Variability of Winter Frosts in central South America: quantifying mechanisms with Decision Trees

Soledad Collazo[1,2,3*], Ricardo García-Herrera[1,3]

[1]Departamento de Física de la Tierra y Astrofísica, Facultad de Ciencias Físicas, Universidad Complutense de Madrid (UCM), Plaza de las Ciencias 1, Madrid, Spain

[2]Departamento de Ciencias de la Atmósfera y los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (FCEN, UBA), Buenos Aires, Argentina

[3]Instituto de Geociencias (IGEO), Consejo Superior de Investigaciones Científicas–Universidad Complutense de Madrid (CSIC–UCM), Madrid, Spain

* scollazo@ucm.es

Tel: +34 91394 4490

Abstract

Agricultural production in Central South America (CSA) is substantially influenced by frost events. This study characterises and quantifies the physical processes leading to frost conditions in CSA from 1979 to 2022, focusing on three innovative aspects: regional frost properties, a novel multi-parametric upper-level jet description, and the quantification of underlying mechanisms through decision trees (DTs). The regionalisation analysis identifies five homogeneous frost regions in CSA. In all regions, the events tend to occur more frequently during the La Niña phase. Moreover, a significant increase in the frequency of widespread frost events has been observed in the Argentinean Pampas during the study period, primarily due to negative trends in minimum temperatures. Furthermore, the synoptic mechanisms triggering frosts, such as cold fronts and post-frontal anticyclones enhanced by subsidence near the subtropical jet (STJ) entrance, have not shown major long-term changes. To describe the jets, we compute six parameters for the STJ and seven for the polar front jet, including latitude, intensity, height, tilting, longitudinal extent, and branch number. DTs are used to identify key jet parameters linked to frost events, such as the latitude, longitudinal extent, and tilt of the Atlantic STJ. Frost likelihood increases when the STJ is north of 31°S, and the extension of the Atlantic STJ is longer than 35° and has a negative tilt. Finally, DTs focused on the onset and end of events highlight geopotential height anomalies and STJ extension as critical variables. These DTs provide concise and accessible information for agricultural decision-makers in CSA.

Keywords: cold temperature extremes, regionalisation, machine learning, upper-level jets

1. Introduction:

Agricultural production is one of the main economic activities in central South America (CSA) (Ferrelli et al., 2021; Mintegui et al., 2019; Taboada et al., 2021). In particular, Argentina and Brazil are among the world's largest exporters of wheat, maize, soybeans and sugar (OECD-FAO, 2019). Brazil has positioned itself as the fourth country in the world with the highest gross value of agricultural production, reaching an average value of USD 94.5 billion between 2013 and 2022 (FAO, 2024). Argentina ranks eleventh, with an average value of USD 32.2 billion. Nevertheless, agricultural production extends to all CSA countries, as can be seen on the Foreign Agricultural Service website (https://ipad.fas.usda.gov/countrysummary/, accessed February 2024). During the austral winter (June-July-August, JJA), wheat and barley are the main crops sown, as they are resistant to low temperatures when the plants are dormant (Crops Randall Schnepf et al., 2001; Skinner and Bellinger, 2016). Winter wheat requires exposure to freezing or near-freezing conditions during its early growth phase to progress to the reproductive stage (Bowden et al., 2008). Without this exposure, winter wheat does not produce seeds.

Therefore, it is important to comprehend the physical processes that cause frost during the austral winter, which are crucial for crop development. Garreaud (2000) identifies three key factors that contribute to the incursion of cold air over subtropical South America. They include a surface cold core anticyclone that moves from the south-eastern Pacific into southern Argentina, a low-pressure centre deepening over the southwestern Atlantic Ocean, and the existence of an upper-level jet entrance over subtropical South America. The transverse, direct circulation induced by the former is instrumental in the occurrence of strong, long-lived cold surges, since this circulation provides additional forcing to the central part of the surface anticyclone. According to Vera and Vigliarolo (2000), the existence of an upper-level jet entrance plays a key role in the cold-surge occurrence over tropical South America. They claim that the

presence of an upper-level subtropical cyclonic perturbation is associated with the location of the subtropical jet (STJ) farther north over South America. Consequently, the subsidence region linked to the descending branch of the Hadley cell moves equatorward, facilitating the propagation of frontal systems into the tropics. A comprehensive examination of the meridional displacement of cold air incursions can be found in Lanfredi and de Camargo (2018), where the authors emphasise the importance of the zonal positioning of the upper-level trough and its interaction with the orography of the Andes and the Brazilian highlands.

These incursions of cold air can lead to widespread frost occurrence over CSA. The higher frequency of generalised frost in Argentina is associated with the presence of a more intense than normal STJ over South America (Müller, 2007; Müller et al., 2005). Moreover, the persistence of frost is also determined by the STJ features. In the case of the less persistent events, a STJ entrance is located farther east over the continent inducing a secondary meridional circulation whose subsiding branch provides additional forcing for the equatorward displacement of the low-level anticyclone (Müller and Berri, 2012, 2007). On the other hand, the most persistent generalised frosts are directly related to the quasi-stationary anticyclonic anomaly over the eastern Pacific Ocean, which deepens in the days preceding the event, and to the wind confluence in the STJ entrance region further west (Müller and Berri, 2012, 2007).

Recently, Collazo et al. (2024) produced a new algorithm to characterise the upper-level jets around South America. The algorithm provides a total of 38 parameters that allow for a detailed and quantitative daily description of the jets. The intensity, location, tilting, longitudinal extent, and number of branches are some of the evaluated parameters. This multi-parametric description covers aspects related to zonal asymmetries, which allows for the consideration of multi-variate configurations of the jet, providing a more complete view of its structure. Finally, these parameters enhance the comprehension of the regional impacts of the jet on climate, as

they permit the consideration of the individual and collective influences of the jet parameters on surface conditions (Barriopedro et al., 2022).

Even though the synoptic conditions driving CSA frosts are well understood, there is significant interannual variability in the region. This variability leads to strong cold air outbreaks in some years (Garreaud et al., 2009; Müller and Ambrizzi, 2010; Müller, 2007; Müller et al., 2005), while no events are recorded in other years (Fernandez-Long et al., 2016). One of the main sources of interannual climate variability in the CSA is the El Niño-Southern Oscillation (ENSO) (Cai et al., 2020). Although the ENSO is typically weaker during the austral winter compared to other seasons and is generally in a neutral phase (Reboita et al., 2021), Müller et al. (2000) find a more frequent occurrence of frost events in the Argentine wet pampas during the La Niña phase (LN), as well as an earlier first frost date than during the El Niño phase (EN).

The relevance of frost events to the agricultural sector, coupled with the better understanding of the physical processes that drive them, has prompted the scientific community to employ machine learning techniques in its description. However, these efforts have typically focused on improving medium-term forecasts rather than using them as a diagnostic tool to identify and quantify the physical characteristics associated with frost in the region. Moreover, they usually study small regions or used deep learning methods that are often difficult for decision makers to interpret or replicate. One of the first attempts to forecast frost in central Argentina using neural networks is documented in Ovando et al. (2005). Using meteorological data such as temperature, relative humidity, cloud cover and wind direction and speed, they predict the occurrence of frost days with an error of between 10 and 23%. Considering similar variables, a neural network is also constructed to predict minimum temperature for the south of Buenos Aires province in Argentina (Hernandez et al., 2020). More recently, Diniz et al. (2021) use spatial information and machine learning techniques to forecast frost risk in southern Brazil. The authors conclude that among the machine learning algorithms tested, random forest provides

the highest accuracy (above 90%) and the lowest class-specific error rates. Other tests, carried out for frost forecasting in southern Brazil, show that neural networks outperformed a statistical product previously developed by the National Space Research Institute (Rozante et al., 2023).

Therefore, the objective of this study is to conduct a diagnostic and climatic analysis to quantify the physical processes leading to frost occurrence in the CSA during the period 1979-2022. To achieve this, we first identify homogeneous regions of frost occurrence and their characteristics. In each region, the atmospheric circulation associated with frost and the role of ENSO are analysed. Particular attention is given to upper-level jets, which are described in detail using a set of recently defined parameters (Collazo et al., 2024). With all this information, decision trees are then used to quantify these relationships and derive simple rules to improve the understanding of the physical mechanisms linked to frost by agricultural users.

1. Data and methodology

An agrometeorological frost occurs when the minimum temperature measured in the meteorological shelter at 2m above the surface is less than 3°C (Fernandez-Long et al., 2016), which would be approximately equivalent to 0°C or below in the outdoor environment at the ground surface due to thermal inversion at night. This is particularly true under anticyclonic conditions as they promote light winds and clear skies, which facilitate radiative cooling of the surface at night.

We use daily minimum temperature from the gridded observational Climate Prediction Center (CPC) dataset (Xie et al., 2007), covering the period 1979-2022 and focusing on the austral winter months (June-July-August) in the CSA. The CPC offers data with a spatial resolution of 0.5° x 0.5°, accessible through their website (https://psl.noaa.gov/data/gridded/, accessed February 2024). This dataset has previously been used in the region for studies related to temperature extremes,

demonstrating good performance (Balmaceda-Huarte et al., 2021). Additionally, we considered gridded minimum temperature data from the Berkeley dataset with a resolution of 1° x 1° (Rohde et al., 2013), which yielded comparable results (not shown). In this work, we have chosen to use gridded observational databases in order to have a better spatial coverage, as the location of meteorological stations in the region is quite heterogeneous and sparse.

