

Pluviometric Regionalization of Catalunya: a Compositional Data Methodology

J. GIBERGANS-BAGUENA¹, M.I. ORTEGO¹ and R. TOLOSANA-DELGADO²

¹Matemàtica Aplicada III - Universitat Politècnica de Catalunya (UPC), Spain, jose.gibergans@upc.edu

²Laboratori d'Enginyeria Marítima - Universitat Politècnica de Catalunya (UPC), Spain

Abstract

The aim of this paper is to introduce a methodology for defining groups from regionalized compositional data, through a hierarchical clustering algorithm aware of both the spatial dependence and the compositional character of the data set. This method is used to define a regionalization of Catalunya (NE Spain) with respect to its precipitation patterns in the Winter season. This region is characterized by a highly contrasted topography, which plays a dominant role in the spatial distribution of precipitation. Each rain gauge station is characterized by the relative frequencies of occurrence of six intervals of daily precipitation amount (classes ranging from “no rain” for precipitation below 3 mm, to “heavy storm” above 50 mm). Recognizing that frequencies are compositional data, the spatial dependence of this data set has been characterized by variograms of the set of all pair-wise log-ratios, in the fashion of the variation matrix. Then, a Mahalanobis distance between stations has been defined using these variograms to ensure that gauges with high spatial correlation get smaller distances. This spatially-dependent distance criterion has been used in a Ward hierarchical cluster method to define the regions. Results reveal 5 quite homogeneous groups of stations, which can be mostly ascribed a physical meaning. Finally, possible links to regional circulation patterns are discussed.

1 Introduction

Many disciplines, like hydrology, climatology, forest management or agriculture, among others, use spatial information of the climatological variables as a basis for the knowledge of the processes they study. Precipitation is one of the main climatic variables used, which distribution is highly variable in time and space. The study of precipitation characteristics in a region is very important for water resources management, drainage of urban and agricultural lands, floods, etc. Thus, an accurate regionalization of precipitation patterns is useful for the optimum design and management of water related activities.

Objective regionalizations of different climatic variables (usually temperature and rainfall) and using different techniques (multivariate analysis, neural networks, etc.), have been performed for many regions of the world: for instance Bonell and Summer (1992), Bonell (1992), White and Richman (1991), Gong and Richman (1995), Fovell and Fovell (1993), Romero et al. (1999), Unal et al. (2003), to mention a few. For the case study of Catalunya, different regionalizations using rainfall as variable have been performed, for instance by Martínez-Albaladejo (1990), Periago et al. (1991), Fernández-Mills et al. (1994), Gibergans-Báguena and Llasat (2007).

In general, methods used in most of these regionalizations have only considered the values of variables at different times and places, but do not take into account their geographical situation, and therefore, explicitly the spatial variability of the phenomenon. In particular, for the reference data set, the precipitation regime of Catalunya is highly irregular. The reasons are the complexity of the general atmospheric circulation and the rich orography (Llasat and Puigcerver, 1992; Martín-Vide, 1995; Doswell et al., 1996; Romero et al., 1997; Rigo and Llasat, 2007). Beyond spatial variability issues, data may also suffer of other sources of inaccuracy, most commonly due to rounded data, censored values and missing data. Censored data is an important problem, as rain gauges typically fail under extreme precipitation conditions.

In order to overcome these problems, daily precipitation has been split in six intervals of daily precipitation intensity, and then frequencies have been computed for each of the intervals. Frequencies may be then considered compositions, and be suitably treated with the log-ratio approach. A Ward hierarchical cluster analysis of the dataset has been carried out, based on a Mahalanobis-Aitchison distance between individual stations. In order to take into account the spatial variability, a generalized Mahalanobis distance is defined using compositional variograms and cross-variograms. All calculations in this contributions have been carried out using the functions of the package *compositions* (van den Boogaart et al., 2010; van den Boogaart and Tolosana-Delgado, 2008) of R

