

ANALYSIS OF FOOD CONSUMPTION IN EUROPE VIA TIME SERIES CLUSTERING

Margherita Gerolimetto, Christine Mauracher

1. Introduction

The second half of the twentieth century has been the theatre of an astonishing rebirth after WWII, which brought to unprecedented generalized socio-economic and wealthy levels, especially for Western countries. Such great economic changes that took place in this period profoundly alter not only the economies of Western countries but many aspects of their social life allowing people to improve their lifestyles and therefore their consumption.

In the last thirty years, the evolution of food consumption in European countries has been analysed by several authors and some general conclusions can be derived from those works. Some focused on the general picture (Blandford, 1984; Gil, et al. 1995; Gracia and Albisu, 2001), others analysed more specific regions (Elsener and Hartmann, 1998; Pingali, 2006) and finally other studies were conducted on the evolution of the food categories (Grigg, 1996; Schneider 2002).

The central aim of this work is to analyse how food consumption patterns have changed from 1961 to 2009 in Europe through the elaboration of FAO data via a statistical tool that, to our knowledge, had never been employed in this kind of studies, *i.e.* the cluster analysis of time series. This statistical tool allows analysing the trend followed by the thirteen food categories in each of the 39 European nations and to group the countries characterized by a similar evolutionary path.

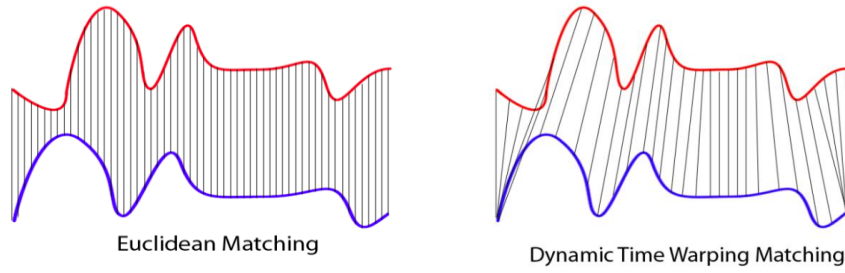
The paper is structured as follows: the second section illustrates clustering time series, the third section presents the data set and the results. The fourth section concludes.

2. Clustering time series with the Dynamic Time Warping (DTW)

To cluster time series there are two approaches (Liao, 2005). In the first approach, existing methods for clustering static data are modified in such a way

that time series data can be handled. In the second approach, time series data are converted into static data so that the traditional algorithm for clustering can be directly used. In this work the former approach has been chosen, also known as raw-data-based approach. It works directly with raw time series data and its major peculiarity, compared to static data cluster analysis, is in that the usual similarity/distance measures for static data are replaced with other measures, specific for time series. In particular, in order to overcome the flaws of the Euclidean Distance, in this paper we employ the Dynamic Time Warping (DTW) distance that, instead, allows an elastic shifting in the time axis, to adjust time series sequences that are similar but out phase, as illustrated in Figure 1.

Figure 1: *Euclidean vs DTW distance*



The principle of the DTW algorithm is to stretch and squeeze locally the two time series, as shown in Figure 1, so that the correct distance between the two time series must be calculated by summing the elements that have been aligned, i.e. between points in time that are in correspondence. Searching for the best match is the central point of the DTW. At the beginning this was done with forms of approximate indexing (among the others, Park et al. 2001), but more recently Keogh and Ratanamahatama (2004) introduce a new method for exact indexing in DTW.

At the core of the DTW technique is the concept of the warping curve $\phi(k) = (\phi_x(k), \phi_y(k))$, $k=1, \dots, n$, where $\phi_x(k) \in \{1, \dots, N\}$ and $\phi_y(k) \in \{1, \dots, N\}$

The warping function maps the time index of time series x_t and y_t and the average cumulated distortions between the two warped time series is

$$d_\phi(x, y) = \sum_{k=1}^T d(\phi_x(k), \phi_y(k)) \frac{m_\phi(k)}{M_\phi}$$

Where $m_\phi(k)$ is a weight and M_ϕ is the normalization constant which guarantees

that comparisons along different time path of the cumulated distortions. The optimal path is the one such that

$$D(x, y) = \min_{\phi} d_{\phi}(x, y)$$

Among all possible warps, only the reasonable ones are of interest, i.e. those such that the deformation of the time axes of x_t and y_t makes them resemble to each other as much as possible. With this in mind, some constraints are usually imposed, for instance monotonicity, to preserve the time order.

