J. Meteorol. Relat. Sci., 9:4
ISSN: 2412-3781
http://dx.doi.org/10.20987/jmrs.4.05.2016

In the special issue of the Journal of Meteorology and Related Sciences

Volume 9, Special issue
May, 2016

Investigating the Homogeneity of Monthly Rainfall Records in Kenya

Andang'o, Hezron Awiti.*, Jully O. Ouma**, Nzioka John Muthama* and Alfred Owuor Opere*

**University of Nairobi, Department of Meteorology, **IGAD Climate Prediction and Applications Centre,*

Corresponding Author:
Hezron, Andang'o Awiti,
University of Nairobi, Department of Meteorology,
P.O. Box 30197-00100, Nairobi
Email: awitihezron@gmail.com

(*Received 21 April 2016, in revised form 27 May 2016, Accepted 31 May 2016*)

ABSTRACT

Homogenization of climate data is of major importance because non-climatic factors make available data unrepresentative of the actual climate variation, and thus the conclusions of climatic and hydrological studies are potentially biased. A great deal of effort has been made to develop procedures to identify and remove non-climatic in-homogeneities. This paper first reviews several widely used statistical techniques then applies statistical simulation approach to precipitation data from different monitoring stations located in Kenya (1950-2006).

Analyses were carried out on several rainfall series in the 12 climatic zones of Kenya. The results of both the Standard Normal Homogeneity Tests (SNHT) and the Buishand Range Test (BR) tests show that, at the 5% significance level, the monthly series have statistically significant trend. Findings from the Standard Normal Homogeneity Test (SNHT) showed that all the monthly rainfall records from the selected synoptic stations were useful and hence could be used for any further analysis. From the Buishand Range (BR) Test done, seven out of the twelve stations were useful while the rest of the stations were doubtful. From the results of the Tests performed it is clear that the Buishand Range (BR) Test was able to detect breaks at the beginning middle and the end of the series. This method was thus recommended for homogeneity testing.

Promising results from the case study open new research perspectives on the homogenization of the Kenyan climate data time series.

Please cite this paper as: Andang'o, H. A., J. O. Ouma, N. J. Muthama and A. O. Opere, 2016. Investigating the Homogeneity of Monthly Rainfall Records in Kenya. *J. Meteorol. Relat. Sci. vol 9 pg 48-54.* http://dx.doi.org/10.20987/jmrs.4.05.2016

1. INTRODUCTION

Research works in Atmospheric Science and other disciplines require vast quantity of quality data in order to come out with better results in their research. Climatological information is widely used in processed form for legal, economical, agricultural and construction industries. (Aura et al. 2015, Muthama et al. 2012, Omeny et al. 2008)

Climate data can provide a great deal of information about the atmospheric environment that impacts almost all aspects of human endeavor. However, for these and other long-term climate analyses, particularly climate change analyses to be accurate, the climate data used must be homogeneous.

Climate data series are based on meteorological observations, following a set of rules, with regard to type of instruments, exposure, representativeness of station location and data recording procedures, amongst others. The history and evolution of observing networks show examples of a variety of changes, for instance, changes on instrument type, on instrument performance (calibration) and data procedures.

*H*omogeneous climate time series is defined as one where variations are caused only by variations in weather and climate (Aguilar *et al. 2003)* .Non-climatic factors may hide the true climatic signals and patterns, and thus potentially bias the conclusions of climate and hydrological studies. Frequent factors are monitoring stations relocations, changes in instrumentation, changes of the surroundings, instrumental inaccuracies, and changes of observational and calculation procedures. Unfortunately, most long-term climatological time series have been affected by a number of non-climatic factors that make these data unrepresentative of the actual climate variation occurring over time.

These factors include changes in: instruments, observing practices, station locations, formulae used to calculate means, and station environment *(Jones et al., 1985; Karl and Williams, 1987; Gullett et al., 990; Heino, 1994*).

Some changes cause sharp discontinuities while other changes, particularly change in the environment around the station, can cause gradual biases in the data. All of these inhomogeneities can cause bias in a time series and lead to misinterpretations of the studied climate.

Removal of these inhomogeneities is of importance or at least determines the possible error they may cause. Several techniques have been developed for the detection of non-climatic inhomogeneities. If the identified irregularities are due to non-climatic factors then adjustments are performed to compensate for the biases produced by the inhomogeneities. Since there is no one single best technique to be recommended, the following four steps are commonly followed (Aguilar et al. 2003); metadata analysis and basic quality control, creation of reference time series, breakpoint detection and data adjustment. Most of the procedures that have been proposed to identify and remove non-climatic inhomogeneities are not proper for immediate application on data with low temporal resolution (i.e., daily or hourly data).

