On the combination of rotated principal component analysis regionalization technique and linear regression in seasonal rainfall prediction

Chibuike Chiedozie Ibebuchi (chibuike.ibebuchi@uni-wuerzburg.de)
Julius Maximilians University Würzburg: Julius-Maximilians-Universitat Wurzburg
https://orcid.org/0000-0001-6010-2330

Original Article

Keywords: regionalization, seasonal rainfall prediction, rotated principal component analysis, Africa south of the equator, linear regression, ocean indices

Posted Date: February 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-164806/v1

Abstract

This study considers the selection of predictors for regional rainfall based on dynamical considerations; for this reason, a regionalization technique that can preserve the underlying physics of rainfall was used in obtaining landmasses and local oceanic domains that are spatially coherent. For the study region (Africa, south of the equator), the adjacent oceans play a vital role in the seasonal rainfall variability at the landmasses; thus uncovering the complex nature of the multivariate relationship between rainfall coherent landmasses and local oceanic domains will enhance the construction of oceanic indices as predictors of seasonal rainfall at specific landmasses using linear regression analysis. Among different cluster analysis techniques, the rotated principal component analysis (PCA) is both fuzzy and allows the overlapping of the classified data set, which makes it a better choice for geophysical research that aims to regionalize continuous data such as rainfall. 10 regions with spatially homogeneous austral summer monthly rainfall totals were classified using the rotated PCA; some classified regions featured landmasses that are spatially coherent with the adjacent ocean, which qualifies them to be further considered on how rainfall anomaly and other physical parameters related to rainfall (e.g. convergence, relative vorticity, and sea level pressure) at the adjacent oceans explain the variations in austral summer rainfall anomaly at the homogeneous landmasses. The analysis of the physical mechanisms associated with the time development of the selected rainfall regions reveals that at the west-central equatorial rainfall region, variations in relative vorticity and convergence are associated with the development of the rainfall region; whereas at the central domains of southern Africa, variations in the patterns of sea level pressure, relative vorticity and convergence at the landmasses, the tropical and the southwest Indian Ocean can be associated with the development of the distinct rainfall sub-regions. The predictability of austral summer rainfall anomaly at the homogeneous landmasses using appropriate predictors at the adjacent local oceanic domains was relatively more accurate at the deep tropics, possibly due to the dominating mechanism of convergence in controlling the tropical rainfall.

1 Introduction

Southern Africa is located between three Oceans, namely, the Southern Ocean, the south Indian Ocean, and the South Atlantic Ocean. The adjacent oceans act as moisture sources to the landmasses, with sea surface temperature (SST) anomaly at the southwest Indian Ocean playing a significant role in the seasonal variability of rainfall in southern Africa (Reason and Mulenga 1999). Evaporation at the Agulhas current modulates rainfall in southern Africa through its influences on the South Indian Ocean Convergence Zone (Cook 2000) which is the major large-scale system that influences the regional rainfall variability in southern Africa. Rapolaki et al. (2020) noted that local oceanic regions surrounding southern African landmasses act as moisture sources in specific seasons. The predictability of rainfall at southern African landmasses, using the statistical approach of linear regression, might be enhanced by uncovering land and maritime domains with homogeneous rainfall regions; such regionalization approach helps in further understanding of the underlying physics (i.e. the causal mechanism) through which rainfall

anomalies at local oceanic domains are directly related to rainfall variability at specific landmasses in southern Africa.

In geophysical research, regionalization involves the grouping of stations (or grid points) into climatologically homogeneous regions with respect to a given input climatic field, for example, precipitation (Johnson and Green 2018); temperature (Yu et al. 2018); surface pressure (Vargas and Compagnucci 1987) among others. Regionalization of climatic data is commonly achieved using cluster analysis, which can be hierarchical or non-hierarchical; or by using rotated principal component analysis (PCA) (Richman and Lamb 1985). In hierarchical clustering, a sequence of groups of patterns is created from a similarity matrix commonly based on the Euclidean distance. In non-hierarchical clustering a set of seed points (i.e. centroids of the clusters) are used as initial conditions for assigning variables to the nearest centroid; a major difference is that in non-hierarchical clustering the number of regions is determined before the clustering process or as a part of the clustering process. Using observed rainfall data and global climate models, Badr et al. (2016) applied the hierarchical clustering technique to regionalize rainfall in Africa. In the rotated PCA method, an orthogonal or oblique rotation of the retained components makes the procedure go beyond variance maximization and a data (dimension) reduction statistical tool, to a technique that holds together, clusters of highly related variables.

Gong and Richman (1995) made a comprehensive comparison of the quality of precipitation regions in North America East of the Rockies, achieved from the cluster analysis techniques based on some selected hierarchical and non-hierarchical clustering algorithms, and the rotated PCA technique; they concluded that the choice of the regionalization technique should be based on the research goal and most importantly on the nature of the real-world data. They strongly recommended that in an applied geophysical research problem, that the physical or causal mechanism associated with the meteorological parameter should first be understood before the decision of a regionalization technique is made, this will enable the underlying physics in the parameter to be preserved, and since the causal mechanism for a meteorological parameter such as precipitation is continuous, then a fuzzy classification technique that allows overlapping of the classified variables, such as rotated PCA, applied at large sample size, yields a physically meaningful result, compared to other clustering techniques. Unlike the rotated PCA, cluster analysis algorithms allow only a rigid/hard classification (i.e. a variable is assigned to only one class). Normally, a meteorological parameter such as precipitation has multiple causal mechanisms that overlap, leading to why a fuzzy classification technique that allows the overlapping of the classified variables might yield the best physical insight.

Bartholv et al. (1987) compared the quality of regionalization of surface pressure when cluster analysis is applied directly on the data set and when the data set is first processed with rotated PCA before the cluster analysis; they concluded that due to the high dimension of the data set, the direct application of cluster analysis to the full original raw variance yielded insufficient stratifications, but the results were improved when the dimension of the data set was reduced (i.e. noise removed) and the feature of the data set extracted with rotated PCA. Harr and Elsberry (1995) showed that fuzzy cluster analysis can as well be applied to climatic data that its major modes of variability have been extracted using empirical

orthogonal function analysis, and this approach has the advantage of yielding physically meaningful classifications. Reflecting on these findings, the rotated PCA will be used in this paper for regionalizing the austral summer (JFM) rainfall totals in southern Africa, since the paper objective has a physical end that should be applicable in the real world. The inclusion of adjacent oceans in the regionalization; construction of local oceanic indices to predict rainfall at specific landmasses; and the test of how seasonal rainfall prediction might benefit from the knowledge of local oceanic regions that are homogenous with specific landmasses is the major added value this study makes in climate predictability studies in southern Africa.

