



Dynamics-based regression models for the South Atlantic Convergence Zone

David Marcolino Nielsen¹ · André Luiz Belém² · Edilson Marton³ · Marcio Cataldi^{1,2}

Received: 24 August 2017 / Accepted: 19 September 2018 / Published online: 9 October 2018
© The Author(s) 2018

Abstract

The South Atlantic Convergence Zone (SACZ) plays a key role in the South American monsoon system (SAMS) precipitation regime, and accounts for around 25% of the rainfall volume over Southeast Brazil between October and April, on average, with peaks of 56% in March and 41% in January. Due to its often varying position and multi-variable structure, diagnosing or quantifying SACZ episodes may lie on subjective criteria. The present study provides a climatological characterization of different SACZ types, based on their position, and differences in dynamics analysed from 19 SAMS wet periods. A cycle in SACZ configurations is identified during the rainy season: northernmost (southernmost) episodes are more likely to occur during the onset/demise (peak) months of the SAMS. Objective SACZ indices are developed taking the principal components of the dynamics of the SACZ as explanatory variables, not including precipitation. The most sensitive threshold proposed for the identification of mean SACZ type yields simultaneous true positive rate of 86%, false alarm rate of 28%, true negative rate of 72% and false negative rate of 15%, for example. The indices have potential to bring gains in predictability to SACZ-related precipitation in Southeast Brazil, apart from allowing the objective diagnosis of SACZ episodes. The indices are also correlated with the South American precipitation dipole and may be used to identify anomalous SAMS precipitation at longer time scales as well.

Keywords SACZ · South American monsoon system · Logistic regression · Objective index · Binary classification · Precipitation anomalies

David Marcolino Nielsen is currently at Institute of Oceanography, CEN, Universität Hamburg, and International Max Planck Research School on Earth System Modelling, Max Planck Institute for Meteorology, Bundesstr. 53, 20146, Hamburg, Germany.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00382-018-4460-4>) contains supplementary material, which is available to authorized users.

✉ David Marcolino Nielsen
david.nielsen@mpimet.mpg.de

¹ Programa de Pós-Graduação em Engenharia de Biosistemas, Universidade Federal Fluminense, Niterói, Brazil

² Departamento de Engenharia Agrícola e Ambiental, Universidade Federal Fluminense, Niterói, Brazil

³ Departamento de Meteorologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

1 Introduction

Two main large-scale characteristics of the South American monsoon system (SAMS) are the reversion in direction of low-level wind anomalies between austral summer and winter followed by an also seasonal meridional movement of rainfall maxima over South America (SA) (Zhou and Lau 1998; Gan et al. 2004; Vera et al. 2006), associated with the annual cycle of mainly two different mechanisms, namely: (1) the Intertropical Convergence Zone (ITCZ) meridional shift over northern SA and (2) the South Atlantic Convergence Zone (SACZ), extending from the southern Amazonian region to central-eastern SA and the neighbouring portion of the Atlantic Ocean, drawing a characteristic quasi-stationary diagonally-oriented cloud band region in the Northwest-Southeast (NW-SE) direction (Kodama 1992, 1993; Quadro 1994).

Some of the features observed during a typical SCAZ episode are schematically represented in Fig. 1. Convection is increased over the continent during summer, leading to the

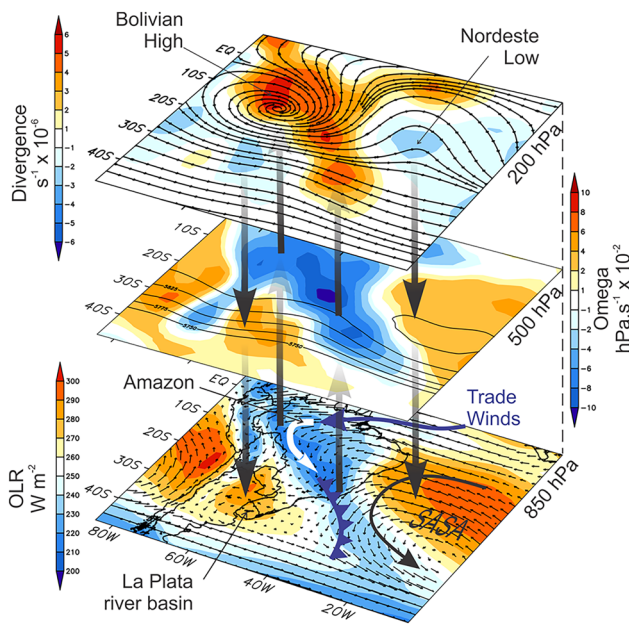


Fig. 1 Schematic representation of the main features of the SACZ dynamics. Bottom: 850 hPa wind vectors with outgoing long-wave radiation (OLR, W m^{-2}); Middle: shaded 500 hPa pressure velocity (ω , hPa s^{-1}) and contours of geopotential height (ϕ , J kg^{-1}); Top: streamlines of the 200 hPa wind and shaded horizontal divergence. Composite of all the 989 days with SACZ configurations between December 1995 and April 2015. Data from NCEP Reanalysis 2 (Kanamitsu et al. 2002) and Climanalise (CPTEC/INPE), further discussed in Sect. 2

seasonal presence of a high-tropospheric anticyclone over SA, the Bolivian High (BH) (Virji 1981; Silva Dias et al. 1983), accompanied by a cyclonic trough or closed cyclonic vortex, the “Nordeste” low (NL), normally observed around 200 hPa (Kousky and Gan 1981; Lenters and Cook 1997). Moisture convergence occurs closer to the surface, supported by Atlantic trade winds and moisture transport from the Amazonian region toward tropical SA, simultaneously with the weakening of the low-level moisture flow maxima pointing southwards to northern Argentina, characterizing the South American precipitation dipole, (Nogués-Paegle and Mo 1997). Figueroa et al. (1995) used a dry eta vertical-coordinate atmospheric model to demonstrate that these characteristic summer features (such as the BH and the NL) and a proper placement of the SACZ depend on the Amazonian latent heat source, a basic zonal flow, and the Andean topography.

Under these conditions, the configuration of a SACZ episode is triggered by the presence of a Frontal System (FS) and its associated trough in around 500 hPa, whose position also responds to remote disturbances in the South Pacific Convergence Zone (SPCZ) region, as early discussed by Grimm and Silva Dias (1995), and to the propagation of extratropical Rossby wave trains in a Pacific-South American

(PSA) teleconnection pattern (Liebmann et al. 1999, 2004; Nogués-Paegle et al. 2000; Muza et al. 2009; van der Wiel et al. 2015). Consequently, a cloud-cover tripolar pattern is seen: higher OLR values (with positive omega at 500 hPa and convergence at 200 hPa) over the La Plata river basin (LPRB) and the NL areas, surrounding the diagonal SACZ region with lower OLR (negative omega at 500 hPa and positive divergence at 200 hPa). The preferred subsidence region, however, is generally located south-westward from the SACZ cloud band, as a response to the enhanced convection in the Amazon and SACZ regions (Gandú and Silva Dias 1998), however also influenced by other tropical heat sources, which is in accordance with the bimodal, or *see-saw* pattern in precipitation between LPRB and the SACZ region, described in many studies (e.g. Nogués-Paegle and Mo 1997; Nogués-Paegle et al. 2000; Salio et al. 2002; Díaz and Aceituno 2003; Liebmann et al. 2004; Marengo et al. 2004; Mattingly and Mote 2016).

The ocean-atmospheric coupling associated with the SACZ has also been explored. Chaves and Nobre (2004) demonstrated that a colder South Atlantic sea surface temperature (SST) pattern is observed as a response to weaker solar radiative forcing during SACZ episodes, while pre-existing warmer (colder) South Atlantic SST anomalies tend to intensify (weaken) the SACZ activity. Jorgetti et al. (2014), studying associations between South Atlantic SSTs and SACZ dynamics, found that northern (southern) placements of the SACZ are associated with tropical (subtropical) warm SST anomalies, for example, and that the consequent 850 hPa circulation anomalies are in accordance with the SAMS active and break phases, as suggested by Jones and Carvalho (2002).

The wet or rainy period of the SAMS occurs typically between October and the end of March/beginning of April (Gan et al. 2004, 2005; Raia and Cavalcanti 2008; Carvalho et al. 2012) and precipitation anomalies in the SACZ region affect the most densely populated regions in Brazil (Carvalho et al. 2002; Seluchi and Chou 2009; Lima et al. 2010; Coelho et al. 2016a, b) during the SAMS wet period. In addition, about 70% of the Brazilian electrical energy sources rely on hydropower generation (ONS 2015). For this reason, relatively small variations in the SACZ position may determine which river basins are supplied with rainfall and which other regions experience drier conditions, influenced by atmospheric subsidence and surface highs. This spatial variability raises concern not only on the frequency and intensity of SACZ episodes, but also on how the SACZ is geographically placed. Therefore, a description of observed latitudinal variations of the SACZ position is provided in this study, and then also taken into account to the further development of predictive models.

Due to the aforementioned multivariable and three-dimensional structure, objectively quantifying the SACZ is

not a trivial task. Some studies have suggested indices often based on rainfall, moisture and OLR anomalies (Silva and Carvalho 2007; Carvalho et al. 2011; Ambrizzi and Ferraz 2015). Quadro et al. (2013) assessed the skill of recent global reanalysis in representing some of the main characteristic of the SAMS, in terms of its water cycle, over the LPBR and SACZ regions. They analysed precipitation, evaporation, vertically integrated moisture flux convergence, runoff, and soil moisture, which are all highly dependent on parameterization schemes. Nielsen et al. (2016) explored indices based on variables more directly derived from the atmospheric dynamical primitive equations, but were very locally limited and did not consider several fundamental systems for the SACZ formation. In the present study, all days in which a SACZ episode was configured (SACZ-days) in 19 wet periods of the SAMS (from October to April), between 1995 and 2015, are examined. Regression models are then designed based on the main features of the SACZ dynamics. Precipitation data and moisture do not compose the indices, due to their strong dependency on parameterization schemes in atmospheric models, often important sources of uncertainties. Therefore, the components of the presented indices are less-dependent, or at least more indirectly dependent, on parameterized variables and more directly associated to wind changes, and may, thus, be used with more confidence in operational weather and climate predictions. Finally, the developed indices are applied in a case-study to the wet period between October 2016 and April 2017, when their predictability is assessed in comparison to rainfall.

In summary, this paper aims to respond to the four following questions and is organized as below:

1. Do latitudinal differences in the SACZ position imply in different dynamical structures?
2. What are the key-variables and regions in atmospheric dynamics that account for most of the variability associated to the presence/absence of the SACZ? How do they relate to different SACZ positions?
3. How can these variables be combined into indices to quantify the probability of a SACZ configuration? How accurate would it be?
4. How good is the predictability of the SACZ, according to this methodology, in comparison with rainfall in Southeast Brazil?

Datasets used are described in Sect. 2; The meridional SACZ types are presented in Sect. 3; Patterns in atmospheric dynamics, associated with the presence and absence of the SACZ, are identified and described in Sect. 4; The model development is described in Sect. 5; A discussion about objective decision thresholds for the indices and their use as proxies for precipitation anomalies is presented in Sects. 6 and 7, followed by a summary of conclusions in Sect. 8.

2 Data

Four sources of data were used in this study:

1. Historical daily configurations of SACZ were compiled from Climanalise monthly reports from October 1995 to April 2015 (comprehending 19 wet periods of the SAMS), published by the Brazilian Center for Weather Prediction and Climate Studies (*Centro de Previsão do Tempo e Estudos Climáticos*, CPTEC 2018b) and the National Institute for Space Research (*Instituto Nacional de Pesquisas Espaciais*, INPE) (<http://climanalise.cptec.inpe.br/~rclimanl/boletim/>). The Climanalise reports state the official SACZ occurrences, as defined by CPTEC/INPE. Once Climanalise reports are available online until December 2014, more recent SACZ episodes were compiled from daily technical reports, which are also provided by CPTEC/INPE (2018a) at their web page (<http://tempo.cptec.inpe.br/boletimtecnico/pt>). Due to the subjectivity associated with defining a SACZ episode, the term *Humidity Convergence Zone* (HCZ) has been coined by Neto et al. (2010) and proposed to classify SACZ phenomena in which some of its typical composing systems are not so clearly or steadily established, sometimes resembling semi-stationary frontal systems, which often result in shorter episodes (< 4 days) and consequently less pronounced rainfall anomalies. Despite the differences in terminology, the HCZ has most of its dynamical structure similar to that of the SACZ's and, therefore, HCZ-days, as also determined by CPTEC/INPE, were here considered as SACZ-days as well. All SACZ-days were visually inspected and the position of its cloud band was compared to the 42°W–44°W longitude stripe (Fig. 2), which served as a meridional axis to classify its position. Five SACZ configuration series, from the northernmost position “A” to the southernmost position “E”, of binary classifications were built: code 0 for SACZ absent and code 1 for SACZ present.
2. 30 years of daily precipitation gridded data, in 0.5° × 0.5° resolution, were obtained from the Climate Prediction Center (CPC) unified gauge-based analysis of global precipitation (Xie et al. 2007) from 01/01/1986 to 31/12/2015.
3. 30 years of daily-averages 2.5° × 2.5° horizontal resolution data were obtained from NOAA/OAR/ESRL PSD AMIP-II NCEP Reanalysis (NCEP2) from their website at <http://www.esrl.noaa.gov/psd/> (Kanamitsu et al. 2002): OLR (W.m^{-2}), zonal (u , m.s^{-1}) and meridional (v , m.s^{-1}) components of wind, vertical velocity omega (ω , hPa.s^{-1}) and geopotential height (ϕ , m), at 850 hPa, 500 hPa and 200 hPa levels, from 01/01/1986

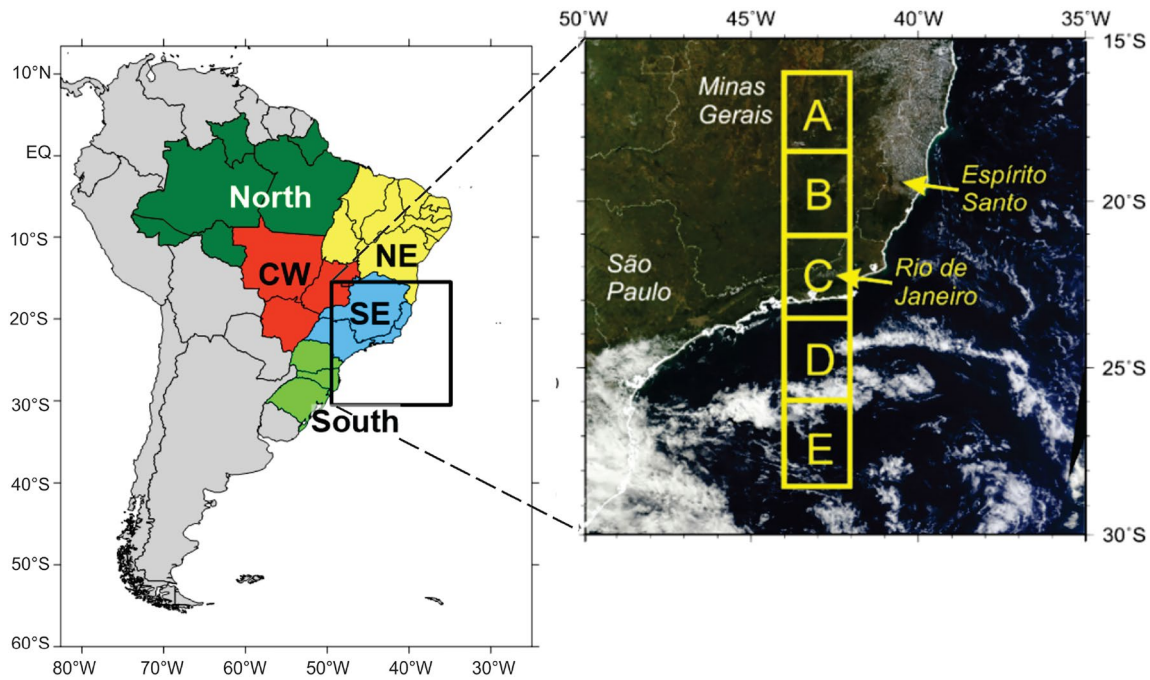


Fig. 2 Brazilian political boundaries (left) and latitude areas (from **a–e**) used for the classification of SACZ-days (right). The areas are 2.0° -longitude and 2.5° -latitude wide, bounded between 16°S – 28.5°S and 42° – 44°W . Adapted from Nielsen et al. (2016)

to 31/12/2015. Daily anomalies were then calculated with respect to monthly means of this 30-year period.

4. Operational $0.5^\circ \times 0.5^\circ$ -horizontal resolution daily forecasts from the Global Forecast System (GFS) from NCEP/NOAA were obtained for the wet SAMS period between 01/10/2016 and 30/04/2017. From GFS, we use the same dynamical variables (horizontal and vertical wind components and geopotential height) and precipitation rate ($\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$). In order to explore different forecasts horizons, results from the daily 00UTC runs are chosen to compose 4-member 6-hourly spaced daily averages, in the same fashion as in NCEPR2, but after 1, 4 and 7 days from the model initialization. In other words, Lead 1 is composed as a mean of the results from 24, 30, 36 and 42 h after the 00UTC initialization, Lead 4: 96, 102, 108 and 114 h, and Lead 7: 168, 174, 180 and 186 h after initialization.

3 Latitudinal variations

The distributions of SACZ-days show two cycles. A first seasonal cycle is characterized by the count of SACZ-days per month, with a peak of 239 days in January (24.1% of all days) and minimums of 57 (5.7%) and 16 (1.6%) SACZ-days in October and April, respectively (Fig. 3a). This distribution is in accordance to the mean onset, peak and demise months of the SAMS wet period as previously reported (e.g.

