Received: 21 December 2019 | Revised: 10 April 2020 | Accepted: 14 April 2020

DOI: 10.1002/joc.6608

RESEARCH ARTICLE

International Journal of Climatology RMetS

A semi-objective circulation pattern classification scheme for the semi-arid Northeast Brazil

Patrick Laux[1,2] | **Brian Böker**[2] | **Eduardo Sávio Martins**[3] | **Francisco das Chagas Vasconcelos Junior**[3] | **Vincent Moron**[4] | **Tanja Portele**[1] | **Christof Lorenz**[1] | **Andreas Philipp**[2] | **Harald Kunstmann**[1,2]

[1]Institute of Meteorology and Climate Research (IMK-IFU), Campus Alpin, Karlsruhe Institute of Technology (KIT), Garmisch-Partenkirchen, Germany

[2]Institute of Geography, University of Augsburg, Augsburg, Germany

[3]Ceará Institute for Meteorology and Water Resources (FUNCEME), Fortaleza, Brazil

[4]Aix-Marseille University, CNRS, IRD, INRAE, Coll. de France, CEREGE, Aix-en-Provence, France

Correspondence
Patrick Laux, Institute of Meteorology and Climate Research (IMK-IFU), Campus Alpin, Karlsruhe Institute of Technology (KIT), Garmisch-Partenkirchen, Germany.
Email: patrick.laux@kit.edu

Funding information
Seasonal water re-sources management in semi-arid regions: Transfer of regionalized global information to practice (SaWaM), Grant/Award Number: 02WGR1421A

Abstract

The semi-arid Northeast Brazil (NEB) is just recovering from a very severe water crisis induced by a multiyear drought. With this crisis, the question of water resources management has entered the national political agenda, creating an opportunity to better prepare the country to deal with future droughts. In order to improve climate predictions, and thus preparedness in NEB, a circulation pattern (CP) classification algorithm offers various options. Therefore, the main objective of this study was to develop a computer aided CP classification based on the Simulated ANnealing and Diversified RAndomization clustering (SANDRA) algorithm. First, suitable predictor variables and cluster domain setting are evaluated using ERA-Interim reanalyses. It is found that near surface variables such as geopotential at 1,000 hPa ($GP_{1,000}$) or mean sea level pressure (MSLP) should be combined with horizontal wind speed at the upper 700 hPa level ($UWND_{700}$). A 11-cluster solution is favoured due to the trade-offs between interpretability of the cluster centroids and the explained variances of the predictors. Second, occurrence and transition probabilities of this 11-cluster solution of $GP_{1,000}$ and $UWND_{700}$ are analysed, and typical CPs, which are linked to dry and wet conditions in the region are identified. The suitability of the new classification to be potentially applied for statistical downscaling or CP-conditional bias correction approach is analysed. The CP-conditional cumulative density functions (CDFs) exhibit discriminative power to separate between wet and dry conditions, indicating a good performance of the CP approach.

KEYWORDS

bias correction, circulation pattern classification, Northeast Brazil, precipitation, SANDRA, statistical downscaling

1 | INTRODUCTION

The semi-arid northeast of Brazil (NEB) is one of the most densely populated dryland regions in the world. It is characterized by recurrent droughts (Marengo *et al*., 2017), and just now still recovering from a series of drought years from 2012 to 2018 (Pilz *et al*., 2019). This has recently led to a severe water crisis, not only in NEB,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2020 The Authors. International Journal of Climatology published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

but also the large metropolitan areas of São Paulo, Belo Horizonte, and Rio de Janeiro.

The NEB can be characterized by a high-rainfall variability, that is, on interannual and intraannual scale, but also on decadal scale (Zhou and Lau, 2001).

According to Kousky (1979), eastern coastal regions may receive even more than 2000 mm/year, while some interior valleys receive less than 400 mm/year. Particularly in these arid regions, the year-to-year variability is higher. They show that frontal systems, which penetrate the southern parts of the NEB, play a crucial role for December–January precipitation in the southern part, but also for fall and winter precipitation along the coast. Such frontal systems lower the surface pressure at low latitudes, and thus favour the southward movement of the equatorial trough zone. Figure 1 illustrates such interannual variability by showing precipitation anomalies for three selected years. Heavy precipitation deficits up to 1,000 mm can be seen during the recent drought period for the years 2012 and 2014, whereas precipitation excess is exemplary shown for the year 1985.

The high-rainfall variability on the different temporal scales poses significant challenges for water resources management in the region. Even in years with average rainfall amounts, such as in 2017 and 2018, NEB was not able to overcome its water scarcity due to the high-spatial variability of the rainfall in the region. In these years, the largest reservoirs could not be filled up by more than 10% of its capacity.

Intraannual or seasonal precipitation variability during the wet season of the interior NEB is impacted by sea surface temperature (SST) anomaly patterns in the Atlantic, the Pacific, and to a lesser extent the Indian Ocean (Ward and Folland, 1991). Likewise, in Hastenrath (2012), drought occurrence in NEB has been linked to SST anomalies in the eastern Pacific. The El Niño Southern Oscillation (ENSO) as well as the northern tropical Atlantic region (i.e., the Tropical Atlantic SST Dipole) is influencing the location of the Intertropical Convergence Zone (ITCZ), which is known to be the main source of rain during the rainy season in this region.

According to Costa *et al*. (2016), most of the drought and wet anomalies could be linked to anomalous states of El Niño Southern Oscillation (ENSO), but not necessarily linked to specific El Niño or La Niña events. In a later study, Costa *et al*. (2018) identified a multi-annual relationship between the state of the sea surface temperature (SST) of the Atlantic and Pacific oceans and anomalous hydrological variability in NEB. Thereby, the northern Tropical Atlantic conditions were shown to play an important role in modulating the long-term variability of the hydrological response of the basins, while only extreme ENSO anomalies seemed to affect the rainy season.

Such large-scale synoptic information, represented by general circulation models (GCMs) can be used for prediction purposes in many water resources management applications. One possibility to implement the large-scale synoptic information is to classify the atmospheric state into circulation patterns or weather types, hereinafter referred to as circulation patterns (CPs) (e.g., Fernández-González *et al*., 2012). In general, one can distinguish between

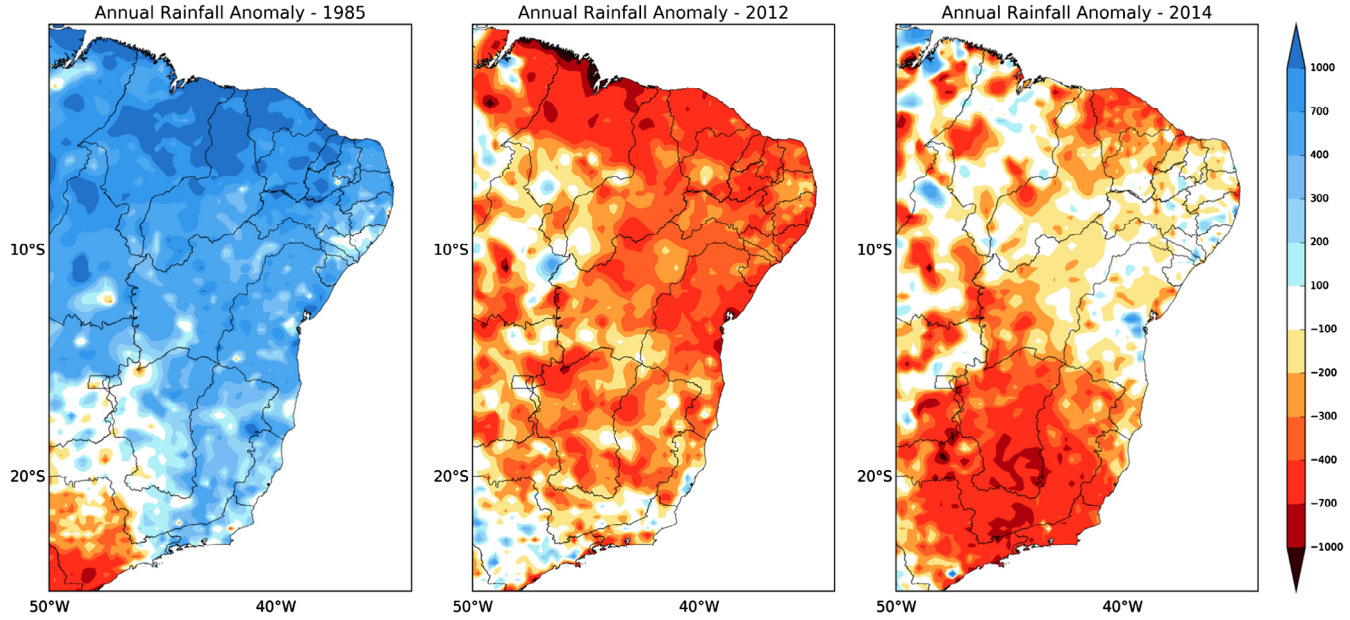

FIGURE 1 Wet precipitation anomaly of 1985 (left) and dry precipitation anomalies of 2012 (middle) and 2014 (right) over the northeast region of Brazil (kriging of observation data, 1981–2010)

manual (subjective) and computer assisted (objective or mostly semi-objective) methods. Often, it is distinguished also between "circulation to environment" and "environment to circulation" approaches (e.g., Dayan *et al.*, 2012). In the former, the classification is performed without consideration of the target variable, whereas the latter accounts for the target variable(s) during the classification. Both approaches are suitable tools in synoptic climatology, and they can be applied both to analyse the link between atmospheric circulation types derived from CP classifications and surface variables such as, precipitation, temperature, or discharge (e.g., Bárdossy *et al.*, 2002; Bárdossy, 2010; Wypych *et al.*, 2018; Bednorz *et al.*, 2019). Such links exist even when the local scale climate is driven by mesoscale events such as convective systems, since these are conditional on the synoptic conditions (Goodess and Jones, 2002). Other authors established links between atmospheric conditions or Sea Surface Temperature (SST) patterns and droughts (e.g., Burgdorf *et al.*, 2019), and demonstrated the usability of their classification for drought prediction. Due to the existence of such links, CP classifications have been originally developed for weather forecasting.

Although GCMs are often strongly biased in prognostic variables such as precipitation, it is widely accepted that GCMs are able to appropriately reflect the large-scale circulation (e.g., Sunyer *et al.*, 2015). This forms the basis for statistical downscaling of hydrometeorological target variables, nowadays, the most widely used application of CPs (e.g., Wetterhall *et al.*, 2007; Bárdossy, 2010; Bárdossy and Pegram, 2011). Apart from dynamical downscaling by using process-based regional climate models (RCMs), in statistical downscaling the large-scale synopic information, represented by CPs is used to derive hydrometeorological variables of interest (e.g., precipitation, temperature, and discharge) on local scales. Moreover, CP classification can be applied for model evaluation and validation, for example, to evaluate the performance of GCMs or RCMs due to their representation of the large-scale information (e.g., Dafka *et al.*, 2018; Prein *et al.*, 2019).

Another very recent application is the identification and correction of the spatio-temporal bias structures in RCMs (Le Roux *et al.*, 2019). The RCMs may have largely different performances depending on the season and the prevailing large-scale atmospheric circulation (Laux *et al.*, 2011; Wetterhall *et al.*, 2012). The inclusion of conditional information about the state of the atmospheric circulation, for example, as CP time series, obtained by classification algorithms may thus help to improve the bias correction. An overview of different classification algorithms, distinguished between subjective, objective, and mixed approaches can be found in Huth *et al.* (2008).

Most approaches are developed for the mid-latitudes, where distinct high- and low-pressure systems are distinguished. A relatively large body of literature exists for Central Europe (e.g., James, 2007; Philipp *et al.*, 2007, 2010) or the Mediterranean region (e.g., Corte-Real *et al.*, 1995; Goodess and Palutikof, 1998; Trigo and DaCamara, 2000; Fernández-González *et al.*, 2012). Within the framework of the *European Cooperation in Science and Technology* (COST) action 733 *Harmonization and Applications of Weather Types Classifications for European Regions*, a bunch of different objective and subjective weather type classifications for Europe have been collected and provided for climate science community (Philipp *et al.*, 2007), and a classification catalogue has been derived for the European domain (Huth *et al.*, 2008; Philipp *et al.*, 2010).

Only few studies exist for tropical and subtropical regions. Amongst them are the study of Ngarukiyimana *et al.* (2018) for Rwanda in East Africa, Moron *et al.* (2008) for the Senegal in West Africa, and Robertson *et al.* (2004) for the NEB. In the latter, a non-homogeneous hidden Markov model was used to downscale daily precipitation occurrence at 10 stations in NEB, using GCM simulations of seasonal-mean large-scale precipitation, obtained with historical sea surface temperatures prescribed globally. It was concluded that their model provides a useful tool for understanding the statistics of rainfall occurrence at stations with respect to the large-scale atmospheric patterns, and for producing local-scale daily rainfall series for impact studies.

