JOURNAL OF ENGINEERING RESEARCH

ISSN: 0795 – 2333

June 2010. JER-15, No. 2

CONTENTS

COMPARATIVE EVALUATION OF CLIMATIC DATA OBSERVATION METHODS IN IBADAN, NIGERIA

T.A. EWEMOJE and O.S. EWEMOOJE***

**Department of Agricultural and Environmental Engineering, Faculty of Technology, University of Ibadan, Nigeria*

***Department of Statistics, Faculty of Science, University of Ibadan, Nigeria*

ABSTRACT

Sustainability of agriculture is dependent upon availability of reliable climatic data for planning. Daily and monthly data were obtained for Ibadan, Nigeria from International Institute of Tropical Agriculture (IITA) automatic weather station and Nigerian Meteorological Agency (NMA) manual weather station. Two methods were employed to analyse the data. First method (M_1) used daily data to generate sets of linear equations. Each equation represents linear relationship between the climate parameter measured at station X (manual) and that same measured at station Y (automatic). Second method (M_2) used monthly values to analyse the data. Accuracy of regression method was analysed by calculating Error Variance (EV) between manual and automatic stations. Errors associated with deviation-based statistics (RMSE) are generally higher than regression-based statistics (EV) for all climate parameters considered. Introduced deviation and correlation based statistics of Mean Squared Deviation (MSD) and its components do not explicitly eliminate error introduced from linearity assumption in regression analysis. Hence, discrepancies have not been adequately explained, but with $r < 0.50$ in all the climate parameters, the model is weakly correlated with measurement. Analyses have shown that manually observed climate data should not be substituted with automatically observed climate data without correcting data bias/errors prior to usage.

Keywords: *Weather data, Comparative evaluation, Manual, Automatic, Statistical analyses.*

1.0 INTRODUCTION

In spite of the importance of meteorological data for agricultural planning and assessment, forecasts, warnings, engineering, environmental impact assessment, and so on, few meteorological stations especially in developing countries measure accurately and continuously these data, but accurate data collection is a primary objective of any field research project and the validity of subsequent analyses is dependent on the accuracy of field data (Wu et al, 2005). Detailed long term field experiments in developing countries are often difficult to conduct due to financial or personnel limitations, hence the application of simulation models which have been developed with data measured under more accommodating conditions and whose mathematical relationships apply to a wide range of conditions according to (Gijsman et al. 2002) is therefore an attractive option as off-the-shelf model developed for a particular location will not necessarily lead to results that are applicable to another location or situation. (Akpabio et al. 2004) proposed use of correlation model in solving these problems of data inadequacy in locations of similar latitude, altitude and climatology. The amount of weather data needed to obtain an adequate estimate of a statistical descriptor of climate has long been debated but (Richardson, 2000) stated that for many climatic variables, no definite length of weather data can be identified as a rule but long records will obviously give more reliable estimates.

Dominant sources causing higher error magnitudes in quantitative data have been carried out and a framework for identification of these error magnitudes in field measured data has been evaluated (Dulal et al. 2006). Accordingly, for precipitation and any other meteorological data (Michelson et al. 2006) classified the causes of these errors as; instrumental, site, human (both systematic and random) and management errors. To avoid errors as a result of insufficient data availability, daily totals of meteorological data from automatic and manual observation stations according to (Bruton et al. 2000) should not be included in a monthly or annual total if either were missing on a given day in order to remove bias from meteorological data comparisons due to missing data.

It is necessary that errors are corrected prior to use in quantitative applications (Michelson, 2004). To account for weather variability in the comparison of automatic and manual measurement, (Bruton et al. 2000) conducted a research over a period of six years which span over several annual crop growing seasons so as to correct the effect of variation in measurement of weather parameters. It was concluded that improved maintenance of automated observations is recommended to justify the replacement of the manual observations.