The paper is structured as follows: Section 2 describes the data and methods on which the analysis is based on. Section 3 presents the classified homogeneous rainfall regions; an analysis of the causal mechanism associated with the development of the regions at a large-scale, and the results of using appropriate predictors at local oceanic domains to predict the JFM monthly rainfall totals at landmasses that are spatially homogeneous with the oceanic domains. Conclusions and discussion of the results are presented in Section 4.

2 Data And Methodology

Gridded precipitation, 2-meter temperature, sea level pressure (SLP), relative vorticity, and divergence at 850 hPa data sets were obtained from NCEP-NCAR (Kalnay et al. 1996) for the 1979-2019 period. The temporal resolution is monthly and the horizontal resolution at which the data sets were used is 2.5° longitude and latitude. Precipitation and 2-meter temperature data sets were obtained at the Gaussian grid and were interpolated to a regular longitude and latitude grid using First order conservative remapping.

Since southern Africa receives most of its rainfall during austral summer, a monthly rainfall totals data set, for the January-March (JFM) months, was used for the regionalization. The precipitation data set was detrended (annual cycle removed), represented in the S-mode structure (i.e. variable is grid points and observation is time series), and standardized before the classification with rotated PCA. The study region for the regionalization study is 0°-36°S and 5.75°E-55.25°E; the region was chosen to include the adjacent oceans which act as moisture sources to landmasses. The grid points were related using the correlation matrix, and the singular value decomposition was used in factorizing the matrix to obtain the eigenvalues and the eigenvector. The eigenvectors were loaded with the square root of their corresponding eigenvalues, this makes them become correlation coefficients between the principal components and each field of the real monthly rainfall anomalies. The loadings present the spatial variability patterns and the scores present the time development associated with the spatial patterns (Compagnucci and Richman 2008). The number of components to retain was approached with scree-test and sensitivity analysis. For the scree-test, the explained variance versus PC number diagram was plotted and the component number after a small slope is followed by a significant drop was considered to be truncated. The scree-test works with the recommendation of North et al. (1982) on cutting components with typically low and close eigenvalues. However, an excessive truncation of the PCs might lead to

information loss. According to Gong and Richman (1995), orthogonal components can be sensitive to the number of retained PCs; to remove the orthogonality constraint - enabling the scores to be correlated - and to maximize the number of near-zero loadings, so that each retained component clusters sub-regions with high spatial coherence, the loadings were rotated obliquely using Promax at a power of 2. The sensitivity analysis implies that the addition of a new component yields a new homogeneous region that has not been already classified by previous vectors, typically assessed by low congruence coefficient between its loadings and the loadings from the already classified patterns.

A disadvantage of using the rotated PCA according to Gong and Richman (1995) is that it yields better results for a large sample size (i.e. greater or equal to 100), thus to bypass it, this study is based on an analysis period of 41 years of monthly JFM data sets, comprising of 123 sample size.

The rotated PCA classification is inherently fuzzy but can be hardened by choosing a cut-off threshold to further cluster grid points under each retained component, with their loadings above the chosen threshold. Nevertheless, the hardening might not hinder overlapping in the classification output but according to Gong and Richman (1995), the accuracy of the regionalization study they conducted dropped after the hardening, though still better than the hard cluster analysis techniques which do not also allow overlapping of the classified variables. For interpreting the patterns, the focus was placed more on the loadings greater than 0.35; because some loadings less than the cut-off threshold could be either due to noise or because another retained component describes the feature there (Barreira and Compagnucci 2011). Considering that the classified data set is not discrete, it might still be physically realistic when the feature in a given region is described by overlapping processes presented by different retained components, so one might not be certain on the actual reason for loadings less than the cut off threshold so that a prior understanding of the underlying physics in the meteorological parameter for a given region, might be necessary for distinguishing noise from phenomena. However, a threshold of 0.2-0.35 according to Richman and Gong (1999) might be sufficient to separate the PCs.

Generally, the method aims at (i) finding land and local oceanic domains that are highly homogeneous based on JFM monthly rainfall totals, (ii) analysis of the (large-scale) phenomena associated with the time development of a given rainfall homogeneous region, (iii) evaluation of meteorological parameters at the local oceanic domains that relate best to rainfall anomalies at specific landmasses to enable the usage of the parameters to construct oceanic indices for seasonal rainfall prediction at the landmasses where the indices work best.

The PC scores present the time development associated with a given region and high PC scores reveal months when the rainfall region is most expressed (Compagnucci and Richman 2008). A subjective threshold of 0.7 was used to cluster months when the analysed rainfall regions were most expressed. The difference between the time average of the clustered months and the climatological normal of 1981-2010 was computed for the analyzed parameters, enabling the investigation of physical processes. Gridpoint Student’s t-test was used in testing for statistical significance of the composite anomalies.

The examined meteorological parameters at the local oceanic regions for rainfall prediction at the landmasses are 2-meter temperature; rainfall, SLP; divergence, and relative vorticity at 850 hPa. All data sets were detrended (annual cycle removed). Correlation analysis and stepwise multiple linear regression were used in relating the predictors at the local oceanic regions to rainfall anomalies at specific homogenous landmasses (i.e. the predictand) at the seasonal and inter-annual time scale. Test of statistical significance for the correlation and regression analysis were made using the Kendall-Tau test and the F-statistics, respectively, at a 95% confidence level. A multiple linear regression equation developed from the predictor-predictand relationship was used in assessing the ability of the method to predict seasonal and inter-annual rainfall at the selected homogeneous rainfall regions.

To effectively delineate land regions and oceanic domains that are highly homogeneous with respect to the JFM monthly rainfall totals, a land mask, and an ocean mask was created for the detrended data sets. For each selected component (loadings), a threshold of 0.35 was used to delineate grid points that highly covary. The highly homogeneous grid points comprise of both land and adjacent ocean grids; thus the spatial average of the clustered grid points for the selected predictors was computed from the ocean mask data sets to obtain the oceanic indices that were used as predictors, similarly, the spatial average of rainfall anomaly computed from the land mask data set was used as the predictand.

3 Results

Fig. 1 shows the explained variance versus the PC number diagram. Based on the recommendation of North et al. (1982) on the separation of eigenvalues for the retained components, it was found that after the 10th component, the eigenvalues were typically close; and with retaining 10 components, almost all the landmasses were classified under a given sub-region so that the addition of further components only uncovers a new rainfall region that has already been delineated. By retaining 10 components, only approximately 60 percent of the total variation was captured. This was not surprising since the addition of adjacent oceans in the classification of the JFM monthly rainfall totals implies also that the complexity of the multivariate relationship in the data set might further increase. However, in line with the study aim to uncover landmasses that are spatially coherent with adjacent oceans, retaining 10 components, is at least satisfactory since the majority of the landmasses were classified.

