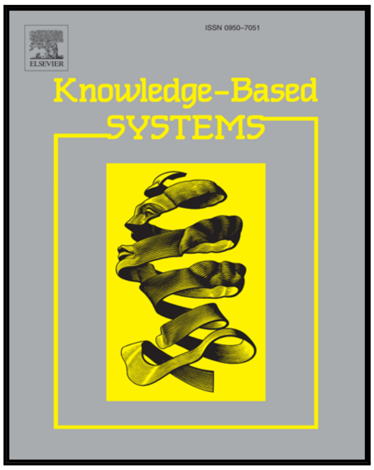


A copula-based clustering algorithm to analyse EU country diets

F.Marta L. Di Lascio, Marta Disegna

PII: S0950-7051(17)30277-0
DOI: 10.1016/j.knosys.2017.06.004
Reference: KNOSYS 3931

To appear in: *Knowledge-Based Systems*

Received date: 30 December 2016
Revised date: 19 April 2017
Accepted date: 2 June 2017

Highlights

- A novel copula-based clustering algorithm is suggested to group EU countries
- Average calories of different food aggregates are used as segmentation variables
- Complex multivariate associations in Countries dietary structures are identified
- Changes towards a (un)common (un)healthier food dietary structure are investigated
- Countries at risk of an increase in obesity and diet-related disease are identified

A copula-based clustering algorithm to analyse EU country diets

F. Marta L. Di Lascio[a,1], Marta Disegna[b,2]

[a] Faculty of Economics and Management, Free University of Bolzano, Piazza Università 1, Bozen-Bolzano, Italy
[b] Faculty of Management, Executive Business Centre, Bournemouth University, 89 Holdenhurst Road, Bournemouth, BH8 8EB, United Kingdom

Abstract

The aim of the paper is to explore the evolution of food diets in 40 European countries according to the common European policies and guidelines on healthy diets. To this end, an innovative clustering method, called CoClust, has been adopted. By means of the copula function, this algorithm is able to find clusters based on the complex multivariate dependence structure of the data generating process, overcoming the limits of classical approaches that cope with only linear bivariate relationships. The analysed database contains information on the average calories from 16 food aggregates in 40 European countries observed over 40 years by the Food and Agriculture Organisation of the United Nations (FAO). Our findings suggest that European country diets are changing, individually or as a group, but not in a unique direction. Central and Eastern European countries are becoming unhealthier, while the tendency followed by the majority of the remaining countries is to integrate the common European guidelines on healthy, balanced, and diversified diets in their national policies.

Keywords: Clustering; CoClust; Healthy diet; Convergence; Dietary energy; EU countries.

1. Introduction

In the literature there is substantial agreement regarding the idea that food consumption patterns, or diets, are changing over time in a non-uniform way, especially showing large spatial variation [1, 2, 3, 4]. However, as regards European Union (EU) countries, [5] discovered an increased homogenisation of diets from 1961 to 2001, even though regional diet differences were still recognisable. This result can be partially attributed to the common food-based dietary guidelines (FBDG) adopted since World War II by EU governments in order to promote healthy diets ensuring adequate daily intakes of both macronutrients

[1] *Email address:* `marta.dilascio@unibz.it`
[2] Corresponding author. *Email address:* `disegnam@bournemouth.ac.uk`

(proteins, carbohydrates and fats) and micronutrients (vitamins and minerals). In 1996 the Food and Agricultural Organisation (FAO) and the World Health Organisation (WHO) published guidelines for the creation of FBDG at the national level, accepted by the EU and subsequently published in 2001. Specifically, WHO/FAO are encouraging and supporting EU countries to develop and implement their own FBDG for healthy, diversified and balanced diets adapted to each country's specific needs (e.g. individual needs, cultural context, locally available foods and dietary customs).

Diets are in fact complex combinations of different food products which do not merely represent regional food consumption patterns but which also describe more widely the social, cultural, political, economic and environmental situation of a country [6].

