Asian Journal of Physical and Chemical Sciences
1(2): 1-9, 2016; Article no.AJOPACS.31545

SCIENCEDOMAIN *international*
www.sciencedomain.org

Predictor Selection Associated With Statistical Downscaling of Precipitation over Zambia

Libanda Brigadier[1*], David Allan[2], Banda Noel[3], Wang Luo[2], Ngonga Chilekana[4] and Linda Nyasa[5]

[1]School of Civil Engineering and Geosciences, Hydrology and Climate Change, Newcastle University, United Kingdom.
[2]School of Geosciences, The University of Edinburgh, United Kingdom.
[3]Department of Climate Change and Meteorological Services, Malawi.
[4]Energy and Water Development, Lusaka, Zambia.
[5]Natural Resources and Environmental Protection, Lusaka, Zambia.

Authors' contributions

This work was carried out in collaboration between all authors. Author LB designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors DA, BN, WL, NC and LN managed the analyses of the study and the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJOPACS/2016/31545
Editor(s):
(1) Thomas F. George, Chancellor / Professor of Chemistry and Physics, University of Missouri- St. Louis One University Boulevard St. Louis, USA.
Reviewers:
(1) Nurul Nadrah Aqilah Binti Tukimat, Universiti Malaysia Pahang, Malaysia.
(2) Charles Bwalya Chisanga, University of Zambia, Zambia.
(3) Marisa Cogliati, Universidad Nacional del Comahue, Argentina.
Complete Peer review History: http://www.sciencedomain.org/review-history/17956

Original Research Article

Received 12th January 2017
Accepted 18th February 2017
Published 24th February 2017

ABSTRACT

A non-generative, analog methodology was used to downscale daily precipitation from CMIP5-CNRM-CM5 developed by Météo-France/CNRS and CMIP5-CANESM2 of the Canadian Centre for Climate Modelling and Analysis. The downscaling reduces the 2° resolution GCM output to point station data. Sensitivity experiments for four different predictor variables (PVs) were carried out to examine the most significant PVs for the case of Zambia. ERA-Interim reanalyses was used for calibration (75%) and validation (25%) for the period 1981 – 2012. The Root Mean Square Error (RMSE) was used to compute the predictive power of CNRM-CM5 and CANESM2 by comparing the difference between their simulation results against ERA-Interim. Pearson correlation coefficient

**Corresponding author: E-mail: B.Libanda2@newcastle.ac.uk;*

(r) was also used to assess the linear relationship between the datasets. Downscaled and observed data were compared and analysed. Results indicate that both CNRM-CM5 and CANESM2 perform well in perfect prognosis over the period 1970 – 2000 averaged over longitude 19°E - 37°E and latitude 22°S - 4°S. Pearson correlation results show that the combination PV2: T850, Q850, and U850 perform well at 95% confidence level. These results fill the knowledge gap of the behaviour of different variables for climate change projections and impact assessment studies in Zambia. Specifically, this study suggests a starting point in the selection of predictor variables for climate change studies in Zambia.

Keywords: Predictor selection; statistical downscaling; precipitation; Zambia.

1. INTRODUCTION

Zambia is a southern African country [1] boarded by 8°S and 18°S, and 21.8°E and 34°E [2]. The country experiences a tropical climate [3] with precipitation patterns mainly influenced by the movements of the Inter-Tropical Convergence Zone (ITCZ) [4,5,6]. The ITCZ is a zone of convergence between the Northeast and Southeast Trade winds [7]. The trekking of the ITCZ begins to affect Zambia from the end of October in an ill-defined manner [8,9] until during the period December – February (DJF) when Zambia experiences much of her rains [10]. Investigating the amount and distribution of rainfall over Zambia is central to the country's management of water resources [11].