2.1 Regionalisation

Homogeneous regions of frost occurrence are identified by analysing binary time series at each grid point. A value of 0 is assigned when no frost occurred and a value of 1 when frost occurred. To calculate the distance between these time series, we use the Jaccard distance (Jaccard, 1908) instead of more commonly used metrics such as the Euclidean distance due to the binary nature of the problem. The Jaccard distance is calculated as one minus the Jaccard similarity. Jaccard similarity considers the simultaneous occurrence of an event in two binary time series as an indication of similarity, while ignoring the similarity when the event is absent (Vandeginste et al., 1998). The index ranges from 0 to 1. A value closer to 1 indicates greater similarity between the two sets of data. The Jaccard similarity can be calculated by dividing the size of the intersection by the size of their union (Baharav et al., 2020).

The distance matrix between all the grid points is then used to cluster them using the K-means algorithm (MacQueen, 1967). The optimal number of clusters can be determined by estimating the cluster stabilities. Stability measures the strength and reproducibility of a cluster and its membership. The stability of each cluster is determined by calculating the mean Jaccard similarity coefficient over multiple bootstrap iterations. Clusters with stability values above 0.85 are considered highly stable. To determine the optimal number of clusters, all clusters must exceed this threshold. We test the stability by considering K clusters between 2 and 10.

Once the homogeneous regions of frost occurrence are obtained, we proceed to characterise the events in each of them using different statistical metrics. The winter frost frequency is

159 estimated by dividing the number of frost days per winter by the total number of winter days
160 (92 days). This frequency is calculated for each grid point, and the regional value is obtained by
161 spatial averaging. From these time series, we estimate the standard deviation of the winter frost
162 frequency, the maximum and minimum winter frost frequency, and the years when they
163 occurred. On the other hand, considering the daily binary time series of frost occurrence at each
164 grid point, we also calculated the percentage of frost events with daily spatial extent over 25%
165 of the grid points in each region. We identified this for subsequent analyses because we wanted
166 to filter out frost events that affected only a few grid points, which may respond to a very local
167 effect. By using this 25% threshold, the events under investigation have areas of approximately
168 250,000 km^2 or larger, which is suitable for examining the synoptics linked to these events. In
169 addition, this threshold is set to avoid excessively restricting the number of events, which can
170 be counterproductive to the application of decision trees (see Section 2.3 for details). In
171 addition, this spatial extent threshold has already been used in previous work in the region, in
172 which frosts affecting at least 25% of the weather stations in central-eastern Argentina are
173 analysed (Müller, 2007; Müller et al., 2000). Finally, it should be noted that the choice of slightly
174 different thresholds (20 and 30%) leads to similar results, with regions with fewer events being
175 slightly more sensitive to changes (not shown).
176 Moreover, we estimate the winter frost frequency trend by region and at grid point level over
177 the period 1979‑2022. The non-parametric Mann‑Kendall trend test (Kendall, 1975; Mann,
178 1945), and Sen's slope estimator are employed to estimate the magnitude of the trend (Sen,
179 1968). Additionally, trends are calculated for the winter frequency of frosts with a spatial extent
180 larger than 25% of the grid points in each region.
181
182 *2.2 Associated atmospheric circulation*
183 To analyse the atmospheric circulation, we generate composites of the mid- and upper-level
184 circulation associated with frost events that have a spatial extent over 25% of the grid points in

185 the region. The daily anomalies of geopotential height at 500hPa (with respect to 1991-2020)
186 and the zonal wind at upper levels of the atmosphere (from 400 to 100 hPa) from ERA5
187 reanalysis (Hersbach et al., 2020) are considered. The data can be accessed from the Copernicus
188 Climate Change Service website
189 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-
190 levels?tab=overview, accessed in January 2023). Geopotential height anomalies are tested for
191 statistical significance using a Student's test at a 95% confidence level.
192 To characterise the upper-level jets, we employ the multi-parametric approach recently
193 proposed by Collazo et al. (2024). Parameters, such as latitude and intensity, are derived from
194 zonal wind peaks exceeding 30 m/s. Moreover, the number of jet divisions, tilting, maximum
195 wind intensity and its location, the longitudinal extent, and the height of the maximum are
196 defined for each jet branch. Due to parameter interdependence, we reduce dimensionality by
197 grouping correlated parameters via the Partitioning Around Medoids algorithm (Kaufman and
198 Rousseeuw, 1990). Further details can be found in Collazo et al. (2024).
199 Finally, to determine the ENSO phase, the National Oceanic and Atmospheric Administration
200 (NOAA) Oceanic Nino Index (ONI) for the JJA season is used
201 (https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php,
202 accessed in January 2024). The ONI is calculated based on the sea surface temperature
203 anomalies in the Nino 3.4 region, which is in the equatorial Pacific Ocean between 120°W and
204 170°W longitude and 5°N and 5°S latitude. EN conditions are indicated by a value of +0.5°C or
205 greater of the ONI that persists for five consecutive quarters, while a value of -0.5°C or less
206 indicates LN phase.
207 Subsequently, we use boxplots to show winter frost frequency in each region based on the ENSO
208 phase. To test the boxplots, we use the non-parametric Kruskal-Wallis test (Kruskal and Wallis,
209 1952) and the Wilcoxon rank sum test (Wilcoxon, 1945). The Kruskal-Wallis test assesses
210 whether population medians are equal across ENSO phases, and the Wilcoxon rank test

211 determines which pairs of populations differ from each other. The Benjamini Hochberg method
212 is used to adjust p-values for multiple comparisons (Benjamini and Hochberg, 1995).
213
214 *1.3 Decision trees*
215 Decision Trees (DTs) are a non-parametric supervised machine learning technique used for
216 classification and regression (Breiman et al., 2017). The goal is to create a model that estimates
217 a target variable by learning simple decision rules from data features. The key factor in
218 developing these DTs is selecting the most appropriate variable for branching. This process is
219 facilitated by the use of impurity measures. One such measure is the Gini impurity index, which
220 is employed in this study to quantify the probability of misclassification when assigning an object
221 to a particular class (Mola and Siciliano, 1997).
222 DTs are popular because they are easy to construct and interpret (especially when they are
223 small), handle mixed data types and missing values, and perform feature selection. They capture
224 complex relationships and can outperform logistic regression when the problem is highly non-
225 linear (Hastie et al., 2009; James et al., 2013; Jiao et al., 2020). These advantages, especially their
226 interpretability, led us to choose them.
227 In our binary classification task of frost occurrence, DTs are preferred despite class imbalance.
228 We use weighting techniques to address the bias towards dominant classes by assigning higher
229 weights to the less frequent category (Pedregosa et al., 2011). The weights are defined as
230 inversely proportional to the frequency of occurrence of each category. Moreover, we split the
231 data into training and test samples. We allocate 80% of the data for training, utilizing random
232 years while ensuring all winter days of each year remain together. The remaining 20% form the
233 test set, with the last year included. To optimize DT depth and prevent overfitting, we employ
234 10-fold cross-validation during training (James et al., 2013).
235 We employ confusion matrices to assess the performance of a DT model on the test dataset.
236 From confusion matrices, we can derive various metrics such as accuracy, sensitivity, and

237 balanced accuracy (Kulkarni et al., 2020). Accuracy is the most commonly used metric, but it can
238 be misleading if used with imbalanced datasets. Sensitivity, also known as true positive rate
239 (TPR), measures the proportion of positive instances that are correctly detected (Géron, 2019),
240 while balanced accuracy measures the average accuracy obtained from both the minority and
241 majority classes. Finally, the area under the curve (AUC) of the receiver operating characteristic
242 (ROC) is a single scalar value that measures the overall performance of a binary classifier and
243 takes values between 0 to 1 (Hanley and McNeil, 1982; Mandrekar, 2010). An AUC of 0.5
244 represents the performance of a random classifier, while values between 0.7 and 0.8 are
245 considered acceptable (Hosmer and Lemeshow, 2000) .
246
247 2. Results
248
249 *3.1 Frost characteristics by homogeneous regions*
250 First, we analysed the frequency of winter frosts for each grid point in CSA (Figure 1a). As
251 expected, the frequency of frosts increases towards the southern part of the region, except for
252 the Andes Mountain range. Moreover, the coastal regions tend to have a lower frequency of
253 these events than continental regions at the same latitude. We then search for homogeneous
254 regions of frost occurrence based on the binary series of the occurrence of these events at each
255 grid point. After analysing the stability of the clusters, we find that the optimal number of
256 regions is five, as all clusters are highly stable (Supplementary Figure 1). Figure 1b shows the
257 regions obtained, which are in agreement with the previous knowledge and with the frequency
258 of frost events in Figure 1a. Each of the regions is assigned a name as shown in Table 1. The
259 region designated as N encompasses areas in Brazil and the Atacama Desert in northern Chile,
260 since these are areas where frost occurrence is highly unlikely (Figure 1a). Also, as mentioned,
261 proximity to the coast and topography influence the frequency of frost. On the other hand, the
262 central region of Chile belongs to the same group as the Pampa, as they are usually affected by

the same pressure systems. During frosts, high-pressure systems moving over the Pacific Ocean and affecting central Chile cross the Andes and extend a ridge to their leeward side, affecting the Pampas in Argentina (Pezza and Ambrizzi, 2005). This mechanism has also been observed in recent cold intrusion events, such as the one that occurred in the winter of 2021 (Marengo et al., 2023).

