INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2016)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4693

Regionalization of rainfall over the Peruvian Pacific slope and coast

Pedro Rau,[a]* Luc Bourrel,[a] David Labat,[a] Pablo Melo,[a] Boris Dewitte,[b] Frédéric Frappart,[a,b] Waldo Lavado[c] and Oscar Felipe[c]

[a] *Géosciences Environnement Toulouse, UMR 5563, Université de Toulouse, CNRS-IRD-OMP-CNES, France*
[b] *LEGOS, UMR 5566, Université de Toulouse, CNRS-IRD-OMP-CNES, France*
[c] *Dirección General de Hidrología y Recursos Hídricos, SENAMHI, Lima, Peru*

ABSTRACT: Documenting the heterogeneity of rainfall regimes is a prerequisite for water resources management, mitigation of risks associated to extremes weather events and for impact studies. In this paper, we present a method for regionalization of rainfall over the Peruvian Pacific slope and coast, which is the main economic zone of the country and concentrates almost 50% of the population. Our approach is based on a two-step process based on k-means clustering followed by the regional vector method (RVM) applied to a network of 145 rainfall stations covering the period 1964–2011. The advantage of combining cluster analysis and RVM is demonstrated compared with just applying each of these methods. Nine homogeneous regions are identified that depict the salient features of the rainfall variability over the study area. A detailed characterization of the rainfall regime in each of the identified regions is presented in response to climate variability at seasonal and interannual timescale. They are shown to grasp the main modes of influence of the El Niño Southern Oscillation (ENSO), that is, increased rainfall over downstream regions in northern Peru during extreme El Niño events and decreased rainfall over upstream regions along the Pacific slope during central Pacific El Niño events. Overall our study points to the value of our two-step regionalization procedure for climate impact studies.

KEY WORDS rainfall; regionalization; k-means; regional vector; Peruvian Pacific slope; Peruvian coast; ENSO

Received 16 January 2015; Revised 20 January 2016; Accepted 21 January 2016

1. Introduction

Rainfall along the Pacific slope and coast of South America is characterized by a complex pattern of spatial and seasonal variability related to its meridional extension and the prominent topography of the Andes Cordillera (Waylen and Poveda, 2002; Garreaud *et al.*, 2009). The Peruvian Pacific slope and coast is located at tropical latitudes and rainfall is mainly influenced by orographic conditions, ocean and atmosphere. The region is characterized by a steep topography that inhibits cross-shore atmospheric flow and disrupt a geotropically balanced zonal wind, inducing a northward sea level pressure gradient along the coast that accelerate the wind northward (Muñoz and Garreaud, 2005). Such a low-level northward mean circulation is associated to cool sea surface temperature (SST) through inducing upwelling and evaporation, which makes this region persistently free of convective rainfall year-around (Takahashi and Battisti, 2007). The Pacific coast of Peru is thus mostly a 'dry zone' that only episodically experiences rainfall events. At interannual timescales, those rainfall events are associated to the El Niño Southern Oscillation (ENSO) phenomenon that is the main climatic influence over rainfall over the Peruvian Pacific coast (Lagos *et al.*, 2008). A rainy season can also be developed owed to a slight weakening of the southeast Pacific anticyclone and the southward displacement of the Pacific Inter-Tropical Convergence Zone (ITCZ) (Lavado *et al.*, 2012).

Although this region concentrates more than 50% of population of Peru, it remains poorly documented in terms of rainfall regionalization. Recent works (Suarez, 2007; Lavado *et al.*, 2012; Ochoa *et al.*, 2014; Bourrel *et al.*, 2015) mostly focused on principal stations or major watersheds. In 1999, a technical report (BCEOM, 1999) proposed a previous rainfall regionalization for the Peruvian Pacific slope and coast based on the Regional Vector Method (RVM) (Brunet-Moret, 1979), which consists in assuming that for the same climatic zone under the same rainfall regime, the annual rainfall is proportional in-between stations, with a little random variation due to rain distribution in the zone (Espinoza *et al.*, 2009). In that report, nine regions were delineated mainly located over the northern coastal region.

Multivariate analysis techniques have proved their efficiency to delineate homogeneous regions based on climatic features such as rainfall data. Many authors have used factor analysis, principal components, clustering techniques or a mixture of all these techniques, to define more precisely climatic zones or rainfall regions (e.g.

* Correspondence to: P. Rau, Géosciences Environnement Toulouse, UMR 5563, Université de Toulouse, CNRS-IRD-OMP-CNES, 14, avenue Édouard Belin, Toulouse 31400, France. E-mail: pedro.rau@get.obs-mip.fr

Ünal *et al.*, 2003; Raziei *et al.*, 2008), to classify rainfall stations (e.g. Stooksbury and Michaels, 1991; Jackson and Weinand, 1995) and to analyse rainfall variability or distribution patterns (Sneyers *et al.*, 1989; Ramos, 2001; Muñoz-Diaz and Rodrigo, 2004; Dezfuli, 2010). Recently Sönmez and Kömüşcü (2011) proposed a rainfall reclassification for Turkey based on k-means methodology highlighting the benefit over prior techniques as regionalization based on topographic and climatic parameters, long-term seasonal rainfall patterns (Türkeş *et al.*, 2002) and hierarchical clustering (Ünal *et al.*, 2003). Sönmez and Kömüşcü (2011) in particular indicate that their method is efficient in grasping shifts in time periods. The advantage of using a cluster analysis in the regionalization procedure stands in the fact that rainfall time series may be non-stationary and/or non-Gaussian due to the complex of influence of climate phenomena (Takahashi and Dewitte, 2015). For instance, the ENSO has a strong positive asymmetry resulting from the fact that strong extreme events are warm events (An and Jin, 2004; Boucharel *et al.*, 2011), resulting in a non-Gaussian distribution of most historical indices (Boucharel *et al.*, 2009). Such non-linearity in the large-scale circulation over the tropical Pacific is likely to influence the rainfall over the Peruvian Pacific slope and coast, which calls for refining regionalization procedures just based on linear techniques (e.g. RVM). This is one of the main motivations of this work that proposes to estimate the improvement brought to rainfall regionalization by applying a hybrid procedure consisting of a combination of two widely used techniques: k-means clustering (Hartigan and Wong, 1979) to obtain a coarse regionalization that is used as first guess in the regionalization using the RVM. Our aim is also to estimate to which extent the decomposition of a large complex narrow area into a reduced number of homogeneous regions can grasp the salient features of the ENSO influence onto rainfall over the Pacific slope and coast of Peru (Horel and Cornejo-Garrido, 1986; Goldberg *et al.*, 1987; Tapley and Waylen, 1990; Takahashi, 2004; Nickl, 2007; Lagos *et al.*, 2008; Lavado *et al.*, 2012; Lavado and Espinoza, 2014; Bourrel *et al.*, 2015). The identified regions are also aimed at being used for ecological and water resources management and easing the interpretation of the manifestation of the main climatic modes in the region as described in Muñoz-Diaz and Rodrigo (2004), Sönmez and Kömüşcü (2011) and Parracho *et al.* (2015).

2. Study area

The study area comprises the Peruvian Pacific slope and coast that covers an area of ~280,500 km^2. This region borders with the Andes mountains by the east (69.8°W), whereas extending west to the Pacific Ocean (81.3°W). It borders with Ecuador in the north (3.4°N) and with Chile in the south (18.4°S). Its maximum width, perpendicular to the coastline reaches 230 km in the southern part and is only 100 km in the northern part. This area is characterized by a significant altitudinal gradient ranging from 0 to ~6500 m asl and includes 54 main river watersheds that cover near 90% of this region. The rivers generally flows from east to west from the Andes towards the Pacific Ocean with bare and steep slopes that favour significant rising, flooding and erosion during highly rainy episodes (Lavado *et al.*, 2012). On the one hand, under normal conditions, this region is influenced by the Southern Pacific Anticyclone in combination with the Humboldt current (cold SSTs) which produces dry and stable conditions to the western central Andes, with moist air trapped below the inversion zone at about 900 hPa–1000 m asl (Vuille *et al.*, 2000; Garreaud *et al.*, 2002), conditions that produce extreme aridity until about that altitude (Lavado *et al.*, 2012). Over this altitudinal limit, it is known that there is an influence of the southward displacement of the ITCZ and is supposed that other mechanisms influencing over the Peruvian Andes, also influence over the Peruvian Pacific slope (i.e. humidity transport from the Amazon, Bolivian High, etc.) (Nickl, 2007; Lagos *et al.*, 2008), nevertheless this has not be studied to date. On the other hand, this region exhibits greater seasonal and interannual rainfall variability than the two main others hydrological regions of Peru: the Amazon and the endorheic Titicaca drainage areas (Lavado *et al.*, 2012), mainly caused by the ENSO influence in the northern areas during the rainy season, with no clear evidence of the ENSO influence for central and southern areas (Lagos *et al.*, 2008; Lavado *et al.*, 2012; Lavado and Espinoza, 2014).

3. Data

3.1. Rainfall dataset

The database includes monthly rainfall records from 139 meteorological stations managed by the SENAMHI (National Meteorological and Hydrological Service of Peru) and 6 meteorological stations managed by the INAMHI (National Meteorological and Hydrological Institute of Ecuador). It was necessary to extend the area into the foothills of the northern Andes, which cover bi-national river watersheds between Peru and Ecuador. Monthly rainfall data are available over 1964–2011 period. Over the 145 stations, 124 stations are located in the Pacific slope and coastal region of Peru (see Figure 1) and 11 belong to the Peruvian Atlantic drainage and 4 to the Titicaca drainage. A careful quality check of this data was performed using the RVM. In this dataset containing 145 stations records, 76% of them present more than 45 years of continuous records, 20% of them between 20 and 45 years of continuous records and only 4% of them between 15 and 20 years of continuous records.

