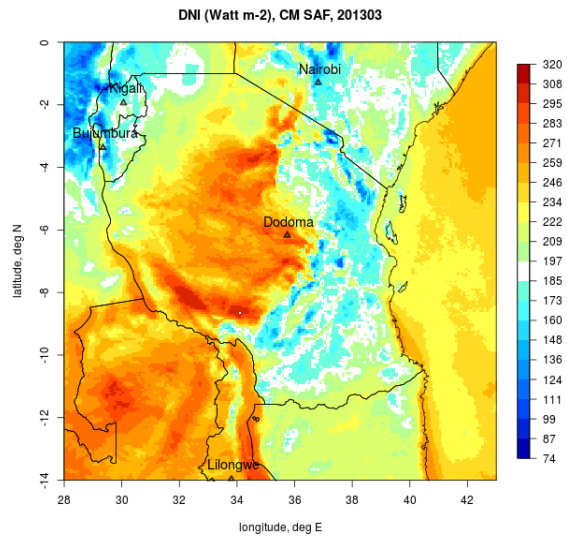

Figure 2.2: Average DNI over Tanzania in 2013

Direct Normalized Irradiance. This is the radiation from the Sun directly to the Earth. It has a fixed direction. It is expressed in Wm^{-2}. In sunny regions or during the summer, it accounts for $70\% - 80\%$ of the total radiation.

Figure 2.1 and Figure 2.2 shows examples of the average SIS and DNI derived from the satellite over Tanzania for March 2013.

The satellite estimates of radiation used in this work was gotten from EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) Trentmann-Jörg (2015). This website has hourly, daily and monthly averages of DNI, SIS and CAL from $1983/01/01$ to $2013/12/31$. For the sake of this work, hourly SIS was selected. The data ranged from $1983/01/01$ to $2013/12/31$ and covering

an area bounded by longitude $35.7°$ E to $35.8°$ E and lattitude $6.2°$ S to $6.1°$ S.This was selected because it is the area that bounds the area from where ground data were derived. The data is provided as a zipped folder.

2.1.2 Dodoma Data on Radiation.

Solar radiation data at the ground stations can be measured directly using instruments such as pyranometers and pyrheliometers or indirectly from the number of sunshine hours or maximum and minimum temperature.

Pyrheliometers are used to measure direct solar radiation. The working of pyrheliometer is such that sunlight enters it through a window and is directed onto a thermopile which converts heat to an electrical signal that can be recorded. A calibration factor is then applied to convert the millivolt signal to an equivalent radiant energy flux, measured in Wm^{-2}. It is sensitive to wavelengths in the band from $0.28\mu m$ to $3\mu m$. A pyranometer is used to measure global solar radiation. It gives the reading in Wm^{-2} and is also sensitive to wavelengths in the band from $0.28\mu m$ to $3\mu m$. Diffuse radiation is measured using a pyranometer with a shading ring.

In Tanzania, the Tanzania Meteorological Agency (TMA) organizes and manages climatic condition records of the country. These records are derived from the weather stations that make hourly observations on variables such as solar radiation, hours of bright sunshine, maximum and minimum temperature and humidity.

The ground data used in this work was obtained from a synaptic station in Dodoma, Tanzania. These data are readily available in R-Instat through being made freely available by TMA. R - Instat is an open source front end to the statistics language R.

The data had other climatic records, rainfall, maximum and minimum temperatures. Figures 2.3 and 2.4 show the first and the last few observations of sunshine duration data.

	Year	Month	Day	Rain	Tmax	Tmin	Sunh
1	1974	Jan	1	7.7	31.2	19.5	10.9
2	1974	Jan	2	0.0	31.6	18.4	11.3
3	1974	Jan	3	0.0	29.9	17.8	10.2
4	1974	Jan	4	0.0	31.0	17.9	11.5
5	1974	Jan	5	0.0	31.7	18.6	11.3
6	1974	Jan	6	0.0	31.9	18.7	11.6

Figure 2.3: The first 6 records of Dodoma sunshine duration hours

	Year	Month	Day	Rain	Tmax	Tmin	Sunh
14240	2012	Dec	26	0.0	30.6	18.1	11.2
14241	2012	Dec	27	0.0	30.7	19.6	11.5
14242	2012	Dec	28	0.0	27.0	20.0	10.0
14243	2012	Dec	29	0.0	29.9	21.2	4.1
14244	2012	Dec	30	8.5	30.5	20.5	8.2
14245	2012	Dec	31	0.0	31.0	19.4	6.6

Figure 2.4: The last 6 records of Dodoma sunshine duration hours

The column of interest for this work is the sunshine duration. This data was from 1983 to make its

comparison with the satellite data easy. The subset data had 10958 entries from Jan 1983 to Dec 2012. Out of these, a total of 3429 observations were missing values, with completely missing values for the years 1994, 1995, 1996 , 1997 and 1998. Leaving out these years gave a total of 9131 observations with 1599 missing values. This makes approximately 18% of the values missing. Figure 2.5 shows the missing sections of the data.

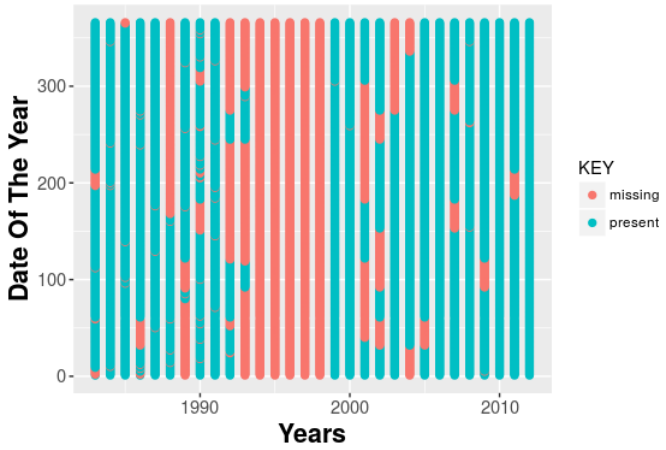

Figure 2.5: Missing values of sunshine hours.

2.2 Data Preparation

2.2.1 Satellite Data.

CMSAF provides, other than the data, R scripts that can be used to unzip the data folder. The script Prep.Data.R was used to uncompress the data. This gave several data files in Network Common Data Format ($netCDF$) with a single file containing all the files merged. To get the data in csv format, the open NetCDF option from climate menu in R-Instat was used. Figure 2.6 shows a snapshot of the climate menu of R-instat $0.2.3$.

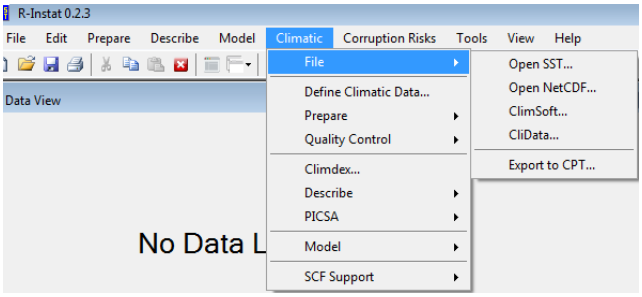

Figure 2.6: Climate menu of R-instat Version $0.2.3$.