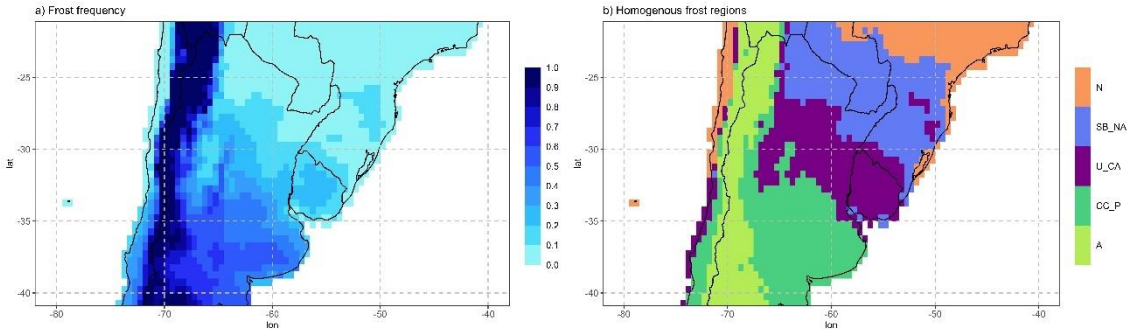


Figure 1: a) Winter relative frequency of frost events in CSA during the period 1979-2022. b) Regionalisation of frost occurrence into five homogeneous clusters.

Table 1: Homogeneous regions of frost events

Acronym	**Region**
N	Northern region
SB_NA	Southern Brazil, northern Argentina, and Paraguay
U_CA	Uruguay and central Argentina
CC_P	Central Chile and humid Pampa
A	Andean region

Once the homogeneous regions are found, we statistically characterise the frost events in each region. The spatial extent of frosts in each region is first analysed based on the proportion of grid points in each region that experienced frosts simultaneously (Figure 2). In the northern region, where frosts are infrequent, they usually affect less than 10% of the grid points. Frost events in the SB_NA and U_CA regions are usually limited to local areas, although there have been instances where they have covered a significant portion of the region. In CC_P, almost all

spatial extensions are equally probable. In contrast, frost events in the Andes are more likely to cover a large geographical area. To summarise, the N, SB_NA, and U_CA regions experience more localised frost events, whereas CC_P and A exhibit a more regional behaviour.

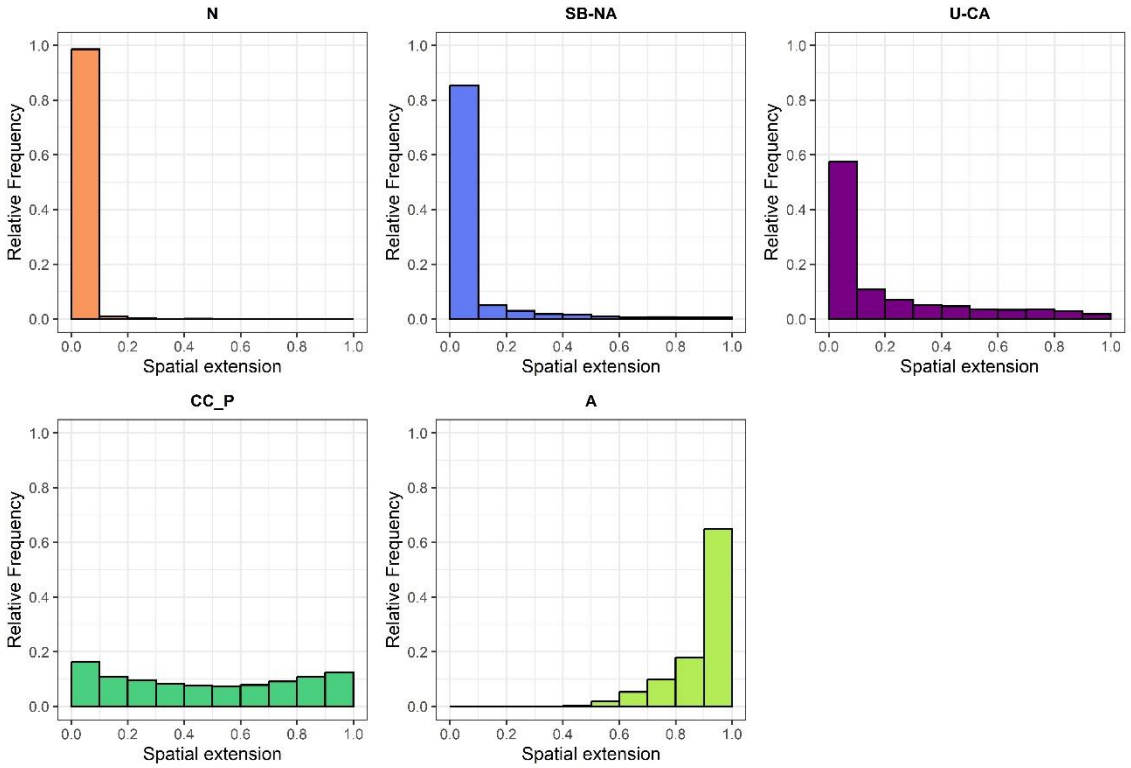


Figure 2: Histograms of the spatial extent of frost days expressed as a proportion of grid points in the region.

Table 2 presents metrics characterising frosts in each region. The mean percentage of winter days with frost, estimated for each grid point and then averaged over each region and time, shows that in N there is less than one frost per winter, while in CC_P almost 50% of winter days have frost events and 90% in the Andes. Additionally, the CC_P region exhibits the highest interannual variability.

In 2007, the maximum frequency of frost was recorded in three regions during an exceptionally cold winter (Collazo et al., 2019). This corresponds to the winter when it snowed in the city of Buenos Aires, the only snowfall in approximately the last 100 years (Pezza et al., 2010). The event was linked to an unusual circulation pattern, which included a dipole around the Antarctic Peninsula with a strong anticyclone to the west and a strong cyclone to the east. This configuration drove the cold advection toward mid-latitudes. Additionally, the Southern Annular Mode was one of the lowest on record during winter 2007, which intensified polar anticyclones (Pezza et al., 2010). Furthermore, we notice that the highest frequency of frost occurred during LN winters in all regions. This is noteworthy because the ENSO is typically weak during the austral winter and is usually in its neutral phase (Reboita et al., 2021). In particular, in our study period (44 winters) there are a total of 8 EN and 9 LN events. In terms of the years with the lowest frost frequency, it is observed that in three regions it occurred in 1982, which was an EN winter. In the other two regions, it also occurred in the 1980s, with one being a neutral winter and the other being during a LN event.

Table 2: Characteristics of daily winter frost events in each region based on the period 1979-2022.

Region	Mean winter frost freq. [%]	Standard deviation of winter frost freq. [%]	Max. winter freq. [%]	Year of max. freq.	Min. winter freq. [%]	Year of min. freq.	Percentage of frosts with extension >25% of grid points in the region [%] (# events)
N	0.81	0.55	2.45	2000	0.05	1982	0.69 (10)
SB_NA	5.85	2.94	12.95	1988	0.82	1982	18.65 (316)
U_CA	19,41	6.86	35.16	2007	4.76	1986	31.18 (1122)
CC_P	48.03	9.82	75.34	2007	26.30	1985	68.71 (2769)
A	90.22	3.06	96.14	2007	82.90	1982	100 (4045)

Considering that the highest frequency of frosts occurs in LN winters across all regions, we undertook a closer examination of the relationship between ENSO phases and frost frequency. Figure 3 shows the box plots obtained to illustrate this relationship. Overall, a higher frequency of frost is observed during LN winters and a lower frequency during EN winters in all regions, which is consistent with the findings of Müller et al. (2000) for central-eastern Argentina. However, the Kruskal-Wallis test indicates a significant difference between the phases only in the N and A regions. Therefore, no significant ENSO signal is found in the regions with the highest interannual variability, CC_P and U_CA (see Table 2), nor in SB_NA. Conducting more detailed studies is necessary to determine whether the low number of ENSO events or the modulation of other modes of climate variability is the cause.

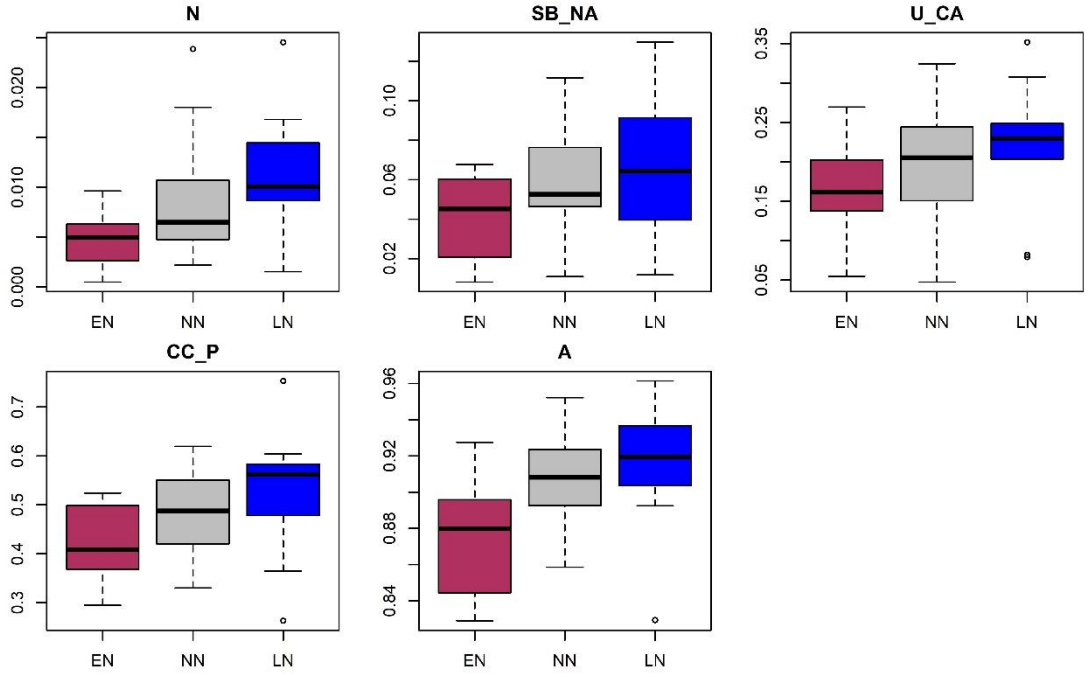


Figure 3: Boxplots for the frost frequency as a function of the ENSO phase in each region. The boxes represent the interquartile range and the line inside the box represents the median. The whiskers extend ±1.5 times the interquartile range. Dots represent outliers.

