

Cluster Analysis of Three-Way Atmospheric Data

Isabella Morlini and Stefano Orlandini

Abstract Classification of meteorological time series is important for the analysis of the climate variability and climate change. The clustering of several years in groups that are homogeneous with reference to the amount of precipitation and to the atmospheric condition can aid in understanding the structure of precipitation and may be important in developing hydrological models. In this paper we propose a cluster analysis of multivariate time series based on a dissimilarity measure that considers the functional form of the data. The unit to be classified are 148 years, from 1861 to 2008, and the variables are the values of precipitation, the minimum temperature, and the maximum temperature in different occasions (days or months) in the province of Modena (Northern Italy).

Keywords Climate change • Clustering • Functional data analysis • Precipitation

1 Introduction

When studying climate change in a spatial area, we may search for typical patterns, common to some time periods, describing the underlying atmospheric process. The analysis and the comparison of these different patterns may give an insight into the long-period changes in meteorological variables, such as rain and temperature. These typical patterns may be thought of as centroids of homogeneous clusters, where the units to be classified are years over a long period of time and the variables are measurements of rain and temperature in different occasions (for example, days

I. Morlini (✉)

Department of Economics “Marco Biagi”, University of Modena and Reggio Emilia, Via Berengario 51, 41100 Modena, Italy
e-mail: isabella.morlini@unimore.it

S. Orlandini

Department of Mechanical and Civil Engineering, University of Modena and Reggio Emilia, Strada Vignolese 905, 41125 Modena, Italy
e-mail: stefano.orlandini@unimore.it

© Springer International Publishing Switzerland 2015

I. Morlini et al. (eds.), *Advances in Statistical Models for Data Analysis*,
Studies in Classification, Data Analysis, and Knowledge Organization,
DOI 10.1007/978-3-319-17377-1_19

177

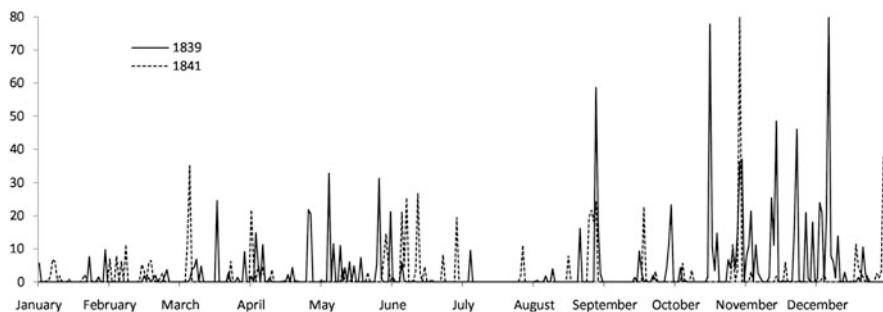


Fig. 1 Daily values of rainfall in Modena in the years 1839 and 1841

or months). Classification of these three-way data (unit \times variables \times occasions) should consider the functional form of the multivariate time series. Indeed, salient features of atmospheric measurements, such as extreme values, maxima or minima, may result shifted in the different series. The transformation of time, that is the warping function from one series to another, must be estimated, before computing the dissimilarity between pairs of series. This function permits a fruitful alignment of the two sequences of measurements. As an example, Fig. 1 reports the daily values (explains) of rainfall intensity in the years 1839 and 1841, in the province of Modena (Northern Italy). The two sequences show a great similarity, considering that both years have a peak around 30 mm in March, three days with more than 20 mm in the period May–June and, in particular, a very rare event such as a daily value near 80 mm in October. The timing of this very rare event is shifted of 13 days in the 2 years (it occurs the 16th of October in the year 1839 and the 29th of October in the year 1841). Cross-sectional similarities, which compare measurements gathered in the same day, produce pessimistic values for these two series. A more comprehensive similarity should align similar events that occur in nearby days. Even the simplest data analysis, such as computing a mean, can require that features be first aligned by a time transformation, a process that is called time-series registration.

Classical functional data analysis [2, 7, 8, 10] interpolates the sequences of values by a smooth curve and assumes that also the time-warping function is a smooth function, differentiable as the curves themselves. Suppose we have n observed values x_{ijt} , $i = 1, \dots, n$, of variable j ($j = 1, \dots, p$) at time t ($t = 1, \dots, T$). In functional data analysis, the model usually assumed is

$$x_{ijt} = s_{ij}(w_{ij}(t)) + \varepsilon_{ijt}, \quad (1)$$

where s_{ij} is the smooth function underlying the time series i of variable j ($i = 1, \dots, n$), ($j = 1, \dots, p$), $w_{ij}(t)$ is the smooth time warping function, and ε_{ijt} is the error term. The errors are assumed independent and identically distributed. The

function $w_{ij}(t)$ is subject to the following constraints:

1. $t_1 < t_2 \iff w_{ij}(t_1) < w_{ij}(t_2)$
2. $w_{ij}(0) = 0$
3. $w_{ij}(T) = T$

To keep the notation simple, in (1) it is assumed that, for each variable j and for each time series i , both the number T and the timing of the sampled values x_{ij} are identical. However, many applications involve variation in locations and numbers of sampling points across replications and formula (1) may be adjusted for these cases. The smooth functions s_{ij} and w_{ij} depend on the time series i and on the variable j and each observation x_{ijt} is associated with the registered curve value $s_{ij}(w_{ij}(t))$. The simplest curve alignment procedure is a landmark registration. A landmark is a feature with a location that is clearly identifiable in all curves. The curves are aligned by transforming the physical time so that the location of the landmarks is the same for all curves. In case of a single landmark, if t_0 is the timing of this landmark in variable j and t_i is the timing of this landmark in curve i , then the time-warping function $w_{ij}(t)$ is specified by fitting a smooth function to the three points $(0, 0)$, (t_0, t_i) , and (T, T) . This function is as differentiable as the curves s_{ij} themselves. According to this definition, $w_{ij}(t_i) = t_0$, and all the registered functions defined as $y_{ij}(t) = s_{ij}(w_{ij}(t))$ will all automatically arrive at the landmark at the same time, namely t_0 . Both the definition of multiple landmarks and their unequivocal identification in individual curves are problematic, especially in long time series of atmospheric data. For example, in Fig. 1, the timing of the rain peak in March may be either the 5th or the 17th. In October, t_0 may be either the 16th or the 29th. Moreover, these peaks may be not so visible in other years. It is evident that both the exact number and the daily locations of landmarks in rain and temperature time series are not objectively identifiable. As an alternative to landmark registration, in this paper we use the dynamic time warping algorithm (dtw). In its original formulation, the dtw estimates a “warping path” for aligning one series to another and minimizes a measure of “discrepancy” between the two series which is called dynamic time warping cost (dtwc). However, if we modify one of the constraints in the classical formulation, the algorithm estimates a path which is a discrete time warping function and minimizes, a cost which is a dissimilarity measure between the two registered series. The dtw algorithm doesn’t require the estimate of the smooth curves interpolating the time series. However, we may estimate these curves and use the smoothed valued $s_{ij}(t)$ in order to have data less noisy than the sampled values x_{ijt} . The main features of dtw are as follows:

- It is a nonparametric procedure which does not require prior assumptions about the form of the warping functions (see, e.g., [9, 12] for the definition of parametric warping functions) or about the number and the timings of salient events (the landmarks).
- It relies on a minimization problem which can be solved efficiently by using dynamic programming.

- When applied to three-way data, the warping functions are estimated by considering the vector-valued time series $\mathbf{x}_{it} = [x_{11t}, \dots, x_{1jt}, \dots, x_{1pt}]$ and not by considering each univariate series x_{ijt} ($j = 1, \dots, p$) ($t = 1, \dots, T$) separately. Therefore, rather than estimating a univariate warping function w_{ij} for each variable j , the dtw estimates a p -variate warping function w_i .

These last two items differentiate the dtw algorithm used in this paper and the algorithm illustrated in [13, 14]. In Wang and Gasser the warping functions are univariate smooth continuous functions.

The paper is organized as follows. In Sect. 2 we illustrate the dtw algorithm used in paper. In Sect. 3 we focus on the application. We first describe the data and the study area and then we show how atmospheric data can be clustered and analyzed to achieve meaningful results.

2 The Warping Function and the Measure of Dissimilarity

The dtw algorithm was originally developed in engineering, for speech analysis and speech recognition, in order to align two sequences of values. Many enhancements of the method have been proposed in the data mining literature. Among other works, we refer to [1, 3–5]. In its original formulation, given the p -dimensional vector-valued series \mathbf{x}_{1t} and \mathbf{x}_{2t} , where $\mathbf{x}_{it} = [x_{11t}, \dots, x_{1jt}, \dots, x_{1pt}]$, $i = 1, 2$ and $t = 1, \dots, T$, the dtw first implies the construction of a $T \times T$ square lattice D , in which the element $d(r, c)$ ($r, c = 1, \dots, T$) is the distance $d(\mathbf{x}_{1r}, \mathbf{x}_{2c})$ between the values of series 1 at time r and the values of series 2 at time c . Any distance may be used in the construction of the square lattice D . However, before computing any Minkowski metric, the p variables should be standardized to take into account the different units of measurements and/or the different variability [6]. Each element $d(r, c)$ corresponds to the alignment between points \mathbf{x}_{1r} and \mathbf{x}_{2c} in the p -dimensional Euclidean space. The dtwc is defined as follows:

$$dtwc = \min \sqrt{\frac{\sum_{k=1}^K d_k}{K}}, \quad (2)$$

where $T \leq K \leq (2T - 1)$, K is determined by the optimization process of the algorithm and the d_k are elements of D subject to the following constraints:

- Boundary condition: $d_1 = d(1, 1) = d(\mathbf{x}_{11}, \mathbf{x}_{21})$ and $d_K = d(T, T) = d(\mathbf{x}_{1T}, \mathbf{x}_{2T})$. This constraint requires that the first time and the last time in one series are aligned with the first time and the last time, respectively, in the other series. So, the first and the last time are not warped.
- Continuity constraint: given $d_k = d(\mathbf{x}_{1r}, \mathbf{x}_{2c})$ then $d_{k-1} = d(\mathbf{x}_{1r'}, \mathbf{x}_{2c'})$ where $(r - r') \leq 1$ and $(c - c') \leq 1$. This condition restricts two successive elements d_k to be adjacent (including diagonally) elements in D .

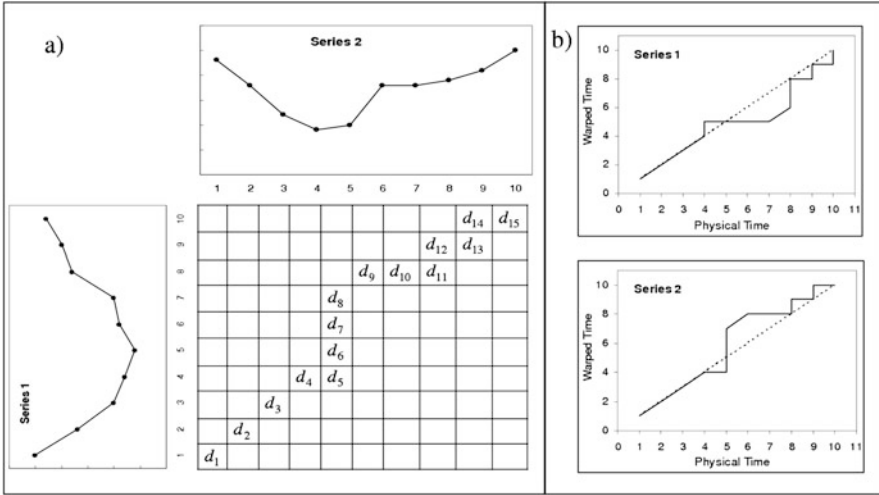


Fig. 2 (a) An example of the distances included in the dtwc; (b) the warping path

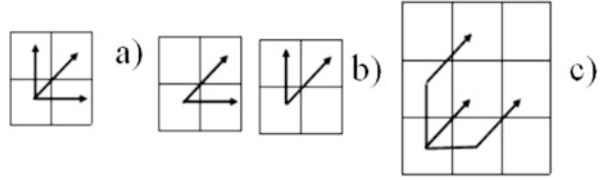
- Monotonicity constrain: given $d_k = d(\mathbf{x}_{1r}, \mathbf{x}_{2c})$ then $d_{k-1} = d(\mathbf{x}_{1r'}, \mathbf{x}_{2c'})$ where $(r - r') \geq 1$ and $(c - c') \geq 1$. This condition forces the couple of points for which the distance is taken into account in the dtwc to be monotonically spaced in time.

The dtw produces a relative shift between the two sampled curves. However, as shown in Fig. 2a, the algorithm defines a warping path and from this path we cannot draw two increasing warping functions to align \mathbf{x}_1 to \mathbf{x}_2 and to align \mathbf{x}_2 to \mathbf{x}_1 , since a single point on one time series may map onto a large subsection of the other series (Fig. 2b). In order to find two monotonic—not strictly increasing—warping functions, one could eliminate the boundary condition $d_k = d(\mathbf{x}_{1T}, \mathbf{x}_{2T})$ and restrict the continuity constrain such that $(r - r') = 1$ for aligning \mathbf{x}_1 to \mathbf{x}_2 and such that $(c - c') = 1$ for aligning \mathbf{x}_2 to \mathbf{x}_1 (Fig. 3b). However, with this restriction, the dtwc becomes asymmetric and cannot be a dissimilarity measure: given two sequences i and i' , and, to keep the notation simple, $\text{dtwc}(\mathbf{x}_i, \mathbf{x}_{i'}) = \text{dtwc}(ii')$, then $\text{dtwc}(ii') \neq \text{dtwc}(i'i)$. In order to define at the same time a dissimilarity measure and a warping function, we use a modified parameterized path. This path is characterized by a weaker continuity constraint, defined as follows:

Continuity constraint: given $d_k = d(\mathbf{x}_{1r}, \mathbf{x}_{2c})$ then $d_{k-1} = d(\mathbf{x}_{1r'}, \mathbf{x}_{2c'})$ where $(r - r') \leq 2$ & $(c - c') < 2$ or $(r - r') < 2$ & $(c - c') \leq 2$ (Fig. 3c).