Gan et al. 2004, 2005; Raia and Cavalcanti 2008; Carvalho et al. 2012). A first difference with respect to the SACZ meridional position is noticed: northernmost regions A and B show the maximum counts of SACZ-days in December, while southernmost regions D and E show this maximum 1 month later, in January. In region C, the amount of SACZ-days is virtually the same during December and January. In regions A and B, there has been less SACZ-days in February than in March, in opposition to regions D and E, which account for more SACZ-days in February in the study period. Important to note that the SACZ cloud band may (and often does) act on more than one of the latitude areas at a time. Therefore, adding up the totals per area implies in erroneously accounting for one SACZ-day more than once.

For a more general view of the SACZ latitudinal variation, SACZ-days are grouped in three sets. Set “AB” is defined as a subset of all SACZ-days containing the days in which its cloud band acted on regions A *or* B, $AB = A \cup B$, not excluding events that occurred in other areas at the same time: $(A \cup B) \cap (C \cup D \cup E) \in AB$. Analogously, we define $DE = D \cup E$, letting $(D \cup E) \cap (A \cup B \cup C) \in DE$. The same applies to the subset C, where $C \cap (A \cup B \cup D \cup E) \in C$.

$P(AB|\text{SACZ} = 1)$ represents the conditional probability of having a SACZ-day in set AB, given that a SACZ is configured. We write $P(AB) = (AB|\text{SACZ} = 1)$ for sake of simplicity in notation, and the same applies to $P(C)$ and $P(DE)$ (Fig. 3b). A second seasonal cycle is observed in the latitudinal variation of the SACZ mean position during

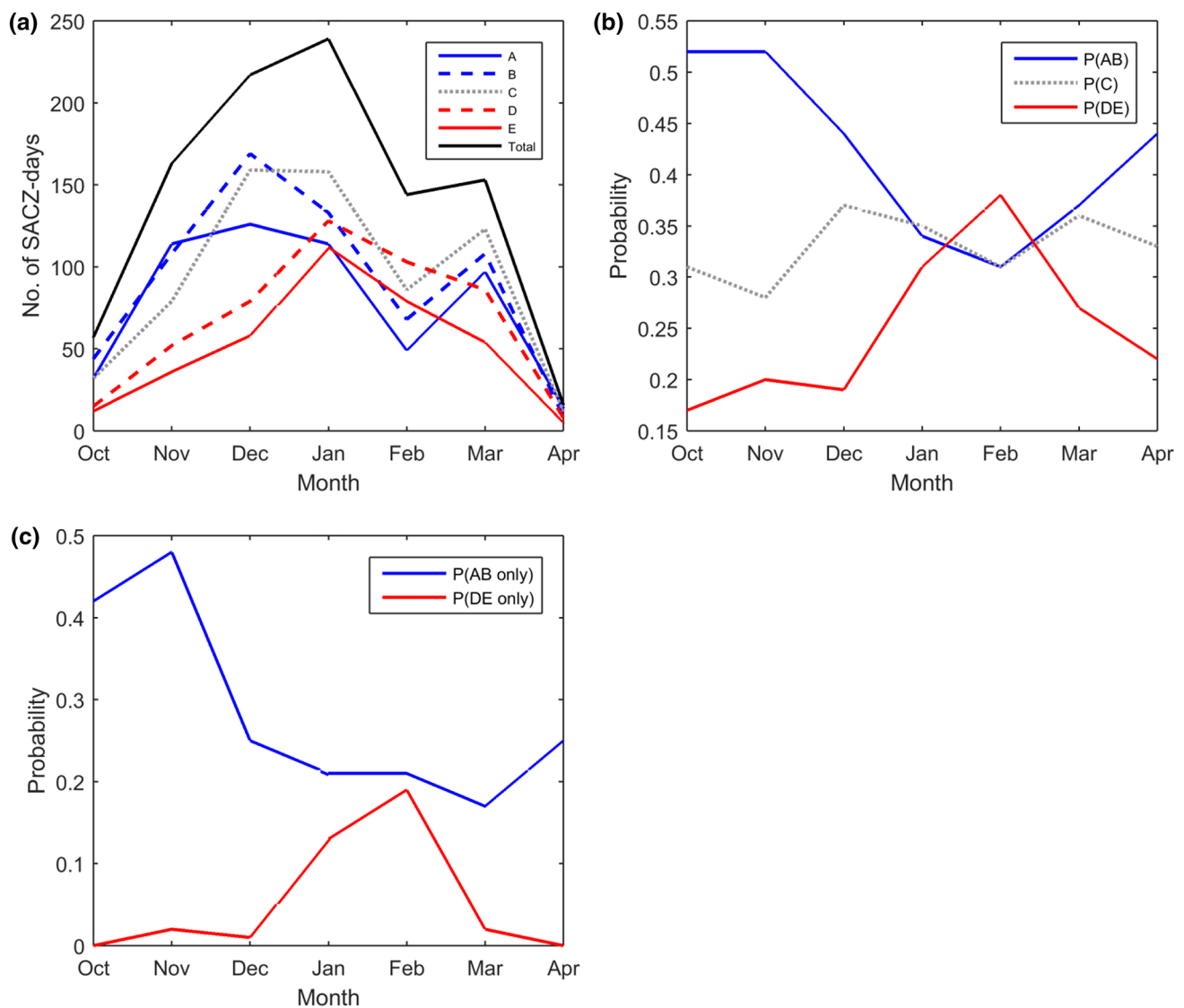


Fig. 3 Distributions of SACZ-days per latitude region between December 1995 and April 2015 (a), proportions of SACZ-days with respect to the totals of SACZ-days per month (b), and the proportion of SACZ-days with cloud-band position restricted to the north-

ernmost and southernmost areas, AB only and DE only, respectively, given the total occurrences per month (c) in the period from December/1995 to April/2015

summer. The probability of having a SACZ-day in the northernmost areas, $P(AB)$, is higher in the months of onset and demise of the SAMS wet period (between 44% and 52% in October, November, March and April) and lower in its peak (with a minimum of 31% February). On the other hand, the probability of having a SACZ-day in the southernmost areas, $P(DE)$, is minimum during the beginning and end of the SAMS wet period (around 17% and 27%) and maximum in February (38%): the only month where $P(DE) > P(AB)$. The middle region C does not present this second mode of latitudinal intraseasonal variation and $P(C)$ accounts for about one-third of all SACZ-days in all months, on average.

We define now the set “AB only” as a subset of all SACZ-days, containing the days in which the cloud band acted *only*¹ in latitude regions A or B: $(AB \text{ only}) = (A \cup B) - (C \cup D \cup E)$. Analogously, $(DE \text{ only}) = (D \cup E) - (A \cup B \cup C)$. The latitudinal variation of the SACZ position within the SAMS wet period is

¹ The term *only* is applied here meaning that the “AB only” set contains SACZ days where the cloud band did not act in boxes C, D or E. However and important to note, this does not mean that the cloud band was restricted to the boundaries of boxes A and B (Fig. 2). Similarly, the cloud band in “DE only” set *did* act in other regions, but strictly not in boxes A, B, or C.

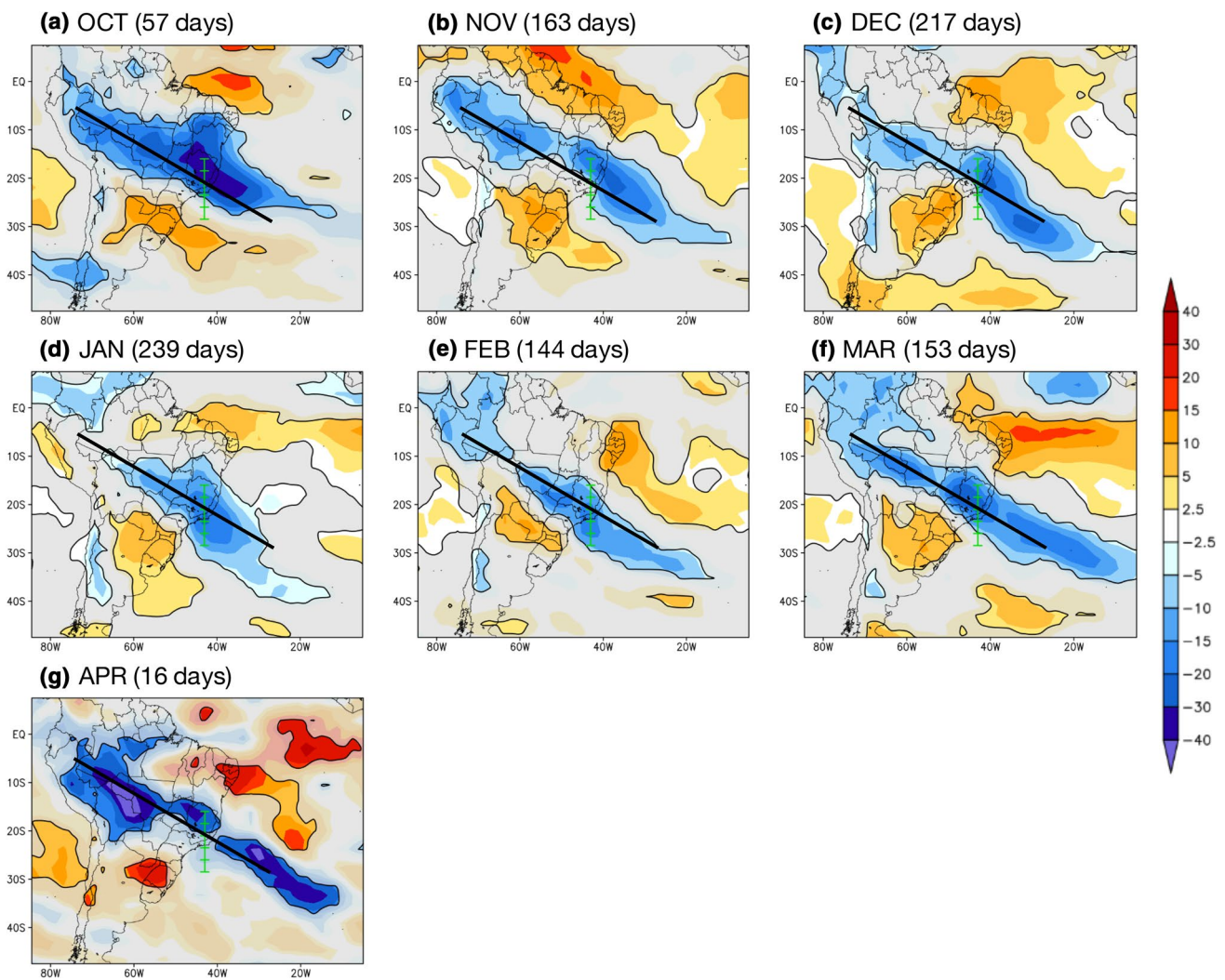


Fig. 4 OLR anomalies (Wm^{-2}) composites of SACZ-days averaged per month. Non-significant averages at a 95% confidence interval are covered with a grey shading. Black diagonal lines are drawn as a fixed reference for the SACZ cloud band mean position

even more clear if probabilities $P(\text{AB only}|\text{SACZ} = 1)$ and $P(\text{DE only}|\text{SACZ} = 1)$ are compared (Fig. 3c). The probability of having a SACZ-day in the onset months of the SAMS wet period (October and November) occurring only in the southernmost region DE, given that a SACZ episode is configured, is virtually null (0%–2%), while maximum in the northernmost region AB only (42% and 48%). This difference gradually decreases to its minimum in February, when SACZ-days are almost as likely to occur in AB only as in DE only, and increases again in March and April, the demise months of the SAMS wet period, where $P(\text{DE only}|\text{SACZ} = 1)$ is once more very low.

The latitudinal shift of the SACZ cloud band during summer is seen in Fig. 4, where a black line is drawn as a fixed reference for the SACZ cloud-band mean position and a green meridional axis is drawn to represent the regions used here to assess its latitudinal variation. Most

of the negative OLR averaged daily anomalies (AOLR) are seen northward from the SACZ mean position in October, November (SAMS onset), March and April (SAMS demise), while closer to its mean position in the remaining months. On the other hand, in February, most of the negative AOLR values lie southward from the reference line. In all months, the surrounding positive AOLR values are visible to northeast and southwest from the cloud band position, as an evidence to subsidence branches of a mean large-scale convection. These results are in accordance with those from Gandú and Silva Dias (1988), who showed that the subsidence movements appear, primarily, as a response to the strong convective activity in the Amazon and SACZ regions, but also influenced by the African and Pacific tropical heat sources during summer.

This meridional cycle is also evident if the complementary anomalies are examined. AOLR fields of all the

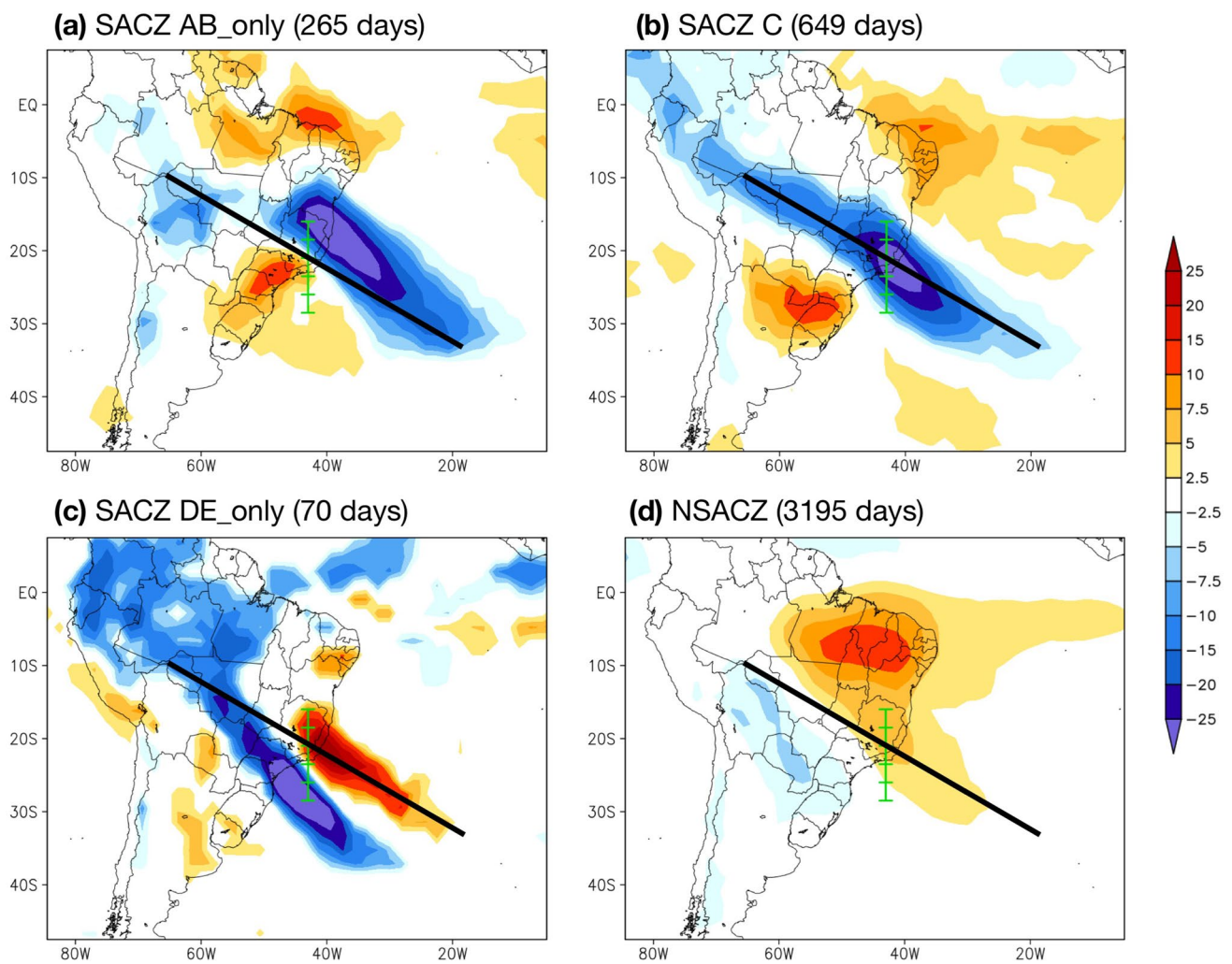


Fig. 5 Composites of OLR anomalies (Wm^{-2}) of SACZ-days averaged by sets “AB only”, “C” and “DE only”: northern (a), mean (b) and southern (c) SACZ types, respectively, and the composites of all

remaining non-SACZ days (d). Non-significant averages at a 95% confidence interval are not shown

remaining days between October and April, from 1995 to 2015 averaged per month, i.e. the days in which the SACZ was not present (NSACZ). The most positive AOLR values are seen northward from the SACZ mean position in October, November and April, while in January and February, positive AOLR values are distributed along the green axes and southward from the reference SACZ mean position (not shown).

Daily SACZ AOLR fields were also averaged by the time-independent sets of days “AB only”, “C” and “DE only”, containing 265, 649 and 70 days, respectively, when the latitudinal variation of the SACZ is clearly visible (Fig. 5). With this composite strategy, SACZ-days in which the cloud band acted in several regions at once are filtered out. The position of the cloud band between the northernmost and southernmost types of SACZ (Fig. 5a, c, respectively) barely intersect. This raises an evident need to define different

position-types of SACZ in order to achieve a more precise quantification method. For this reason, these three sets of days will serve as *reference sets* (samples of SACZ-days) for further composite analysis and hereafter be referred to as Northernmost (AB), Mean (or Central, C) and Southernmost (DE) SACZ types.