However, currently, products of General Circulation Models (GCMs) are too coarse [11-17] to be used in predicting rainfall patterns at local level. Statistical downscaling is therefore used as a technique that bridges the gap between GCM outputs and products useful for impact studies. This is done through the establishment of statistical relationships between local climate variables, in this case precipitation, and large-scale predictors. The developed relationships are applied to the outputs of global climate model experiments [18]. Statistical downscaling approaches have thus been used in many studies to produce reliable station-based rainfall data with varied assignments of statistical confidence. For example [19] used statistical downscaling in a precipitation study over Campbell River basin, British Columbia, Canada. [20] also used a stochastic weather generator to statistically downscale precipitation in a study on seasonal precipitation and temperature forecasts produced by the International Research Institute for Climate and Society (IRI). Several studies [21,22] show that at the core of statistical downscaling methods lies the establishment of statistical relationships between large-scale atmospheric variables (Predictors) and local-scale variables (Response). The dynamics of local meteorological variables are then projected by using large scale information at the local level [idem, 21].

Despite these advances in statistical downscaling methods, the use of statistically downscaled data in understanding the variability of rainfall in Zambia is still at its infancy. Precipitation has higher variability as compared to temperature. It is therefore more difficult to project than temperature. For this reason, caution needs to be attached to all factors related to it's projection. The objective of this paper is to show the behaviour of different Predictor Variables (PVs) in describing rainfall patterns over Zambia. Results documented herein will be useful in studies relating to precipitation projection over Zambia.

2. DATA AND METHODOLOGY

2.1 Response Variables

In the present study, Response is used after [22] to denote the use of the dependant precipitation variable as observed by thirty nine Meteorological stations (Fig. 1) sourced from the Zambia Meteorological Department and predicted by independent variables (PVs) shown in Table 2.

The list of stations with the exact longitudes and latitudes used in this study is given in Table 1.

2.2 Predictor Variables

[idem, 23,24,25] have all suggested the use of moisture, wind, and temperature as key PVs. The choice of predictors was based on large-scale circulations simulated by CNRM-CM5 and CMIP5-CANESM2. To understand the most significant PVs in downscaling precipitation over Zambia, temperature, humidity, Zonal wind, and

Geo-potential were put in four different class combinations (Table 2) and tested for sensitivity over longitude 19°E - 37°E and latitude 22°S – 4S° (Zambia).