Distribution of agricultural systems on the globe is largely a function of climate (i.e., the long-term average meteorological conditions that favor one kind of farming system over another). Agriculture (including the choice of crop varieties) is adapted to climate**.** At the same time farming at a given location is subject to the impact of year-to-year (inter-annual) variability in climate: the outcome of agriculture at a given site is affected by weather.

Climate variability can occur at different temporal and spatial scales. Short-term variability in rainfall onset may impact planting. Long-term protracted episodes of drought may impact the general water balance in a region (e.g. water for irrigation systems). Weather data and information are factors in the decision-making process that can be used to reduce uncertainty and improve economic and other decisions. Public emergency services, having received timely and accurate warnings about storms and floods can provide personnel and equipment to minimize damage. (Muthama et al. 2012, Aura et al.2015, Omeny et al. 2008). Economic value or benefit of the warnings, forecasts and climate information consists of the improvements in economic and related outcomes resulting from the use of these services.

The aim of this study was to investigate the homogeneity of monthly rainfall records in the 12 climatic zones of Kenya.

2. DATA AND METHODS

We used long term series of monthly precipitation totals (in millimeters) covering the 12 climatic zones of Kenya (Aura et al. 2015, Muthama et al. 2003, Omeny et al. 2008). Quantitative climate analyses require a good basis of reliable and consistent climate data. HoAnd

wever, several factors affect the quality of data and should be considered for any analyses (Sahin and Cigizoglu, 2010). The common methods of assessing the data quality are single and double-mass curves. Coaster and Soares (2009) hold that these methods are subjective and should only be used for experimental purposes without any scientific importance.

This study therefore applies much robust methods; the Standard Normal Homogeneity Test (SNHT) and the Pettit test. The duo have been found useful for testing the homogeneity of climate dataset (; Orlowsky, 2015; Kang and Yusof, 2012; Costa and Soares, 2009; Costa *et al.,* 2008; Wijngaard *et al.,* 2003 and Pettitt, 1979.

These techniques involve transforming the data to a value statistic that can be given a critical value subject to the dataset size. The Standard Normal Homogeneity Test is more sensitive to detect inhomogeneity near the beginning and the end of the dataset. According to this test, a statistic T_k is used to compare the mean of the first *y* years with the last (*n-y*) years and can be written as below;

$$T_k = k\bar{z}_1 + (n-k)\bar{z}_2, \qquad k = 1,2,3 \ldots n \ldots \ldots Equation\ 1$$

Where

$$\bar{z}_1 = \frac{1}{k}\frac{\sum_{i=1}^{k}(Y_i - \bar{Y})}{s} \ldots\ldots\ldots\ldots\ldots\ldots\ldots Equation\ 2$$

And

$$\bar{z}_2 = \frac{1}{n-k}\frac{\sum_{i=k+1}^{n}(Y_i - \bar{Y})}{s} \ldots\ldots\ldots\ldots Equation\ 3$$

In Equation 2, $\bar{Y}$ is the mean, *S* the standard deviation, Y_i the annual series to be tested and *k* the years of record with that of the last $(n-1)years$. The i^{th} year consists of a break if the value of *T* is maximum. To reject null hypothesis, the test statistic, $T_O = \max T_i$ is greater than the critical value, which depends on the sample size.

According to Pettitt (1979), the Pettitt test can be used to detect a single breakpoint in a time series.

The test is based on the rank, r_i of the Y_i and does not consider the normality of the series. The critical values given in table 1 can then be compared with the analyzed values.

Von Neumann ratio test uses the ratio of the successive mean square (year to year) difference to the variance (Costa and Sores, 2009; Kang and Yusof, 2012). When the sample is homogeneous, the expected value is two (N=2). The value of N is lower than 2 if there is a break in the sample. The Von Neumann ratio test is given by equation 4.

$$N = \frac{\sum_{i=1}^{n-1}(Y_i - Y_{i+1})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad \ldots\ldots\ldots Equation\ 4$$

Table 1: 1% and 5% critical values for X_k of the Pettitt test as a function of n.

N	20	30	40	50	70	100
1%	71	133	208	293	488	841
5%	57	107	167	235	393	677

3. RESULTS AND DISCUSSION

The observed rainfall datasets for the representative station for each of the 12 climatic zones in Kenya that were subjected to both the Standard Normal Homogeneity Test homogeneity test and the Pettit. Although these tests have many characteristics in common, the results obtained indicated few differences. The Standard Normal Homogeneity Test detected breaks near the beginning and the end of the data series, whereas the Pettit test detected a single breakpoint in a time series.

