Accuracy of Giovanni and Marksim Software Packages for Generating Daily Rainfall Data in Selected Bimodal Climatic Areas in Tanzania

Kahimba, F.C.[1], Tumbo, S.D. [1], Mpeta, E. [2], Yonah, I.B. [2], Timiza, W.,[2] and Mbungu, W. [1]

[1]Department of Agricultural Engineering and Land Planning, Sokoine University of Agriculture
P. O. Box 3003, Morogoro, Tanzania
[2]Tanzania Meteorological Agency,
P. O. Box 3056, Dar Es Salaam, Tanzania

Corresponding Author: *kahimbafcs@yahoo.com*

Abstract

Agricultural adaptation to climate change requires accurate, unbiased, and reliable climate data. Availability of observed climatic data is limited because of inadequate weather stations. Rainfall simulation models are important tools for generating rainfall data in areas with limited or no observed data. Various weather generators have been developed that can produce time series of climate data. Verification of the applicability of the generated data is essential in order to determine their accuracy and reliability for use in areas different from those that were used during models development. Marksim and Giovanni weather generators were compared against 10 years of observed data (1998-2007) for their performance in simulating rainfall in four stations within the northern bimodal areas of Tanzania. The observed and generated data were analyzed using climatic dialog of the INSTAT program. Results indicated that during the long rain season (masika) Giovanni predicted well the rainfall amounts, rainy days, and maximum dry spells compared to Marksim model. The Marksim model estimated seasonal lengths much better than the Giovanni model during masika. During short rain season (vuli), Giovanni was much better than Marksim. All the two software packages had better predictions during masika compared to vuli. The Giovanni model estimated probabilities of occurrence of rainfall much better (RMSE = 0.23, MAE = 0.18, and d =0.75) than Marksim (RMSE = 0.28, MAE = 0.23, and d = 0.63). The Marksim model over-predicted the probabilities of occurrence of dry spells greater than seven days (MBE = 0.17) compared to the Giovanni model (MBE = 0.01). In general the Giovanni model was more accurate than the Marksim model in most of the observed weather variables. The web based Giovanni model is better suited to the northern bimodal areas of Tanzania. The Marksim model produced more accurate climatic data when the long-term average climate data are used as input variables. This study recommends the use of rainfall data generated using Giovanni software over Marksim, for areas receiving bimodal rainfall regimes similar to the northern bimodal areas of Tanzania.

Key words: Bimodal rainfall; long-term average climate data; dry spells; Giovanni; INSTAT; Marksim

Introduction

Background

Crop simulation models are necessary tools for quantification of the effects of the current and future climate change in crop production. Accurate quantification of the effects will help to manage the potential risks of climate change in agricultural production and the environment (Howden *et al.,* 2007; Tingem *et al.,* 2007). However, for crop simulation models to work they require information on crop, soil, and climate (Howden *et al.,* 2007). Among important and challenging parameters in climate information is rainfall.

Rainfall is highly variable in space and time due to the complexity and rapid variation of the vertical structure of the precipitation cloud (Bruhn *et al.,* 1980; Liu *et al.,* 1999). Therefore, getting accurate information for a specific location is a challenge. The situation is even more magnified for areas with no meteorological stations, or where the stations are sparsely located (Lanza, 2000).

The common method that has been used in getting weather information in areas with no observed data is through interpolation of observed data from nearby stations (Jones and Thornton, 2000). This method however is prone to errors since topographic uniformity cannot be guaranteed at longer distances. In addition, observed data in recording stations may be available only for a specific period of time; may contain gaps; or may not be available on a long-term basis (Semenov *et al.,* 1998; Jones and Thornton, 2000).

Development of weather generators such as WGEN and LARS-WG (Semenov *et al.,* 1998), Marksim (Jones and Thornton, 2000), Giovanni (Acker *et al.,* 2005), WRF and MM5 (Kusaka *et al.,* 2005), and ClimGen (Tingem *et al.,* 2007), have helped to alleviate the interpolation problem and assist in risk assessment in the fields of hydrology, agriculture, and environment. Weather generators can produce time series of several long-term weather data. Other advantages of weather generators are possibilities of interpolating observed data from a known station to get data on an ungauged station, and filling-in missing data on an existing station. Weather generators are now being used in many researches around the world in the fields of forestry, ecology, environment, hydrology, meteorology, and agriculture (Tingem *et al.,* 2007).

Despite their acceptability and wide use, the main problem with weather generators such as Giovanni and Marksim is low accuracy and generality of the generated data as compared to the measured values. Data sets from weather generators may also not be specific to a location; hence if proper validation based on specific climatic condition is not performed, the generated data may not be representative of local conditions (Jones and Thornton, 2003; Kusaka *et al.,* 2005; Tingem *et al.,* 2007). These tools need to be evaluated in locations where they are to be used, especially when the areas are meteorologically different from those that were used during model calibrations and validations (e.g. Bruhn *et al.,* 1980). Hence, the main objective of this paper was to assess the performance of Marksim and Giovanni software packages in simulating rainfall in bimodal climatic areas. Specific objectives were to:

1. Evaluate the performance of Giovanni and Marksim software packages in simulating rainfalls within the northern bimodal areas of Tanzania.
2. Compare the accuracy of the model simulations during the long *(Masika)* and short *(Vuli)* rain seasons.
3. Propose a better weather generator for areas receiving bimodal rainfall similar to the northern bimodal areas of Tanzania.

Validation of these software packages in the northern bimodal areas of Tanzania will help researchers, decision makers, and farmers in deciding on the use of a weather generator that is better suited to bimodal climatic conditions.

The Marksim software

Marksim model was developed by Jones and Thornton (2000) for the purpose of generating daily weather data that can be used for running DSSAT (Decision Support System for Agro technology Transfer) crop models and for risk analyses in Latin America and Africa. The model uses a third order Markov process to stochastically generate daily weather data, taking into account events that have occurred in the previous three days (Jones and Thornton, 2000). Bruhn *et al.* (1980) and Rahman (2000) used a first order Markov chain in developing a rainfall simulation model.

The Marksim model generates daily rainfall, maximum and minimum air temperatures, and global solar radiation for a series of several years. It has a full window interface. The model uses three options of input variables as follows (Jones and Thornton, 2000):

(i) Defining a point by latitude and longitude only (pixel size of 18 by 18 km; may not be very accurate especially in areas with varying topography).