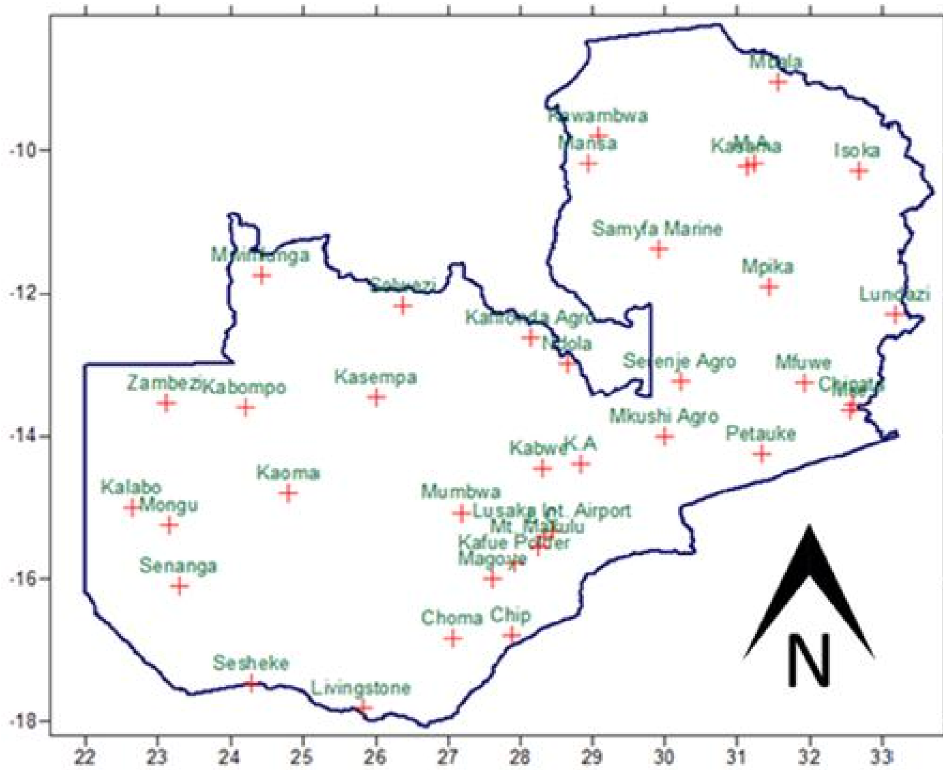

Fig. 1. Map of Zambia showing stations used in the study

Table 1. Geographical location of meteorological stations used in this study

Name	Latitude (°S)	Longitude(°E)
Kawambwa	9.793	29.076
Mbala	9.028	31.553
Mwinilunga	11.74	24.431
Mansa	10.173	28.942
Samyfa Marine	11.371	29.911
Kasama	10.224	31.14
Misamfu Agro	10.171	31.225
Mpika	11.901	31.433
Isoka	10.272	32.68
Zambezi	13.534	23.108
Kasempa	13.457	26
Kabompo	13.596	24.208
Solwezi	12.171	26.367
Ndola	12.994	28.659
Kafironda Agro	12.614	28.148
Serenje Agro	13.227	30.215
Mkushi Agro	14	29.996
Msekera	13.646	32.563
Chipata	13.564	32.589
Lundazi	12.294	33.175
Mfuwe	13.255	31.931

Name	Latitude (°S)	Longitude(°E)
Kalabo	14.997	22.646
Mongu	15.254	23.151
Kaoma	14.795	24.804
Mumbwa	15.075	27.189
Kafue Polder	15.777	27.921
Kabwe Agro	14.395	28.828
Kabwe Main	14.448	28.302
Lusaka Int. Airport	15.324	28.448
Lusaka City	15.417	28.321
Mt. Makulu	15.548	28.248
Petauke	14.251	31.339
Senanga	16.111	23.298
Sesheke	17.477	24.301
Livingstone	17.823	25.82
Magoye	15.998	27.617
Choma	16.838	27.07
Chipepo	16.795	27.879