Fig. 2 shows the sub-regions with coherent JFM rainfall totals for the 1979-2019 period. The rainfall regions in Fig. 2 were also reproduced when the analysis period was divided into two halves (not shown), however, the stability of the classification shifted for the classification done before the satellite age (i.e. before 1979), suggesting the sensitivity of the regionalization output to the quality of the input data set. From Fig. 2, there is considerable overlapping in the classification output and also some regions tend to have both high negative and positive loadings (e.g. R5); an interpretation to this is that the causal mechanism associated with positive rainfall anomaly at a given region might as well overlap, in causing negative rainfall anomaly at another region. For example, in Fig.1 a dominant positive anomaly is evident at southern African landmass while at Madagascar and northern Mozambique a weak negative anomaly can be seen; this is typically what happens during an inactive state of the Mozambique Channel Trough

when rainfall is enhanced at southern African landmasses relative to Madagascar and northern Mozambique (Barimalala et al. 2019). The weak structuring of the rainfall regions is physically realistic, but the focus for further analysis will be placed more on coherent regions where positive loadings dominate on the landmasses and adjacent oceans. Also, to aid robust physical interpretability tropical and some subtropical regions will be considered since the rainfall climatology of the mid-latitudes is much more complex (e.g. the activity of cold fronts). For this reason, R7 will not be considered.

R2, R4, R6, and R8 are selected for further analysis since these sub-regions features a considerable number of grid points at the landmasses homogeneous with adjacent oceans. R2 features high positive loading dominant at some central regions of southern Africa, Mozambique Channel, and northern Agulhas current; Fig. 3 and 4 clearly show that the spatial patterns of SLP and relative vorticity anomalies play a vital role in the development of the spatial structure of R2. Also, the wind anomalies indicate enhanced convergence of moist southeast, cross-equatorial northeast, and northwest winds at the central landmasses. R4 features a positive loading dominant at Madagascar (except for the northernmost region) and the adjacent oceans to the east coast and the west coast. From Fig.3 and 4, R4 is characterized by a strong cyclonic and negative relative vorticity anomaly dominant in oceans adjacent to Madagascar; as a result, southeast and northeast winds are adjusted to westerly towards Madagascar where they converge. It is evident that the casual mechanism of R2 and R4 partly overlaps based on enhanced cyclonic activity at the east coast of Mozambique. In R6, a positive loading dominates at the west-central equatorial landmasses and parts of the tropical South Atlantic east coast. Fig. 3 shows that wind anomalies are predominantly southerly towards the equator and from Fig. 4 enhanced relative vorticity and convergence are evident at R6. Enhancement of convergence and relative vorticity during the active months of R6 is quite conceivable since enhanced convergence by the Inter-Tropical Convergence Zone is the principal mechanism that controls the tropical rainfall. R8 is similar to R4 except that in the former enhanced cyclonic activity, cyclonic relative vorticity anomaly, and the convergence of northeast and southeast anomalous winds shift towards northern Madagascar.

Anomalies of monthly JFM divergence field, relative vorticity, 2-meter temperature, rainfall totals (i.e. the predictors) were spatially averaged at local oceanic domains that are homogeneous with the landmasses in R2, R4, R6, and R8, respectively, for the 1979-2019 period; also anomalies of JFM rainfall totals were spatially averaged for the coherent landmasses (predictand) in R2, R4, R6, and R8, respectively. To uncover the statistical relationship between the JFM rainfall totals at the landmasses and oceanic domains, in addition to the underlying dynamics, correlation and regression analysis were made. Correlations that are statistically significant at a 95% confidence level are presented in Table 1, dash in the table show correlations that are not significant. For R2, Table 1 shows that variations in SLP, relative vorticity, 2-meter temperature, and divergence at the local oceanic domains that R2 landmass is coherent with, are related to the variations in rainfall anomaly at the landmasses in R2. Cyclonic circulation, cyclonic relative vorticity, lower temperature, and enhanced convergence anomalies during austral summer at the local oceanic domains are related to enhanced rainfall in R2 landmasses. The correlation coefficient between rainfall anomalies at the local oceanic domains and the landmasses at R2 was rather found to be relatively higher at a lag of 1 month compared to lag 0. In R4, variations in SLP and rainfall

anomalies at the adjacent oceans are well related to rainfall anomalies at the landmasses. Cyclonic anomaly and positive rainfall anomaly at the adjacent oceans of R4 relates to enhanced rainfall at the landmasses in R4. In R6, divergence/convergence at the local oceanic domains (i.e. parts of the tropical South Atlantic east coast) is strongly related to negative/positive rainfall anomaly at the west-central equatorial landmasses. Also, cyclonic relative vorticity and positive rainfall at the local oceanic domains are equally related to enhanced rainfall at the landmasses in R6. In R8, cyclonic anomaly and positive rainfall anomaly at the concerned local oceanic domains are well related to enhanced rainfall anomaly at northern Madagascar. Table 2 shows the percentage of variability in JFM monthly rainfall at the homogeneous landmasses explained by the oceanic indices, based on linear and multiple linear regression analysis. It can be seen that relatively, the predictors jointly captured best the variations in R6 due to the high relationship between convergence anomaly at the local oceanic domains and rainfall anomaly at the west-central equatorial landmasses.

Table 3 shows the error (i.e. predicted minus observed, in absolute value), in some summary statistics for the predicted JFM monthly rainfall totals anomaly at the landmasses, using multiple linear regression and the appropriate predictors in Table 1 and 2 (i.e. those that the correlations and regressions are statistically significant). It can be seen that prediction accurately captured the long-term mean in JFM rainfall anomaly at the respective landmasses and also relatively, the distribution of JFM rainfall anomaly at the west-central equatorial landmasses (Fig. 5) is well captured using the selected predictors. Fig. 6 shows the time series for annual (JFM) rainfall anomaly averaged over the coherent landmasses for the actual and predicted values; it can be seen that generally the year to year variability of JFM rainfall anomaly averaged over the specified coherent landmasses might be predicted using the local oceanic indices, especially at the deep tropics. The mean absolute error for predicted values in R6 is 35 mm/month but higher for the other regions (i.e. within the range of 60 – 100 mm/month).

