Theor. Appl. Climatol. 50, 169–183 (1995)

Theoretical and Applied Climatology

551.577.2(46)

Departament de Física i Enginyeria Nuclear, ETSEIB, Universitat Politècnica de Catalunya, Barcelona, Spain

Principal Component Analysis of Precipitation and Rainfall Regionalization in Spain

G. Fernández Mills

With 7 Figures

Received June 26, 1994
Revised September 19, 1994

Summary

A T-mode Principal Component Analysis (PCA) based on a network including 68 pluviometric gauges and their 12 mean monthly amounts of rainfall is attempted in order to describe the main patterns governing precipitation in Spain. The procedure is applied to a 12×12 intermonth covariance matrix; the unrotated components and two additional solutions deduced after varimax and oblimin rotations are presented and discussed. In all cases component scores are computed and their spatial distribution is discussed. Two regionalizations of Spanish rainfall are then obtained and compared in terms of group homogeneity. There follows a discussion concerning the main pluviometric characteristics of each region deduced from the best division.

1. Introduction and Data Base

The orographic complexity of the Iberian Peninsula and its mid-latitude situation between the Atlantic Ocean and the Mediterranean Sea justify detailed studies concerning spatial rainfall distribution. As will be seen, very different precipitation regimes coexist in a relatively small area, including gauges where mean annual precipitation reaches 2000 mm and regions with annual records smaller than 200 mm (Table 1).

From a general point of view, Autumn, Winter and Spring rainfall in Spain can be related to baroclinic disturbances linked to the southern winter jet (Lawson, 1989). These may originate from the Atlantic Ocean, tracking eastwards between Spain and north Africa, or develop on trailing cold fronts that have moved down from northern Europe. The contribution to precipitation of African desert depressions (Prezerakos, 1990; Prezerakos et al., 1990) is remarkable in the south eastern part of the country, and disturbances generated in the western Mediterranean basin itself (Catalano-Balearic cyclogenesis) (Ramis et al., 1984; Jansà, 1978) can lead to Eastern advections carrying warm and humid air masses that produce important records in eastern sided areas. By contrast, westerlies and atlantic depressions with northern tracks tend to affect the northwest and the north of Spain, which have a greater exposure to western and northwestern flows. Atlantic air masses lose their humidity mainly in these regions, and reach the Mediterranean coast without significant activity. In summer, rainfall is only relevant in the mountainous areas of the North. It is very scarce in the South where June, July, August and September monthly amounts are extremely low and can be compared with North African summer records (Maheras, 1985).

The objectives of the present study are to analyze the patterns governing the pluviometric regime and to identify regions of coherent precipitation. In the first stage, a T-mode PCA (Daultrey, 1976; Harman, 1976; Richman, 1986 and Preisendorfer, 1988; among others) is applied to a 12×12 intermonth covariance matrix deduced from the 12 mean monthly precipitation of 68 synoptic

Table 1. *The Stations of the Synoptic Raingauge Network with Their Altitude and Their Mean Annual Rainfall*

Number	Rain Gauge	Annual Rainfall (mm)	Altitude (m)
1	Toledo	366	551
2	Cáceres	467	453
3	Ciudad Real	462	601
4	Huelva	483	10
5	Córdoba	618	107
6	Jaén	588	503
7	Cádiz	589	6
8	Granada	392	710
9	Ceuta	537	206
10	Tarifa	716	46
11	Málaga	451	12
12	Almería	207	6
13	Melilla	356	35
14	Vitoria	858	452
15	Logroño	431	450
16	Pamplona	715	440
17	Huesca	579	503
18	Girona	787	95
19	Montseny	985	1712
20	Tarragona	473	25
21	Cuenca	549	955
22	Teruel	383	915
23	Castellón	426	47
24	Valencia	450	12
25	Albacete	365	699
26	Murcia	305	110
27	San Javier	376	2
28	Ibiza	408	12
29	Gando	124	4
30	Arrecife	124	6
31	Los Rodeos	678	500
32	Izaña	369	2000
33	Lérida	364	180
34	Barcelona	606	12
35	Tortosa	561	48
36	Sta Cruz Tenerife	231	8
37	La Coruña	969	14
38	Mahón	640	47
39	Palma de Mallorca	443	10
40	Valladolid	417	690
41	Madrid	474	642
42	Badajoz	523	200
43	Sevilla	573	20
44	Zaragoza	312	200
45	Alicante	323	8
46	Labacolla	1913	260
47	Pollensa	813	123
48	Burgos	638	861
49	Vigo	2030	44
50	Lugo	1088	424
51	Pontevedra	1556	68
52	Orense	817	148
53	Ponferrada	603	540
54	Gijón	986	22
55	Oviedo	1009	248
56	Santander	1202	10
57	Donosti	1621	210
58	Bilbao	1300	12
59	León	583	910
60	Palencia	392	738
61	Zamora	403	660
62	Soria	576	1079
63	Salamanca	436	782
64	Avila	327	1131
65	Segovia	442	1015
66	Navacerrada	1469	1800
67	Guadalajara	444	694
68	Molina de A.	506	1073

stations of the Instituto Nacional de Meteorologia (INM). This analysis is followed by the interpretation of the intensity of the patterns at different months in terms of the dominant synoptic situations. The monthly averages were computed using the records of the standard period 1961–1990. The situation of the (gauges) is shown in Fig. 1, and Table 1 also includes the altitude of the stations considered. Two rotation methods (Varimax and Oblimin) are then performed. The first one preserves the orthogonality of the rotated components and, by contrast, the oblimin algorithm leads to oblique solutions with significant correlations between the resulting factors.

The second step will consist of an application of a Davis and Kalkstein clustering procedure (Davis and Kalkstein, 1990; Lana and Fernández Mills, 1994; Fernández Mills et al., 1994) leading to the identification of regions of coherent precipitation. Component scores are then considered as elements characterizing each raingauge, and allow the building of such a regionalization. Similar studies related to rainfall regionalization of some parts of the Spanish area are those of Periago et al. (1991), Summer et al. (1993) and Fernández Mills et al. (1994, in press). However, regionalizations of precipitation in the whole country including discussions about spatial patterns of this variable are not present in the literature. This is the main reason why this subject is presented and discussed in this paper.