(ii) Defining a point by latitude, longitude, and elevation (produces better estimates than the first option).

(iii) Using long-term average monthly climate data of a given station as input variables. This option gives a better estimate of the daily weather data for a specific location.

Hence the shortcoming of using this weather generator is that if long-term average climate data are not available for a station then the generated data will not be very accurate (Jones and Thornton, 2000). In addition, the software does not generate rainfall data for specific years, rather data which is statistically correct within the 10 or 15 years. Hence it is difficult to perform yearly pair-wise comparisons of annual weather data with observed data or data generated by other weather generators. Weather data generated using Marksim software has been used in several researches (e.g. Jones and Thornton, 2003). More details on the physics of the model and model descriptions are available in Jones and Thornton (2000).

The Giovanni software

Giovanni is an acronym for the **G**ES-DISC (Goddard Earth Science Data and Information Services Centre) **I**nteractive **O**nline **V**isualization **A**nd a**N**alyis **I**nfrastructure. It is an online tool for exploring, comparing, and analyzing remote sensing data using a web browser. The Giovanni software has a global partial-temporal coverage and can provide data in several formats such as HDF, ASCII, net CDF, and kml/kmz. The software accelerates conventional discovery, acquisition, management, and analysis of remote sensing data by allowing direct web visualization and analysis. The software performs the basic analytical functions using Grid Analysis and Display system (GrADS) (Acker *et al.*, 2005; Acker and Leptoukh, 2007; Leptoukh *et al.,* 2007). Giovanni can be used to provide near-real-time 3-hourly, Multi-Satellite precipitation analysis, and rainfall ground observation data. The model can also generate rapidly time series data including improved time resolution with 8-day data sets (Acer *et al.,* 2005; Leptoukh *et al.,* 2007).

The Giovanni system can be used to generate climatic datasets such as precipitation, atmospheric chemistry, and sea surface and air temperatures. The Giovanni software has been used in many researches. For example, Kaufman *et al.* (2005) used the software to investigate dust transport over the Atlantic Ocean. Acker *et al.* (2006) used the software for coastal zone remote sensing research in the Algerian coast. Some of the data products include 3-hourly rainfall, daily rainfall, 10-day rainfall, monthly rainfall, and monthly rainfall anomaly. Details of the software and additional data are available in the NASA's GES-DISC webpage (http://giovanni.gsfc.nasa.gov), and in Acker *et al.* (2005). The model has also been used in exploring meteorological variables by Ahmad *et al.,* (2007) and Rui *et al.,* (2007).

Materials and methods

Location and characteristics of the study area

The meteorological stations whose observed data were used in this study are monitored by the Tanzania Meteorological Agency (TMA). The study area lies between latitudes 3.4°S and 6.8°S, and longitudes 36.6°E and 39.0°E. In the bimodal rainfall regime, short rains or *Vuli* start from October and ends in December (OND), and long rains or *Masika* start from March to May (MAM). Bimodal rainfall regime areas used in this study were in the North-eastern highlands and North-eastern coast. Short rains are highly variable in space and time (Liu *et al.,* 1999; Jones and Thornton, 2003). Four stations namely, Arusha, Morogoro, Same and Tanga, were selected for this study, which are within the eastern and North-eastern bimodal areas of Tanzania (Table 1, Figure 1). These were stations with long-term and complete data sets needed by Marksim and Giovanni software packages for the selected 10-year period.

Data sources and collection

Rainfall data was used to compare the performance of Giovanni and Marksim weather generators. The observed data (over 10-year period) were obtained from the Tanzania Meteorological Agency (TMA) database. Measurements of daily and hourly rainfall at the stations were performed using both standard and automatic rain gauges.

Table 1: Geographical descriptions of the experimental stations

Region	Latitude (°S)	Longitude (°E)	Elevation (m amsl)	Annual rain (mm)	Max temp (°C)	Min temp (°C)	Year range (year)
Arusha	3.37	36.63	1372.0	809.9	25.8	14.2	1971-2000
Morogoro	6.83	37.65	512.0	848.6	30.2	18.9	1971-2000
Same	4.08	37.73	860.0	562.5	29.0	17.6	1971-2000
Tanga	5.05	39.04	49.0	1329.0	30.7	22.0	1971-2000

Figure 1: Map of Tanzania showing the four meteorological stations used in the study

Data generation and analysis

Rainfall data was generated using Marksim V. 1.0 (Jones and Thornton, 2000) and Giovanni V. 3.0.1 (NASA, 2007), which were then compared against the observed data. The input data for the Maksim software were latitude, longitude, and elevation of the station. The Marksim software generates daily rainfall data with random years (Jones and Thornton, 2000, 2002, 2003). The Giovanni software generates daily rainfall data of interest by specifying the time of interest (i.e. the start and end dates), and latitudes and longitudes of the area. Analysis of the 10-year averages for Maksim and Giovanni software packages were used during comparison with the observed data. Marksim is one of the most popular software compared to other generators in climate change studies, while Giovanni gives true/near-real time data that are comparable to actual measured data (Leptoukh *et al.,* 2007; Prados *et al.,* 2007; Mazandarani et al., 2013).

Annual and monthly rainfall totals, rainy days for rainfall greater than 2 mm, rainy days for rainfall 10 mm, the start and end of the season, seasonal lengths, dry spell runs, and the probability of occurrences of dry spells longer than 7 days and 10 days for each station were determined during OND and MAM crop growing seasons. Ten-year simulated (Giovanni and Marksim) and observed daily rainfall data were analyzed using climatic dialog of the INSTAT program version 3.36 (Stern *et al.,* 2006). The start of the season was defined as the first date from 1st October/ 1st March getting more than 20.0 mm of rainfall in 1, or 2 days for OND and MAM, respectively. A rainy day was defined as a day with more than 2.0 mm of rainfall.

The end of the season (cessation date) was determined based on simple water balance as it was derived by Dennett *et al.* (1983). The amount of water in the soil on day *i+1* is:

$$W_{i+1} = W_i + R_i - E \qquad (1)$$

Where R_i is daily rainfall and E is daily evapotranspiration, taken here as 5.0 mm per day throughout the season (Allen *et al.,* 2006), and W_i is the amount of water in the soil on day i. The maximum water storage capacity of the soil was taken to be 60 mm (Dennett *et al.,* 1983). The end of the season was defined as the first day that W_i becomes zero and remains at zero for more than five days (Dennett *et al.,* 1983).