Clicking on open NetCDF opens a dialogue box that allows one to select the file to be opened. The opened file can then be exported as a CSV file.

Because the ground data had missing values from 1994 to 1998, the satellite data was split into two, the lower data from 1983 to 1992 and the upper data ranging from 1999 to 2012. Figure 2.7 and Figure 2.8 show the first 6 elements of the lower and upper datasets, respectively.

	period	lat	lon	station	SIS
1	0	-6.149998	35.70	S6.1499981880188_E35.7000045776367	NA
2	0	-6.149998	35.75	S6.1499981880188_E35.75	NA
3	1	-6.149998	35.70	S6.1499981880188_E35.7000045776367	0
4	1	-6.149998	35.75	S6.1499981880188_E35.75	0
5	2	-6.149998	35.70	S6.1499981880188_E35.7000045776367	0
6	2	-6.149998	35.75	S6.1499981880188_E35.75	0

Figure 2.7: The first 6 elements of the lower data

	period	lat	lon	station	SIS
1	140256	-6.149998	35.70	S6.1499981880188_E35.7000045776367	NA
2	140256	-6.149998	35.75	S6.1499981880188_E35.75	NA
3	140257	-6.149998	35.70	S6.1499981880188_E35.7000045776367	0
4	140257	-6.149998	35.75	S6.1499981880188_E35.75	0
5	140258	-6.149998	35.70	S6.1499981880188_E35.7000045776367	0
6	140258	-6.149998	35.75	S6.1499981880188_E35.75	0

Figure 2.8: The first 6 elements of the upper data

The period columns in both datasets were converted to respective dates indicating the year, month, date and hour. The data was then subset by removing the night hours ($7P.M - 5A.M$) because no radiation is received. Figures 2.9 and 2.10 show the lower and upper data with the dates after subsetting.

	lat1	lon1	station1	period	split	date	year	month	day	hour	SIS
15	-6.149998	35.70	1	4	1	1983-01-01	1983	1	1	7	4
16	-6.149998	35.75	2	4	2	1983-01-01	1983	1	1	7	5
17	-6.149998	35.70	1	5	1	1983-01-01	1983	1	1	8	NA
18	-6.149998	35.75	2	5	2	1983-01-01	1983	1	1	8	NA
19	-6.149998	35.70	1	6	1	1983-01-01	1983	1	1	9	48
20	-6.149998	35.75	2	6	2	1983-01-01	1983	1	1	9	44

Figure 2.9: First 6 elements of the lower dataset after subsetting.

	lat1	lon1	station1	period	split	date	year	month	day	hour	SIS
15	-6.149998	35.70	1	140261	1	1999-01-01	1999	1	1	7	167
16	-6.149998	35.75	2	140261	2	1999-01-01	1999	1	1	7	237
17	-6.149998	35.70	1	140262	1	1999-01-01	1999	1	1	8	393
18	-6.149998	35.75	2	140262	2	1999-01-01	1999	1	1	8	431
19	-6.149998	35.70	1	140263	1	1999-01-01	1999	1	1	9	708
20	-6.149998	35.75	2	140263	2	1999-01-01	1999	1	1	9	699

Figure 2.10: First 6 elements of the upper dataset after subsetting

.

The values were then aggregated to daily data for easy comparison with the daily sunshine hours. The lower dataset had a total of 3652 observations with 147 missing, while the upper, out of 5479 obser-

vations, 80 were missing. This give a total of 227 missing values out of 9131 observations, translating to 2.5% of the satellite observations missing. Figure 2.11 shows the missing values in SIS derived from the satellite.

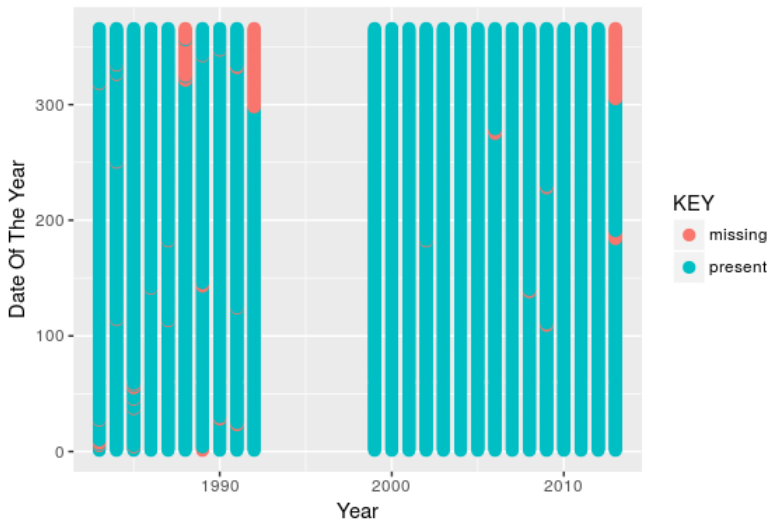

Figure 2.11: Missing values of SIS

2.2.2 Dodoma Data on Radiation.

Other than sub-setting, there were no further preparations that were done on it, since it was in a format that can be analysed.

3. METHODOLOGY

Here we elucidate the methods used to address the main objective of this work. The methods used to compare satellite data and ground data are described. The chapter also describes the probability distributions functions that were used to describe the distribution of solar radiation.

3.1 Comparison of Radiation Data sets

The comparison of the two radiation datasets (Satellite and Ground Data) is to be done in two ways;

(i) Drawing of comparative plots of surface solar irradiance from the satellite and the sunshine hours from Dodoma.

(ii) Calculation of statistical parameters, mean percentage error (MPE), correlation (r), mean absolute deviation (MAD) and mean bias error (MBE) were calculated from the mean monthly SIS data as follows (Journée and Bertrand, 2010);

Let D_{gi} be measurements from the ground and D_{si} be measurements from the satellite;

- MPE

$$MPE = \frac{1}{n}\left[\sum_{i=1}^{n}\left(\frac{D_{gi} - D_{si}}{D_{gi}} \times 100\right)\right], \tag{3.1.1}$$

- correlation (r)

$$r = \frac{n\sum_{i=1}^{n} D_{gi}D_{si} - \left(\sum_{i=1}^{n} D_{gi}\right)\left(\sum_{i=1}^{n} D_{si}\right)}{\sqrt{n\sum_{i=1}^{n} D_{gi}^2 - \left(\sum_{i=1}^{n} D_{gi}\right)^2}\sqrt{n\sum_{i=1}^{n} D_{si}^2 - \left(\sum_{i=1}^{n} D_{si}\right)^2}}, \tag{3.1.2}$$

- MBE

$$MBE = \frac{1}{n}\left[\sum_{i=1}^{n}(D_{gi} - D_{si})\right], \tag{3.1.3}$$

- MAD

$$MAD = \frac{1}{n}\left[\sum_{i=1}^{n}|(D_{gi} - D_{si})|\right], \tag{3.1.4}$$

The statistical parameters described are to be used to give an idea of whether satellite-derived measurements can be used to fill the gap in the ground measurements.