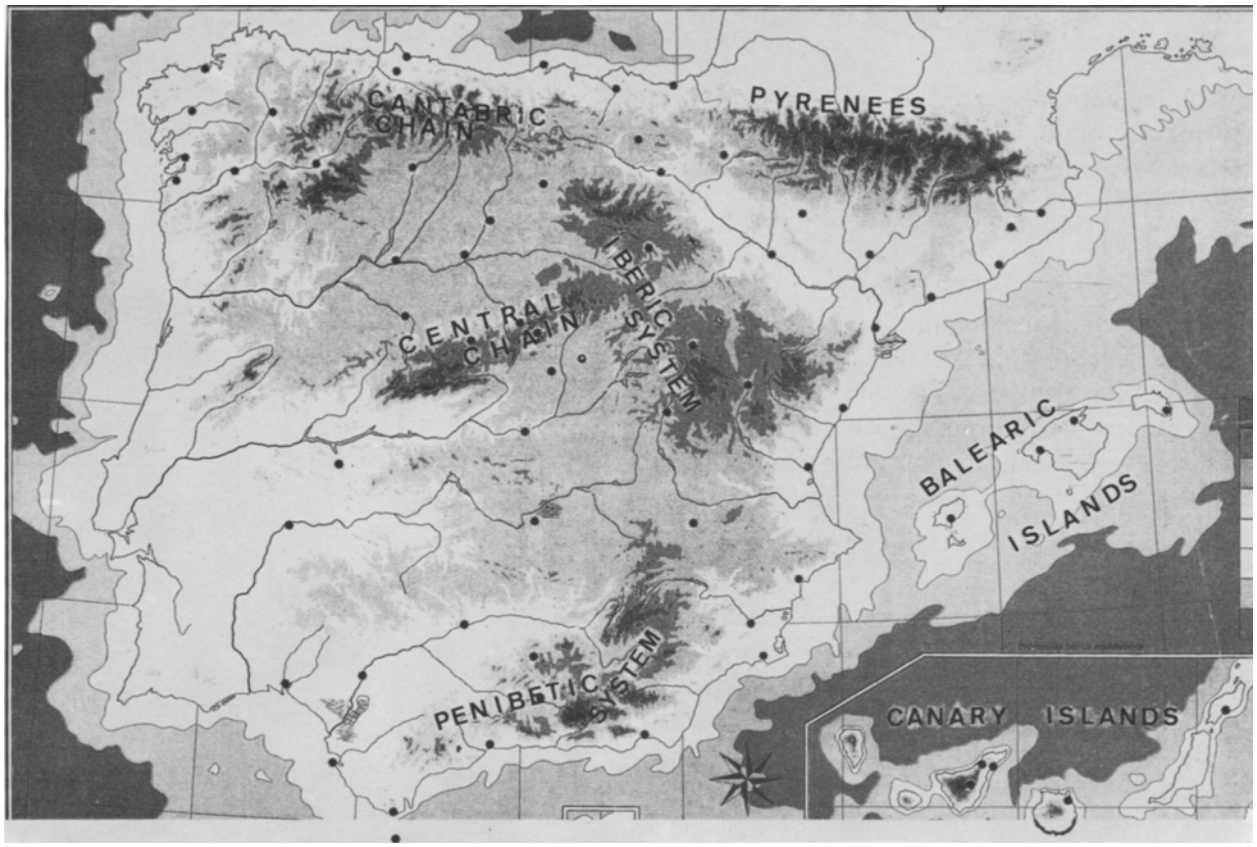


Fig. 1. Spanish relief and location of the 68 raingauges of the synoptic network

2. PCA Results

PCA can be applied to either correlation or covariance matrices. If the problem considered is univariate as in our application, the algorithm can start from both types of similarity matrices. Nevertheless, several authors have suggested (Wilmott, 1978; Legates, 1991) that covariance is a better measure of similarity in univariate works because it contains all of the advantages inherent to correlation coefficients, and in addition it preserves the metric and retains more information than correlation. Following these considerations, the component extraction was done from the inter-month covariance matrix. Consequently, a T-mode PCA is performed as this is the natural way to proceed to get a covariance matrix of maximum rank if the number of locations (stations) exceeds the number of times (months).

Let $x_{i\mu}$ be the observation at the location indicated by the index i with $1 \leqslant i \leqslant N$ and at the time indicated by the index μ with $1 \leqslant \mu \leqslant M$. An element $c_{\mu\tau}$ of the covariance matrix considered can be calculated by:

$$c_{\mu\tau} = \sum x_{i\mu} x_{i\tau} \tag{1}$$

The determination of the eigenvectors of this matrix leads finally to the T-mode Principal Components (PC). Each variable (month) can be expressed as a linear combination of orthogonal components. A high loading of one PC at a certain time (month) and low loadings of the others means, that at this time, the situation is explained by one score pattern, and the corresponding variable (month) shows a simple structure. By contrast, if the loadings of all the components are similar, the variable behaviour is really a linear combination of patterns leading to a complex structure. As it was previously mentioned in the introduction, detailed descriptions about PCA can be extensively found in the literature.

The starting data matrix includes 68 cases (stations) characterized by 12 variables (12 mean

monthly rainfall). Three types of solutions have been analyzed:

- Unrotated components.
- Varimax orthogonally rotated components
- Oblimin obliquely rotated components.

Sometimes, unrotated components exhibit some characteristics that hinder the isolation of individual modes of variation (Richman, 1986; White et al., 1991). The interpretation of these modes is easier and clearer if rotations are performed. Consequently, two well known rotation algorithms have been applied in order to achieve a simpler structure in component space leading to meaningful patterns.

2.1 *Unrotated Solution*

Table 2a summarizes eigenvalues and the cumulative proportion of explained variance related to the significant unrotated principal components. We must emphasize that a solution including 3 components explains 96% of total variance. Table 2b shows unrotated component loadings (in mm). Pairwise plots of the deduced components (Fig. 2a–c) show the existence of an important

Table 2a. *Eigenvalues and Cumulative Proportion of Explained Variance Corresponding to Three Unrotated Principal Components*

Component	Variance Explained Eigenvalues	Cumulative Proportion in Data Space (%)
1	14401	83.22
2	1732	93.23
3	493	96.08

Table 2b. *Unrotated Component Loadings (in mm)*

Variable	Component 1	Component 2	Component 3
January	52.9	−12.5	3.3
February	47.1	−11.3	6.7
March	38.2	−6.4	6.0
April	26.4	5.5	2.7
May	25.0	11.9	9.1
June	14.6	13.8	8.6
July	10.0	11.8	2.3
August	13.4	18.3	−2.9
September	23.5	20.7	−1.7
October	34.7	9.9	−9.2
November	43.2	−2.4	−8.8
December	49.9	−5.9	−7.7