Next, we look for trends in the series of frost frequencies per winter over the period 1979-2022. No significant trends in frost frequency by region are found. At the grid point level, only a few positive trends are detected, indicating an increased frequency of winter frost over time, mainly in the southern CC_P region (Supplementary Figure 2). Using data from meteorological stations and over a longer period (1960-2018), Ferrelli et al. (2021) found a predominance of negative trends in the annual number of days with meteorological frost in the humid Pampa, although positive trends were observed at some stations located in the south of the region. The differences between the results observed here and this previous study may be due to several factors: the definition of frost (agrometeorological vs. meteorological), the season considered (austral winter vs. annual accumulation), the study period, the database (local vs. gridded).

Finally, we calculate the percentage of frost events that had a spatial extent greater than 25% of the grid points in the region (Table 2). In N, there are only ten events that meet this criterion. Conversely, in the Andes, all the winter days in the period satisfy this requirement, so an analysis of the associated circulation is not carried out for this region. Furthermore, we estimate trends in the frequency of these extensive winter frosts (Supplementary Figure 3). We find a significant long-term change in the number of frosts with these characteristics in the CC_P region, with an increase of 3.62% per decade. This might be due to changes in temperature, atmospheric circulation, or both. Supplementary Figure 4 shows the warming of minimum temperatures in the N region (which does not lead to a significant frost trend due to the low number of events recorded in this region) and significant negative trends in mean winter minimum temperatures in several areas belonging to CC_P. The latter could potentially contribute to an increase in the frequency of widespread frosts in the region. On the other hand, possible changes in atmospheric circulation patterns leading to frosts in CC_P are discussed in the next section.

3.2 Atmospheric circulation associated with frost events

Previous works have well-defined the conceptual model that leads to the incursion of cold air in CSA (e.g., Garreaud 2000; Marengo et al., 1997). However, we are interested in differences in the atmospheric circulation associated with frost among the different regions. Figure 4 shows composites of the atmospheric circulation associated with frost events with a spatial extent of more than 25% of the grid points in each region. The patterns of geopotential height anomalies at 500 hPa (Z500a) obtained are quite similar in all regions (Figure 4a) and consistent with Garreaud (2000). They show a cyclonic system over the Atlantic Ocean near the Brazilian coast and anticyclonic anomalies southwest of this centre over central and southern Argentina and Chile. Consequently, the occurrence of frost in different regions is due to a common physical process: the arrival of a cold front followed by the establishment of an anticyclone over the area. The disparities among these regions are mainly caused by small variations in the position and intensity of these systems. Frosts in the southernmost region require on average a cyclone located further southwest and of lower intensity than in the other regions. We analyse and quantify these differences in more detail in the next section.

With respect to long-term changes in the mid-level atmospheric circulation associated with extensive freezing in CC_P, we find significant increases in Z500a in subtropical regions of South America and polar latitudes over the ocean (Supplementary Figure 5a). However, no significant changes are observed over CC_P. Similarly, the frequency of cyclonic anomalies (in the green box in Fig. 4a) associated with frost events in the region does not show significant trends either (Supplementary Figure 6). The migratory post-frontal anticyclone, a key element, also does not present significant changes. Therefore, we conclude that the observed changes in the mid-level circulation do not make a relevant contribution to the increase in extensive frosts in CC_P.

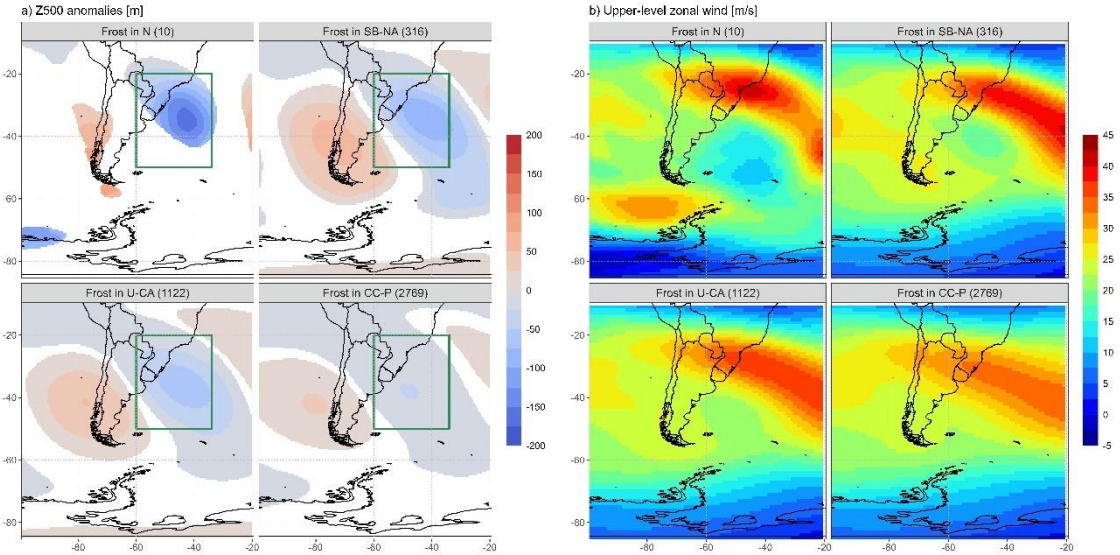


Figure 4: Composites of a) geopotential height anomalies at 500hPa (Z500a) [m], and b) upper-level zonal wind vertically averaged between 400 and 100hPa [m/s] associated with frost occurrence with a spatial extent greater than 25% of the grid points in each region. For Z500a, only significant anomalies are shown according to a 95% confidence t-test. The green rectangle indicates the region where the geopotential height will be averaged in the next analyses. Panel titles indicate the number of frost events considered in brackets.

Associated with the mid-level cyclonic system, a STJ is present in the upper-level atmosphere (Figure 4b). This jet has its entry region over northern Argentina, Paraguay, and southern Brazil, and then extends over the Atlantic Ocean with a negative tilting. As noted earlier, the existence of an inflow area in the jet stream is a crucial factor in the occurrence of severe and prolonged frost. However, the position of the STJ entry varies regionally, as does the degree of tilt. In N, the tilt over the continent does not appear to be particularly pronounced. Additionally, there are variations in the intensity and location of the jets over the Pacific.

To quantify these differences, we use the algorithm developed by Collazo et al. (2024). For the first time, we apply it to the austral winter and obtain a comprehensive, multi-parametric characterisation of the upper-level jets (Table 3). A total of thirteen main variables are detected,

six to describe the STJ and seven to describe the polar front jet (PFJ). The parameters identified include latitude, intensity, longitudinal extent, tilt, height of the jet maximum, and the number of branches. The statistical distributions of these parameters are shown in Supplementary Figure 7.

In contrast to summer, when there are many days when the STJ is absent (Collazo et al., 2024), in winter both jets tend to be present every day. Moreover, during the cold season, the jets show a greater variability in their location and intensity, corresponding to the higher baroclinicity during this season (Dos Santos et al., 2023). Therefore, more parameters are needed to characterise the jets in winter than in summer.

Finally, with respect to the long-term changes of the winter jet, we find only significant trends in the Pacific branch of the STJ (Supplementary Figure 5b). Consequently, the only parameter that shows significant increases is ext.lon.pac.stj. Nevertheless, the position of the Atlantic branch of the STJ, which is one of the most important factors affecting cold air incursions (as we discuss later), does not show any discernible trend.

Table 3: Subset of the jet parameters used to characterise the subtropical jet (STJ) and polar front jet (PFJ) during austral winter.

	Parameter	Definition	Unit
STJ	lat.stj	Latitude at which the STJ presents the maximum of the zonal mean of the zonal wind	°
	int.stj	Wind speed at lat.stj	m/s
	ext.lon.atl.stj	Extension in the longitudinal direction of the STJ Atlantic branch	°
	ext.lon.pac.stj	Extension in the longitudinal direction of the STJ Pacific branch	°
	tilt.atl.stj	Slope of the linear regression between latitudes and longitudes of the STJ Atlantic branch	°/°
	hei.pac.stj	Pressure level at which the maximum zonal wind is located for the STJ Pacific branch	hPa
PFJ	lat.pfj	Latitude at which the PFJ presents the maximum of the zonal mean of the zonal wind	°
	shar.pfj	Difference between the intensity of the PFJ and the meridional mean of the zonal wind	m/s
	ext.lon.pac.pfj	Extension in the longitudinal direction of the PFJ Pacific branch	°

	tilt.pac.pfj	Slope of the linear regression between latitudes and longitudes of the PFJ Pacific branch	°/°
	tilt.atl.pfj	Slope of the linear regression between latitudes and longitudes of the PFJ Atlantic branch	°/°
	hei.pac.pfj	Pressure level at which the maximum zonal wind is located for the PFJ Pacific branch	hPa
	branches.pfj	Number of PFJ divisions	

411

412

413 *3.3 Quantification of the physical processes leading to frosts*

414 In order to quantify the role of the physical processes linked to the occurrence of frosts in each
415 region, we apply DTs. To describe these processes, we consider several variables, including the
416 day of the season, the mid-level geopotential height anomalies in the region indicated in Figure
417 4, and the 13 jet parameters. DTs may not use all of these variables, but rather select the most
418 relevant ones to describe the problem.