3.2. Sea surface temperature and ENSO indices

We used global values of *in situ* monthly SST obtained from the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) dataset (Rayner *et al.*, 2003) over 1964–2011 time interval at 150°–0°E, 25°S–25°N of the Pacific and Atlantic basins, which can be downloaded at: http://www.metoffice.gov.uk/hadobs/hadisst/data/

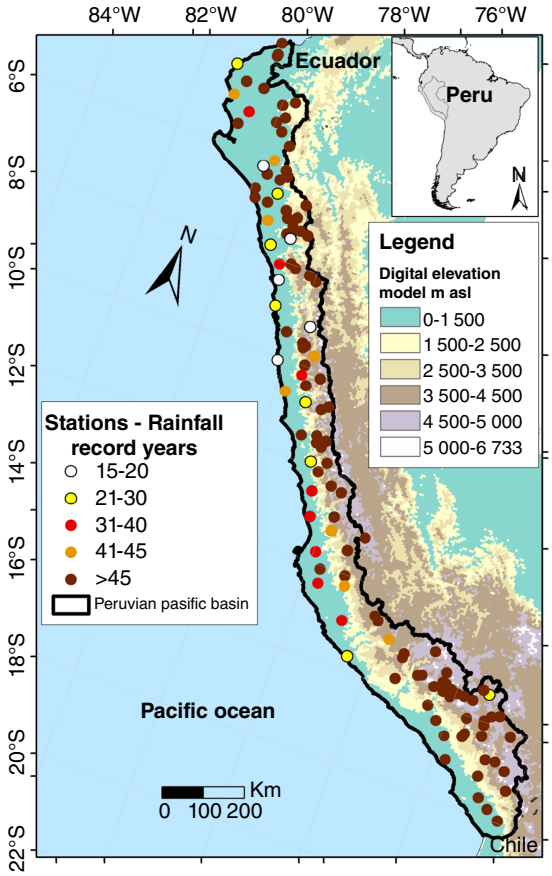

Figure 1. Geographical distribution of stations over the Peruvian Pacific slope and coast. The contour in black thick line delimits the region of interest. Rainfall record length of the stations is indicated by the colours of the dots (five stations are represented by white points and have between 15 and 20 years of data record, eight by yellow that have between 21 and 30 years, eight by red that have between 31 and 40 years, eight by orange that have between 41 and 45 years and 95 stations have more than 45 years of data records, represented by brown points). A Digital Elevation Model (SRTM – 90 m) shows the topographical characteristics and altitudes of the study area.

download.html available on a $1^\circ \times 1^\circ$ grid. The evaluation of the relationship between rainfall variability and Tropical Pacific SST was conducted by using two oceanic indices (the E and C indices) proposed in a recent study by Takahashi *et al.* (2011). These indices are by construction independent (orthogonal) and describe the two main modes of ENSO related to the variability of Eastern equatorial Pacific (E) and Central equatorial Pacific (C). These indices were also used in Lavado and Espinoza (2014) and Bourrel *et al.* (2015) for depicting the influence of the tropical Pacific SST onto rainfall over Peru.

4. Methods

The methodology is composed of three steps summarized in Figure 2: the first one corresponds to the reviewing of rainfall data, their homogenization by RVM and the filling of missing monthly rainfall data; the second one corresponds to the regionalization process including k-means clustering and RVM analysis using an iterative process by trial and error with the goal of searching regions that present similar annual rainfall amount, and interannual variability; and the last one corresponds to a detailed characterization of the rainfall regime in each region in response to climate variability at seasonal and interannual timescales.

4.1. Data homogenization and validation

It was carried out in three steps:

1 The analysis period was chosen to be as long as possible for a significant number of stations over the Pacific slope and coast and complementary stations presented in data section. We also impose that the selected stations should have at least continuous records longer than 15 years.

2 To evaluate the homogeneity of datasets for identifying inconsistent information in terms of quality issues as: station microenvironment, instrumentation, variations in time and position (Changnon and Kenneth, 2006); we used here the RVM. It relies on the principle of annual rainfall proportionality between neighbouring stations represented as rainfall indexes which characterize the rainfall pattern of a predetermined area. The RVM is based on the calculation of an extended rainfall vector within the study period. This concept refers to the calculation of a weighted average of rainfall anomalies for each station, overcoming the effects of stations with extreme and low values of rainfall. Then, the regional annual pluviometric indices Z_i and the extended average rainfall P_j are found by using the least squares technique. This could be obtained by minimizing the sum of Equation (1).

$$\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\frac{P_{ij}}{P_j} - Z_i\right) \tag{1}$$

where i is the year index, j the station index, N the number of years, and M the number of stations. P_{ij} stands for the annual rainfall in the station j, year i; P_j is the extended average rainfall period of N years; and finally, Z_i is the regional pluviometric index of year i. The complete set of Z_i values over the entire period is known as 'regional annual pluviometric indexes vector'. Being an iterative process, this method allows to calculate the vector of each of the predefined regions, then provides a stations – vector interannual variability comparison, for finally discards those that are not consistent with the regional vector (RV). This process is repeated as much as necessary. See more details of this method in Espinoza *et al.* (2009).

3 For the stations that were selected during the homogenization process and also had missing monthly data, once their spatial representation proved significant, were subjected to a process of information completion. In this case, this procedure was performed using the values of rainfall index calculated from the RV and the mean value of rainfall monthly data of the concerned station.

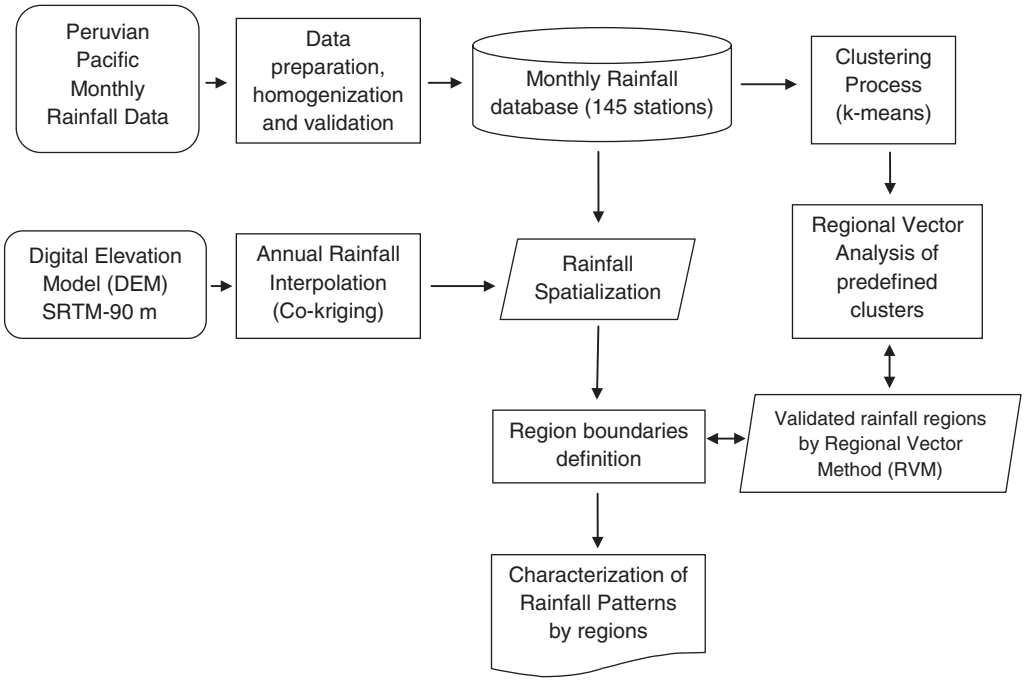

Figure 2. Schematic of the methodological steps for the regionalization of the rainfall time series.

A more detailed description can be found in Bourrel *et al.* (2015).

Through these three stages, 145 pluviometric stations were validated. The geographical location of the 124 Peruvian Pacific slope and coastal stations is depicted in Figure 1, which also mentions the rainfall record length for each station.

4.2. Classification and regionalization process

In this section, we described the regionalization process using the RVM approach, which required a first guess to initialize the process. In this study, this first guess is obtained performing a k-means clustering as a classification of rainfall data from the stations selected in 3.1.

4.2.1. k-Means clustering technique

k-Means clustering is a statistical technique designed to assign objects to a fixed number of groups (clusters) based on a set of specified variables. One of the principal advantages of k-means technique consists in its cluster's identifying performance which allows ranking the obtained clusters as a function of their representativeness. The process involves a partitioning schema into k different clusters previously defined. Objects that are within those k clusters must be as similar as possible to those that belongs to its own group and completely dissimilar to the objects that are in the other clusters. Similarity depends on correlation, average difference or another type of metrics. By definition each cluster is characterized by its own centroid with the cluster members located all around it. The algorithm used at annual rainfall timescale was the Hartigan–Wong which adopts the squared Euclidean distance as a dissimilarity measurement. See more details of this method in Hartigan and Wong (1979).

A key part of the k-means application is to define an optimum number of clusters. In order to succeed in the definition of partitioning groups, an estimation of the silhouette number must be performed for each desired number of groups. The silhouette width is used to evaluate the statistical significance of each identified cluster (Rousseeuw, 1987). The silhouette value is obtained following Rousseeuw (1987) as:

$$S(i) = \frac{\min\{b(i,k)\} - a(i)}{\max\{a(i), \min(b(i,k))\}} \tag{2}$$

where $a(i)$ corresponds to the average similarity between the ith object and the other objects of the same group and $b(i,k)$ is the average similarity between the ith object and the members of the kth clusters. The range of variation for this silhouette index is between -1 and $+1$, when the silhouette value is close to $+1$ means that there is a better member correspondence to its own cluster, whereas a negative value represents the object this is not well located in the appropriate cluster. Meanwhile the value of 0 means that objects could belong to any k cluster. We also compute an average silhouette width for the whole k clusters which represents the mean of $S(i)$, and it can be used to choose the best number of clusters, by taking the value of k for which $S(i)$ is maximal.

4.2.2. Regionalization analysis

There are classical ways to predefine regions; it can be based on stations proximity and homogeneity, physiographic patterns or topographical constraints related to isohyets (Espinoza *et al.*, 2009; Bourrel *et al.*, 2015). Here,

rainfall stations grouped by k-means clustering are set up as predefined regions. The criteria for using k-means clustering as first step of regionalization is based on the advantages in time solving and the preset number of groups at the beginning of the process whilst RVM requires defining the stations grouped into a predefined region, being a long and exhaustive methodology if it is not provided an accurate number of groups.

Regionalization was performed using the RVM, which is generally oriented to: (a) assess rainfall data quality based on the homogeneity within a predetermined region (Espinoza *et al.*, 2009) and (b) achieve rainfall regionalization processes (establishment of representative vectors of homogeneous rainfall zones) to gather the stations exhibiting the same interannual variability. The process for regionalization is similar to the process explained in Section 4.1 (item 2). It depends on the computation of a 'mean station' or 'vector' from all data involved in the study area that will be compared with each pluviometric station (Brunet-Moret, 1979). Prior to the use of the RVM, it is necessary to group stations into predefined regions.