The profitability and sustainability of agriculture are dependent upon climatic patterns and in particular rainfall (Bosch et al. 1999), and temperature (Trewin, 2010). Rain fed agriculture is mostly practiced in Nigeria and according to (Idowu and Gbuyiro 2002) it is the most variable of all the climatic elements. Combination of evapotranspiration and rainfall is fundamental in water balance of an environment. Variability of evapotranspiration patterns over a 9-year period between years and different months was investigated by (Enciso et al. 2005) and it was concluded that irrigation guidelines can be developed based on historical data since coefficient of variability is less than 15%.

Rainfall analysis for irrigation management was examined by (Akinyemi et al, 2006) over a 10-year period from IITA Ibadan automatic weather station data and it was discovered that there was a significant difference ($p<0.05$) in rainfall amounts from year to year for the period 1991-2002 investigated. Precipitation distributions are highly skewed since they commonly consist of many low values and a few high values. Hence it is therefore unwise to use linear regression and correlation coefficient as a measure of agreement because few high values will disproportionately affect derived relationship. Also, existence of a linear relationship assumption between independent (x) and dependent (y) data can not be ascertained, hence unnecessary for x-y comparison (Kobayashi et al, 2000). Therefore, a means of approximating normal distribution with such data is to transform to decibel scale (Michelson, 2004).

Correlation and regression coefficients are not explicitly related to other commonly used statistics such as Root Mean Square Deviation (RMSD) which is also known as Root Mean Square Error (RMSE) according to (Retta et al. 1996). RMSE represents the mean distance between simulation and measurement. Deviation-based statistics are often used in conjunction with correlation and regression-based statistics (Retta et al. 1996 and Kiniry et al,1997). To correct the assumption of the regression that an output (y) is linearly related to a measurement (x), an approach based on Mean Squared Deviation (MSD) which is better suited to the x-y comparison than regression was proposed by Kobayashi and Salam, (2000). This MSD approach is the sum of three components: Squared Bias (SB), Squared Difference between Standard Deviations (SDSD), and Lack of Correlation weighted by the Standard Deviation (LCSD).

2.0 METHODOLOGY

The study area lies roughly between latitudes 6^0-8^0N and longitudes 3^0-6^0E. The two weather stations are the IITA and the Nigerian Meteorological Agency station, all in Ibadan. Daily meteorological data for 2003-2004 and mean monthly meteorological data for 1998-2005 were obtained from IITA automatic (model) weather station and also the manual (measured) observation station of Nigerian Meteorological Agency (NMA). The two weather observation stations were located approximately 2Km apart, at an altitude of 228m above mean sea level.

In the first analysis, two methods were employed to analyse the following data; minimum and maximum Temperature, Radiation, Wind speed, Relative humidity, Evaporation and Rainfall. The first method (M_1) used two years of daily weather data to generate a table of 365linear equations. Each equation represents the linear relationship between the climate parameter measured at station X (manual observation station - measured) and that same climate parameter measured at station Y (automatic observation station - model). Daily totals of meteorological data from both stations are excluded if either were missing on a given day as recommended by (Bruton, J.M., Hoogenboom, G. and McClendon, R.W., 2000). The regression coefficients were;

$$A_k = (A_{K1} + A_{K2})/2 \text{ and } B_k = (B_{K1} + B_{K2})/2 \quad \text{.....(1)}$$

Where; k represents the days of the year from 1 to 365. Subscript 1 and 2 represents the first and second years.respectively.

For second method (M_2), mean monthly values of weather data for 8-year period were computed in order to analyse the weather data. Accuracy of linear regression method of the weather data was analysed by calculating the Error Variance (EV) between manual observation data station and automatic observation data station (Whitmore, 1991).

$$EV = SS/(n-2) \quad (2)$$

where $$SS = \sum_{i=1}^{n} (Y_i - A - BX_i)^2 \quad \text{.....(3)}$$

n = Number of data points; X_i = Manual observation data; and Y_i = Automatic observation data.