The direct and indirect effects of diet on both environment and health are in fact so strong and important that it is undoubtedly necessary to do an in-depth analysis of regional food consumption and production patterns [6, 7, 8, 9]. As underlined by [10], everyone needs food to live but a poor diet (in terms of either quantity or quality) can lead to negative consequences on health. Therefore, it is necessary to improve governments' knowledge on where they are and where they are going in terms of dietary patterns in order to prevent and control unhealthy situations before the socio-economic costs fully develop.

To this aim, the main focus of this study is to explore how EU diets have changed between 1970 and 2011 in accordance with the WHO/FAO guidelines on healthy diet. Besides the use of traditional measures of convergence in food consumption, such as gamma and sigma convergence [3, 9, 11, 12], different econometric models [9, 13, 14, 15, 16, 17], cluster analysis and data mining [18, 19, 20, 21, 22, 23, 24] have been adopted to analyse EU countries' progress towards sustainable development and common food consumption patterns. In this study, we suggest the use of *i*) the Mediterranean Adequacy Index (MAI) [25] to measure the health of each country diet, *ii*) the gamma convergence [9] to get the most recent and long period of convergence in food consumption, and *iii*) the CoClust [26, 27] to identify sets of countries characterised by complex associations in their dietary structures.

The CoClust is an innovative model-based clustering algorithm that assumes data are generated by a multivariate copula model [28]. The copula [29] is a well-known multivariate tool for generating multivariate joint distributions with a variety of complex dependence structures. Hence, the CoClust is theoretically able to discover complex multivariate relationships that are not possible to identify using more traditional dependence measures (for example, the linear correlation coefficient is only able to capture linear bivariate dependence relationships). Recently, other clustering techniques based on copula have been proposed in the literature. In particular, based on a probabilistic interpretation of canonical correlation analysis, [30] introduced a copula mixture model to capture dependencies in the joint space of multiple views to perform a clustering of objects. More recently, [31, 32] proposed a clustering method based on a mixture of copulas where each cluster is described by a copula and the copulas are linearly combined. This is a generalisation of the classical model-based approach [33] that uses normal distributions and only accounts for linear dependencies. In addition, [34] suggested a copula-based network clustering techniques in which the aim is to identify clusters of objects sharing a common dependence structure. However, the idea underlying the CoClust and the peculiarities of the clustering it identifies make it different from the other clustering approaches based on copula. In particular, the interpretation of the CoClust is based on within-group independence and the among-group dependence, thus the focus is on the relationships across clusters and not on the relationship within clusters, as the other approaches do.

In computer science and information literature, many studies have approached the mathematical aspects of copulas [see for instance: 35, 36, 37, 38, 39]. At the same time, copulas have been adopted to study the dependence structure of financial markets, i.e. to measure the co-movements among financial time series, and nowadays there is a vast and growing literature on this topic [see for instance the following recent studies: 40, 41, 42, 43, 44, 45, 46, 47]. Copulas have also been used in studies related to applied economics, such as tourism [48, 49, 50] and agriculture. Focusing on economic agriculture, the field of this study, copulas have been adopted to study the co-movements between time series regarding prices for food (corn, soyabean, wheat, and rice) and either oil prices [51] or US dollar (USD) exchange rate [52]. Furthermore, at the micro-level copula models have been integrated to censored equation systems [53] and nonparametric median regression [54] to study meat consumption and total food expenditure respectively. However, to the best of our knowledge, copulas have been never used to perform cluster analysis using food consumption as segmentation variables.

The paper is organised as follows. Section 2 describes the CoClust algorithm and illustrates pros and cons related to its use. In Section 3 data has been presented focusing on the description of the evolution of EU countries' diets towards the MD. Section 4 presents the results with a focus on which countries are evolving towards a (un)healthy diet. The paper concludes in Section by offering some final remarks.