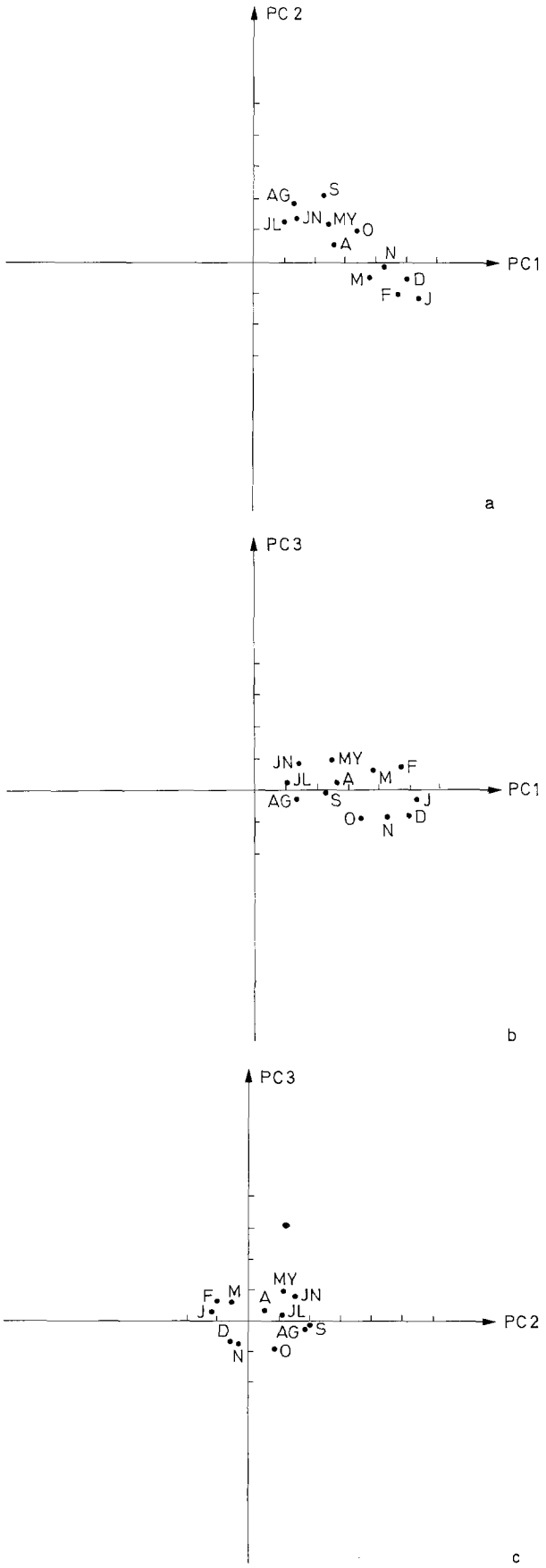


Fig. 2a, b, c. Pairwise plots of the unrotated principal components

number of complex variables loading highly in more than one component. This is specially true when summer rainfall is considered, and will hamper factor interpretation. Figure 3a represents the spatial distribution of component 1 scores. Component 1 loads highly in winter precipitation and poorly in summer rainfall. A close look at Fig. 3a allows us to conclude that maximum scores correspond to stations located at the N and NW of the country, where the rainfall regime is mainly governed by westerlies and Atlantic perturbations. In these stations, maximum rainfall is recorded in Winter and a deep minimum exists in July and August, when Atlantic depressions and frontal waves are deviated to higher latitudes. Nevertheless, there are no dry months and January and December totals frequently reach 350 mm. Scores are negative in the rest of the country excepting western Andalusia, the Northern areas of Catalonia and the Balearic Islands and some mountains of the Central Chain with relevant winter amounts. These regions are exposed to disturbances linked to the southern winter jet. The western lands of Andalusia and the Central Chain are often affected by cyclonic SW flows caused by atlantic depressions which have southern tracks in winter. This synoptic situation produces important showers in the SW of the Iberian Peninsula and on the southern sided ranges of the Central Chain. Positive scores corresponding to the North of Catalonia and the North of Balearic Islands can be associated with the effects of trailing cold fronts which often lead to cyclogenesis in the Lion Gulf in winter. Consequently, the resultant northern surface flows tend to produce rainfall episodes in the vicinity of the Pyrenees and in Menorca and the northern side of the mountains of Mallorca. Negative PC1 scores are linked to regions where maximum rainfall usually appears in Autumn or Spring as a consequence of eastern advections produced by Mediterranean depressions in the first case. Spring maxima associated with inland stations are due to other factors that will be discussed later on.

Component 2 higher loadings correspond to rainfall in September, August, June, May and July. By contrast, negative loadings are associated with winter precipitation. High scores (Fig. 3b) are present in NE areas and in the eastern part of the Cantabric regions close to the western Pyrenees, where monthly precipitation is relevant in the hot months. In fact, Catalan stations (NE Spain) often

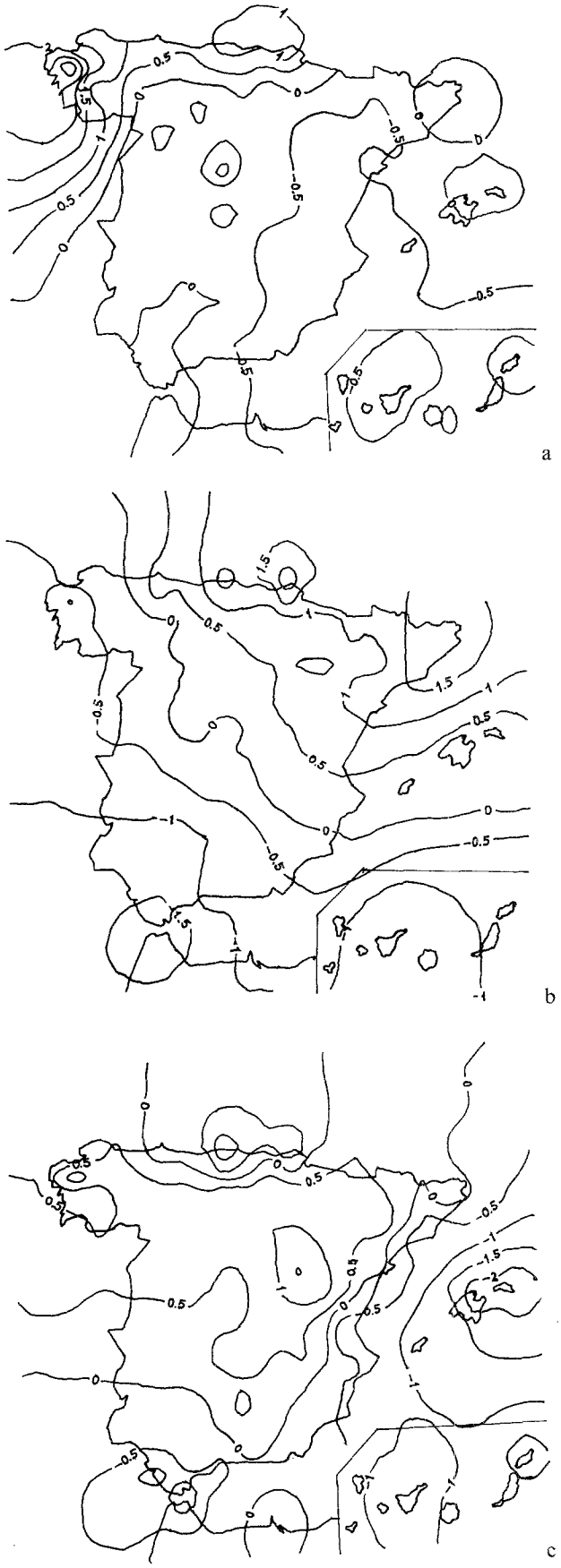


Fig. 3a, b, c. Spatial distribution of unrotated component scores (PC1: 3a, PC2: 3b, PC3: 3c)

reach their maximum monthly values in September when Mediterranean thunderstorms occasionally cause important damage over the coast and in the pre-littoral ranges. On the other hand, high negative scores are located in the South of Iberian Peninsula and over the Canary Islands, where summer rainfall is extremely scarce or non-existent and the number of dry months is significant. This behaviour is also typical in Mediterranean dry climates of North African countries which are far away from the summer tracks of frontal waves and Atlantic depressions.