Both tests assumed that the annual series being tested were normally distributed.

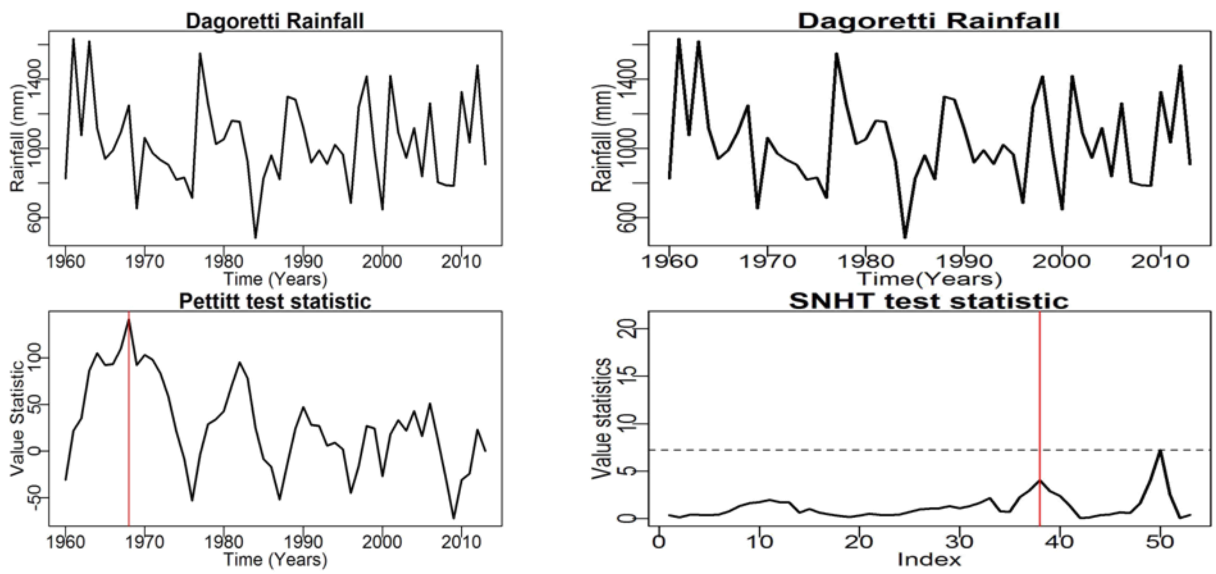
The results include graphical plots and the tables for the tests performed.Two significant levels were used in these tests i.e. 1% significant level (p1), 5% significant level (p5) while NS means Not Significant in Table 2.

The outcome of the test were characterised into either "useful", "doubtful" and "suspect" depending on the number of rejected null hypothesis which state that, the annual values Y_i of the testing variable Y are identically distributed, independent and it's homogeneous at 1% significant level. The results was classified as useful if it rejected one or none null hypothesis under the four tests, it was then considered as homogeneous and can be used for further analysis. If the series reject the two null hypotheses of the four tests, it was then considered doubtful and was inspected before further analysis. A data series was considered suspect if it rejects three or the four null hypotheses and therefore was not considered for further analysis.