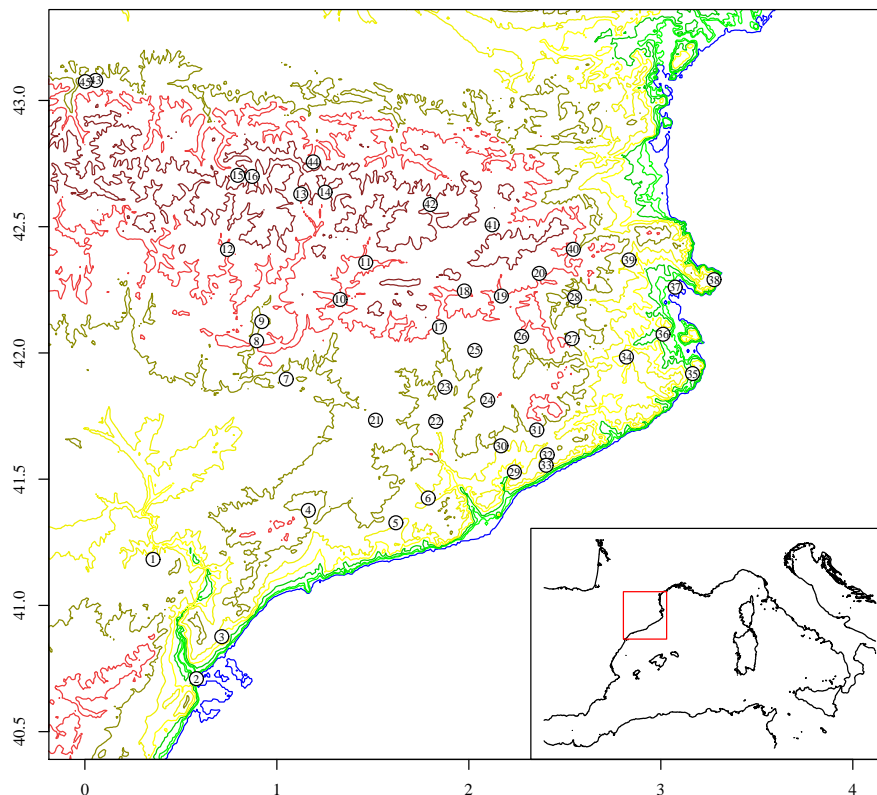


Figure 1: Orographic characteristics of the study area, and distribution of rain gauges. The lower right inset shows the location in the Mediterranean region. Gauge numbers are reported in Table 1.

A detailed description of the used dataset is provided in section 2. Cluster analysis for pluviometric regionalization is discussed in section 3. A methodology for compositional geostatistics is briefly presented in section 4. Section 5 presents the regionalization obtained for the winter season.

2 Case study

Catalunya is located on the NE corner of the Iberian Peninsula (Fig. 1) and has an area of 35000 km^2 . Its coastline runs along the NE–SW direction and its main orographic features are a complex coastal mountain system (*Serralada Litoral*, *Serralada Pre-Litoral* and *Serralada Transversal*), with some peaks exceeding 1000 m, and the Pyrenees, roughly lying along the E–W direction on the northern border of Spain. The main rivers are the Ebro with its tributaries, and the so-called *inner basins* (Llobregat, Besós, Fluvià, etc.) where most floods occur. The proximity to the Mediterranean Sea in combination with its specific orography have a determinant role in the pluviometric regime variability. The Pyrenees shelter the southern basins against Atlantic and northerly advection, but at the same time their relief enhance the development of storms in the temperate and hot months of the year affecting the northern face of the Pyrenees. The coastal chain system enhance the pluviometric effects of Mediterranean cyclogenesis along the coast, forming for moderate easterly advection a pluviometric screen between the coast and the rest of the country. The Ebro basin pluviometric regime is strongly conditioned by the precipitation deficit due to these mountain chains and the relative remoteness of the Atlantic Ocean.

The original dataset consists of daily precipitation amounts recorded at 46 rain gauges (Fig. 1, Table 1) from 1970 to 1990, provided by the *Instituto Nacional de Meteorología* (Spanish Met Office).

Due to its inaccuracies, the presence of censored data and its many missing values, the original precipitation data have been applied a series of thresholds to convert them into a categorical variable. The selected thresholds appear naturally taking into account the characteristics of its

Height	Code	Name	Height	Code	Name
357	1	La Pobla de Masaluca	327	23	Balsareny
8	2	Amposta	717	24	Moià
142	3	Perelló	707	25	Prats de Llu canès
350	4	Montblanch	555	26	Borgonyà
198	5	Castellv{1} de la Marca	370	27	Les Planes d'Hostoles
162	6	Sant Sadurn{1} d'Anoia	296	28	Castellfollit de la Roca
318	7	Artesa de Segre	96	29	Martorelles
380	8	Terradets, embassament	203	30	Caldes de Montbui
421	9	Gavet, Central	346	31	Cánoves, Can Garriga
558	10	Organyà	147	32	Dosrius
691	11	La Seu d'Urgell	88	33	Argentona
838	12	Port de Suert	70	34	Girona, Bell-Lloch
957	13	Esterra d'Àneu	64	35	Calella de Palafrugell
1052	14	Tabescan, presa	44	36	Jafre
974	15	Viella	17	37	Castelló d'Empúries
1144	16	Arties	23	38	Cadaqués
704	17	Berga	193	39	Darnius
843	18	La Pobla de Lillet	1570	40	Le Tec
738	19	Campdevàdol	1562	41	Mont-Louis
947	20	Camprodon	2047	42	Hospitalet
680	21	Calaf	628	43	Usson
238	22	Manresa	860	44	Salou
			673	45	Artigues