Once calculated, the RV is compared iteratively with data station for discarding those stations whose data are not consistent with the RV and reprise the process. The rejection of a given station could mean that this station belongs to a neighbouring region that could present greater consistency. Therefore in many cases, stations or areas are re-grouped or divided in order to obtain regions that show homogeneous features. The main statistical criteria for regrouping stations into homogeneous regions is based on thresholds applied to the standard deviation of the differences between annual pluviometric indices of stations and the RV indices; and to the correlation coefficient between RV and annual pluviometric values of stations. These thresholds are fixed to the standard deviation lower than 0.4 and correlation coefficient greater than 0.7. Rainfall database management and RVM were carried out using the HYDRACCESS software (Vauchel, 2005).

4.2.3. Rainfall data interpolation

After regionalization based on punctual information (i.e. rainfall stations), it was done a rainfall spatialization by isohyets allowing to delimit polygonal regions. Annual rainfall was interpolated incorporating elevation data using the co-kriging classical geostatistical approach, which is widely used in the hydrometeorological field (Goovaerts, 2000; Diodato, 2005; Buytaert *et al.*, 2006). Co-kriging, which is a multivariate version of kriging technique, took into account the digital elevation model (DEM) provided by NASA-NGA, Shuttle Radar Topographic Mission (SRTM – 90 m) data (http://srtm.csi.cgiar.org) as correlated secondary information based on a spherical variogram (Goovaerts, 2000; Mair and Fares, 2011). This rainfall interpolation map was used as a background raster guide for delineating polygonal regions involving the station points grouped with regionalization analysis. These polygons follow the isohyets shape with geometrical approach (perpendicular and bisector criteria of boundaries of regions traversing isohyets and stations) and a statistical approach (revalidation of new defined areas with the RVM with proper fit of stations inside each region).

Finally, representative monthly rainfall time series of each region were obtained with the co-kriging methodology because of better performance than other techniques (e.g. Thiessen Polygons, Inverse Distance Weighted and Kriging) over mountains areas (Hevesi *et al.*, 1992a, 1992b; Goovaerts, 2000; Diodato, 2005). Time series were assigned to centroids as representative points for obtain mean latitude, longitude and altitude of each region.

4.3. Rainfall variability and sea surface temperature anomalies

In order to investigate the relationship between rainfall and ENSO, a covariance analysis (i.e. singular value decomposition – SVD) is used, which consists in deriving the eigenvectors and eigenvalues of the covariance matrix between rainfall anomalies (December–January–February–March–April mean) over the Peruvian Pacific slope and coast and the SST anomalies over the Pacific and Atlantic basins (December–January–February mean) that maximizes the fraction of the cumulative squared covariance (Yang and Lau, 2004). Data were previously detrended in the period 1964 to 2011. More details and comments about this technique can be found in Bretherton *et al.* (1992) and Cherry (1997). In order to provide an estimate of the statistical significance of the SVD modes, a Monte–Carlo test is performed that consists in creating a surrogate data, a randomized dataset of rainfall and sea surface temperature by scrambling 40 yearly maps among the 48 years in the time domain. The SVD is then performed on the scrambled dataset. The same procedure of scrambling the dataset and performing the analysis is repeated 500 times, each time keeping the value of the explained covariance of the first two dominant modes and comparing the SVD modes of the original dataset and the ones of the scrambled dataset. The method is described in Björnsson and Venegas (1997). The 90% confidence level of the mode patterns is defined so as to the 10 and 90% percentiles of the ensemble correspond to a value that differs from the estimated mode by less than 0.5 times the standard deviation among the ensemble.

5. Results

5.1. Rainfall classification

A cluster analysis of the annual rainfall data was performed by applying k-means technique on the 124 rainfall stations previously selected. The optimal value for the cluster numbers was determined by an average silhouette value and a negative silhouette number for a number of cluster groups varying from 3 to 10 (see Table 1).

Maximum silhouette values are obtained for cluster 3 group (0.64), cluster 4 group (0.60) and cluster 6 group (0.55), considering as a reasonable structure a cluster having a silhouette value greater than 0.50 and as a weak structure a silhouette value less than 0.50 following Kononenko

Table 1. Results of the k-means analysis for number of cluster groups varying from 3 to 10.

Number of cluster groups	**3**	4	5	**6**	7	8	9	10
Average silhouette value	**0.64**	0.60	0.54	**0.55**	0.54	0.54	0.46	0.45
Negative silhouette number	**6**	4	9	**6**	8	6	11	9

Optimal values for selecting the number of cluster groups are shown in bold.

and Kukar (2007). The number of negative silhouette values is minimal for cluster 3 group (6), cluster 4 group (4) and cluster 6 group (6). After plotting the cluster groups into a map showing their spatial distribution, we select the cluster 3 and cluster 6 groups among them, as these two clusters show certain arrangement of rainfall stations according to topographical and latitudinal variation (Figure 3(a) and (b)). Cluster 4 group was an intermediate group that corresponds to one sub-region in the north.

The two cluster groups (clusters 3 and 6) exhibit a similar spatial distribution. Pluviometric stations from both groups present an altitudinal distribution along the Pacific slope and coast, defining three regions: the stations located in lowlands (green triangles), in middle altitude basin (white circles) and in highlands (black points). Cluster 6 group presents three additional regions, two of them closely related to northern rainfall features for the middle altitude basin (cluster 4 of cluster 6 group represented by red triangles) and highlands (cluster 6 of cluster 6 group represented by yellow circles). Two stations are considered as isolated (cluster 5 of cluster 6 group represented by blue circles).

Even if cluster 6 group is less representative than cluster 3 group in terms of silhouette value, cluster 6 group is considered acceptable for representing correctly the variability of northern rainfall, offering an initial classification of rainfall or initial approach of rainfall regionalization over the Peruvian Pacific slope and coast.

5.2. Regionalization

After cluster definition, the RVM was performed over these preliminary regions using an iterative process by trial and error, that adds and deletes stations from neighbouring regions considering the criteria described in Section 4.2.2. This process could be also verified by their interannual coefficient of variation (CV). In Figure 4, the stations located in the western area of the coast (lowlands) present greater values of CV (>1.8) than those located in middle altitude basin and in highlands. Northern region presents higher CV values in lowlands and in the middle altitude basin. Highlands present lower CV values (<0.8) along the Pacific slope independently of the latitude.

High CV values in the northern region correspond to strong interannual rainfall variability with anomalies greater than 1000 mm year^{-1}. High CV values are also observed along southern latitude which are mostly caused by small values around the near zero annual average. These values are due to the large-scale mid-tropospheric subsidence over the southeastern subtropical Pacific Ocean, enhanced by the coastal upwelling of cold water (Enfield, 1981; Virji, 1981; Vuille *et al.*, 2000; Garreaud *et al.*, 2002; Lavado *et al.*, 2012).

Based on the iterative process of the RVM, we identify nine homogeneous rainfall patterns (see Figure 5). Comparing with the initial cluster groups derived from k-means, rainfall stations from clusters 1, 2 and 4 located in the coastal zone and northern Andes (see Figure 3(b)) exhibit higher coefficients of variation in coastal proximity (see Figure 4). Cluster 1 includes the regions 1, 4 and 7 along the coastal zone. Cluster 4 defines region 2: in this case, clustering process successfully assigned each station as well as RV reported them as separate from other regions. Cluster 5 and 6 are regrouped into region 3. Finally, cluster 3 defines regions 5, 6, 8 and 9: in this case, the low variability, their high altitude as the latitudinal extension, defines these four regions.

k-Means methodology and RVM did not provide a final regionalization by their own. For clustering method, some groups are not well defined because of isolated stations to be included in other groups, associated to low silhouette values (see cluster 2 from cluster 3 group in Figure 3(a) and clusters 2, 3 and 6 from cluster 6 group in Figure 3(b)). This can be explained by the characteristics of the annual rainfall database used, related to the presence of non-globular clusters with a chain-like shape or with not well defined centres (see regions 3, 8 and 9 in Figure 5), which are one of the principal disadvantages using this technique (Kaufman and Rousseeuw, 1990). For the RVM, it is possible to obtain grouped regions following only the statistical criteria with the thresholds presented in Section 4.2.2.. However, there is the risk of increasing computing time and obtaining unrealistic groups because of using only a statistical criteria and not an initial arrangement inferred in this case from the k-means clustering. k-Means inferred three regions for lowlands, middle altitude basin and highlands (see Figure 3(b)) as a first guess to the final regionalization by the RVM in the north (see regions 1, 2 and 3 in Figure 5) and in the south (see regions 7, 8 and 9 in Figure 5), which it not was possible to identify using only the RVM (not shown). The two-step methodology (k-means and RVM) has also presented a slight improvement in the thresholds of Section 4.2.2. with respect to the thresholds obtained with RVM only, with about +6% for the standard deviation of the differences between annual pluviometric indices of stations and the RV indices (from 0.39 to 0.42); and about +0.5% for the correlation coefficient between RV and annual pluviometric values of stations (from 0.78 to 0.79).

Following the approach summarized in the flow chart presented in Figure 2 and applying the methodology described in Section 4.2.3.; the nine regions were well delineated taking into account the rainfall interpolation map as shown in Figure 5. Annual rainfall in each region exhibits a relationship with altitude and latitude, rainfall is higher at low latitudes and at southern latitudes in high altitudes as shown in Figures 5 and 6.