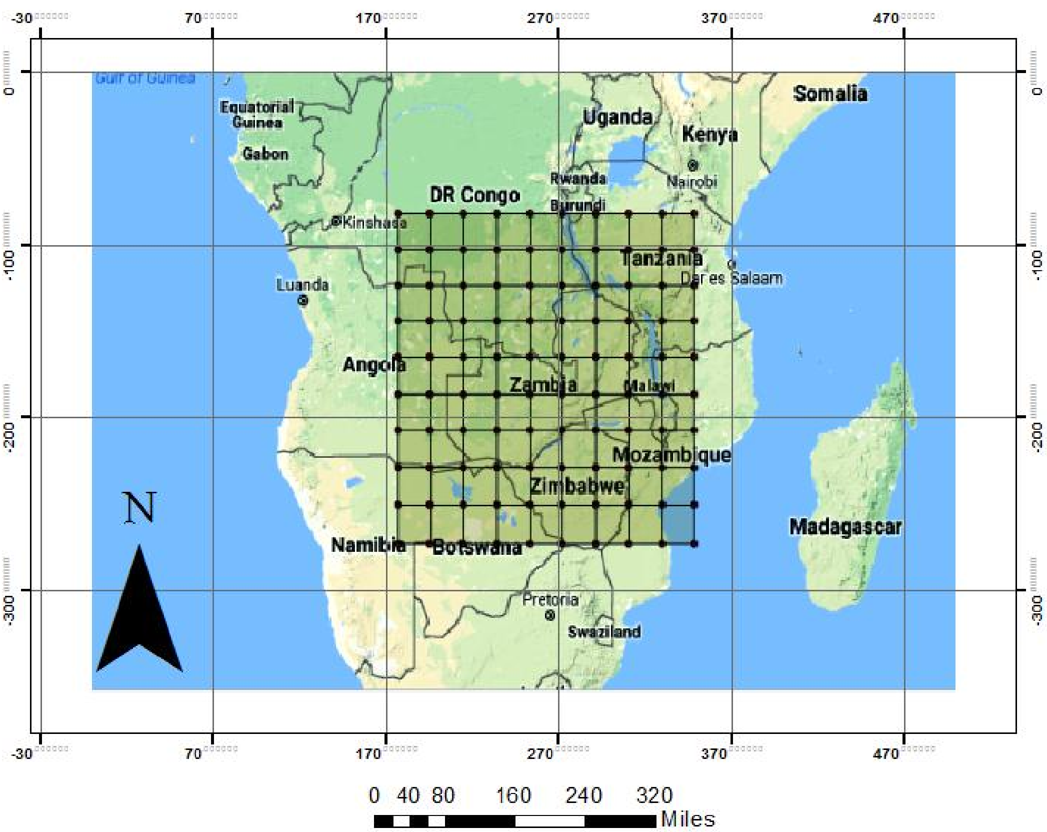

Fig. 2. Study domain averaged over longitude 19°E - 37°E and latitude 22°S - 4°S

2.3 Downscaling Approach

The downscaling approach employed in this study is based on an analog statistical downscaling method. This technique associates the simulations of large-scale synoptic circulations with local variables in check with station observation datasets. ERA-Interim reanalyses is used for calibration (75%) and validation (25%) for the period 1981 – 2012.

Fig. 2 shows the domain of the study. The training (calibration) is based on 75% of the period in order to use the remaining 25% to ascertain (test/validate) that the calibration was correctly done. Studies that have used the analog technique include [26].

Table 2. Predictor Variables (PVs) used in this study

	PV	Level
	Temperature	T700
PV1	Q humidity	Q700
	U ZonalWind	U850
	Temperature	T850
PV2	Q humidity	Q850
	U ZonalWind	U850
	Temperature	T850
	Temperature	T700
PV3	Q humidity	Q850
	Q humidity	Q700
	U ZonalWind	U850
	GPM	Z850
	Temperature	T850
	Temperature	T700
PV4	Q humidity	Q850
	Q humidity	Q700
	U Wind	U850

2.4 Statistical Analysis

2.4.1 Root mean square error (RMSE)

In order to understand the ability of CMIP5-CNRM-CM5 and CMIP5-CANESM2 to reproduce precipitation trends over Zambia, Root Mean Square Error (RMSE) was used. RMSE gives a measure of the departure of predicted data from observed. Many studies have used RMSE in precipitation studies over different regions [27].

The RMSE of a model's prediction with respect to a given PV X_{model} is defined:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \quad (1)$$

Where X_{obs} is observed values and X_{model} is modelled values at time/place i.

2.4.2 Pearson correlation coefficient (r)

Pearson's correlation (r) was used to scrutinize the strength of the linear relationship (if any) between GCMs and ERA-Interim. The Pearson correlation is obtained by dividing the covariance of the two variables by the product of their standard deviations and is thus given in Eq. 2 below:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (2)$$

Pearson gives 1 in the case of a perfect upward linear relationship, and -1 in case of a downward linear relationship, and any values in between indicate the magnitude of relationship between GCMs and ERA-Interim. A correlation coefficient of 0 means no linear relationship exists.

3. RESULTS AND DISCUSSION

3.1 Annual Cycle and Interannual Variability

Outputs of CMIP5-CNRM-CM5 and CMIP5-CANESM2 were investigated in comparison to observed station data (1970 – 2000) to gauge the strength of the Models and hence their usefulness in consequent analysis. Results (Fig. 3) show that both CNRM-CM5.1 and CANESM2 are able to capture the annual cycle of the rain season over Zambia. The rainy season which begins in October (transitional month) and goes on to April (transitional month) was well represented by both models. Similarly the dry season from May to September was captured.

Apart from the amount of rainfall received, spatial distribution is key to the agricultural and water resources sectors of Zambia. Therefore, this study also investigated the ability of CNRM-CM5 and CANESM2 to reproduce the distribution in comparison to observed data over Zambia. Results (Fig. 4) show that generally both GCMs are able to capture the downward gradient from the North to the south of the country. However, in comparison to observed data (a), CNRM-CM5 (b) performs better than CANESM2 (c) which seems to be observing much lower precipitation over Chipata and parts of Eastern Province. Standardized anomaly of these results is given in Fig. 5.