Due to the above mentioned physical link between synoptic-scale patterns such as ENSO and the hydrological situation in the NEB, the main objective of this study is to develop a computer aided circulation pattern (CP) classification approach for applications in water resources management. For this reason, it is aimed at a throughout evaluation based on different predictor variables as well as domain settings for the classification with respect to its ability to discriminate between dry and wet conditions in the NEB. Thus, it may provide the base for many applications in water resources management in the region.

The paper is structured as follows: in the following Section 2, we elucidate more detailed climatic features of the study region within NEB. We selected the São Francisco (SF) river basin to highlight the complex climatology and the specific challenges for water resources management. In Section 3, the SANDRA CP classification method, the Markov Chain approach as well as the data for the classification and the validation used in this study are explained. Section 4 is subdivided into the description of the classification and the selected features of the classified CPs, that is, Occurrence Probability of CPs, Persistence of CPs, Transition Probability of CPs, Wetness Indices of CPs, and the CDF of CPs, followed by Section 5.

2 | STUDY REGION

We selected the São Francisco (SF) river basin to highlight the complex climatology and specific challenges with respect to water resources management in NEB, which helps to explain our setup for the classification in Section 3.

The SF basin (Figure 2) covers an area of approximately 630,000 km^2, and has its sources in the southeast of Brazil, in the state of Minas Gerais. From there, the SF river flows approximately 2,700 km until it meets the Atlantic Ocean in the northeastern region. Starting in Minas Gerais, it crosses the states of Bahia, Pernambuco, Alagoas and Sergipe. The SF river basin partly covers also the states of Goiás and the Federal District Brasilia. The large size of SF river basin (it corresponds approximately to the size of France) and the fact that it is an interstate basin makes water resources management a particularly challenging task.

The precipitation regime in the SF basin is complex as it is modulated by different meteorological systems, such as the South Atlantic Convergence Zone in the southernmost part of the basin and ITCZ, leading to a relatively high-spatiotemporal variability of precipitation (see Figure 3). The figure shows the monthly precipitation climatology for Três Marias, Sobradinho, Aracaju (from the South to the North). The upper SF basin, represented by Três Marias, provides the highest rainfall amounts to the SF river basin, the main season lasts from October to February. The main generating rainfall system is the incursion of cold fronts, that is, the South Atlantic Convergence Zone. For the northernmost area, the lower SF basin, represented by Aracuju, the main season peaks from may to July. This region is affected by the the movement of the ITCZ. The middle SF basin, as depicted by Sobradinho, has an intermediate precipitation climatology. Besides the onset of the rainy season, the overall rainfall amounts decrease from the upper to the lower SF basin. During the dry season (June–August), significant amounts of rains (and thus streamflow) can be observed across the SF basin due to convective systems and penetrating cold fronts. For this reason, it should be noted that water resources management in the SF basin requires weather and climate forecasts not only for the rainy season, but for the entire year. This need is also corroborated by the fact that water from the SF River is

FIGURE 2 The São Francisco (SF) river basin in NEB. The figure is obtained from https://de.m.wikipedia.org/wiki/Datei:São_Francisco_basin_map.png, released under the licence CC BY-SA 4.0

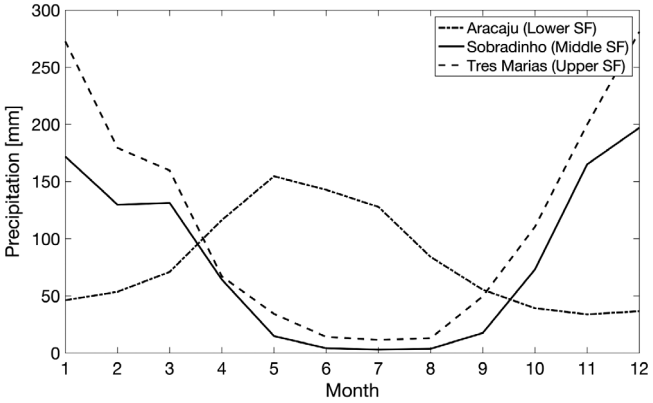

FIGURE 3 The monthly precipitation climatology for Três Marias, Sobradinho and Aracaju, derived from GPCP, representing the upper-, middle-, and lower-SF basin, respectively. The location of the stations can be obtained from Figure 2

diverted to the north of the more semi-arid NEB, in order to improve water availability there. In particular, seasonal climate forecasts for the forthcoming months can help decision makers in water resources management in such semi-arid environments (e.g., Siegmund *et al.*, 2015).

3 | DATA AND METHODS

Figure 4 provides a work flow of the procedure applied in this study. The work flow (from top to bottom) is indicated by the arrows, the main methods are given in the grey shaded boxes, that is, (i) the SANDRA CP Classification and (ii) the Markov Chain approach. Both methods are applied based on ERA-Interim data. The CPs are finally linked to local weather conditions (see Table 1) by calculating the wetness index and the CP-conditional CDFs, based on observation data. In the following the data are described first, followed by the methods.

3.1 | Data

3.1.1 | ERA-Interim reanalyses

The ERA-Interim project, initiated in 2006, aimed to provide a bridge between ECMWF's previous reanalysis, that is, ERA-40 (1957–2002), and the next-generation extended reanalysis envisaged at ECMWF (ERA5). Improvements compared to ERA-40 mainly comprised the representation of the hydrological cycle, the quality of the stratospheric circulation, and the handling of biases and changes in the observing system. The ERA-Interim atmospheric model and reanalysis system has

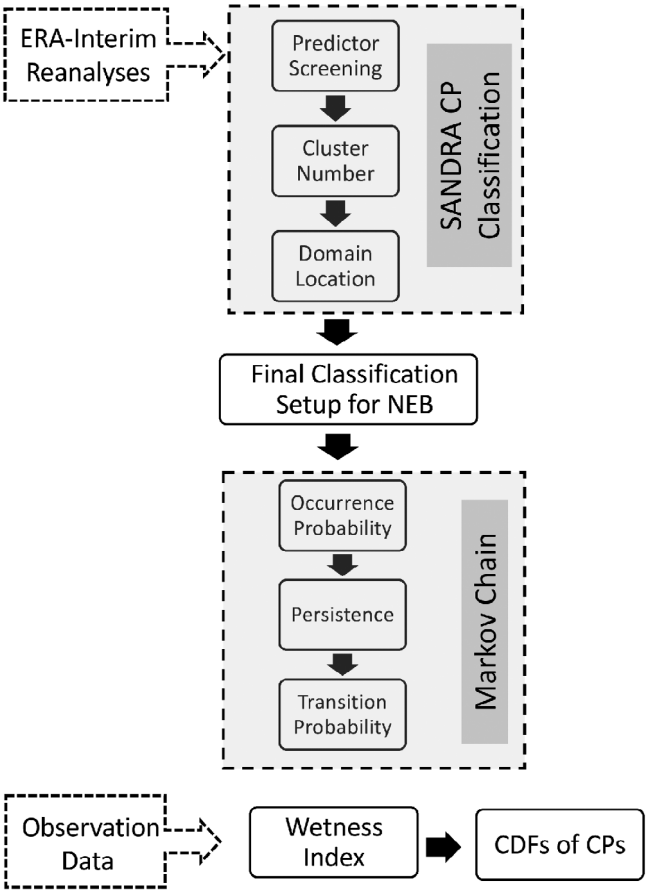

FIGURE 4 Flow chart of applied methods and data. The applied work flow (from top to bottom) is indicated by the arrows, the main methods are given in the grey shaded boxes, that is, the SANDRA CP classification and the Markov chain approach. Both methods are applied based on ERA-Interim data. The CPs are finally linked to local weather conditions by calculating the wetness index and the CP-conditional CDFs, based on observation data

used cycle 31r2 of ECMWF's Integrated Forecast System (IFS). Data are provided on a reduced Gaussian grid with approximately uniform horizontal 79 km spacing (approximately 0.75° of latitude and longitude), at 60 vertical levels (with top level at 0.1 hPa). A detailed description can be found in Dee *et al.* (2011). ERA-Interim is a reanalysis product, frequently used for validation purposes and known to be skillful in representing atmospheric processes (Lin *et al.*, 2014). Note that we did not use the more recent high-resolution (approximately 31 km) ERA5 product by intention, since the higher resolution would have tremendously increased the computational demands without expected significantly improved results. This is due to the fact that the spatial variability in the upper-level atmospheric fields (as usually applied for CP classifications) is relatively low.

TABLE 1 Station IDs, coordinates, and missing values (%)

Station ID	Name	Lat	Lon	Missing values
101	JACOBINA	−11.2	−40.5	7.7
102	REMANSO	−9.6	−42.1	3.8
110	JURAMENTO	−16.8	−43.7	12.2
114	IBIRITE	−20.0	−44.0	13.9
115	SAO MATEUS	−18.7	−39.9	9.5
120	JOAO PINHEIRO	−17.7	−46.2	11.2
127	ITABAIANINHA	−11.1	−37.8	17.0
135	BOM JESUS DO PIAUI	−9.1	−44.1	14.9
161	BOM JESUS DA LAPA	−13.3	−43.4	15.3
166	PAULISTANA	−8.1	−41.1	14.6
171	MONTE AZUL	−15.2	−42.9	12.4
179	POSSE	−14.1	−46.4	8.8
182	BARRA	−11.1	−43.2	10.5
183	CAETITE	−14.1	−42.5	17.4
184	LENCOIS	−12.6	−41.4	15.1
202	PAO DE ACUCAR	−9.8	−37.4	17.3
203	DIAMANTINA	−18.2	−43.6	12.8
207	PESQUEIRA	−8.4	−36.8	9.3
244	PROPRIA	−10.2	−36.8	8.7
248	MONTES CLAROS	−16.7	−43.8	20.0
255	VALE DO GURGUEIA CRISTIANO CASTRO	−8.4	−43.7	13.0
256	JANAUBA	−15.8	−43.3	8.7
265	PIRAPORA	−17.4	−44.9	18.2
290	CORRENTINA	−13.3	−44.6	7.6
297	JANUARIA	−15.4	−44.0	6.4
309	ITAMARANDIBA	−17.9	−42.9	7.8
334	UNAI	−16.4	−46.9	7.5
364	ITUACU	−13.8	−41.3	14.1
467	BURITIS	−15.5	−46.4	17.7
744	PATROCINIO	−19.0	−47.0	14.5

The predictor variables used for the CP classification (see Table 2) are daily ERA-Interim data at 12:00 UTC. The number of grid cells used in the classification experiments depends on the domain size, for the final classification setup in the domain 10°N–30°S and 60–20°W, 54 × 54 grid cells have been used.

3.1.2 | Observation data

The precipitation data used in this study consists of meteorological stations from INMET (National Institute for Meteorology), ANA (National Water Agency) and various other meteorological state agencies such as Meteorology and Water Resource Center of Ceara State (FUNCEME) in northeastern Brazil). FUNCEME has collected and provided the data for this study, and has been responsible to quality check and correct the data.

From the provided observation data, a small subset of 30 stations within the domain 8–21°S and 47–36°W have been pre-selected for evaluating the CP classification on local scale. The selection has been made based on the fraction of allowed missing values not exceeding 20% during the period 1980 to 2016. The location of the selected stations and the fraction of missing values (%) is shown in Figure 5. The exact coordinates and station names are given in Table 1.

TABLE 2 Explained cluster variance (ECV) [0, ..., 1] of the performed SANDRA classification experiments using the domain 10°N–30°S and 60–20°W and geopotential height (GP) in 1,000, 700, 500, and 300 hPa), mean sea level pressure (MSLP), and horizontal wind component (UWND) in 700 hPa (upper part)

Number CPs	5	6	7	8	9	10	11	12	13	14	15	16
Predictor variables												
$GP_{1,000}$						0.69						
GP_{700}						0.61						
GP_{500}						0.61						
GP_{300}						0.64						
$GP_{1,000} + GP_{700}$						0.60						
$GP_{1,000} + GP_{500}$						0.54						
$GP_{1,000} + GP_{300}$						0.57						
(MSLP + $UWND_{700}$)			0.65	0.67	0.68	0.69	0.70	0.71	0.72			
($GP_{1,000}$ + $UWND_{700}$)	0.61	0.63	0.65	0.67	0.68	0.69	**0.70**	0.71	0.72	0.72	0.73	0.73
($GP_{1,000}$ + $UWND_{700}$) 10°N–30°S and 60–10°W						0.67						
($GP_{1,000}$ + $UWND_{700}$) 10°N–30°S and 60–0°W						0.65						
($GP_{1,000}$ + $UWND_{700}$) 10°N–30°S and 60°W–10°E						0.65						
($GP_{1,000}$ + $UWND_{700}$) 15°N–35°S and 65–15°W						0.59						
($GP_{1,000}$ + $UWND_{700}$) 10°N – 40°S and 60–10°W						0.59						
($GP_{1,000}$ + $UWND_{700}$) 10°N–40°S and 70–10°W						0.56						

Note: Domain experiments by using GP in 1,000 hPa and UWND in 700 hPa with different domain extensions (lower part).