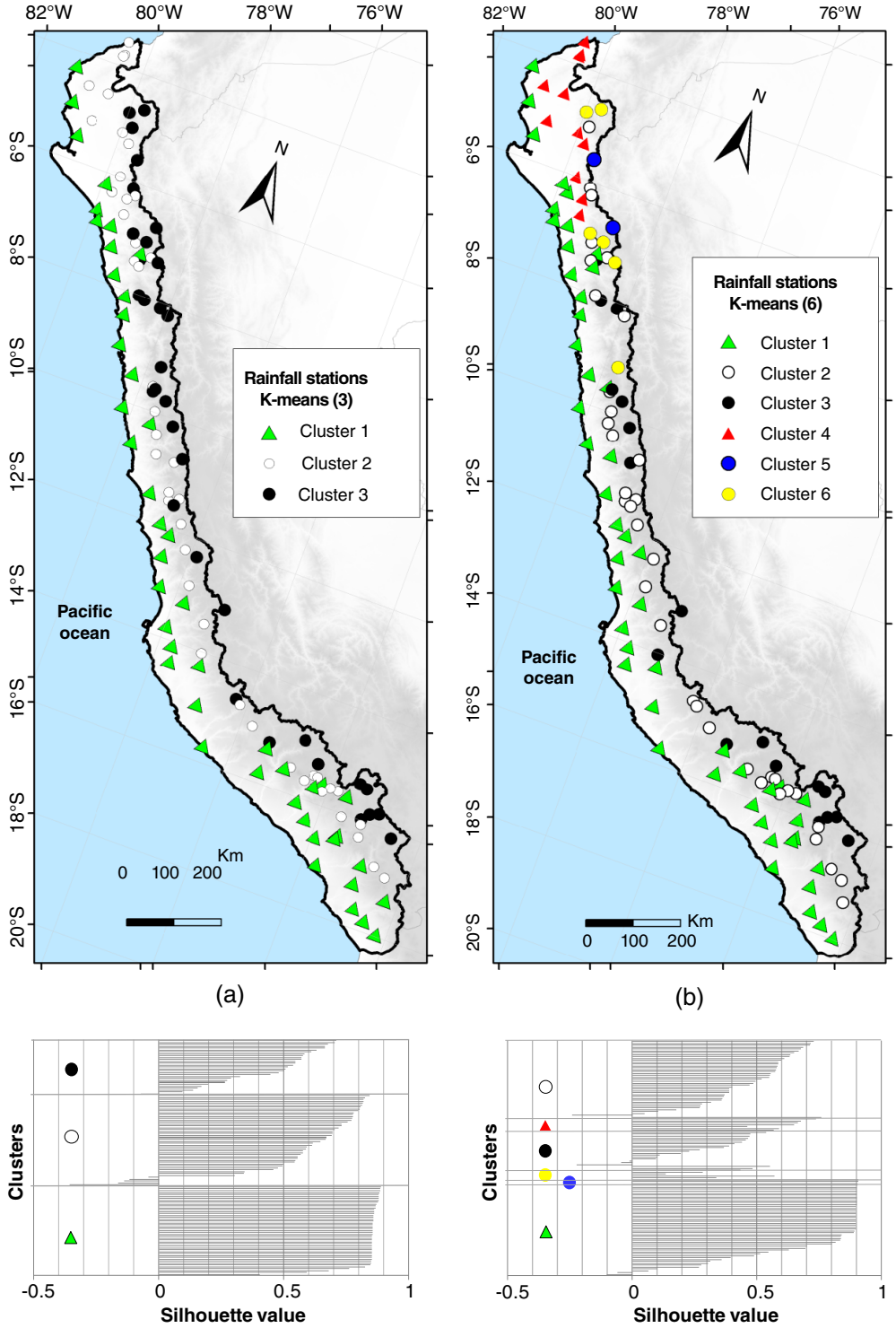

Figure 3. Spatial distribution of (a) cluster 3 group and (b) cluster 6 group obtained with the k-means process. Silhouette value for each cluster group is provided in the bottom panels.

Correlation coefficient between the stations and the regional vector of each region was calculated separately and the spatial distribution of these coefficients of correlation is shown in Figure 7. The purpose of this analysis is to emphasize the level of representation of the regional vector and identify locally the areas within a region where this vector is more representative. Considering regions 4 and 7, the coefficient of correlation is less than 0.7 and greater than 0.5. These coefficients are considered as acceptable considering the dryer conditions with more than 90% of the rainfall records near zero throughout the year due to hydroclimatic features, where any value greater than zero causes a strong variability reducing the relationship with its RV. For the northern regions 1 and 2, the mean correlation is more than 0.9 being a very good representation of RV and the more representative areas are shown

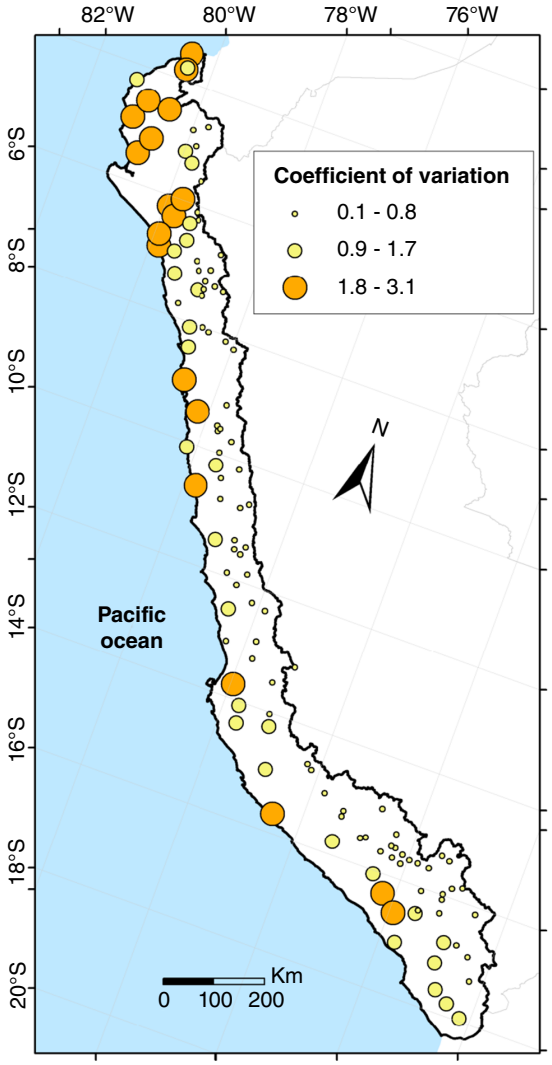

Figure 4. Spatial distribution and range of coefficient of variation (CV) for all of the pluviometric stations of the Peruvian Pacific slope and coast network.

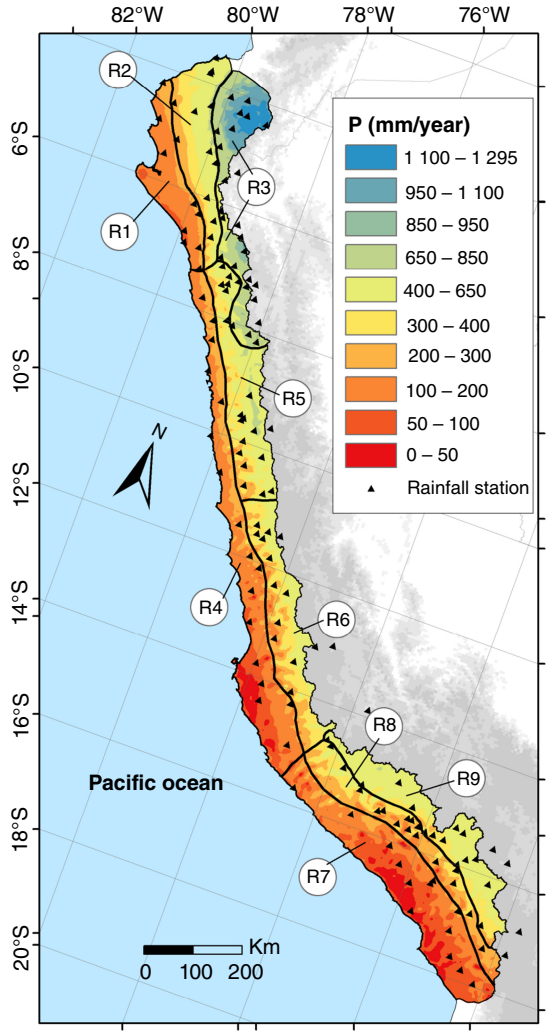

Figure 5. The nine homogeneous rainfall regions after the regionalization process of clustering and RVM. Interpolated surface of annual rainfall (isohyets obtained using co-kriging method) is also shown to demonstrate rainfall differences between regions.

in red coloration. The strong correlation values are due to extreme rainfall events related with ENSO strong events increasing the association level between stations and RV. Regions 3, 5, 6, 8 and 9 located in highlands, have correlations greater than 0.7 being a good representation of the RV with the more representatives areas in orange coloration.

5.3. Regions characterization

In this section, we document the rainfall seasonal distribution and interannual variability over the nine identified regions. Some of geographical features (area, latitudinal and altitudinal ranges) are presented in Table 2. All regions present a unimodal rainfall seasonal distribution (see Figure 8) and differ from their peak calendar month, intensity and duration of the rainy season.

Region 1 extends over northern lowlands including drier areas as the Sechura desert (79°–81°W and 5.5°–6.5°S) where the average interannual rainfall is about 90 mm year^{-1}. A maximum seasonal rainfall is observed in March (see Figure 8(a)1) with a rainy season from January to May (JFMAM) with values less than 50 mm month^{-1} which represent near to 90% of the annual rainfall. The rest of the year is considered as dry due to values near or equal to zero, corroborating the irregularity in the seasonal rainfall pattern (see Figure 8(a)1) and in the interannual variability of monthly rainfall (see Figure 8(a)2) at the coast (Garreaud *et al.*, 2002; Lavado *et al.*, 2012).

Region 2 comprises a large part that belongs to the foothills of the northern Andes covering bi-national river watersheds of Peru and Ecuador. This zone exhibits an irregular seasonal rainfall pattern (see Figure 8(b)1) and an irregular interannual variability of monthly rainfall (see Figure 8(b)2). Average interannual rainfall value is around 370 mm year^{-1}. The wettest period occurs between January and April (JFMA) cumulating near to 90% of total rainfall.