Also following (Whitmore, A.P., 1991) recommendation, the Root Mean Square Error (RMSE) which quantifies the dispersion between data was also used.

$$RMSE = \left[\sum_{i=1}^{n} (X_i - Y_i)^2 / N\right]^{0.5} \quad \text{.....(4)}$$

The precipitation in mm/day is transformed to decibel scale in order to approximate it to a normal distribution using;

$$dBR = 10 \times \log(R) \quad \text{.....(5)}$$

The second analysis was used to correct the assumption of linearity in the regression based statistics. Hence, the difference between the model output and measurement was computed from;

$$MSD = \frac{1}{n}\sum_{i=1}^{n} (x_i - y_i)^2 \quad \text{.....(6)}$$

This implies that MSD=RMSE^2. Partitioning MSD into two components gave

$$MSD = (\bar{x} - \bar{y})^2 + \frac{1}{n}\sum_{i=1}^{n} [(x_i - \bar{x}) - (y_i - \bar{y})]^2 \quad \text{....(7)}$$

Where $\bar{x}$ and $\bar{y}$ are the means of x_i and y_i (i=1, 2, 3 ... n). The first term of the equation (7) denoted as Squared Bias (SB) represents the bias of the model from measurement and given as

$$SB = (\bar{x} - \bar{y})^2 \quad \text{.....(8)}$$

The second term is the difference between the model and the measurement with respect to deviation from the means given as the Mean Squared Variation (MSV)

$$MSV = \frac{1}{n}\sum_{i=1}^{n} [(x_i - \bar{x}) - (y_i - \bar{y})]^2 \quad \text{.....(9)}$$

Partitioning the MSV, standard deviation of the model is denoted as SD_{mo}, the measurement is denoted as SD_{me}, and correlation coefficient (r) between the model and measurement is given as

$$SD_{me} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{.....(10)}$$

$$SD_{mo} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \text{.....(11)}$$

$$r = \left[\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right] \div (SD_{me} SD_{mo}) \quad \text{.....(12)}$$

Substituting equations (10) and (11) in equation (9) on expansion, we have:

$$MSV = SD_{mo}^2 + SD_{me}^2 - 2\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \quad \text{....(13)}$$

Expressing MSV in equation (13) in terms of the correlation coefficient r, gives:

$$MSV = (SD_{mo} - SD_{me})^2 + 2SD_{me}SD_{mo}(1-r) \ldots\text{(14)}$$

Hence,

$$SDSD = (SD_{mo} - SD_{me})^2 \ldots\ldots\text{(15)}$$

$$LCSD = 2SD_{me}SD_{mo}(1-r) \ldots\ldots\text{(16)}$$

SDSD in equation (15) is the difference in the magnitude of fluctuation between the model and measurement. A bigger value of SDSD shows that the model failed to suggest the magnitude of fluctuation among the n measurements. LCSD in equation (Mott, P.; Sammis, T.W and Southward, G.M., 1994.) is the lack of positive correlation weighted by the standard deviations. A bigger value of LCSD shows that the model failed to suggest the pattern of fluctuation across the n measurements.

Combining the above expressions, MSV and MSD can be rewritten as:

$$MSV = SDSD + LCSD \ldots\ldots \text{(17)}$$

$$MSD = SB + SDSD + LCSD \ldots\ldots\text{(18)}$$

Equation (16) shows the role of the correlation coefficient, r, in LCS and hence in MSD. A bigger value of r would reduce MSD and therefore increase the model accuracy. Nevertheless, r is only a component out of the many in MSD; other components may be more significant in determining MSD than r.

3.0 RESULTS AND DISCUSSIONS

Error analyses of regression-based statistics (EV) were compared with the deviation-based statistics (RMSE) over a 2-year period with daily meteorological data and over 8-year period with mean monthly meteorological data in Ibadan. Minimum and maximum temperature error analysis of EV and RMSE shows statistically at $P \leq 0.05$ that relative difference in error associated with EV is generally higher than RMSE when manual and automatic data are compared. This implies that dispersion between manual and automatic recorded climate parameter is negligible.