3.2 RMSE

The Fig. 6 of the root mean square error (RMSE) show that PV2 i.e. T850, Q850, and U850 performs better, with an RMSE value of 4.1, than

PV1, PV3, and PV4 which gave 4.3, 4.5, and 4.6 respectively. These results also suggest that PV1 i.e. T700, Q700, and U850 is the closest combination to being able to predict precipitation over Zambia. This analysis also revealed that an increase in the number of PVs reduced the predictive strength of the models. For example, PV4 which included geo-potential at the 850 level in addition to T850, Q850, and U850 performed poorly.

3.3 Pearson Correlation Coefficient (r)

Table 3 gives the Pearson correlation results (r) which shows that the calibration and validation of the statistical downscaling approach was correctly done with all PVs giving an r value of over 0.8. These results have also complimented the findings of RMSE with PV2 being the best correlated combination. It is therefore the finding of this experiment that for optimum results in downscaling precipitation over Zambia in perfect prognosis using analog method, the most informative predictors are: T850, Q850, and U850.

Table 3. Pearson correlation (r) of GCMs with ERA-interim

	PV1	PV2	PV3	PV4
CNRM	0.81	0.84	0.81	0.82
CANES2	0.8	0.83	0.8	0.81

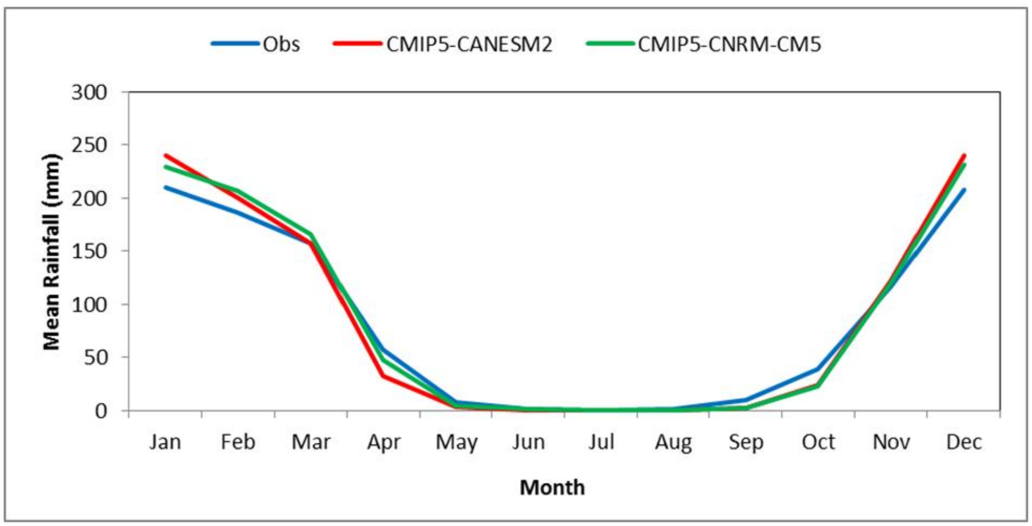

Fig. 3. Mean annual cycle of rainfall (mm) over Zambia for station data (blue), CANESM2 (red), and CNRM-CM5 (green), averaged over longitudes 21E–34E and latitudes 190S–7.5S for the period 1970 – 2000

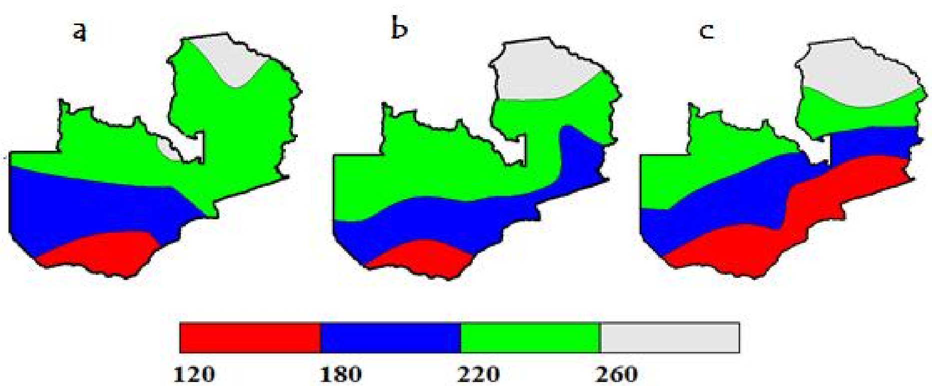

Fig. 4. Monthly mean DJF rainfall (mm) during the period 1970 – 2000 over Zambia with (a) being observed, (b) CNRM-CM5 and (c) CANESM2