419 Figure 5 displays the DT used to diagnose the occurrence of frost in the CC_P region. The DT
420 selects the first variable to be used (also known as the root node or splitting feature) in order to
421 maximise the separation of the data at that node. For the CC_P region, the variable that provides
422 the purest initial split is the averaged geopotential height anomalies at 500hPa over the region
423 shown in Figure 4 (z500a). If z500a is greater than 12m, we must go to the left branch of the
424 tree, and if it is less, to the right. The nodes show the assigned class, the estimated probability
425 of frost occurrence and the percentage of observations at the node. After this first division of
426 the data according to z500a values, we obtain two nodes indicating that there is a 36%
427 probability of frost with anticyclonic anomalies higher than 12m, whereas when z500a is cyclonic
428 or slightly positive, the probability of frost rises to 62%. This is consistent with what is observed
429 in the z500a composites in Figure 4. Continuing with the DT branches, it is possible to further
430 refine the mechanisms that contribute to the occurrence of frost. In the left sector of the tree
431 (under anticyclonic anomalies), frosts occur mainly with an STJ located north of 30°S and certain
432 combinations of the longitudinal extension of the Atlantic and Pacific branches of the STJ. The
433 fact that the STJ is located at low latitudes combined with a positive z500a would indicate that

the cold front has already passed through the CC_P region and that the post-frontal migrating anticyclone is established.

The right branch of the DT indicates that cyclonic or slightly positive anomalies combined with an ext.lon.atl.stj above 35° generate favourable conditions for the occurrence of frost in the region, with an estimated probability of 74% (Figure 5). Moreover, this combination of variables is the one most frequently associated with a frost event, as can be seen by comparing the estimated frost probabilities and the percentage of observations at each of the terminal nodes. Again, this is consistent with Figure 4, where slight cyclonic anomalies and an extended STJ over the Atlantic and the continent are observed to be associated with frost in the CC_P region. It is important to note that ext.lon.atl.stj determines the location of the STJ entrance, which is a key element for cold air incursions in CSA. If the ext.lon.atl.stj is shorter than 35°, there is still a 66% chance of frost if all the following conditions are met: it is not the last week of winter, the STJ intensity is below 44 m/s, and the tilt of the Atlantic branch of the STJ is negative (less than -0.11°/°). A negative tilt implies a NW‑SE orientation of the jet branches and is compatible with the expected inclination at the front of the trough (Collazo et al., 2024). Furthermore, Müller (2007) points out that a more intense STJ than normal results in an increased frost frequency due to the associated strong thermal gradient. However, we find an upper limit for this relationship set at 44 m/s, which is located at the right tail of the empirical distribution of int.stj (Supplementary Figure 7). A more intense STJ is associated with a poleward displacement of the STJ (Manney and Hegglin, 2018), so that cold air may be confined to higher latitudes and therefore no extensive frosts would be recorded in CC_P. In summary, the DT allows us to quantify the previous conceptual model defined by Garreaud (2000), to identify the most important variables for the description of frost, and to establish thresholds on these variables that contribute to increase the probability of these events.

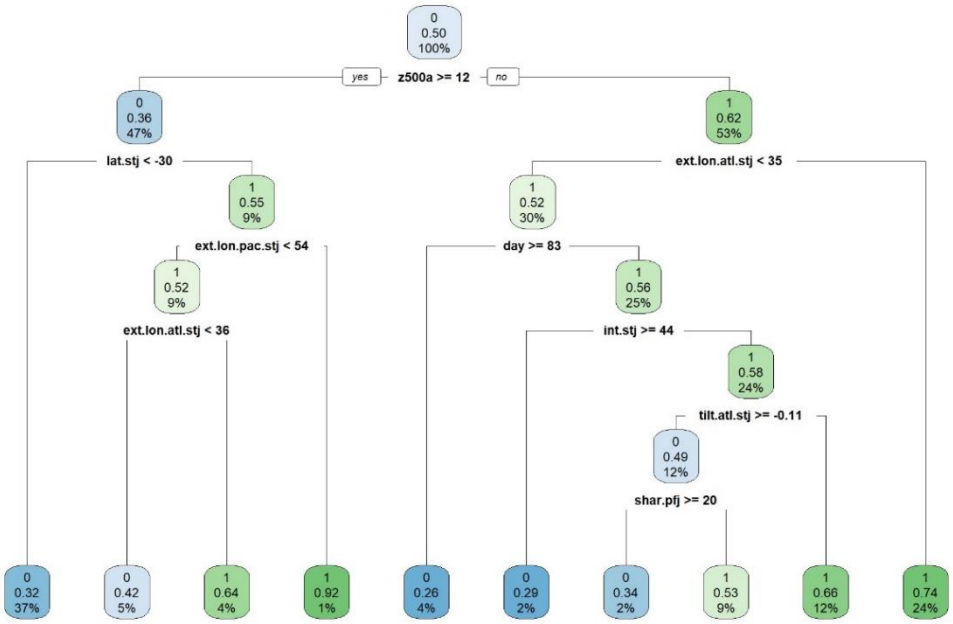


Figure 5: Decision tree to diagnose the physical processes involved in the frost occurrences in the CC_P region. The nodes show the assigned class, the estimated probability of frost occurrence, and the percentage of observations at the node, respectively.

The DT for the U_CA region is shown in Figure 6. This tree is less complex than that obtained for CC_P. The DT for U_CA uses five variables to perform the classification while CC_P uses eight. In addition, the DT of U_CA contains one fewer tree level. For both regions, the DTs starts by separating the training samples according to z500a. Furthermore, when the cyclonic anomalies are not so intense, it is again observed that the latitude of the STJ is a determining factor in the probability of frost occurrence, with thresholds similar to those previously found. Another similarity is observed when anticyclonic anomalies are recorded. In both regions, these anomalies must be accompanied by extensive ext.lon.atl.stj to favour frost, in the case of U_CA above 37°. Alternatively, when z500a falls within the range of -49m to 14m, the latitude of the PFJ becomes relevant in U_CA, leading to a distinction between U_CA and CC_P. If the PFJ is present south of 53°S or north of this latitude but not during the last three weeks of winter, it can cause widespread frosts in the U_CA region. The role of PFJ in promoting frost occurrence

has been studied by Müller and Berri (2012, 2007). They suggest that when the STJ is situated at approximately 30°S and the PFJ is positioned at around 60°S, this configuration plays a substantial role in conducting Rossby waves to the continent that favours widespread and persistent frosts. Therefore, this is consistent with our findings through the DT, although the latter provides more accurate thresholds and other variables that need to be considered, such as z500a.

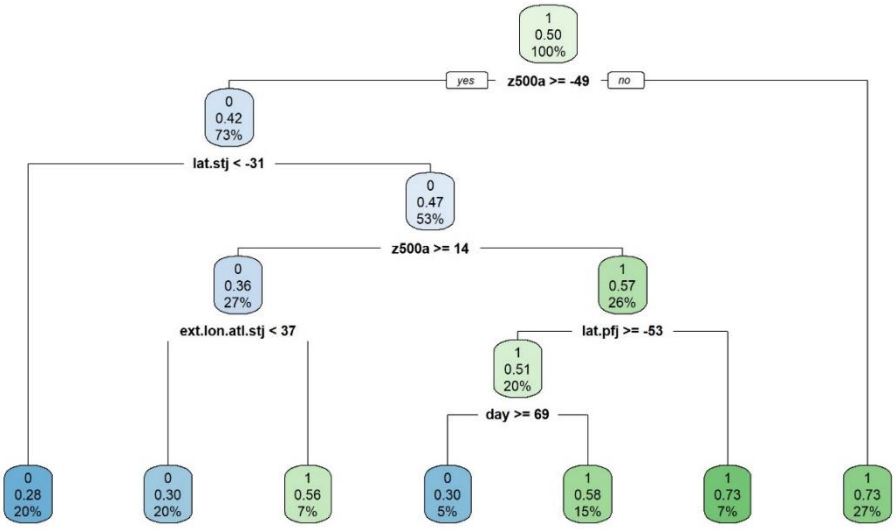


Figure 6: Same as Figure 5 for the U_CA region.

As in the other two regions, the DT for SB_NA starts by separating the training samples according to z500a (Figure 7). Another coincidence is that the occurrence of frost is found to be favoured by cyclonic or slightly positive anomalies. Together with these z500a, several requirements linked to the jets increase the possibility of recording an extended frost towards the beginning and the end of winter, even though these are different from each other. In the early winter, a short Pacific branch of the STJ and a not extremely intense PFJ are required. On the other hand,

towards the end of winter, frosts are more likely with the following two combinations: 1) the STJ located south of 27°S, or 2) the STJ located north of that latitude but with an extremely negative Atlantic branch tilting. Finally, in contrast to the previous regions, if anticyclonic anomalies higher than 14m are recorded, the DT does not find a jet configuration contributing to rise the frost probability.