Table 1: Meteorological Data Error Analyses

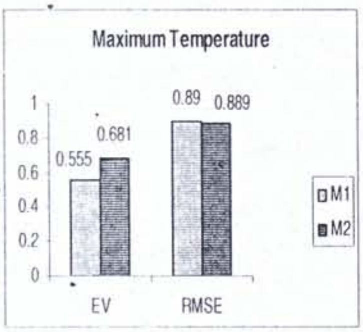
	M_1		M_2		$\Delta = \lvert M_1 - M_2 \rvert$	$\Delta = \lvert M_1 - M_2 \rvert$
Climate Parameter	EV	RMSE	EV	RMSE	EV Relative Difference	RMSE Relative Difference
Max Temp	0.555	0.89	0.681	0.889	0.126	0.001
Min Temp	0.828	1.358	1.084	1.211	0.256	0.147
Radiation	2.699	2.549	5.787	2.965	3.088	0.416
Wind Speed	0.72	2.222	1.415	2.769	0.695	0.547
Relative Humidity	18.609	16.74	1.918	14.877	16.691	1.963
Evaporation	0.895	1.672	2.592	1.605	1.697	0.067
Rainfall	32.686	7.677	23.383	4.886	9.303	2.791

From Figure 1(a, b), it was observed that higher error values were associated with the two methods (M_1 and M_2) based on deviation statistic (RMSE) than regression statistics (EV). However, other factors could have been responsible for the minimal error observed from the regression statistics based on the assumption of the linear relationship between the climate parameter measured at a manual observation station and that same climate parameter measured at automatic observation station.

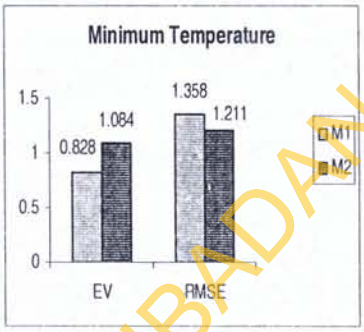
Considering Radiation, Wind speed, Relative humidity, and Evaporation, analysis of error variance of the two methods suggests that the difference between manually and automatically observed climate parameters is high. Hence, replacing a manual data with an automatic data (and vice versa) will introduce greater errors or uncertainties in measurement and evaluation. Similarly, relative differences observed in EV and RMSE analyses (Table 1) implies that using either daily or mean monthly climate data, large discrepancies are associated with these climate parameters measured for the same location as shown in Figure 1 (c-g).

As a result of the assumption that a linear relationship exists between model and measured values, there is the need to estimate error variance independently from this Assumption because each measurement are based on replicated measurements of these climatic data over a long period.

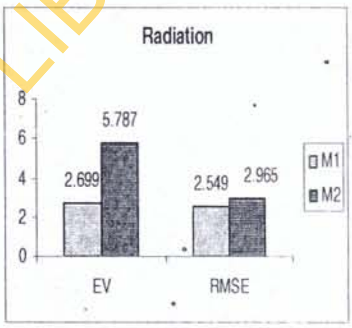
Rainfall gave the highest EV values for the two methods used when compared to RMSE values for all climate parameters considered. This may be attributed to transformation to decibel scale in order to approximate it to a normal distribution thereby removing skewing effects of few high rainfall events.