With this continuity constraint, the classical boundary condition, and the monotonicity constraint, the dtw algorithm estimates a $w d_i(t)$ warping function, with the

Fig. 3 Representation of the dtw step. (a) The classical dtw step; (b) step with restrictions on the continuity constraint; (c) step with a weaker continuity constraint



following properties:

1. $t_1 < t_2 \Rightarrow wd_i(t_1) \leq wd_i(t_2)$ (the function is monotonic increasing but not strictly increasing and it is not smooth)
2. $wd_i(0) = 0$
3. $wd_i(T) = T$

and a dtwc dissimilarity measure, satisfying the following conditions:

1. $dtwc(ii') \geq 0, i, i' = 1, \dots, N$ (nonnegativity)
2. $dtwc(ii) = 0, i = 1, \dots, N$ (this a condition weaker than the identity condition required for distance measures)
3. $dtwc(ii') = dtwc(i'i), i, i' = 1, \dots, N$ (symmetry)

As outlined in the Introduction, wd_i is equal for every variable j ($j = 1, \dots, p$), since it is estimated considering the vector-valued series $\mathbf{x}_{it} = [x_{11t}, \dots, x_{1jt}, \dots, x_{1pt}]$, $t = 1, \dots, T$ and not by considering each univariate series x_{ijt} ($j = 1, \dots, p$), ($t = 1, \dots, T$) separately. Another feature that characterizes the warping function wd_i and that may be useful in applications with meteorological data, is the possibility to define the maximum number of time-lags between the physical time and the warped time. Indeed, considering for example daily series, only similar events that occur in nearby days are likely to be expression of the same feature (for example, a peak or an extreme value, in a certain period) and should be aligned. Salient events that occurs in days which are faraway, should be considered as two “different” features in the two series and should not be aligned. The maximum number of days between the timing of two events that are likely to be logically compared depends on the application and on the aim of the data analysis. In general, if u is the maximum number of lags for which we assume the same event may be timed differently in the different series, the simplest strategy is to introduce the following “windowing condition” in the dtw algorithm:

$$d_k = d(\mathbf{x}_{1r}, \mathbf{x}_{2c}) \text{ with } |r - c| \leq u$$

We refer to [11] for the definition of more refined constraints on the warping path, aimed at preventing unrealistic warping.

The dtwc dissimilarity matrix may be used for classifying time series with the following hierarchical methods: the average linkage, the complete and the single linkages. The centroid method is not appropriate, since the dendrogram obtained with this method with a dissimilarity measure is a non-monotonic cluster tree.

3 Classification of Meteorological Time Series

We perform a cluster analysis of atmospheric measurements gathered by a historical weather station in the urban area of the province of Modena, in the Emilia Romagna Region (Northern Italy). The station is the geophysical observatory of Modena. Even though the weather station does not conform to the W.M.O. regulations for the position of the instruments (which have been emitted many years after the construction of the geophysical observatory) it does permit the collection rainfall data, in the same location, from 1831. Information about the history of the geophysical observatory may be found in the web page <http://www.ossgeo.unimo.it>. Here we only report the main coordinates of the station:

- Boreal latitude: $44^{\circ}38'50.76''$
- East longitude from Greenwich: $10^{\circ}55'45.50''$
- Height of the barometric cockpit from the sea level: 64.2 m
- Height of the rain gauge from the ground: 41.9 m
- Height of the ground from the sea level: 34.6 m

The (cross-sectional) mean values and the maximum values of the total rainfall for the day (in mm) of the period 1831–2008 are reported in Fig. 4. The minimum daily value is always equal to 0. In average, the total amount of rain in a day is less than 4 mm and reaches the highest peaks in October and November and the minimum values in August. The pattern of the maximum values is different: salient peaks are present in quite all months. In some years, the total rain in a day has reached values higher than 75 mm. The series reported in Fig. 4 shows that the variable has a high variability between years and between days. There are many years presenting anomalous extreme values and it is clear that the cross-sectional mean underestimates both the value of the peaks and the order of magnitude of the phenomenon. We consider the available “three way” data set, with $p = 3$. The three variables are:

- X_1 : minimum air temperature (in Celsius degree)
- X_2 : maximum air temperature (in Celsius degree)
- X_3 : total rainfall (in mm)