Component 3 loads highly in May, June, February and March rainfall and negatively in October, November and December. Consequently, high scores are related to inland stations with relevant peaks of rainfall activity in late winter, spring and June (Fig. 3c). The thermal contrast between the continent and the Atlantic sea surface temperature is maximum in this period and can explain an increase in frontal activity in the inner regions. Negative component values correspond to gauges close to the Mediterranean coast recording important autumn amounts with maxima in October or November, where cyclonic situations and eastern advections are common in the western mediterranean basin.

2.2 *Varimax Rotated Solution*

The agreement between rotated and unrotated solutions can be measured if the congruence coefficient (Harman, 1976) is computed:

$$g_{ab} = \sum(b_{ja}b_{jb})/((\sum b_{ja}^2)(\sum b_{jb}^2))^{0.5} \quad (2)$$

where b_a is a loading from eigenvector a of one solution, and b_b is a loading from eigenvector b of a second solution. The summation occurs for n elements of the eigenvector. Consequently, in this study n equals 12, the number of variables. Table3a–c show these coefficients when the unrotated and the varimax solutions are compared (3a), when the unrotated and the oblimin components are considered (3b) and finally when the two rotated solutions are tested (3c). According to specified ranges of the congruence coefficient (Richman, 1986) there is a perfect agreement between unrotated PC1 and varimax rotated PC1, and a good inverse agreement between unrotated PC3 and varimax rotated PC3. In the latter case, a change in sign is detected in the loading field when varimax rotation is performed.

Table 3a. *Harman Congruence Coefficients Between Unrotated and Varimax Rotated PC*

		Varimax PC1	Rotated PC2	PC3
Unrotated	PC1	0.99	0.81	0.23
	PC2	−0.17	0.58	0.23
	PC3	0.00	0.05	−0.94

Table 3b

		Oblimin PC1	Rotated PC2	PC3
Unrotated	PC1	0.96	0.39	0.10
	PC2	−0.29	0.91	0.30
	PC3	−0.04	0.16	−0.95

Table 3c

		Oblimin PC1	Rotated PC2	PC3
Varimax	Rot	0.99	0.24	0.06
	ated	0.61	0.85	0.21
		0.19	0.15	0.99

This is not a shortcoming, as loadings signs are arbitrary. By contrast, the "goodness of match" attached to unrotated and varimax rotated PC2 is very poor (0.58). In fact, rotated PC2 shows a better agreement with unrotated PC1 (0.81). Consequently, differences in the physical meaning of the former component can be expected. Figure 4a–c show pairwise plots of varimax rotated significant components. If the goal of rotation is to achieve a simpler structure, the results obtained are disappointing. A relevant number of variables

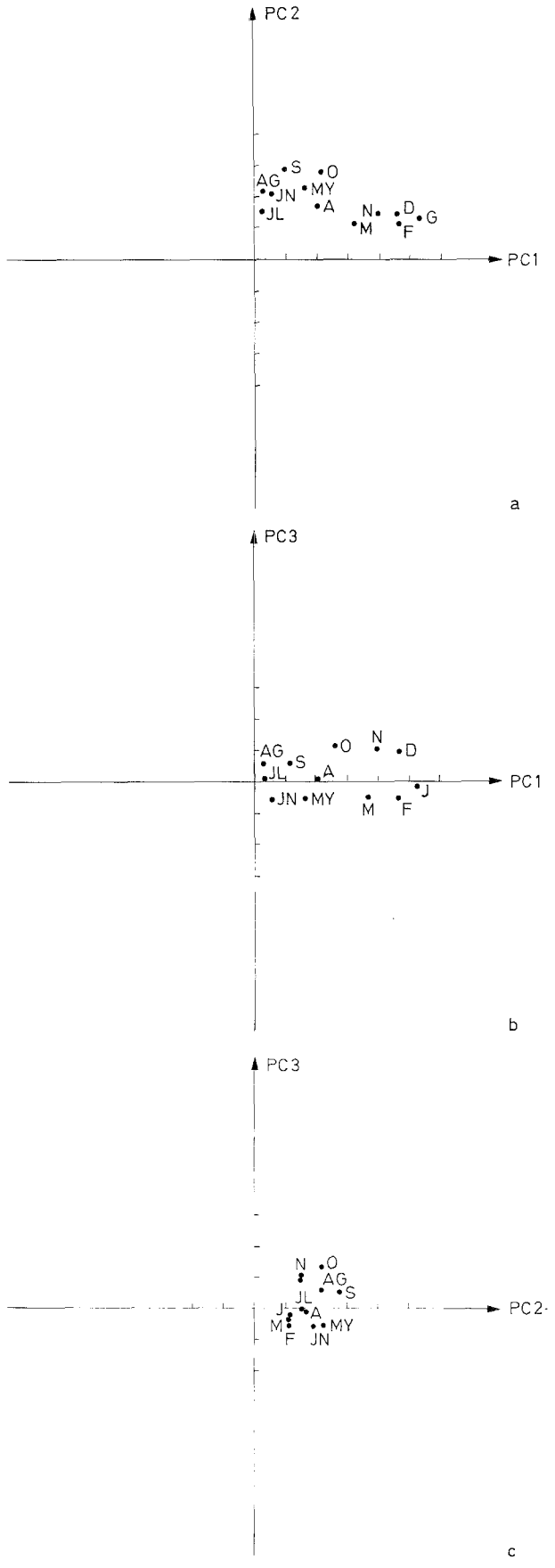


Fig. 4a, b, c. Pairwise plots of the varimax rotated principal components

Table 4. *Varimax Rotated Component Loadings (in mm)*

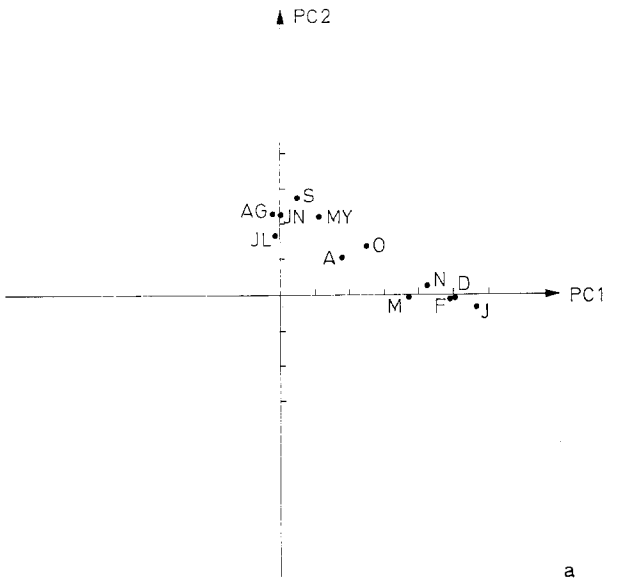
Variable	Component 1	Component 2	Component 3
January	53.0	12.2	−2.6
February	47.2	11.3	−6.0
March	37.1	11.7	−5.1
April	21.3	16.8	−0.8
May	17.2	22.7	−6.3
June	7.0	19.8	−6.0
July	3.8	15.2	−0.3
August	4.1	21.7	5.8
September	12.2	28.4	5.4
October	27.0	22.6	12.0
November	40.1	15.4	10.3
December	47.6	15.4	9.1