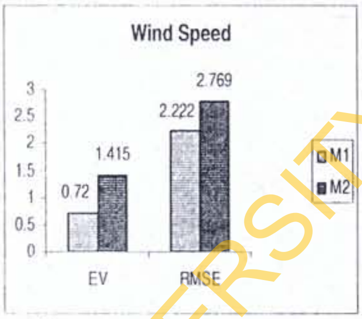
a



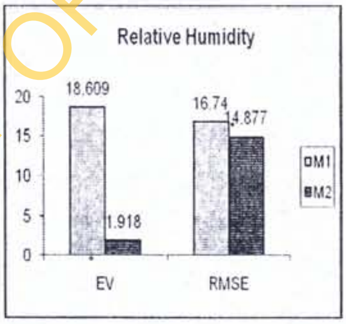
b



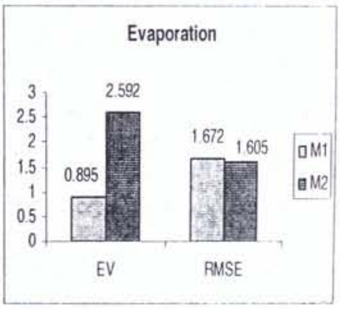
c



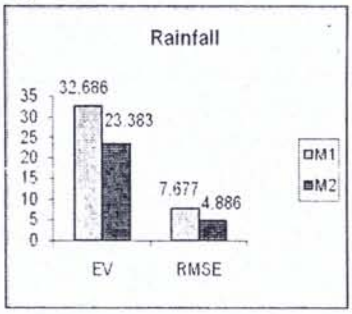
d



e



f



g

Figures 1(a-g): Comparison of Climate Parameters using Regression-Based (Error Variance) and Deviation-Based (RMSE) Statistics

Hence, mean squared deviation (MSD) and its components were introduced because error from the linearity relationship in regression analysis could not be taken for granted. However, because of the different aspects of the model-measurement discrepancies which may not have been adequately covered, deviation based statistics were used in conjunction with correlation based statistics. The analyses of the MSD and its components, with the inclusion of correlation coefficients were presented in Tables 2 and 3. Graphical representation of the climate parameters analysed was given in Figure 2(a-d).