Air temperature is known only from the year 1861. We then cluster 148 sequences: the years 1861 to 2008. We perform a cluster analysis of the 148 years on the basis of the minimum temperature, the maximum temperature, and the total rainfall for the month. We will refer to these data, with $T = 12$, as monthly values of X_1 , X_2 , and X_3 . Figure 5 reports time series of the minimum, the maximum, and the (cross-sectional) mean of the monthly values of X_1 , X_2 , and X_3 . This figure shows that Modena experiences a “mediterranean” climate with mild wet winters and hot, less rainy, summers. While the temperature shows a clear seasonal pattern, and both the maximum and the minimum values follow the same average pattern, the amount of rainfall has a more irregular trend and the minimum and the maximum values show different patterns. Before computing the dtw dissimilarity measure, data are

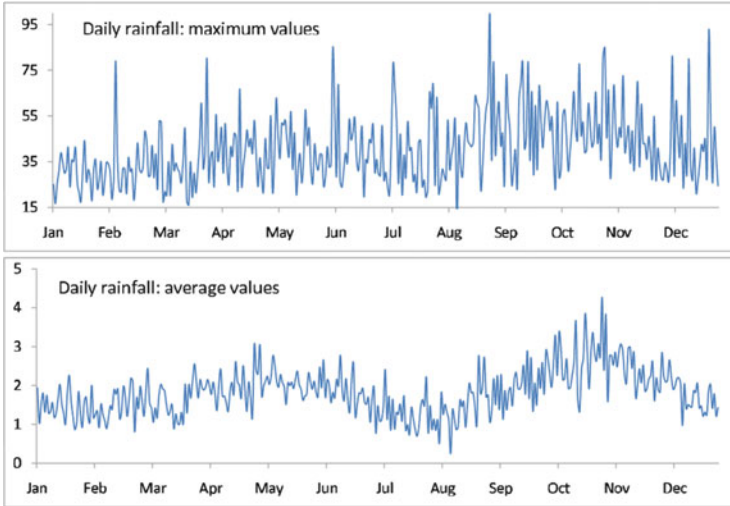


Fig. 4 Time series of the maximum (*top*) and the average (*bottom*) daily values of rain

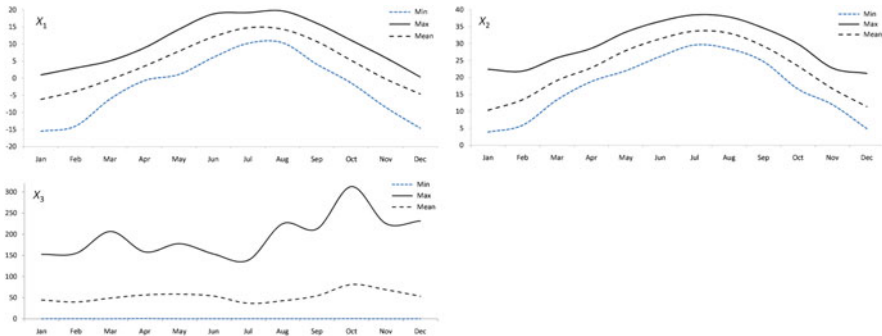


Fig. 5 Time series of the minimum, the maximum and the average monthly values of X_1, X_2, X_3

standardized so that in each t ($t = 1, \dots, T$ with $T = 12$) each variable has 0 mean and unit variance.

In the warping function, we set $u = 2$, allowing for a maximum shift of 2 months. Figure 6 reports dendrograms obtained with the single, the complete, and the average linkages. The trees show that the single and the average linkages exhibit less ability to provide separation than the complete linkage. The single linkage is greatly affected by the “chain effect”. The average linkage is less influenced by this effect but it still tends to aggregate single observations or very small groups in each stage of the hierarchy and many single observations remain isolated till the last stages. The complete linkage readily distinguishes clusters with more than one or two observations. On the basis of the ratio between the within variance and the total variance (which has a relatively high increase from partition in six clusters