The consequent rainfall of the northernmost type of SCAZ (AB) affects mostly the northernmost portion of the Southeast and Northeast Brazilian regions (Fig. 5a), which are characterized by a drier climate than its surroundings (Reboita et al. 2010). These regions are less frequently reached by frontal systems, thus the SACZ represents a very important source of rainfall. However, this type of SACZ would not cause rainfall in most of the important basins in Brazilian SE, regarding population density and hydraulic electricity generation. The middle SACZ type (C) represents its mean position and its cloud

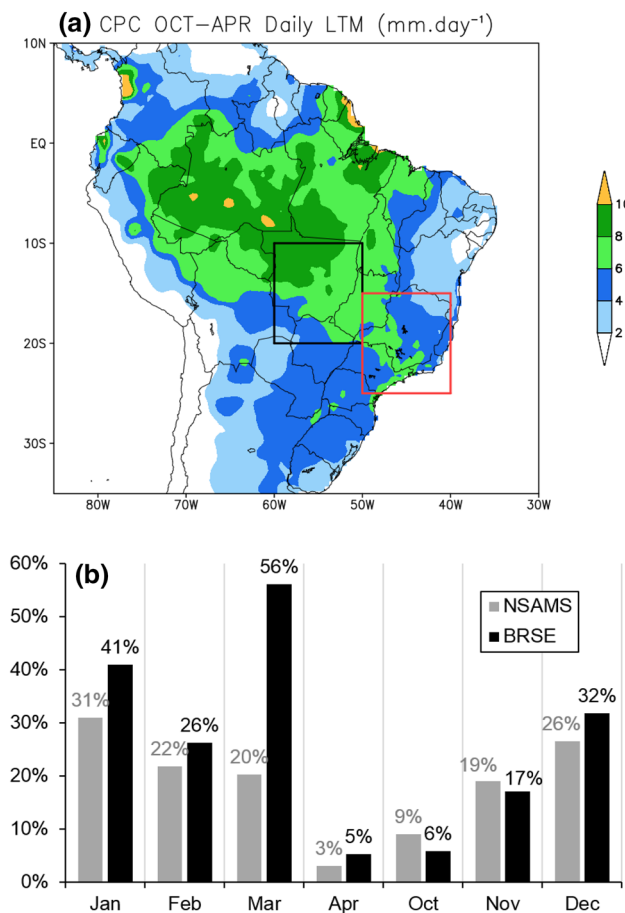


Fig. 6 The delimited areas of the nucleus of the SAMS (NSAMS, in black) and the Brazilian Southeast region (BRSE, in red) over the 1981–2010 CPC gauge-based October–April mean precipitation climatology (mm day^{-1}) (a); and proportions of rainfall volumes occurred in SACZ-days within the NSAMS and BRSE, between December 1995 and April 2015 (b)

band is almost entirely restricted to the SE and Central-West (CW) regions of Brazil (Fig. 5b). The southernmost SACZ (DE) is the only type that causes rainfall in the Southern region of Brazil, besides SE and CW. In addition, the more positive OLR anomalies suggest that the mean preferred subsidence of the convection associated to this type of SACZ (DE only) is located north-eastward from the diagonal cloud band (Fig. 5c), while the other types of SACZ (C and AB) present more positive AOLR values south-westward from the diagonal band (Fig. 5a, b), as more commonly observed in mid-latitude troughs in a baroclinic atmosphere (Holton 2004). These results are also in accordance with those from Gandú and Silva Dias (1998), where they suggest that the compensating subsidence branch, acting to compensate a SACZ-located heat source, is nearly confined to its West.

The region defined between 10°S – 20°S and 60°W – 50°W is referred to as the nucleus of the SAMS (NSAMS, Fig. 6a)

(Gan et al. 2004, 2005), where the annual rainfall cycle is directly influenced by the SAMS, and where around 20% of the rainfall volumes occur during SACZ configurations on average. The maximum proportion of rainfall during SACZ episodes in this region occurs in January (31%) and minima in October (9%) and April (3%), when SACZ configurations are less frequent (Fig. 7b). The Brazilian Southeast (BRSE) region is also dependent on rainfall associated with the SAMS cycle, and around 26% of the rainfall volumes in this area occur during SACZ episodes: maxima of SACZ-associated rainfall percentages of 56% and 41% are observed in March and January, and minimum of 5% and 6% in April and October, respectively (Fig. 6b), in accordance to the seasonal cycles observed in Fig. 3 and to SAMS precipitation climatology (as in Fig. 1 from Vera et al. 2006). Noteworthy, the amount of SACZ-associated rainfall in February is less than half of that in March (26% and 56%, respectively, of total rainfall between October 1995 and April 2015).

4 Dynamical patterns

In order to identify patterns in the atmospheric dynamics associated with the presence and absence of the SACZ, the same above described anomaly-calculation applied to OLR was used with all other dynamical variables: SACZ-days anomalies were individually calculated with respect to their long-term monthly means (1986–2015) and averaged by grouping the reference sets (AB only, C and DE only). This relatively simple methodology allowed the identification of variables and areas in which SACZ-day composites were significantly different from long-term means at a 95% confidence interval. In this section, some of the regions of significant anomalous patterns are delimited with numbered boxes, from which time series of *raw* predictors will be extracted. These series will then be transformed to build sets of uncorrelated and linearly independent explanatory variables for the development of models in Sect. 5.

Large-scale cyclonic gyres at 850 hPa are observed in both composites of anomalies of the northernmost and mean SACZ types (Fig. 7a, b). This cyclonic anomaly is in accordance with peak-monsoon circulation (as in Fig. 13b in Grimm et al. 2007) and intraseasonal SAMS wet periods, or active phases during the SAMS wet period, driving the Amazonian moisture flow towards the SACZ influence regions (Jones and Carvalho 2002; Gan et al. 2004; Carvalho et al. 2011; Ma et al. 2011). However, the cyclonic anomaly in the southernmost SACZ-type composite is much less pronounced and mostly constrained over the continent: its east-side branch does not flow south-eastward as in the other cases, but shows a much stronger southward component, pointing to the southernmost states of Brazil. In Fig. 7c, an anti-cyclonic anomaly in mid-latitudes over the

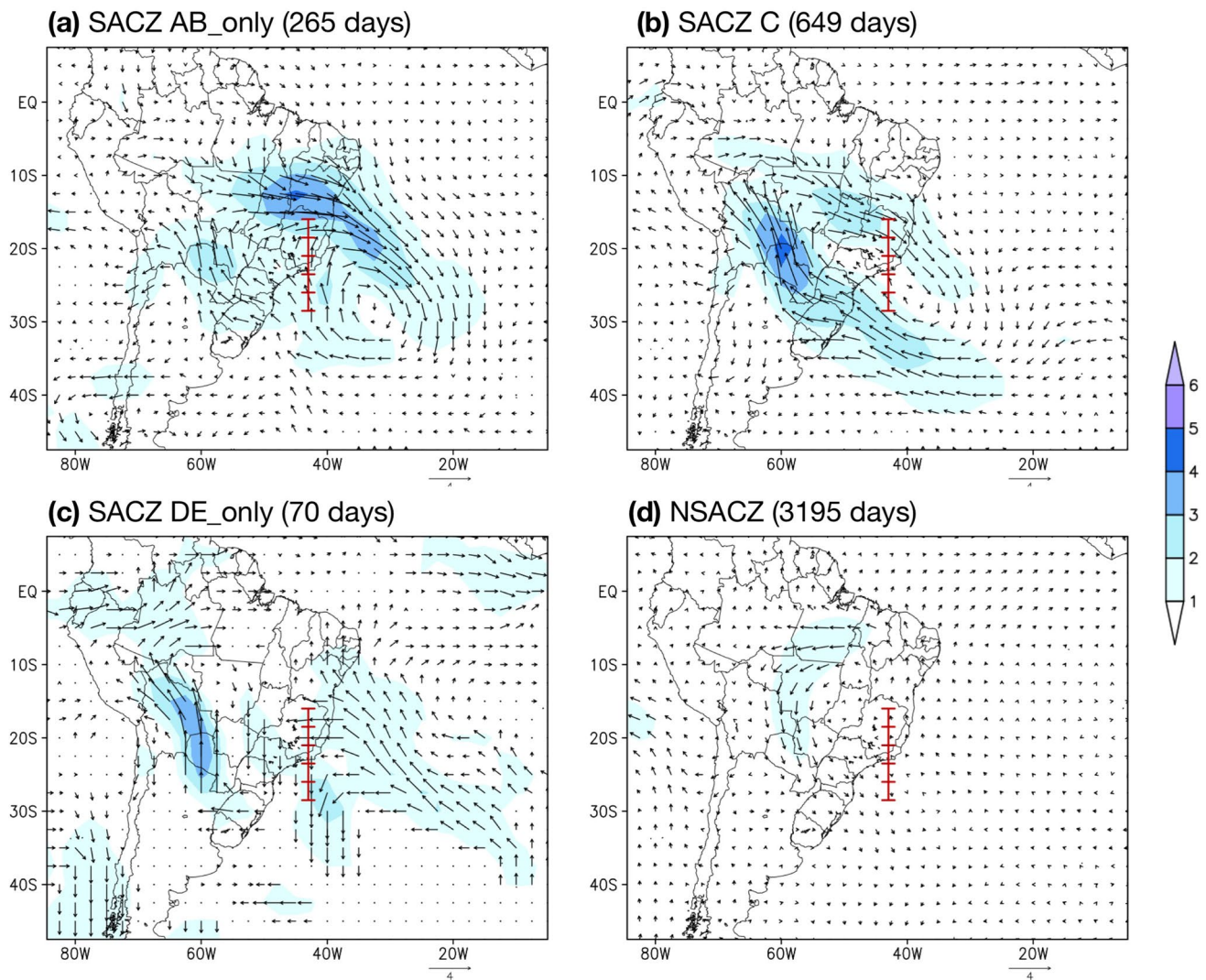


Fig. 7 Composites of daily 850 hPa wind anomalies (ms^{-1}), averaged by the reference sets: the northernmost (a), mean (b), southernmost (c) types of SACZ, and the averaged daily anomalies of all the remaining days between October 1995 and April 2015 in which the SACZ was not configured (NSACZ, d). Vectors are only drawn where

both meridional and zonal components of the wind are significant at a 95% level. A red vertical axis is drawn as a reference for the latitude regions from A to E. This set up of figures representing the reference samples will be adopted throughout this section with all other composites

Atlantic Ocean resembles a prefrontal surface high, typical of cold-frontal systems acting over southern Brazil. All three SACZ (NSACZ) composites show positive (negative) anomalies of the meridional component of wind over Paraguay and Bolivia, the west-side northward branch of the trough anomalies, which is in accordance to previous studies. Liebmann et al. (2004), demonstrated the relationship between weakened (strengthened) SALLJ, or East-of-Andes pole-ward transport of Amazonian moisture, and positive (negative) precipitation anomalies in the SACZ region. The configuration of SACZ episodes is associated with Amazonian moisture transport deflected toward the East, pointing to the Atlantic Ocean, instead of southwards toward the LPRB

(Herdies et al. 2002), justifying the positive anomalies in the SALLJ region in Fig. 7a–c, which would then help configure the South American precipitation seesaw between SACZ and LPRB regions (Nogués-Paegle and Mo 1997; Díaz and Aceituno 2003; Herdies et al. 2002). Herdies et al. (2002) also showed that moisture is more efficiently transported poleward in SACZ than NSACZ conditions, despite the short period analysed in their study.

In order to express the circulation imprints into the dynamical indices, the anomalous 850 hPa wind is decomposed. Clear dipoles of zonal and meridional wind components are visible in all SACZ types, except for the southernmost case, where positive zonal anomalies are not

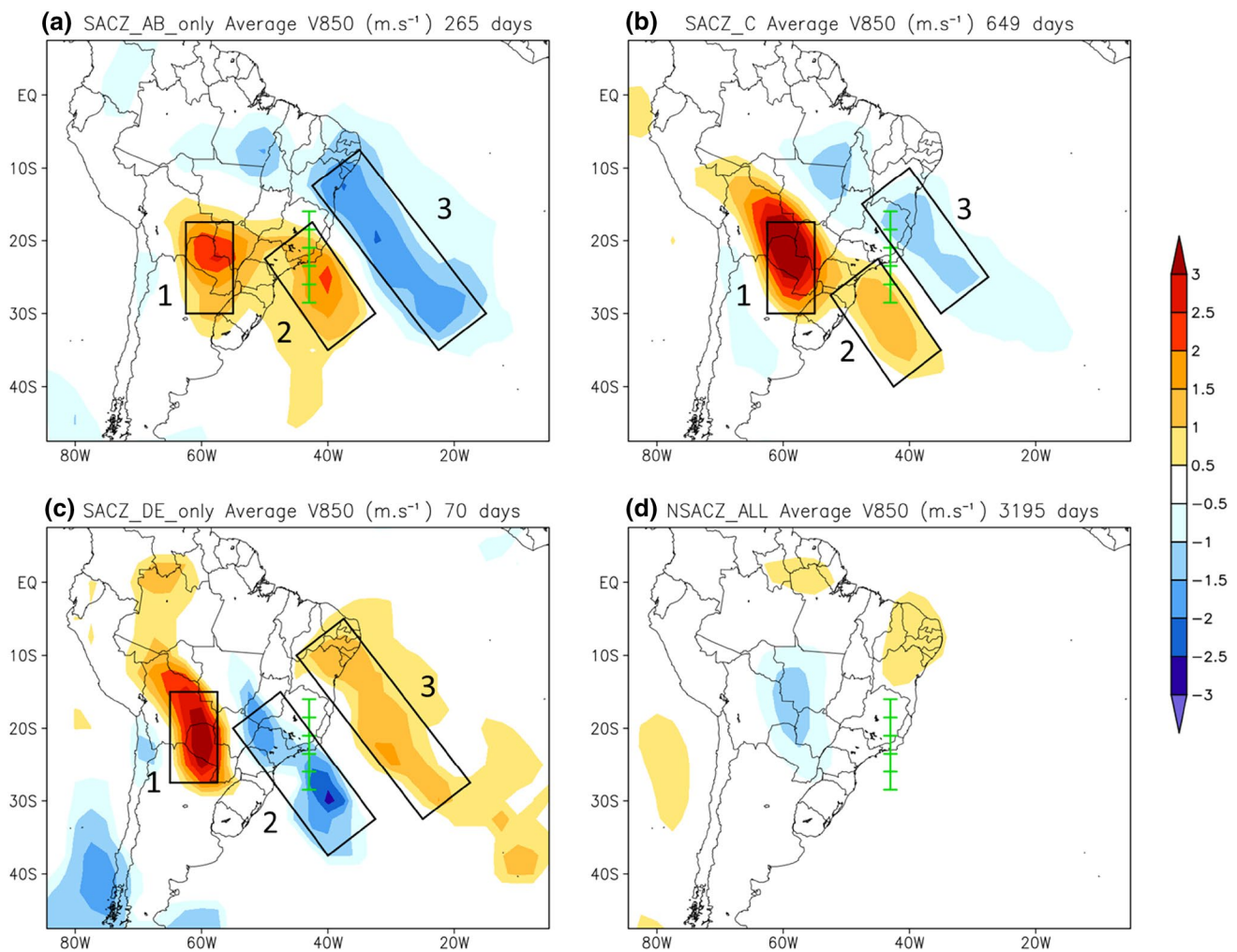


Fig. 8 Same as in Fig. 7, but only for the meridional component of the 850 hPa wind

observed, but only a negative anomalous region associated to the north-westward branch of the anti-cyclonic anomalous circulation over the Atlantic (not shown). The composites of anomalies for the meridional component of the 850 hPa wind is shown in Fig. 8. Once more, a clear common feature of all SACZ types is the weakened mass transport from the Amazonian region towards the LPRB, highlighted in boxes 1. On the other hand, the NSACZ pattern (Fig. 8d) shows an enhanced southward transport in the SALLJ region. This bimodal pattern is especially in accordance to results brought by Herdies et al. (2002), who showed that the poleward moisture transport from the Amazon is displaced eastward (westward) in SACZ (NSACZ) phases within the wet season. The differences in the 850 hPa circulation are made evident in Fig. 8, since boxes 2 and 3 in the southernmost SACZ type show meridional wind anomalies opposite in signal, when compared to the mean and northernmost SACZ type. Once more, the northernmost and southernmost SACZ

patterns show signals in antiphase over boxes 2 and 3 (see Fig. 5 for similar anti-phase pattern in OLR). The negative meridional anomalies in Fig. 8a–c suggest the direction in which the Amazonian moisture is transported, in substitution to the East-of-Andes SALLJ-like pattern during SACZ episodes. The poleward transport of moisture is then accompanied by winds with a northward anomalous component located south-westward (north-eastward) to the SACZ cloud band in its southernmost (mean and northernmost) types.