3.2 Distribution of solar irradiance

Because of their reliability when it comes to solving modern technology systems problems whose functionality depends on the reliability of its components, Log-normal, Gamma and Weibull distributions are tested to get the best fit for solar radiation in a given month.

(i) **Log-normal distribution.** If a random variable $Y = \ln(X)$ has a normal distribution with mean μ and standard deviation σ, the continuous random variable X is said to have a log normal distribution, and its density function is

$$f(X|\mu,\sigma) = \frac{1}{\sigma X\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[lnX-\mu]^2} \quad \text{for } x \geq 0. \tag{3.2.1}$$

The mean and the variance of the distribution are

$$\text{mean } = \exp(\mu + \frac{1}{2}\sigma^2) \quad \text{and} \quad \text{variance } = (\exp(2\mu+\sigma^2))(\exp(\sigma^2)-1).$$

(ii) **Gamma distribution.** If a random variable X has a gamma distribution, its density function is

$$f(x|a,b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-\frac{x}{b}}, \quad X > 0, \text{With} \tag{3.2.2}$$
$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\, dx,$$

where a and b are parameters of the gamma distribution. The mean and variance of gamma distribution are

$$\mu = ab, \qquad \text{and } \sigma^2 \ = ab^2.$$

(iii) **Weibull distribution.** This distribution has become more popular due to its ability to fit the data from many fields such as engineering, weather, and life data, to mention but a few. The general Weibull distribution has the density function

$$f(x|a,b,c) = \frac{a}{b}\left(\frac{x-a}{b}\right)^{c-1} \exp\left(-\left(\frac{x-a}{b}\right)^c\right) \quad x \geq a, \quad a \in \mathbb{R}, \quad \text{and} \quad b, c \in \mathbb{R}^+, \tag{3.2.3}$$

where a is the location parameter, b is the scale parameter and c the shape parameter. The Weibull distribution used in this work assumes that a failure can occur at point zero. Therefore, the location parameter is zero. This gives a distribution that has the density distribution

$$f(x|a,b) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}. \tag{3.2.4}$$

The mean and variance of the Weibull distribution are

$$\mu = b\Gamma\left(\frac{1}{a}+1\right), \quad \text{and} \quad \sigma^2 = b^2\left\{\Gamma\left(\frac{2}{a}+1\right) - \Gamma\left(\frac{1}{a}+1\right)^2\right\}.$$

The probability distribution function that best describes the distribution of the solar irradiance in a given month, in Dodoma, was found to be weibul distribution.

4. ANALYSIS AND RESULTS

The results of the analysis are presented. The data were processed and subsequently analysed in response to the main objective of this work. The findings presented in this chapter demonstrate the potential of using satellite-derived radiation data as a substitute of scarce radiation data measured from the ground. We start with the analysis of the ground data followed by satellite data and finally the comparison of the two datasets.

4.1 Fitted PDFs for solar radiation

When the PDFs of lognormal, weibull and gamma distributions were fitted, weibull was found to describe the data much better than the other two.

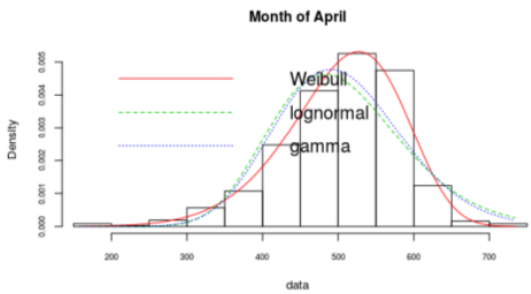

Figure 4.1: fitted distribution for April

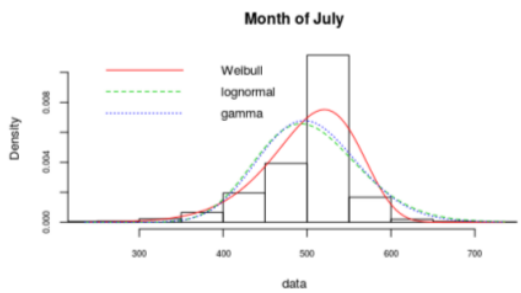

Figure 4.2: fitted distribution for July

Figure 4.1 and Figure 4.2 shows the distributions fitted for the months of April and July.

4.2 Sunshine Hours

Here, the analysis of the ground data is presented. Yearly plots are first presented. These include the box-plots and a time series plot showing the daily variation of sunshine hours over the years. The monthly plots and finally a histogram are plotted.

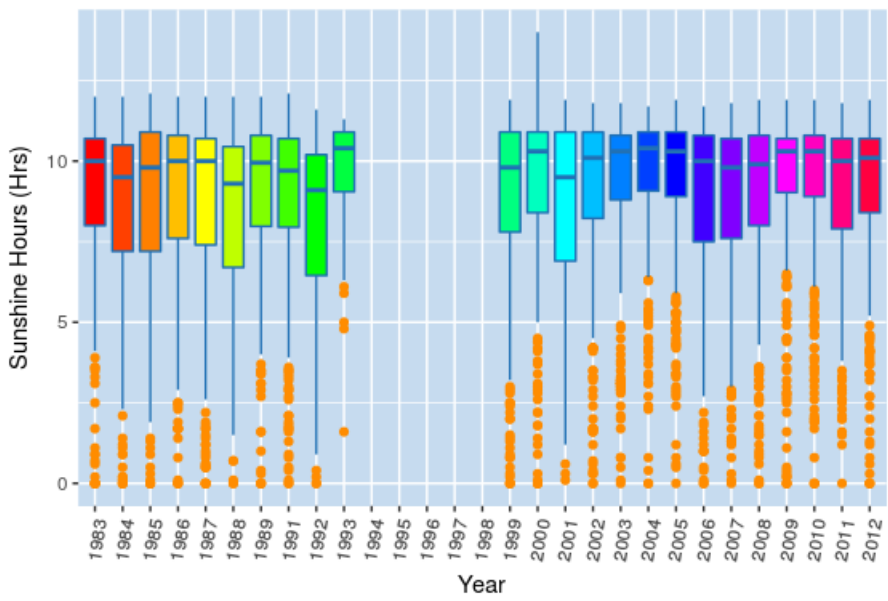

Figure 4.3: Sunshine hours by years

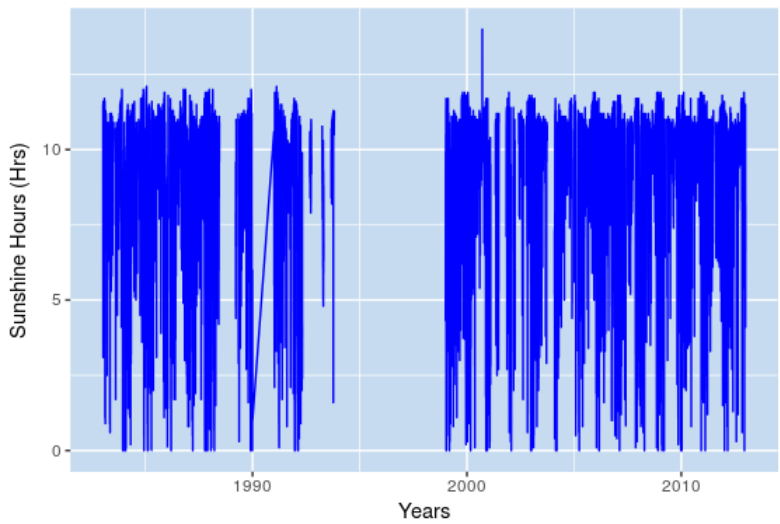

Figure 4.4: Daily variation of sunshine hours over the years

The following observations were made from the yearly plots for the ground data that were drawn;