4 Discussions And Conclusions

Several studies have indicated the role SST anomalies at the south Indian Ocean and the South Atlantic Ocean play in the seasonal rainfall variability at the southern African landmasses (Walker 1990; Reason and Mulenga 1999; Cook 2000; Viguad 2009). During austral summer, warm SST at the southwest Indian Ocean correlates with enhanced rainfall in southern Africa. A study by Rapolaki et al. (2020) reported the specific oceanic domains at the south Indian Ocean and the South Atlantic Ocean that act as moisture sources to southern Africa. Since the adjacent oceans in southern Africa play a vital role in rainfall variability in the sub-continent, it will be conceivable that most rainfall that falls on the landmasses starts in the adjacent oceans; this also includes the equatorial regions. Thus this study uses the concept of regionalization to delineate landmasses and local adjacent oceanic domains that are spatially homogeneous based on JFM monthly rainfall totals, to further understand the phenomena that relate to rainfall at specific landmasses and local oceanic domains. The understanding of the phenomena helps to achieve the best physical insight on using appropriate meteorological parameters (predictors) related to rainfall at the local oceanic domains to predict rainfall anomaly at the landmasses that fall under the same homogeneous rainfall regions with the local oceanic domains, using linear regression.

Rotated PCA was applied on the S-mode matrix of detrended monthly JFM rainfall totals in Africa, south of the equator, for the regionalization. The technique was selected since it is fuzzy and allows overlapping of the classified grid points so that it yields a physical insight (Gong and Richman 1995). According to Huth et al. (2008), the output of fuzzy classification techniques though of interest in the climatological society is cumbersome and difficult to apply; some recent developments in clustering of geophysical data sets seem to at times neglect the fuzzy and overlapping nature of the data sets, perhaps the result of the negligence in getting the best physical insight from a cluster analysis technique in geophysical research, and the prioritizing of rigid cluster analysis algorithms that yields very high internal cohesion and external isolation have led to unusual randomness in the outputs of both regionalization and synoptic climatological classifications. According to Gong and Richman (1995), rigid cluster analysis techniques for regionalization purposes might: (i) be sensitive to the choice of dissimilarity matrix and the number of retained clusters, (ii) distort the multidimensional space of the input parameter, (iii) omit useful information on the probability of group membership, (iv) result in irregularly shaped regions; in general, the result might lack consistency. The output of regionalizing geophysical data or synoptic classifications nevertheless is supposed to show some level of stability because the causal mechanism associated with the meteorological parameter is at least statistically stationary in the sense that the physical laws do not change with time (Lorenz 1956). In geophysical research, the ranking of a classification output as suggested by Gong and Richman (1995) and Huth (1996) should be based on the ability of the technique to preserve the underlying physics and not the quality of stratification it yields in the mathematical sense (i.e. very well separated classes). A major argument on the preference of hard/rigid cluster analysis algorithms in geophysical research is to reduce internal variability in the classified data, however when this is done at the expense of the underlying physics in the climatic system then it might as well be unnecessary. A major disadvantage of rotated PCA applied to a meteorological parameter represented in the S-mode structure is that the eigenvalues are related to the number of grid points, perhaps this might contribute to why altering the target study domain tend to cause some shift in the classification output. As Richman and Gong (1995) suggested, an examination of physical mechanisms before embarking on the classification will be appropriate to ensure the results are realistic.

10 regions with coherent JFM monthly rainfall totals were classified. The time behavior of the respective regions is categorized by the associating PC scores (Compagnucci and Richman 2008) and this enabled further examination of the large-scale causal mechanism associated with the development of a given region. When the physical mechanism associated with rainfall variability in the study region is known *a priori*, uncovering the physical insight associated with the development of the classified rainfall regions and relating it to the already known mechanisms associated with rainfall climatology at the study region, might be the best way to validate if the classifications are meaningful, compared to the popular usage of statistical metrics such as the Euclidean distance.

Of interest are the rainfall homogeneous sub-regions at the tropical and some subtropical landmasses that are coherent with adjacent oceans. The analysis of such regions showed that at the west-central equatorial regions, convergence and vorticity play the leading role in the time development of the rainfall

region, and the result is quite reasonable in line with the already known climatology of the tropics. For the homogeneous sub-regions comprising of the central domains in southern Africa and Madagascar, variations in the patterns of SLP, vorticity, and convergence in the landmasses, the tropical and the southwest Indian Oceans explains well the time development of the rainfall sub-regions. This is equally physically reasonable since SST anomaly (at the tropical and the southwest Indian Ocean) which plays a vital role in the austral summer rainfall climatology in the study region, is related to SLP gradients, adjustment of winds, moisture convergence, and so rainfall at the preferred homogeneous domains.

Finally, using multiple linear regression analysis with appropriate predictors (related to rainfall) at the local oceanic regions, the seasonal and inter-annual variability in JFM rainfall anomaly averaged over the coherent landmasses were captured especially for the west-central equatorial landmasses. To this end, using the method outlined here, future studies might consider how the addition of more predictors related to rainfall at the local oceanic regions might reduce the error in the seasonal rainfall predictions to enable forecast.

Declarations

Acknowledgments

Thanks to NOAA/OAR/ESRL PSL, Boulder, Colorado, USA for the NCEP/NCAR data sets provided on their website at https://psl.noaa.gov/.

Conflict of interest: There are no conflicts of interest in this paper

Funding statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author's contribution: work was designed and executed by Chibuike Chiedozie Ibebuchi

References

1. Badr HS, Dezfuli AK, Zaitchik BF, Peters-Lidard CD (2016) Regionalizing Africa: Patterns of Precipitation Variability in Observations and Global Climate Models. J Clim 29 24:9027–9043. https://doi.org/10.1175/JCLI-D-16-0182.1
2. Barimalala R, Blamex RC, Desbiolles F, Reason CJC (2019) Variability in the Mozambique Channel Trough and impacts on Southeast African rainfall. J Clim 33 2:749-765. https://doi.org/10.1175/JCLI-D-19-0267.1
3. Barreira S, Compagnucci RH (2011) Spatial fields of Antarctic sea-ice concentration anomalies for summer–autumn and their relationship to Southern Hemisphere atmospheric circulation during the period 1979–2009. Ann Glaciol 52 57:140-150. https://doi.org/10.3189/172756411795931741
4. Bartholv J, Barnston AG, Livezey RE (1987) The use of cluster analysis and rotated empirical orthogonal function analysis in describing the macrocirculation pattern of the northern hemisphere

and the Atlantic-European region on different heights. In Preprints 3rd International Meeting on Statistical Climatology, Vienna, 23–27