Note that no DT is obtained for the N region because there were only ten extensive frost events, too few to train and test the tree.

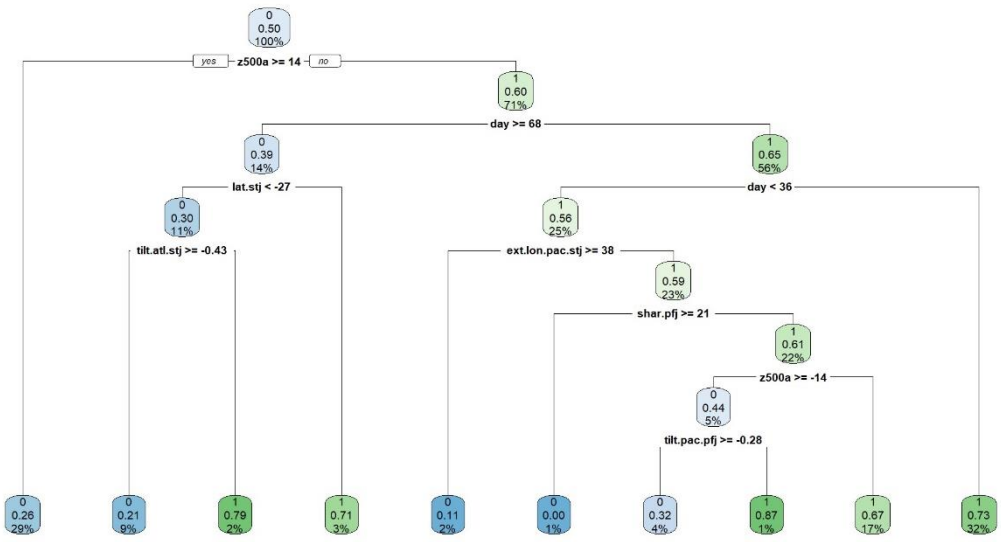
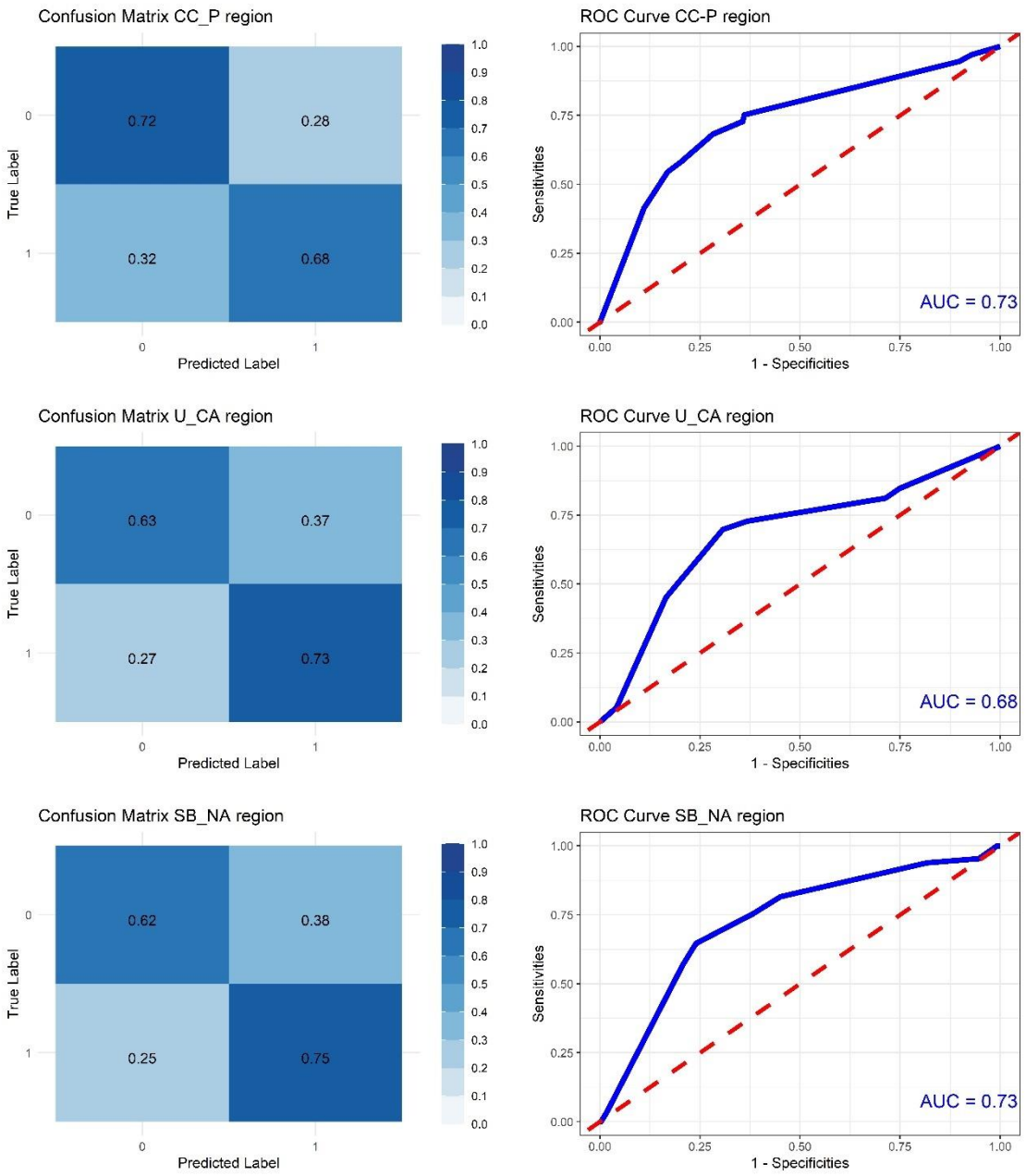


Figure 7: Same as Figure 5 for the SB_NA region.

To assess that the DTs are indeed capturing the physical processes linked to frost occurrence, the performance of the DTs on the test sample is estimated from confusion matrices and ROC curves (Figure 8). In all regions, the percentage of correctly classified observations is higher than the failures. Moreover, the AUC exceeds the threshold of 0.5, set as a lower limit to indicate the discrimination ability between the two classes, showing values between 0.68 and 0.73.

Therefore, the DTs have skill to classify the occurrence or non-occurrence of frost. In the CC_P region, the balanced accuracy is 0.70. Therefore, 70% of the overall observations are correctly classified, and the TPR indicates that the proportion of observed frosts that are correctly classified as frost by the model is 68%. The U_CA and SB_NA regions have similar skills, with a balanced accuracy of 68 and 69%, respectively. In addition, the rate of correctly classified frosts is 73 and 75% respectively.

Another important aspect of frost that has been addressed in previous work is the persistence of the events (Müller and Berri, 2012, 2007). Supplementary Figure 8 shows the cumulative distribution function of frost persistence with a spatial extent of more than 25% of the grid points in each region. As expected, persistence increases as we move southwards. In particular, for the CC_P region, there are frosts that persisted for more than 20 consecutive days. Given this persistence in the time series, we evaluate whether the DTs can be improved by also incorporating the occurrence of frost on the previous day as a variable (Supplementary Figures 9-12). In all regions we observe that once the DTs divide the samples according to whether they recorded frost the previous day, the physical mechanisms that lead to an increased probability of frost are similar. Regarding the skill of the DTs, a slight improvement of the balanced accuracy is found when persistence is taking it into account, although it is not substantial (3.67% higher on average). In fact, the TPR of the DTs for the U_CA and SB_NA regions is higher without including persistence (Figure 8 and Supplementary Figure 12). Therefore, we conclude that persistence does not provide additional relevant information about the mechanisms that lead to frost in the DTs. Furthermore, previous work has already established that frost persistence can be explained by the position of the entrance region of the jet (Müller and Berri, 2012, 2007). An STJ located further east on the continent, associated with shorter ext.lon.atl.stj, results in less frost persistence. Consequently, including information on both persistence and STJ position becomes redundant in our model.



534

535 Figure 8: Confusion matrix, expressed as frequencies, and ROC curves for each region used to evaluate
536 the performance of DTs showed in Figures 5-7. In the latter, the area under the ROC curve (AUC) is also
537 indicated.

538

539 Finally, an exploratory analysis of a more difficult problem is carried out: obtaining DTs that try
540 to explain the mechanisms that lead to the onset and end of a frost (Supplementary Figure 13).
541 The difficulty lies in the fact that this is a more specific question and therefore the number of

samples to train and test is reduced. Moreover, it changes the imbalance between the categories of our binary problem. To address the onset of frost, the difference between a day without frost (d_i) and the following day (d_{i+1}) is considered, so that if this difference is 0 it implies that no frost has started, while if it is 1 it indicates the frost beginning. In addition, the variables incorporated into the DTs are considered in the d_i. Following this idea, to analyse the end of a frost, the difference between a day with frost and the following day is considered. In this problem, 0 indicates when the frost persists and -1 indicates the end of the frost. Regarding the complexity of the trees, a compromise had to be made between being too complex and difficult to interpret, and also prone to overfitting, or simpler and therefore easier to interpret, but with less skill (Supplementary Figure 14). We have chosen to prioritise the latter option, as our main objective is interpretability.

Although the DT skill for frost onset in CC_P is low (balanced accuracy = 0.56), the importance of ext.lon.atl.stj for the occurrence of these events is again highlighted, as in previous analyses (Supplementary Figure 13). Likewise, the end of frosts in CC_P is more likely with ext.lon.atl.stj below 49° combined with different tilting thresholds for the different branches of the STJ and PFJ. A similar conclusion can be reached by analysing the right DT branch of Supplementary Figure 9.