The vertical motion patterns expressed in omega (pressure velocity) at 500 hPa are also examined in the same fashion and shows patterns in accordance to the 850 hPa anomalies. The northernmost and mean SACZ composites show similar vertical motion patterns, with the main subsidence located south-westward from the diagonal cloud-band position, as in Gandú and Silva Dias (1998), while this signal is inverted for the southernmost type of SACZ, where the main subsidence is located north-eastward from

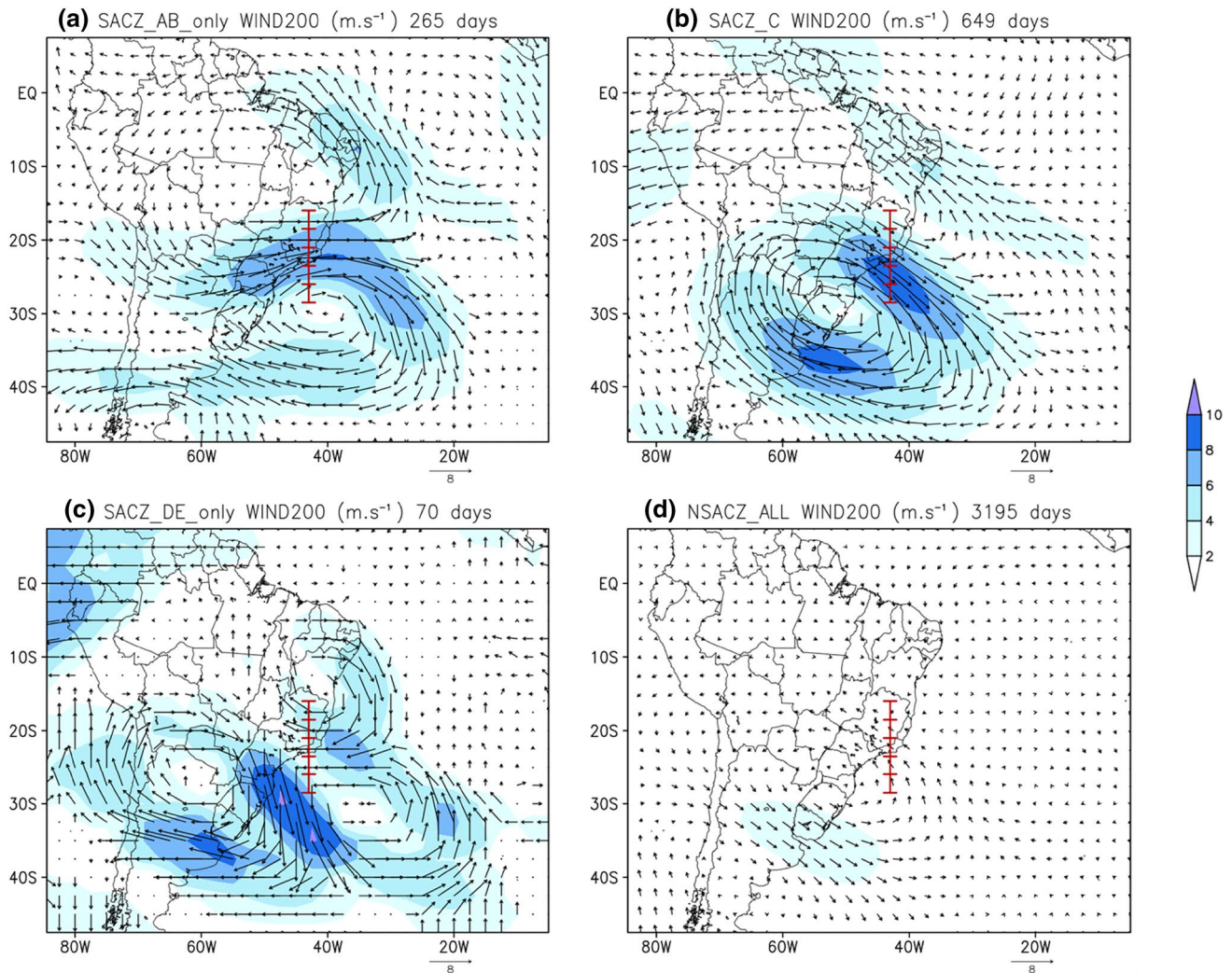


Fig. 9 As in Fig. 7, but for the 200 hPa level

the convection diagonal band and the south-westward subsidence is less pronounced (not shown). The pattern presented for the southernmost SACZ type, once more differing from the others, is not in accordance with Gandu and Silva Dias (1998), who show that the preferred subsidence branch associated to the SACZ is located south-westward from the heat source.

The 850 hPa predominant cyclonic anomalous patterns seen in the northernmost and mean SACZ types (Fig. 7a, b) also appear in 200 hPa (Fig. 9a, b), slightly displaced south-westward, evidencing a baroclinic character of the system. In addition, the northernmost and mean SACZ composite anomalies also show anti-cyclonic gyres, suggesting strengthened BH configurations, along with a cyclonic vortex in the Equatorial portion of the Atlantic Ocean. Overall, SACZ (NSACZ) wind patterns are in accordance to the monsoonal especially wet (dry) periods as presented by Gan et al. (2004). However, the 200 hPa anomalous pattern of the

southernmost SACZ type differs from the two others. An enhanced BH circulation is not clear and a cyclonic anomaly is observed at around 12°S 40°W, suggesting a NL-like configuration displaced/strengthened onto the continent. The nucleus of the cyclonic anomalous pattern over LPRB in 200 hPa in Fig. 9c is also displaced westward in comparison to 850 hPa in Fig. 7c, which also indicates a mean baroclinic structure of the system.

The anomalous patterns identified in the 200 hPa-level vertical component of the relative vorticity, associated with the presence of the SACZ, have already been noticed in some studies, and are likely associated with the propagation of extratropical Rossby wave trains, moving eastward from the southern Pacific Ocean and equatorward over South America (Kalnay et al. 1986; Liebmann et al. 1999). Van der Wiel et al. (2015) more recently demonstrated the role of dry non-divergent barotropic Rossby-wave dynamics and the 200 hPa-jet horizontal wind shear in generating the

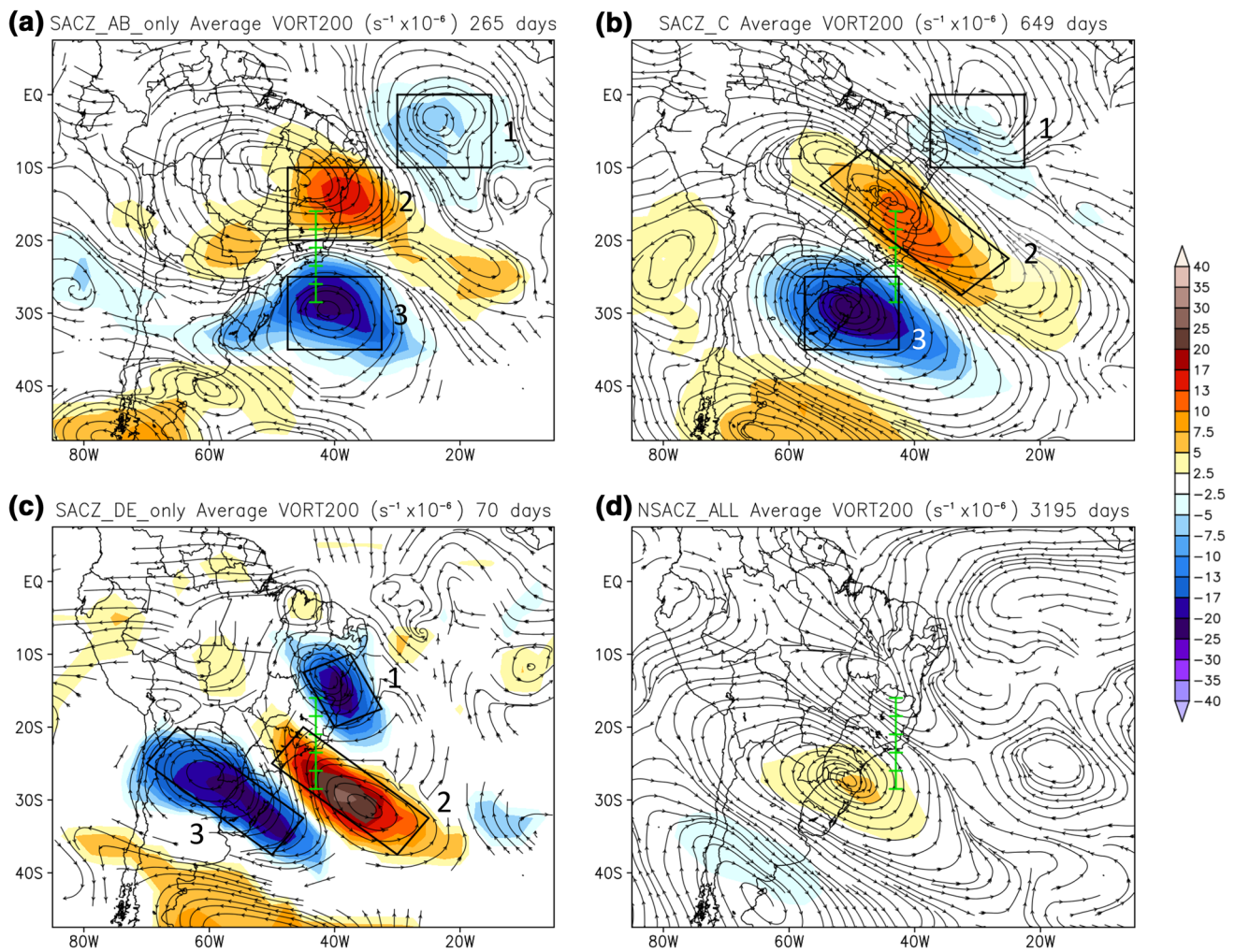


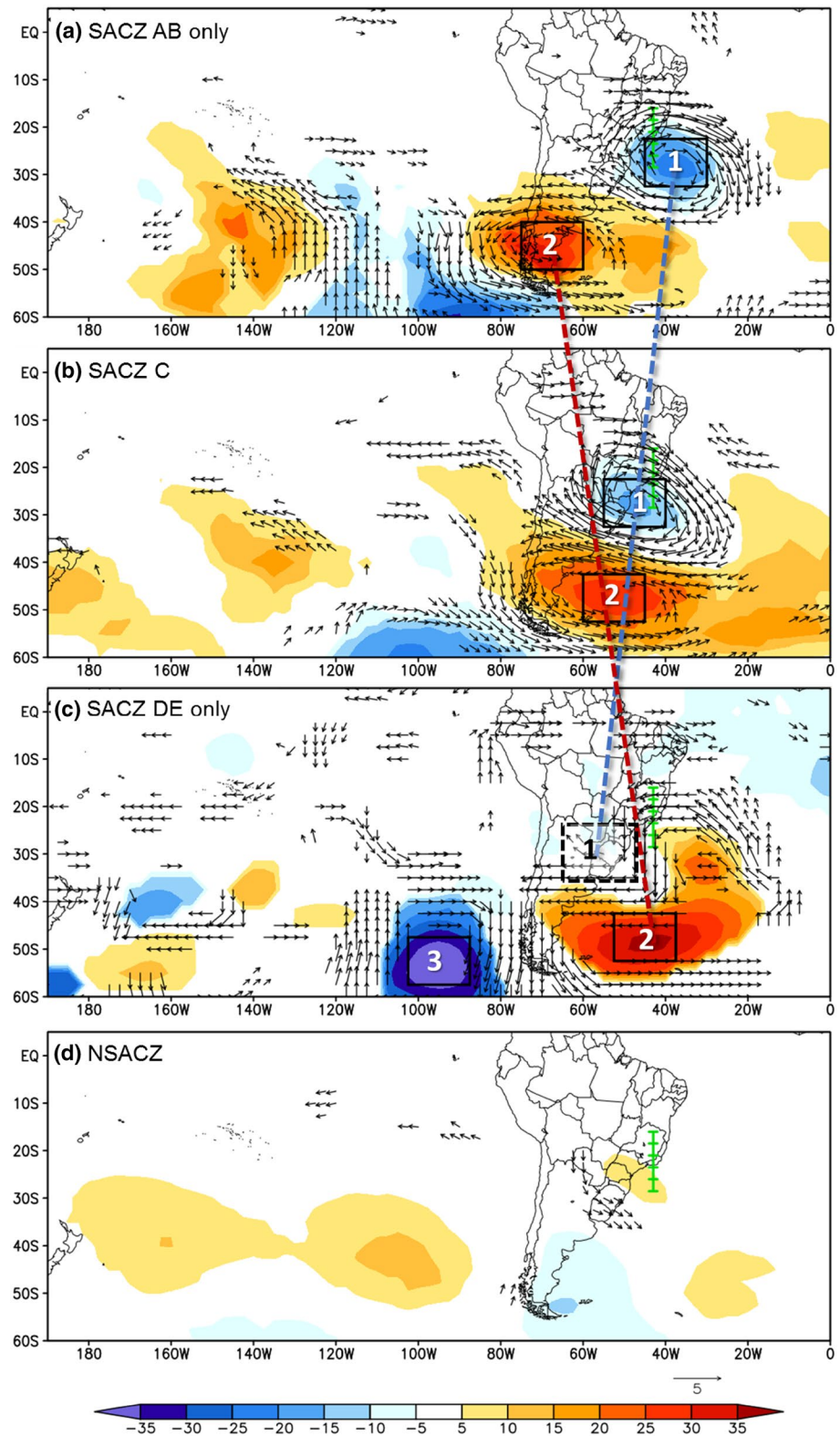
Fig. 10 Same as in Fig. 7, but for the vertical component of the relative vorticity at 200 hPa

diagonally oriented and elongated 200 hPa vorticity anomalies during SACZ episodes. Three nuclei of this wave train are comprehended in our study domain and identified in Fig. 10. The positions of the boxes vary among SACZ types and in accordance to the previous figures. Boxes numbered as 1 and 2 suggest strengthened and/or displaced BH and NL, respectively, while boxes 3 are associated with the position of the surface and mid-atmospheric trough. Most of the variation in position of box number 3, among the three SACZ types, occurs in the zonal direction, while relatively null in latitude. In addition, an anti-phase in anomalous vorticity maxima is visible between the NSACZ and SACZ averages.

Rossby wave trains are also identified in the alternating signals of anomalies of 500 hPa geopotential height and wind fields over the southern Pacific Ocean (Fig. 11), more clearly visible in the northernmost and mean SACZ types. The SACZ pattern is in accordance with the evolution of

the westward wave train associated to precipitation positive anomalies in the SACZ region presented by Liebmann et al. (2004), who also show that precipitation events in the vicinity of the SALLJ, pointing southwards to the LPRB, are associated to wave trains in anti-phase with the SACZ case. In agreement, the SACZ pattern shown here seems to be in phase with the dry-period pattern in Uruguay presented by Díaz and Aceituno (2003). The mean areas of cyclonic (box 1) and anti-cyclonic (box 2) anomalies appear to zonally move in opposite directions among the meridional SACZ configuration types. The negative (positive) geopotential anomaly center moves westward (eastward) while the SACZ meridional types switch southward. These centers of anomalous maxima and minima are disposed more sparsely, with the greatest zonal distance, in the northernmost SACZ pattern. In the mean SACZ average, areas 1 and 2 are closer in longitude, which gives the wave train a more meridional than a zonal character over

Fig. 11 Composites of reference samples for the 500 hPa geopotential height. The 500 hPa wind is only drawn where anomalies are significant in both zonal and meridional components (95% level) and vector magnitudes are larger than 1 ms^{-1}



SA, in comparison with the northern SACZ type. In the intraseasonal time scale, Grimm and Silva Dias (1995) show that anomalous convective activity in the SACZ and SPCZ regions are dynamically connected, via Rossby wave

dispersion. They used a divergent barotropic model to show that a slight change in positioning of the region of divergence caused by the SPCZ may weaken the trough associated with the SACZ. In their results, it is also possible to

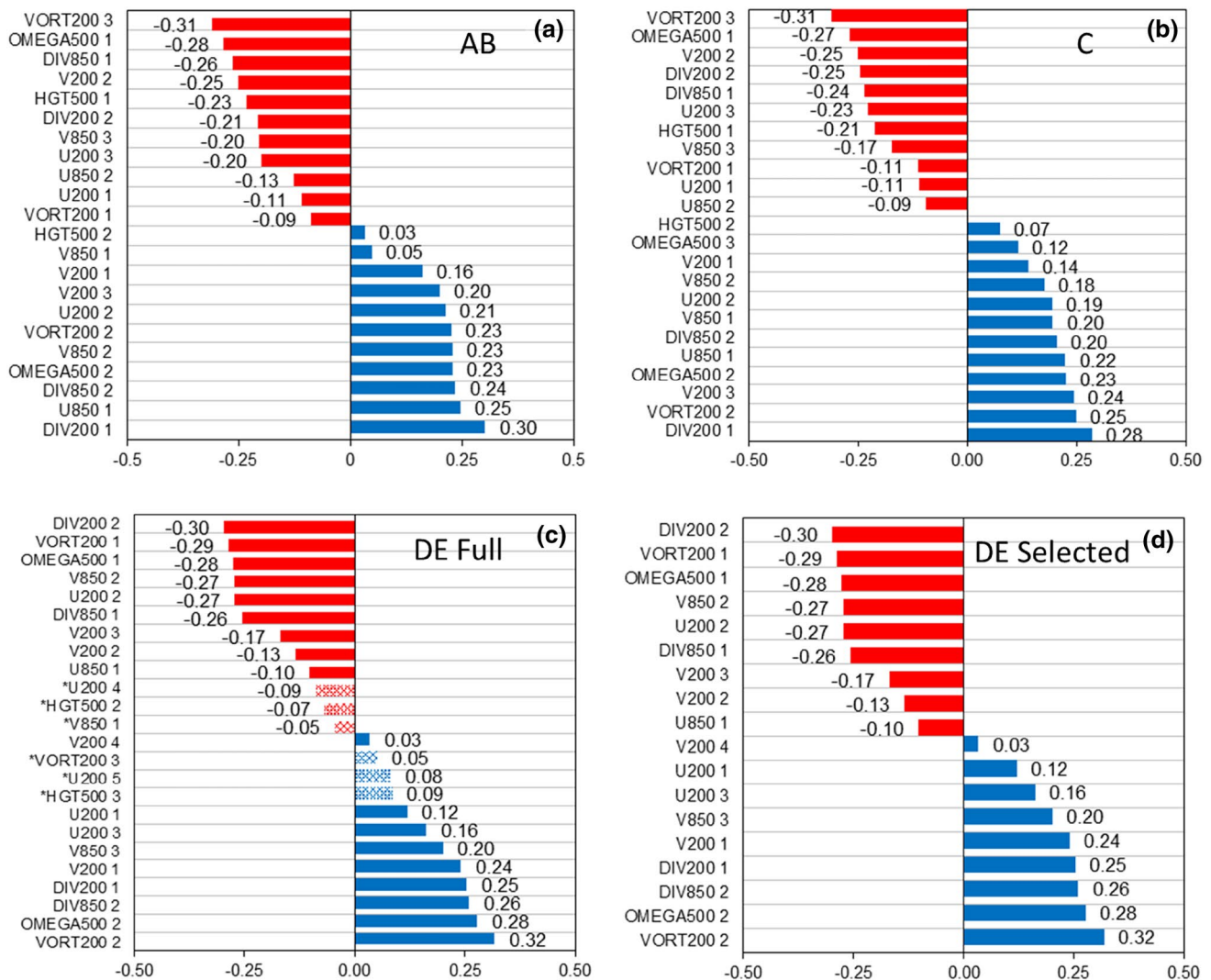


Fig. 12 PC1 coefficients (loadings) of datasets AB, C, DE Full and DE Selected. The hatched bars in DE Full indicate coefficients with opposite signals from the SACZ mode, which are removed to obtain DE Selected

observe a southward displacement of this trough, although this is not clearly mentioned in their work.