Statistical analysis

Statistical analyses were performed to verify the authenticity of daily rainfall data generated using Marksim or Giovanni. Both relative and absolute scalar accuracy measures were used to check the accuracy, bias, and reliability of the generated data compared to the observed data from gauged stations (Willmott *et al.,* 1985; Wilks, 1995; Legates and McCabe Jr., 1999).

The absolute scalar accuracy measures used were the root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE) (Hargreaves and Samani, 1985; Willmott *et al.,* 1985; Wilks, 1995). The relative accuracy measures used were the coefficient of determination (R^2), and modeling index or index of agreement (d). The statistical significance was reported on the basis of level of significance (p-value or alpha) of 0.05, with sample means with no significant difference being indicated by common letters "a, b, c, d" (SAS Institute Inc., 2004).

The mean absolute error is the arithmetic average of absolute differences between the observed (O_i) and predicted (P_i) values (Wilks, 1995; Legates and McCabe Jr., 1999). It is expressed as:

$$MAE = \frac{1}{n}[\sum_{i=1}^{n} | P_i - O_i |] \quad (2)$$

Where n is the number of observations, P_i is the predicted value, and O_i is the observed value. For perfect prediction the MAE ranges between zero and large positive values (Willmott *et al.,* 1985). The RMSE is the square root of average squared differences between P_i and O_i (Wilks, 1995; Steel *et al.,* 1997). It is expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (3)$$

The square function in the RMSE makes the measure to be more sensitive to extreme errors than the MAE measure (Wilks, 1995). The MBE is the average of the differences between the P_i and O_i pairs. It indicates average interpolation 'bias'; that is, average over- or under-estimation by an interpolator (Willmott and Matsuura, 2006). A value close to zero indicates equal distribution between negative and positive errors. It is expresses as (Wilks, 1995; Steel *et al.,* 1997):

$$MBE = \frac{1}{n}[\sum_{i=1}^{n}(P_i - O_i)] \quad (4)$$

Since the MBE averages the sum of errors, it does not give a better indication of the magnitude of individual prediction errors (Wilks, 1995).
The index of agreement (d) was calculated as follows (Willmott *et al.,* 1985):

$$d = 1.0 - \left[\frac{\sum_{i=1}^{n}(p_i - O_i)^2}{\sum_{i=1}^{n}(| P_i - \bar{P} | + | O_i - \bar{O} |)^2} \right] \quad (5)$$

Where n is the total number of observations, P_i is the predicted value, O_i is the observed value, and $\bar{P}$ and $\bar{O}$ are the means of the predicted and observed values, respectively. The relative accuracy measure d accounts for the differences between means and variances of P_i and O_i. It ranges between 0.0 and 1.0. Values close to 1.0 indicates better agreement between the O_i and P_i (Hargreaves and Samani, 1985; Willmott *et al.,* 1985; Legates and McCabe Jr., 1999). The general linear model (GLM) procedure of the SAS statistical package (SAS Institute Inc., 2004) was also used to compare the variable means for the two software packages.

Results and discussions

Comparison of the rainfall amounts

Table 2 presents a summary of statistical comparison of rainfall amounts predicted by the Giovanni and Marksim software packages. The monthly rainfall amounts were compared for two rain seasons, *masika* (March, April, and May - MAM), and *vuli* (October, November and December - OND). Comparison was also made on annual basis.

During the first rain season (masika or MAM), there was no significant difference in generated/estimated rainfall amounts ($\alpha = 0.05$) between the Marksim and Giovanni software packages on three out of four stations (Arusha, Morogoro, and Same). At Tanga, Marksim rainfall amounts were significantly lower ($\alpha = 0.05$) than the observed and Giovanni rainfall amounts.

During the OND season Marksim rainfall amounts were significantly lower ($\alpha = 0.05$) than the observed rainfall on two stations (Arusha and Tanga), while Giovanni amounts were significantly higher than the observed in one station (Morogoro). Considering the annual rainfall, both Marksim and Giovanni rainfall amounts were significantly lower than the observed rainfall at Arusha. There was also a significant difference between the observed and predicted values for Marksim at

Table 2. Comparison of rainfall amounts for the observed, Giovanni, and Marksim data.

		Rainfall amounts from experimental stations (mm)**			
Season*	**Data set**	**Arusha**	**Morogoro**	**Same**	**Tanga**
MAM	Observed	341.8±156 a	400.6±87 a	214.7±92 a	557.7±210 a
	Giovanni	266.4±93 b	373.3±87 a	205.6±96 a	469.1±220 a
	Marksim	234.4±145 b	465.7±181 a	275.7±76 a	198.6±111 b
OND	Observed	189.3±135 a	175.8±134 a	168.1±132 a	275.4±154 a
	Giovanni	147.2±130 a	323.1±151 b	146.2±134 a	209.0±183 ab
	Marksim	37.7±43 b	151.7±48 a	190.7±100 a	111.6±76 b
Annual	Observed	695.4±293 a	796.4±223 a	530.8±223 a	1174.2±310 a
	Giovanni	479.3±253 b	1042.1±211 b	471.7±202 a	904.2±407 a
	Marksim	352.6±121 b	931.6±295 ab	621.7±132 a	427.3±101 b

* MAM = March, April and May; OND = October, November and December.

** Means with the same letter are not significantly different at $\alpha = 0.05$.

Tanga and Giovanni at Morogoro. In summary, Giovanni rainfall amounts were significantly different ($\alpha = 0.05$) from observed values in 3 out of 12 observations while Marksim values were significantly different in 5 out of 12 observations. The two weather generators had better simulations of rainfall amounts during MAM season compared to OND and annual rainfalls.