2. Methodology

In order to explore how the EU countries' diets evolved over time with respect to the common European policies and guidelines on healthy diets we follow a three-step process: firstly, we identify a suitable index to measure the health of a country's diet; secondly, we measure the convergence in food consumption among countries through a suitable index; finally, we perform a cluster analysis to identify profiles of countries characterised by complex multivariate relationships.

In the literature, different food indexes have been proposed to evaluate the health of a country diet. In this paper, the Mediterranean Adequacy Index (MAI), developed by [25], has been adopted to assess how close each country diet is to the healthy Mediterranean Diet (MD) over time. The MD is commonly considered a healthy and prudent diet since it is plant-centered (i.e. it is characterised by a high consumption of legumes, whole grains, fruits and vegetables, nuts and seeds) and the consumption of meat and dairy products is moderately low, as also recommended in the WHO/FAO guidelines.

The MAI is easily obtained by dividing the sum of the percentages of the calories from Mediterranean food aggregates (M), by the sum of the percentages of the calories from Non-Mediterranean food aggregates $(\overline{M})$, which is as follows:

$$\mathrm{MAI}_t = \sum_{i \in n} \frac{\sum_{j \in M} y_{ijt}}{\sum_{j \in \overline{M}} y_{ijt}} \tag{1}$$

where y_{ijt} is the per capita per day calories from the j-th food aggregate observed in the t-th year for the i-th EU country. The higher the value of the MAI, the higher the adherence to the MD.

The convergence in food consumption, i.e. the tendency toward homogenisation in nutrient supply among different countries over time [13], has been detected through the coefficient of variation (CV) of total calories from food products computed per each year as follows:

$$\mathrm{CV}_{jt} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{ijt} - \overline{y}_{jt})^2}}{\overline{y}_{jt}} \tag{2}$$

where $\overline{y}_{jt}$ is the average per capita per day calories from the j-th food aggregate observed in the t-th year for all EU countries. As defined by [9], reductions in the CV in food consumption over time are identified as periods of convergence.

Cluster analysis is a data-driven approach that attempts to discover structures and potentially meaningful relationships within data itself, grouping together objects into clusters. The extensive literature of clustering includes both methods based on distance/dissimilarity measures and methods based on probability models [55]. Generally speaking, distance-based clustering techniques group objects into the same cluster on the basis of their similarity computed through a suitable distance or dissimilarity measure between two objects, like the Euclidean distance or the one minus the squared correlation coefficient. Hence, in this case clusters are generated in a way which maximises homogeneity within-cluster and the separation between-cluster. On the contrary, model-based clustering techniques [33] assume that data are generated by a finite mixture of probability distributions. This means that objects are grouped in the same $k-$th cluster if they come from the same specific density function f_k that is generally a Gaussian one. In this case the operational definition of clusters is based on the internal linear dependence among objects. In practice, both distance-based and model-based methods are able to cope only with pairwise and/or linear relationships between objects, but they are not suitable to model multivariate complex dependence. To overcome these limits, it is possible to adopt the CoClust algorithm, a model-based technique that assumes data are generated by a copula function.

2.1. Copula background

Copula function is born with Sklar's theorem [29] that states that every joint distribution function $F(\cdot)$ can be expressed in terms of K marginal distribution function F_k and the copula distribution function C as follows:

$$F(x_1, \ldots, x_k, \ldots, x_K) = C(F_1(x_1), \ldots, F_k(x_k), \ldots, F_K(x_K)) \tag{3}$$

for all $(x_1, \ldots, x_k, \ldots, x_K) \in \bar{\mathbb{R}}^K$ (where $\bar{\mathbb{R}}$ denotes the extended real line). According to this theorem we can split any joint probability function $f(\cdot)$ into the margins and a copula, so that the latter represents the association between variables, e.g. the multivariate dependence structure of a joint density function [56, for details]:

$$f(x_1, \ldots, x_k, \ldots, x_K) = c(F_1(x_1), \ldots, F_k(x_k), \ldots, F_K(x_K)) \prod_{k=1}^{K} f_k(x_k). \tag{4}$$