In the southernmost SACZ type, the PSA wave train configuration is less clearly visible and the continental trough anomaly is much less pronounced, where the third center of cyclonic anomaly emerges to lower latitudes in the southern Pacific Ocean at around 100°W . This is yet another evidence that the southernmost SACZ days considered in this study may include characteristic of other systems, such as transient mid-latitude baroclinic waves.

This composite analysis was performed until significant anomalous regions were defined in boxes (as in Fig. 8, for example) from a comprehensive set of dynamical variables: the zonal and meridional components of the 850 hPa and 200 hPa wind, horizontal wind divergence at 850 hPa and 200 hPa, the vertical component of the relative vorticity at 200 hPa, pressure velocity and geopotential height at

500 hPa. In total, 22, 23 and 24 regions of significant anomalies associated with the SACZ were defined for its northernmost, mean and southernmost positions, respectively. In the next section, 30-year daily time series of these variables are extracted and combined into SACZ indices.

5 Model development

5.1 Explanatory variables

The problem of colinearity (or multicollinearity, in case of more than one pair of correlated explanatory variables) is a well-known issue inherent of regression analysis, although there exist ways to treat it (Farrar and Glauber 1967; Schaefer 1986; Belsley 1991). The use of Principal Component Analysis (PCA) for the orthogonalization of a dataset

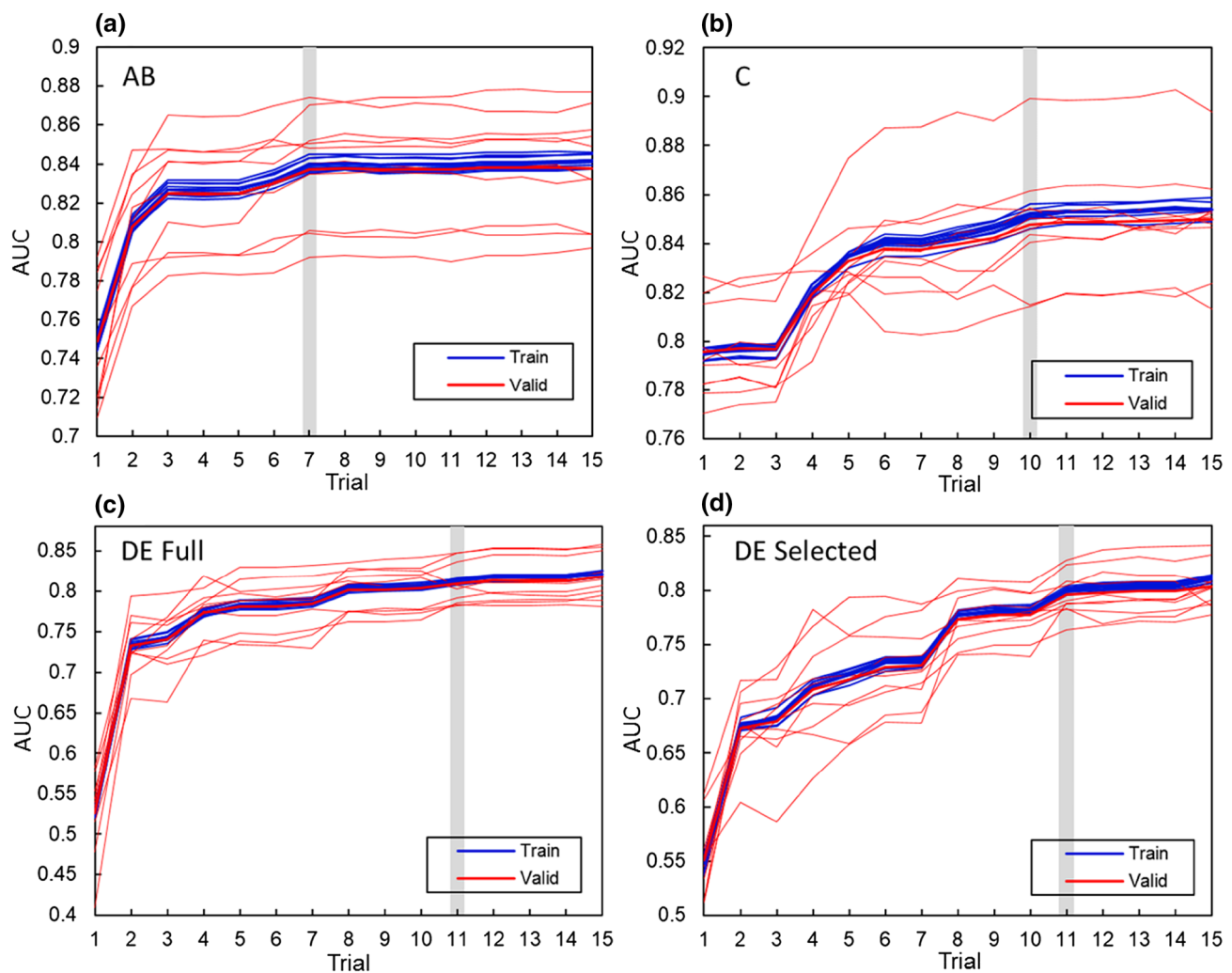


Fig. 13 Area under the ROC curve (AUC) of each stepwise model building trial (horizontal axes) and the tenfold cross validation (red lines) and training (blue lines) runs for each of the four sets of princi-

pal components. Thicker lines represent the calibration and validation averages. The grey bars indicate the step of the best chosen model configuration

prior to the adjustment of regression models (Principal Component Regression, PCR) solves the problem of multicollinearity, once centered principal components are linearly independent and perfectly uncorrelated, which allows a more proper estimation and analysis of regression coefficients (Massy 1965; Jolliffe 1982; Dormann et al. 2013).

The series of all 4245 days of SAMS wet periods (October–April) between 1995 and 2015 from the selected areas described in the previous section are originally correlated and therefore centered (long-term mean removed) and submitted to a PCA performed on the correlation matrix. The first principal component (PC1) is physically meaningful and associated to the SACZ mode for the northern and mean types (AB and C, respectively, Fig. 12a, b), accounting for 36.9% and 34.5% of the total variance, respectively. For the southernmost case (DE), the first PC of the full set of variables explained 26.5% of the variance, but did not represent a SACZ mode, once 6 of

the 24 selected variables presented opposite signals in loadings, namely: geopotential height, the zonal wind at 200 hPa in areas 4 and 5, meridional wind at 850 hPa in area 1 and the relative vorticity at 200 hPa in area 3 (Fig. 12c). These 6 variables were all gradually removed from the full set (not shown) until a meaningful SACZ-mode PC1 was obtained, accounting for 34.5% of the total variance (Fig. 12d). Therefore, two sets of variables were defined to compose the uncorrelated dataset for the southernmost SACZ type: the *DE Full* and *DE Selected* sets, being the first principal component of the latter physically meaningful and associated to the SACZ. However, the use of scores of PC1 only was never a most accurate option to identify SACZ-days (further explained in Sect. 5.4) and the inclusion of higher-order principal components as explanatory variables appeared to be important to identify SACZ events in a daily-scale, showing that, even with the very careful selection of variables and geographical areas, other modes of

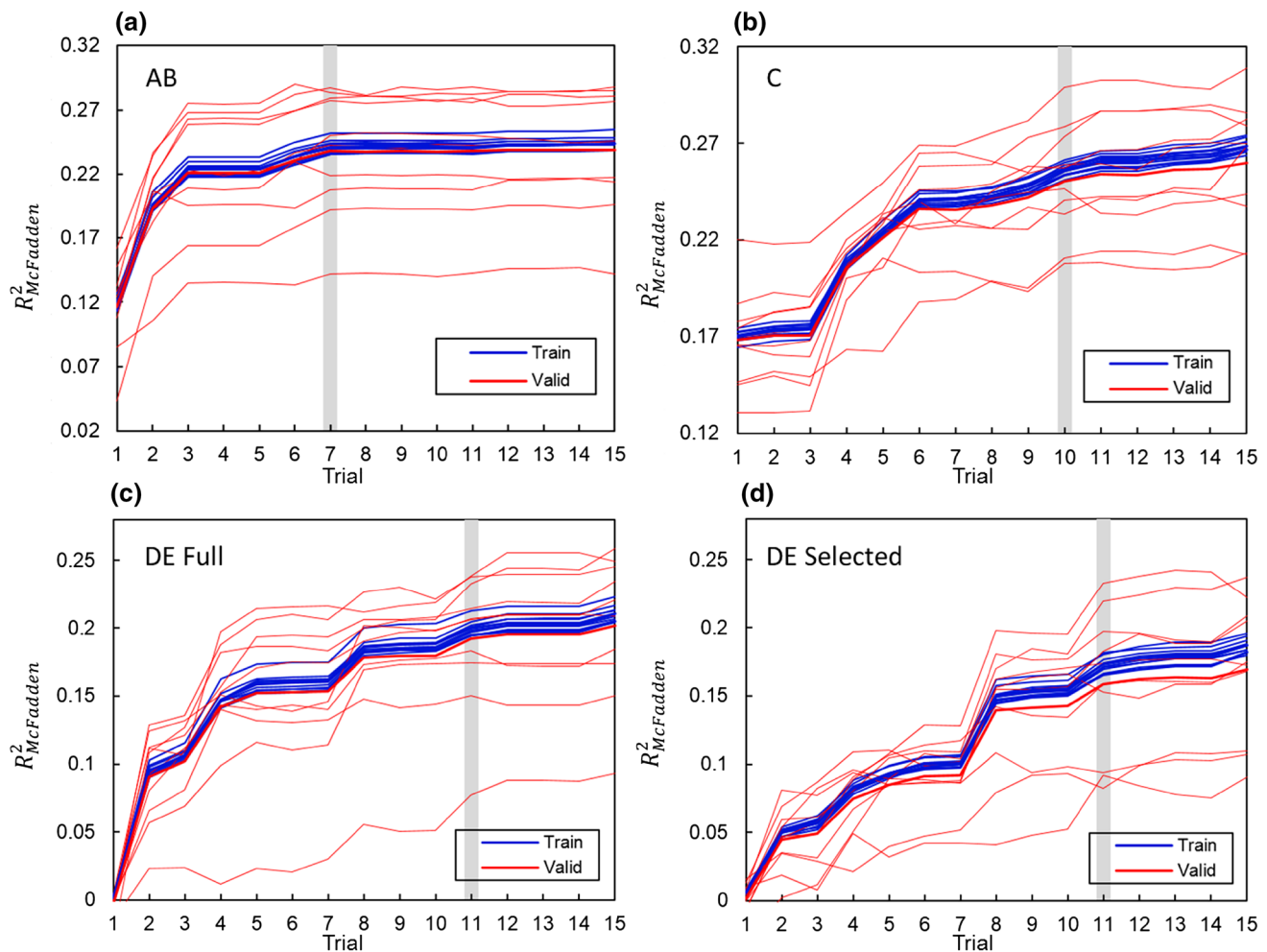


Fig. 14 Same as in Fig. 13, but for \bar{R}^2

variability than those identified in the previous section may, to some extent, be associated to SACZ events as well.

In addition, the original variables (before PCA) are approximately normally distributed and normalized prior to the PCA transformation. Original values of $0+2\sigma$ and $0-2\sigma$ (zero plus and minus 2 standard deviations) were scaled to correspond to -1 and 1 , respectively, which allowed that the original signals were maintained after normalization, except for the all-positive geopotential height series, whose original values of $\text{mean}+2\sigma$ and $\text{mean}-2\sigma$ were chosen to correspond to -1 and 1 , respectively. The factor loadings for PC1 of the four data sets are shown in Fig. 12. The 200 hPa relative vorticity, wind divergence (both in 850 hPa and 200 hPa) and omega in mid-levels tend to account for a larger portion of the variance of PC1, generally figuring among the 3 or 4 largest loadings in module, while the pure meridional and zonal components of the wind tend to yield smaller weights.

The four sets of uncorrelated explanatory variables (namely: the scores of AB, C, DE Full and DE Selected datasets), each containing 4,245 observations (days) from

October to April, 1995–2015, were then used in the adjustment and validation of logistic regression models to predict SACZ episodes in a daily scale, as follows.

5.2 Binary logistic regression

While ordinary linear regression aims to describe the relationship between continuous variables, logistic regression is applicable to cases when the outcome variable is discrete, taking a certain number of possible values (Hosmer and Lemeshow 2004). This is especially appealing when modeling dichotomous or binary variables, such as positive/negative diagnosis, passing/failing tests and many other possible Boolean features, such as the presence/absence of the SACZ on a given day. For this reason, binary logistic regression has been widely explored more traditionally in health (Arshad and Hide 1992; Brough et al. 2015; Lomholt et al. 2016), but also in environmental sciences (e.g.: Flantua et al. 2007; Guanche et al. 2014; Raja et al. 2016).

Consider we take one of the meridional types of SACZ (AB for instance) and let the SCAZ configuration series be vectors of n days, $Y = (y_1, y_2, \dots, y_n)$, whose values are coded as $y_i = 1$ for a day with the presence of SACZ and $y_i = 0$ for its absence, where $i = 1, 2, \dots, n$, and $x = (x_1, x_2, \dots, x_p)$ a matrix of p independent variables of length n (i.e. the scores of PCA). In a general regression point of view, $E(Y_i|x_i)$ represents the model outcome: the estimate of y given the explanatory data x on day i . This is commonly obtained by a linear combination of the explanatory variables in multiple linear regression, $E(Y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$, where $E(Y|x)$ may assume any value in $[-\infty, +\infty]$, if x also varies in this range. In our case, the dependent variable Y is dichotomous and the n quantities $E(Y|x)$ should assume values within the interval $[0, 1]$, interpreted as the conditional probability that $Y = 1$ (SACZ be present) given the data x on day i , written as $P(Y = 1|x)$. The regression model is then expressed in the form of Eq. 1, in which $\pi(x)$ represents $E(Y|x)$ when the logistic distribution is used and $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ is the vector of regression parameters.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}, \quad 0 \leq \pi(x) \leq 1 \quad (1)$$

For an arbitrary combination of parameters β , the conditional probability $P(Y = 1|x)$ is provided by $\pi(x)$, while $P(Y = 0|x)$ is equivalent to $[1 - \pi(x)]$. Still for a certain β , the contribution of the i^{th} collected data to the model's likelihood is written as $\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$. In a series of n observations, the likelihood function $L(\beta)$ is expressed by the joint probability of all n subjects given the vector of parameters β , as in Eq. 2.

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2)$$

The best combination of parameters is the one that maximizes the likelihood function. However, it is mathematically cumbersome to find the maxima of the expression in Eq. 2, thus derivatives are normally calculated on its log, so that products are expressed as summations, and the Newton–Raphson algorithm is commonly applied (Hosmer and Lemeshow 2004). This method of adjustment of model coefficients is referred to as the Maximum Likelihood Estimation (MLE) and the vector of parameters that maximize Eq. 2 (and its log) is called the maximum likelihood estimate of β , denoted by $\hat{\beta}$.

5.3 Goodness-of-fit and performance analysis

The goodness-of-fit (GOF) of the models was here assessed using three different criteria: the area under the receiver operating characteristic (ROC) curve (AUC) (Bradley 1997),

the Akaike's information criterion (AIC) (Akaike 1974) and McFadden's pseudo R^2 (McFadden 1973).

Once logistic regression outcomes $\pi(x)$ are continuous between 0 and 1, it is often necessary to determine a cut-off point, or *threshold*, to decide whether, given a regression result, the occurrence of the modelled event should be classified as *positive* or *negative* (i.e. SACZ present or absent, respectively). Only after the establishment of this threshold, becomes the binary classification of model results possible and more objectively interpretable. For an arbitrary threshold h , true positives (TP) and true negatives (TN) represent the correct classifications of the model estimates. TP equals to the count of days in which the model outcome $\pi(x)$ was greater than the established threshold h , $\pi(x) > h$, and the SACZ was observed, $Y = 1$, simultaneously. TN equals to the count of days in which the model outcome was less than h and the SACZ was not configured: $\pi(x) < h$ and $Y = 0$. On the other hand, the amount of mistaken classifications are expressed in counts of false positives (FP, or type I errors), when $\pi(x) > h$ and $Y = 0$, and false negatives (FN, or type II errors), when $\pi(x) < h$ and $Y = 1$.

Still for an arbitrary threshold, the proportions of correct classifications are measured in terms of the true positive rate (TPR), or Sensitivity, the percentage of positives correctly classified as such (Eq. 3); and the true negative rate (TNR), or Specificity, the percentage of negatives correctly classified (Eq. 4). These are two key-quantities in the performance assessment of a classifier and may be interpreted as the conditional probabilities of a correct positive classification, given the presence of the modelled event and the conditional probability of a correct negative classification, given the absence of the event, respectively. Analogously, the complementary proportions of incorrect classifications, types I and II error rates, are expressed by the false positive rate (FPR), the percentage of positives incorrectly classified as such (Eq. 5); and the false negative rate (FNR), the percentage of incorrectly negative classifications (Eq. 6).

$$\begin{aligned} \text{TPR} &= \frac{\text{Correct positive classifications}}{\text{Total positive observations}} \\ &= \frac{TP}{TP + FN} = P(\pi(x) > h | Y = 1) \end{aligned} \quad (3)$$

$$\begin{aligned} \text{TNR} &= \frac{\text{Correct negative classifications}}{\text{Total negative observations}} \\ &= \frac{TN}{TN + FP} = P(\pi(x) < h | Y = 0) \end{aligned} \quad (4)$$

$$\begin{aligned} \text{FPR} = 1 - \text{TNR} &= \frac{\text{Incorrect positive classifications}}{\text{Total negative observations}} \\ &= \frac{FP}{TN + FP} = P(\pi(x) > h | Y = 0) \end{aligned} \quad (5)$$

$$FNR = 1 - TPR = \frac{\text{Incorrect negative classifications}}{\text{Total positive observations}} \\ = \frac{FN}{TP + FN} = P(\pi(x) < h|Y = 1) \quad (6)$$

The ROC graph exhibits pairs of TPR and FPR for the entire domain of thresholds. The AUC equals to the area under the ROC curve and serves as a “one-number” quantity to represent a classifier’s performance, proportional to its overall accuracy, useful when a threshold is not yet set. AUC assumes values between 0 (worst performance) and 1 (best performance). Random guessing yields $AUC \approx 0.5$ (Swets 1988).