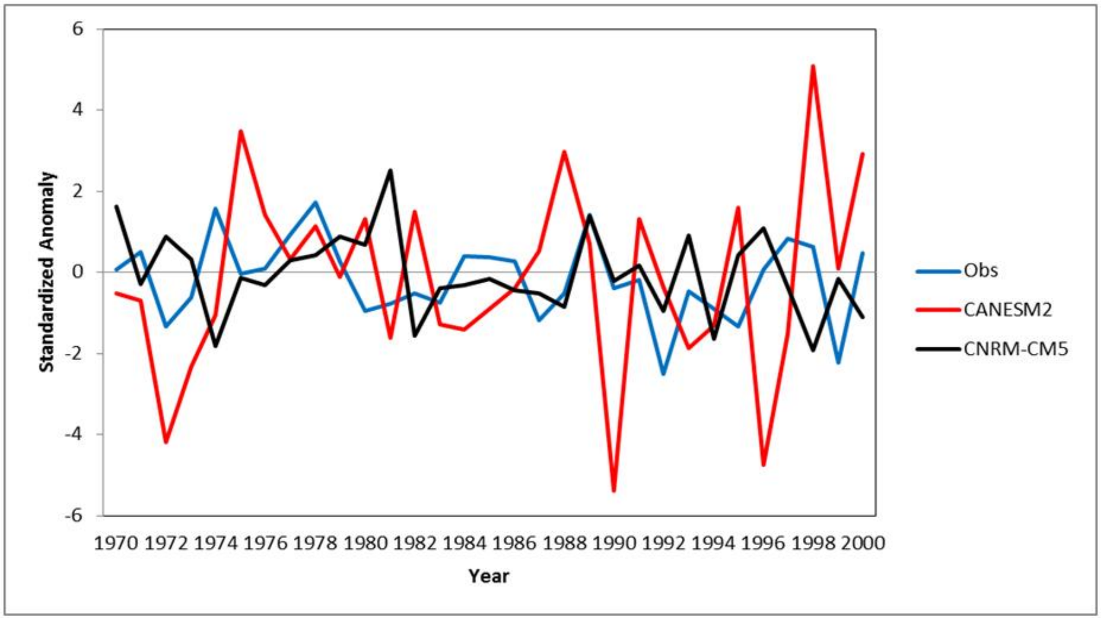

Fig. 5. Interannual variability of the mean DJF rainfall over Zambia, averaged over the period 1970 – 2000