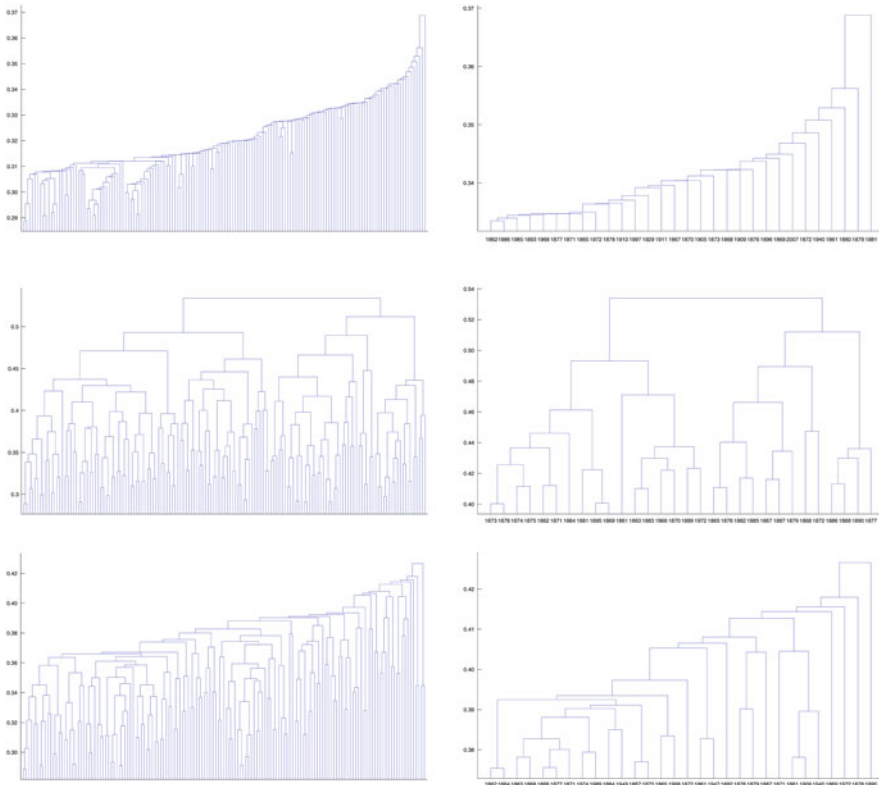


Fig. 6 Data set with $n = 148$, $T = 12$, and $p = 3$: full dendrogram (on the *left*) and dendrogram with 30 leaf nodes (on the *right*) resulting by collapsing lower branches of the full dendrogram, obtained with the single linkage (in the *top*), the complete linkage (in the *middle*), and the average linkage (in the *bottom*)

to partition in five clusters) we consider the classification in six groups. Analyzing cluster means, we see that partitions with less than six groups aggregate years with a very different behavior, while partitions with more than six groups lead to different clusters with similar average behavior. Groups, in the six-clusters partition, are as follows:

- Cluster 1: {1861, 1917, 1941, 1942, 1947, 1953, 1980, 1985}
- Cluster 2: {1863, 1866, 1883, 1889, 1892, 1898, 1900, 1902, 1904, 1905, 1910, 1912, 1914, 1915, 1919, 1920, 1923, 1924, 1925, 1926, 1927, 1928, 1930, 1933, 1934, 1936, 1937, 1939, 1943, 1944, 1951, 1954, 1955, 1956, 1964, 1965, 1969, 1970, 1971, 1972, 1973, 1974, 1977, 1978, 1982, 1984, 1986, 1991, 1996}
- Cluster 3: {1868, 1929, 1938, 1963, 1993, 2002}
- Cluster 4: {1865, 1867, 1869, 1870, 1872, 1876, 1879, 1882, 1885, 1887, 1890, 1896, 1897, 1911, 1913, 1916, 1921, 1931, 1948, 1952, 1957, 1958, 1961, 1967, 1976, 1979, 1981, 1987, 1988, 1990, 1992, 2006}

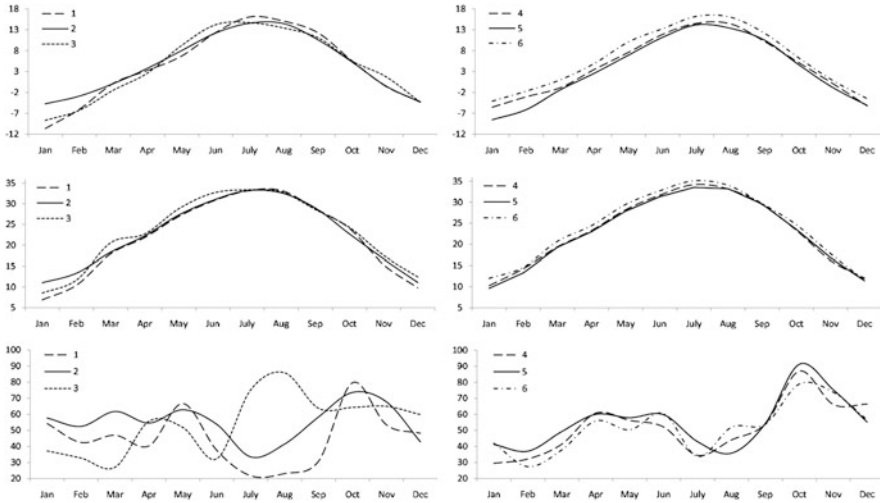


Fig. 7 Data set with $n = 148$, $T = 12$, and $p = 3$: group means in the partition in six clusters. Variable X_1 is reported in the *top*, variable X_2 in the *middle*, and variable X_3 in the *bottom*