In semi-arid NEB, improved knowledge of the expected precipitation amounts a few months ahead may help water managers for decision support. CP classification can potentially contribute in various ways: it can be used as a tool for statistical downscaling of GCMs, or as a CP-conditional bias correction tool for GCMs and RCMs on various time scales (short-, seasonal-, and long-term scales). The SANDRA CP classification approach can be used therefore and is described in the following subsection. Next, the setup for the classification for the study region is given, followed by a description of the data used in this study.

Note that the period for the circulation pattern classification as well as for the analyses of the results is split into a calibration period (1980–2009) and an independent, but shorter validation period (2010–2016).

3.2 | Methods

3.2.1 | SANDRA circulation pattern classification

Simulated ANnealing and Diversified RAndomization clustering (SANDRA) cluster analysis, developed by Philipp *et al*. (2007), is applied for the NEB for the first time. It has been selected due to the improved performance of the simulated annealing algorithm compared to the conventional k-means classification (e.g., Lutz *et al*., 2012). The latter has no strategy to avoid of getting stuck in local optima during the optimization procedure. Simulated annealing has been designed to better approximate the global optimum, that is, it does not converge to a local optimum which cannot be left anymore during the optimization procedure (Philipp *et al*., 2010; Bárdossy *et al*., 2015). The method allows data objects to be assigned to a wrong cluster during the iteration process, meaning that the object is not necessarily assigned to its closest cluster centroid. Each object can be classified into a wrong cluster, if the acceptance probability P is larger than a random number between 0 and 1, where P is given by:

$$P = \exp\left(\frac{\mathrm{ED}_{\mathrm{old}} - \mathrm{ED}_{\mathrm{new}}}{T}\right), \quad (1)$$

where

$$\mathrm{ED} = \sqrt{\|x_1 - x_2\|^2}. \quad (2)$$

All data objects x are normalized before the Euclidean distance ED is calculated. $\mathrm{ED}_{\mathrm{old}}$ is the ED between the

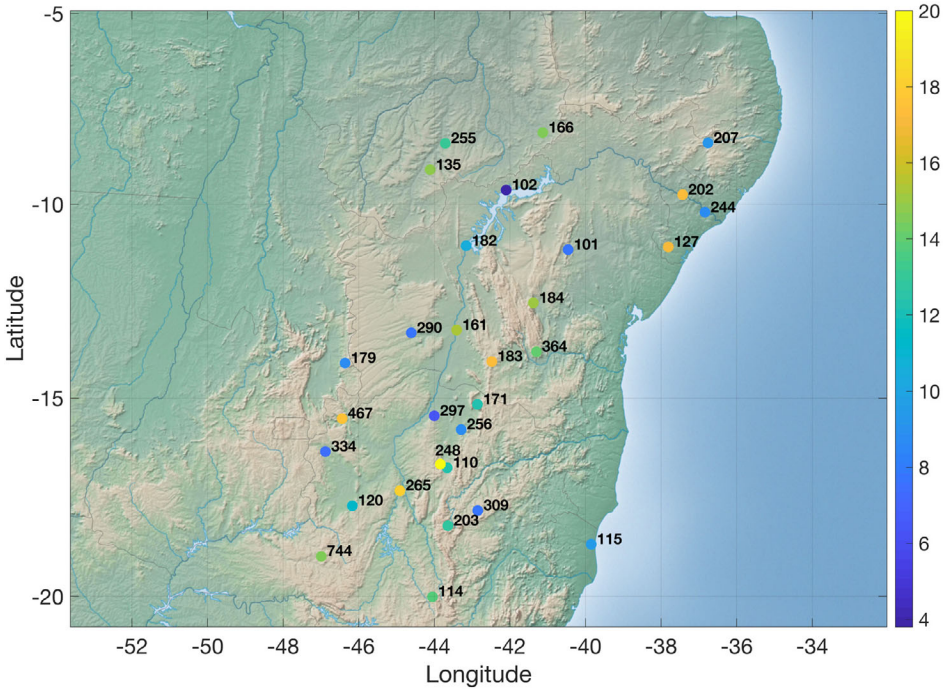

FIGURE 5 Map showing the station ID numbers and the number of missing values (%) for each station. The exact coordinates and number of missing values can be obtained from Table 1

data object and the old cluster, ED_{new} is the ED to the potentially new cluster, T is a control parameter (the annealing temperature) that is reduced after each iteration step by a constant factor a, the so-called cooling rate:

$$T_{i+1} = a \times T_i, \tag{3}$$

where i is the iteration step. The initial value of T is automatically chosen for each run by the used clustering software and its typical value is around 10^8–10^9. A constant value of 0.99 is used for the cooling rate a for all iteration steps. With the above algorithm and setting of values, the initial iteration steps of a single classification run are often characterized by wrong assignments due to the high-annealing temperature value. As annealing temperature decreases in the following iteration steps, the probability of wrong assignments decreases accordingly. During the final iteration steps reassignments mostly take place if the resulting distance between object and cluster centroid is smaller than the current distance. The process of reassignments repeats until no further changes, neither wrong nor right assignments, are taking place. The limit of iterations is infinite, thus the process repeats until convergence is reached. Then, the algorithm calculates the explained cluster variance (ECV) of the current clustering for comparison of different runs.

The classification is repeated a predefined number of times with randomized starting partitions of the data (called diversified randomization, here: 1,000 times) to overcome issues of the simulated annealing algorithm with respect to the optimization starting position. The run with the best ECV is chosen as final result of the SANDRA clustering.

The final classification setup is obtained by evaluating the performance of multiple runs with different predictor variables (i.e., predictor screening), different cluster numbers, and different domain locations (Table 2, upper part). An 11-cluster-solution of both, the horizontal wind at 700 hPa ($UWND_{700}$) and the geopotential at 1,000 hPa ($GP_{1,000}$) has been selected as suitable solution. An 11-cluster-solution meets the predefined threshold of 0.7 (corresponding to 70%) of the ECV while keeping the number of classes relatively small. A too large number of classes is expected to reduce the sample size for certain CPs and would thus hinder to derive robust results in the CP-conditional analyses. The ECV in relation to an increasing number of classes is describing a logarithmic function. An increase of the cluster numbers from 11 to 19 would result in an increase of ECV from 0.7 to 0.75 only (Figure 6). Note that a classification based on the mean sea level pressure (MSLP) combined with the $UWND_{700}$ would result in very similar performances.

The classification has been tested using different domain settings for the two predictand variables $GP_{1,000}$ and $UWND_{700}$ and a 10-cluster solution (see Table 2, lower part). Based on the performed domain settings, the domain between 10°N–30°S and 60–20°W is found to deliver good results.

In SANDRA, the number of clusters used for the classification has to be predefined. The selection of the numbers of clusters depends on different criteria and can be understood as a trade-off between the interpretability of

the cluster centroids and the overall explained cluster variance (ECV) of the predictor field(s). The ECV is defined as:

$$\mathrm{ECV} = 1 - \frac{\mathrm{WSS}}{\mathrm{TSS}}, \tag{4}$$

where WSS is the so-called within-cluster sum of square, and TSS is the total sum of squares. WSS is calculated as follows:

$$\mathrm{WSS} = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - z_i\|^2, \tag{5}$$

where x denotes all data objects who belong to the Cluster C_i, z_i refers to the ith corresponding cluster centroid, and k denotes the number of clusters. More technical information about the SANDRA classification is given in Philipp *et al.* (2007, 2010) and Hoffmann and Schlünzen (2013).

A small number of clusters increase the interpretability while reducing the fraction of the ECV. Higher cluster numbers may also increase the possibility to better link the prevailing circulation with unusual conditions, such as dry or wet conditions.

3.2.2 | Markov chains

The occurrence and sequence of CPs can be modelled as a Markov chain, which is a stochastic model describing the sequence of possible events in which the probability of each event depends only on the state attained in the previous event. A Markov Chain can be used for describing systems that follow a chain of linked events such as, a sequence of rainfall events or, as used here, a sequence of CPs. A Markov chain consists of a finite set of states (e.g., dry or wet, different CPs), whose transitions, that is, the change from a certain CP state into another (e.g., from CP1 to CP2) occurring with certain (transition) probabilities. Markov chains are frequently applied in modelling and statistical downscaling precipitation (e.g., Robertson and Smyth, 2003; Laux *et al.*, 2009; Greene *et al.*, 2011). In this study, occurrence probabilities of the CPs are calculated. Moreover, their persistences and transition probabilities and their transition pathways are derived. Graphically, a Markov chain can be represented by a directed graph, which illustrates transition probabilities and pathways. The states are represented by vertices, and transitions are described by arrows. The Markov chain is strongly connected if there is a directed path from each vertex to every other vertex (e.g., Gansner *et al.*, 1993).

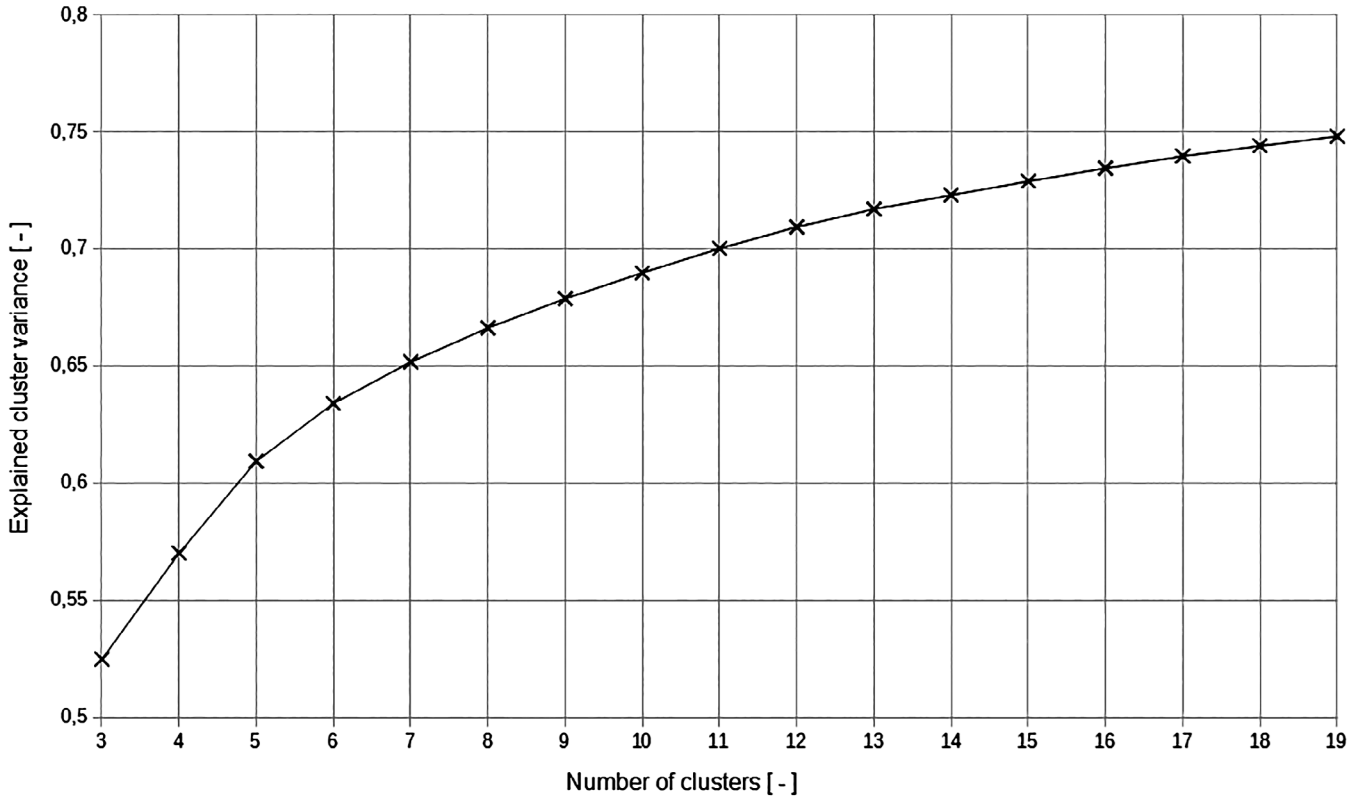

FIGURE 6 Explained variance of joint horizontal wind component at 700 hPa and geopotential (GP) at 1,000 hPa as a function of the cluster number

3.2.3 | Linking CPs to precipitation

In order to study the impact of the CP on the weather state (i.e., wet or dry), the wetness index I_{wet} (Bárdossy, 2010) is applied. It is defined as follows:

$$I_{\text{wet}_i}(\text{CP}) = \frac{z_{\text{d},i}(\text{CP})}{\bar{z_{\text{d},i}}}, \tag{6}$$

where $z_{\text{d},i}(\text{CP})$ is the normalized daily (d) precipitation amount of a certain *CP* at station *i* and $\bar{z_{\text{d},i}}$ is the overall mean daily precipitation amount at station *i*, irrespectively of the prevailing CP. The wetness index thus offers one possibility to compare the clusters with respect to their precipitation amount (wetness) between different locations (stations). A value higher than unity of any given CP means that the CP is much wetter than the average precipitation, and a value significantly smaller than unity that the CP is much drier than the average precipitation at this station. In turn, a value around unity means that the CP is representing normal, that is, approximately average conditions.