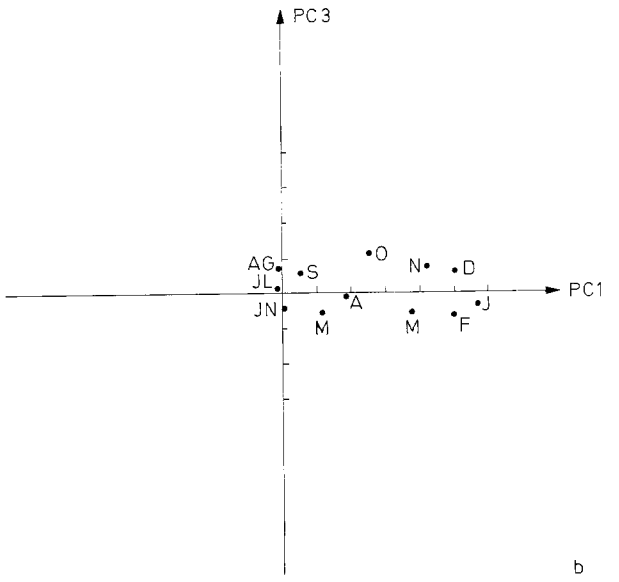
are still complex and will hamper the interpretations.

Table 4 shows varimax rotated factor loadings corresponding to components 1, 2 and 3. Rotated PC1 and rotated PC3 represent similar variation patterns to those of the unrotated solution; nevertheless, the physical meaning of the rotated PC2 is now less clear because the differences between high and low loadings are smaller (see Table 4 and Table 2b). In short, varimax rotation has added confusion and must not be recommended in this case. By this reason the spatial distribution of varimax rotated component scores is not included.

2.3 Oblimin Rotated Solution

Table 3b shows the congruence coefficients of the unrotated and the oblimin rotated components. There is a good agreement between unrotated and rotated PC1 and the same conclusion can be drawn with unrotated and rotated PC2. The latter congruence was worse when a varimax rotation was performed. By contrast, a good inverse agreement appears when unrotated and rotated PC3 are compared. Consequently, the patterns that have been established from the unrotated solution will be confirmed. Pairwise plots of rotated components are drawn in Fig. 5a–c and show that the degree of simple structure has been slightly improved. A comparison between Fig. 5a–c (oblimin rotated solution) and Fig. 2a–c (unrotated solution) leads to the conclusion that the number of variables that fall near a factorial axis is greater in the first set of figures. That means that in the oblimin rotated solution, the differences





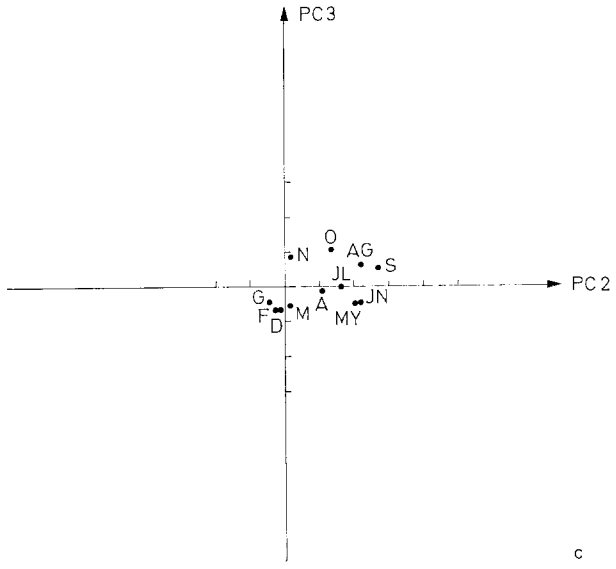


Fig. 5a, b, c. Pairwise plots of the oblimin rotated principal components

Fig. 6a, b, c. Spatial distributuion of oblimin rotated component scores (PC1: 6a, PC2: 6b, PC3: 6c)

between high and low loadings are greater than in the unrotated solution. Consequently, after an oblimin rotation, the previous component interpretations are still valid and more clear.

Figure 6a–c show the spatial distribution of oblimin rotated component scores. With respect to unrotated scores, only a change in sign in PC3 scores field is remarkable.

3. The Method of Rainfall Regionalization

The Davis and Kalkstein classification algorithm works with a previous division based on an Average Linkage method which has been extensively applied in climatology (Kalkstein et al., 1987; Serra et al., 1991 etc). The number of groups is determined in terms of the variations in the similarity index (Lana and Fernández Mills, 1993; Fernández Mills et al., 1992)

$$L_{nm} = D^2_{nm} + V_n + V_m \tag{3}$$

where D^2_{nm} represents the square Euclidian distance between centroids of groups n and m, and V_n and V_m represent the within-group variances for groups n and m. Two groups are fused when their similarity index is minimum. Taking into account that average linkage is an aggregative algorithm, the similarity index L_{nm} corresponding to each fusion will increase with a decreasing number of remaining clusters. A fusion leading to a sharp increase in L_{nm} is rejected because two dissimilar groups have been merged. Consequently, the optimal number of groups is the one coresponding to the previous fusion.

In a second stage, a K means clustering (Dixon, 1985) is applied to the configuration obtained in the first step using the centroid coordinates of each group as seeds. Three possible regionalizations can be derived if three different starting data matrices, including unrotated, varimax rotated and oblimin rotated component scores, are considered. Nevertheless, it must be underlined that the similarity indices of Average Linkage and K means procedures are built up from the Euclidean distance between elements, and this variable will remain unchanged if an orthogonal rotation is performed. Consequently, the classifications deduced by means of unrotated and varimax rotated component scores will be identical, and only regionalizations starting from unrotated and oblimin rotated scores will be presented and compared.

3.1 Group Homogeneity

The homogeneity of each cluster can be studied by computing pairwise correlation coefficients within the group. Their minimum (R_{min}), average ($\langle R \rangle$), and maximum (R_{max}) values can be good estimators of cluster coherence. If these coefficients are high, the corresponding group must be considered as homogeneous. By contrast, if a cluster includes one or several dissimilar elements, at least one of the above mentioned values will be small. All these concepts have been applied to the clusters derived in the two regionalization procedures, and the quality of the resulting divisions is then compared.

4. Rainfall Classification Based on the Unrotated PCA Solution

Table 5 shows the evolution of the similarity index for the last iterations of the average linkage process. According to the rules mentioned above, an initial sharp increase appears when a reduction from 10 to 9 groups is attempted. An average linkage solution including 10 clusters is then assumed to be the best one. A K means algorithm is then applied to the configuration previously deduced and the final results are shown in Table 6 and Fig. 7. The main characteristics of the clusters obtained are summarized in the following section. Table 7 includes the 12 centroid coordinates of each group in mm.