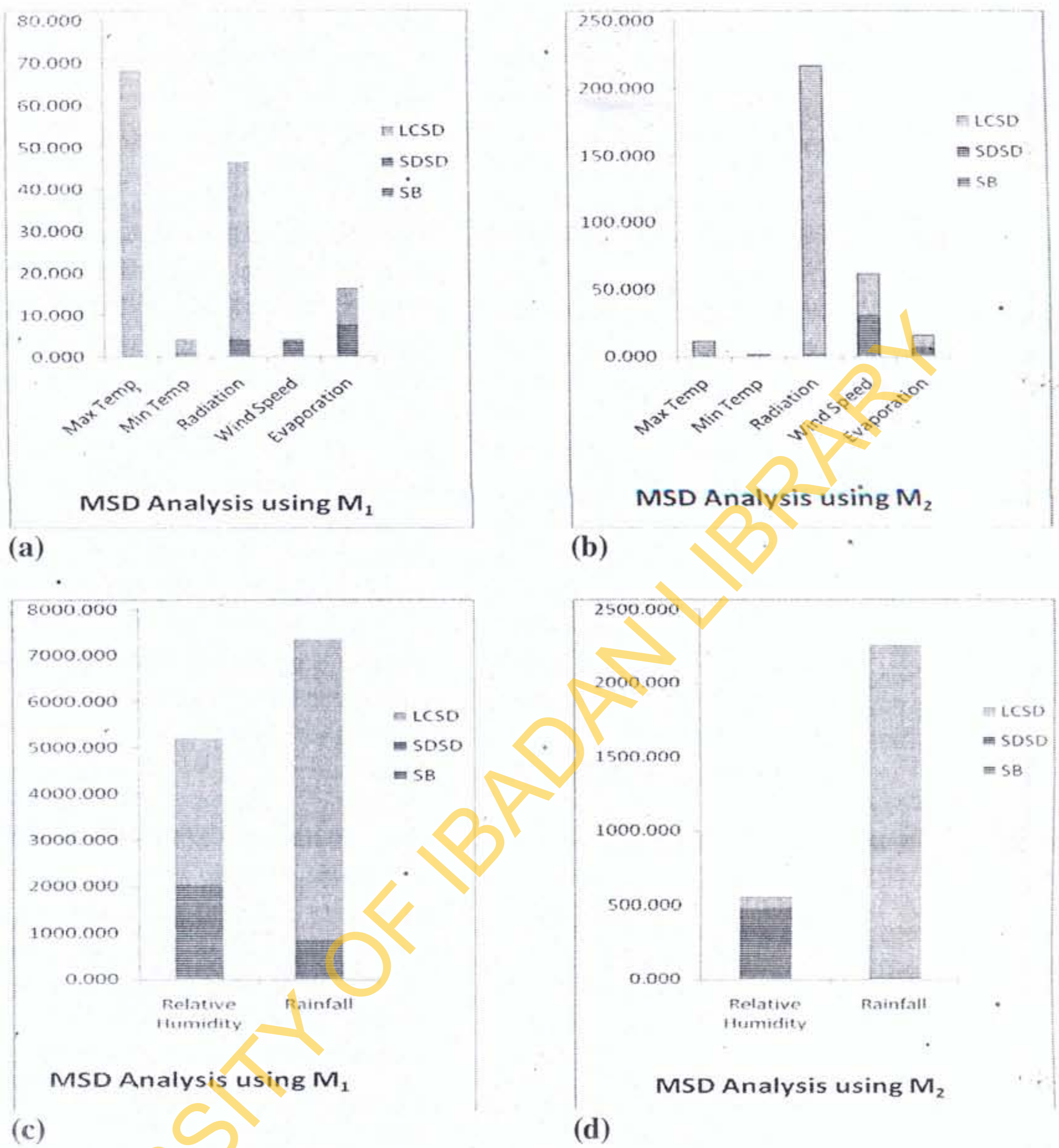
The results in tables 2 and 3 shows the values obtained for the various climate parameter in the analysis of the mean squared deviation and its components using the first method (M_1) with two years of daily weather data and the second method (M_2) with mean monthly values of weather data for the 8-year period. The larger the correlation coefficient (r) the smaller the MSD value for all the climate parameter being considered. Considering M_1, rainfall and minimum temperature with the smallest and largest r values of 0.009 and 0.417, have the corresponding MSD values of 7364.012 and 4.460 are the largest and smallest respectively, while under M_2 rainfall and minimum temperature have the smallest and largest r values of 0.019 and 0.356, with the corresponding MSD values of 2263.609 and 2.126 as the largest and smallest respectively. Similar results are obtained when r was compared with deviation statistics of MSV, SD_{me} and SD_{mo} in Tables 2 and 3. The results showed that all the models were weakly correlated with the measurement, $r < 0.50$.

Table 2: Analysis of Mean Squared Deviation Components using M_1

Climate Parameter	MSD	MSV	SD_{me}	SD_{mo}	r	SB	SDSD	LCSD
Max Temp	68.387	68.135	6.226	6.442	0.151	0.252	0.047	68.088
Min Temp	4.460	3.566	1.743	1.753	0.417	0.893	0.000	3.566
Radiation	46.735	44.097	5.605	4.331	0.125	2.638	1.624	42.473
Wind Speed	4.572	0.882	0.780	0.740	0.237	3.690	0.002	0.880
Relative Humidity	5196.145	4967.563	66.506	24.155	0.012	228.582	1793.665	3173.898
Evaporation	16.544	16.405	4.167	1.398	0.250	0.139	7.666	8.740
Rainfall	7364.012	7363.644	73.708	44.615	0.009	0.368	846.412	6517.232

Table 3: Analysis of Mean Squared Deviation Components using M_2

Climate Parameter	MSD	MSV	SD_{me}	SD_{mo}	r	SB	SDSD	LCSD
Max Temp	11.920	11.875	2.992	2.828	0.300	0.045	0.027	11.848
Min Temp	2.126	2.056	1.166	1.347	0.356	0.070	0.033	2.023
Radiation	216.507	214.825	10.988	10.410	0.062	1.682	0.334	214.491
Wind Speed	61.804	61.275	7.816	2.331	0.144	0.529	30.082	31.193
Relative Humidity	554.270	352.123	18.705	1.917	0.020	202.147	281.842	70.281
Evaporation	16.216	15.740	1.549	4.111	0.280	0.476	6.568	9.172
Rainfall	2263.609	2263.228	32.424	35.444	0.019	0.382	9.119	2254.108