Another possibility to link the CPs to the weather state is to analyse the CP-conditional cumulative density functions (CDFs). This indicates whether or not the different CPs have discriminative power to describe the weather state, and thus can be applied in applications such as bias correction or statistical downscaling.

4 | RESULTS

In the following, the derived CPs are grouped according to their major atmospheric conditions. The CP are then described according to their major characteristics and their relation to wet and dry states in NEB, as analysed based on the wetness index I_{wet}. Then, the occurrence frequencies, the persistences, and the transition pathways and -probabilities of the CPs are analysed. Finally, the value of the classification is discussed with respect to the development of a CP-conditional bias correction.

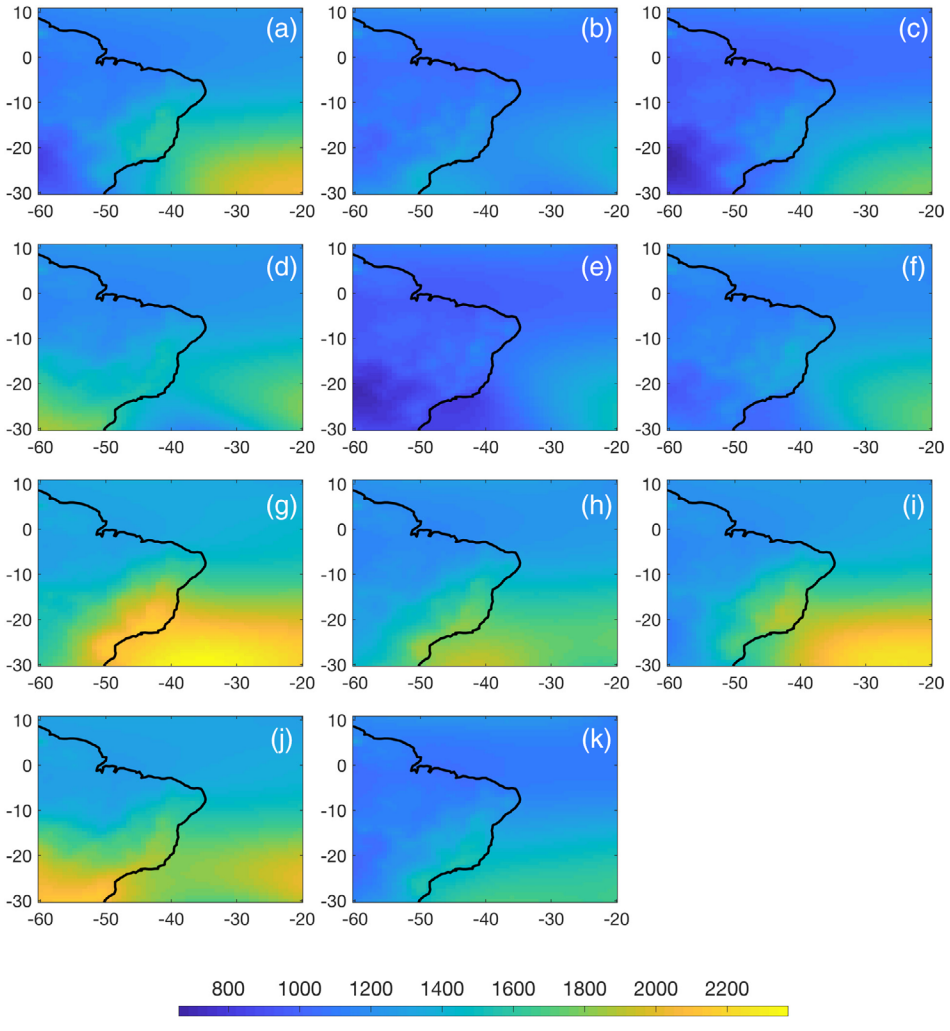

FIGURE 7 Cluster centroids of geopotential (GP) at 1,000 hPa·($m^2 \cdot s^{-2}$) of the selected 11-cluster-solution

4.1 | Circulation patterns (CPs) of the cluster solution

The classification scheme is based on the dominant centers of $GP_{1,000}$ within the domain (Figure 7), which are virtually existing only in the southern, that is, the subtropical part of the domain. The CPs can be grouped into three major groups (A, B, C), corresponding to a CP with a clear contrast between a maritime High- and a continental Low-pressure system (group A), a CP with a maritime High and continental ridge (group B), and a CP with a High pressure bridge (group C). The clusters can then be further subdivided according to the magnitude/degree of the derived groups (Table 3) as well as in combination with $UWND_{700}$, as shown in Figure 8. Thus, in this study, we combined atmospheric information from lower (1,000 hPa) and middle troposphere (700 hPa), which is often seen as more credible for, for example, determining convective events, caused by high reaching clouds (e.g., Dayan *et al*., 2012). Since the grouping is

TABLE 3 Unique character combination consisting of a 6-character string to label and characterize the derived CPs based on the $GP_{1,000}$ field

Position in name string	Abbreviations		
1. & 4.	m: Maritime	c: Continental	b: Bridge
2. & 5.	H: High	L: Low	E: Extension
3. & 6.	w: Weak	m: Moderate	s: Strong

Note: Positions 1 & 4 define the position of the Centre of action, 2 & 5 the group or type of the Centre of action, and 3 & 6 their degree/magnitude. As an example, mHscLm means *maritime High strong*, *continental Low moderate*.

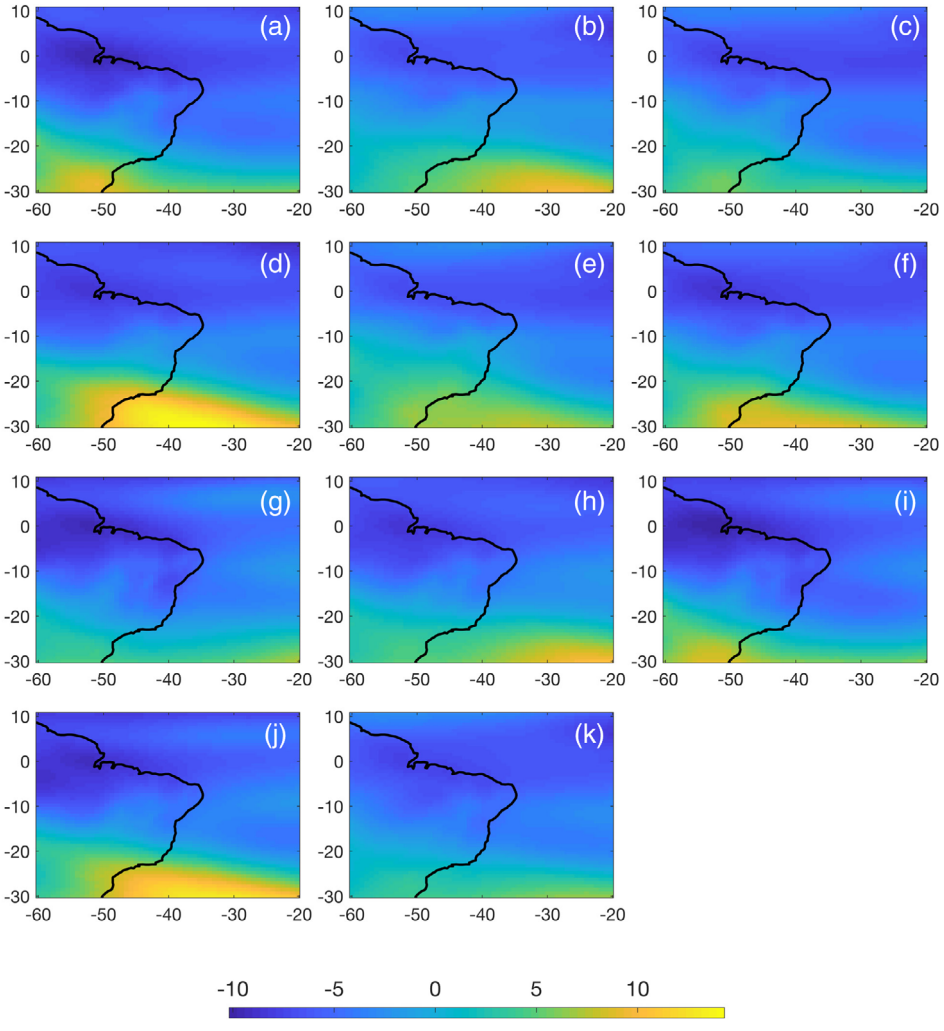

FIGURE 8 Cluster centroids of horizontal wind speed at 700 hPa·(m·s^{-1}) of the selected 11-cluster-solution

based on the atmospheric conditions and does not include the local weather situation, it cannot be fully avoided that wet and dry conditions can occur within a certain group. Subsequent calculations are done based on the individual 11 CPs.

In the following, the assignment of the 11 CPs to the three groups is described and each CP is briefly characterized:

4.1.1 | Group a: CP With a clear contrast between a maritime high- and a continental low-pressure system

- CP1 (mHscLm, *maritime High strong*, *continental Low moderate*): CP1 (Figures 7a and 8a) is characterized by a strong subtropical High over the South Atlantic, which extends to the coastal regions of Central Brazil

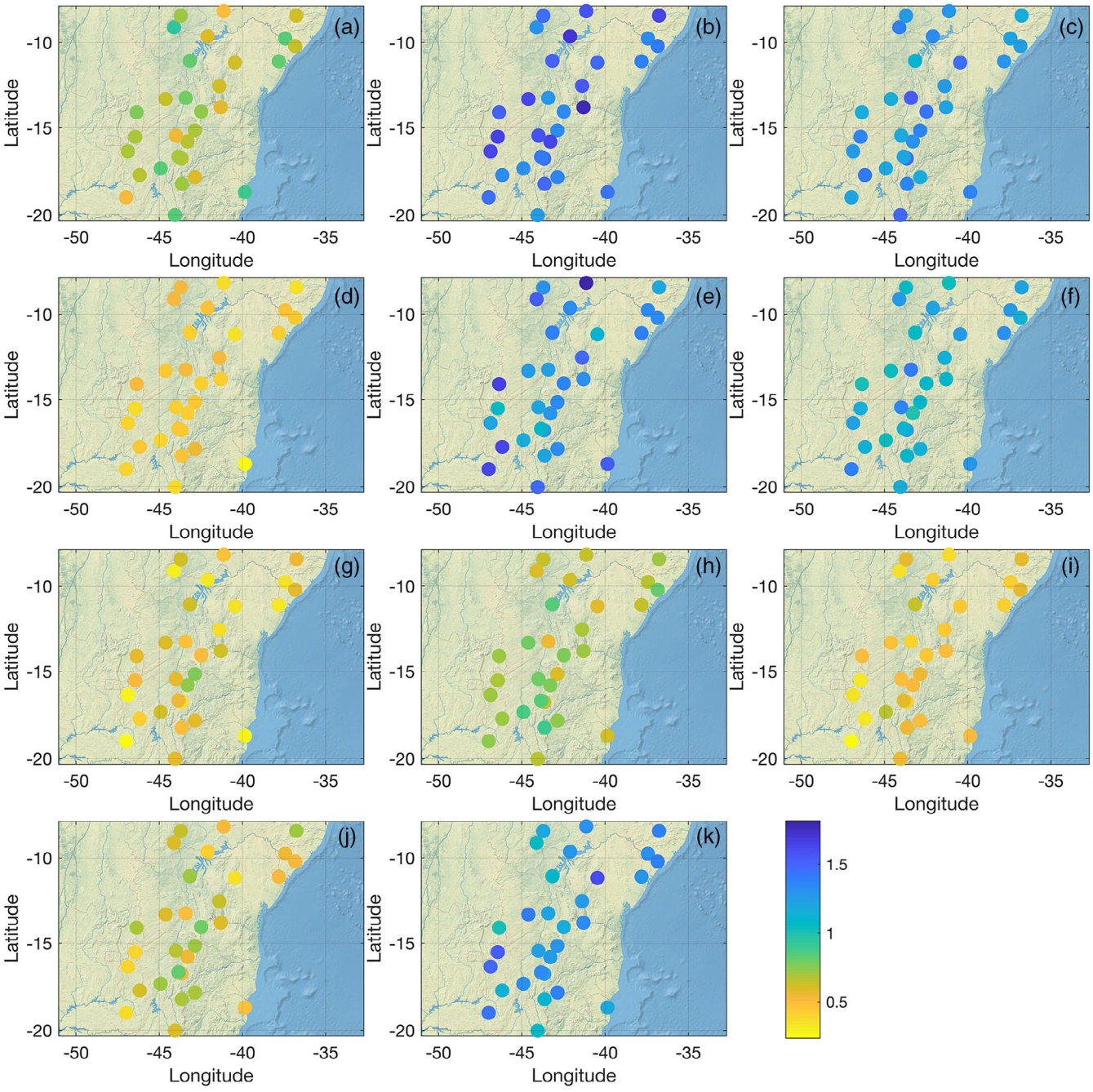

FIGURE 9 Annual consideration of the wetness index (WI) for 30 selected stations based on the selected 11-cluster-solution

and a continental Low over Southwest Brazil. Only marginally differentiated $GP_{1,000}$ fields prevail over the Amazon and the equatorial Atlantic. The maximum westerly flow of the $UWND_{700}$ is in the coastal region of South Brazil and the maximum easterly flow is in the coastal region of the Amazon. This CP is linked with wetness indices *WIs* ≈ 1 or smaller in NEB, which leads to rather dry to average rainfall amounts (Figure 9a).