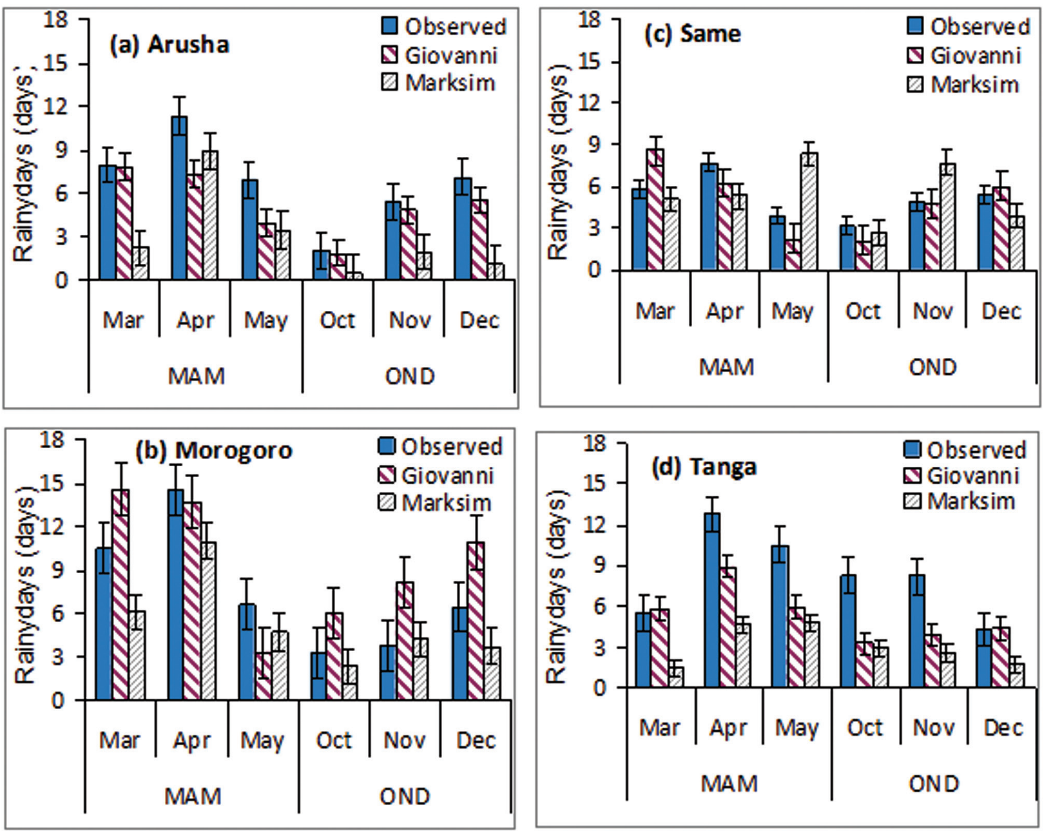

Figure 2: Comparison of rainy days (> 2 mm) for the Observed, Giovanni, and Marksim data sets at the four experimental stations. Error bars indicate standard errors of measurements.

Rainy days analysis

Rainy days were determined by considering a day with more than 2 mm (Figure 2) and 10 mm of rainfall (Figure 3) as described in Tumbo *et al.,* (2010). Ten-year average rainy days for each month were used to compare the two software packages. Considering the rainy days greater than 2 mm; during the OND season, the Giovanni overestimated the rainy days in one station (Morogoro) and under estimated at Tanga (Figure 2). It performed well on the rest of the two stations. For the same season, the Marksim software underestimated rainy days in three out of four stations (Arusha, Morogoro, and Tanga). During the MAM season all the two software packages predicted the rainy days well in three out of four stations (Arusha, Morogoro, and Same). At Tanga both the Giovanni and Marksim software packages underestimated the rainy days.

Considering rainfall greater than 10 mm; the Marksim software underestimated the rainy days in three out of four stations (Arusha, Morogoro, and Tanga) during the second rainy season (OND) (Figure 3). During the MAM season, all the two software packages predicted the rainy days well in three out of four stations (Arusha, Morogoro, and Same). At Tanga the Marksim software underestimated the rainy days. In summary, during MAM the Giovanni software predicted well the rainy days in 7 out of 8 observations while Marksim predicted well in 4 out of 8 observations. During OND better predictions were 5 out of 8 and 4 out of 8 for Giovanni and Marksim software packages, respectively.

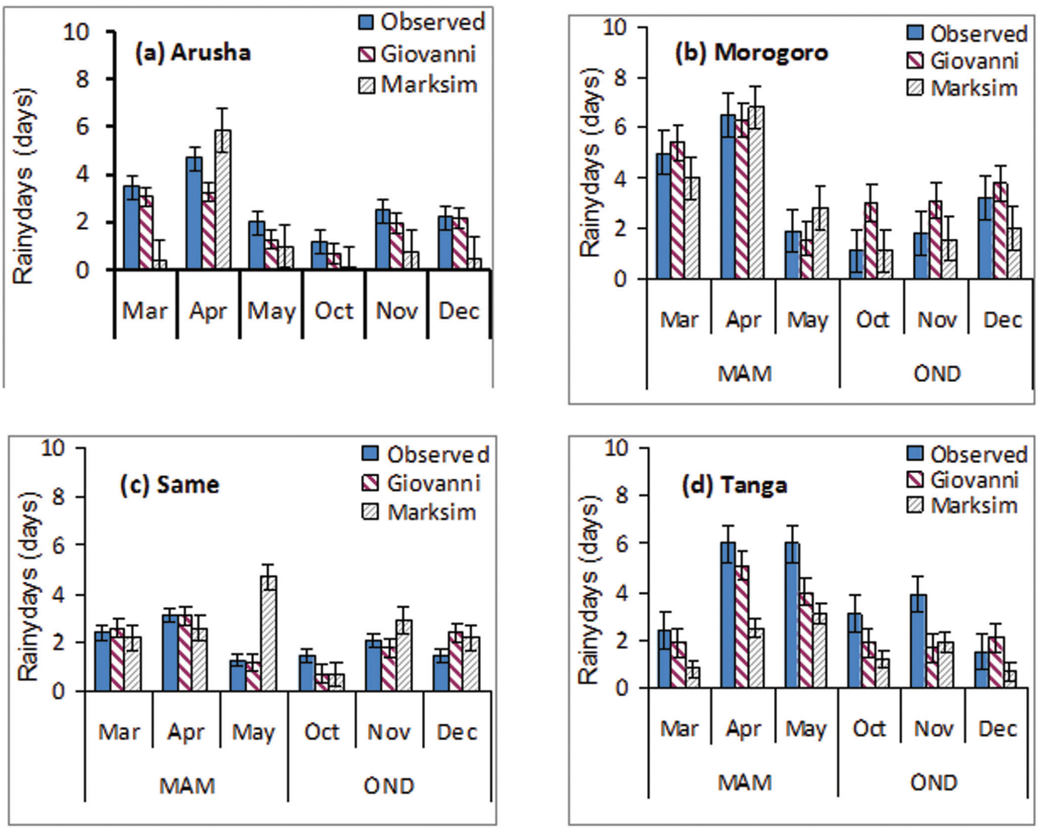

Figure 3: Comparison of rainy days (> 10 mm) for the Observed, Giovanni, and Marksim data sets at the four experimental stations. Error bars indicate standard errors of measurements.