3. Empirical analysis with time series clustering

This study examines the 39 European countries, whose territories belong to Europe and with the boundaries and names used internationally in 2009.¹

The dataset used in this work contains information on the average daily per capita calorie (Kcal) consumption of 14 food aggregates² annually detected from 1961 to 2009. Data come from the Food Balance Sheets calculated by FAO³ and the food aggregates are: alcoholic beverages, animal fats, cereals, eggs, fish and seafood, fruits and tree nuts, meat and offals, milk, miscellaneous, starchy roots and pulses, stimulants and spices, sugar and sweeteners, vegetable oils and oilcrops, vegetables.

The use of caloric equivalents facilitates the analysis, since they are assumed to be common across countries for a given food product and it simplifies the cross country comparisons. However, the use of caloric weights implies greater emphasis on food products which are high in calories per unit e.g. fats, oils, cereal and starchy foods (Blandford, 1984).

We now move to the results of the analysis, using the hierarchical agglomerative clustering algorithm and the DTW distance, implemented, in particular, with the complete linkage method. Among all food categories, here for

¹ They are: Albania, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Moldova, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, The Former Yugoslav Republic of Macedonia (the FYROM), Ukraine and United Kingdom.

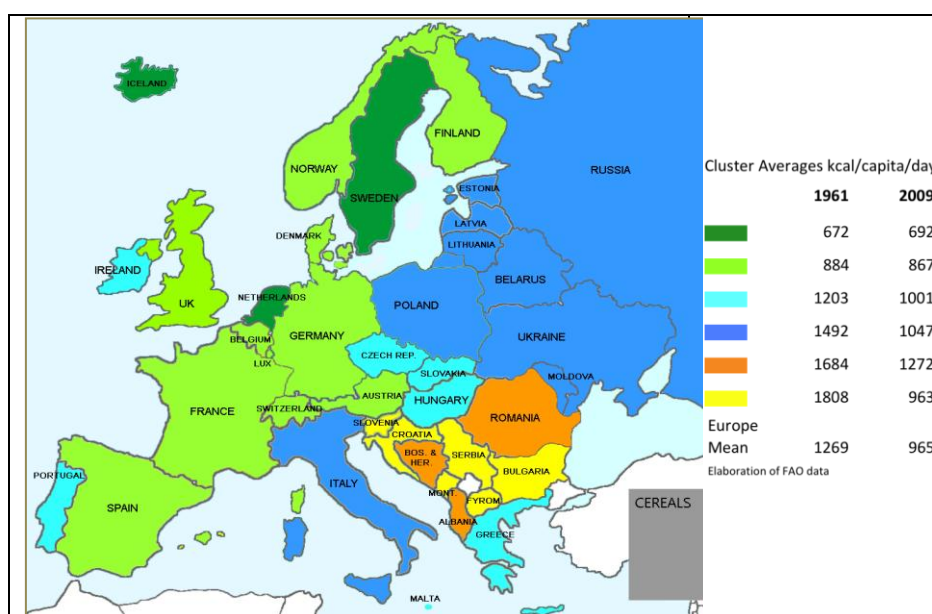
² The sum of these fourteen categories gives the 100 per cent of the daily total caloric intake per capita of every country analysed in this work.

³ It is important to clarify that these are not consumption data, but estimates of the available supplies, calculated from national statistics of production with the exclusion of exports, food put into stores and used for seed, feed and industrial purposes. A further reduction of 10% is made for waste estimation between farms and retail outlets. The final figures are food supplies available at a retail level in calories per capita.

space reasons, we focus only on cereals and meat, i.e. two important categories in the European diet, but other results relatively to the other food categories are available upon request from authors.

As for cereals consumption, the dendrogram has been cut into six groups and to each of them it is assigned a colour in order to obtain the European map of Cereal consumption (Figure 2).