- CP3 (mHmcLs, *maritime High moderate*, *continental Low strong*): CP3 (Figures 7c and 8c) is characterized by a relatively weak subtropical High over the South Atlantic and a strong continental Low over Southwest Brazil. Like for CP1, only marginally differentiated $GP_{1,000}$ fields prevail over the Amazon and the equatorial Atlantic. Overall, the horizontal wind component is moderate and the westerly flows are shifted to the Southwest of the continental region, the easterly flows prevail over the Amazon and the equatorial Atlantic. CP3 indicates rather wet conditions in NEB (*WIs* > 1), see Figure 9c.
- CP5 (mHwcLs, *maritime High weak*, *continental Low strong*): CP5 (Figures 7e and 8e) is characterized by a subtropical High shifted far into the South Atlantic and a strong continental Low over Southwest Brazil as well as low pressure in front of the coast of southern Brazil. In general, the magnitude of the $GP_{1,000}$ fields are low in magnitude and only poorly differentiated over the Amazon and the Equatorial Atlantic. There is a moderate westerly flow in front of the south Brazilian coast and over southern Brazil as well as a moderate easterly flow over the Amazon and the Equatorial Atlantic. Similarly to CP3, CP5 indicates wet conditions (*WIs* >> 1, see Figure 9e).
- CP6 (mHmcLw, *maritime High moderate*, *continental Low weak*): CP6 (Figures 7f and 8f) is characterized by a moderately pronounced subtropical High shifted to the South Atlantic and a weak continental Low over Southwest Brazil. The $GP_{1,000}$ fields are low in magnitude and only poorly differentiated over the Amazon and the Equatorial Atlantic. The maximum of the westerly flow is in front of the coast of southern Brazil,

TABLE 4 Occurrence frequencies of CPs during the year of the selected 11-cluster-solution for the calibration period 1980–2009 and the validation period 2010–2016 (in brackets)

	J	F	M	A	M	J	J	A	S	O	N	D
CP1	4.4	5.7	3.2	4.7	8.5	13.3	9.9	14.6	**20.7**	**23.8**	11.4	4.5
	(9.7)	(7.7)	(5.9)	(5.3)	(6.5)	(7.8)	(8.3)	(9.3)	(11.2)	(12.8)	(12.7)	(12.2)
CP2	9.8	13.3	**25.4**	**21.6**	11.8	0.6	0.2	0.1	2.1	6.6	9.6	11.8
	(7.8)	(8.4)	(11.7)	(14.4)	(13.0)	(11.3)	(9.6)	(8.4)	(7.7)	(7.4)	(8.2)	(8.1)
CP3	**24.6**	18.4	12.9	10.4	4.8	0.3	0.6	0.8	2.0	13.8	**27.7**	**30.3**
	(**26.3**)	(**31.6**)	(**25.3**)	(**21.5**)	(17.8)	(15.2)	(13.1)	(11.5)	(10.4)	(10.6)	(12.0)	(13.5)
CP4	0.3	2.0	1.2	5.6	19.2	12.8	8.2	8.4	13.0	5.2	1.8	0.1
	(0.5)	(0.5)	(1.7)	(3.3)	(7.2)	(7.6)	(7.7)	(7.8)	(8.0)	(7.9)	(7.4)	(6.8)
CP5	**24.9**	15.3	14.5	5.9	2.3	0.4	0.0	0.1	1.0	5.3	17.1	**24.5**
	(14.7)	(12.8)	(11.6)	(9.1)	(7.3)	(6.1)	(5.2)	(4.5)	(4.1)	(4.5)	(5.6)	(7.6)
CP6	**22.4**	**21.2**	13.2	8.9	5.3	3.8	4.0	2.3	5.6	10.6	12.2	16.8
	(**25.3**)	(19.8)	(19.1)	(15.9)	(13.4)	(12.0)	(10.5)	(9.3)	(8.8)	(8.8)	(8.6)	(9.5)
CP7	0.0	0.0	0.0	0.0	3.0	11.0	**21.9**	17.3	7.4	0.8	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.7)	(3.2)	(6.1)	(7.8)	(8.1)	(7.2)	(6.6)	(6.0)
CP8	1.2	2.2	2.8	9.8	**20.4**	19.0	11.0	10.2	16.8	13.0	3.0	1.5
	(1.4)	(0.7)	(2.8)	(4.4)	(8.3)	(9.7)	(9.5)	(9.4)	(9.9)	(10.1)	(9.5)	(8.8)
CP9	0.1	0.0	0.1	0.3	5.0	19.1	**23.4**	**30.4**	18.7	7.1	0.9	0.2
	(0.5)	(0.2)	(0.3)	(0.2)	(0.8)	(2.9)	(6.8)	(9.5)	(10.8)	(10.4)	(9.6)	(8.8)
CP10	0.0	0.0	0.0	0.9	4.1	17.0	**20.2**	14.9	7.3	1.2	0.1	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.4)	(3.2)	(5.1)	(6.8)	(6.7)	(6.1)	(5.6)	(5.1)
CP11	12.3	**21.8**	**26.7**	**32.0**	15.5	2.7	0.5	0.9	5.4	12.8	16.2	10.2
	(13.8)	(18.3)	(**21.5**)	(**25.8**)	(**24.6**)	(**21.1**)	(18.1)	(15.9)	(14.3)	(14.2)	(14.2)	(13.6)

Note: Frequencies >20% are bolded.

and a weak maximum of the easterly flow prevails in the region of the Amazonian coast. CP6 shows *WIs* > 1, and thus rather wet conditions in NEB (Figure 9f).

4.1.2 | Group B: CP With a maritime high and a continental ridge

- CP2 (mHwcLw), *maritime High weak*, *continental Low weak*: CP2 (Figures 7b and 8b) CP2 shows *WIs* >> 1 and leads to wet conditions in NEB (Figure 9b).
- CP7 (mHscEs, *maritime High strong*, *continental Extension strong*: CP7 (Figures 7g and 8g) is characterized by a pronounced subtropical High over the South Atlantic extending to continental continental South Brazil. The magnitude of the $GP_{1,000}$ field is high, but poorly differentiated over the Amazon and the Equatorial Atlantic. The westerly flow is weak and shifted to the South. The easterly flow has a pronounced maximum over the Amazon. This CP, in general, can be linked to dry conditions (Figure 9g).
- CP8: mHmcEm, *maritime High moderate*, *continental Extension moderate*: CP8 (Figures 7h and 8h) captivates with a moderately pronounced subtropical extending to continental continental South Brazil. The magnitude of the $GP_{1,000}$ is relatively high, but only poorly differentiated over the Amazon, the Equatorial Atlantic and Southwest Brazil. CP8 is linked to normal (average) conditions in NEB (*WIs* ≈ 1), see Figure 9h.
- CP9: mHscEm, *maritime High strong*, *continental Extension moderate*: CP9 (Figures 7i and 8i) is characterized by a pronounced subtropical High with the maximum over the South Atlantic and a high-pressure ridge over Central Brazil. In general, the $GP_{1,000}$ is relatively high, but poorly differentiated over the Amazon, the Equatorial Atlantic and Southwest Brazil. The westerly flows are moderate with their maximums shifted to Southwest Brazil. The easterly flows are pronounced and have their maximums over the Amazon. CP9 is linked with anomalous dry conditions (*WIs* << 1, see (Figure 9i).
- CP11: mHwcEw, *maritime High weak*, *continental Extension weak*: CP11 (Figures 7k and 8k) shows a subtropical High over the South Atlantic, extended to continental southern Brazil. In general, the $GP_{1,000}$ fields shows low values and is only poorly differentiated over the Amazon, the Equatorial Atlantic and Southwest Brazil. The westerly flow is weak only and the maximum is shifted to the South Atlantic. The easterly flow is moderate with its maximum shifted to the Equatorial Atlantic. CP11 brings wet conditions to NEB (Figure 9k).

4.1.3 | Group C: CP with a high-pressure bridge

- CP4 (bHw, *bridge High weak*): CP4 (Figures 7d and 8d) is characterized by a moderate subtropical High over the South Atlantic and a pronounced continental High over Southwest Brazil. There is a tendency for a high-pressure bridge formation over the coast of Central Brazil and a lower $GP_{1,000}$ in the coastal region of South Brazil. Overall, only moderate and poorly differentiated $GP_{1,000}$ fields over the Amazon and Equatorial Atlantic prevail. The maximum of the westerly flows is shifted to the South Atlantic and the easterly flow is moderate and has its maximum over the eastern equatorial Atlantic. CP4 can be linked with dry conditions (Figure 9d).
- CP10 (bHs, *bridge High strong*): CP10 (Figures 7j and 8j) can be described by a strong high-pressure bridge with its two maximums over the South Atlantic and

TABLE 5 Average persistence (days) of the CPs for the selected 11-cluster-solution for the calibration period 1980–2009 and the validation period 2010–2016 (values for the dry and the wet season are in brackets)

CP	Calibration (1980–2009) (wet, dry)	Validation (2010–2016) (wet, dry)
CP1	1.9	2.1
	(1.7, 2.0)	(2.1 2.1)
CP2	1.9	2.0
	(1.9, 1.7)	(2.0, 1.9)
CP3	2.1	2.2
	(2.1, 1.7)	(2.3, 1.7)
CP4	1.6	1.8
	(1.3, 1.6)	(1.4, 1.8)
CP5	2.6	2.9
	(2.7, 1.8)	(3.0, 2.0)
CP6	1.7	1.9
	(1.9, 1.4)	(2.1, 1.4)
CP7	2.8	2.6
	(−, 2.7)	(−, 2.6)
CP8	1.8	1.8
	(1.6, 1.8)	(1.8, 1.7)
CP9	2.4	2.2
	(1.5, 2.4)	(1.3, 2.2)
CP10	2.0	2.0
	(1.3, 2.0)	(1.0, 2.0)
CP11	2.0	1.9
	(2.1, 1.6)	(2.0, 1.6)

Southwest Brazil. The $GP_{1,000}$ has relatively high values, but is poorly differentiated over the Amazon and the Equatorial Atlantic. The westerly flows are strong and have their maximum in front of the coast of South Brazil. The easterly flow is also strong with its maximum over the Amazon. For most of the stations, CP10 is linked with weakly dry conditions (Figure 9j).

4.2 | Seasonal consideration of selected features of the CPs

In semi-arid regions such as the NEB with a distinct dry and rainy season during the year, one would expect a certain level of discriminative power of the classification based on a seasonal consideration. The applied classification is thus being separately analysed for the rainy season from November to April, and the dry season from May to October by considering the occurrence probability, the persistence as well as the spatial distribution of the wetness index:

4.2.1 | Occurrence probability of CPs

In the 11-cluster-solution, it is found that the occurrence of certain CPs heavily depends on the time of the year (Table 4). While some of the CPs predominantly occur during the rainy season (CP2, CP3, CP5, CP6, CP11), others like CP1, CP4, CP7, CP8, CP9, CP10 predominantly occur during the dry season. CP7, CP9, and CP10 are typical dry season patterns and occur nearly exclusively during the dry season. CP5 virtually does not occur during the dry season. This indicates that the identified cluster solution based on the atmospheric fields is potentially able to discriminate between dry and wet states in NEB.

4.2.2 | Persistence of CPs

Overall, the mean persistence of the classified CPs is about 2.1 days with a standard deviation of 0.4 days for both, the calibration and the validation period (Table 5). This indicates that the obtained results are robust for independent periods.