5. Compagnucci RH, Vargas WM (1987) Patterns of surface pressure fields during July 1972–1983 in southern South America and the Antarctic Peninsula. In Preprints 3rd International Meeting on Statistical Climatology, Vienna, 1–14
6. Compagnucci RH, Richman MB (2008) Can principal component analysis provide atmospheric circulation or teleconnection patterns? Int J Climatol 28 6:703–726. https://doi.org/1002/joc.1574
7. Cook KH (2000) The South Indian Convergence Zone and Interannual Rainfall Variability over Southern Africa. J Clim 13 21:3789–3804. https://doi.org/10.1175/1520-0442(2000)013<3789:TSICZA>2.0.CO;2
8. Gong X, Richman MB (1995) On the Application of Cluster Analysis to Growing Season Precipitation Data in North America East of the Rockies. J Clim 8 4:897-931. https://doi.org/10.1175/1520-0442(1995)008<0897:OTAOCA>2.0.CO;2
9. Huth R (1996) An intercomparison of computer-assisted circulation classification methods. Internation. Int J Climatol 16:893-922. https://doi.org/10.1002/(SICI)1097-0088(199608)16:8<893::AID-JOC51>3.0.CO;2-Q
10. Huth R, Beck C, Philipp A, Demuzere M, Ustrnul Z, Cahynová M, Kysely J, Tveito OE (2008) Classifications of atmospheric circulation patterns: recent advances and applications. Ann N Y Acad Sci 1146:105–152. https://doi.org/10.1196/annals.1446.019
11. Johnson F, Green J (2018) A comprehensive continent-wide regionalisation investigation for daily design rainfall. J Hydrol Reg Stud 16 67-79. https://doi.org/10.1016/j.ejrh.2018.03.001
12. Kalnay E et al (1996) The NCEP/NCAR 40-year reanalysis project. Bull Amer Meteor Soc 77 3:437-472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2
13. Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, Contract AF19 (604)-1566, Meteorology Department, Massachusetts Institute of Technology, Cambridge, MA.
14. North Gerald, Bell T, Cahalan FR, Moeng FJ (1982) Sampling errors in the estimation of empirical orthogonal functions. Monthly Mon Wea Rev 110 7:699-706. https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2
15. Rapolaki RS, Blamey RC, Hermes JC et al. (2020) Moisture sources associated with heavy rainfall over the Limpopo River Basin, southern Africa. Clim Dyn 55:1473–1487. https://doi.org/10.1007/s00382-020-05336-w
16. Reason CJC, Mulenga H (1999) Relationships between South African rainfall and SST anomalies in the southwest Indian Ocean. Int J Climatol 19:1651–1673
17. Richman MB, Gong X (1999) Relationships between the definition of the hyperplane width to the fidelity of principal component loadings patterns. J Clim 12 6:1557–1576. https://doi.org/10.1175/1520-0442(1999)012<1557:RBTDOT>2.0.CO;2

18. Vigaud N, Richard Y, Rouault M, Fauchereau N (2009) Moisture transport between the South Atlantic Ocean and Southern Africa: relationships with summer rainfall and associated dynamics. Clim Dyn 32 1:113-123. https://doi.org/1007/s00382-008-0377-7
19. Walker ND (1990) Links between South African summer rainfall and temperature variability of the Agulhas and Benguela Current systems. J Geophys Res Oceans 95 C3:3297-3319. https://doi.org/10.1029/JC095iC03p03297
20. Yu Y, Shao Q, Lin Z (2018) Regionalization study of maximum daily temperature based on grid data by an objective hybrid clustering approach. J Hydrol 564 149-163 https://doi.org/10.1016/j.jhydrol.2018.07.007

Tables

Table 1 Correlation coefficients between the selected predictors averaged over the local oceanic regions and monthly rainfall anomaly averaged over the landmasses coherent with the oceanic regions for the (JFM) 1979-2019 period

Region	SLP	2-meter temperature	Relative vorticity	Divergence	Precipitation
R2	-0.48	-0.52	-0.48	-0.52	0.22
R4	-0.57	-	-0.31	-0.38	0.61
R6	-	0.3	-0.49	-0.75	0.47
R8	-0.41	0.33	-0.31	-0.34	0.53

Table 2 Regression coefficients between the selected predictors averaged over the local oceanic regions and monthly rainfall anomaly averaged over the landmasses coherent with the oceanic regions for the (JFM) 1979-2019 period

Region	SLP (%)	2m temperature (%)	Relative vorticity (%)	Divergence (%)	Precipitation (%)	Joint variance explained (%)
R2	23	27	24	27	5	44
R4	33	-	10	15	37	46
R6	-	10	25	57	23	69
R8	17	11	9	-	29	42

Table 3 Error, in absolute value, for the summary statistics between the actual and predicted JFM monthly rainfall anomaly

Region	Mean (mm/month)	Standard deviation (mm/month)	95 percentile (mm/month)	5 percentile (mm/month)
R2	0.0	36.5	48.9	48.2
R4	0.0	37.9	49.8	85.0
R6	0.0	13	0.3	29
R8	0.0	58.5	104	87