Table 5. *Evolution of the Similarity Index in the last Stages of the Average Linkage Clustering Algorithm Applied to a Data Matrix Including Unrotated Component Scores*

Remaining Clusters	L_{mn}
18	1.4
17	1.4
16	1.6
15	1.6
14	1.7
13	1.9
12	2.1
11	2.3
10	2.4
9	3.4
8	3.5
7	3.9
6	4.4
5	6.0
4	6.2
3	7.6

Table 6. *Final Number of Elements of each Cluster After a Davis and Kalkstein Regionalization Based on Unrotated Component Scores*

Group	Number of Elements	(%)
1	24	35.3
2	14	20.6
3	5	7.4
4	5	7.4
5	5	7.4
6	8	11.8
7	2	2.9
8	1	1.4
9	3	4.4
10	1	1.4

Cluster 1 includes 24 raingauges located at inner areas of the country (Fig. 7). Their PC3 scores are significantly positive (Fig. 3c). Mean maximum values are frequently reached in May when Atlantic frontal waves show increased activity caused by a maximum thermal contrast between continents and oceans. On the other hand, summer rainfall is scarce and mainly due to occasional air mass storms generally leading to modest amounts. The Mediterranean influence is not very relevant, as the increase in Mediterranean cyclogenesis in autumn is not followed by important monthly rainfall increases in September and October. The group homogeneity is not very high as $\langle R \rangle$ reaches the lowest value (0.664) (see Table 8) and

Table 7. *Centroid Coordinates in mm for each Group Deduced from a Davis and Kalkstein Classification Procedure Applied to a Starting Data Matrix Including Unrotated Component Scores*

GR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1	45	43	46	45	51	44	19	20	40	48	49	47
2	52	46	53	38	28	11	2	3	17	46	51	59
3	96	79	70	45	27	12	1	4	17	62	105	109
4	36	38	60	56	66	49	33	55	89	96	60	60
5	124	117	109	88	87	58	29	31	72	108	127	132
6	41	33	38	40	31	23	8	20	50	85	60	57
7	300	269	220	140	137	83	38	47	118	183	204	244
8	160	165	152	104	93	59	39	41	90	141	183	228
9	120	93	81	91	86	65	48	73	99	124	135	142
10	149	108	93	121	121	89	77	128	155	143	167	188

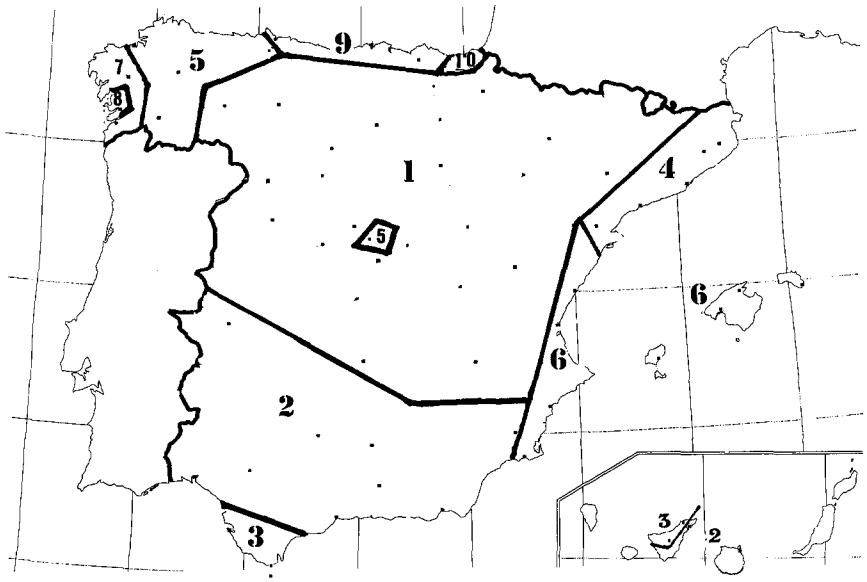


Fig. 7. The Spanish rainfall regions deduced after a Davis and Kalkstein clustering procedure applied to an initial data matrix including unrotated component scores

Table 8. *Analysis of Group Homogeneity in the First Regionalization* (based on unrotated component scores). R_{min}, R_{max} and $\langle R \rangle$ are minimum, maximum and mean pairwise correlation coefficients within each group. N is the number of elements of each group

Group	N	R_{min}	R_{max}	$\langle R \rangle$
1	24	−0.305	0.986	0.664
2	14	0.379	0.989	0.834
3	5	0.894	0.992	0.963
4	5	0.762	0.959	0.827
5	5	0.822	0.956	0.910
6	8	0.693	0.982	0.859
7	2	0.973	0.973	0.973
8	1	1.000	1.000	1.000
9	3	0.886	0.948	0.916
10	1	1.000	1.000	1.000

there are elements which are inversely correlated (Ponferrada and Teruel (−0.305) for example). Nevertheless, the number of negative correlations is small, and Table 9 shows that 57% of the values range between 0.6 and 1.0.

Cluster 2 is formed by elements of S and SE peninsular areas, and also includes the drier raingauges located in the Canary Islands (Fig. 7). Their rainfall regime is characterized by significant winter precipitation together with a deep minimum in the hot season (Table 7). In fact, the records are extremely poor in June, July, August and September, where baroclinic disturbances linked to westerlies affect regions with higher latitudes. In addition, the storm activity related to the hot season appearing occasionally in cluster 1, is non-existent here, and the cyclonic waves which develop as separate configurations of the lower mid-latitudes affect the region in late autumn and in the winter season. These patterns explain the distribution of cluster 2 monthly rainfall along the year. Negative scores in PC2 are a common feature of cluster 2 stations. The group homogeneity is remarkable ($\langle R \rangle = 0.834$).

Cluster 3 includes stations with a similar behaviour but with greater winter records (Fig. 7, Table 7). The synoptic patterns explaining the distribution of monthly rainfall throughout the year are the same as in group 2. Nevertheless, the high amounts recorded in late autumn and winter result from a greater exposure of the raingauges included in group 3 to oceanic air masses carried by Atlantic depressions with southern tracks. Cluster 3 is very homogeneous ($\langle R \rangle = 0.963$).

Cluster 4 shows a moderate degrees of coherence ($\langle R \rangle = 0.827$), and is associated with littoral and pre-littoral areas of the NE of Spain (Fig. 7). The corresponding rainfall regime is very similar to the one governing stations of S or SE France (Dauphiné, 1976). Maximum monthly values are recorded in Autumn (September and October) and are caused by an increase in frontal storm activity in September and in Mediterranean cyclogenesis in October. In fact, the presence of troughs or cut-offs in upper level layers becomes significant in this season and can lead to heavy rainfall episodes when eastern advection dominates at the surface. The potential instability of mediterranean air masses is important in autumn as sea surface temperatures reach their highest values in this period. Minimum monthly values appear in July but summer precipitation is not so scarce as in southern groups; in fact August amounts

Table 9. *Distribution of Pairwise Correlation Coefficients within Groups in %, Corresponding to the Regionalization Based on Unrotated Component Scores.* Correlation intervals are 1 (< 0), 2 (0–0.1), 3 (0.1–0.2),..., 10 (> 0.8)

Correlation Intervals										
Group	1	2	3	4	5	6	7	8	9	10
1	2.5	4.0	3.3	5.4	9.8	4.3	12.7	14.9	27.9	15.2
2	0.0	0.0	0.0	1.1	2.2	4.4	8.8	15.4	29.7	38.5
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	90.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	40.0	50.0	10.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0	80.0
6	0.0	0.0	0.0	0.0	0.0	0.0	3.6	21.4	42.9	32.1
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.3	66.7
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

are comparable to those recorded in the Spring months. This behaviour can be interpreted easily if we take into account that the interaction of warm and humid surface air masses and cold advection in upper layers is not rare in late summer, and can lead to significant storm activity, mainly in the second half of August.