Figures 2(a-d): Graphical representation of MSD Analysis using the two Methods

Comparing model with measurement is quite straightforward by using MSD. Here, the components SB, SDSD and LCSD are simply added; therefore, the user can identify the major component of MSD and investigate it further. It has been reported in previous research works that if SB is the major component of MSD, maximizing r values does not improve the model accuracy much. However, figure 2(a-d) shows that LCSD is the major contributor to the magnitude of the MSD statistics and hence, maximizing the r values would lead to minimizing MSD values and hence an increase in model accuracy. In the case of manual (measured) and automatic (model) weather observation station being analysed, maximizing the correlation coefficient r values of the climate parameter will result in a good comparison between the manual and automatic weather data for the daily and mean monthly observed data for different periods in this work. From the relationship in Equation (18), the results in Figure (2) showed that LCSD is the major component influencing MSD followed by SDSD and SB respectively.

The low MSD values for the following climate parameters; minimum temperature, maximum temperature and Evapotranspiration, is an indication that model values from automatic weather station is close in magnitude to the measured values from the manual weather station. The SB values which represent the bias between the means of the measurement and model showed that the bias in all the climate parameters considered was negligible with the exception of relative humidity values as seen in Tables 2 and 3.

The difference between model and measurement with respect to deviation from means denoted by MSV showed that manual values cannot be used to represent/replace automatic values because of the bigger MSV values observed in Table 2 for maximum temperature, radiation, relative humidity and rainfall. Moreover radiation, relative humidity and rainfall have bigger MSV values (Table 3). The difference in MSV for maximum temperature may be attributed to large variation in the daily values of this climate parameter in Table 2 with 2-year daily weather data while not much variability occurred in mean monthly weather data for the 8-year period (Table 3).

Assessing the magnitude of fluctuation among n measurements and the pattern of fluctuation across n measurements using equation 15, climate parameters investigated showed a very small SDSD values excepting relative humidity and rainfall for the 2-year daily weather data (Table 2) while only relative humidity values was relatively high in Table 3. Hence for relative humidity, the model values failed to simulate the magnitude of fluctuation in the measured values. Similarly for the LCSD, larger values are recorded for the radiation and rainfall values (Table 3). This implied that for radiation and rainfall, the model failed to simulate the pattern of the fluctuations across the measured values.

4.0 CONCLUSION

Errors associated with deviation-based statistics (RMSE) are generally higher than regression-based statistics (EV) for all climate parameters considered from the relative difference in data error analyses (table1). First method (M_1) values which used daily weather data are lower than second method (M_2) values which used mean monthly values of the weather data in the computation of Error Variance and Root Mean Square Error. This implies that analysis and evaluation of large meteorological data are better carried out with daily time-step than monthly time-step which is the bases of M_1 and M_2 respectively.

The introduced MSD and its components from the foregoing analysis does not explicitly give a result to show that the error introduced from the assumption of linear relationship in regression analysis could be taken care of by the deviation and correlation based statistics. Hence, a concrete conclusion cannot be drawn because of the different aspects of the model-measurement discrepancies which have not been adequately explained. But with $r<0.50$ in all the climate parameters, the model is weakly correlated with the measurement.

Hence, analyses have shown that manual station data should not be substituted with automatic station data without first correcting bias/errors prior to usage.

5.0 REFERENCES

(1) Akinyemi, J. O. and Akinyemi, M. L. (2006), "*Rainfall Analyses for Irrigation Management in Ibadan South-western Nigeria*". St. Joseph, Mich. ASABE Paper No. 062031.