However, the persistences vary for the dry and the wet season. Average over all CPs, the differences of the mean persistences between dry and wet season is about 0.5 days. This is true for both, the calibration and the validation period. However, there are large discrepancies for different CPs. While the differences can be approximately 0.9 days (for calibration period, and even 1.0 day for the validation period) for CP5 and CP9, only small differences can be found for CP2 and CP8 (0.2 days for calibration, 0.1 days for validation).

4.2.3 | Transition probability of CPs

Based on Markov chains, the directed graph of the CPs of the 11-cluster-solution is calculated using the transition matrices (Figure 10). This figure represents the state changes from one CP to another for the whole year as

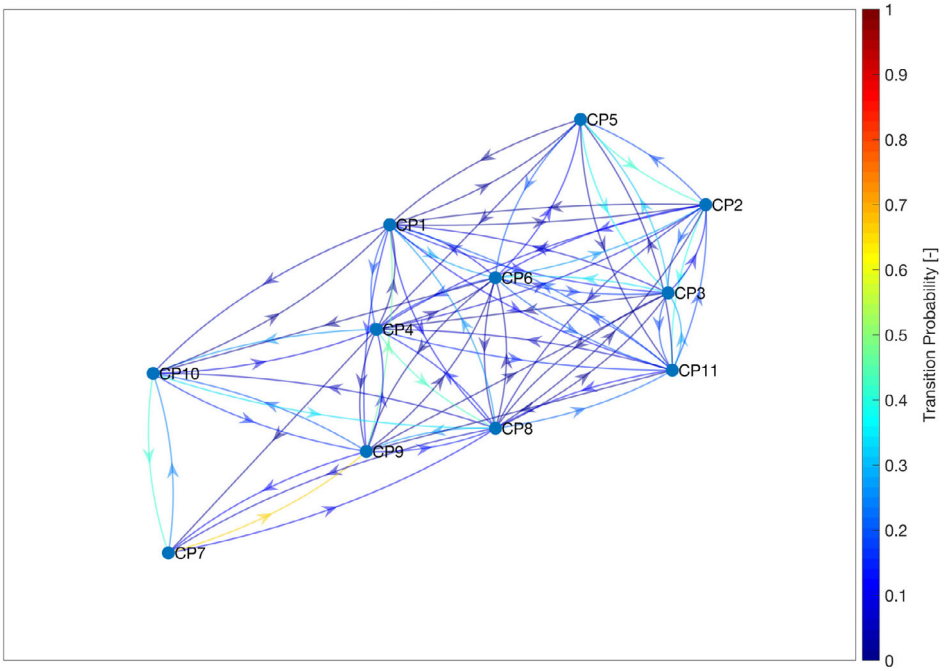

FIGURE 10 Directed graph of the selected 11-cluster-solution w/o discrimination between wet and dry season. The arrows represent the transitions from a certain CP to another, and the colours of the arrows represent the corresponding transition probabilities

well as the corresponding transition probabilities, coded as colours of the arrows. It can be seen that the graph is strongly connected, that is, most of the states (CPs) are connected to each other. On the other hand, it can be seen that some of the transition probabilities from some CPs to others are increased. This includes, for example, the transitions from CP4 to CP8, from CP7 to CP9, or from CP10 to CP7. This goes along with physically plausible transitions, that is, sequences from one CP to another. The transitions from CP4 to CP8 and CP10 to CP7 are essentially comparable shifts of patterns. In both cases a high-pressure bridge pattern transitions into a single but large high-pressure field pattern. The main difference between these transitions is the intensity of the starting high-pressure bridge and the resulting single high-pressure field. CP4 is essentially a less intense CP10 and CP8 is a less intense CP7, so the CP4 to CP8 transition can simply be considered as equal to the CP10 to CP7 transition under less intense pressure fields. The CP7 to CP9 transition reflects a change from the strongest single

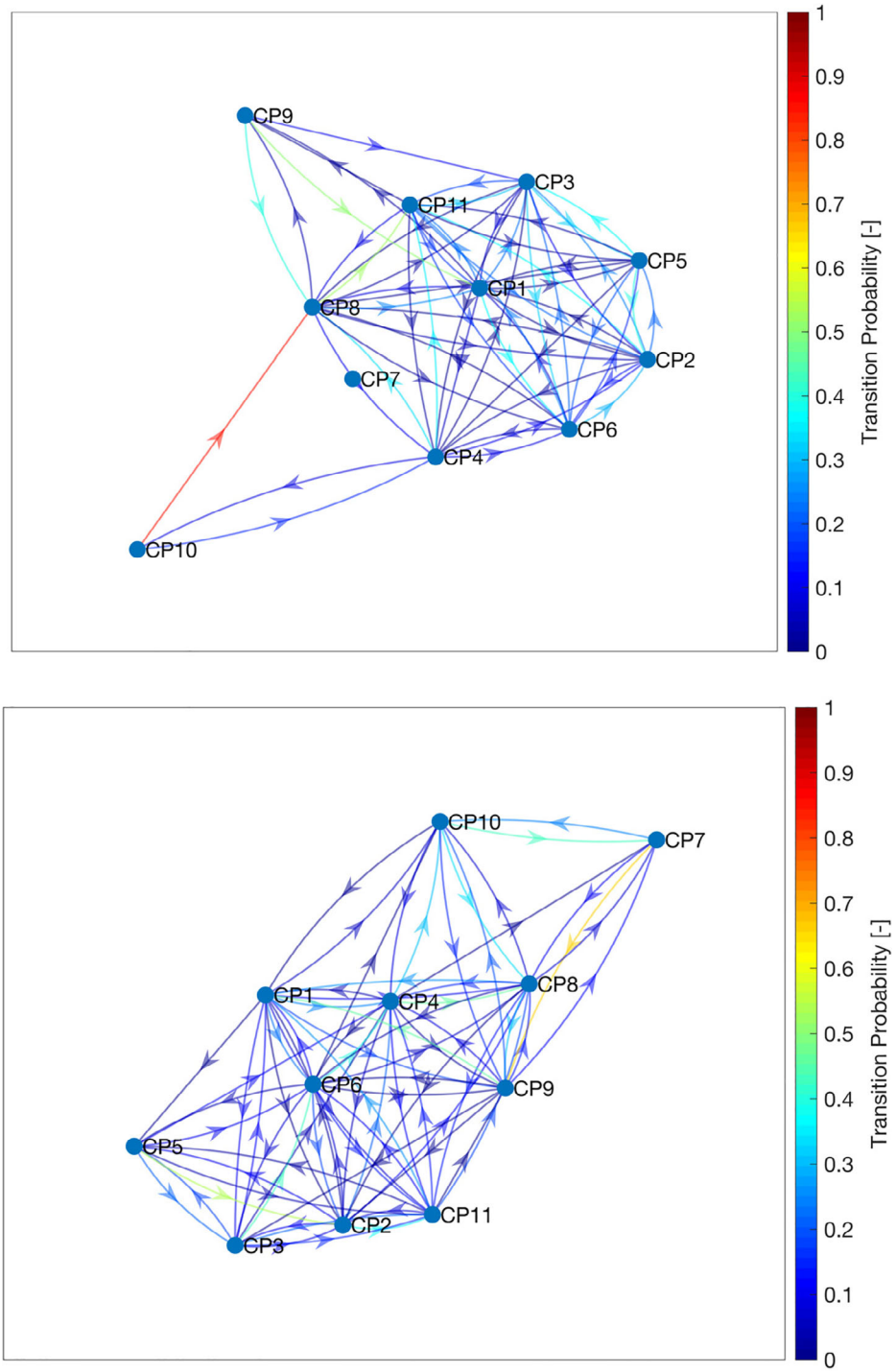

FIGURE 11 Directed graph of the selected 11-cluster-solution for the wet (top) and the dry season (bottom). The arrows represent the transitions from a certain CP to another, and the colours of the arrows represent the corresponding transition probabilities

high-pressure field pattern into a less intense version of itself. Since all of these before mentioned transitions are from one high-pressure state into another and occur in an area considered as part of the subtropical high-pressure area resulting from the tropical circulation, the increased transition probabilities are considered as plausible. A direct transition into a non-high pressure CP would imply the occurrence of a sudden change of the whole tropical circulation, which can be considered as a low-probability event. The CP10 to CP7 and the CP7 to CP9 transitions described above are a sequence of high-probability transitions. Due to their high-occurrence probabilities and the fact that these CPs are considered as dry CPs (with a low WI) this sequences in particular might be a valuable for future research on drought prediction.

The directed graphs differ significantly for the different seasons (Figure 11). For the wet season (Figure 11, top), the graph is less connected compared to that of the dry season (Figure 11, bottom). This is due to the fact that CP7 virtually does not occur during the wet season. In addition, some of the transitions occur only in one

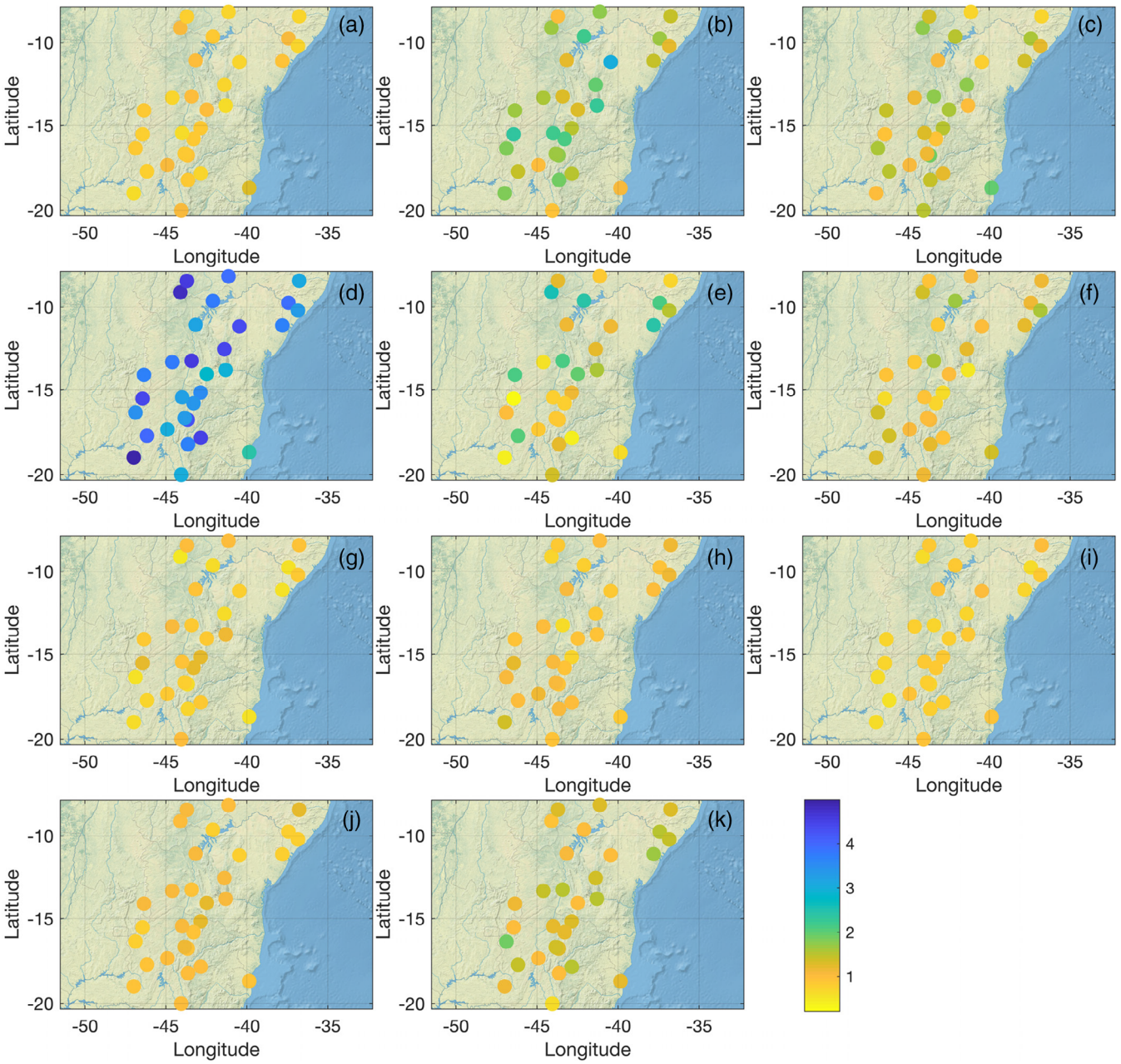

FIGURE 12 Dry season consideration of the wetness index (WI) for 30 selected stations based on the selected 11-cluster-solution

direction. This is the case for the transition from CP10 to CP8 as well as from CP9 to CP3.