The box-plots in Figure 4.3 show the daily sunshine hour split by years. It shows that over the years, the sunshine hours has been stationary, there is no pattern from one year to another. This is confirmed by the time series plot of the sunshine hours as shown in Figure 4.4. The Ljung-Box test confirmed this from its small p-values which suggest that the series is stationary. The box-plots also show unusual features in the data like gaps and outliers. The years 1994 to 1998 had no records and thus were omitted from the analysis.

Other than yearly plots, monthly box-plots of the sunshine hours was also plotted,

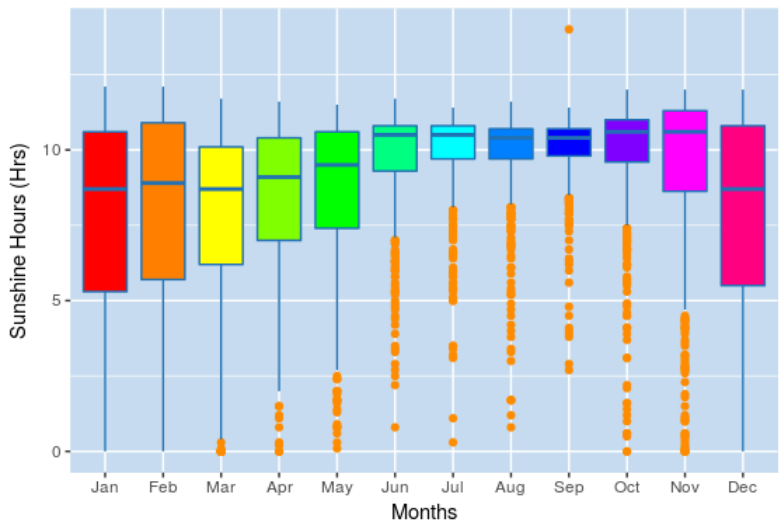

Figure 4.5: Sunshine hours by months

The box-plots of sunshine hours split by months (Figure 4.5) showed some possibility of seasonality in sunshine hours; the minimum values are recorded in the months of March and April, and the maximam in the months of June to October. The month of September shows less variability in sunshine hours received. This is depicted by the smaller length of the boxplot. The opposite is seen in the months of December, January and February which show high variability in the sunshine hours.

In an attempt to know the distribution of the sunshine hours (where the majority of values falls and how much variation is in the data), a histogram was also plotted.

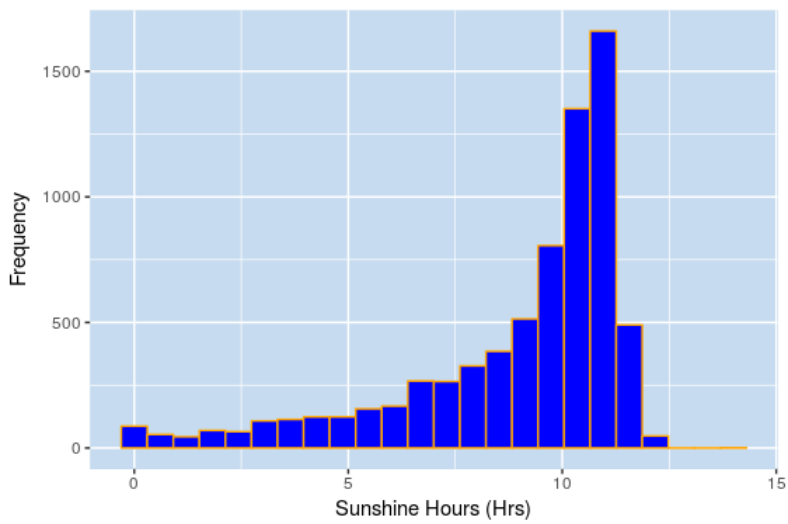

Figure 4.6: Histogram of sunshine hours

The distribution of sunshine hours in Figure 4.6 shows heavy skew to the left. This is an indication that majority of values fall between 5 and 11, an evidence that Dodoma receives alot of solar radiation.

This section presented the analysis that were done using the ground data. For easy comparison, the same analysis were done for the satellite derived estimates as shown in section 4.3.

4.3 Satellite estimates of solar radiation

This section describes the analysis done using the satellite-derived estimates. The yearly plots followed by the monthly box-plots and finally the histogram of the satellite-derived measurements are presented.

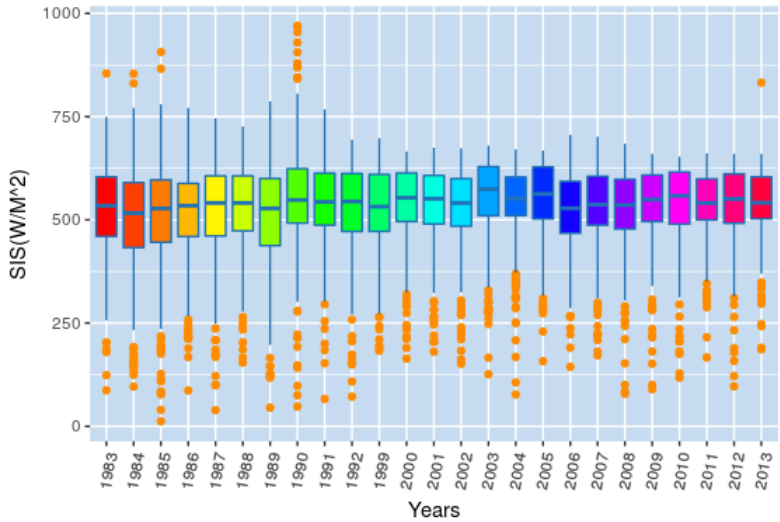

Figure 4.7: SIS radiation by Years

The box-plots for daily SIS split by years, like those for sunshine hours, show that the amount of solar radiation received over the years is stationary.

The box-plots also depicts unusual features, outliers and gaps, in the data. Most of these outliers are lower outliers, apart from a few especially in the year 1990, which shows very extreme values and thus are worth investigation. These outliers were investigated further in the comparison of the two datasets.

The years 1994 to 1998 were omitted to allow easy comparison with the sunshine hours.