(2) Akpabio, L. E.; Udo, S. O. and Etuk, S. E. (2004), "*Empirical Correlations of Global Solar Radiation with Meteorological data for Onne, Nigeria*". Turk. J. Phys. Vol 28, pp.205-212.

(3) Bosch, D. D.; Sheridan, J. M. and Davis, F. M. (1999), "*Rainfall Characteristics and Spatial Correlation for the Georgia Coastal Plain*". ASAE Trans. Vol. 42(6), pp.1637-1644.

(4) Bruton, J. M.; Hoogenboom, G. and McClendon, R. W. (2000), "*A Comparison of Automatically and Manually Collected Pan Evaporation Data*". ASAE Trans. Vol. 43(5), pp.1097-1101.

(5) Dulal, K. N.; Takeuchi, K. and Ishidaira, H. (2006), "*A Framework for the Analysis of Uncertainty in the Measurement of Precipitation Data: a Case Study for*

Nepal". Agric Engrg International: the CIGR Ejournal. September, Vol. VIII.

(6) Enciso, J. and Wiedenfeld, B. (2005), "*Irrigation Guidelines Based on Historical Weather Data in the Lower Rio Grande Valley of Texas*". Agric Water Management Vol. 76, pp. 1-7.

(7) Gijsman, A. J.; Hoogenboom, G.; Parton, W. J. and Kerridge, P. C. (2002), "*Modifying DSSAT Crop Models for Low-Input Agricultural Systems Using a Soil Organic Matter-Residue Module from CENTURY*". Agron. J. Vol. 94, pp.462-474.

(8) Idowu, O. S. and Gbuyiro, S. O. (2002), "*Agricultural Planning of Farm Operations in the Tropical Rainforest and Guinea Savanna Area of Nigeria*". Central Forecast Office, Federal Department of Meteorological Services, Lagos.

(9) Kiniry, J. R.; Williams, J. R.; Vanderlip, R. L.; Atwood, J. D.; Reicosky, D. C.; Mulliken, J.; Cox, W. J.; Mascagni, H. J.; Hollinger, S. E. and Wiebold, W. J. (1997) "*Evaluation of Two Models for Nine U.S. Locations*". Agron. J. Vol. 89, pp.421-426.

(10) Kobayashi, K. and Salam, M. U., (2000) "*Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components*". Agron. J. Vol. 92, pp.345-352.

(11) Michelson, D. B. (2004), "*Systematic Correction of Precipitation Gauge Observations using Analyzed Meteorological Variables*". J. of Hydrology Vol. 290, pp.161-177.

(12) Mott, P.; Sammis, T. W. and Southward, G. M. (1994), "*Climate Data Estimation Using Climate Information from Surrounding Climate Stations*". Applied Engrg. in Agric. Vol. 10 (1), pp.41-44.

(13) Retta, A.; Vanderlip, R. L.; Higgins, R. A. and Moshier, L. J. (1996) "*Application of SORKAM to Simulate Shattercane Growth using Forage Sorghum*". Agron. J. Vol. 88, pp.596-601.

(14) Richardson, C. W. (2000), "*Data Requirements for Estimation of Weather Generation Parameters*". ASAE Trans. Vol. 43(4), pp.877-882.

(15) Trewin, B. (2010), "*Exposure, Instrumentation and Observing Practice effects on Land Temperature Measurements*". WIRES Climate Change, John Wiley & Sons, Ltd. Doi: 10.1002/wcc.46.

(16) Whitmore, A. P. (1991): "*A Method for Assessing the Goodness of Computer-Simulation of Soil Processes*". J. of Soil Science. 1991 Vol. 42(2), pp.289-299.

(17) Wu, H.; Hubbard, K. G. and You, J. (2005) "*Some Concerns when Using Data from the Cooperative Weather Station Networks: A Nebraska Case Study*" Journal of Atmospheric and Oceanic Technology. Vol. 22, pp.592-6024.