Table 1: Codes and names of rain-gauges locations. Height is in meters above sea level.

histogram (Fig. 2): as can be seen, too many readings are concentrated at 5 and 10 mm/24h, which suggests a sort of rounding in the manual recording process. Therefore, we have chosen six intervals of daily precipitation: no rain (below 3), weak for [3, 5), moderate weak for [5, 10), moderate for [10, 20), moderately heavy for [20, 40), and heavy (40 and over). Moreover, these intervals are equally spaced logarithmically to honor the ratio scale of precipitation (Egozcue et al., 2006; Ortego et al., 2010). A clustering of a regionalized categorial variable would require hierarchical Bayesian models, beyond the scope of this paper. Instead, we take a simplified approach by working with the frequency vectors as compositions.

3 Cluster analysis for spatially dependent data

In order to detect changes in the precipitation frequency distribution patterns, a cluster analysis will be performed. Cluster analysis (Everitt, 1993) is a multivariate technique used for grouping objects into homogeneous subgroups according to similarities. Euclidean and Mahalanobis distances are frequently used to measure the similarity between objects or individuals. Let $\mathbf{x}_n = (x_{n1}, x_{n2}, \dots, x_{np})$, $n = 1, 2, \dots, N$, be the vector of p characteristics observed for the n -th sample in a dataset with N individuals. The squared Euclidean distance is defined as

$$d_E^2(n, m) = (\mathbf{x}_n - \mathbf{x}_m)' (\mathbf{x}_n - \mathbf{x}_m) ,$$

where the prime indicates transpose vector. The Mahalanobis distance is defined as

$$d_M(n, m) = (\mathbf{x}_n - \mathbf{x}_m)' \mathbf{C}^{-1} (\mathbf{x}_n - \mathbf{x}_m) , \quad (1)$$

where \mathbf{C} is the estimated covariance matrix defined as:

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})' (\mathbf{x}_i - \bar{\mathbf{x}}) ,$$

and $\bar{\mathbf{x}}$ is the estimated vector of means. Unlike the Euclidean distance, the Mahalanobis distance takes into account any possible dependence between variables. However, this distance ignores the

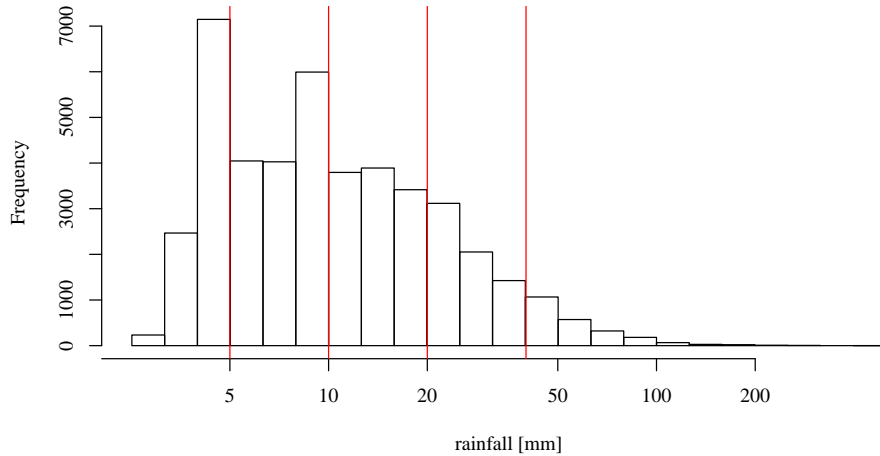


Figure 2: Daily rainfall histogram

spatial structure, which may affect the relationship between samples. The Mahalanobis distance (Eq. 1) can be then generalized replacing the covariance matrix, \mathbf{C} by any positive definite matrix. In particular, one can choose a suitable matrix that takes into account spatial dependence, following Pawlowsky-Glahn et al. (1997); Jiménez-González et al. (1998).

Geostatistical techniques model spatial data as the realizations of random functions (Matheron, 1965; Olea, 1999; Pawlowsky-Glahn and Olea, 2004). Important tools for measuring the spatial dependence among samples are the (semi)variogram and cross-(semi)variogram. Let $S = \mathbf{x}(s_1), \mathbf{x}(s_2), \dots, \mathbf{x}(s_N)$ be a regionalized data set of size N , where s_i are the spatial coordinates of the i -th sample, and $\mathbf{x}(s_i) := \mathbf{x}_i$ is a column vector of p observations at that location. Assuming some sort of stationarity of the considered variables, the (semi)variogram of the i -th variable is defined as

$$\gamma_i(h) = \frac{1}{2} E [(x_i(s) - x_i(s+h))^2].$$

Alternatively, the auto-covariance function of variable i may be defined as

$$C_i(h) = Cov [x_i(s), x_i(s+h)].$$

These two functions are symmetric on h and satisfy that $\gamma_i(h) = C_i(0) - C_i(h)$.