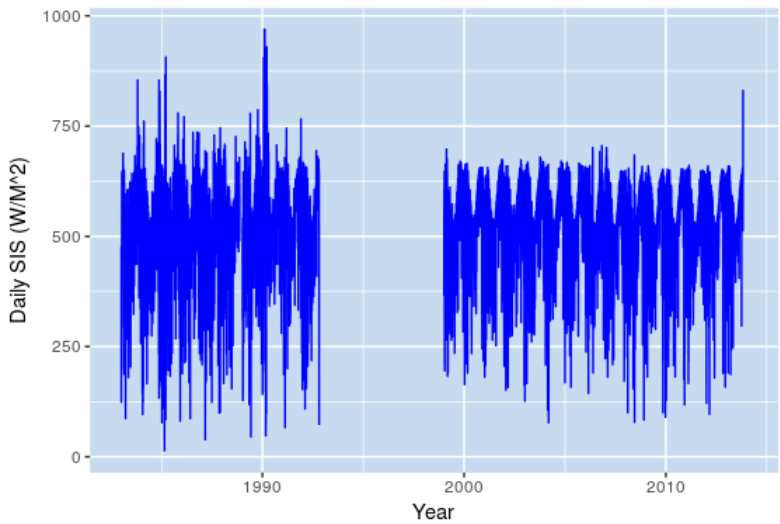

Figure 4.8: Time series plot of the SIS

These time series plots also depict SIS stationarity as the SIS variations does not show any pattern over the years. It also show the gap between 1994 to 1998, the years that were left out for easy comparison with the ground data.

These are indications that the area under study receives on average, the same sunshine durations each year.

The monthly variations were also investigated to get an insight of the distribution of the sunshine hours in each month.

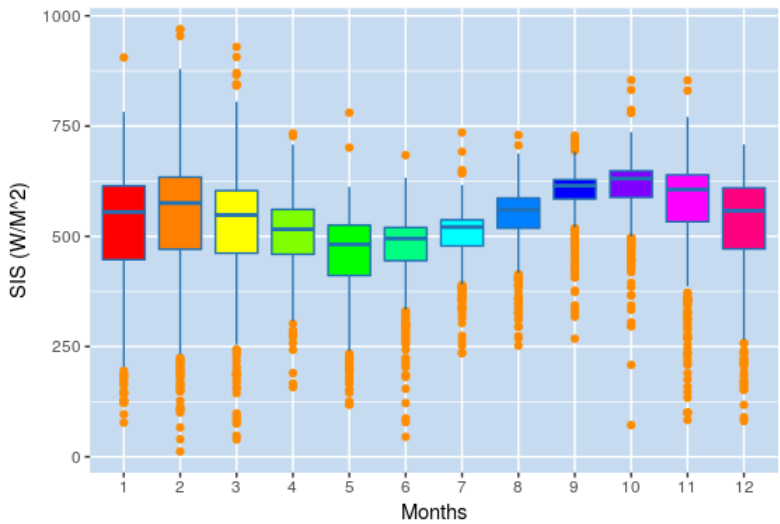

Figure 4.9: SIS radiation by months

The box-plots of SIS split by months shows that the months of January, February, March, August, September, October, November and December record larger measurements of solar radiation, with the month of September and October recording the highest amount. It also show that the month

of September receive almost the same amount of radiation throughout. This is shown by the small length of the box-plots, which is an indicator of less variability. The months of April to July receives lower surface solar irradiance, with the month of May recording the lowest readings. This may be because these are wet months.

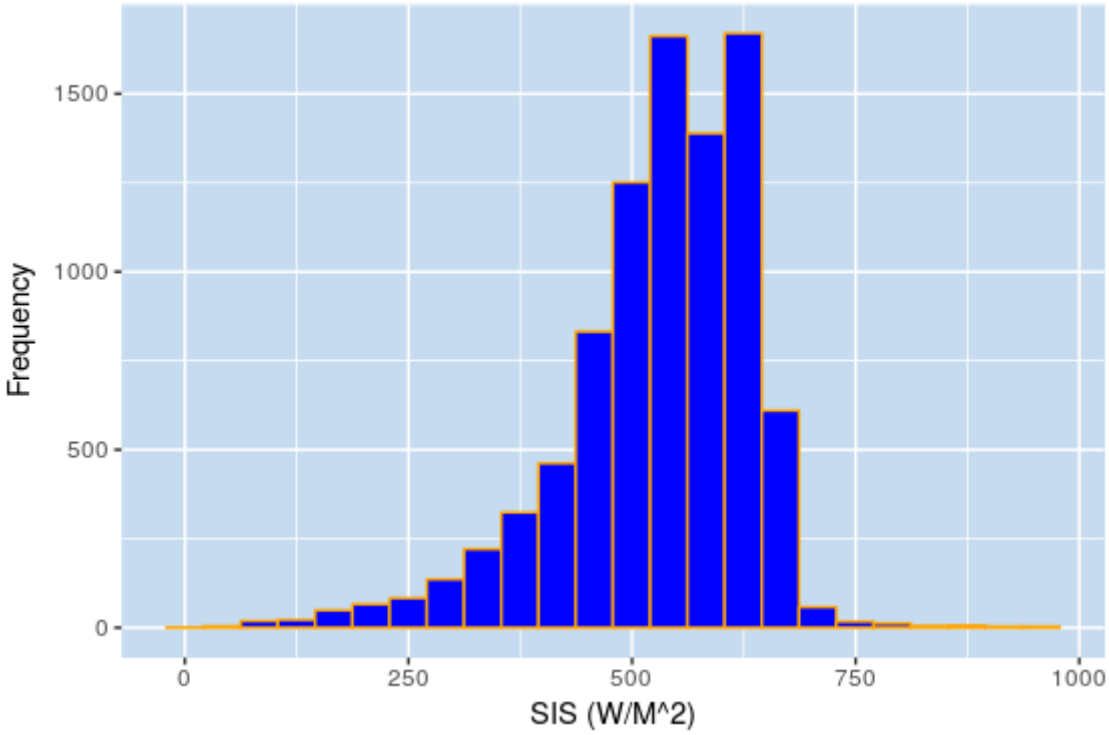

Figure 4.10: Histogram of SIS

The histogram of SIS (Figure 4.10), just like Figure 4.6, shows that solar surface radiation are heavily skewed to the left. This is an indication that Dodoma receives alot of solar radiation. with the majority of values falling between 400 and 750.

To find the extent of the difference between the datasets, other statistical parameters, Correlation r, MPE $(\%)$, MAD (WM^{-2}), and MBE (WM^{-2}) were calculated.

4.4 Comparison of Sunshine Hours and Solar Radiation Estimates

The exploratory analysis conducted shows that the two datasets, sunshine hours and the solar radiation estimates from satellite, are similar in distribution. Even though the two datasets are measured in different units, the similarity may be attributed to the fact that the underlying process generating the two datasets is the same.

The comparison of the two datasets was done in two levels, daily data and monthly data.