Figure 2: Map of Cereal Consumption in Europe (kcal/capita/day)

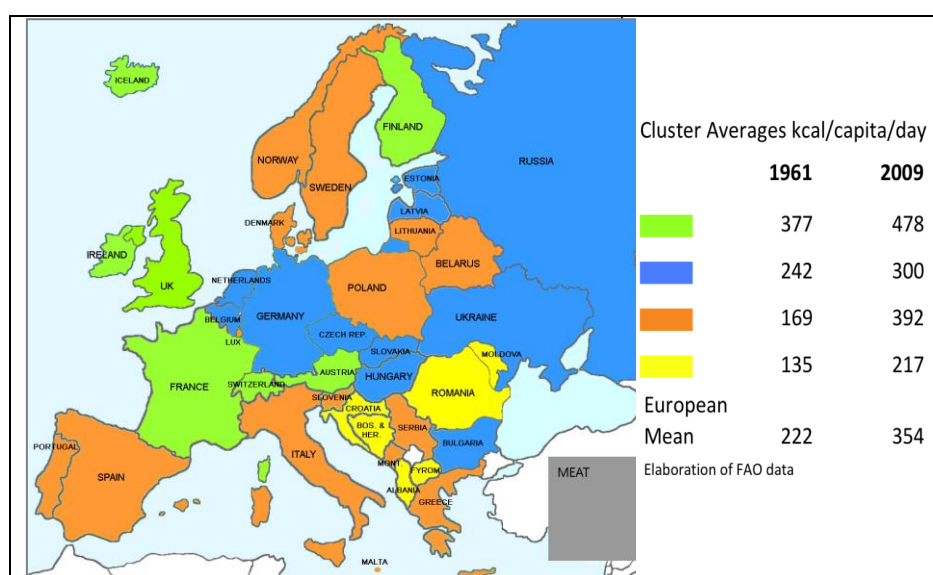


The countries painted in blue share almost the same pattern: they start with a high consumption, above the European mean, and the values decrease constantly over the period stabilizing in the last years around 1100 kcal/capita/day, still larger than the European mean. It is worth noticing that the majority of the members of this group are countries formerly belonging the Soviet Bloc and located in the North East of Europe. Unexpectedly, Italy also belongs to this cluster although it follows a slightly different path, with a sharp drop in early 1980s and from that point until nowadays the values are almost stable. Countries belonging to the light blue group follow a decreasing but less steep path. The light green group includes countries following a U-shape curve evolution. In particular, in the early years the cereal intake was lower than the average, however, with the amazing growth of the last decades, the cereal consumption increased and got closer to the European

mean.⁴

Moving to meat consumption, via the clustering analysis, yet again hierarchical agglomerative clustering algorithm, implemented, in particular, with the complete linkage method and the DTW distance, four clusters have been identified (Fig. 3).

Figure 3: Map of meat Consumption in Europe (kcal/capita/day)



Countries in this cluster mostly belong to Western and Northern Europe; Spain has a fairly different path since its consumption declined for the whole period. The dark green group is characterized by a pretty stable trend, with consumption values generally lower than the European mean ranging around 600 and 700 calories per capita per day. Finally, the yellow group shares a pattern which is slightly decreasing from 1961 to 1991 and shows a dramatic fall in consumption in 1992 and from that shock all the data stabilized to a level that is just above the European mean, except for Bulgaria where cereals consumption decreased over the period. The orange group is composed by countries whose pattern is peculiar but all of them share a high consumption of cereals. It is worth noting that in the 1960s, the diet in the Balkans is highly based on cereals consumption, which accounts for more than 50 per cent of the daily caloric per capita intake. This characteristic still holds afterwards for Albania, in which the share of cereals in the diet is even above

⁴ Such increase could be due to an increasing concern in health and cereals are very good against fibre-deficiency diseases (Grigg, 1996)

60 per cent until the early 1990s.