Dry spells analysis

Analysis of the mean maximum dry spell runs was based on days with less than 1 mm of rainfall. Figure 4 presents a comparison of the two software packages (with respect to the observed data) in simulating the dry spell runs during MAM and OND seasons.

fairly accurate during MAM compared to OND. Table 3 presents summary of statistical comparison of the Marksim and Giovanni software packages in simulating probability

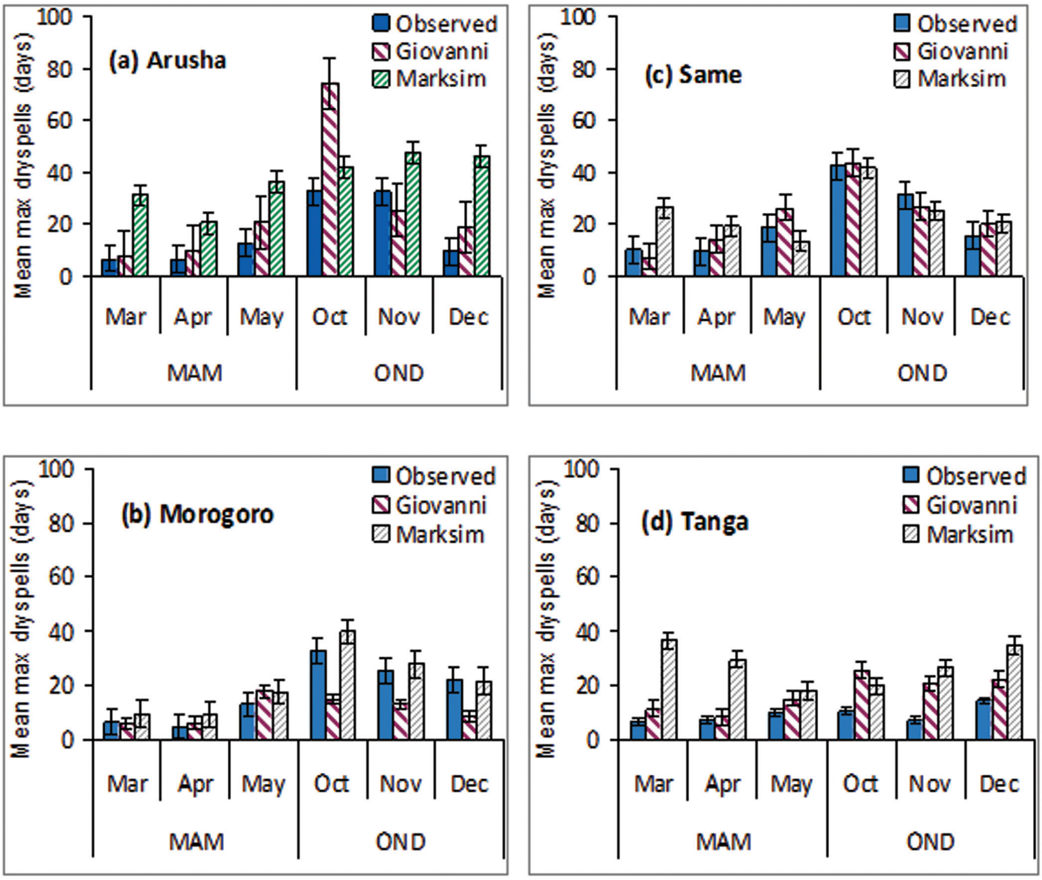

Figure 4: Dry spells analysis for the Observed, Giovanni, and Marksim data sets at four experimental stations. Error bars indicate standard errors of measurements.

During the MAM season, the Giovanni software simulated well the dry spell runs in all the four stations. Marksim software overestimated the dry spells in two out of four stations (Arusha and Tanga). For the vuli season, Marksim software predicted well the dry spells in two out of four stations, and over predicted at Arusha and Tanga. The Giovanni software predicted fairly well the dry spells in two out of four stations, under predicted at Tanga, and over predicted at Arusha. While all the two software packages performed equally during the OND season, the Giovanni datasets were more comparable to the observed data than the Marksim data sets during the MAM season. Hence the two software packages were of occurrence of dry spells greater than 7 days. Comparison was performed during the masika (MAM) season. The Giovanni software estimated the probabilities of dry spell occurrence much better (RMSE = 0.23, MAE = 0.18, R^2 = 0.46 and d =0.75) compared to the Marksim software (RMSE = 0.28, MAE = 0.23, R^2 = 0.30 and d =0.63). The Marksim software tended to overestimate the probabilities (MBE = 0.17) compared to the Giovanni software (MBE = 0.01).

Seasonal start dates, end dates, and length

Table 4 presents summary of statistical analysis of seasonal start dates and end dates simulated

Table 3. Comparison of the average seasonal start and end dates for the Observed, Giovanni and Marksim data sets.

			Experimental stations mean start and end dates (DoY)**			
Season*	Dates	Data set	Arusha	Morogoro	Same	Tanga
MAM	Start dates	Observed	87±18 a	68±6 a	83±15 a	83±22 a
		Giovanni	74±11 b	75±16 a	86±13 a	95±23 a
		Marksim	98±9 a	76±12 a	82±17 a	94±21 a
	End dates	Observed	130±8 ab	132±9 a	128±8 a	145±12 a
		Giovanni	142±19 a	159±21 b	139±16 a	181±36 b
		Marksim	129±9 b	123±23 a	133±13 a	130±9 a
OND	Start dates	Observed	302±14 a	307±17 a	303±15 a	291±9 a
		Giovanni	306±22 a	293±23 a	307±27 a	299±21 a
		Marksim	324±14 b	304±28 a	316±14 a	304±15 a
	End dates	Observed	337±3 a	338±5 a	339±5 a	342±8 a
		Giovanni	343±13 a	341±10 a	353±17 b	345±12 a
		Marksim	338±9 a	340±9 a	343±9 a	336±2 a

* MAM = March, April and May; OND = October, November and December.