4.2.4 | Wetness indices of CPs

The wetness indices (WIs) without discrimination between wet and dry season is already described in combination with the occurrence of the different CPs. Figures 12 and 13 depict the WIs for 30 different stations in NEB, separately for the dry and the wet season in the calibration period. One can see that CP4 shows extraordinary high values of WI (WI >> 1) during the dry season, but also enhanced values for CP2 and CP5, and for fewer stations also for CP11. During the wet season, on the other hand, CP4 leads to very dry conditions for all stations. CP7 does not occur and CP10 shows a highly variable pattern, which is related to the very small number of

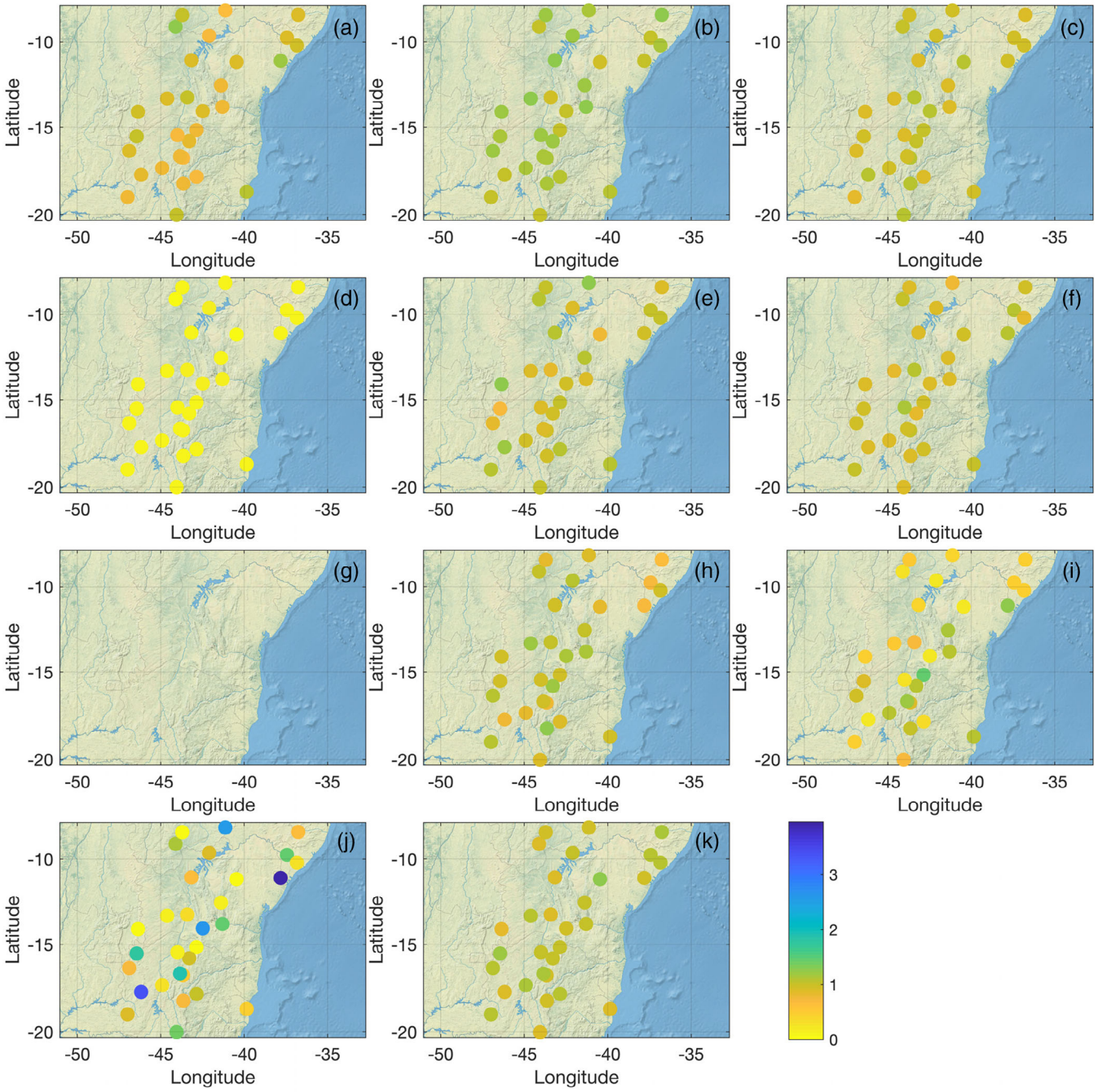

FIGURE 13 Wet season consideration of the wetness index (WI) for 30 selected stations based on the selected 11-cluster-solution. Please note that CP7 (shown in subplot g) does not occur during the wet season

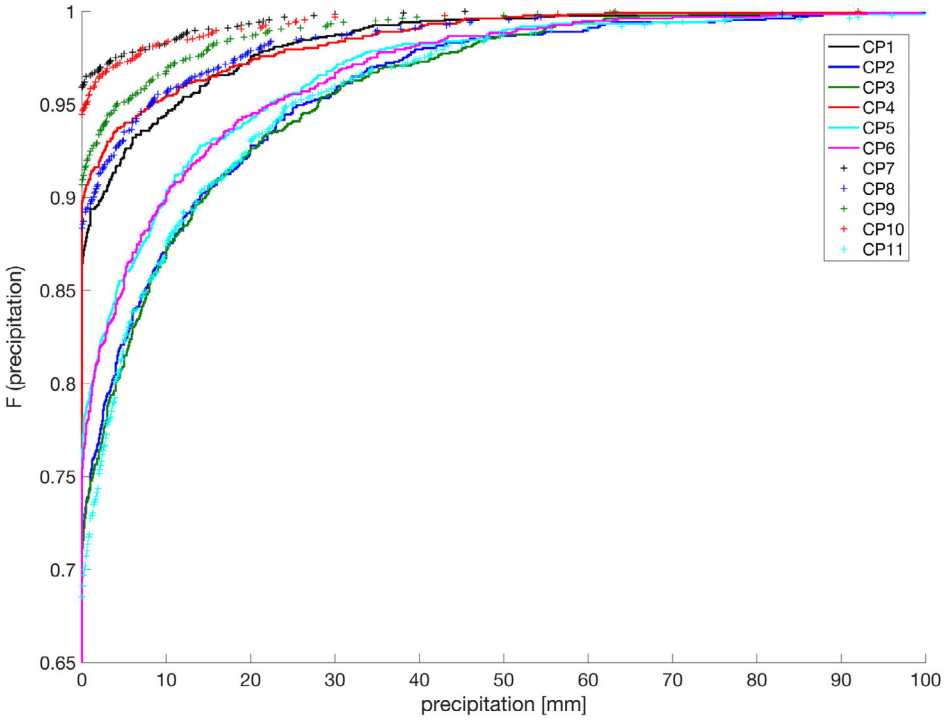

FIGURE 14 Empirical cumulative distribution function of the selected 11-cluster-solution for one arbitrarily selected precipitation gauge in the region

frequencies during the rainy season. In this case, few rainfall occurrences may cause very high WIs. For the remaining CPs, WIs slightly larger than unity are found. For both, the dry and the wet season, no clear spatial patterns of the stations across the NEB as, for example, related to topography for any given CP is found. The results are very similar and could thus be confirmed for the validation period (not shown).

4.2.5 | CDFs of CPs

The required level of discriminative power of the identified CPs is assessed for one arbitrarily selected precipitation gauge in the region. The empirical cumulative density function is shown for the different CPs for the annual consideration (Figure 14). It can be seen that wet CPs (e.g., CP2, CP3, CP5, CP6, CP11) exhibit a different shape compared to the remaining CPs that are linked to dry situations. This, in combination with different bias correction approaches, such as quantile mapping, potentially allows to derive transfer functions with a higher performance.

5 | DISCUSSION AND CONCLUSIONS

A semi-objective circulation pattern classification has been performed and analysed for the semi-arid NEB region. In this study, we followed a typical "circulation to environment" approach, in which the CP classification (here: the SANDRA algorithm) is performed as a separate step that typify significant modes of the atmospheric circulation. After the classification, the CPs are linked to environmental phenomena at surface, such as heavy precipitation events causing high-runoff episodes (e.g., Prudhomme and Genevier, 2011; Wypych *et al.*, 2018; Bednorz *et al.*, 2019) or droughts (e.g., Dayan *et al.*, 2012; Burgdorf *et al.*, 2019).

Although the classification algorithm per se is objective, a lot of subjective decisions have to be met and may heavily impact on the results. The decisions comprise the selection of suitable predictor variables, the domain size and -location, the number of cluster centroids, and others. Therefore, tremendous efforts have been made for predictor screening and domain settings, making this a valuable base for further studies in NEB. For NEB, the only available study on circulation pattern classifications the authors are aware of is Robertson *et al.* (2004). In our study, for instance, it is found that near surface pressure patterns, such as MSLP and GP, reveal better classification performances, expressed as total explained cluster variances, when combined with upper level variables, such as the horizontal wind component.

It is stressed that also the grouping of the CPs is based on subjective decisions, and these decisions are based on the atmospheric conditions exclusively, which means that the observed weather states at surface have not been considered. However, the discriminative power of the resulting groups confirmed their potential applicability: it

is found that group A (CPs 1, 2, 3, 5, 6) is the wettest (WI = 1.19 on average for all stations), followed by group B (CPs 7, 8, 9, 11) with an average WI = 0.82, and group C (CPs 4, 10), with an average WI = 0.70 (not shown). Other grouping approaches, for example, by considering the wetness index would possibly lead to different groups. For the case of this study, CP1 could be (re-)grouped into group B and CP11 into group A, thus increasing the differences in the WIs between both groups and better discriminating between dry and wet conditions in NEB. In this study, however, it is intentionally decided to group due to the atmospheric conditions exclusively. Moreover, it is worth mentioning that the grouping has been performed only to aid interpretability of the results, all considerations for potential applications are done based on the 11 individual CPs.

The findings of this study open various avenues for applications in water resources management in NEB, such as a tool for statistical downscaling of GCMs based on CPs, or a CP-conditional bias correction algorithm of RCM output. Therefore, one may capitalize on the fact that a typical "wet" CP has a different precipitation distribution in time and space compared to a "dry" CP, which has been demonstrated in this study. This may finally lead to more efficient and robust model corrections when compared to those using the full distribution for correction. A correction based on separate months does usually implicitly accounts for this, however, this may not be a suitable solution when the same transfer functions are applied in future climate impact studies since the timing of seasons might shift (Wetterhall *et al*., 2012). As such, a CP-conditional bias correction approach can be seen as one potential solution to partly overcome the caveats due to the stationary assumptions, which exist for other bias correction approaches.

In addition to that, the CP classification can also be used to study drought dynamics in the region. Global warming may not only amplify drought duration and severity, but possibly changing drought dynamics in complex ways (Burgdorf *et al*., 2019). Although droughts are recurrent phenomena (Marengo *et al*., 2017), the recent multi-annual drought from 2012 to 2018 might be an indication for a change in the drought dynamics in NEB.

Since droughts in NEB are known to be linked to the SST anomaly patterns in the Atlantic, the Pacific, and to a lesser extent in the Indian Ocean (e.g., Ward and Folland, 1991; Hastenrath, 2012; Costa *et al*., 2018), future works can focus on drought prediction using CP classification based on *SST* as predictor. Due to the strong link between *SST* anomalies and MSLP patterns, the presented SANDRA classification can also be tested for other domains covering also parts of the Pacific or the Indian Ocean. As suggested by the results of this study, near surface pressure patterns, such as MSLP or GP, reveal better classification performances, when combined with upper level variables, such as the horizontal wind component in 700 hPa. The choice of the low-level variable, that is, if the $GP_{1,000}$ or the MSLP is applied, does not affect the overall classification performance.

The development of statistical downscaling or CP-conditional bias correction approaches for NEB is beyond the scope of this paper, however, will be subject of further studies. Different characteristics with respect to occurrence, persistence, and transition probability during the dry and the wet season are promising indicators for discriminative power of the identified CPs. Therefore, it will be evaluated to which extend the occurrence- and transition probabilities as well as persistences of the CPs may help to improve statistical downscaling- and bias correction approaches. Finally, another classification strategy following the "environment to circulation" approach can be applied, in which the CP classification is considered along a specific environment-based criteria set for a particular phenomenon, such as a drought or a flood (Dayan *et al*., 2012). An example is the classification of Bárdossy and Pegram (2012).

If the focus is on drought prediction, the presented CP classification approach itself can be further improved by including information about the local weather state (e.g., precipitation) in the classification approach directly (Bárdossy, 2010). This is subject of ongoing research and may lead to better results.