As seen in the left DT branch of Supplementary Figure 10, frost onset in U_CA is more likely under cyclonic conditions and with an extensive Atlantic branch of the STJ. However, Supplementary Figure 13 provides a little more detail on the mechanisms that favour frost initiation. For example, it is observed that towards the end of winter a frost can start if the cyclonic anomalies are combined with a short Pacific branch of the STJ. Regarding the end of a frost event in U_CA, the requirements depend on the day of the season. In the first two months of winter, anticyclonic anomalies and not extremely negative tilt.atl.stj lead to a 79% chance of frost termination, while if the anomalies are cyclonic (between -84 and 2 m) and the

ext.lon.atl.stj is very short, the chance of frost termination is 65%. In the last month of winter, unless the condition that z500a is extremely cyclonic is met, the frost is probably over.

For the onset of a frost in SB_NA a very good skill is obtained only considering the z500a (Supplementary Figure 13 and 14). Cyclonic anomalies result in a higher probability of an event initiation. The DT for frost end in SB_NA also considers only one variable, the ext.lon.atl.stj. Short extensions increase the probability of frost end, however in this case the skill is considerably lower than for frost onset.

3. Conclusions

This study characterises frost events in CSA during the austral winter from 1979 to 2022. While previous literature has provided insight into the physical processes behind frost events in the region (Garreaud, 1999; Lanfredi and de Camargo, 2018; Müller and Ambrizzi, 2010; Müller, 2007; Müller et al., 2005; Vera and Vigliarolo, 2000), this work introduces three significant advancements to the field. These include an exploration of regional frost characteristics, a novel multi-parametric approach for describing upper-level jets, and the quantification of underlying mechanisms using decision trees. The following presents the main findings of this study.

- *Regionalisation:* Previous analyses of frost events in CSA have often delineated regions based on productive agricultural zones (e.g., Hernandez et al., 2020; Mintegui et al., 2019; Müller and Berri, 2007). However, it is important to note that these regions may not exhibit a uniform frost behaviour. Our research has identified five homogeneous regions of frost occurrence in CSA: north (N); southern Brazil, northern Argentina, and Paraguay (SB_NA); Uruguay and central Argentina (U_CA); central Chile and humid Pampa (CC_P); and Andean region (A). Frost events are rare in the N region, with less than one event per winter on average. In contrast, A experiences freezing conditions on

around 90% of winter days, while CC_P has such conditions on almost 50% of the days. Despite these differences, A and N regions experience a significant increase in the frequency of frost events during La Niña winters. This pattern is also observed in other regions, although it is not statistically significant.

- *Trends:* Our analysis reveal a significant long-term increase in extensive frost events within the CC_P region, averaging 3.62% more events per decade. To understand the drivers of this trend, we investigate the relationship with changes in minimum temperatures and atmospheric circulation. The results suggest that negative trends in winter minimum temperatures across various areas of the CC_P region are the primary mechanism behind these long-term changes.
- *Winter upper-level jet parameters:* Numerous previous studies have demonstrated the importance of upper-level jets in determining the occurrence of frost in the CSA (Garreaud, 2000; Müller, 2007; Müller et al., 2005; Müller and Berri, 2012). Therefore, in this work we present a comprehensive and quantitative characterisation of the jets, which is essential for a more detailed understanding of their role in cold extremes and for their incorporation as features in machine learning models. Building on the methodology of Collazo et al. (2024) designed for summer jets, we adapt it to characterise upper-level jets over South America during winter. Our findings indicate that six parameters are required to describe the STJ, and seven to represent the PFJ, including latitude, intensity, tilting, longitudinal extent and height. Furthermore, we observe that the number of parameters needed is higher than those observed during summer. The greater complexity of the jets during winter is attributed to the increased variability observed in this season in both the position and intensity of the jets.
- *Quantification of underlying mechanisms using decision trees:* The high interpretability of the decision trees has enabled the identification of a subset of variables and their thresholds that are associated with a higher probability of frost occurrence for each

region in CSA. We find that cyclonic or slightly positive z500a (generally below 14m) in the Atlantic Ocean tend to favour frosts. Among the jet parameters the most outstanding are: latitude of the STJ, longitudinal extent and tilting of the Atlantic branch of the STJ. The conditions that favour frost are an STJ north of 31°S, an extension above 35°, and negative tilts (which are more pronounced in the northernmost regions). The prolonged duration of these conditions over several days contributes to the persistence of frost. However, incorporating a variable specifically indicating this persistence into decision trees results in only a marginal enhancement of skill. Finally, we observe that that frost events are more likely to occur in July.

- *Variables relevant for the onset/offset of frosts:* An exploratory analysis is conducted, focusing exclusively on the beginning and end of the events. It is, however, important to note that the skill of the DTs is generally lower in this analysis due to the smaller sample size and the greater imbalance between categories. Again, the most relevant variables are z500a and ext.lon.atl.stj. In particular, we find an interesting result in the SB_NA region. Using only the z500a in the DT, we achieved a balanced accuracy of 0.76 in diagnosing the onset of frost. This high accuracy suggests the potential for this method to be a valuable tool across various applications.

The results of this work are intended to bring agricultural users closer to understanding the physical processes that lead to the occurrence of winter frosts in CSA. This knowledge will help to improve decision-making in this sector of the economy, which is so crucial for the countries of the region. Furthermore, this study holds particular interest for the forecasting community. It has the potential to enable a more rigorous evaluation of how well numerical weather prediction (NWP) models capture the physical processes that lead to frost in CSA. Specifically in Argentina, the National Meteorological Service has identified considerable errors in the two NWPs used for minimum temperature forecasts - GFS and WRF. According to the verification carried out for the period 2020-2022, these models show the largest biases during the austral

winter, with more than 40% of winter days with errors above 2°C for forecast day+1 and day+2 (Matsudo and García Skabar, 2023). A comparable behaviour is observed in southern Brazil, where the ETA model underestimates the temperature by more than 20% for the three lead times studied (24, 48 and 72 h) (Rozante et al., 2023). Therefore, there is an urgent need for a future study that could focus on a detailed analysis of the specific meteorological conditions and model configurations that contribute to these forecast errors.

Acknowledgements

This work was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 847635 (UNA4CAREER) through the SAFETE project (code 4230420).

References

Baharav, T.Z., Kamath, G.M., Tse, D.N., Shomorony, I., 2020. Spectral Jaccard Similarity: A New Approach to Estimating Pairwise Sequence Alignments. Patterns 1, 100081. https://doi.org/10.1016/j.patter.2020.100081

Balmaceda-Huarte, R., Olmo, M.E., Bettolli, M.L., Poggi, M.M., 2021. Evaluation of multiple reanalyses in reproducing the spatio-temporal variability of temperature and precipitation indices over southern South America. International Journal of Climatology 41, 5572–5595. https://doi.org/10.1002/joc.7142

Barriopedro, D., Ayarzagüena, B., García-Burgos, M., García-Herrera, R., 2022. A multi-parametric perspective of the North Atlantic eddy-driven jet. Clim Dyn. https://doi.org/10.1007/s00382-022-06574-w

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300.

Bowden, P., Edwards, J., Ferguson, N., Mc Nee, T., Manning, B., Roberts, K., Schipp, A., Schulze, K., Wilkins, J., 2008. Wheat growth and development. Procrop, NSW Department of Primary Industries District Agronomists.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification And Regression Trees. Routledge. https://doi.org/10.1201/9781315139470

Cai, W., McPhaden, M.J., Grimm, A.M., Rodrigues, R.R., Taschetto, A.S., Garreaud, R.D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., Ng, B., Anderson, W., Wang, G., Geng, T.,

Jo, H.-S., Marengo, J.A., Alves, L.M., Osman, M., Li, S., Wu, L., Karamperidou, C., Takahashi, K., Vera, C., 2020. Climate impacts of the El Niño–Southern Oscillation on South America. Nat Rev Earth Environ 1, 215–231. https://doi.org/10.1038/s43017-020-0040-3

Collazo, S., Barrucand, M., Rusticucci, M., 2019. Variability and predictability of winter cold nights in Argentina. Weather Clim Extrem 26, 100236. https://doi.org/10.1016/j.wace.2019.100236

Collazo, S., García-Herrera, R., Barriopedro, D., 2024. Summer upper-level jets modulate the response of South American climate to ENSO. Clim Dyn 62, 1031–1054. https://doi.org/10.1007/s00382-023-06955-9

Crops Randall Schnepf, Field D, Dohlman, E., Bolling Market, C., 2001. Agriculture in Brazil and Argentina: Developments and Prospects for Major.

Diniz, É.S., Lorenzon, A.S., de Castro, N.L.M., Marcatti, G.E., dos Santos, O.P., de Deus Júnior, J.C., Cavalcante, R.B.L., Fernandes-Filho, E.I., Hummeldo Amaral, C., 2021. Forecasting frost risk in forest plantations by the combination of spatial data and machine learning algorithms. Agric For Meteorol 306. https://doi.org/10.1016/j.agrformet.2021.108450

Dos Santos, J.D., Machado, J.P., Saraiva, J.M.B., 2023. The Response of Southwest Atlantic Storm Tracks to Climate Change in the Brazilian Earth System Model. Atmosphere (Basel) 14, 1055. https://doi.org/10.3390/atmos14071055

FAO, 2024. FAOSTAT Value of Agricultural Production Database [WWW Document]. URL https://www.fao.org/faostat/en/#data/QV (accessed 4.9.24).