Table 2: Results summary of the data

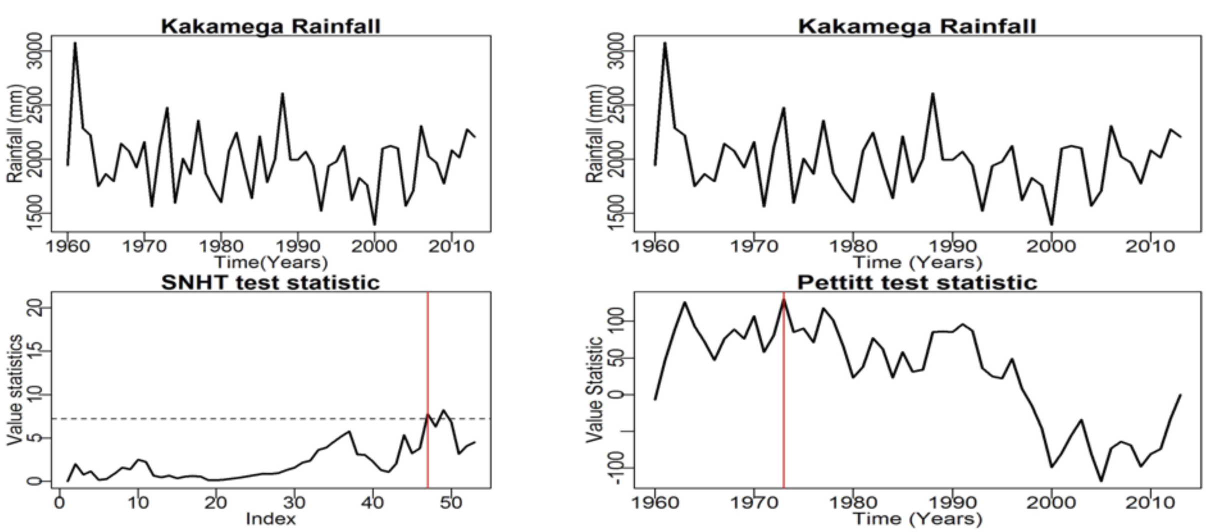
Test	Lodwar	Kakamega	Dagoretti	Ki-sumu	Narok	Wajir	Garissa	Makindu	Mombasa	Lamu
SNH	NS	NS	NS	NS	p5	NS	NS	NS	NS	NS
BHR	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
PET	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
VON	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Break										
SNH	51	4	4	2	4	9	2	42	48	2
BHR	24	5	10	10	21	11	10	43	41	38
PET	20	14	9	34	20	10	9	42	40	9
VON	1.96	1.9	1.98	1.88	1.66	2.05	1.67	1.73	1.73	2.15
Class	useful	useful	useful	useful	useful	useful	useful	useful	useful	useful

The following were some of results of the test statistics of annual mean for the Pettit test and the Standard Normal Homogeneity Test of the various stations



(a) **(b)**

Figure 1: Test statistics of annual mean for the (a) Pettit test and (b) the Standard Normal Homogeneity Test for Dagoretti station.

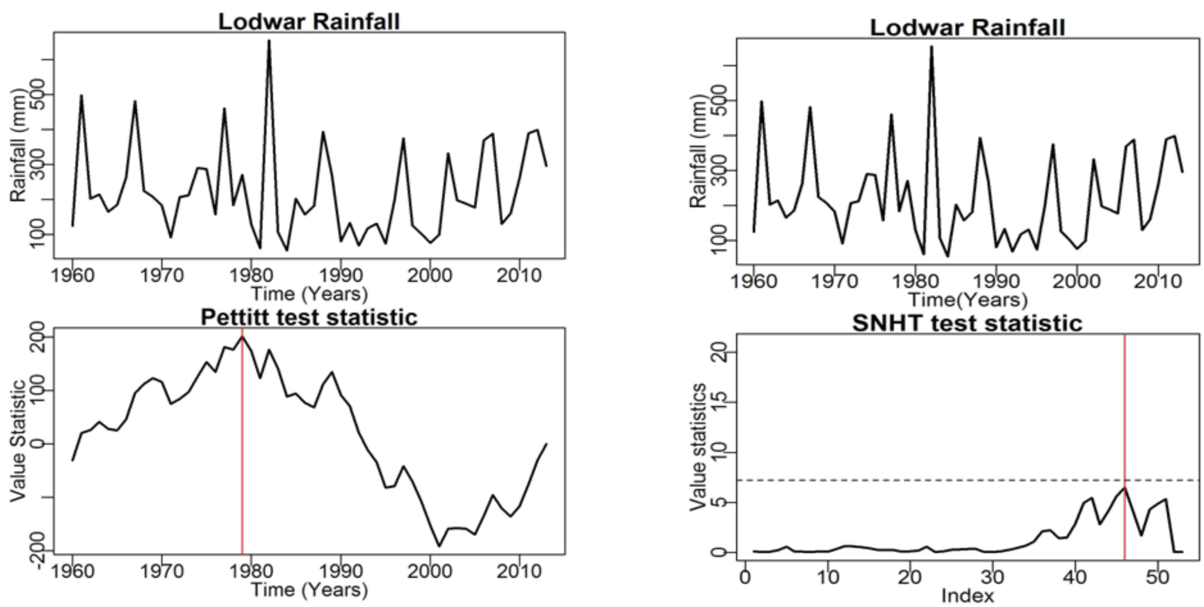


(a) **(b)**

Figure 2: Test statistics of annual mean for (a) the Pettit test and (b) the Standard Normal Homogeneity Test for Kakamega station.

The two methods involve transforming the data to a value statistic that can be assigned a critical value. From Figure 1, it is evident that SNHT is more sensitive to detect inhomogeneity near the beginning and the end of the dataset. Petit test on the other hand detected a breakpoint towards the middle of the time series.