For easy comparison, the sunshine hours were converted to the same unit as the radiation data from satellite using *Angstrom's formulae* (NRMED),

$$R_s = \left(a_s + b_s \frac{n}{N}\right) R_a, \tag{4.4.1}$$

where, R_s is the solar radiation in $MJm^{-2}day^{-1}$, n is the actual duration of sunshine hours, N is the maximum possible duration of sunshine, a_s is the regression constant showing the fraction of extraterrestrial radiation reaching the earth on overcast days ($n = 0$), $a_s + b_s$ is the fraction of the extraterrestrial radiation reaching the surface on a clear day and R_a

$$R_a = \frac{24(60)}{\pi} G_{sc} d_r \left[\omega_s \sin(J)\sin(\delta) + \cos(J)\cos(\delta)\sin(\omega_s)\right], \tag{4.4.2}$$

where G_{sc} is the solar constant, J is latitude in radians $d_r = 1 + 0.033\cos\left(\frac{2\pi}{365}J\right)$,

$$\delta = 0.409\sin(\frac{2\pi}{365}J - 1.39) \text{ and } \omega_s = \arccos(-\tan(J)\tan(\delta))$$

For the area under study, $J = -0.0342\pi$, $G_{sc} = 1369\ Wm^{-2}$, $N = 12.1$, the maximum possible sunshine duration. Because there were no actual solar radiation data, no calibration was done to change a_s and b_s, therefore, the default values of $a_s = 0.25$ and $b_s = 0.50$ were used.

4.4.1 Evaluation of Monthly Data.

The comparison of monthly ground data calculated from sunshine hours and monthly satellite data was done using the monthly sums. It showed a bias of 727.2, mean percentage error of 4.04 and mean absolute deviation of 1194.40, as shown in Table 4.1. This small MPE indicates that the two datasets are similar.

The correlation between satellite data and ground data is strong with a value of 0.687. However, the months of March, April and September showed lower values of 0.402, 0.572 and 0.25, respectively, possibly because these are wet months. The months of May, June, July and August recorded the highest bias between the two datasets, probably because of the higher number of sunshine hours during this period.

Table 4.1: Bias, MAD, MPE and correlation for the comparison of Monthly ground data and satellite data

	JAN	FEB	MAR	APRIL	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	MEAN
r	0.858	0.8734	0.402	0.572	0.735	0.6211	0.867	0.71	0.25	0.7167	0.779	0.855	0.687
MPE	-2.85	-3.78	-2.81	4.37	15.94	18.07	14.24	7.62	0.34	-2.09	1.51	-2.10	4.04
MBE	-406.5	-569.2	-402.3	722.5	2577.1	3178.4	2542.5	1357.8	64.7	-351.1	274.0	-261.1	727.2
MAD	617.0	668.8	841.3	835.2	2577.1	3178.4	2542.5	1357.8	303.8	357.2	542.1	511.66	1194.40

Table 4.2: Bias, MAD, MPE and correlation for the comparison of daily ground data and satellite data

	JAN	FEB	MAR	APRIL	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	MEAN
r	0.872	0.818	0.708	0.682	0.708	0.648	0.651	0.633	0.454	0.384	0.757	0.877	0.683
MPE	-4.91	-4.09	-4.90	2.44	14.0	16.4	12.10	6.76	-0.21	-3.64	-0.11	-3.90	2.495
MBE	-15.28	-15.27	-13.47	23.12	80.72	97.23	73.99	42.60	1.19	-10.42	7.22	-8.24	21.95
MAD	44.38	41.63	56.70	50.98	89.12	98.31	77.64	49.67	30.74	37.57	41.23	42.40	55.03

Table 4.3: Monthly statistical parameters after removing outliers

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	MEAN
r	0.86	0.87	0.44	0.77	0.74	0.68	0.87	0.69	0.21	0.718	0.79	0.86	0.71
MPE	-4.30	-5.37	-2.76	4.49	14.93	17.30	13.29	6.57	-0.69	-2.92	-0.48	-3.13	3.08
MBE	-628.22	-771.76	-394.79	733.04	2405.14	3008.21	2338.72	1157.39	-120.92	-496.87	-71.08	-440.44	559.87
MAD	711.73	786.57	747.38	738.29	2405.14	3008.21	2338.72	1157.39	317.64	496.87	454.05	469.15	1135.93

Table 4.4: Daily statistical parameters after removing outliers

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	MEAN
r	0.86	0.87	0.44	0.77	0.74	0.68	0.87	0.69	0.21	0.718	0.79	0.86	0.71
MPE	-4.32	-4.03	-3.18	3.40	13.79	15.94	11.52	5.99	-1.06	-4.03	-0.98	-3.55	2.46
MBE	-16.63	-17.0	-9.91	24.47	78.04	93.01	68.62	36.54	-5.22	-15.24	0.07	-10.62	18.84
MAD	39.36	33.34	43.02	43.78	84.75	94.05	71.61	43.44	30.23	38.36	38.0	39.0	50.32

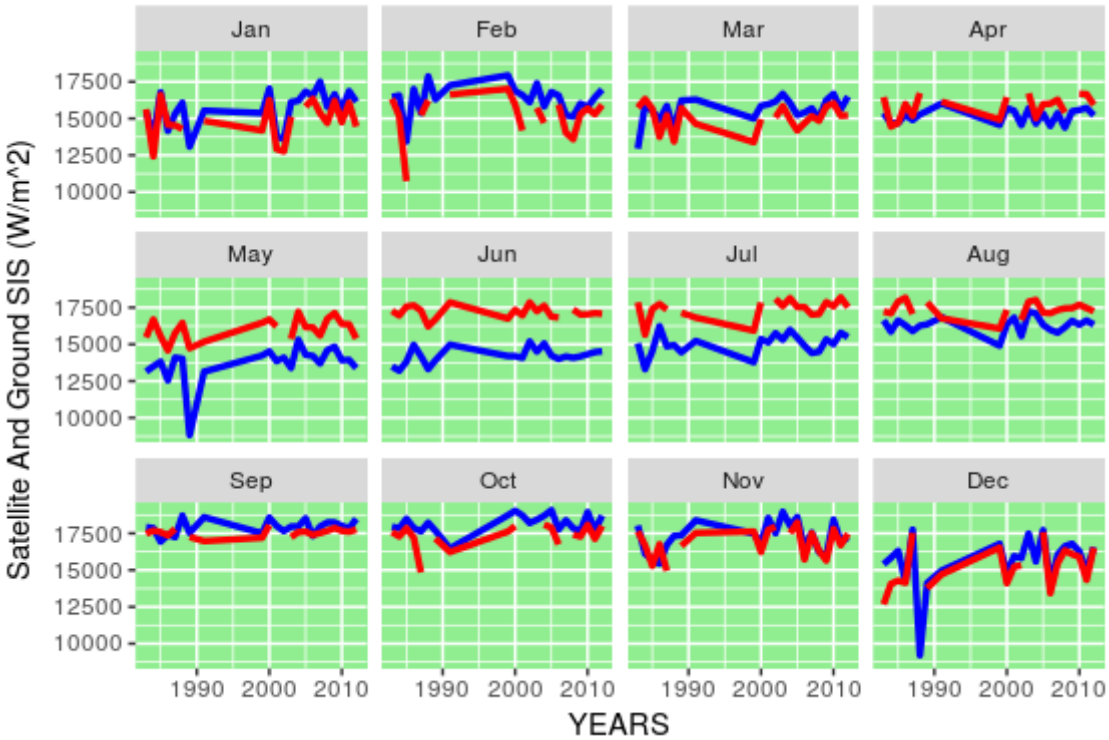

Figure 4.11: SIS variation by month. Combined line graphs of ground and satellite data. The red lines represents the data collected from the ground while the blue lines represent the satelite data.