Northern coastal regions as regions 1 and 2 are significantly affected by strong events represented by two peaks reaching 413 mm month^{-1} in March 1983

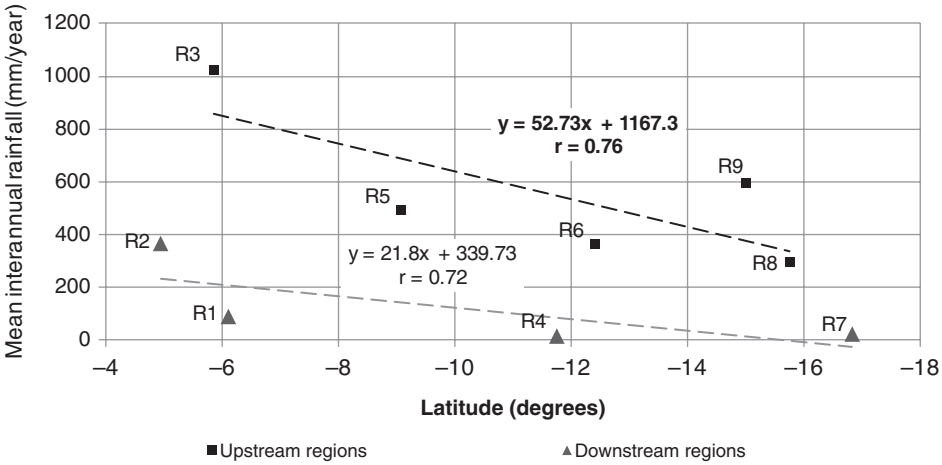

Figure 6. Relationship between mean interannual rainfall and latitude for the nine identified regions grouped in upstream and downstream regions. Representation for upstream regions is significant at the 90% level using a Student's *t*-test.

and 299 mm month^{-1} in March 1998 for region 1 (see Figure 8(a)2); and 746 mm month^{-1} in March 1983 and 708 mm month^{-1} in March 1998 for region 2 (see Figure 8(b)2). A summary of rainfall statistics is given in Table 2 and a boxplot representation of monthly rainfall in Figure 9. Outliers from Figure 9, represented by small circles, correspond to values exceeding 1.5 times the interquartile range (IQR). All regions have observations that exceed Q3 + 1.5(IQR), however, northern coastal regions 1 and 2 differs from the rest for having greater number of outliers values (14 and 17%, respectively) with the largest rainfall anomalies reaching 56 and 25 times of mean monthly rainfall for regions 1 and 2, respectively. Most of the interannual variability in rainfall, reflected as well in higher CV values (see Table 2), is directly due to the occurrence of the strong El Niño events indicating also a high intensity of interannual variability than other regions. This is particularly obvious for region 1 where three extreme rainfall events are observed corresponding to the year 1972, 1982 and 1997, known as strong El Niño years. Interestingly the more inland region 2 exhibits interannual variations of rainfall that does not necessarily corresponds to the strong El Niño years. These events may correspond to local convective events associated to coastal warm oceanic conditions related mainly to Kelvin waves and the Madden and Julian Oscillation (MJO) (Woodman, 1985; Bourrel *et al.*, 2015).

Region 3 covers bi-national river watersheds of Peru and Ecuador bordering with the Amazon Basin by the east. This is also the wettest region (see Figures 8(c)1, (c)2 and 9). On the other hand, rainfall amount decreases southward with rainfall regularity in the seasonal pattern (see Figure 8(c)1) and in the interannual variability of monthly rainfall (see Figure 8(c)2), with a rainy season from January to April (JFMA) that represents almost 70% of the annual rainfall. Mean interannual rainfall reaches 1024 mm year^{-1}, representing 11 times of the mean interannual rainfall of region 1 and 3 times of region 2 (see Table 2). The rainfall interannual variations are weakly associated to the extreme El Niño events (the correlation between the E index and rainfall is 0.2) but is negatively correlated to the C index ($r = -0.4$) indicating that the R3 region is sensitive to cool enhanced coastal conditions during Central Pacific El Niño events (Bourrel *et al.*, 2015). The inter-events fluctuations are also noticeable which are related to local convective events not related to ENSO but mostly by the ITCZ and the large-scale atmospheric variability associated to the MJO (Tapley and Waylen, 1990; Takahashi, 2004; Bourrel *et al.*, 2015). Also noteworthy, there is an increase of rainfall peaks frequency over the last two decades (see Figure 8(c)2).

Region 4 is the longest region located between the coastal plain and the foothills of the western Andes and contains some of the major coastal cities as the capital Lima. This region corresponds to a zone influenced by the large-scale mid-tropospheric subsidence of the south-eastern subtropical Pacific Ocean, enhanced by the coastal upwelling of cold water (Vuille *et al.*, 2000; Garreaud *et al.*, 2002; Lavado *et al.*, 2012) without presenting a relationship between strong rainfall peaks and strong ENSO events. Then, mean interannual rainfall reaches a value of 16 mm year^{-1} defining the driest region in the country (see Table 2) with rainfall irregularity in the seasonal pattern (see Figure 8(d)1) and in the interannual variability of monthly rainfall (see Figure 8(d)2) very common in coastal regions (see Figure 9). The wet period from January to March (JFM) represents near to 75% of the annual rainfall. In the southern part, drier areas are found such as the Nazca desert (74.5°–75.5°W and 14.5°–15.5°S).

Region 5 comprises a border with region 3 and the Amazon Basin by the east. The mean interannual rainfall reaches 492 mm year^{-1} and the wet period occurs between December and April (DJFMA) cumulating near to 80% of total rainfall. No rainfall peaks were identified during strong El Niño events (see Figure 8(e)2) as those in regions 1 and 2, suggesting that rainfall in regions 4 and 5 are probably to be affected by others processes, either local (e.g. coastal SST) or non-local (e.g. dry air transport from the southern region that reduce the rainfall), resulting in a heterogeneous interannual variability of monthly rainfall (see

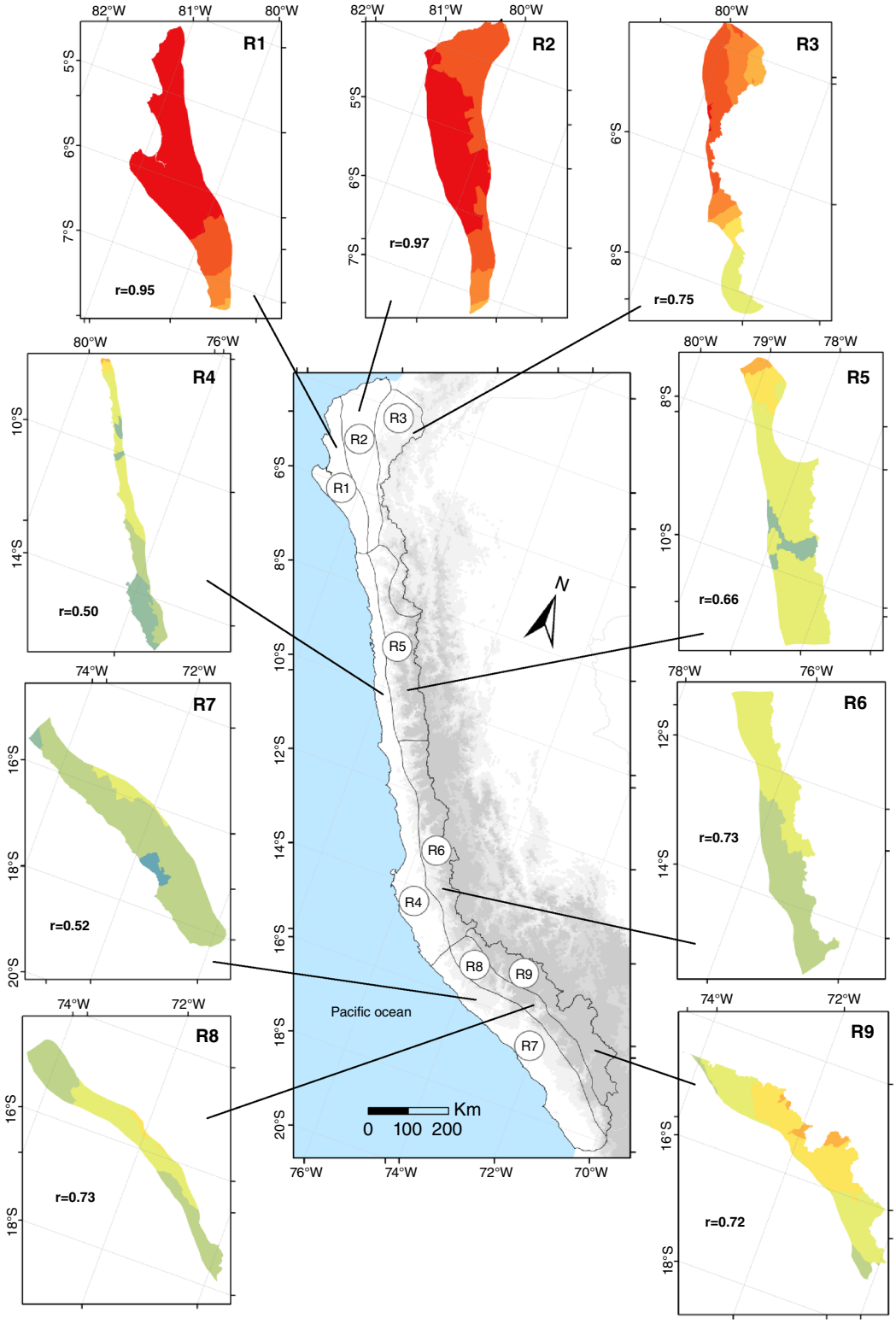

Figure 7. Coefficients of correlation of the stations and the final regional vector of each region identified after the regionalization process (k-means and RVM). A mean value of correlation is also provided by region in bold as well as the spatial distribution of correlation with the regional vector. Correlations are significants at the 90% level using a Student's *t*-test.

Figure 8(e)2) with a low value of coefficient of variation around 0.3 (see Table 2).

Region 6 borders with the Amazon Basin by the east and shows a heterogeneous rainfall pattern without distinguishing any peak corresponding to the strong El Niño events (see Figure 8(f)2). Rainfall distribution is well defined with a rainy season from December to March (DJFM) that represents near to 85% of the annual rainfall (see Figure 8(f)1) and with a mean interannual rainfall reaching 366 mm year^{-1}.

Table 2. Geographical features and annual rainfall values for the nine identified regions.