ACKNOWLEDGEMENTS

The study was mainly funded by the BMBF project *Seasonal water resources management in semi-arid regions: Transfer of regionalized global information to practice* (SaWaM, project number: 02WGR1421A). The clustering software has been produced in the EU funded *COST Action 733* program and is available under: http://cost733.geo.uni-augsburg.de (last access: 10/2019). We would also like to thank the ECMWF for providing the ERA-Interim reanalysis data as well as FUNCEME for the precipitation observations. The authors also thank the IT team of the KIT/IMK-IFU for providing access to and for maintaining the HPC-environment. Finally, we wish to thank the two anonymous reviewers for their valuable comments and suggestions.

ORCID

Patrick Laux https://orcid.org/0000-0002-8657-6152
Eduardo Sávio Martins https://orcid.org/0000-0002-9858-2541
Francisco das Chagas Vasconcelos Junior https://orcid.org/0000-0002-1558-8383
Harald Kunstmann https://orcid.org/0000-0001-9573-1743

REFERENCES

Bárdossy, A. (2010) Atmospheric circulation pattern classification for south-West Germany using hydrological variables. *Physics and Chemistry of the Earth*, 35, 498–506. https://doi.org/10.1016/j.pce.2010.02.007.

Bárdossy, A. and Pegram, G. (2011) Downscaling precipitation using regional climate models and circulation patterns toward hydrology. *Water Resources Research*, 47, W04505. https://doi.org/10.1029/2010WR009689.

Bárdossy, A. and Pegram, G. (2012) Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. *Water Resources Research*, 48 (9), 1–13. https://doi.org/10.1029/2011WR011524.

Bárdossy, A., Stehlík, J. and Caspary, H.J. (2002) Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules. *Climate Research*, 23, 11–22. https://doi.org/10.3354/cr023011.

Bárdossy, A., Pegram, G., Sinclair, S., Pringle, J. and Stretch, D. (2015) Circulation patterns identified by spatial rainfall and ocean wave fields in southern Africa. *Frontiers in Environmental Science*, 3, 31. https://doi.org/10.3389/fenvs.2015.00031.

Bednorz, E., Wrzesiński, D., Tomczyk, A.M. and Jasik, D. (2019) Classification of synoptic conditions of summer floods in polish Sudeten Mountains. *Water (Switzerland)*, 11(7), 1450. https://doi.org/10.3390/w11071450.

Burgdorf, A.M., Brönnimann, S. and Franke, J.R. (2019) Two types of North American droughts related to different atmospheric circulation patterns. *Climate of the Past*, 15(6), 2053–2065. https://doi.org/10.5194/cp-15-2053-2019.

Corte-Real, J., Zhang, X. and Wang, X. (1995) Large-scale circulation regimes and surface climatic anomalies over the Mediterranean. *International Journal of Climatology*, 15(10), 1135–1150. https://doi.org/10.1002/joc.3370151006.

Costa, D.D., da Silva Pereira, T.A., Fragoso, C.R., Madani, K. and Uvo, C.B. (2016) Understanding drought dynamics during dry season in eastern Northeast Brazil. *Frontiers in Earth Science*, 4 (June), 69. https://doi.org/10.3389/feart.2016.00069.

Costa, D.D., Uvo, C.B., Da Paz, A.R., Carvalho, F.D.O. and Fragoso, C.R. (2018) Long-term relationships between climate oscillation and basin-scale hydrological variability during rainy season in eastern Northeast Brazil. *Hydrological Sciences Journal*, 63(11), 1636–1652. https://doi.org/10.1080/02626667.2018.1523614.

Dafka, S., Toreti, A., Luterbacher, J., Zanis, P., Tyrlis, E. and Xoplaki, E. (2018) On the ability of RCMs to capture the circulation pattern of Etesians. *Climate Dynamics*, 51(5-6), 1687–1706. https://doi.org/10.1007/s00382-017-3977-2.

Dayan, U., Tubi, A. and Levy, I. (2012) On the importance of synoptic classification methods with respect to environmental phenomena. *International Journal of Climatology*, 32(5), 681–694. https://doi.org/10.1002/joc.2297.

Dee, D.P., et al. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. https://doi.org/10.1002/qj.828.

Fernández-González, S., Del Río, S., Castro, A., Penas, A., Fernández-Raga, M., Calvo, A.I. and Fraile, R. (2012) Connection between NAO, weather types and precipitation in León, Spain (1948-2008). *International Journal of Climatology*, 32, 2181–2196. https://doi.org/10.1002/joc.2431.

Gansner, E.R., Koutsofios, E., North, S.C. and Vo, K.P. (1993) A technique for drawing directed graphs. *IEEE Transactions on Software Engineering*, 19, 214–230. https://doi.org/10.1109/32.221135.

Goodess, C.M. and Jones, P.D. (2002) Links between circulation and changes in the characteristics of Iberian rainfall. *International Journal of Climatology*, 22, 1593–1615. https://doi.org/10.1002/joc.810.

Goodess, C.M. and Palutikof, J.P. (1998) Development of daily rainfall scenarios for Southeast Spain using a circulation-type approach to downscaling. *International Journal of Climatology*, 18(10), 1051–1083. https://doi.org/10.1002/(SICI)1097-0088(199808)18:10<1051::AID-JOC304>3.0.CO;2-1.

Greene, A.M., Robertson, A.W., Smyth, P. and Triglia, S. (2011) Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. *Quarterly Journal of the Royal Meteorological Society*, 137(655), 347–359. https://doi.org/10.1002/qj.254.

Hastenrath, S. (2012) Exploring the climate problems of Brazil's Nordeste: a review. *Climatic Change*, 112(2), 243–251. https://doi.org/10.1007/s10584-011-0227-1.

Hoffmann, P. and Schlünzen, H. (2013) Weather pattern classification to represent the urban heat Island in present and future climate. *Journal of Applied Meteorology and Climatology*, 52 (12), 2699–2714. https://doi.org/10.1175/JAMC-D-12-065.1.

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, J. and Tveito, O.E. (2008) Classifications of atmospheric circulation patterns: recent advances and applications. *Annals of the New York Academy of Sciences*, 1146, 105–152. https://doi.org/10.1196/annals.1446.019.

James, P.M. (2007) An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe. *Theoretical and Applied Climatology*, 88, 17–42. https://doi.org/10.1007/s00704-006-0239-3.

Kousky, V.E. (1979) Frontal influences on Northeast Brazil. *Monthly Weather Review*, 107(9), 1140–1153. https://doi.org/10.1175/1520-0493(1979)107<1140:FIONB>2.0.CO;2.

Laux, P., Wagner, S., Wagner, A., Jacobeit, J., Bárdossy, A. and Kunstmann, H. (2009) Modelling daily precipitation features in the Volta Basin of West Africa. *International Journal of Climatology*, 29, 937–954. https://doi.org/10.1002/joc.1852.

Laux, P., Vogl, S., Qiu, W., Knoche, H.R. and Kunstmann, H. (2011) Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. *Hydrology and Earth System Sciences*, 15(7), 2401–2419. https://doi.org/10.5194/hess-15-2401-2011, http://www.hydrol-earth-syst-sci.net/15/2401/2011/hess-15-2401-2011.html.

Le Roux, R., Katurji, M., Zawar-Reza, P., Quénol, H. and Sturman, A. (2019) Analysis of spatio-temporal bias of weather research and forecasting temperatures based on weather pattern classification. *International Journal of Climatology*, 39(1), 89–100. https://doi.org/10.1002/joc.5784.

Lin, R., Zhou, T. and Qian, Y. (2014) Evaluation of global monsoon precipitation changes based on five reanalysis datasets. *Journal of Climate*, 27, 1271–1289. https://doi.org/10.1175/JCLI-D-13-00215.1.

Lutz, K., Jacobeit, J., Philipp, A., Seubert, S., Kunstmann, H. and Laux, P. (2012) Comparison and evaluation of statistical downscaling techniques for station-based precipitation in the Middle East. *International Journal of Climatology*, 32, 1579–1595. https://doi.org/10.1002/joc.2381.

Marengo, J.A., Torres, R.R. and Alves, L.M. (2017) Drought in Northeast Brazil—past, present, and future. *Theoretical and Applied Climatology*, 129(3-4), 1189–1200. https://doi.org/10.1007/s00704-016-1840-8.

Moron, V., Robertson, A.W., Ward, M.N. and Ndiaye, O. (2008) Weather types and rainfall over Senegal. Part I: observational analysis. *Journal of Climate*, 21(2), 266–287. https://doi.org/10.1175/2007JCLI1601.1.

Ngarukiyimana, J.P., Fu, Y., Yang, Y., Ogwang, B.A., Ongoma, V. and Ntwali, D. (2018) Dominant atmospheric circulation patterns associated with abnormal rainfall events over Rwanda, East Africa. *International Journal of Climatology*, 38, 187–202. https://doi.org/10.1002/joc.5169.

Philipp, A., et al. (2010) Cost733cat—a database of weather and circulation type classifications. *Physics and Chemistry of the Earth*, 35(9-12), 360–373. https://doi.org/10.1016/j.pce.2009.12.010.

Philipp, A., Della-Marta, P.M., Jacobeit, J., Fereday, D.R., Jones, P. D., Moberg, A. and Wanner, H. (2007) Long-term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering. *Journal of Climate*, 20(16), 4065–4095. https://doi.org/10.1175/JCLI4175.1.

Pilz, T., Delgado, J.M., Voss, S., Vormoor, K., Francke, T., Cunha Costa, A., Martins, E. and Bronstert, A. (2019) Seasonal drought prediction for semiarid Northeast Brazil: what is the added value of a process-based hydrological model? *Hydrology and Earth System Sciences*, 23(4), 1951–1971. https://doi.org/10.5194/hess-23-1951-2019.

Prein, A.F., Bukovsky, M.S., Mearns, L.O., Bruyère, C.L. and Done, J.M. (2019) Simulating North American weather types with regional climate models. *Frontiers in Environmental Science*, 7(April), 1–17. https://doi.org/10.3389/fenvs.2019.00036.

Prudhomme, C. and Genevier, M. (2011) Can atmospheric circulation be linked to flooding in Europe? *Hydrological Processes*, 25, 1180–1190. https://doi.org/10.1002/hyp.7879.

Robertson, A.W. and Smyth, S.K.P. (2003) Hidden Markov Models for Modeling Daily Rainfall Occurrence over Brazil. Tech. Rep. Tech. Rep. UCI-ICS 03-27, Information and Computer Science, University of California, Irvine, CA, 36p.

Robertson, A.W., Kirshner, S. and Smyth, P. (2004) Downscaling of daily rainfall occurrence over Northeast Brazil using a hidden Markov model. *Journal of Climate*, 17(22), 4407–4424. https://doi.org/10.1175/JCLI-3216.1.

Siegmund, J., Bliefernicht, J., Laux, P. and Kunstmann, H. (2015) Toward a seasonal precipitation prediction system for West Africa: performance of CFSv2 and high-resolution dynamical downscaling. *Journal of Geophysical Research Atmospheres*, 120 (15), 7316–7339. https://doi.org/10.1002/2014JD022692.

Sunyer, M.A., et al. (2015) Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe. *Hydrology and Earth System Sciences*, 19(4), 1827–1847. https://doi.org/10.5194/hess-19-1827-2015.

Trigo, R.M. and DaCamara, C.C. (2000) Circulation weather types and their influence on the precipitation regime in Portugal. *International Journal of Climatology*, 20(1315), 1559–1581. https://doi.org/10.1002/1097-0088(20001115)20:13<1559::AID-JOC555>3.0.CO;2-5.

Ward, M.N. and Folland, C.K. (1991) Prediction of seasonal rainfall in the north nordeste of Brazil using eigenvectors of sea-surface temperature. *International Journal of Climatology*, 11(7), 711–743. https://doi.org/10.1002/joc.3370110703.

Wetterhall, F., Halldin, S. and Xu, C.Y. (2007) Seasonality properties of four statistical-downscaling methods in Central Sweden. *Theoretical and Applied Climatology*, 87, 123–137. https://doi.org/10.1007/s00704-005-0223-3.

Wetterhall, F., Pappenberger, F., He, Y., Freer, J. and Cloke, H.L. (2012) Conditioning model output statistics of regional climate model precipitation on circulation patterns. *Nonlinear Processes in Geophysics*, 19(6), 623–633. https://doi.org/10.5194/npg-19-623-2012.

Wypych, A., Ustrnul, Z., Czekierda, D., Palarz, A. and Sulikowska, A. (2018) Extreme precipitation events in the polish Carpathians and their synoptic determinants. *Idojaras*, 122 (2), 145–158. https://doi.org/10.28974/idojaras.2018.2.3.

Zhou, J. and Lau, K.M. (2001) Principal modes of interannual and decadal variability of summer rainfall over South America. *International Journal of Climatology*, 21(13), 1623–1644. https://doi.org/10.1002/joc.700.

How to cite this article: Laux P, Böker B, Martins ES, *et al.* A semi-objective circulation pattern classification scheme for the semi-arid Northeast Brazil. *Int J Climatol*. 2020;1–22. https://doi.org/10.1002/joc.6608