Supposing a real process f (i.e. the SACZ configuration) and two candidate models g_1 and g_2 , the AIC quantifies the amount of information lost in representing f by g_1 in respect with g_2 . Derived from information theory and based on the Kullback–Leibler divergence, AIC is a function of a model’s likelihood, L , and its number of parameters, K = number of covariates + 1 (the intercept), as in Eq. 7.

$$AIC = -2 \log(L) + 2K \quad (7)$$

AIC may be also calculated as a function of (and proportional to) the residual sum of squares (RSS) as in Eq. 8. Note that, in both manners, AIC is penalized by increased number of explanatory variables ($K-1$).

$$AIC = n \log\left(\frac{RSS}{n}\right) + 2K \quad (8)$$

where

$$RSS = \sum_{i=1}^n [y_i - \hat{\pi}(x)_i]^2$$

and y_i is the i^{th} observed value of the variable to be predicted and $\hat{\pi}(x)_i$ its estimated value by the regression model.

Although single AIC measures may be obtained for each model, individual values are meaningless and should be only applied to compare the relative fit of models trained on the same set of observations. Therefore, AIC variations among model setups, $\Delta AIC_{\text{model}}$, are analysed with respect to the minimum among comparable peers, $\Delta AIC_{\text{model}} = AIC_{\text{model}} - \min(AIC)$, instead of their absolute values, AIC_{model} . The Brier score (BS) (Brier 1950) is another commonly applied GOF assessment measure in logistic regression (Eq. 9). It represents the mean squared error and is therefore proportional to AIC as a function of RSS, differing from it mainly at lacking the term that penalizes the metrics as the number of covariates used increases.

$$BS = \frac{1}{n} \sum_{i=1}^n [\hat{\pi}(x)_i - y_i]^2 \quad (9)$$

Lastly, McFadden’s pseudo R^2 (hereafter \bar{R}^2) is a measure of proportional increase in a model’s maximized likelihood, L , with respect to the maximum likelihood of the model without predictors (the null model, only with intercept), L_0 (Eq. 10). Important to note that \bar{R}^2 is not directly comparable to ordinary least squared R^2 from linear regression. Its values are typically much lower and only applicable to relative assessment of GOF of models adjusted on the same set of observations.

$$\bar{R}^2 = 1 - \frac{\log(L)}{\log(L_0)} \quad (10)$$

5.4 Stepwise building, training and validation

A stepwise method was applied to define the best combination of explanatory variables for the regression models: PCs were added one at a time to each model configuration, and their statistical significance was assessed, as suggested by (Hosmer and Lemeshow 2004). Explanatory variables were discarded upon a $p > 0.05$. The order of inclusion of variables in a setup was the proportion of variance explained by each PC. In addition, for each new configuration trial, a tenfold cross-validation methodology was followed: all observations (days) were randomly assigned to ten same-size sets (folds) of 424 days, from which nine were used for the regression parameters estimation (training set) and the one remaining set was used to validate the model fit (test set). This was repeated ten times, in order that each set is once used for validation and every data point is used for training nine times (an important and common strategy to avoid the overfitting of predictive models).

The optimum model setups are defined considering a trade-off between minimum number of covariates and best fit, assessed in conjunction by the metrics strategies already described. The evolution of AUC, \bar{R}^2 and ΔAIC along the step-wise building trials are presented in Figs. 13, 14 and 15, respectively. A more rapid increase in fit of models is seen in the first stepwise building trials at all three GOF metrics used, while asymptotic fit increase is observed along with the addition of all possible predictors in the last steps, especially observing AUC and \bar{R}^2 . Once AIC expresses a decrease in GOF proportional to the number of covariates in the model configuration, ΔAIC steps present an actual minimum at the optimum fit, once it increases with the addition of more explanatory variables. This is visible in the validation runs of all datasets, which allows a more objective definition of the best model configurations. These patterns are observed within the tenfold training/validation runs and not any substantial discrepancies are noticed. Although the dispersion of GOF metrics at validation sets is always larger than their dispersion at training sets, their means vary accordingly along model building steps. For the northernmost type

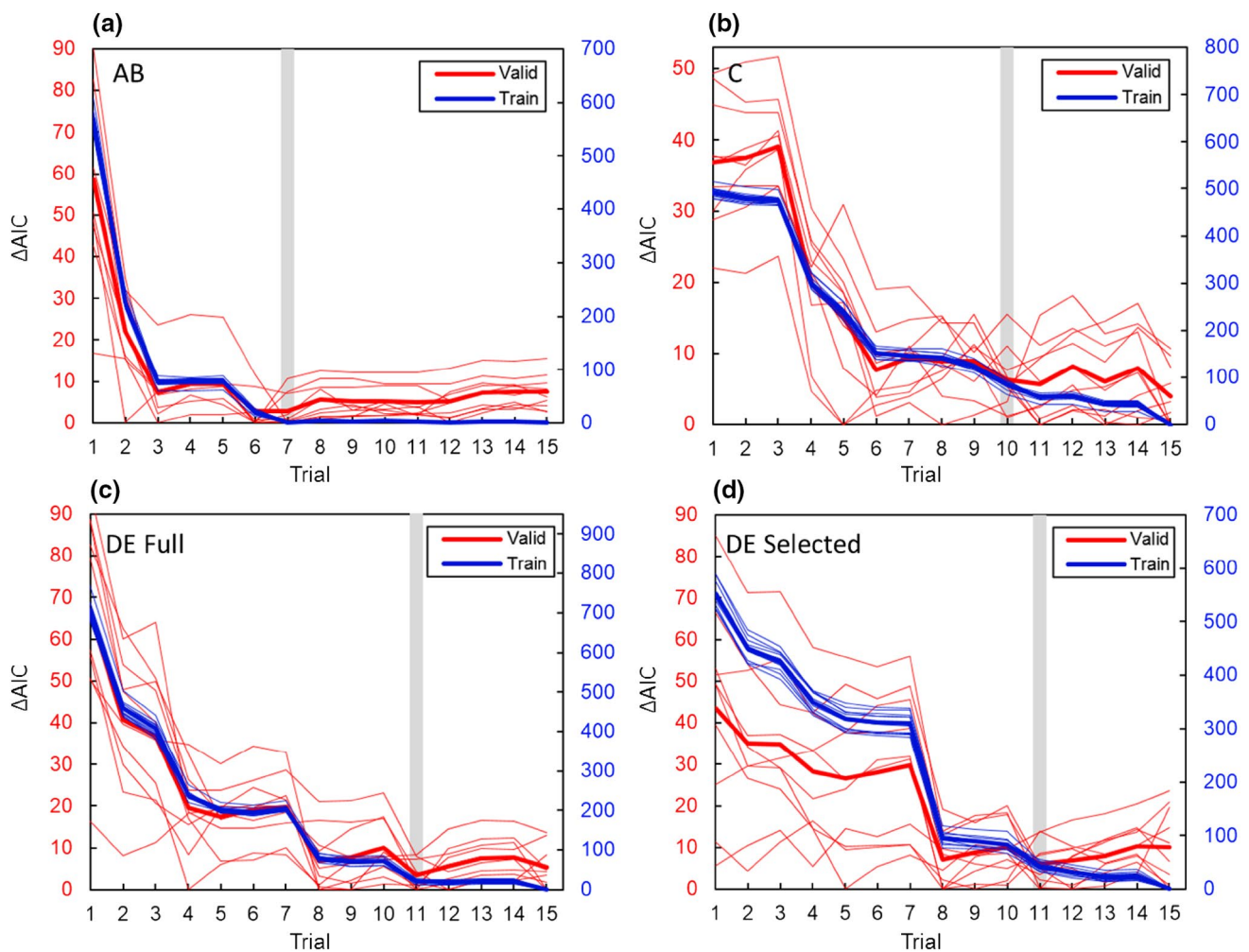


Fig. 15 Same as in Fig. 13, but for ΔAIC . Note that although magnitudes vary significantly between training and validation sets (values are plotted on separate axes), the shape of curves is similar along the model building steps

of SACZ (AB), the best model configuration was composed by 5 explanatory variables (PCs 1, 2, 3, 6 and 7) defined in the 7th trial of the stepwise model-building method. For the mean type of SACZ (C), the best model configuration was composed by 8 predictors (PCs 1, 2, 4, 5, 6, 8, 9 and 10) defined in the 10th trial. For the southernmost type of SACZ (DE), both datasets achieved the best model configuration at the 11th trial composed by 8 PCs for the Full data (1, 2, 3, 4, 5, 8, 9 and 11) and 9 PCs for the Selected data (1, 2, 3, 4, 5, 6, 8, 9 and 11). Following the 10-fold cross-validation methodology, these optimum model configurations are taken for the final 11th parameter estimation, in which the entire series are used and not any data points are left out for validation.

The average AUCs obtained from cross-validation (CV) and its associated standard deviations are compared to the final models' in Table 1, where substantial discrepancies are not observed. Comparison of the DE CV AUCs indicate a slightly better performance of the model composed by the

full dataset than the selected dataset. Among different latitudinal types of SACZ, the mean C SACZ type yielded a subtle better performance model, while the southernmost DE SACZ type presented the worst AUC metrics, though all values are greater than 0.8. The estimated parameters $\hat{\beta}$ are also compared in this same manner and all final parameters are shown within the range of the CV mean \pm one standard deviation (Table 2).

6 Threshold analysis

A daily operational use of the indices requires that their continuous values are transformed back into a binary response, which demands that objective decision thresholds, or operating cut-off limits (hereafter, h), are determined: outcomes greater than h are classified as positive (SACZ occurs) and outcomes less than h are classified as negative (SACZ does

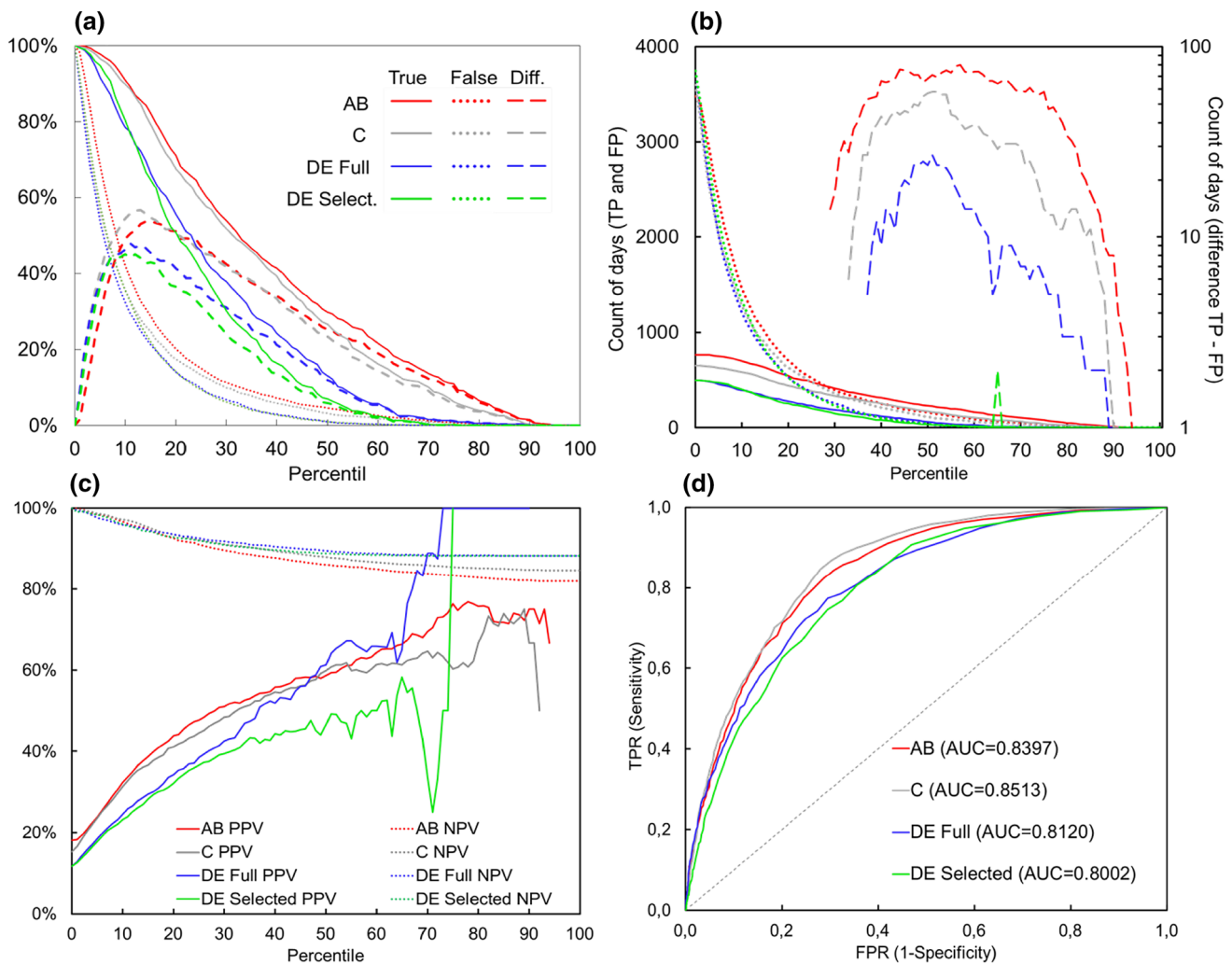


Fig. 16 **a** TPR (full lines), FPR (dotted lines) and Youden's statistics (TPR – FPR, dashed lines) on the vertical axis evaluated for each index threshold (percentiles) along the horizontal axis; **b** TP (full lines), FP (dotted lines) plotted on the left-hand vertical axis and the difference TP – FP (dashed) on the logarithmic right-hand vertical axis, per index percentile; **c** PPV (full lines) and NPV (dotted lines);

and **d** ROC curves with corresponding AUC values. All graphs show curves for the four defined model setups and three SACZ types, following the same colour code: northernmost type (AB, red), mean type (C, grey), southernmost type with full dataset (DE Full, blue) and southernmost type with selected PCs set (DE Selected, green)

not occur). For this purpose, the $[0,1]$ range of the indices was equally divided into 100 percentiles. All the 99 values between percentiles, plus the boundary values of 0 and 1,

were used as thresholds and tested for sensibility, specificity and other metrics, as following described.

The variation of TPR and FPR among 100 possible index thresholds is shown in Fig. 16a. The difference between these two quantities, which also equals to the difference $TNR - FNR$, is known as J statistics, or Youden's index (Youden 1950, hereafter YI), and measures the difference between proportions of correctly and incorrectly classified events. YI is null in the lowest and highest possible thresholds. Not any negative values are observed and maximum YI values are reached around the tenth threshold. The thresholds in which YI values are maxima are hereafter referred to as H_1 .

Table 1 Averaged cross-validation (CV) AUCs, their associated standard deviations and the final model's AUCs (parameters estimated on entire datasets) for each SACZ-type

	CV AUC average	CV AUC STD	Final AUC
AB	0.8372	0.0267	0.8397
C	0.8478	0.0228	0.8513
DE full	0.8094	0.0246	0.8120
DE selected	0.7956	0.0190	0.8002

Table 2 Mean and standard deviations of cross-validation (CV) parameters in comparison with the final model

	Intercept	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
AB										
CV mean	-2.084	-0.689	0.956	0.736	0.609	-0.547	-	-	-	-
CV STD	0.040	0.028	0.042	0.048	0.050	0.074	-	-	-	-
Final	-2.083	-0.689	0.955	0.736	0.609	-0.546	-	-	-	-
C										
CV mean	-2.434	-0.956	0.217	0.804	-0.679	-0.787	-0.451	-0.558	0.815	-
CV STD	0.032	0.024	0.033	0.040	0.040	0.077	0.075	0.100	0.078	-
Final	-2.433	-0.956	0.217	0.803	-0.678	-0.786	-0.451	-0.557	0.815	-
DE Full										
CV mean	-2.690	0.150	-1.001	-0.357	-1.028	-0.468	-1.125	-0.332	-0.946	-
CV STD	0.070	0.030	0.039	0.038	0.075	0.074	0.080	0.080	0.075	-
Final	-2.689	0.150	-1.000	-0.357	-1.027	-0.468	-1.125	-0.331	-0.946	-
DE Selected										
CV mean	-2.637	0.234	-0.868	-0.416	-0.839	0.643	-0.607	-1.848	0.522	1.665
CV STD	0.046	0.031	0.048	0.080	0.050	0.095	0.098	0.124	0.145	0.185
Final	-2.636	0.234	-0.867	-0.415	-0.838	0.642	-0.606	-1.845	0.522	1.663

However, important to note that the amount of days with SACZ and without SACZ configurations differ considerably (e.g. 499 SACZ-days and 3746 non-SACZ days in AB). Therefore, a maximum difference in proportions (TPR – FPR) does not necessarily mean a positive difference in count of days. For example, while taking the 30% or 40% lowest SACZ index values as thresholds, the amount of incorrectly classified positives (FP) is remarkably greater than the amount of the correctly classified positives (TP), although FP decreases rapidly and equals to TP, which continues to decrease, but in a less steep rate (Fig. 16b). The threshold in which (TP = FP), or in other words, where the number of correctly and incorrectly classified events is the same, is here defined as H_2 . The difference (TP – FP) reaches maximum values at around the 60th and 70th percentiles, and this positive difference in number of days also vary somewhat considerably different among SACZ types: 80 days for AB while only 2 days for DE yielded with the selected-PCs dataset. This threshold limit where (TP – FP) is maximum is here defined as H_3 , which it reflects the idea of the maximum YI threshold, H_1 , however maximizing the difference of correct and incorrect positive classifications in terms of absolute count of days, instead of proportions of different populations.