In Figure 2 the results obtained for the Kakamega station indicate the test's as valuable tools for inhomogeneities detection in climate time series with several breakpoints detected. Thus, their application for operational purposes within the National Meteorological Services in the GHA, or elsewhere and for external users may be recommended.



(a) **(b)**

Figure 3: Test statistics of annual mean for (a) the Pettit test and (b) the Standard Normal Homogeneity Test for Lodwar station.

The results of the homogeneity test of the Lodwar rainfall station in Figure 3 (a) and (b) above confirm the Standard Normal Homogeneity Test's sensitivity to detect inhomogeneity near the beginning and the end of the dataset and Petit test's detection of a single breakpoint in a time series.

4. CONCLUSIONS

Homogeneity of the daily rainfall series was detected successfully by using annual mean, annual maximum and median as the testing variables. The results were assessed by classifying the stations into 3 categories, which are useful, doubtful and suspect.

The results showed that some of the stations in Kenya are homogeneous when annual mean is used as testing variables.

History of the station relocation, observing practices and instruments used are important in analyzing the homogeneity of the stations. Unfortunately, these data are not available in the stations studied. Therefore it does not have evidence to evaluate the breaks detected and correct the series.

The methodology applied gave a better result compared to the results of the mass curves normally applied in inhomogeneity detection.

Hence it is clear that the use of the mass curve is weaker method compared to the Statistical Methods as it is unable to detect breaks in the middle of the data series. The methods therefore proved to be more rigorous and hence recommended for homogeneity testing and analysis.

REFERENCES

Alexandersson H. (1986): A Homogeneity Test Applied to Precipitation Test. J.Climatol., 6, 661 -675.

Aura S., S. Kahuha, B. Chanzu., N. J. Muthama., F. K. Karanja (2015): Making Meteorological Services more Beneficial to Farmers. WMO Bulletin 64(1).

Brunetti M., M. Maugeri, F. Monti, T. Nanni (2006): The Variability and Change of Italian Climate in the Past 160 Years. IL Nuovo Cimento, 29 C, 3-12.

Buishand T.A. (1986): Some Methods for Testing the Homogeneity of Rainfall Records. J. Hydrol., 58, 11-27.

Costa A. C., Negreiros J, Soares A. (2008): Identification of inhomogeneities in precipitation time series using stochastic simulation. Geostatistics for Environmental Applications. Springer, Berlin, pp 275–282

Costa A. C., A. Soares (1986): Homogenization of Climate Data: Review and New Perspectives Using Geostatistics. Math Geosc 41, 291-305.

Deni S.M., A. A. Jemain, K. Ibrahim (2008): The Spatial Distribution of Wet and Dry Spells over Peninsular Malaysia. Theor. Appl. Climatol., 94, 163-173

Khaliq M. N. and T. B. M. J. Ouarda (2007): On the Critical Values of the Standard Normal

Homogeneity Test (SNHT). Int. J. Climatol, 27, 681-687.

Menne M.J. and C.N.Williams (2005): Detection of undocumented change points using multiple test statistics and composite reference series. J Climate 18(20):4271–4286

Muthama N. J., A. O. Opere and C.B. Lukorito (2003): Utilization of Meteorological products in agriculture and water sectors in Central and Eastern Kenya. J. Afric.Met.Soc. Vol. 6. No. 1, pp 58-64.

Muthama N. J., Masieyi W. B., Okoola R.E., Opere A. O., Mukabana J. R., Nyakwada W., Aura S.,, Chanzu B. A., and Manene M. M., 2012. Survey on the Utilization of Weather Information and Products for Selected Districts in Kenya. *J. Meteorol. Relat. Sci*., vol 6 pg 51 –58. http://dx.doi.org/10.20987/jmrs.08.2012.605

Omeny P. A, Ogallo L. A., Okoola R. Hendon H., and Wheeler M., 2008. East African Rainfall Variability Associated with the Madden-Julian Oscillation. *J. Kenya Meteorol. Soc*., vol 2 pg 105–114

Stepanek P., Zahradnicek P. and Skalak P. (2009): Data Quality Control and Homogenization of Air Temperature and Precipitation Series in the Area of the Czech Republic in the period of 1961-2007. Advances in Science and Research, 3, 23-26.

Wijngaard J. B., A. M. G. Kleink Tank and G. P. Konnen (2003): Homogeneity of 20th Century European Daily Temperature and Precipitation Series. Int. J. Climatol, 23, 679-692. 14(2001), 964-978.