Region	Area (km^2)	Altitudinal range (m asl)	Latitudinal range (°S)	Annual minimum rainfall (mm $year^{-1}$)	Annual maximum rainfall (mm $year^{-1}$)	Annual average rainfall (mm $year^{-1}$)	CV	SD (mm $year^{-1}$)
1	20 300	0–500	4.2–7.3	3.2	1345.2	89.7	2.6	233.3
2	27 600	0–1500	3.4–7.3	17.3	2772.2	366.5	1.5	534.2
3	27 200	1500–3500	3.6–8.3	533.0	1812.9	1023.7	0.3	294.4
4	48 600	0–1500	7.3–15.5	1.6	62.2	15.5	0.7	11.4
5	32 500	1000–5000	7–11	174.1	825.8	492.4	0.3	145.8
6	30 400	2000–5000	11–15	75.0	693.5	365.9	0.4	133.3
7	49 300	0–2500	15.5–18.4	5.1	54.9	23.2	0.6	13.5
8	25 400	2500–4000	14.6–17.8	23.2	528.8	296.1	0.4	111.8
9	30 100	3500–5500	14.4–17.7	220.5	833.2	594.0	0.2	143.2

Region 7 is characterized by lower rainfall regime with a rainy season from JFM accounting for 65% of the annual rainfall. Furthermore, this region is one of the driest areas in the country where the interannual rainfall (23 mm $year^{-1}$) presenting rainfall irregularity in the seasonal pattern (see Figure 8(g)1) and in the interannual variability of monthly rainfall (see Figure 8(g)2). This region could be considered as an extension of region 4, also influenced by the large-scale mid-tropospheric subsidence of the southeastern subtropical Pacific Ocean but differing in the increase of rainfall peaks frequency in the last decade unlike region 4 as can be seen in Figure 8(g)2.

Region 8 comprises an area thus belongs to the foothills of the southern Andes. This zone exhibits irregular rainfall in the seasonal pattern (see Figure 8(h)1) and in the interannual variability of monthly rainfall (see Figure 8(h)2). The mean interannual rainfall presents a higher value than region 7, reaching 296 mm $year^{-1}$. The wettest period occurs between December and March (DJFM) cumulating near to 90% of total rainfall (see Figure 8(h)1).

Finally, region 9 borders with the Titicaca Basin in the south and east and with the Amazon Basin by the east. The mean interannual rainfall reaches 594 mm $year^{-1}$ and the wet period occurs between December and March (DJFM) cumulating near to 80% of total rainfall. Similar to region 8, region 9 presents a deficit in rainfall during strong El Niño events (see Figure 8(h)2 and (i)2). However, unlike region 8, it presents rainfall regularity in the seasonal pattern (see Figure 8(i)1) and in the interannual variability of monthly rainfall (see Figure 8(i)2) associated with a low value of coefficient of variation around 0.2 (see Table 2) indicating also the lowest intensity of interannual variability. Up to this point, we propose a co-variability analysis between rainfall and tropical SST (see paragraphs below) to deepen the understanding of the relationship between regions and ENSO. Other climatological variables mentioned in Section 2 will need further research and are out of scope of this work.

In order to estimate the value of the regionalization for interpreting the impact of climatic variability over rainfall along the Pacific slope and coast of Peru, a covariance analysis is performed between the rainfall time series of the nine regions and the SST anomalies over the Tropical Pacific and Atlantic Oceans. For clarity, the SST anomalies are considered for the peak ENSO season (i.e. December–January–February mean, hereafter DJF) whereas the rainfall fluctuations are considered for the approximate rainy season (i.e. December-January-February-March-April season, hereafter DJFMA). The results of the covariance analysis (see Section 4.3 for details) are presented in Figures 10 and 11, showing the patterns and time series of the first (Figure 10) and second (Figure 11) SVD modes between SST in the tropical Pacific and Atlantic over DJF and rainfall over the regions over DJFMA. Values of the mode patterns, significant at the 90% level, are indicated by the colour shading (Figures 10(b) and 11(b)) and the red colour (Figures 10(a) and 11(a), see method in Section 4.3). The results indicate a significant relationship between both fields because the percentage of covariance is 66% and 23% for the first and second modes, respectively, and the associated time series of the mode patterns are significantly correlated [r value reaches 0.59 (0.54) for mode 1 (2) in Figure 10(c) (Figure 11(c)), respectively]. The first mode for SST accounts for the strong eastern Pacific El Niño variability as suggested by the large positive skewness of the principal component time series associated to the two strong El Niño events of 1997/1998 and 1982/1983. The correlation between the time series associated to the SST mode pattern and the E index reaches 0.80. The second mode is reminiscent of the central Pacific El Niño variability because it has a strong positive loading near the deadline. Its associated principal component time series is strongly correlated to the C index reaching a correlation of 0.96, significant at the 95% level. Interestingly the time series associated to the first mode for SST is also highly correlated with the C index ($r = 0.73$), which indicates that extreme rainfall events are related to both the E and C modes. It explains in particular why the SST mode pattern has a significant loading in the central Pacific which is not the case for the E mode pattern with is more confined towards the coast of Ecuador (see Takahashi *et al.*, 2011).

The analysis of the mode patterns for rainfall clearly indicates that the first mode accounts for extreme rainfall events in the northern part of Peru (regions 1 and 2) whereas the second mode pattern has a larger loading (negative value) for the upstream regions (region 3, 6, 8 and

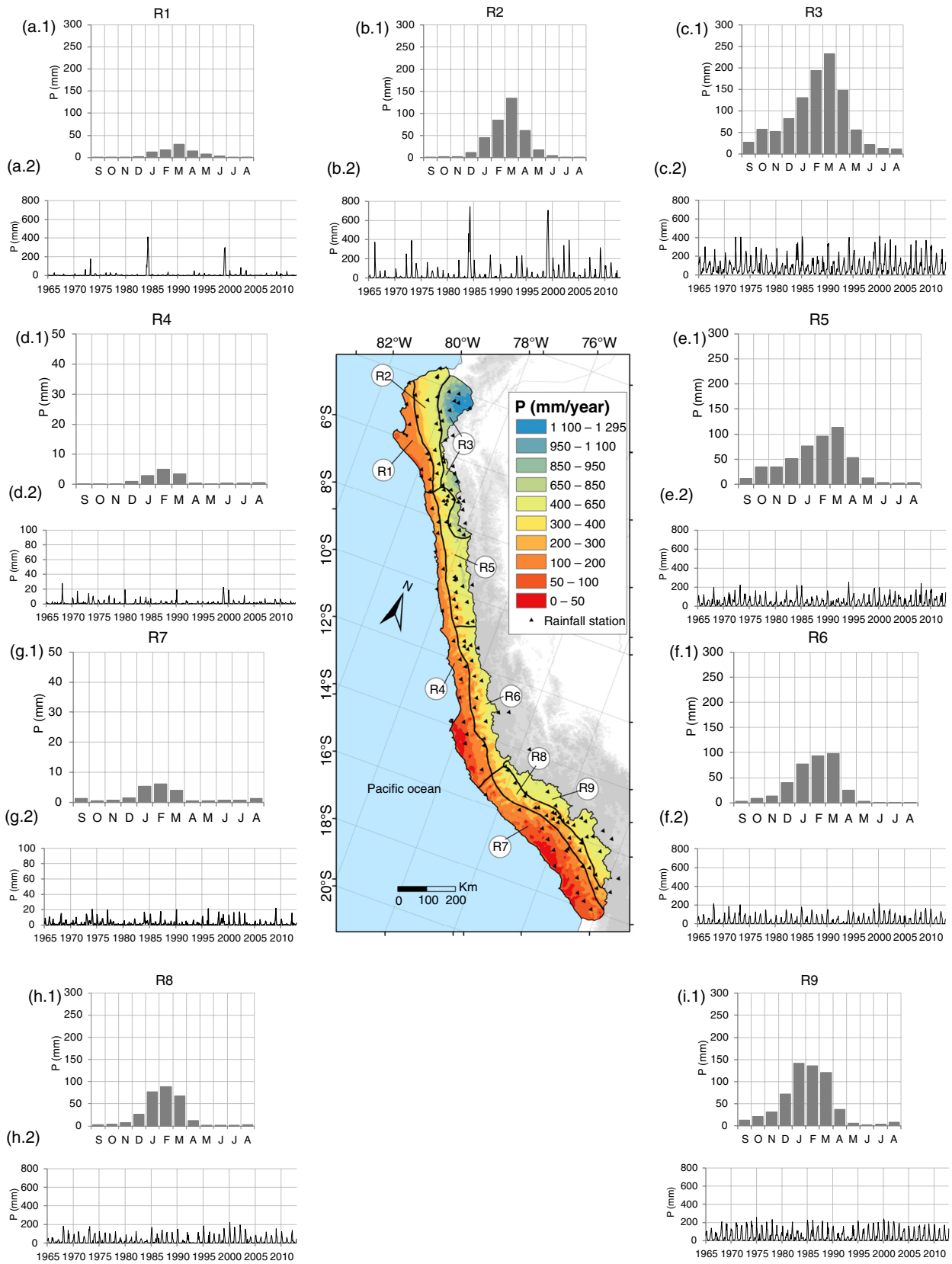

Figure 8. Monthly rainfall regime (1964–2011) for the nine identified regions. A rainfall time series is shown by region. Regions 4 and 7 are shown in a different rainfall scale.

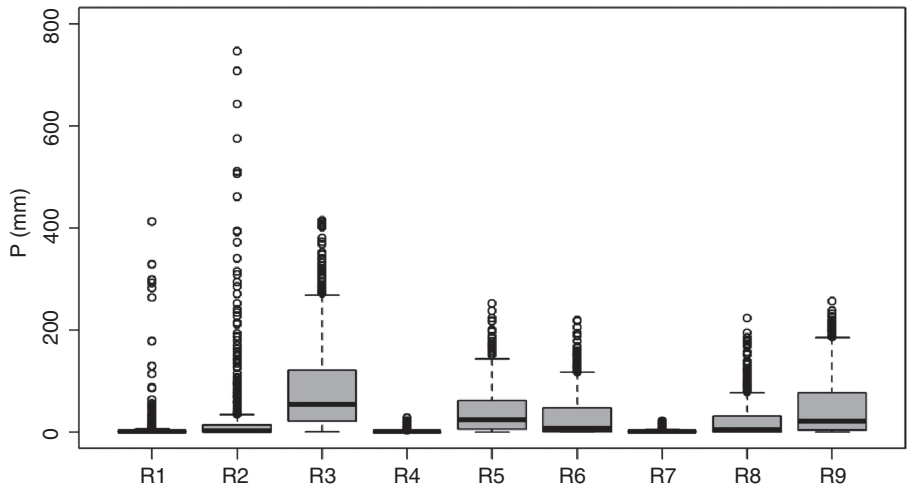

Figure 9. Boxplot of monthly rainfall for the nine identified regions.