The definition of variogram can be extended to the bivariate case. The cross-variogram reflects the covariance of the increments of two variables between two locations placed a vector h apart,

$$\gamma_{ij}(h) = Cov [x_i(z) - x_i(z+h), x_j(z) - x_j(z+h)].$$

Cross-covariances are also easy to define as functions $C_{ij}(h)$ giving the covariance between $x_i(z)$ and $x_j(z+h)$. Note that in general $C_{ij}(h) \neq C_{ji}(h)$; instead, it is always true that $C_{ij}(h) = C_{ji}(-h)$. Moreover

$$2\mathbf{\Gamma}(h) = 2\mathbf{C}(0) - (\mathbf{C}(h) + \mathbf{C}^t(h))$$

where $\mathbf{\Gamma}(h)$ (and $\mathbf{C}(h)$) is a matrix-valued function with variograms (autocovariances) in the main diagonal and cross-variograms (cross-covariances) in the off-diagonal elements. For $h \neq 0$, $\mathbf{C}(h)$ describes the spatial dependence among the considered variables, whereas at $h = 0$ it reflects the classical covariance between them. Under second-order stationarity hypothesis, it is verified that $\lim_{h \rightarrow \infty} \mathbf{\Gamma}(h) = \mathbf{C}(0)$ and $\lim_{h \rightarrow \infty} \mathbf{C}(h) = \mathbf{0}$.

From Jiménez-González et al. (1998), the following generalization of the Mahalanobis distance for data with spatial dependence is defined:

$$d(n, m) = (x(z_n) - x(z_m))' [2\mathbf{C}(0) - \mathbf{\Gamma}(h)]^{-1} (x(z_n) - x(z_m)) . \quad (2)$$

This distance has the effect of reducing the Mahalanobis distance for spatially *dependent* samples. This can be seen by considering two extreme cases:

- taking h “large enough”, samples may be thought to be spatially independent; the metric matrix becomes then $[2\mathbf{C}(0) - \mathbf{\Gamma}(h)]^{-1} \rightarrow \mathbf{C}(0)^{-1}$ and the proposed Mahalanobis distance converges to the classical one;
- taking $h = 0$, the variogram $\mathbf{\Gamma}(0) = \mathbf{0}$ and the metric matrix is then $[2\mathbf{C}(0)]^{-1} = 1/2\mathbf{C}(0)^{-1}$, and the resulting Mahalanobis distance is half of the classical one.

The cluster technique used in this work is Ward’s method (Ward, 1963). This method produces a hierarchical clustering, i.e. initially each object is assumed to form a separate group, and then objects or groups close to one another are successively merged. Hierarchical methods usually produce a graphical output known as a dendrogram that shows this hierarchy and indicates at which level of similarity any two clusters were merged. The choice of a suitable number of clusters is a subjective task.

4 Variograms for compositional data

Spatial dependence has been already introduced into the cluster methodology through the introduction of spatial dependence in the Mahalanobis distance (Eq. 2). In order to introduce the compositional characteristics of our input data, the Mahalanobis distance is computed using the *ilr* coordinates of the vectors of frequencies $\mathbf{z}(s_n) = \mathbf{z}_n = (z_{n1}, z_{n2}, \dots, z_{nD})$, i.e. by means of a matrix of contrasts \mathbf{V} ,

$$\mathbf{x}_n = \mathbf{V}^t \cdot \log(\mathbf{z}_n) = \text{ilr}(\mathbf{z}_n).$$

As usual, \mathbf{V} columns sum up to zero (hence they are contrast) and these columns form a system of $D - 1$ orthonormal vectors, thus $\mathbf{V}^t \cdot \mathbf{V} = \mathbf{I}$ and $\mathbf{z}_n = \mathcal{C} \exp(\mathbf{V} \cdot \mathbf{x}_n) = \text{ilr}^{-1}(\mathbf{x}_n)$.