Figures

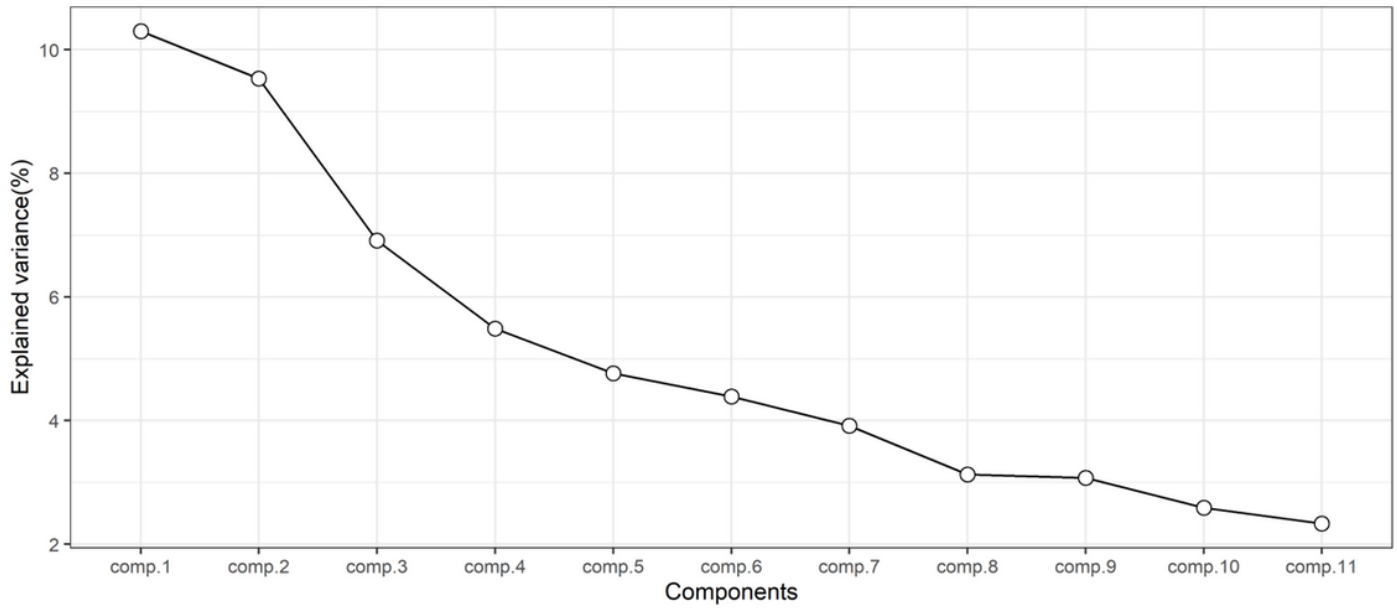

Figure 1

Explained variance versus PC number diagram

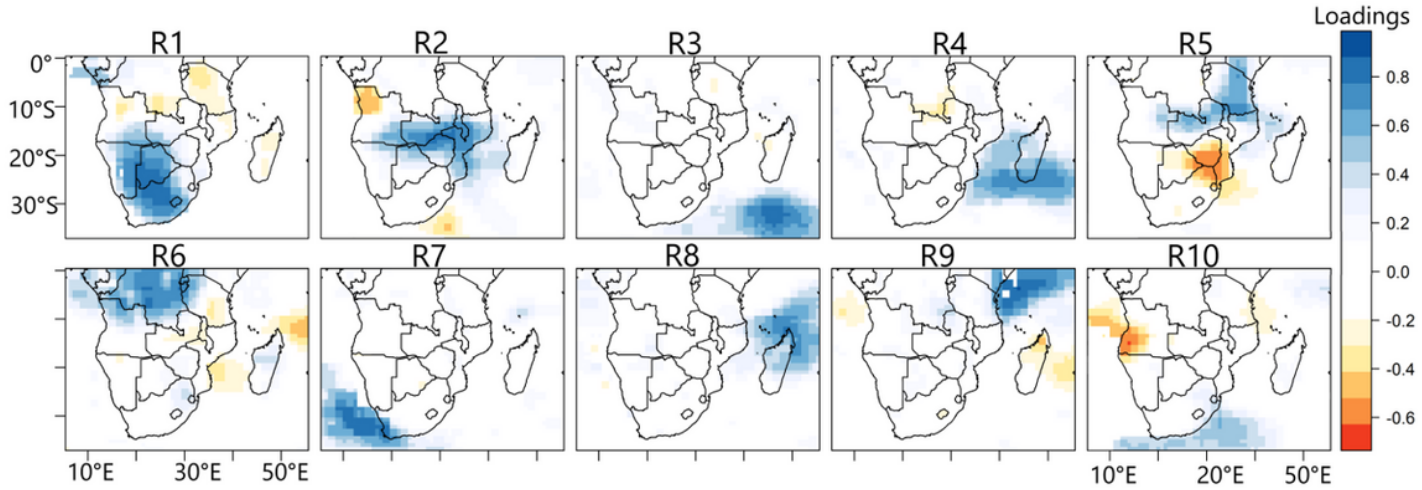

Figure 2

Domains with homogeneous monthly JFM rainfall totals

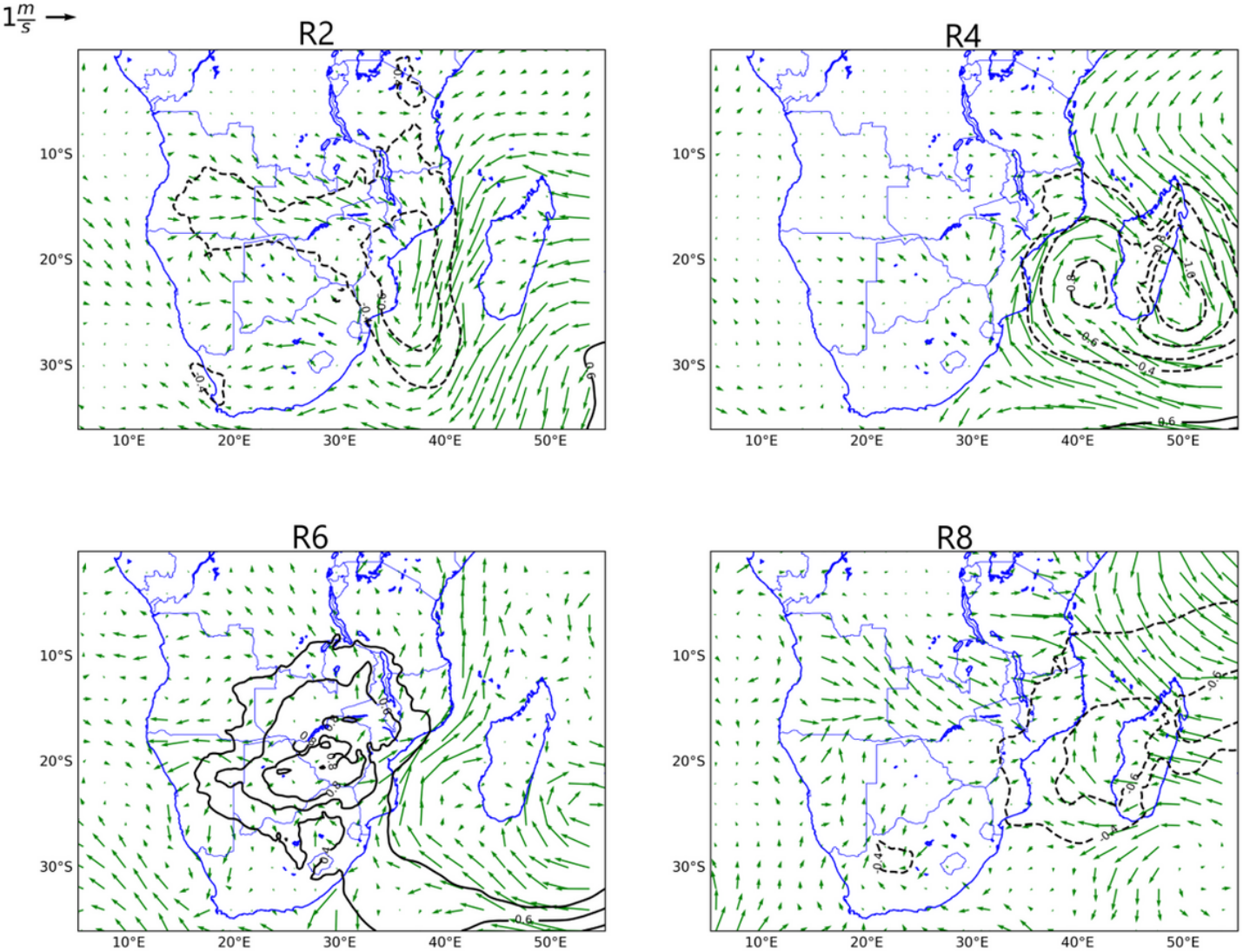

Figure 3

Composite anomalies for 10 meter wind in m/s and sea level pressure for the selected regions. Contour lines show SLP anomalies in hPa and the contour interval is 0.2hPa. Only values exceeding the 95% confidence limit from the Student's grid-point t-test were plotted. Green vectors are wind anomalies in m/s. Dash contour lines show negative SLP anomalies and thick contour lines show positive SLP anomalies. Scale of the vector is written on top of the map