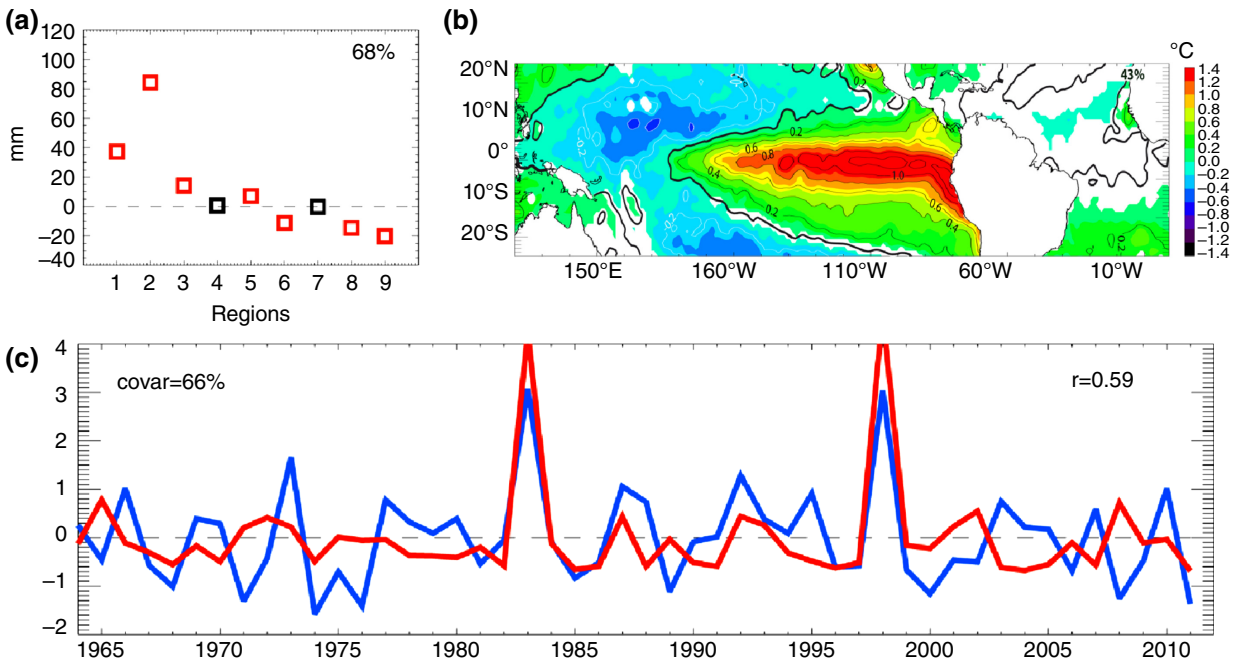

Figure 10. Dominant SVD mode between rainfall in DJFMA over the 9 regions and SST anomalies in DJF over the tropical Pacific and Atlantic. (a) Mode pattern for rainfall. (b) Mode pattern for SST (contour interval is every 0.2 °C). (c) Associated SVD time series for rainfall (red colour) and SST (blue colour). The percentage of variance of the modes is indicated in panels a and b, whereas the percentage of covariance (covar) is indicated in panel c. The correlation (r) between the SVD time series is also indicated in panel c. The contour in thick black line in panel b indicates the zero-contour. The shading (red) colour in panel b (a) respectively, indicates where the mode pattern is statistically significant at the 90% level.

9), clearly indicative that during central Pacific El Niño events, the Pacific slope of Peru experiences a deficit in rainfall that increases with altitude. Note that this analysis is consistent with results from previous works (Lavado and Espinoza, 2014; Bourrel *et al.*, 2015), which analysed the relationship between E and C indices and stations over the Peruvian territory and over the North to Centre of the Peruvian Pacific coast and slope, respectively. We here provide a more quantitative estimate of this relationship through the covariance analysis, which indicates its potential for climate impact studies. In particular, the SVD modes would allow building a linear statistical model of rainfall over the Peruvian Pacific coast using SST as a predictor. The regionalization procedure prior to conducting the SVD analysis is also valuable in easing the interpretation of the ENSO impact on rainfall, in particular, by avoiding probable spurious effects associated to outliers or multiple atmospheric influences. Another important result arising from this analysis is that the extreme rainfall events over the Peruvian Pacific coast are not solely influenced by extreme El Niño events (accounted for by the E mode) but are also influenced by SST in the central equatorial Pacific, as evidenced by the strong correlation between the principal component of the first SVD mode for SST and the C index. This suggests that the magnitude and location along the equator of the SST anomalies in the central Pacific are

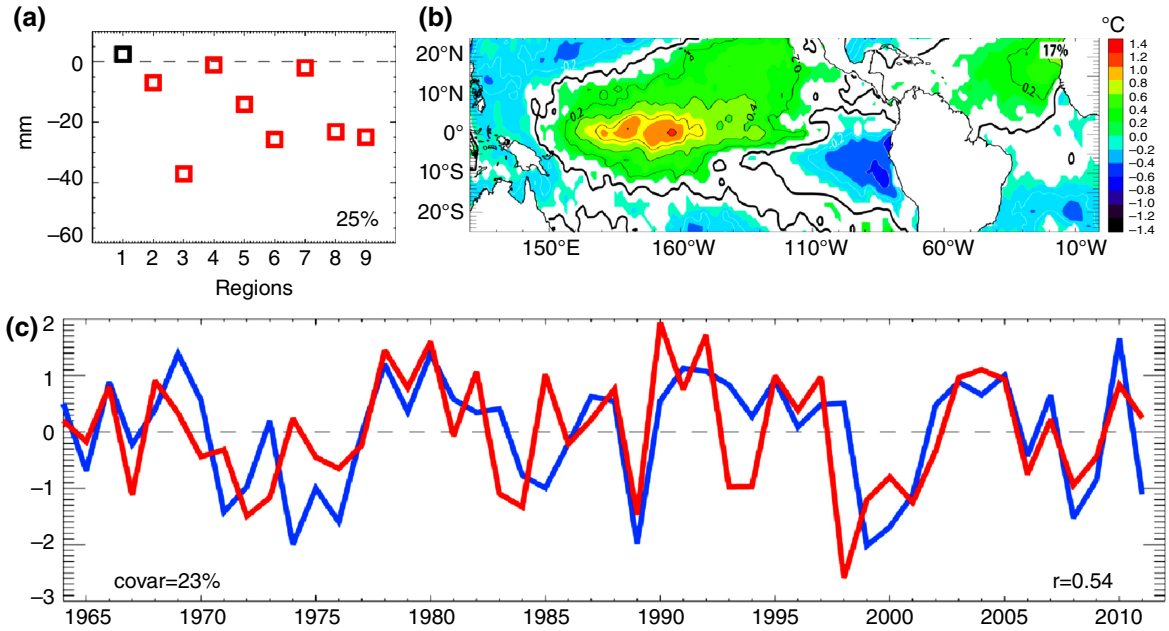

Figure 11. Same as Figure 10, but for the second SVD mode.

important parameters to determine the ENSO impact of rainfall over the Peruvian Pacific coast.

6. Conclusions

This study proposes a method for the regionalization of the rainfall in the Peruvian Pacific slope and coast that consists in a two-step procedure: a preliminary cluster analysis (k-means) followed by the RVM analysis. Using this procedure, nine regions are identified that depicts synthetically the relationship between rainfall variability and altitude and latitude. In particular, rainfall variability is higher at the northern latitudes and it decreases to the south in high altitudes. The motivation for performing a classification using cluster analysis prior to the regionalisation by RVM stands in the complex of processes influencing rainfall variability over this region. In particular, previous studies (Lavado and Espinoza, 2014; Bourrel *et al.*, 2015) have shown that rainfall along the Pacific slope and coast of Peru experiences the influence of both type of El Niño, and due to the strong positive skewness of strong El Niño events, the distribution of rainfall data is not Gaussian, limiting to some extents the linear analysis approach (i.e. RVM). It was in particular verified that our approach leads to a different definition of the regions than an approach based only on RVM. We inferred three regions for lowlands, middle altitude basin and highland in the northern and southern Pacific slope and coast, which was not possible to identify using the method based on the RVM only. The k-means clustering analysis allows for a preliminary grouping of station data that is used as a first guess for the RVM and this step constrains to a large extend the regionalization procedure. The proposed two-step methodology also leads to a slight improvement in the thresholds estimated with the RVM only.

The nine identified regions are shown to grasp the salient features of the influence of ENSO onto rainfall along the Pacific slope and coast of Peru (Horel and Cornejo-Garrido, 1986; Goldberg *et al.*, 1987; Tapley and Waylen, 1990; Takahashi, 2004; Nickl, 2007; Lagos *et al.*, 2008; Lavado *et al.*, 2012; Lavado and Espinoza, 2014; Bourrel *et al.*, 2015), which illustrate its potential for climate impact studies. The dominant co-variability mode between SST in the tropical Atlantic and Pacific Oceans and the reduced set of time series associated to the nine regions has a strong positive loading over the northern part of Peru (regions 1 and 2) for precipitation and over the eastern tropical Pacific for SST, thus accounting for extreme El Niño events. On the other hand, the second mode pattern has a larger loading (negative value) for the upstream regions along the Pacific slope (region 3, 6, 8 and 9), clearly indicative that during central Pacific warming, these regions experience a deficit in rainfall that tends to increase with altitude (more negative in the north than in the south). This is consistent with Lavado and Espinoza (2014) which analysed the relationship between the two types of ENSO and stations over the Peruvian territory, while providing a more synthetic picture of the ENSO influence. In addition, the first co-variability mode between rainfall and SST indicates that extreme rainfall events take place in the North (regions 1 and 2) and are influenced by SST anomalies in the central Pacific (i.e. SST anomalies that project on the C mode), which was not identified in previous works. We attribute this discrepancy between our result and the one by Lavado and Espinoza (2014) to the regionalization procedure that we perform prior to the statistical analysis with ENSO indices. In particular, our regionalization product accounts exclusively for rainfall variability over the Peruvian Pacific continental slope and coast, and is not influenced by stations located

at high-altitude regions that might be influenced by inland circulation patterns. The regionalisation procedure has also the advantage of reducing the influence of outliers in the covariance analysis.

Future work will be dedicated to further investigate the ENSO/rainfall relationship based on the nine identified regions, incorporating other atmospheric and oceanic key indices (cf. Bourrel *et al.*, 2015). Our product will also provide valuable information for hydrological sensitivity analysis over Peruvian Pacific watersheds (through hydrological modelling) for quantifying the effects of climate variability and human activities on runoff with the aim of improving ecological and water resources management.