Pawlowsky-Glahn and Olea (2004) present different specifications of the covariance/variogram structure of a regionalized composition. We nevertheless follow Tolosana-Delgado et al. (2011) idea of using the intrinsic variation matrix (also known as *lr*-variograms or variation-variograms, $\mathbf{\Psi} = (\psi_{ij})$), a generalization of the concept of variation matrix. The variation-variograms are defined as the matrix of direct variograms of all pairwise log-ratios of any two variables (i, j), and can be estimated from a regionalized compositional data set by

$$\hat{\gamma}_{ij}(h) = \frac{1}{2N(h)} \sum_{n,m \in N(h)} \left(\log \frac{z_{ni}}{z_{nj}} - \log \frac{z_{mi}}{z_{mj}} \right)^2.$$

This matrix contains the same information as *ilr* variograms, but to estimate them we only use two parts at a time. This reduces the influence of a bad frequency estimation of all parts, a sensible issue given the small number of large precipitation storms observed. The fitting of the variogram model may be performed using variation-variograms with respect to a relative scale, so that the behavior at the origin is better captured. That implies choosing a parametric correlogram model $\rho(h, \theta)$ depending on a parameter θ , and optimizing a logarithmic measure of discrepancy between the empirical variograms $\hat{\mathbf{\Psi}}(h)$ and the model

$$\mathbf{\Psi}(h|\theta, \mathbf{A}, \mathbf{B}) = \mathbf{A}\delta(h \neq 0) + \mathbf{B}(1 - \rho(h, \theta))$$

under the restriction that both matrices \mathbf{A} (nugget effect) and \mathbf{B} (variogram sill) are negative semi-definite. According to (Tolosana-Delgado et al., 2011), this is necessary in order to ensure that the model $\mathbf{\Psi}(h|\theta, \dots)$ is compatible with a covariance function model for the *ilr*-transformed compositions. This *ilr*-model can be obtained simply as

$$C(h|\theta, \dots) = -\frac{1}{2}\mathbf{V} \cdot \mathbf{A} \cdot \mathbf{V}^t \delta(h \neq 0) + \mathbf{V} \cdot \mathbf{B} \cdot \mathbf{V}^t (1 - \rho(h, \theta))$$

This will be the model used to compute spatially-dependent Mahalanobis distances with Eq. (2).

5 Results

The winter season (December-January-February) is used for illustration. After summer, winter is the season with the lowest precipitation, though these are very important in the water cycle, providing a valuable storage for dry periods of spring and summer. Catalunya climate is characterized by the high irregularity of precipitation, in both spatial and temporal scale. In particular, winter is the most irregular season of the year (Martín-Vide and López-Bustins, 2006).