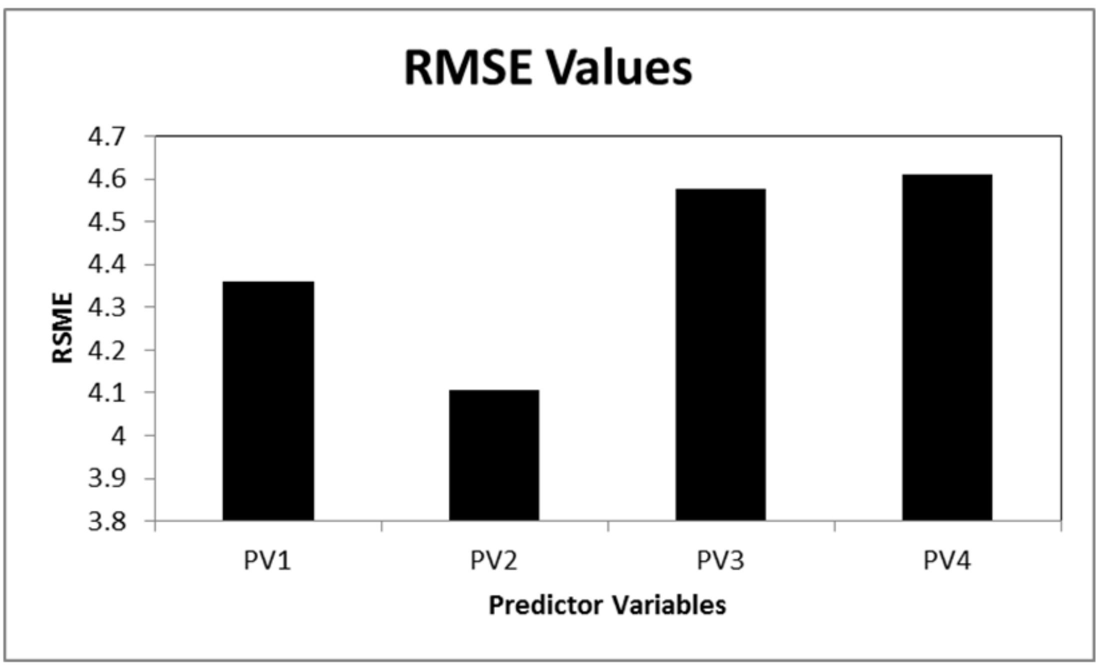

Fig. 6. Root mean square error results for mean rainfall during the period 1970 – 2000 averaged over 19°E - 37°E and latitude 22°S - 4°S

4. CONCLUSION

The analysis was carried out in order to identify suitable predictors associated with statistical downscaling of precipitation in Zambia. CNRM-CM5 and CANESM2 were trained (calibrated) and tested (validated) using ERA-Interim reanalyses for the period 1981 – 2012. The downscaled products of CNRM-CM5 and CANESM2 were then compared to observed precipitation data. Results showed that both CNRM-CM5 and CANESM2 performed well in perfect prognosis over the period 1970 – 2000 averaged over longitude 19°E - 37°E and latitude 22°S - 4°S. However, in comparison to observed data CNRM-CM5 captured the trend of precipitation better than CANESM2 which seemed to be observing much lower precipitation over Chipata and parts of Eastern Province.

Sensitivity experiments for four different predictor variables (PVs) were carried out to examine the most significant PVs for the case of Zambia. A combination of PV2: T850, Q850, and U850 performed better than any other combinations at 95% confidence level. These results show that CNRM-CM5 and CANESM2 are useful for climate change studies over Zambia. Additionally, PV2 has shown to be the most significant PV when using the analog approach to statistically downscale precipitation over Zambia. These results will be useful in forth coming research lines relating to precipitation projection over Zambia.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Nkolola B, Linda N, Ngonga C.'Drivers of deforestation in the Miombo Woodlands and their impacts on the environment. Advances in Research. 2016;6(5):1–7.
2. Libanda B, Ogwang BA, Ongoma V, Ngonga C, Nyasa L. Diagnosis of the 2010 DJF flood over Zambia. Natural Hazards. 2015;81(1):189–201
3. Lal R, Cummings DJ. 'Clearing a tropical forest I. Effects on soil and micro-climate. Field Crops Research. 1979;2: 107–91.
4. Libanda B, Nkolola B, Musonda B. 'Rainfall variability over northern Zambia. Journal of Scientific Research & Reports. 2015;6(6):416–425.
5. Baumberg V, Weber T, Helmschrot J. Assessing the change in rainfall characteristics and trends for the southern Afr', EGU General Assembly Conference Abstracts. 2015;17. Available:http://adsabs.harvard.edu/abs/2015EGUGA..17.1434B
6. Klutse NAB, Abiodun BJ, Hewitson BC, Gutowski WJ, Tadross MA. Evaluation of two GCMs in simulating rainfall inter-annual variability over southern Africa', Theoretical and Applied Climatology. 2015;123(3):415–436
7. Libanda J, Nkolola B, Linda N. Economic significance of agriculture for poverty reduction: The case of Zambia' Archives of Current Research International. 2016;5(2): 1–9.
8. Shongwe ME, van Oldenborgh GJ, Van den Hurk BJJM, De Boer B, Coelho CAS. Van Aalst MK. 'Projected changes in mean and extreme Precipitation in Africa under global warming. Part I: Southern Africa'. Journal of Climate. 2009;22(13):3819–3837.
9. Tyson PD. Climatic change in southern Africa: Past and present conditions and possible future scenarios. Climatic Change. 1991;18(2-3):241–258.
10. Banda, Titus D. Drainage system and storm water management for Ndola Central Business Area, Zambia; 2016. Available: https://lup.lub.lu.se/student-
11. Hachigonta S, Reason C. Interannual variability in dry and wet spell characteristics over Zambia. Climate Research. 2006;32:49–62.
12. Turco M, Hardenberg, Provenzale A. 'Comparison of bias correction methods for the RegCM4-ICTP daily precipitation si', EGU General Assembly Conference Abstracts. 2015;17. Available:http://adsabs.harvard.edu/abs/2015EGUGA..1711016T
13. Schmatz DR, Luterbacher J, Zimmermann NE, Pearman PB. 'Gridded climate data from 5 GCMs of the last glacial maximum downscaled to 30 arc s for Europe', Climate of the Past Discussions. 2015; 11(3):2585–2613
14. Wiebe K, Lotze-Campen H, Sands R, Tabeau A, Van der Mensbrugghe D, Biewald A, Bodirsky B, Islam S, Kavallari A, Mason-D'Croz D, Müller C, Popp A, Robertson R, Robinson S, van Meijl H, Willenbockel D. Climate change impacts on agriculture in 2050 under a range of plausible socioeconomic and emissions scenarios', Environmental Research Letters. 2015;10(8):085010.
15. Nungesser M, Saunders C, Coronado-Molina C, Obeysekera J, Johnson J, McVoy C, Benscoter B. 'Potential effects of climate change on Florida's Everglades'. Environmental Management. 2014;55(4): 824–835.