Acknowledgments

This work was supported by Peruvian Ministry of Education (MINEDU-PRONABEC, scholarship). Authors would like to thank SENAMHI (Meteorological and Hydrological Service of Peru) for providing complete rainfall raw dataset. Boris Dewitte thanks the CNES for support through the TOSCA/CNES project Modokalt.

References

An SI, Jin FF. 2004. Nonlinearity and asymmetry of ENSO. *J. Clim.* **17**: 2399–2412.

BCEOM. 1999. Estudio hidrológico-meteorológico en la vertiente del Pacífico del Perú con fines de evaluación y pronóstico del fenómeno El Niño para prevención y mitigación de desastres. Asociación BCEOM-Sofi Consult S.A – ORSTOM, Programa de apoyo a la emergencia Fenómeno del Niño. Volumen I, Contrato de préstamo n°4250-PE-BIRF, Presidencia de la Republica, Perú, Lima.

Björnsson H, Venegas SA. 1997. A manual for EOF and SVD analyses of climatic data. CCGCR Report No. 97-1, McGill University, Montréal, Québec.

Boucharel J, Dewitte B, Garel B, du Penhoat Y. 2009. ENSO's non-stationary and non-Gaussian character: the role of climate shifts. *Nonlinear Process. Geophys.* **16**: 453–473.

Boucharel J, Dewitte B, Du Penhoat Y *et al.* 2011. ENSO nonlinearity in a warming climate. *Clim. Dyn.* **37**(9–10): 2045–2065.

Bourrel L, Rau P, Dewitte B, Labat D, Lavado W, Coutaud A, Vera A, Alvarado A, Ordoñez J. 2015. Low-frequency modulation and trend of the relationship between ENSO and precipitation along the northern to centre Peruvian Pacific coast. *Hydrol. Process.* **29**(6): 1252–1266.

Bretherton CS, Smith C, Wallace JM. 1992. An intercomparison of methods for finding coupled patterns in climate data. *J. Clim.* **5**: 541–560.

Brunet-Moret Y. 1979. Homogénéisation des précipitations. *Cahiers ORSTOM*. *Serie. Hydrol.* **16**: 3–4.

Buytaert W, Celleri R, Willems P, De Biévre B, Wyseure G. 2006. Spatial and temporal rainfall variability in mountainous areas: a case study from the south Ecuadorian Andes. *J. Hydrol.* **329**: 413–421.

Changnon S, Kenneth K. 2006. Changes in instruments and sites affecting historical weather records: a case study. *J. Atmos. Ocean. Technol.* **23**: 825–828.

Cherry S. 1997. Some comments on singular value decomposition analysis. *J. Clim.* **10**: 1759–1761.

Dezfuli AK. 2010. Spatio-temporal variability of seasonal rainfall in western equatorial Africa. *Theor. Appl. Climatol.* **104**(1–2): 57–69.

Diodato N. 2005. The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain. *Int. J. Climatol.* **25**: 351–363.

Enfield D. 1981. Thermally driven wind variability in the planetary boundary layer above Lima, Peru. *J. Geophys. Res.* **86**(C3): 2005–2016.

Espinoza JC, Ronchail J, Guyot JL, Cochonneau G, Naziano F, Lavado W, De Oliveira E, Pombosa R, Vauchel P. 2009. Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia and Ecuador). *Int. J. Climatol.* **29**: 1574–1594.

Garreaud R, Rutllant J, Fuenzalida H. 2002. Coastal lows along the Subtropical West Coast of South America: mean structure and evolution. *Mon. Weather Rev.* **130**: 75–88.

Garreaud RD, Vuille M, Compagnucci R, Marengo J. 2009. Present-day South American climate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **281**(3–4): 180–195.

Goldberg RA, Tisnado G, Scofield RA. 1987. Characteristics of extreme rainfall events in north-western Peru during the 1982–1983 El Niño period. *J. Geophys. Res.* **92**(C14): 225–241.

Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **228**: 113–129.

Hartigan JA, Wong MA. 1979. Algorithm AS 136: a K-means clustering algorithm. *J. R. Stat. Soc. Ser. C* **28**(1): 100–108.

Hevesi J, Istok J, Flint A. 1992a. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis. *J. Appl. Meteorol. Climatol.* **31**: 661–676.

Hevesi J, Flint A, Istok J. 1992b. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part II: isohyetal maps. *J. Appl. Meteorol. Climatol.* **31**: 677–688.

Horel JD, Cornejo-Garrido AG. 1986. Convection along the coast of northern Peru during 1983: spatial and temporal variation of clouds and rainfall. *Mon. Weather Rev.* **114**: 2091–2105.

Jackson IJ, Weinand H. 1995. Classification of tropical rainfall stations: a comparison of clustering techniques. *Int. J. Climatol.* **15**(9): 985–994.

Kaufman L, Rousseeuw P. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons Inc: Hoboken, NJ.

Kononenko I, Kukar M. 2007. *Machine Learning and Data Mining: Introduction to Principles and Algorithms*. Horwood Publishing: Chichester, UK.

Lagos P, Silva Y, Nickl E, Mosquera K. 2008. El Niño – related precipitation variability in Peru. *Adv. Geosci.* **14**: 231–237.

Lavado W, Espinoza JC. 2014. Impactos de El Niño y La Niña en las lluvias del Peru. *Rev. Bras. Med.* **29**: 171–182.

Lavado WS, Ronchail J, Labat D, Espinoza JC, Guyot JL. 2012. Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages. *Hydrol. Sci. J.* **57**(4): 1–18.

Mair A, Fares A. 2011. Comparison of rainfall interpolation methods in a Mountainous Region of a Tropical Island. *J. Hydrol. Eng.* **16**(4): 371–383.

Muñoz R, Garreaud R. 2005. Dynamics of the low-level Jet off the West coast of subtropical South America. *Mon. Weather Rev.* **133**: 3661–3677.

Muñoz-Diaz D, Rodrigo F. 2004. Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and principal component analysis: comparison. *Ann. Geophys.* **22**: 1435–1448.

Nickl E. 2007. *Teleconnections and Climate in the Peruvian Andes*. MSc thesis, Department of Geography, University of Delaware: Newark, DE.

Ochoa A, Pineda L, Crespo P, Willems P. 2014. Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific–Andean region of Ecuador and Peru. *Hydrol. Earth Syst. Sci.* **18**(3179–3193): 2014.

Parracho A, Melo-Gonçalves P, Rocha A. 2015. Regionalization of precipitation for the Iberian Peninsula and climate change. *Phys. Chem. Earth* , doi: 10.1016/j.pce.2015.07.004.

Ramos M. 2001. Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region. *Atmos. Res.* **57**: 123–138.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**(D14): 4407.

Raziei T, Bordi I, Pereira LS. 2008. A precipitation-based regionalization for Western Iran and regional drought variability. *Hydrol. Earth Syst. Sci.* **12**: 1309–1321.

Rousseeuw P. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**: 53–65.

Sneyers R, Vandiepenbeeck M, Vanlierde R. 1989. Principal component analysis of Belgian rainfall. *Theor. Appl. Climatol.* **204**: 199–204.

Sönmez İ, Kömüşcü AÜ. 2011. Reclassification of rainfall regions of Turkey by K-means methodology and their temporal variability in relation to North Atlantic Oscillation (NAO). *Theor. Appl. Climatol.* **106**(3–4): 499–510.

Stooksbury D, Michaels P. 1991. Cluster analysis of southeastern US climate stations. *Theor. Appl. Climatol.* **150**: 143–150.

Suarez W. 2007. *Le bassin versant du fleuve Santa (Andes du Pérou): Dynamique des écoulements en contexte glacio-pluvio-nival*. PhD thesis, Université Montpellier II, France.

Takahashi K. 2004. The atmospheric circulation associated with extreme rainfall events in Piura, Peru, during the 1997–1998 and 2002 El Niño events. *Ann. Geophys.* **22**: 3917–3926.

Takahashi K, Battisti D. 2007. Processes controlling the mean Tropical Pacific precipitation pattern. Part II: the SPCZ and the Southeast Pacific Dry Zone. *J. Clim.* **20**: 5696–5706.

Takahashi K, Dewitte B. 2015. Strong and moderate nonlinear El Niño regimes. *Clim. Dyn.*, doi: 10.1007/s00382-015-2665-3.

Takahashi K, Montecinos A, Goubanova K, Dewitte B. 2011. ENSO regimes: reinterpreting the canonical and Modoki Niño. *Geophys. Res. Lett.* **38**: L10704.

Tapley T, Waylen P. 1990. Spatial variability of annual precipitation and ENSO events in western Peru. *Hydrol. Sci. J.* **35**(4): 429–446.

Türkeş M, Sümer UM, Kılıç G. 2002. Persistence and periodicity in the precipitation series of Turkey and associations with 500 hPa geopotential heights. *Clim. Res.* **21**: 59–81.

Ünal Y, Kindap T, Karaca M. 2003. Redefining the climate zones of Turkey using cluster analysis. *Int. J. Climatol.* **23**: 1045–1055.

Vauchel P. 2005. *Hydraccess: Software for Management and Processing of Hydro–meteorological Data Software, Version 2.1.4. Free Download*. www.ore-hybam.org/index.php/eng/Logiciels/Hydraccess (accessed 1 September 2014).

Virji H. 1981. A preliminary study of summertime Tropospheric circulation patterns over South America estimated from cloud wind. *Mon. Weather Rev.* **109**: 599–610.

Vuille M, Bradley RS, Keimig F. 2000. Interannual climate variability in the Central Andes and its relation to tropical Pacific and Atlantic forcing. *J. Geophys. Res.* **105**: 12447–12460.

Waylen P, Poveda G. 2002. El Niño-Southern Oscillation and aspects for western South American hydro-climatology. *Hydrol. Process.* **16**: 1247–1260.

Woodman R. 1985. *Recurrencia del Fenómeno El Niño con intensidad comparable a la del Niño 1982–1983*. *Concejo Nacional de Ciencia y Tecnología* (CONCYTEC): Lima, Peru, 301–332.

Yang F, Lau K-M. 2004. Trend and variability of China precipitation in spring and summer: linkage to sea-surface temperature. *Int. J. Climatol.* **24**: 1625–1644.