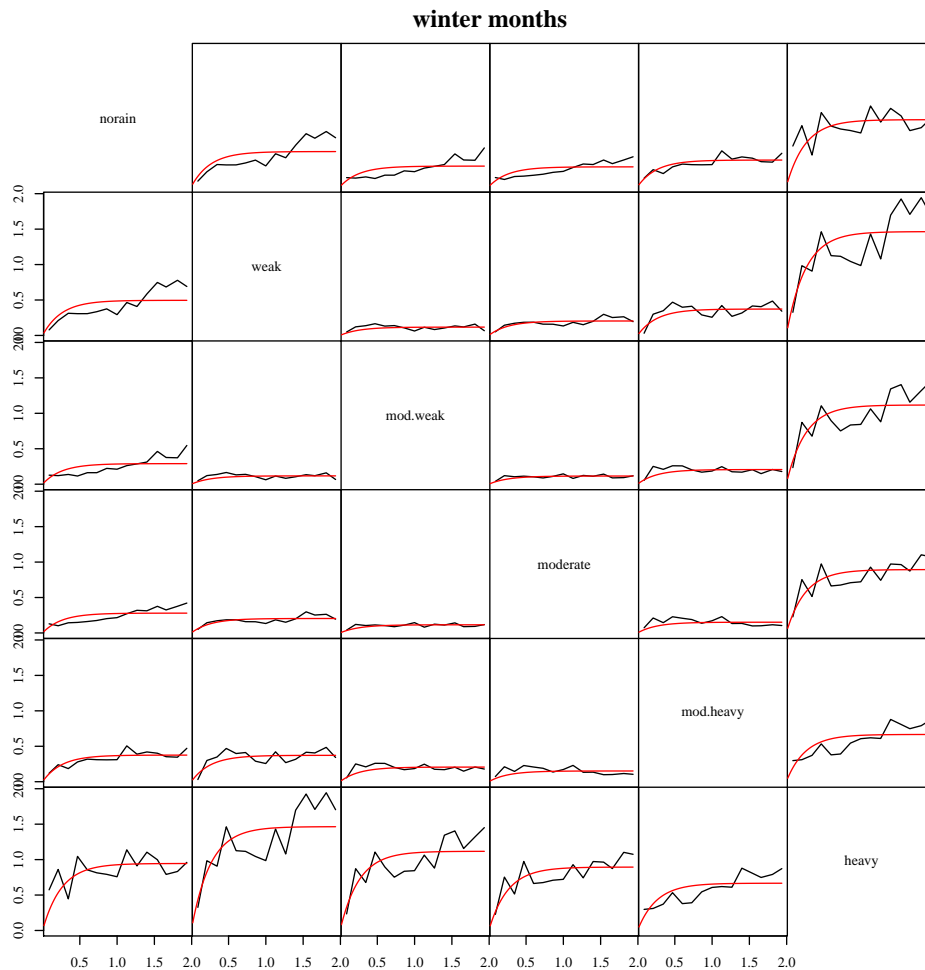


Figure 3: Variograms

To characterize the compositional variography, all pair-wise log-ratio variograms have been estimated with standard methods. These have been then fitted an exponential variogram model, forcing the corresponding sill matrix to be negative semi-definite (as necessary). This was automatically done using geostatistical functionalities of the package “compositions”. Results of the estimation and fitting can be seen in Fig. 3.

Then, a hierarchical clustering has been performed following section 3. We can see in Fig. 4 the stations cluster into five groups. The geographical representation of this clustering is shown in Fig. 5, where colors indicate different groups:

- *green*: stations on W Catalunya. The southward retreat of the subtropical high pressure zones facilitates the arrival to Catalunya of unstable Atlantic air masses that cause these precipitations. However, having crossed the whole Iberian Peninsula and its mountain ranges, they arrive too weak and produce light precipitation.
- *orange*: Winter, after summer, is the season with less amount of rain in the coastal zone. This low precipitation is mainly due to the presence of a low pressure area on the Iberian Peninsula which favors circulation from the E over the Catalan coast. Since during this season the sea temperature is not high, there is no important contribution of water vapor and, consequently, there are no heavy rainfalls from convective origin (as happens in autumn).
- *blue*: stations placed in the central basins, these are affected by situations of E and mostly SW winds, where the mass of air is channelled through the valley of the Llobregat river. Precipitation occurs due to the rise of these air masses over the mountains of the Serralada Pre-litoral.
- *magenta*: similar to the previous case, but now the stations are located at the NE basins of Catalunya. These basins are affected by E storms. Air masses enter through the valleys