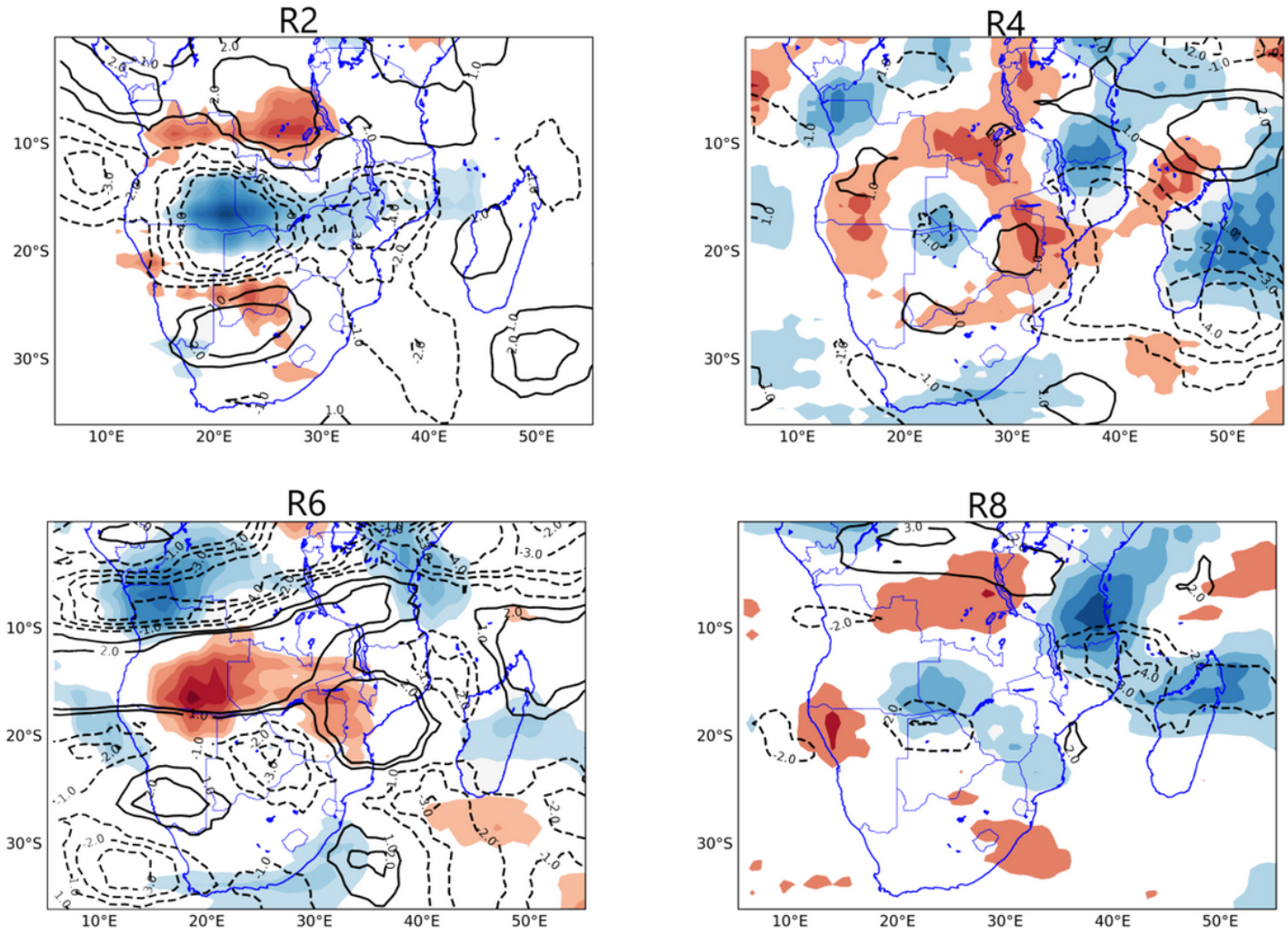

Figure 4

Composites anomalies of relative vorticity and divergence for the selected regions. Contour lines show relative vorticity anomaly at 10^6/s and shading shows divergence anomaly. Only regions with values exceeding the 95% confidence limit using Student's grid-point t-test were plotted. Blue colors indicate negative divergence (i.e. convergence) and red colors indicate positive divergence. Dash contour lines show negative relative vorticity anomalies and thick contour lines show positive relative vorticity anomalies