** Means with the same letter are not significantly different at α = 0.05. Means are averages of 10-year observations. DoY = Day of the Year.

by the two software packages, as compared to the observed values. Ten-year averages of seasonal lengths were compared during the MAM and OND seasons. Statistical analysis of the ten-year average seasonal lengths is presented in Table 4. During the MAM season, there was no significant difference (α = 0.05) between the Marksim and observed start dates in all the four stations (Table 4). Giovanni start dates were significantly different (α = 0.05) from the observed values in one station (Arusha) out of four stations.

For the end dates, there was a significant difference (α = 0.05) between the observed and simulated end dates in two out of four stations for Giovanni (Morogoro and Tanga), and in one out of four stations (Arusha) for Marksim software.

During OND, Giovanni simulated well the start dates in all the four stations, and had significantly different end dates (α = 0.05) at one station (Same) out of the four stations (Table 4). The Marksim software simulated well the end dates in all the stations, but had one significantly different start date (α = 0.05) at Arusha. In summary, all the software packages simulated well the start dates and end dates in all the seasons, with the Giovanni having excellent start

Table 4. Comparison of the average seasonal lengths for the Observed, Giovanni and Marksim data sets.

		Mean seasonal lengths on experimental stations (days)**			
Season*	Data set	Arusha	Morogoro	Same	Tanga
MAM	Observed	47±20 a	64±12 a	45±15 a	61±21 a
	Giovanni	68±22 b	85±27 b	53±15 a	92±36 b
	Marksim	31±13 a	47±26 a	51±21 a	36±17 c
OND	Observed	35±12 a	31±16 a	36±16 ab	51±14 a
	Giovanni	37±26 a	48±14 b	46±24 a	46±24 ab
	Marksim	15±8 b	36±22 ab	27±11 b	32±15 b

* MAM = March, April and May; OND = October, November and December.

** Means with the same letter are not significantly different at α = 0.05. Means are averages of 10-year observations.

and end dates prediction during OND, while the Marksim software had excellent start dates and end dates predictions during the MAM season. All the two software packages predicted the start dates much better than the end dates (Table 3). Better simulations of start dates compared to the end dates by weather generators have also been reported in the literature (e.g., Smith *et al.,* 2005).

The Marksim software predicted well the seasonal lengths in three out of four stations (Arusha, Morogoro, and Same) during MAM. However, the software significantly underestimated (α = 0.05) the seasonal lengths in three out of four stations during OND (Table 5). The Giovanni software significantly overestimated (α = 0.05) the seasonal lengths in three out of four stations during MAM, and in one station (Morogoro) during OND. Therefore, the Marksim software estimated the seasonal length much better than the Giovanni software during MAM, while during OND, Giovanni was much better than Marksim. The Marksim model gives more accurate climatic data if long-term average climate data are used as input variables. Even without the long-term average climate data, the Marksim model can still be used to generate daily data in bimodal climatic areas where there are no observed data.Similar results on seasonal start and end dates are also reported by Tumbo *et al.,* 2010, and Mazandarani *et al.,* 2013.

Time series analysis of Giovanni rainfall data
Figure 5 presents the time series analysis of the Giovanni vs. observed values for the 10-year period from 1998 to 2007. The Marksim data sets were not used for time series analysis since the data are generated with random years as stated earlier. The Giovanni software had better predictions of seasonal rainfall during OND compared to MAM season.

The web based Giovanni software uses satellite data to generate the 3-hourly and daily weather data (Acker *et al.,* 2006). It could be a better option for use in the tropical bimodal areas that have neither observed data nor long-term average climate data. However if there is no internet connection, this weather generator cannot be useful. Considering its seasonal performance, the Giovanni software is better suited to the MAM than OND in bimodal climate areas. However, the recent improvements in OND predictions as shown in Figure 5 might make the Giovanni data also useful during OND. According to previous studies the Marksim software gives more accurate data if long-term (30-years) average climate data are used as input variables (Jones and Thornton, 2000). Hence

Table 5. Statistical analysis of probability of occurrence of dry spells greater than 7 days for the Giovanni and Marksim datasets vs. observed data during MAM season.

				Absolute error Measures*			Relative error Measures*	
Model	Station	N	Mean	RMSE	MAE	MBE	R2	d
Giovanni	Arusha	37	0.68	0.17	0.13	0.07	0.62	0.87
	Morogoro	37	0.56	0.34	0.27	-0.02	0.36	0.77
	Same	37	0.74	0.26	0.17	-0.12	0.10	0.54
	Tanga	37	0.78	0.16	0.13	0.12	0.77	0.84
	Average***	37	0.69	0.23	0.18	0.01	0.46	0.75
Marksim	Arusha	37	0.94	0.39	0.33	0.33	0.20	0.50
	Morogoro	37	0.63	0.26	0.20	0.05	0.55	0.86
	Same	37	0.89	0.13	0.10	0.02	0.17	0.64
	Tanga	37	0.93	0.33	0.27	0.27	0.26	0.52
	Average	37	0.85	0.28	0.23	0.17	0.30	0.63

* RMSE = root mean square error; MAE = mean absolute error; and MBE = mean bias error.
** R2 = coefficient of determination; and d = index of agreement.
*** Average statistic for the four experimental stations.