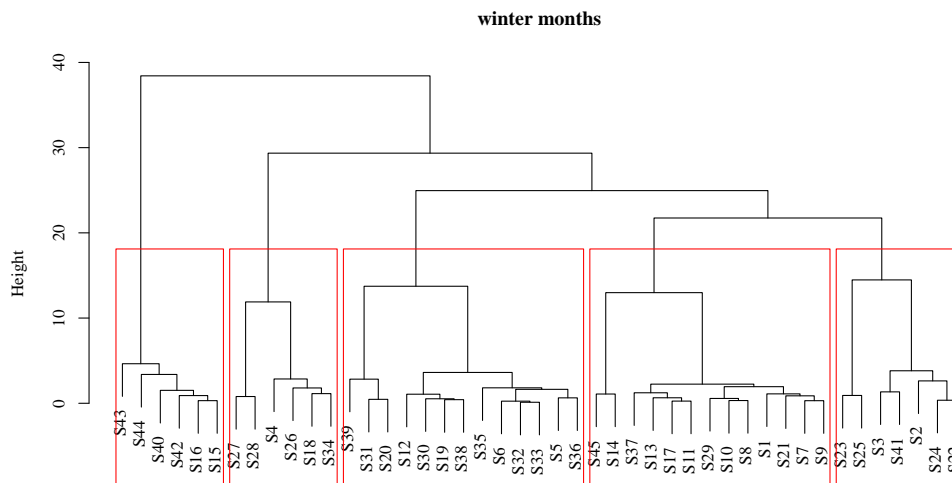


Figure 4: Winter Dendrogram

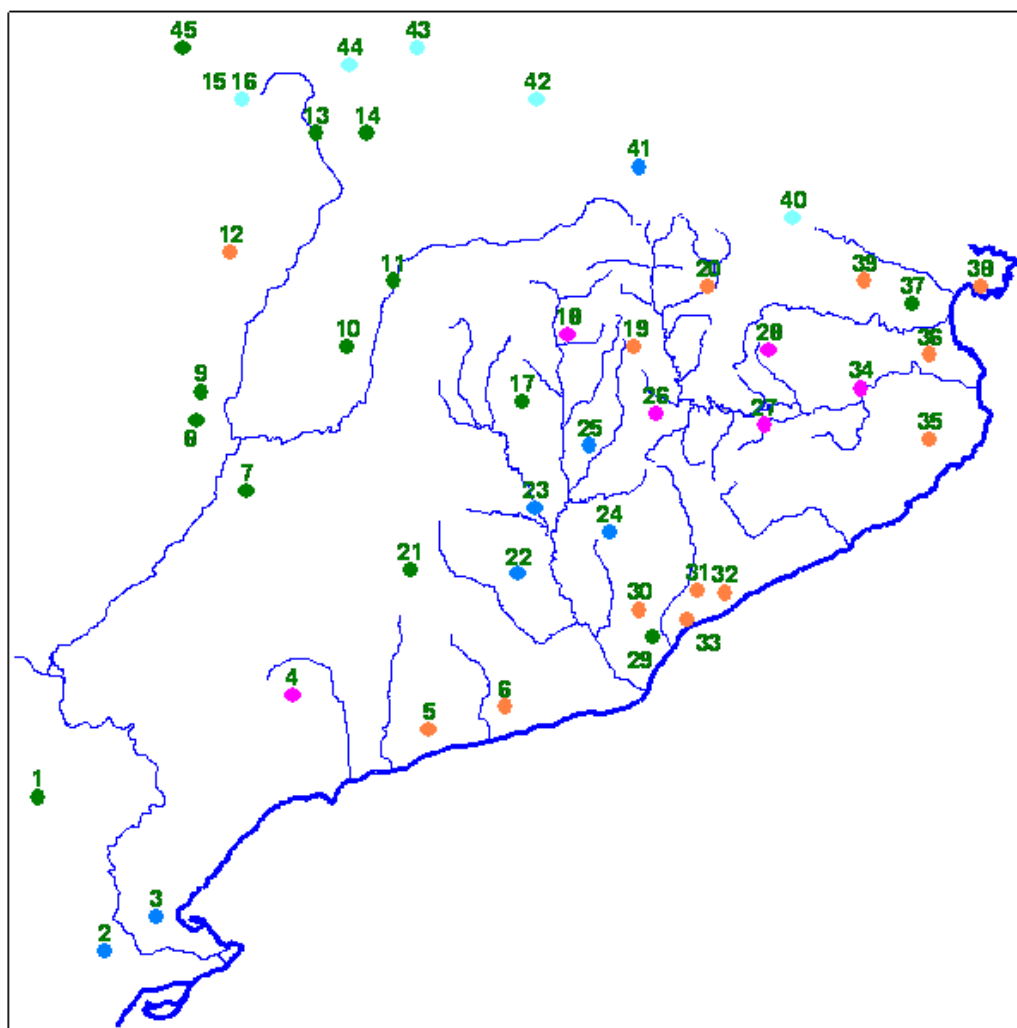


Figure 5: Winter regionalization

of rivers like Ter and Fluviá, producing rainfall by ascension of the water masses over the mountains of the *Serralada Transversal*.

- *cyan*: Corresponds to the stations located in the north face of the Pyrenees. These are open to the NW and rather sheltered from S and SW, facilitating the passage of frontal systems and low pressure areas from the Atlantic ocean that follow the general West-East circulation of these latitudes. This causes an Atlantic climate in contrast to the typical Mediterranean climate of Catalunya.

Acknowledgements

This research has been partially funded by the Spanish Ministry of Education and Science under project 'CODA-RSS' Ref. MTM2009-13272; and by the *Agència de Gestió d'Ajuts Universitaris i de Recerca* of the *Generalitat de Catalunya* under the project Ref: 2009SGR424. The third author acknowledges also funding within the "Juan de la Cierva" program of the Spanish Ministry of Education and Science (ref. JCI-2008-1835).