Yet another useful approach to assess the performance of a binary classifier is to define the *predictive values* associated to each possible threshold. While sensitivity and specificity are defined as ratios of correct classifications over the actual observations of the modelled event, positive and negative predictive values (PPV and NPV, respectively) are defined as the ratios of the count of correct classifications over the total classifications, both correct and incorrect (Eqs. 11 and 12). These quantities express the probability of a classification to be correct among the classifications of the same kind.

$$PPV = TP / (TP + FP) \quad (11)$$

$$NPV = TN / (TN + FN) \quad (12)$$

The graph in Fig. 16c presents the variation of PPV and NPV along 100 increasing possible indices values taken as thresholds. Note that NPV figures frequently between 90% and 100% and never lower than 80%, suggesting very reliable predictions of non-SACZ days. PPV starts at around 10% and 20% with $h=0$ and reaches maxima between 70% and 80% for the northernmost SACZ types (AB and C). For the southernmost SACZ type DE, PPV reaches 100%, meaning that above this thresholds not any incorrect positive classifications of SACZ occur. ROC curves are presented in Fig. 16d, showing all AUC values (as in Table 1) greater than 0.8.

Thresholds H_1 , H_2 and H_3 , identified as critical points in the above mentioned curves, are suggested as key values for the operational predictive use of the model outputs. Their corresponding TPR, TNR, FPR, FNR, PPV and NPV are

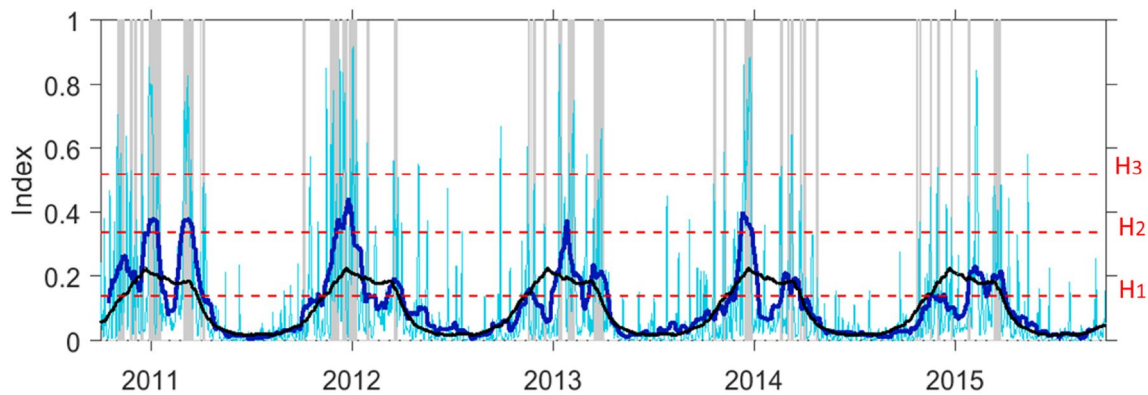


Fig. 17 Mean SACZ type (C) daily index (dark blue), its 31-day moving average (black heavy line) and observed SACZ-days in which the cloud band directly acted on areas A and B (light blue bars). Dashed red lines indicate the critical thresholds H1, H2 and H3

presented in Table 3. This is done here to avoid the arbitrary classification of model outputs, however other thresholds may and should be also defined, based on a predictor's particular interest, such as to select a very specific classification, where PPV is maximum or FPR is set to 10%, or a very sensitive classification, setting TPR = 90%, for example.

7 SACZ indices as proxies for precipitation anomalies

Once defined the best model configurations for the different SACZ types, it was possible to calculate the SACZ indices from the entire 30-year NCEPR2 daily series, from January 1986 to December 2015, and its respective daily climatological means, including months from May to September, and years from 1986 to 1994 which were not used in the model

adjustment and former PCA-based transformation. Analogously, this calculation is possible at any other period when the required dynamical fields are available.

An example period of the daily index and its long-term means for the mean SACZ type, which affects BRSE, comprehending the states of São Paulo and Rio de Janeiro, is presented in Fig. 17, from October/2010 to April/2015, along with a 31-point smooth, the observed SACZ events recorded for this region and the proposed thresholds. The SAMS shows significant intraseasonal variability (Nogués-Paegle et al. 2000; Jones and Carvalho 2002; Gan et al. 2004; Muza et al. 2009) and the index is capable of capturing its aspects associated with SACZ episodes, as in the 2010/11 and 2012/13 SAMS wet periods, when SACZ configurations and break phases followed by the index are clearly visible. At a first look at the SAMS's interannual variability, the SACZ index shows, in comparison to its

Table 3 TPR, FPR, TNR, FNR, PPV and NPV for the indices critical thresholds and each SACZ type

SACZ type	SACZ index value	TPR (%) (sensitivity)	FPR (%)	TNR (%) (specificity)	FNR (%)	PPV (%)	NPV (%)
H₁ (maximum YI)							
AB	0.15	84.4	30.7	69.3	15.6	37.7	95.3
C	0.14	84.6	27.9	72.1	15.4	36.1	95.7
DE full	0.12	77.4	29.5	70.5	22.6	28.7	95.0
DE selected	0.13	74.5	29.5	70.5	25.5	27.0	94.5
H₂ (TP=FP)							
AB	0.30	55.3	11.8	88.2	44.7	50.8	89.9
C	0.34	48.1	8.5	91.5	51.9	50.5	90.7
DE full	0.38	27.7	3.6	96.4	72.3	50.9	90.9
DE selected	0.60	2.6	0.3	99.7	97.4	50.0	88.5
H₃ (maximum TP-FP)							
AB	0.58	25.2	3.2	96.8	74.8	63.1	85.4
C	0.52	25.3	2.9	97.1	74.7	60.7	87.8
DE full	0.52	12.2	0.9	99.1	87.8	64.2	89.4
DE selected	0.66	1.4	0.1	99.9	98.6	58.3	88.4

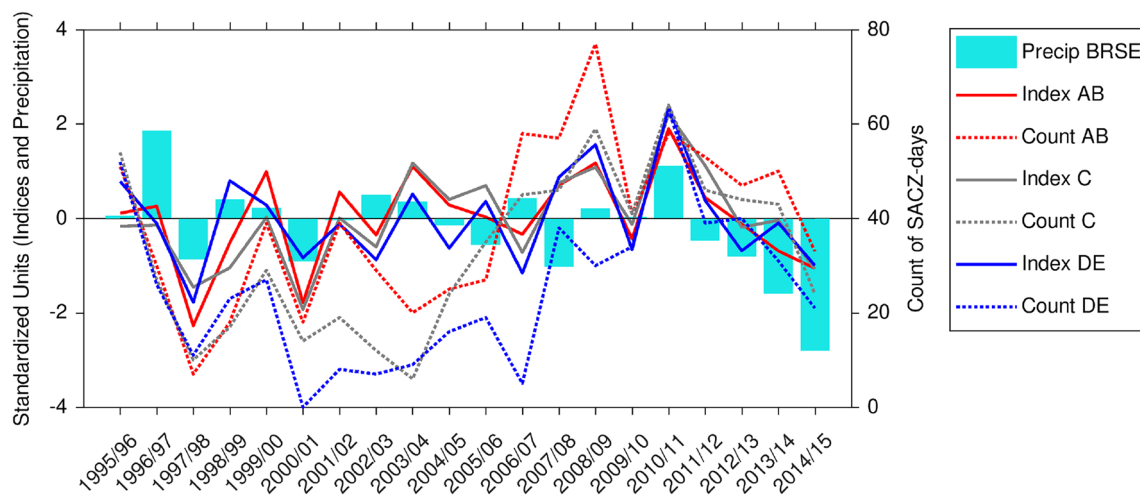


Fig. 18 Annual series of: standardized October–April (SAMS wet period) averages of the SACZ indices anomalies (left axis), standardized southeast Brazil (BRSE) precipitation anomalies (left axis) and

count of SACZ-days between December 1995 and April 2015 per wet period (right axis). Correlation coefficients among these yearly series are presented in Table 4

calculated climatology, both wet and dry anomalies. The summer of 2014–2015, for example, is known for the Brazilian southeast region having experienced severe droughts and public water and energy rationing programmes (Coelho et al. 2016b), where the index presented values of magnitude consistently lower than its long term mean.

October–April annual means of daily anomalies of the SACZ indices, as well as the count of SACZ days and BRSE (Fig. 6a) precipitation anomalies are compared and displayed in Fig. 18. The persistent under-average rainfall anomalies after summer 2012–2013 are followed by consistent negative SACZ index anomalies. The indices show stronger correlations with precipitation anomalies in BRSE than the count of SACZ days, in the interannual scale, although indices and count of SACZ days also present significant correlations (Table 4). In addition, although not all SACZ types directly affect BRSE regarding rainfall anomalies, here we show that all indices present stronger correlation with precipitation anomalies in BRSE than the sole count of SACZ-days in this region.

The indices were also compared to daily gridded precipitation anomalies from CPC (Xie et al. 2007). Correlation coefficients were calculated between the indices and precipitation daily anomalies (both with respect to the 1986–2015 monthly climatology) taking only periods between October and April from 1986 to 2015, and presented moderate strength, up to around 0.5 as shown in Fig. 19. The nuclei of maximum correlation are in agreement the influence regions of the northernmost, mean and southernmost SACZ types. Not only are the different meridional SACZ-types clear in the spatial correlations, but also their inversely proportional relationship with LPRB precipitation becomes spatially evident, to characterize the “seesaw”, or dipole in precipitation

pattern in SA (Nogués-Paegle and Mo 1997; Díaz and Aceituno 2003; Liebmann et al. 2004; Mattingly and Mote 2016).

8 Summary and conclusions

The SACZ has direct impacts on the rainfall regime of the most densely populated and economically important regions in Brazil, accounting for around one quarter of the rainfall volumes in BRSE during the SAMS wet period (October–April) on average, with peaks of 56% in March and 41% in January. In the present study, 19 wet periods of the SAMS are examined and three modes of SACZ configurations are defined in terms of the latitudinal variation of its cloud band position: northernmost (AB), mean (C) and southernmost (DE) SACZ types. This spatial categorization of SACZ-days allowed the identification of a latitudinal shift behaviour of the SACZ cloud band, which, on average, shows itself in its northernmost position in the months of the SAMS onset (October and November), gradually moves southward to reach its southernmost mean position in February, and returns to its mean northernmost initial position during the SAMS demise (March and April). Quadro (1994) has suggested a northernmost preferential position of the SACZ during December, and a southernmost preferential position later during summer, but not yet the higher probabilities of northernmost configurations by the end of the SAMS wet period. As far as the authors are concerned, this is the first time this full one-period intraseasonal cycle (southwards and northwards shift between October and April) in the SACZ climatological position is identified and quantified.

Among the different latitudinal SACZ types, differences in dynamics are also identified. Dynamical patterns of

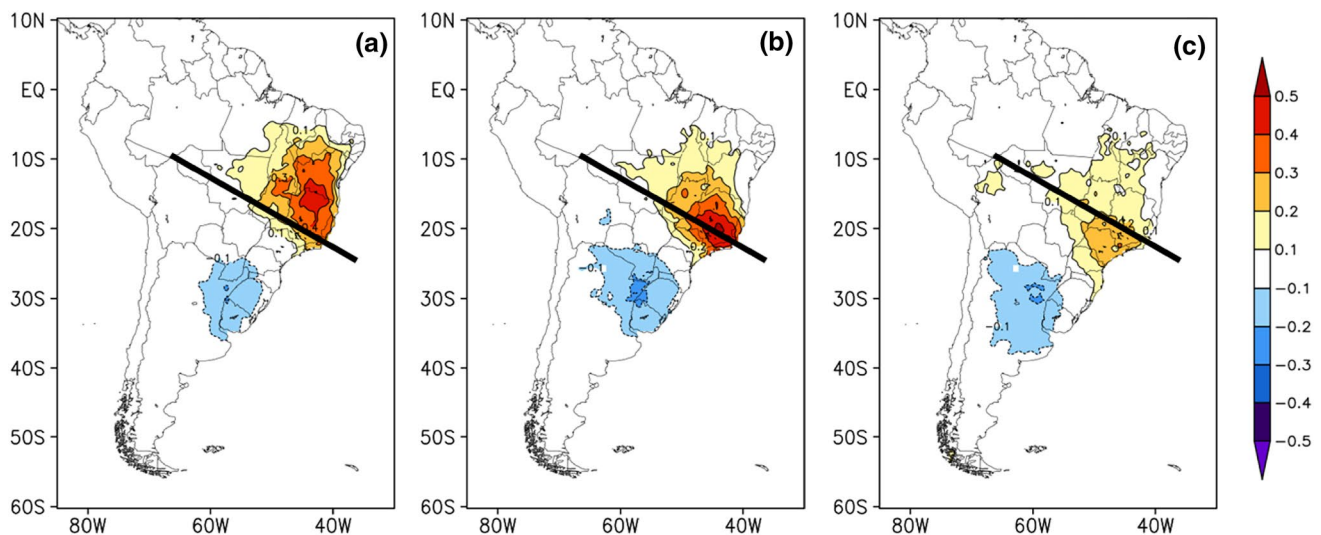


Fig. 19 Maps of correlation coefficients between the full 30-year daily SACZ indices and rain gauge-based precipitation anomalies from CPC (Xie et al. 2007)

the northernmost and mean SACZ types are in agreement with the South American precipitation dipole, in which the SACZ presence (absence) inhibits (supports) convection in the LPRB, in accordance with several studies (e.g. Nogués-Paegle and Mo 1997; Salio et al. 2002; Díaz and Aceituno 2003; Liebmann et al. 2004; Marengo et al. 2004). Results showed anomalous weakened (enhanced) low level Amazonian moisture transport via SALLJ toward the LPBR region during the presence (absence) of all three SACZ types, in accordance to Herdies et al. (2002), and suggesting that the precipitation dipole and the eastward deflection of moisture transport direction is not so strongly affected by latitudinal variations of the SACZ position.

Anomalous circulation patterns observed in all three meridional SACZ types show cyclonic anomalies with baroclinic structure from 850 to 200 hPa. These results are in agreement with several other studies, in which trough-like anomalies are associated with active phases within the

SAMS wet period and SACZ configurations (e.g. Jones and Carvalho 2002; Gan et al. 2004; Carvalho et al. 2011; Ma et al. 2011). The position of the SACZ cloud band moves meridionally, while the trough anomaly moves zonally among the three SACZ modes: the cyclonic centers associated to the SACZ appear displaced westward onto the continent while the SACZ places itself southwards. The predominant cyclonic anomalies in 850 hPa are also more clearly distinguishable in the North and Mean-SACZ modes. In the South-SACZ mode, the cyclonic anomaly is restricted over the continent, and an anti-cyclonic anomaly, north-eastward from the trough is identified.

The northernmost and mean SACZ composites show a Rossby wave train propagating from southern Pacific Ocean, in accordance to previous studies (e.g. Grimm and Silva Dias 1995; Liebmann et al. 1999). In the southernmost SACZ mode, this pattern is not as clearly visible.

Table 4 Correlation coefficients between SACZ indices' anomalies yearly averaged from October to April, count of SACZ days per SAMS wet period (October–April) and precipitation anomalies in BRSE also averaged in October–April (annual series presented in Fig. 18)

	SACZ Index AB	SACZ Index C	SACZ Index DE	Count SACZ AB	Count SACZ C	Count SACZ DE	BRSE Precip.
SACZ Index AB	1						
SACZ Index C	0.889	1					
SACZ Index DE	0.795	0.765	1				
Count SACZ AB	0.566	0.544	0.519	1			
Count SACZ C	0.482	0.558	0.566	0.892	1		
Count SACZ DE	0.492	0.576	0.652	0.584	0.794	1	
BRSE Precip.	0.510	0.332	0.372	0.071	0.081	0.106	1

Bold coefficients are statistically significant at a 95% confidence level

Northernmost and mean SACZ types presented their associated preferred subsidence branch located south-westward from the convective heat source, in agreement to Gandú and Silva Dias (1998). However, southernmost SACZ-type configurations did not fully follow this pattern, once their preferred subsidence branch is located north-eastward from the cloud band.

The aforementioned atmospheric dynamical patterns bring to evidence fundamental differences observed between the North and Mean SACZ types, in comparison with the South SACZ type. The latter shows characteristics of mid-latitude baroclinic waves, leading to the occurrence of cold frontal systems acting over southern Brazil. It is therefore likely that some of the days with SACZ configurations, as officially determined by the National Weather Forecast and Climate Studies (CPTEC/INPE), presented dynamics more closely related to cold fronts, which frequently act in southern Brazil.

The identification of significantly anomalous patterns in dynamics allowed the selection of explanatory variables, whose principal components were used to train and (cross-) validate logistic regression models, for each of the three meridionally different SACZ types. Predictors include the vertical and horizontal components of wind, horizontal wind divergence, vorticity and geopotential height anomalies, and do not include humidity, precipitation and OLR, for these being generally represented with lower skill by numerical models, and more directly influenced by parameterization schemes, important sources of uncertainties.

The developed indices are continuous between 0 and 1, thus objective thresholds, or cut-off values, are proposed to allow their operational use for the SACZ daily identification/forecast. Three thresholds are suggested in terms of sensitivity and specificity critical points. For the most sensitive threshold, H_1 , the index proposed for the mean SACZ type presents a simultaneous true positive rate of 85% and a false alarm rate of 28%, for example. Playing a trade-off in order to yield less false alarms, for example, the more specific threshold suggested, H_3 , presents a false alarm rate of less than 3% and a simultaneous true positive rate of 25%. These are possible cut-off points, although other values may be chosen based on fixed accuracy or predictive values and rates, as in accordance to the predictor's wish. For instance, fixing a negative predictive value at 90% for the mean SACZ type (i.e. the probability that a negative classification is incorrect, given all negative classifications, equals to 10%) yields a positive predictive value of 55% (i.e. the probability that a positive classification being correct, given all positive classifications). This translates to a true positive rate of 39% and a simultaneous false alarm rate of 6% (i.e. the probability that a positive classification matches a SACZ occurrence and the probability that a negative classification matches a SACZ absence, respectively).