Cluster 5: {1862, 1864, 1871, 1873, 1874, 1875, 1877, 1878, 1880, 1881, 1884, 1888, 1891, 1893, 1894, 1895, 1899, 1901, 1903, 1906, 1907, 1908, 1909, 1918, 1922, 1932, 1935, 1940, 1949, 1959, 1960, 1962, 1989}

Cluster 6: {1886, 1945, 1946, 1950, 1966, 1968, 1975, 1983, 1994, 1995, 1997, 1998, 1999, 2000, 2001, 2003, 2004, 2005, 2007, 2008}

Cluster means are reported in Fig. 7. Clusters 1 and 3 represent two small groups with anomalous years. Cluster 1 groups together former years (the most recent one is 1985), which are characterized by low maximum temperatures in quite all months, by a very dry summer season and dry months in the second part of autumn. This kind of climate is completely absent in the two last decades. A similar pattern characterizes group 5, in which are clustered several years from 1962 to 1989. In this group, the minimum temperatures are very low, the summer season is dry but the autumn months are extremely wet. Cluster 3 groups together 6 years (with the recent 2002) with a large amount of rain in the summer season and relatively dry spring months. The minimum and maximum temperatures in these years are in line with the average values. Cluster 6 groups many of the most recent years and the cluster means may be considered as representative of the actual climate situation. This group is characterized by high maximum and minimum temperatures and by a relatively large amount of rain in summer, in autumn, and in the beginning of the winter season. The time series of the group means (as long as the composition of the clusters) lead to the evidence that a climate change is present, at the beginning of the twentieth century. Both the minima and the maxima temperatures are higher, all over the years, and the seasonality in the rain is less evident, since the average amount of rain shows less variability across months.

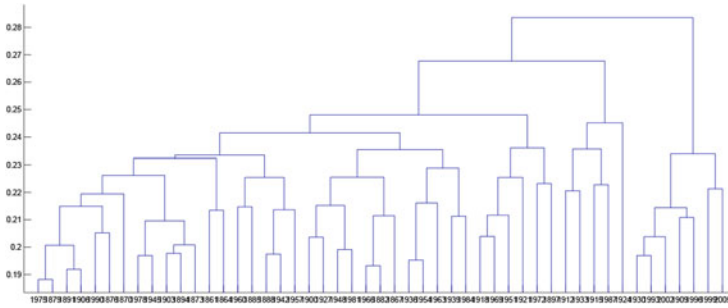


Fig. 8 Data with $n = 49$, $T = 12$, and $p = 3$: dendrogram obtained with the complete linkage

In order to gain insights into the climate change, we perform a second analysis, considering sequences of 3 years. Each series has 36 monthly values ($T = 36$) and $n = 49$. The first series is the triennium 1861–1863, the last series is the triennium 2005–2007. The label of each series is the second year (for example, for the first series the label is 1862 and for the last series the label is 2006). We consider triennium in order to allow a larger shift in the warping function and to allow the shift for the month of January (for the second and the third year) and for December (for the first and the second year). Indeed, considering series of 1 year, the warping in the winter months of January and December is not possible. We set $u = 3$ (the same length of a season). Figure 8 reports the dendrogram obtained with the complete linkage. Here again, the complete linkage seems less affected by the “chain effect” than the single and the average linkages and the tree shows the presence of well-separated clusters. We consider partitions in six and three groups. The cluster means of partition in six groups are shown in Fig. 9 and the group memberships are:

- Cluster 1: {1861, 1900, 1912, 1915, 1918, 1921, 1924, 1930, 1933, 1936, 1939, 1951, 1954, 1960, 1963, 1969, 1972, 1975, 1978 }
- Cluster 2: {1864, 1888, 1891, 1897, 1903, 1942, 1945, 1957, 1966, 1984, 1987, 1990 }
- Cluster 3: {1879 }
- Cluster 4: {1867, 1870, 1885, 1948 }
- Cluster 5: {1873, 1876, 1894, 1906, 1909, 1927 }
- Cluster 6: {1882, 1981, 1993, 1996, 1999, 2002, 2005 }

This partition reveals the presence of an outlier, the triennium 1878–1880 in group 3, which is characterized by extreme (both very high and very low) values in the temperatures and in the rain. This group is merged with group 4 in partition in three clusters. Group 4 contains early years and is characterized by very low temperatures in winter and large amounts of rain in spring and autumn. Groups 1, 2, and 5, contain non-recent years and are merged together in partition in three clusters. The time series of the average values of these groups are smoother than the other series: the seasonality in the temperatures is more evident and the amount of rain