16. Hill RA, Hawkins CP, Jin J. Predicting thermal vulnerability of stream and river ecosystems to climate change. Climatic Change. 2014;125(3-4):399–412.
17. Seo GH, Cho YK, Choi BJ, Kim KY, Kim B, Tak Y. Climate change projection in the northwest pacific marginal seas through dynamic downscaling. Journal of Geophysical Research: Oceans. 2014; 119(6):3497–3516.
18. Zhao N, Yue T, Zhou X, Zhao M, Liu Y, Du Z, Zhang L. Statistical downscaling of precipitation using local regression and high accuracy surface modeling method. Theoretical and Applied Climatology; 2016;1–12.

    DOI: 10.1007/s00704-016-1776-z
19. Mandal S, Srivastav RK, Simonovic SP. Use of beta regression for statistical downscaling of precipitation in the Campbell river basin, British Columbia, Canada. Journal of Hydrology. 2016;538: 49–62.
20. Kim Y, Rajagopalan B, Lee G. Temporal statistical downscaling of precipitation and temperature forecasts using a stochastic weather generator. Advances in Atmospheric Sciences. 2015;33(2):175–183.
21. Pour SH, Shahid S, Chung ES. A hybrid model for statistical Downscaling of daily rainfall. Procedia Engineering. 2017;154: 1424–1430.
22. Vaittinada P, Vrac M, Bastin S, Carreau J, Déqué M, Gallardo C. Intercomparison of statistical and dynamical downscaling models under the Euro- and Med-Cordex initiative framework: Present climate evaluations. Climate Dynamics. 2015; 46(3-4):1301–1329.
23. Li L, Schmitt RW, Ummenhofer CC, Karnauskas KB. North atlantic salinity as a predictor of Sahel rainfall. Science Advances. 2016;2(5):e1501588-e1501588.
24. Zorita E, von Storch H. The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. J. Clim. 1999;12: 2474–2489.
25. Sudheer KP, Gosain AK, Ramasastri KS. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrological Processes. 2002; 16(6):1325–1330.
26. Charles A, Timbal B, Fernandez E, Hendon H. ‘Analog downscaling of seasonal rainfall forecasts in the Murray darling basin. Monthly Weather Review. 2013;141(3):1099–1117.
27. Du J, Mullen SL, Sanders F. ‘Short-range ensemble forecasting of quantitative Precipitation. Monthly Weather Review. 1997;125(10):2427–2459.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/17956