For all proposed indices and still for a daily-scale analysis, the negative predictive values never reached rates lower than 80%, for any possible thresholds. For the southernmost SACZ types, the positive predictive values reached 100% at more specific thresholds, meaning that the probability of an incorrect positive classification, given all positive classifications, is null above certain index thresholds.

It was verified that error yielded in precipitation forecasts obtained from the model Global Forecast System (GFS) were often greater than those obtained from dynamical variables (i.e. wind components and geopotential height) (see Supplementary Material). Moreover, the SACZ indices, as forecasted by the model GFS, presented greater accuracy in comparison with precipitation forecast in Southeast Brazil. This translates to a gain in predictability of precipitation (associated to SACZ events) in important regions of Brazil. When testing the objective identification of SACZ days from GFS outputs, the rates of true positives were always greater or equal than the rates of false positives. In addition, a further investigation of a strong false alarm event brought to evidence that there were days, not officially classified as with SACZ configurations, which showed characteristics in dynamics very closely associated to a traditional SACZ event.

Apart from the daily forecast application, the SACZ indices presented here showed agreement with the SAMS wet period interannual precipitation anomalies. This raises the possibility for the quantification of the SACZ and the SAMS in different time scales, from weather prediction to climate studies.

Acknowledgements Open access funding provided by Max Planck Society. This project has been supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) from “Edital Universal” Grant number 454397/2014-3.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Ambrizzi T, Ferraz SET (2015) An objective criterion for determining the South Atlantic convergence zone. *Front Environ Sci* 3:1–9. <https://doi.org/10.3389/fenvs.2015.00023>
- Arshad SH, Hide DW (1992) Effect of environmental factors on the development of allergic disorders in infancy. *J Allergy Clin Immunol* 90:235–241. [https://doi.org/10.1016/0091-6749\(92\)90077-F](https://doi.org/10.1016/0091-6749(92)90077-F)

- Belsley DA (1991) A Guide to using the collinearity diagnostics. *Comput Sci Econ Manag* 4:33–50. <https://doi.org/10.1007/BF00426854>
- Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 30:1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Brier GW (1950) Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 78:1–3. <https://doi.org/10.1126/science.27.693.594>
- Brough HA, Liu AH, Sicherer S et al (2015) Atopic dermatitis increases the effect of exposure to peanut antigen in dust on peanut sensitization and likely peanut allergy. *J Allergy Clin Immunol* 135:164–170. <https://doi.org/10.1016/j.jaci.2014.10.007>
- Carvalho LM V, Jones C, Liebmann B (2002) Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic convergence zone. *J Clim* 15:2377–2394. [https://doi.org/10.1175/1520-0442\(2002\)015%3C2377:EPEISS%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C2377:EPEISS%3E2.0.CO;2)
- Carvalho LMV, Silva AE, Jones C et al (2011) Moisture transport and intraseasonal variability in the South America monsoon system. *Clim Dyn* 36:1865–1880. <https://doi.org/10.1007/s00382-010-0806-2>
- Carvalho LMV, Jones C, Posadas AND et al (2012) Precipitation characteristics of the South American monsoon system derived from multiple datasets. *J Clim* 25:4600–4620. <https://doi.org/10.1175/JCLI-D-11-00335.1>
- Chaves RR, Nobre P (2004) Interactions between sea surface temperature over the South Atlantic Ocean and the South Atlantic convergence zone. *Geophys Res Lett* 31:L03204. <https://doi.org/10.1029/2003GL018647>
- Coelho CAS, Cardoso DHF, Firpo MAF (2016a) Precipitation diagnostics of an exceptionally dry event in São Paulo, Brazil. *Theor Appl Climatol* 125:769–784. <https://doi.org/10.1007/s00704-015-1540-9>
- Coelho CAS, de Oliveira CP, Ambrizzi T et al (2016b) The 2014 southeast Brazil austral summer drought: regional scale mechanisms and teleconnections. *Clim Dyn* 46:3737–3752. <https://doi.org/10.1007/s00382-015-2800-1>
- CPTEC/INPE (2018a) Boletim Técnico. Retrieved from: <http://tempo.cptec.inpe.br/boletimtecnico/pt> (in Portuguese)
- CPTEC/INPE (2018b) Climanálise: Boletim de Monitoramento e Análise Climática. Retrieved from: <http://climanalise.cptec.inpe.br/~rcimanl/boletim/> (in Portuguese)
- Díaz A, Aceituno P (2003) Atmospheric circulation anomalies during episodes of enhanced and reduced convective cloudiness over Uruguay. *J Clim* 16:3171–3185. [https://doi.org/10.1175/1520-0442\(2003\)016%3C3171:ACADEO%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016%3C3171:ACADEO%3E2.0.CO;2)
- Dormann CF, Elith J, Bacher S et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:027–046. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Farrar DE, Glauber RR (1967) Multicollinearity in regression analysis: the problem revisited. *Rev Econ Stat* 49:92–107
- Figuerola SN, Satyamurty P & Silva Dias PL (1995) Simulations of the summer circulation over the South American region with an eta coordinate model. *J Atmos Sci* 52(10), 1573–1584. [https://doi.org/10.1175/1520-0469\(1995\)052%3C1573:SOTSCO%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052%3C1573:SOTSCO%3E2.0.CO;2)
- Flantua SGA, Boxel JH, Hooghiemstra H, Smaalen J (2007) Application of GIS and logistic regression to fossil pollen data in modelling present and past spatial distribution of the Colombian savanna. *Clim Dyn* 29:697–712. <https://doi.org/10.1007/s00382-007-0276-3>
- Gan MA, Kousky VE, Ropelewski CF (2004) The South America Monsoon circulation and its relationship to rainfall over west-central Brazil. *J Clim* 17:47–66. [https://doi.org/10.1175/1520-0442\(2004\)017%3C0047:TSAMCA%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017%3C0047:TSAMCA%3E2.0.CO;2)
- Gan MA, Rao VB, Moscati MCL (2005) South American monsoon indices. *Atmos Sci Lett* 6:219–223. <https://doi.org/10.1002/asl.119>
- Gandú AW, Silva Dias PL (1998) Impact of tropical heat sources on the South American Tropospheric Upper circulation and subsidence. *J Geophys Res* v 103:6001–6015. <https://doi.org/10.1029/97JD03114>
- Grimm AM, & Silva Dias PL (1995) Analysis of tropical–extratropical interactions with influence functions of a barotropic model. *J Atmos Sci* 52(20), 3538–3555. [https://doi.org/10.1175/1520-0469\(1995\)052%3C3538:AOTWI%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052%3C3538:AOTWI%3E2.0.CO;2)
- Grimm AM, Pal JS, Giorgi F (2007) Connection between spring conditions and peak summer monsoon rainfall in South America: role of soil moisture, surface temperature, and topography in eastern Brazil. *J Clim* 20:5929–5945. <https://doi.org/10.1175/2007JCLI1684.1>
- Guanche Y, Mínguez R, Méndez FJ (2014) Autoregressive logistic regression applied to atmospheric circulation patterns. *Clim Dyn* 42:537–552. <https://doi.org/10.1007/s00382-013-1690-3>
- Herdies DL, da Silva A, Silva Dias MA, Nieto Ferreira R (2002) Moisture budget of the bimodal pattern of the summer circulation over South America. *J Geophys Res* 107(D20):8075. <https://doi.org/10.1029/2001JD000997>
- Holton JR (2004) An introduction to dynamic meteorology. 4th edn. Elsevier Inc., Amsterdam
- Hosmer DW, Lemeshow S (2004) Applied logistic regression. 2nd edn. John Wiley & Sons, Hoboken
- Jolliffe IT (1982) A note on the use of principal components in regression. *J R Stat Soc Ser C (Applied Stat)* 31:300–303. <https://doi.org/10.2307/2348005> doi
- Jones C, Carvalho LM V (2002) Active and break phases in the South American monsoon system. *J Clim* 15:905–914. [https://doi.org/10.1175/1520-0442\(2002\)015%3C0905:AABPI%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C0905:AABPI%3E2.0.CO;2)
- Jorgetti T, da Silva Dias PL, de Freitas ED (2014) The relationship between South Atlantic SST and SACZ intensity and positioning. *Clim Dyn* 42(11–12):3077–3086
- Kalnay E, Mo KC, Paegle J (1986) Large-amplitude, short-scale stationary rossby waves in the southern hemisphere: observations and mechanistic experiments to determine their origin. *J Atmos Sci* 43:252–275. [https://doi.org/10.1175/1520-0469\(1986\)043%3C0252:LASSSR%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1986)043%3C0252:LASSSR%3E2.0.CO;2)
- Kanamitsu M, Ebisuzaki W, Woollen J et al (2002) NCEP–DOE AMIP-II reanalysis (R-2). *Bull Am Meteorol Soc* 83:1631–1643 + 1559. <https://doi.org/10.1175/BAMS-83-11-1631>
- Kodama Y (1992) Large-scale common features of subtropical precipitation zones (the Baiu Frontal Zone, the SPCZ, and the SACZ) Part I: Characteristics of subtropical frontal zones. *J Meteorol Soc Japan* 70:813–836. <https://doi.org/10.1248/cpb.37.3229>
- Kodama Y-M (1993) Large-scale common features of sub-tropical convergence zones (the Baiu Frontal Zone, the SPCZ, and the SACZ) Part II: conditions of the circulations for generating the STCZs. *J Meteorol Soc Japan* 71:581–610. <https://doi.org/10.1248/cpb.37.3229>
- Kousky VE, Gan MA (1981) Upper tropospheric cyclonic vortices in the tropical South Atlantic. *Tellus* 33:538–551. <https://doi.org/10.1111/j.2153-3490.1981.tb01780.x>
- Lenters JD, Cook KH (1997) On the origin of the Bolivian high and related circulation features of the south American climate. *J Atmos Sci* 54:656–678. [https://doi.org/10.1175/1520-0469\(1997\)054%3C0656:OTOOTB%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1997)054%3C0656:OTOOTB%3E2.0.CO;2)
- Liebmann B, Kiladis GN, Marengo JA, et al (1999) Submonthly convective variability over South America and the South Atlantic convergence zone. *J Clim* 12:1877–1891. doi: [https://doi.org/10.1175/1520-0442\(1999\)012%3C1877:SCVOSA%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012%3C1877:SCVOSA%3E2.0.CO;2)

- Liebmann B, Kiladis GN, Vera CS, et al (2004) Subseasonal variations of rainfall in South America in the vicinity of the low-level jet east of the Andes and comparison to those in the South Atlantic convergence zone. *J Clim* 17:3829–3842. [https://doi.org/10.1175/1520-0442\(2004\)017%3C3829:SVORI%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017%3C3829:SVORI%3E2.0.CO;2)
- Lima KC, Satyamurty P, Fernández JPR (2010) Large-scale atmospheric conditions associated with heavy rainfall episodes in Southeast Brazil. *Theor Appl Climatol* 101:121–135. <https://doi.org/10.1007/s00704-009-0207-9>
- Lomholt FK, Nielsen SF, Nordestgaard BG (2016) High alcohol consumption causes high IgE levels but not high risk of allergic disease. *J Allergy Clin Immunol* 138:1404–1413.e13. <https://doi.org/10.1016/j.jaci.2016.05.022>
- Ma HY, Ji X, Neelin JD, Mechoso CR (2011) Mechanisms for precipitation variability of the Eastern Brazil/SACZ convective margin. *J Clim* 24:3445–3456. <https://doi.org/10.1175/2011JCLI4070.1>
- Marengo JA, Soares WR, Saulo C, Nicolini M (2004) Climatology of the low-level jet east of the Andes as derived from the NCEP-NCAR reanalyses: Characteristics and temporal variability. *J Clim* 17:2261–2280. [https://doi.org/10.1175/1520-0442\(2004\)017%3C2261:COTLJE%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017%3C2261:COTLJE%3E2.0.CO;2)
- Massy WF (1965) Principal components regression in exploratory statistical research. *J Am Stat Assoc* 60:309:234–256. <https://doi.org/10.1080/01621459.1965.10480787>
- Mattingly KS, Mote TL (2016) Variability in warm-season atmospheric circulation and precipitation patterns over subtropical South America: relationships between the South Atlantic convergence zone and large-scale organized convection over the La Plata basin. *Clim Dyn* 48:1–23
- McFadden D (1973) Conditional logit analysis of qualitative choice behavior. *Front. Econom.* 105–142
- Muza MN, Carvalho LMV, Jones C, Liebmann B (2009) Intraseasonal and interannual variability of extreme dry and wet events over southeastern South America and the subtropical Atlantic during austral summer. *J Clim* 22:1682–1699. <https://doi.org/10.1175/2008JCLI2257.1>
- Neto OBS, Escobar J, Silva PES (2010) Método objetivo para identificar episódios de Zonas de Convergência de Umidade (ZCOU) no ambiente operacional do Centro de Previsão de Tempo e Estados Climáticos - CPTEC. In: Congresso Brasileiro De Meteorologia. pp 1–5
- Nielsen DM, Cataldi M, Belém AL, Albuquerque ALS (2016) Local indices for the South American monsoon system and its impacts on Southeast Brazilian precipitation patterns. *Nat Hazards* 83:909–928. <https://doi.org/10.1007/s11069-016-2355-4>
- Nogués-Paegle J, Mo KC (1997) Alternating wet and dry conditions over South America during summer. *Mon Weather Rev* 125:279–291. [https://doi.org/10.1175/1520-0493\(1997\)125%3C0279:AWADCO%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125%3C0279:AWADCO%3E2.0.CO;2)
- Nogués-Paegle J, Byerle L a, Mo KC, et al (2000) Intraseasonal modulation of South American summer precipitation. *Mon Weather Rev* 128:837–850. [https://doi.org/10.1175/1520-0493\(2000\)128%3C0837:IMOSAS%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128%3C0837:IMOSAS%3E2.0.CO;2)
- ONS (2015) Annual Report and Financial Statements. Operador Nac do Sist Elétrico 206
- Quadro MFL de (1994) ESTUDO DE EPISÓDIOS DE ZONA DE CONVERGÊNCIA DO ATLÂNTICO SUL (ZCAS) SOBRE A AMÉRICA DO SUL. INPE Instituto Nacional de Pesquisas Espaciais
- Quadro MFL, Berbery EH, Silva Dias MA, Herdies DL, Gonçalves LG (2013) The atmospheric water cycle over South America as seen in the new generation of global reanalyses. In *AIP Conf Proc* 1531(1): 732–735. <https://doi.org/10.1063/1.4804874>
- Raia A, Cavalcanti IFA (2008) The life cycle of the South American monsoon system. *J Clim* 21:6227–6246. <https://doi.org/10.1175/2008JCLI2249.1>
- Raja NB, Çiçek I, Turkoglu N et al (2016) Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Nat Hazards* 1–24. <https://doi.org/10.1007/s11069-016-2591-7>
- Reboita MS, Gan MA, Rocha RP, Da Ambrizzi T (2010) Regimes de precipitação na América do Sul: uma revisão bibliográfica. *Rev Bras Meteorol* 25:185–204. <https://doi.org/10.1590/S0102-77862010000200004>
- Salio P, Nicolini M, Saulo AC (2002) Chaco low-level jet events characterization during the austral summer season. *J Geophys Res Atmos* 107:1–17. <https://doi.org/10.1029/2001JD001315>
- Schaefer RL (1986) Alternative estimators in logistic regression when the data are collinear. *J Stat Comput Simul* 25:75–91. <https://doi.org/10.1080/00949658608810925>
- Seluchi ME, Chou SC (2009) Synoptic patterns associated with landslide events in the Serra do Mar, Brazil. *Theor Appl Climatol* 98:67–77. <https://doi.org/10.1007/s00704-008-0101-x>
- Silva AE, Carvalho LMV (2007) Large-scale index for South America monsoon (LISAM). *Atmos Sci Lett* 8:51–57. <https://doi.org/10.1002/asl.150>
- Silva Dias PL, Schubert WH, DeMaria M (1983) Large-scale response of the tropical atmosphere to transient convection. *J Atmos Sci* 40:2689–2707
- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293. <https://doi.org/10.1126/science.3287615>
- van der Wiel K, Matthews AJ, Stevens DP, Joshi MM (2015) A dynamical framework for the origin of the diagonal South Pacific and South Atlantic Convergence Zones. *Q J R Meteorol Soc* 141:1997–2010. <https://doi.org/10.1002/qj.2508>
- Vera C, Higgins W, Amador J et al (2006) Toward a unified view of the American monsoon systems. *J Clim* 19:4977–5000. <https://doi.org/10.1175/JCLI3896.1>
- Virji H (1981) A preliminary study of summertime tropospheric circulation patterns over South America estimated from cloud winds. *Mon Weather Rev* 109:599–610. [https://doi.org/10.1175/1520-0493\(1981\)109%3C0599:APSOST%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109%3C0599:APSOST%3E2.0.CO;2)
- Xie P, Chen M, Yang S et al (2007) A gauge-based analysis of daily precipitation over East Asia. *J Hydrometeorol* 8:607–626. <https://doi.org/10.1175/JHM583.1>
- Youden WJ (1950) Index for rating diagnostic tests. *Cancer* 3:32–35. [https://doi.org/10.1002/1097-0142\(1950\)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3)
- Zhou J, Lau KM (1998) Does a monsoon climate exist over South Am J Clim 11:1020–1040. [https://doi.org/10.1175/1520-0442\(1998\)011%3C1020:DAMCEO%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011%3C1020:DAMCEO%3E2.0.CO;2)