Other than the monthly analysis, daily analysis were also done. This is shown in subsection 4.4.2.

4.4.2 Evaluation of Daily Data.

The comparison of ground data calculated from sunshine hours and satellite data showed a bias of 21.95, Mean percentage error of 2.495 and mean absolute deviation of 55.03, as shown in Table 4.2. This is an indication that there is no big difference between the two datasets and therefore, the satellite data are a good approximation of the ground truth.

Also, the correlation between satellite data and ground data is strong with a value of 0.683, but is lower than the correlation from the monthly data. However, the months of September and October showed lower values of 0.454 and 0.384, respectively, an indication that the satellite data and ground data correlate well in some months and not others, probably because of outliers or the season. This was further investigated.

As stated earlier, there were a number of outliers with very extreme values in 1990, the upper outliers of this year, which start at about 820, corresponds to very high sunshine hours, 15.1, that can not be reached at any particular time in the area under study. These outliers, together with those that are very low, below 150, were left out, and the statistical parameters calculated again.

The monthly and daily statistical parameters after removing outliers are shown in Tables 4.3 and 4.4

The new average correlation increased to 0.71, and the other statistical parameters reduced considerably. However, the months of January, February and December consistently showed high correlation, and the month of September consistently low correlations. Because of this, the months of February

and September were investigated further to get an insight of what can be the cause of these differences.

For the month of February, it was found that there is linear spread of sunshine hours from minimum value of 0.0 to the maximum value of 12 as depicted by the mean 7.9. Figure 4.12 shows the spread of the radiation data from the two sources for the month of February. The solar radiation estimates from the satellite were also found to have a good spread with a mean of 533.9. The trend of the two sources of radiation were found to be similar throughout the month Figure 4.14. The high correlation can therefore be attributed to the linear trend of data in this month.

In the month of September, out of 688 observations, with the maximum record as 11.4, only 267 are less than 10. This is a clear indication of clustering of observations. High average record of 10.07 justify the clustering observed. The same pattern was seen in solar radiation estimates from satellite, where only 218 observations out of 688 had values less than 600. Figure 4.13 shows the spread of radiation data from the two sources for the month of September. The low correlation can be explained by the non-uniform distribution in the recorded data. In spite of the low correlation and the uneven spread of records, some general trends were observed. The records from the two data sources showed some general trend over the years (Figure 4.11) and the month Figure 4.15.

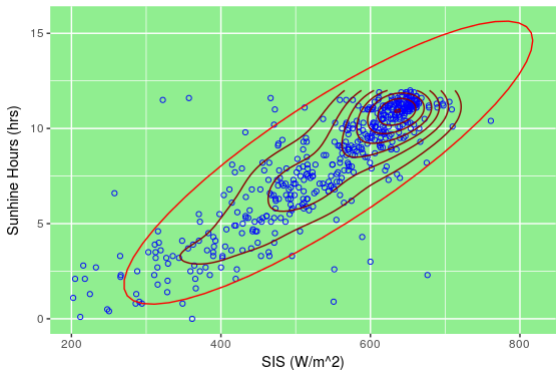

Figure 4.12: Scatter plot in February

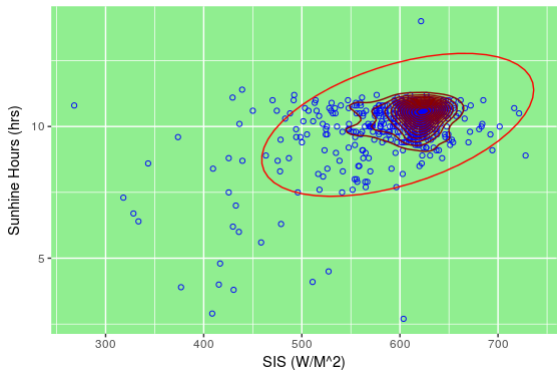

Figure 4.13: Scatter plot in September

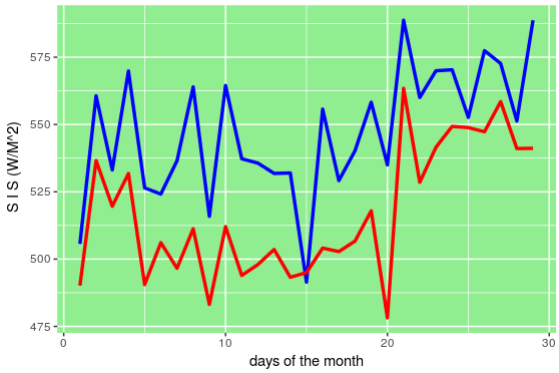

Figure 4.14: Variation of SIS over February

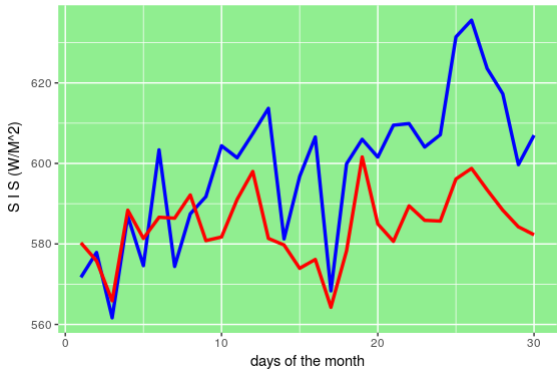

Figure 4.15: Variation of SIS over September

In this chapter, we compared the radiation data from satellite and data from the ground. Some months showed very high correlation, while others recorded low correlations. The high correlation were found

to be due to the good spread and similar general trends. Low correlations were found to be mainly due to clustered observations. However, the records in the months with low correlation showed similar general trend over the years as well as daily observations within the months. Our general feeling therefore is that the the two data source shows a good similarity, and satellite data can be used for places where the data is not available to provide continuous data.

5. CONCLUSION

The satellite data from EUMETSAT is the first freely available radiation data record for hourly, daily and monthly surface solar irradiance covering a region bounded between latitude $\pm 65^\circ$ and longitude $\pm 65^\circ$ for the time period 1983 to 2013. It has high quality, seen from fewer missing values and high spatial resolution $0.05^\circ \times 0.05^\circ$. Hence, it is qualified for many climate applications such as trend analysis and variation. This work has looked at the possibility of using satellite data to fill the gap of weather data shortage in Africa to help in making well informed climatic decisions.

In this work, we mainly considered solar radiation as a climatic element in Dodoma, Tanzania. The satellite data that was freely obtained from CMSAF was compared with the actual measurements from Dodoma. The comparison showed that the two data sets are similar in distribution. The correlation of the two data sets was also found to be 0.71 on average for both monthly and daily records, which is an indication that the two data sets are similar, and satellite data can be used to fill in the gaps in ground measured records as well as be used in places where the records are not available.