References

- Bonell, M. (1992). Atmospheric circulation and daily precipitation in wales. *Theoretical and Applied Climatology* 46, 3–25.
- Bonell, M. and G. Summer (1992). Autumn and winter daily precipitation areas in Wales, 1982-1983 to 1986-1987. *International Journal of Climatology* 12, 77–102.
- Doswell, C., H. Brooks, and R. Maddox (1996). Flash flood forecasting: an ingredients-based methodology. *Weather Forecast* 11, 560–581.
- Egozcue, J. J., V. Pawlowsky-Glahn, M. I. Ortego, and R. Tolosana Delgado (2006). The effect of scale in daily precipitation hazard assessment. *Natural Hazards and Earth System Science* 6, 459–470.
- Everitt, B. S. (1993). *Cluster Analysis*. Edward Arnold, Cambridge (UK). 170 p.
- Fernández-Mills, G., X. Lana, and C. Serra (1994). Catalanian precipitation patterns: Principal component analysis and automated regionalisation. *Theoretical and Applied Climatology* 49, 201–212.
- Fovell, R. G. and M. Fovell (1993). Climate zones of the conterminous United States defined using cluster analysis. *Journal of Climate* 6, 2103–2135.
- Gibergans-Báguena, J. and M. Llasat (2007). Improvement of the analog forecasting method by using local thermodynamic data. Application to autumn precipitation in Catalonia. *Atmospheric Research* 86, 173–193.
- Gong, X. and M. B. Richman (1995). On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Journal of Climate* 8, 897–931.
- Jiménez-González, S., V. Pawlowsky-Glahn, J. J. Egozcue, and R. Ramírez-Llorente (1998). Ward's method adapted to spatially dependent data. In A. Buccianti, G. Nardi, and R. Potenza (Eds.), *Proceedings of IAMG'98, The Fourth Annual Conference of the International Association for Mathematical Geology*, Volume 1, Napoli, Italia, pp. 445–450. De Frede Editore.
- Llasat, M. and M. Puigcerver (1992). Pluies extremes en Catalogne: Influence orographique et caractéristiques synoptiques (extreme rainfall in Catalonia: orographical influence and synoptic characteristics). *Hydrologie Continentale* 2, 99–115.
- Martín-Vide, J. (1995). La pluviometría en Catalunya. *Butlletí de les Societats Catalanes de Ciències Físiques, Químiques i Matemàtiques : segona època XV*, 135–141.
- Martín-Vide, J. and J.-A. López-Bustins (2006). The Western Mediterranean oscillation and rainfall in the Iberian Peninsula. *International Journal of Climatology* 26, 1455–1475.

- Martínez-Albaladejo, A. (1990). Les comarques pluviomètriques de la conca dels Pirineus Orientals. una aplicació de l'anàlisi factorial. *Revista Catalana de Geografia* 12(V), 55–71.
- Matheron, G. (1965). *Les variables régionalisées et leur estimation—une application de la théorie des fonctions aléatoires aux sciences de la nature*. Masson et Cie., Paris (F). 305 p.
- Olea, R. A. (1999). *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers, Norwell, MA (USA). 303 p.
- Ortego, M. I., J. Gibergans-Báguena, R. Tolosana-Delgado, J. J. Egozcue, and M. C. Llasat (2010). Bayesian trend analysis for daily rainfall series of Barcelona. *Adv. in Geosciences*. 26: 71-76.
- Pawlowsky-Glahn, V. and R. Olea (2004). *Geostatistical Analysis of Compositional Data*. London, G.B.: Oxford University Press.
- Pawlowsky-Glahn, V., G. Simarro, and J. Martín-Fernández (1997). Spatial cluster analysis using a generalised mahalanobis distance. In V. Pawlowsky-Glahn (Ed.), *Proceedings of IAMG'97, The Third Annual Conference of the International Association for Mathematical Geology*, Barcelona, Spain, pp. 175–180. CIMNE (Centro Internacional de Métodos Numéricos en Ingeniería).
- Periago, M., X. Lana, C. Serra, and G. Fernández-Mills (1991). Precipitation regionalization: An application using a meteorological network in Catalonia (NE Spain). *International Journal of Climatology* 11, 529–543.
- Rigo, T. and M. Llasat (2007). Analysis of mesoscale convective systems in Catalonia (NE of Spain) using radar for the period 1996-2000. *Atmospheric Research* 83, 458–472.
- Romero, R., C. Ramis, and S. Alonso (1997). Numerical simulation of an extreme rainfall event in Catalonia: role of orography and evaporation from the sea. *Quarterly Journal of the Royal Meteorological Society* 123, 537–559.
- Romero, R., C. Ramis, J. A. Guijarro, and G. Sumner (1999). Daily rainfall affinity areas in Mediterranean Spain. *International Journal of Climatology* 19, 557–578.
- Tolosana-Delgado, R., K. G. van den Boogaart, and V. Pawlowsky-Glahn (2011). Geostatistics for compositions. In: Pawlowsky-Glahn, V. and A. Buccianti (eds), *Compositional Data Analysis at the beginning of the XXI century*, Wiley. In press
- Unal, Y., T. Kindap, and M. Karaca (2003). Redefining the climate zones of Turkey using cluster analysis. *International Journal of Climatology* 23(9), 1045–1055.
- van den Boogaart, K. G. and R. Tolosana-Delgado (2008). “compositions”: a unified R package to analyze Compositional Data. *Computers and Geosciences* 34(4), 320–338.
- van den Boogaart, K. G., R. Tolosana-Delgado, and M. Bren (2010). *compositions: Compositional Data Analysis*. R package version 1.10-1.
- Ward, Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58, 236–244.
- White, D. and M. Richman (1991). Climate regionalization and rotation of principal components. *International Journal of Climatology* 11(9), 1–25.