Fernandez-Long, M.E., Barnatan, I., Murphy, G., 2016. Generation and use of frost information in Argentina. Meteorologica 41, 7–31.

Ferrelli, F., Brendel, A.S., Perillo, G.M.E., Piccolo, M.C., 2021. Warming signals emerging from the analysis of daily changes in extreme temperature events over Pampas (Argentina). Environ Earth Sci 80. https://doi.org/10.1007/s12665-021-09721-4

Garreaud, R.D., 2000. Cold Air Incursions over Subtropical South America: Mean Structure and Dynamics. Mon Weather Rev 128, 2544–2559. https://doi.org/10.1175/1520-0493(2000)128<2544:CAIOSS>2.0.CO;2

Garreaud, R.D., 1999. Cold Air Incursions over Subtropical and Tropical South America: A Numerical Case Study. Mon Weather Rev 127, 2823–2853. https://doi.org/10.1175/1520-0493(1999)127<2823:CAIOSA>2.0.CO;2

Garreaud, R.D., Vuille, M., Compagnucci, R., Marengo, J., 2009. Present-day South American climate. Palaeogeogr Palaeoclimatol Palaeoecol 281, 180–195. https://doi.org/10.1016/j.palaeo.2007.10.032

Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn and TensorFlow, 2nd Edition. ed. O'Reilly.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7

Hernandez, G., Müller, G.V., Villacampa, Y., Navarro-Gonzalez, F.J., Aragonés, L., 2020. Predictive models of minimum temperatures for the south of Buenos Aires province. Science of The Total Environment 699, 134280. https://doi.org/10.1016/j.scitotenv.2019.134280

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146, 1999–2049. https://doi.org/10.1002/qj.3803

Hosmer, D., Lemeshow, S., 2000. Applied Logistic Regression, 2nd Edition. ed. John Wiley and Sons, New York.

Jaccard, P., 1908. Nouvelles Recherches Sur La Distribution Florale. Bulletin de la Société vaudoise des Sciences Naturelles 44, 223–270.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning : with applications in R. Springer, New York.

Jiao, S.R., Song, J., Liu, B., 2020. A Review of Decision Tree Classification Algorithms for Continuous Variables. J Phys Conf Ser 1651, 012083. https://doi.org/10.1088/1742-6596/1651/1/012083

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/9780470316801

Kendall, M.G., 1975. Rank Correlation Methods, 4th edition. ed. Charles Griffin, London.

Kruskal, W.H., Wallis, W.A., 1952. Use of Ranks in One-Criterion Variance Analysis. J Am Stat Assoc 47, 583–621. https://doi.org/10.1080/01621459.1952.10483441

Kulkarni, A., Chong, D., Batarseh, F.A., 2020. Foundations of data imbalance and solutions for a data democracy, in: Data Democracy. Elsevier, pp. 83–106. https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Lanfredi, I.S., de Camargo, R., 2018. Classification of Extreme Cold Incursions over South America. Weather Forecast 33, 1183–1203. https://doi.org/10.1175/WAF-D-17-0159.1

MacQueen, J.B., 1967. Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability 1, 281–297.

Mandrekar, J.N., 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology 5, 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Mann, H.B., 1945. Nonparametric Tests Against Trend. Econometrica 13, 245. https://doi.org/10.2307/1907187

Manney, G.L., Hegglin, M.I., 2018. Seasonal and Regional Variations of Long-Term Changes in Upper-Tropospheric Jets from Reanalyses. J Clim 31, 423–448. https://doi.org/10.1175/JCLI

Marengo, J., Cornejo, A., Satyamurty, P., Nobre, C., Sea, W., 1997. Cold Surges in Tropical and Extratropical South America: The Strong Event in June 1994. Mon Weather Rev 125, 2759–2786. https://doi.org/10.1175/1520-0493(1997)125<2759:CSITAE>2.0.CO;2

Marengo, J., Espinoza, J.C., Bettolli, L., Cunha, A.P., Molina-Carpio, J., Skansi, M., Correa, K., Ramos, A.M., Salinas, R., Sierra, J.P., 2023. A cold wave of winter 2021 in central South America: characteristics and impacts. Clim Dyn 61, 2599–2621. https://doi.org/10.1007/s00382-023-06701-1

Matsudo, C., García Skabar, Y., 2023. Verificación de los pronósticos del Sistema de Asimilación y Pronóstico numérico del SMN para el período 2020-2022. Nota Técnica SMN 2023-145.

Mintegui, J.M., Puhales, F.S., Boiaski, N.T., Nascimento, E. de L., Anabor, V., 2019. Some mean atmospheric characteristics for snowfall occurrences in southern Brazil. Meteorology and Atmospheric Physics 131, 389–412. https://doi.org/10.1007/s00703-018-0578-5

Mola, F., Siciliano, R., 1997. A fast splitting procedure for classification trees. Stat Comput 7, 209–216. https://doi.org/10.1023/A:1018590219790

Müller, G., Ambrizzi, T., 2010. Rossby wave propagation tracks in southern hemisphere mean basic flows associated to generalized frosts over southern South America. Atmósfera 23, 25–35.

Müller, G.V., 2007. Patterns leading to extreme events in Argentina: Partial and generalized frosts. International Journal of Climatology 27, 1373–1387. https://doi.org/10.1002/joc.1471

Müller, G. V., Ambrizzi, T., Núñez, M.N., 2005. Mean atmospheric circulation leading to generalized frosts in central southern South America. Theor Appl Climatol 82, 95–112. https://doi.org/10.1007/s00704-004-0107-y

Müller, G. V., Berri, G.J., 2012. Atmospheric circulation associated with extreme generalized frosts persistence in central-southern South America. Clim Dyn 38, 837–857. https://doi.org/10.1007/s00382-011-1113-2

Müller, G. V., Berri, G.J., 2007. Atmospheric circulation associated with persistent generalized frosts in Central-Southern South America. Mon Weather Rev 135, 1268–1289. https://doi.org/10.1175/MWR3344.1

Müller, G. V, Núñez, M.N., Seluchi, M.E., 2000. Relationship between ENSO cycles and frost events within the Pampa Húmeda region. International Journal of Climatology 20, 1619–1637.

OECD-FAO, 2019. OECD-FAO Agricultural Outlook 2019-2028 Chapter 2. Latin American Agriculture: Prospects and Challenges. OECD Publishing, Paris/Food and Agriculture Organization of the United Nations, Rome.

Ovando, G., Bocco, M., Sayago, S., 2005. Redes neuronales para modelar predicción de heladas. Agricultura Técnica 65. https://doi.org/10.4067/s0365-28072005000100007

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pezza, A.B., Ambrizzi, T., 2005. Cold Waves in South America and Freezing Temperatures in São Paulo: historical background (1888-2003) and case studies of cyclone and anticyclone tracks, Revista Brasileira de Meteorologia.

Pezza, A.B., Simmonds, I., Coelho, C.A.S., 2010. The unusual Buenos Aires snowfall of July 2007. Atmospheric Science Letters 11, 249–254. https://doi.org/10.1002/asl.283

Reboita, M.S., Ambrizzi, T., Crespo, N.M., Dutra, L.M.M., Ferreira, G.W. de S., Rehbein, A., Drumond, A., da Rocha, R.P., Souza, C.A. de, 2021. Impacts of teleconnection patterns on South America climate. Ann N Y Acad Sci 1504, 116–153. https://doi.org/10.1111/nyas.14592

Rohde, R., Muller, R.A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., Wickham, C., 2013. A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. Geoinformatics & Geostatistics: An Overview 01. https://doi.org/10.4172/2327-4581.1000101

Rozante, J.R., Ramirez, E., Ramirez, D., Rozante, G., 2023. Improved frost forecast using machine learning methods. Artificial Intelligence in Geosciences. https://doi.org/10.1016/j.aiig.2023.10.001

Sen, P.K., 1968. Estimates of the Regression Coefficient Based on Kendall's Tau. J Am Stat Assoc 63, 1379. https://doi.org/10.2307/2285891

Skinner, D., Bellinger, B.S., 2016. Freezing tolerance of winter wheat as influenced by extended growth at low temperature and exposure to freeze-thaw cycles. Canadian Journal of Plant Science CJPS-2016-0154. https://doi.org/10.1139/CJPS-2016-0154

Taboada, M.Á., Costantini, A.O., Busto, M., Bonatti, M., Sieber, S., 2021. Climate change adaptation and the agricultural sector in South American countries: Risk, vulnerabilities and opportunities. Rev Bras Cienc Solo 45. https://doi.org/10.36783/18069657rbcs20210072

Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J., 1998. Cluster analysis. pp. 57–86. https://doi.org/10.1016/S0922-3487(98)80040-3

Vera, C.S., Vigliarolo, P.K., 2000. A Diagnostic Study of Cold-Air Outbreaks over South America. Mon Weather Rev 128, 3–24. https://doi.org/https://doi.org/10.1175/1520-0493(2000)128<0003:ADSOCA>2.0.CO;2

Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1, 80. https://doi.org/10.2307/3001968

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., Liu, C., 2007. A Gauge-Based Analysis of Daily Precipitation over East Asia. J Hydrometeorol 8, 607–626. https://doi.org/10.1175/JHM583.1

Statements and Declarations

Funding

This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 847635 (UNA4CAREER) through the SAFETE project (code 4230420).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions statement

Soledad Collazo: *Conceptualization, Formal Analysis, Visualization, Writing – Original Draft.*
Ricardo García-Herrera: *Conceptualization, Supervision, Writing - Review & Editing*

Data availability

All data used in this study is publicly accessible online