From the work, a Weibull probability distribution best describes the distribution of solar radiation of all the months for the area under study. This can be used to estimate the energy output a given photovoltaic cell can produce and therefore it is possible to select a solar panel that is suitable for a particular region.

ACKNOWLEDGEMENT

I am grateful to GOD for giving me good health, wisdom, knowledge and understanding that has made me complete this essay.

I am also grateful to my colleagues Prof. Roger Stern and Mr. Samson Wanyonyi for guidance and support throughout this work. The two were very instrumental in providing useful information that was used in this project and also in reviewing the report and providing valuable comments that enriched the script.

I am beholden to AIMS Tanzania centre staff and tutors.

Special thanks to my family for constant encouragement and prayers

References

ACPC. United nations economic commission for african climate policy centre. Available from http://www1.uneca.org/Portals/acpc/documents/working_papers/WP3-Climate%20data%20network%20and%20rescuing%20draft%20final.pdf, 2011.

L. Bechini, G. Ducco, M. Donatelli, and A. Stein. Modelling, interpolation and stochastic simulation in space and time of global solar radiation. *Agriculture, Ecosystems & Environment*, 81(1):29–42, 2000.

J. S. Bojanowski. *Quantifying Solar Radiation at the Earth Surface with Meteorological and Satellite Data*. University of Twente, ITC Faculty of Geo-Information Science and Earth Observation, 2014.

J. S. Bojanowski, A. Vrieling, and A. K. Skidmore. Calibration of solar radiation models for europe using meteosat second generation and weather station data. *Agricultural and Forest Meteorology*, 176:1–9, 2013.

A. Brisson, P. Le Borgne, and A. Marsouin. Development of algorithms for surface solar irradiance retrieval at o&si saf low and mid latitudes. *Eumetsat Ocean and Sea Ice SAF internal project team report*, 1999.

E. Cuce and P. M. Cuce. A comprehensive review on solar cookers. *Applied Energy*, 102:1399–1421, 2013.

F. Evrendilek and C. Ertekin. Assessing solar radiation models using multiple variables over turkey. *Climate Dynamics*, 31(2-3):131–149, 2008.

B. Geerts and E. Linacre. *Climates and Weather Explained*. Routledge, 2002.

A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, and H. G. Beyer. Solar energy assessment using remote sensing technologies. *Remote Sensing of Environment*, 86(3):423–432, 2003.

L. Hunt, L. Kuchar, and C. Swanton. Estimation of solar radiation for use in crop modelling. *Agricultural and Forest Meteorology*, 91(3):293–300, 1998.

P. Ineichen. Satellite based short wave irradiance validation over africa. 2010.

C. Jacovides, F. Tymvios, V. Assimakopoulos, and N. Kaltsounides. Comparative study of various correlations in estimating hourly diffuse fraction of global solar radiation. *Renewable Energy*, 31 (15):2492–2504, 2006.

M. Jarraud. Guide to meteorological instruments and methods of observation (wmo-no. 8). *World Meteorological Organisation: Geneva, Switzerland*, 2008.

M. Journée and C. Bertrand. Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measurements. *Remote Sensing of Environment*, 114(11): 2692–2704, 2010.

S. A. Kalogirou. *Solar Energy Engineering: Processes and Systems*. Academic Press, 2013.

H. S. Mavi and G. J. Tupper. *Agrometeorology: Principles and Applications of Climate Studies in Agriculture*. CRC Press, 2004.

T. Mavromatis. Estimation of solar radiation and its application to crop simulation models in greece. *Climate Research*, 36(3):219–230, 2008.

M. Medany, I. Niang-Diop, T. Nyong, and R. Tabo. Background paper on impacts, vulnerability and adaptation to climate change in africa. In *UNFCCC Convention, Ghana*, pages 21–23, 2006.

V. Mking'imle. *Assessing the Potential of Solar Energy Utilization in Central Tanzania*. PhD thesis, University Of Nairobi, 2013.

R. Mueller, J. Trentmann, C. Träger-Chatterjee, R. Posselt, and R. Stöckli. The role of the effective cloud albedo for climate monitoring and analysis. *Remote Sensing*, 3(11):2305–2320, 2011.

R. Mueller, T. Behrendt, A. Hammer, and A. Kemper. A new algorithm for the satellite-based retrieval of solar surface irradiance in spectral bands. *Remote Sensing*, 4(3):622–647, 2012.

R. Müller, U. Pfeifroth, C. Träger-Chatterjee, R. Cremer, J. Trentmann, and R. Hollmann. Surface solar radiation data set-heliosat (sarah)-, 2015a.

R. Müller, U. Pfeifroth, C. Träger-Chatterjee, J. Trentmann, and R. Cremer. Digging the meteosat treasure—3 decades of solar surface radiation. *Remote Sensing*, 7(6):8067–8101, 2015b.

R. R. Nemani, C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, 300(5625):1560–1563, 2003.

NRMED. Meteorological data. Meteorological Data, http://www.fao.org/docrep/x0490e/x0490e07.htm#radiation, Accessed May 2017.

R. Posselt, R. Müller, J. Trentmann, and R. Stöckli. Meteosat (mviri) solar surface irradiance and effective cloud albedo climate data sets, doi: 10.5676. Technical report, EUM_SAF_CM/RAD_MVIRI, 2011.

C. W. Richardson. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190, 1981.

Z. M. Salameh, B. S. Borowy, and A. R. Amin. Photovoltaic module-site matching based on the capacity factors. *IEEE Transactions on Energy Conversion*, 10(2):326–332, 1995.

A. Skartveit, J. A. Olseth, and M. E. Tuft. An hourly diffuse fraction model with correction for variability and surface albedo. *Solar Energy*, 63(3):173–183, 1998.

R. Sozzi, A. Salcido, R. S. Flores, and T. Georgiadis. Daytime net radiation parameterisation for mexico city suburban areas. *Atmospheric Research*, 50(1):53–68, 1999.

Spectral. Solar energy. Solar Energy, http://www.solarsystemcentral.com/sun_page.html, Accessed May 2017.

TAHMO. Weather station in homabay. http://tahmo.org/weather-stations-installed-in-kenya/, Accessed May 2017.

Tanzania Shapefile. Tanzania. http://www.maplibrary.org/library/stacks/Africa/Tanzania/index.htm, Accessed May 2017.

M. . R. P. U. T.-C. C. C. R. . H. R. Trentmann-Jörg. Surface solar radiation data set - heliosat (sarah). Available from http://dx.doi.org/10.5676/EUM_SAF_CM/SARAH/V001, 2015.

UNDP. A new vision for weather and climate services in africa. Available from https://www.thegef.org/sites/default/files/publications/WeatherAndClimateServicesAfrica.pdf, 2016.

F. Whisler, B. Acock, D. Baker, R. Fye, H. Hodges, J. Lambert, H. Lemmon, J. McKinion, and V. Reddy. Crop simulation models in agronomic systems. *Advances in Agronomy*, 40:141–208, 1986.

M. Wild, D. Folini, C. Schär, N. Loeb, E. G. Dutton, and G. König-Langlo. The global energy balance from a surface perspective. *Climate Dynamics*, 40(11-12):3107–3134, 2013.