Cluster 5 can be considered as very coherent ($\langle R \rangle = 0.910$, Table 8). This class is related to some parts of NW Spain and includes one mountain station (Navacerrada) located in the Central Chain (Fig. 7 and Fig. 1). The rainfall regime is characterized by maximum monthly amounts in winter and a deep minimum in summer (Table 7). The annual variations in monthly precipitation show the same patterns as those of cluster 3 but the recorded amounts are now greater. All these features can be interpreted in a similar way, and the conclusions that have been drawn with cluster 3 are also valid. Nevertheless, it must be underlined that Cluster 5 NW raingauges are more exposed to North-Atlantic air masses carried by depressions with northern tracks. The northwestern flow following a cold front passage produces frequent showers on the north side of the Cantabric System and in NW Spain (Fig. 1). The southern side and the Meseta are sheltered by the chain when this circulation type takes place. The station of Navacerrada (Central Chain) belongs surprisingly to this cluster. Nevertheless, its altitude (1800 m) and situation (in the top of the chain, Fig. 1) can explain a behaviour which is similar to the one governing northern raingauges. When the case of Navacerrada is considered, orographic influences are critical in the intensification of rainfall when the Iberian Peninsula is affected by SW or NW cyclonic flows. These synoptic situations are frequent in the cold season and may explain the existence of greater amounts than in clusters 3 and 1.

Cluster 6 ($\langle R \rangle = 0.859$) includes E and SE areas of the Iberian Peninsula and the Balearic Islands, which are governed by a typical west Mediterranean precipitation regime. Maximum monthly rainfall is mostly recorded in October and, like in cluster 4, associated with Mediterranean lows producing eastern advection and carrying warm and humid air masses to the coast and the nearer ranges. These surface situations are often accompanied by the existence of troughs or cut-offs in the upper layers and can be responsible for flood damage. Although clusters 4 and 6 correspond to Mediterranean regions, there are some differences between them. First of all, summer rainfall in the latter group is scarcer and this is probably due to the southern latitudes of cluster 6 elements which are far away from the usual tracks of surface and upper level disturbances. Secondly, spring precipitation is systematically higher in cluster 4 and finally, September amounts lead to a wider autumn peak in the NE.

The two rainiest stations of the data base belong to cluster 7 ($\langle R \rangle = 0.973$, Table 8). As in groups 2, 3 and 5, maximum rainfall is recorded in winter and a deep minimum appears in summer. The synoptic situations responsible for rainfall episodes are similar to those of cluster 5; that means that the contribution of baroclinic disturbances related to westerlies is important. In fact, the exposure of cluster 7 stations to W, SW or NW flows that enable a frontal overrun is maximum, and may explain the magnitude of winter rainfall. With the possible exception of July and August, monthly precipitation is always relevant.

Cluster 8 is formed by one Northwestern station with similar patterns to group 7 elements (Fig. 7). Only the smaller amounts recorded in the former region can explain the separation between these groups.

Cluster 9 ($\langle R \rangle = 0.916$, Fig. 8) includes raingauges located on the Cantabric littoral and which are very exposed to north and northwesterly flows. The group is comparable with cluster 5, but two differences must be underlined. Firstly, summer precipitation is more relevant than in the latter class and secondly, a secondary maximum appears in April.

Cluster 10 is formed by a single station (Donosti) on the eastern part of the Cantabric coast. The rainfall patterns are similar to those of group 9. However, the nearness of the Pyrenean chain may explain the systematic existence of higher monthly amounts in this eastern raingauge.

5. Rainfall Classification Based on the Oblimin Rotated PCA Solution

Table 10 shows the evolution of the similarity index in the last iterations of the average linkage process. As in the previous case, an initial sharp increase is detected when a transition from 10 to

Table 10. *Evolution of the Similarity Index in the last Stages of the Average Linkage Procedure Applied to an Initial Data Matrix Including Òblimin Rotated Component Scores*

Remaining Clusters	L_{mn}
18	1.0
17	1.1
16	1.1
15	1.4
14	1.7
13	1.8
12	2.0
11	2.1
10	2.1
9	2.9
8	3.2
7	3.4
6	3.8
5	4.0
4	5.0
3	8.2

9 remaining clusters is attempted. Consequently, 10 groups will also be considered in this regionalization. Table 11 includes the size of each deduced cluster when a K Means procedure is applied to the first configuration, and the results of the analysis of group homogeneity are summarized in Tables 12 and 13. The comparison between the two classifications can be made easily as the number of groups is the same. A close look at the latter table and Table 8 suggests that the first division is better and is formed by groups with a high degree of coherence. In fact, clusters 1 and 3 of the second regionalization include elements

Table 11. *Final Number of Elements After a Davis and Kalkstein Regionalization Based on Oblimin Rotated Component Scores*

Group	Number of Elements	(%)
1	24	35.3
2	13	19.1
3	8	11.8
4	4	5.9
5	9	13.2
6	3	4.4
7	2	2.9
8	2	2.9
9	2	2.9
10	1	1.5

Table 12. *Analysis of Group Homogeneity in the Second Regionalization* (based on oblimin rotated component scores). R_{min}, $\langle R \rangle$ and R_{max} are minimum, maximum and mean pairwise correlation coefficients within a group. N is the number of elements of each group

Group	N	R_{min}	R_{max}	$\langle R \rangle$
1	24	−0.305	0.973	0.644
2	13	0.379	0.992	0.847
3	8	−0.105	0.956	0.575
4	4	0.487	0.948	0.725
5	9	0.485	0.959	0.775
6	3	0.960	0.988	0.973
7	2	0.981	0.981	0.981
8	2	0.973	0.973	0.973
9	2	0.908	0.908	0.908
10	1	1.000	1.000	1.000

Table 13. *Distribution of Pairwise Correlation Coefficients within Groups Corresponding to the Regionalization Based on Oblimin Rotated Component Scores.* Correlation intervals are 1 (< 0), 2 (0.0–0.1), 3 (0.1–0.2), ..., 10 (>0.8).

Correlation	Intervals									
Group	1	2	3	4	5	6	7	8	9	10
1	5.4	3.6	4.0	5.1	5.4	7.6	4.0	12.0	14.1	42.0
2	0.0	0.0	0.0	0.0	1.3	5.1	2.6	6.4	10.3	74.3
3	10.7	7.1	10.7	7.1	0.0	10.7	0.0	7.1	10.7	42.9
4	0.0	0.0	0.0	0.0	0.0	16.7	33.3	0.0	0.0	50.0
5	0.0	0.0	0.0	0.0	0.0	2.8	11.1	8.3	27.8	50.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

which are inversely correlated, and this shortcoming was only present in cluster 1 of the first division. Furthermore, the latter is formed by 9 groups with mean pairwise correlation coefficients greater than 0.8, while only 6 clusters of the second classification exceed this threshold. Consequently, the oblimin rotation has slightly improved the interpretations of the Spanish precipitation patterns and leads to a simpler structure. However, when a Davis and Kalkstein regionalization procedure is attempted starting from oblimin rotated component scores, the results are disappointing. This is the main reason why the features of the second regionalization are not discussed.