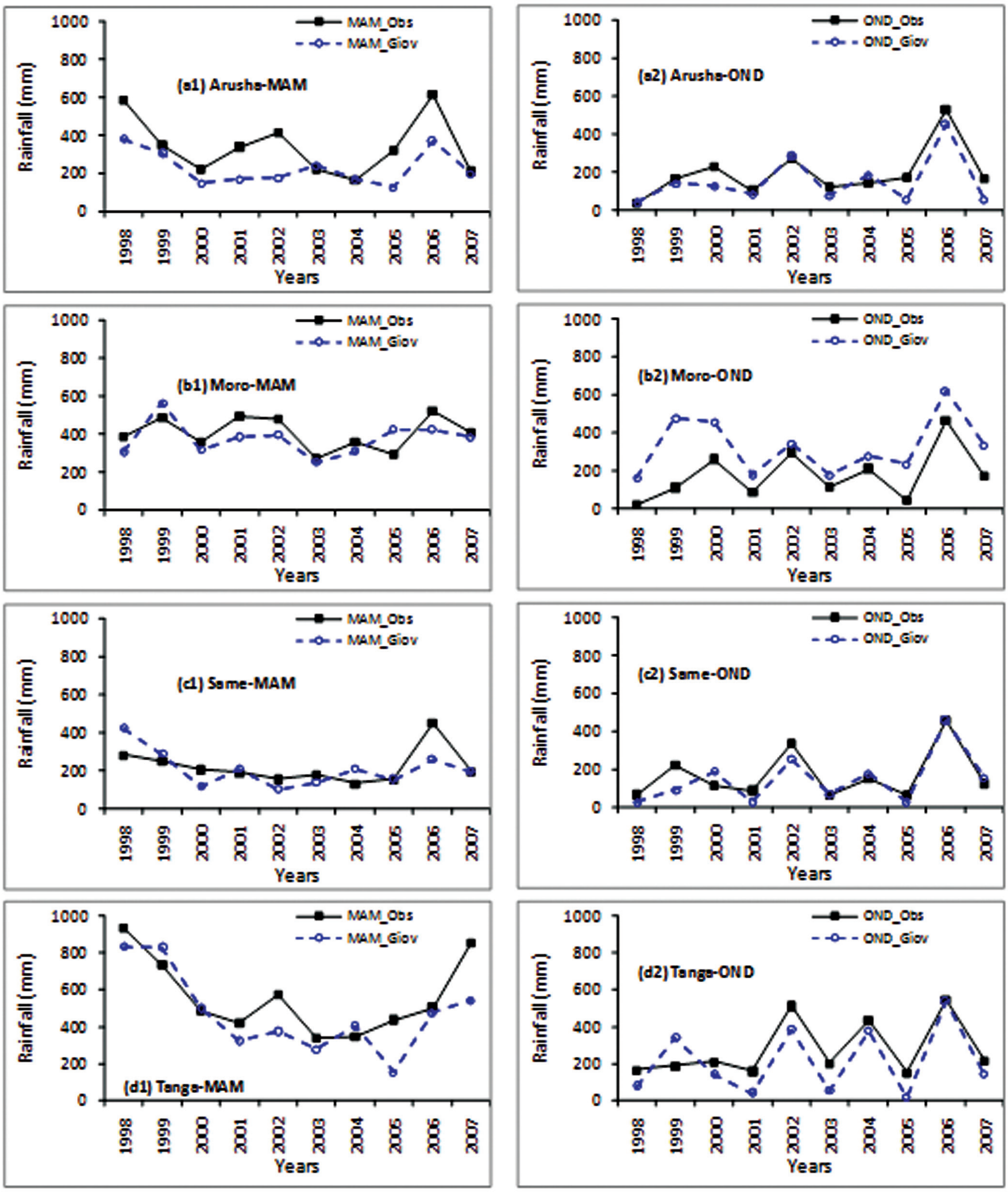

Figure 5. Time series analysis of the Giovanni vs. Observed data sets of seasonal rainfall during (1) MAM, and (2) OND.

the software is recommended to be used in the bimodal climate areas where long-term average climate data are available.

Summary and Conclusions

Marksim and Giovanni weather generators were compared for their accuracy in predicting rainfalls in the northern bimodal areas of Tanzania. Ten-year rainfall data generated by the two software packages were compared against observed data from four stations within the study area. Rainfall amounts, generated using the Giovanni software, are much better than those generated using the Marksim software during both *masika* (MAM) and *vuli* (OND) seasons. Also, the software packages simulated rainfall amounts better during the MAM compared to OND seasons. Therefore, this study recommends the use of rainfall data generated using Giovanni software over Marksim especially in areas similar to the northern bimodal areas of Tanzania.

The Giovanni software can be used to provide estimates of rainfall in areas with no rainfall station, which is typical of most of the sub-Saharan African countries including Tanzania. Even in some areas where observation stations are available, much of the data tend to have gaps of up to one year. One can fill in the gaps using estimates from the Giovanni software. Access to daily rainfall data is another challenge. Most agencies have very strict policies in providing their daily meteorological data. For example, the price is very high for daily rainfall data in Tanzania for stations monitored by Tanzania Meteorological Agency such that most researchers in agriculture cannot afford. With Giovanni software, one can easily generate three hourly and daily rainfall data free of charge as long as access to the Internet is available. With easy and faster access to such data it will be easier to monitor conditions of agricultural crop during the growing season, estimate potential seasonal production, simulate the crop production potential of an area and investigate the effects of the current climate change using crop simulation models. Access to the Giovanni software for rainfall data is through the website: http://agdisc.gsfc.nasa.gov/Giovanni/aovas/. The site provide access to current global and regional conditions (near-real-time and experimental), and research quality data (global and regional archives), which this study used. The software needs fewer inputs of latitude and longitude data that define the area of interest, and specification for either map/plot or time series.

Acknowledgements
This publication was supported by the Climate Change Adaptation in Africa (CCAA) program, a joint initiative of Canada's International Development Research Centre (IDRC) and the United Kingdom's Department for International Development (DFID). The images and data used in this study were acquired using the GES-DISC Interactive Online Visualization AndaNalysis Infrastructure (Giovanni) as part of the NASA's Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The views expressed are those of the authors and do not necessarily represent those of DFID, IDRC, or GES-DISC.

References

Acker, J., Berrick, S., Kempler, S., Leptoukh, G., Liu, Z., Rui, H., Teng, W., Shen, S., Zhu, T. and Johnson, J. (2005). Giovanni: A vital tool enabling rapid and accurate climate data analysis for evaluation, assessment, and research. U.S. Climate Change Science Program Workshop. Arlington, Virginia: November 14-16.