The green-painted group is characterised by a high consumption of meat, always higher than the European mean. On the opposite side, there is the yellow group including countries whose consumption of meat is always under the mean line. For the first three decades the trend is generally growing for most of the countries, but in the early 1990s there was a sharp collapse in consumption, probably this is linked to the Balkan wars and the following changes in political regimes. Between these two extreme groups there are other two clusters characterised by middle consumption levels: the orange-painted group, including countries whose initial meat consumption was quite low and that reached the European mean and eventually even crossed it over. The blu group includes almost all the countries linked to the Soviet Bloc. They started from middle consumption levels, which were generally a little higher than the orange cluster, and grew until 1990. Then the consumed meat amount dropped in the early 1990s but in the last fifteen years consumption has increased again.

It needs to be pointed out that the increasing pattern is common for almost all the members of the group, but with different rates. It is not by chance that the collapse in consumption happened just after the fall of the Berlin Wall. In most of these countries the consequent transition phase from a planned economy to a market one involved several changes in consumers' purchasing powers and so their food habits. In particular, the consumption of meat is one of the categories which were affected the most since it is a luxury good, due to its very high cost.

5. Conclusions

The main findings of this paper are the similarities and peculiarities in food consumption patterns between 1961 and 2009 in Europe. Food consumption is variably affected by different factors including food availability, food accessibility and food choice, which in turn may be influenced by geography, demography, disposable income, urbanization, globalization, marketing, religion, culture and consumer attitudes (Kearney, 2010). Moreover changes in the distribution system have had an important effect on food consumption patterns. Also trade liberalization is another relevant factor that has lead to changes in food intake. In the considered period several events marked the history of Europe and deeply influenced the eating habits of the European population, e.g. two oil crises, the fall of the Berlin wall, the Balkan wars.

It is important to underline that the data used were not enough to perform disintegrated analyses which would give in further research more clarifying results.

References

- BLANDFORD D. 1984. Changes in food consumption patterns in the OECD area, *Review of Agricultural Economics*, Vol. 11, No. 1, pp. 43-64.
- ELSNER K., HARTMANN M. 1998. *Convergence of Food Consumption Patterns Between Eastern and Western Europe*. IAMO Discussion paper No. 13-1998, (Institut für Agrarentwicklung in Mittel und Osteuropa).
- GIL J. M., GRACIA A., PÉREZ y PÉREZ L., 1995, Food Consumption and Economic Development in the European Union, *European Review of Agricultural Economics*, Vol. 22 , No. 3, pp. 385-399.
- GRACIA A., ALBISU L. M. 2001. Food Consumption in the European Union: Main Determinants and Country Differences, *Agribusiness*, Vol. 17 , No. 4, pp. 469-488.
- GRIGG D. 1996. The Starchy Staples in Food Consumption, *Annals of the association of American Geographers*, Vol. 86, No. 3, pp. 412-431.
- KEARNEY J. 2010. Food consumption trends and drivers, *Philosophical Transactions of the Royal Society B*, Vol. 365, No. 1554, pp. 2793-2807.
- KEOGH E.J, RATANAMAHATAMA C.A. 2004. Exact indexing of Dynamic Time Warping. *Knowledge and Information Systems*, Vol. 7, pp. 358-386.
- LIAO T. W. 2005. Clustering of Time Series Data - a survey, *Pattern Recognition*, Vol. 38, pp. 1857-1874.
- PARK S., KIM S., CHU W. 2001. Segment-based approach for subsequence searches in sequence databases. In: *Proceedings of the 16th ACM symposium on applied computing*, Las Vegas, NV, pp 248-252.
- PINGALI P. 2006. Westernization of Asian Diets and the Transformation of Food Systems: Implications for Research and Policy, *Food Policy*, Vol. 32, No. 3, pp. 281-298.
- SCHNEIDER A. 2002. Overview of the market and consumption of pulses in Europe. *British Journal of Nutrition*, Vol. 88, No. 3, pp. 243-250.

SUMMARY

The purpose of the paper is to investigate the evolution of food consumption in the European countries. In particular the attention will be focused on the food consumption patterns and the major differences occurred in the period between 1961 and 2009. The statistical technique, that has been used, is the Cluster Analysis of Time Series, with the measure of the Dynamic Time Warping (DTW).

Margherita GEROLIMETTO, Lecturer in Economics Statistics, University of Venice

Christine MAURACHER, Professor in Agricultural Economics, University of Venice