Some comments are necessary about the general motivation for PCA in this particular case. Generally, one of the goals of PCA is to compress large data sets, where there is no chance to look at all the information in detail. This argument does not hold in this case since it is not very difficult to study 12 patterns and their differences instead of the 3 pc-score patterns. Nevertheless, some of the seasonal changes found by PCA cannot be detected more easily by simply looking at the mean monthly precipitation data. In short, PCA allows the identification of three basic seasonal patterns:

- those related to high PC1 scores where maximum monthly values are recorded in winter with a deep minimum in summer.
- those associated with high PC2 scores which correspond to stations where maximum rainfall is recorded in late summer or early autumn.
- Finally, patterns characterized by high positive PC3 scores in the unrotated solution (high negative in the oblimin rotated) and linked with inner raingauges where maximum monthly values are recorded in late spring (May). In this case, autumn rainfall is not specially remarkable as in Mediterranean areas. These three patterns explain 96% of the total variance. The seasonal rainfall behaviour of a Spanish station is a linear combination of these. These conclusions justify the PCA approach prior to the identification of coherent rainfall clusters.

6. Conclusions

In the first part of the present study, Spanish precipitation patterns were derived starting from an initial data matrix of 68 synoptic stations characterized by 12 mean monthly precipitation amounts. PCA is applied to the interstation covariance matrix and three unrotated, varimax rotated and oblimin rotated components accounting for 96% of the total variance were considered. Their pairwise plots showed that a slightly simpler structure is achieved when an oblimin rotation is performed. By contrast, varimax results are disappointing.

The interpretations given to unrotated and oblimin rotated component score fields allow the identification of different modes of variation in Spanish precipitation. A significant part of the country (with significant PC1 scores) is governed by the Atlantic and shows peaks of rainfall activity in winter months followed by deep minima in the hot season. On the other hand, stations belonging to eastern and northeastern regions (with significant PC2 scores) reach their maximum values in autumn, in agreement with an increased frequency of west Mediterranean cyclogenetic processes. Positive PC3 scores in the unrotated solution and negative ones in the oblimin rotated factorization are related to inland stations recording maximum rainfall values in spring, when frontal activity is enhanced by the existence of a maximum thermal contrast between the continent and the Atlantic ocean. All these features are taken into account when a Davis and Kalkstein regionalization is performed starting from unrotated or oblimin rotated component scores. In both cases 10 groups are considered, and a subsequent study of group homogeneity reveals that the classification based on unrotated component scores is more reliable. The characteristics of each cluster have been extensively discussed and illustrate the climatic variety of the rainfall regimes that coexist in Spain (subtropical, Mediterranean, Atlantic, continental). Some elements of the described patterns exist in other European latitudes (Western Europe, Southern France, Italy). Nevertheless, we can also find features inherent to North African precipitation, mainly in the Canary Islands and in south and southeastern parts of the Iberian Peninsula.

References

Daultrey, S., 1976: *Principal Component Analysis.* Norwich, Great Britain: Geo Abstracts Ltd.

Dauphiné, A., 1976: Les précipitations dans les midis français, étude de Climatologie inductive. Thèse d'Etat. Université de Nice.

Davis, R. E., Kalkstein, L. S., 1990: Development of an automatic spatial synoptic climatological classification. *Int. J. Climatol.*, **10**, 769–794.

Dixon, W. J., 1985: *BMDP Statistical Software*. Berkeley, Los Angeles: University of California Press.

Fernández Mills, G., Serra, C., Lana, X., Periago, M. C., 1992: Une méthode de classification pour les types de temps à Barcelone. *La Météorologie 7 ème série*, **43**, 43–51.

Fernández Mills, G., Lana, X., Serra, C., 1994: Catalonian Precipitation patterns: Principal Component Analysis and automated regionalization. *Theor. Appl. Climatol.*, **49**, 201–212.

Harman, H. H., 1976: *Modern Factor Analysis*. Chicago, IL: The University of Chicago Press.

Jansà, A., 1978: Inestabilidad baroclina y ciclogénesis en el Mediterráneo occidental. Ph.D. thesis, Univ. Autònoma de Barcelona.

Kalkstein, L. S., Tan, G., Skindlov, J. A., 1987: An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Climate Appl. Meteor.*, **26**, 717–730.

Lana, X., Fernández Mills, G., 1994: Minimum sample size for synoptic weather type classification: application to a winter period data recorded on the catalonian coast (NE Spain). *Int. J. Climatol.* (In Press).

Lawson, J. S., 1989: The wet spell in southern Spain mid-October 1988. *Weather*, **44**, 475–478.

Legates, D. R., 1991: The effect of domain shape on Principal Component Analysis. *Int. J. Climatol.*, **11**, 135–146.

Maheras, P., 1985: A Factorial analysis of mediterranean precipitation. *Arch. Met. Geoph. Biocl., Ser B*, **36**, 1–14.

Periago, M. C., Lana, X., Serra, C., Fernández Mills, G., 1991: Precipitation regionalization: an application using a meteorological network in Catalonia (NE Spain). *Int. J. Climatol.*, **11**, 529–543.

Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography*. New York: Elsevier.

Prezerakos, N. G., 1990: Synoptic flow patterns leading to the generation of northwest African depressions. *Int. J. Climatol.*, **10**, 33–47.

Prezerakos, N. G., Michaelidis, S. C., Vlassi, A. S., 1990: Atmospheric synoptic conditions associated with the initiation of north-west African depressions. *Int. J. Climatol.*, **10**, 711–730.

Ramis, C., Ballester, M., 1984: Ciclogénesis catalano-balear. Estudio del temporal de Abril de 1978. *Rev. Geof.*, **40**, 243–258.

Richman, M. B., 1986: Rotation of Principal Components. *Int. J. Climatol.*, **6**, 293–335.

Serra, C., Fernández Mills, G., Lana, X., Periago, M. C., 1991: Regionalización pluviométrica de Cataluña a partir de valores medios mensuales des precipitación. *Anales de Física*, **B-87**, 284–293.

Sumner, G., Ramis, C., Guijarro, J. A., 1993: The spatial organization of daily rainfall over Mallorca, Spain. *Int. J. Climatol.*, **13**, 89–109.

White, D., Richman, M., 1991: Climate regionalization and rotation of Principal Components. *Int. J. Climatol.*, **11**, 1–25.

Wilmott, C. J., 1978: P-mode principal component analysis, grouping and precipitation regions in California. *Arch. Met. Geoph. Biocl., Ser. B.*, **26**, 277–295.

Author's address: G. Fernández Mills, Departament de Física i Enginyeria Nuclear, ETSEIB, Universitat Politècnica de Catalunya, Avda Diagonal 647, E-08028 Barcelona, Spain.