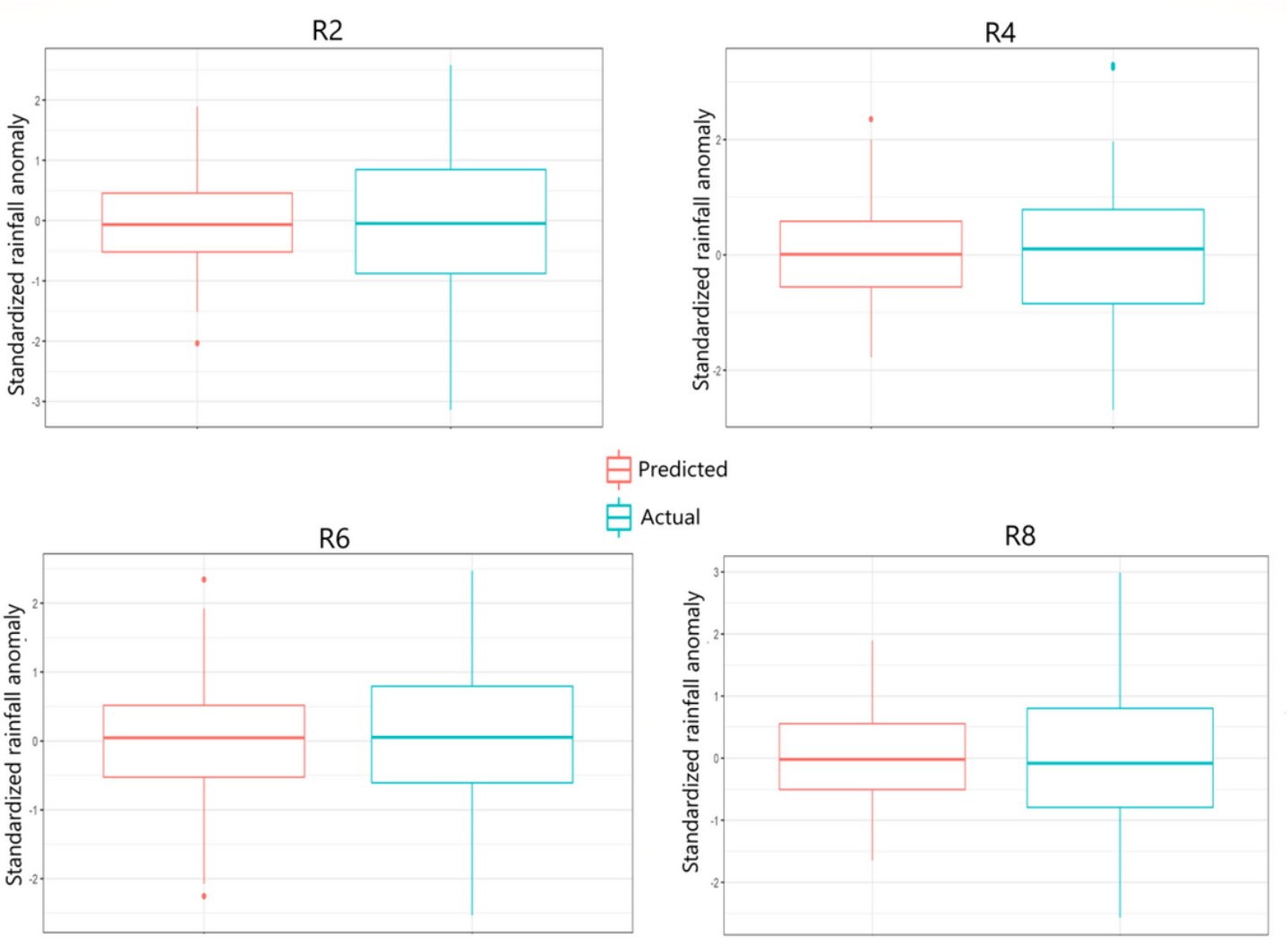

Figure 5

Predicted and actual distribution of JFM rainfall anomaly averaged over the landmasses with homogeneous JFM monthly rainfall totals for the 1979-2019 period

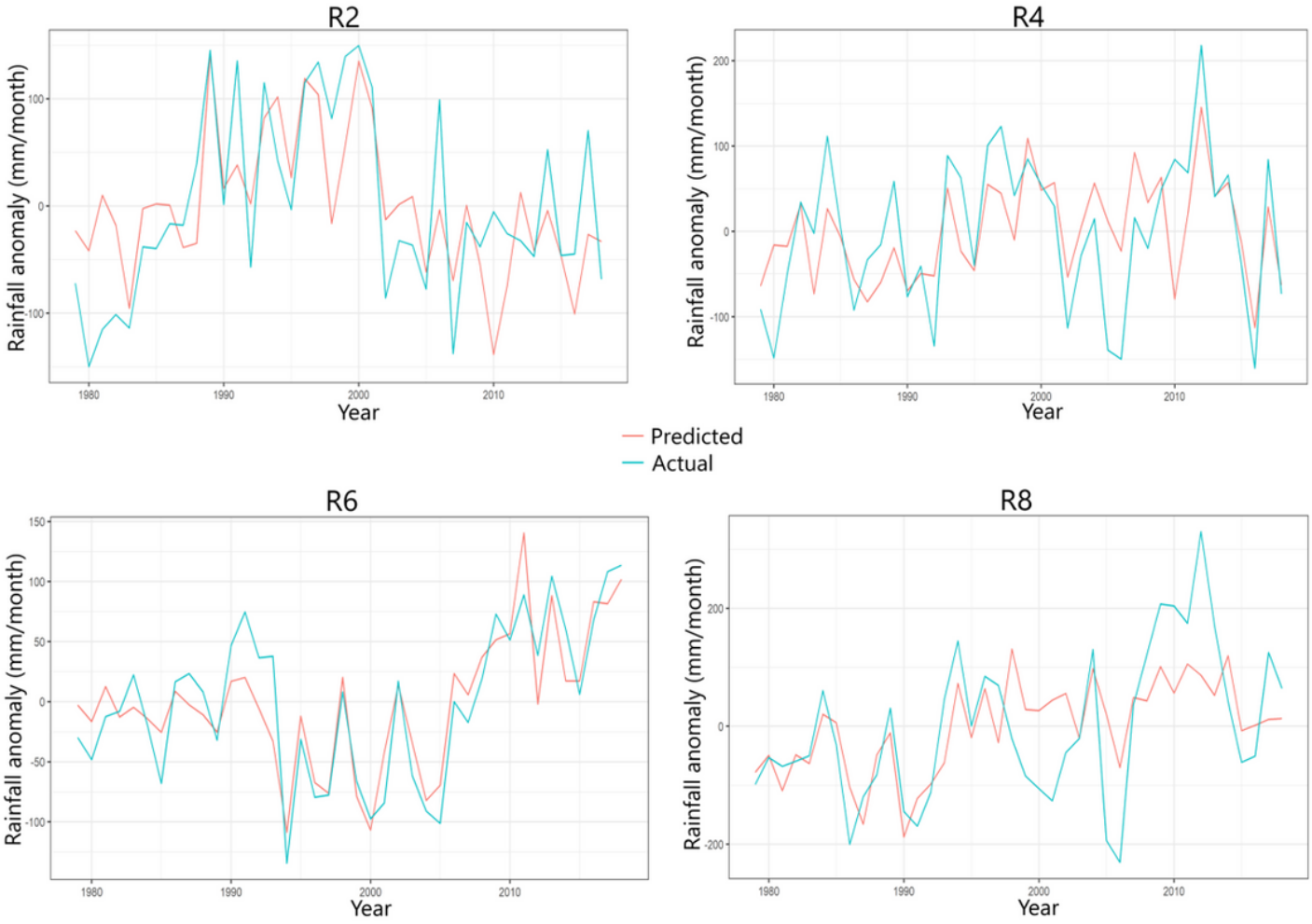

Figure 6

Time series of the actual and predicted annual mean JFM rainfall totals for the 1979-2019 period from the selected homogeneous rainfall regions