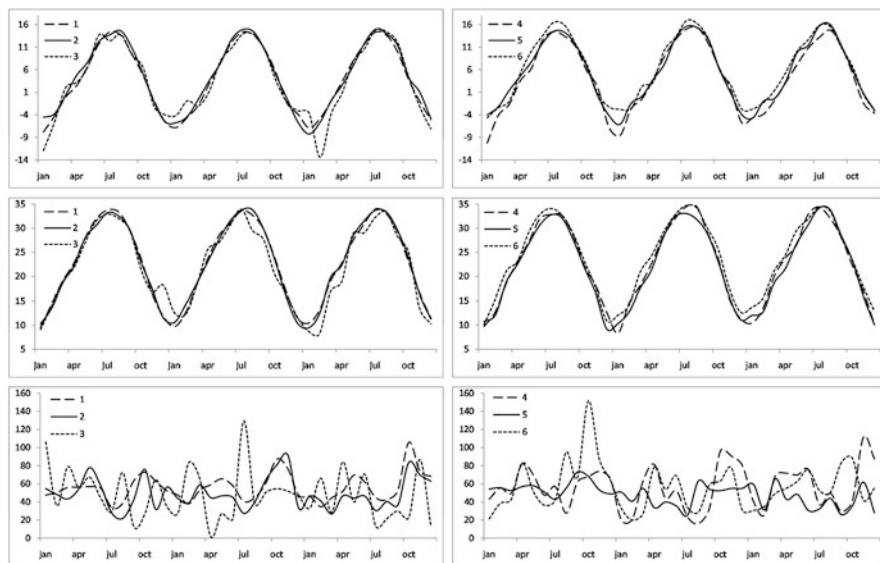


Fig. 9 Data set with $n = 49$, $T = 12$, and $p = 3$; group means in the partition in six clusters. Variable X_1 is reported in the *top*, variable X_2 in the *middle*, and variable X_3 in the *bottom*

across months presents less variability. Group 6 contains recent years. It remains a single group in partition in six clusters and it is not merged with other groups until the top level of the hierarchy. This feature gives evidence of the peculiarity of the years contained in the group. The average values of the temperatures (both the minima and the maxima) are higher than the values in the other groups. In particular, the minima temperatures are much higher than in the other groups. The amount of rain is greatly variable across months and shows anomalous peaks in the first year of the triennium. In general, the amount of rain is higher around April and October and the summer months are wetter than in other groups.

The climate change is more evident in this second analysis, since all recent years (after 1991) are clustered together. The group containing these years remains isolated until the last level of the dendrogram and the times series of the average values show peculiar patterns.

Acknowledgements This study was carried out under the research program PRIN 2010–2011 (grant 2010JHF437) funded by the Italian Ministry of Education, University, and Research.

References

1. Chu, S., Keogh, E., Hart, D., Pazzani, M.: Iterative deepening dynamic time warping. In: Second SIAM International Conference on Data Mining (2002)
2. Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York (2006)

3. Keogh, E.: Exact indexing of dynamic time warping. In: 28th International Conference on Very Large Data Bases, Hong Kong, pp. 406–417 (2002)
4. Keogh, E., Pazzani, M.: Scaling up dynamic time warping for data mining applications. In: Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Boston, pp. 285–289, 20–23 August 2000
5. Keogh, E., Pazzani, M.: Dynamic time warping with higher order features. In: First SIAM International Conference on Data Mining (SDM'2001), Chicago (2001)
6. Milligan, G.W., Cooper, M.C.: A study of standardization of variables in cluster analysis. *J. Classif.* **5**, 181–204 (1988)
7. Ramsay, J.O.: Functional components of variation in handwriting. *J. Am. Stat. Assoc.* **95**, 9–15 (2000)
8. Ramsay, J.O., Dalzell, C.J.: Some tools for functional data analysis (with Discussion). *J. Roy. Stat. Soc. Ser. B* **53**, 539–572 (1991)
9. Ramsay, J.O., Li, X.: Curve registration. *J. Roy. Stat. Soc. Ser. B* **60**, 351–363 (1998)
10. Ramsay, J.O., Silverman, B.W.: *Functional Data Analysis*, 2nd edn. Springer, New York (2005)
11. Ratanamahatana, C.A., Keogh, E.: Making Time-series classification more accurate using learned constraints. In: Proceedings of SIAM International Conference on Data Mining (SDM '04), Lake Buena Vista, pp. 11–22, 22–24 April 2004
12. Silverman, B.: Incorporating parametric effects into functional principal components analysis. *J. Roy. Stat. Soc. Ser. B* **57**, 673–689 (1995)
13. Wang, K., Gasser, T.: Alignment of curves by dynamic time warping. *Ann. Stat.* **25**(3), 1251–1276 (1997)
14. Wang, K., Gasser, T.: Synchronizing sample curves nonparametrically. *Ann. Stat.* **27**(2), 439–460 (1999)