Acker, J., Leptoukh, G., Kempler, S., Berrick, S., Rui and H. Shen, S. (2006). Application of NASA Giovanni to coastal zone remote sensing research. EARSeL Workshop Proceedings. Bolzano, Italy, 7-9 June.

Acker, J.G. and Leptoukh, G. (2007). Online analysis enhances use of NASA earth science data. Eos, Transactions, American Geophysical Union 88(2): 14-17.

Ahmad, S., Leptoukh, G., Johnson, J., Gerasimov, I. and Kempler, S. (2007). Exploring meteorological variables in hurricanes using Giovanni. In: Proc. Amer. Meteorological Soc. 23rd Int. Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, San Antonio, Texas, 2007 Jan 15-18.

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. (2006). Crop Evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56. Rome, Italy, FAO.

Bruhn, J.A., Fry, W.E. and Fick, G.W. (1980). Simulation of daily weather data using theoretical probability distributions. Journal of Applied Meteorology 19(9): 1029-1036.

Dennett, M.D., Rodgers, J.A. and Stern, R.D. (1983). Independence of rainfall through the rainy season and the implications for the estimation of rainfall probabilities. Royal Met Society. Journal of Climatology 3: 375-384.

Hargreaves, G.H. and Samani, Z.A. (1985). Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture 1: 96-99.

Howden, S.M., Soussana, J.F., Tubiello, F.N., Chhetri, N., Dunlop, M. and Meinke, H. (2007). Adapting agriculture to climate change. PNAS 104(50); 19691-19696.

Jones, P.G. and Thornton, P.K. (2000). MarkSim: Software to generate daily weather data for Latin America and Africa. Agronomy Journal 92: 445 - 453.

Jones, P.G. and Thornton, P.K. (2002). Spatial modeling of risk in natural resource management.Conservation Ecology 5(2): 27.

Jones, P.G. and Thornton, P.K. (2003). The potential impacts of climate change on maize production in Africa and Latin America in 2055. Global Environmental Change 13: 51-59.

Kaufman, Y.J., Koren, I., Remer, L.A., Tanré, D., Ginoux, P. and Fan, S. (2005). Dust transport and deposition observed from the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) spacecraft over the Atlantic Ocean., J. Geophys. Res. 110: D10S12.

Kusaka, H., Crook, A., Dudhia, J. and Wada, K. (2005). Comparison of the WRF and MM5 models for simulation of heavy rainfall along the Baiu front. SOLA 1: 197-200.

Lanza, L.G. (2000). A conditional simulation model of intermittent rain fields. Hydrology and Earth Systems Sciences 4(1): 173-183.

Legates, D.R. and McCabe Jr., G.J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydro-climatic model validation. Water Resources Research 35(1): 233-241.

Leptoukh, G., Cox, S., Farley, J., Gopalan, A., Mao, J. and Berrick, S. (2007). Exploring NASA and ESA atmospheric data using Giovanni, the online visualization and analysis tool. In: Proc. Envisat Symposium. (2007,Montreux, Switzerland: 23–27 April 2007.

Liu, J., Cui, Z., Zhang, L., Dou, X. and Lu, D., (1999). Comparison study on rainfall retrieval algorithm: model simulation and armar data. Poster session 6. In: Proceedings, GIS Development, ACRS.

Mazandarani, G.A., Ahmadi, A.Z. and Ramazani, Z. (2013). Investigation Analysis of the Agronomical Characteristics of the Daily Rainfall in Rain-Fed Agriculture (Case study: Tehran). International Journal of Agriculture and Crop Sciences 5(6): 612-619.

NASA. (2007). Giovanni-3 Online Users Manual. Goddard Earth Sciences Data and Information Services Centre, National Aeronautics and Space Administration [online] http://disc.sci.gsfc.nasa.gov/giovanni/G3_manual_table_of_contents.shtml (accessed 12/02/2009).

Prados, A., Leptoukh, G., Alston, E.J. and Sokolik, I.N. (2007). Assessing U.S. air quality using CALIPSO and MODIS data via Giovanni. In: Proc. Amer. Geophysical Union Fall Meeting, San Francisco, California, 2007 Dec 10-14.

Rahman, M.S. (2000). A Rainfall simulation model for agricultural development in Bangladesh. Discrete Dynamics in Nature and Society 5(1): 1 - 7.

Rui, H., Farley, J., Leptoukh, G. and Berrick, S. (2007). New data and capabilities in the NASA Goddard Giovanni system. In: Proc. Amer. Geophysical Union Fall Meeting, San Francisco, California, 2007 Dec 10-14.

SAS Institute Inc. (2004). Statistical analysis system (SAS) software version 9.1.2, Cary, NC: SAS Institute, Inc.

Semenov, M.A., Brooks, R.J., Barrow, E.M. and Richardson, C.W., (1998). Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Climate Research 10; 95–107.

Smith, M.A., Thurow, T.L. and Legg, D. (2005). Predicting herbage standing crop on Wyoming rangeland from winter and spring precipitation.58th Annual meeting of the Society for Range Management. Ft. Worth, TX., 5-11 February.

Steel, R.G.D., Torrie, J.H. and Dickey, D.A. (1997). Principles and Procedures of Statistics: A Biometrical Approach, 3rd edition. New York, NY: McGraw-Hill Inc.

Stern, R.D., Rijks, D., Dale, I. and Knock, J. (2006). INSTAT Climatic Guide.INSTAT Statistical Services Centre, University of Reading. Reading, UK.

Tingem, M., Rivington, M., Azam-Ali, S. and Colls, J. (2007). Assessment of the ClimGen stochastic weather generator at Cameroon sites. African Journal of Environmental

Science and Technology 1(4): 086-092.

Tumbo, S.D., Mpeta, E.,Tadross, M., Kahimba, F.C., Mbillinyi, B.P. and Mahoo, H.F. (2010). Application of self-organizing maps technique in downscaling GCMs climate change projections for Same, Tanzania. J. Physics and Chemistry of the Earth 35: 608–617.

Wilks, D.S. (1995). Statistical Methods in Atmospheric Sciences.John Wiley & Sons Ltd.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnel, J. and Rowe, C.M. (1985). Statistics for the evaluation and comparison of models.Journal of Geophysical Research 90(C5): 8995-9005.

Willmott, C.J. and Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate theperformance of spatial interpolators. International Journal of Geographical Information Science 20(1): 89–102.