Such separation determines the modelling flexibility given by copulas since it is possible to decompose the estimation problem in two steps: in the first step margins are estimated; and in the second step the copula model is estimated. The most used estimation method is the two-stage inference for margins method [57] that employs the log-likelihood estimation method to estimates both the parameter(s) of each margin and the copula parameter θ. This method can be used in a semi-parametric approach [58] that does not require distributional assumptions on the margins since these are modelled through the empirical cumulative distribution functions $\hat{F}_k(X_{ki})$ with $k = 1, \ldots, K$. Then, the log-likelihood copula function is used to estimate θ as follows:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log c \left\{ \hat{F}_1(X_{1i}), \ldots, \hat{F}_k(X_{ki}), \ldots, \hat{F}_K(X_{Ki}); \theta \right\} \tag{5}$$

where n is the sample size. In the literature, many different copula models are available [28, for details] but it has been demonstrated that the Elliptical and the Archimedean families are the most useful in empirical modelling. The Elliptical family includes the Gaussian copula and the t-copula: both copulas are symmetric; they exhibit the strongest dependence in the middle of the distribution; and they can take into account both positive and negative dependence since $-1 \leq \theta \leq 1$. As usual, the t-copula is characterised by two parameters, the dependence parameter θ and the number of degrees of freedom, and it converges to a Gaussian copula as the number of degrees of freedom approaches infinity. The Archimedean family, by comparison, enables us to describe both left and right asymmetry as well as weak symmetry among the margins by employing Clayton's, Gumbel's and Frank's model, respectively. Clayton's copula has the parameter $\theta \in (0, \infty)$ and as θ approaches zero, the margins become independent. The dependence parameter θ of a Gumbel model is restricted to the interval $[1, +\infty)$ where the value 1 means to independence. Finally, the dependence parameter θ of a Frank copula may assume any real value and as

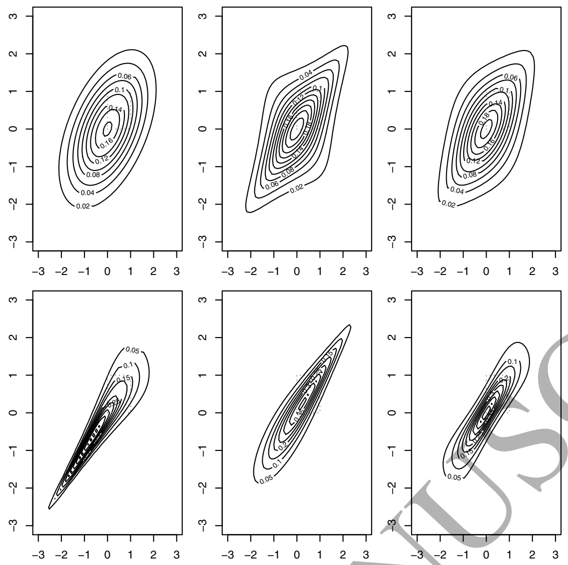


Figure 1: Contour plots of bivariate copula models with normal standard margins and dependence parameter θ such that the Kendall's correlation coefficient is $\tau = 0.7$; upper panel: Gaussian and t-Student copula models for two number of degrees of freedom: 2 and 4; lower panel: Clayton, Gumbel and Frank copula models.

θ approaches zero, the marginal distributions become independent. Figure 1 shows the contour plots of the bivariate density functions defined by the above five copula models with standard normal margins and a level of θ such that the Kendall's correlation coefficient is $\tau = 0.7$. According to the kind of copula model, the value of θ will have a specific meaning. However, it is always true that the greater the value of the dependence parameter, the stronger the association among the margins.

2.2. The CoClust algorithm

The CoClust algorithm assumes that data are generated by a K-dimensional copula function C where each margin F_k is the probability-integral transform of the density function f_k that generates the k-th cluster. Hence, a K-dimensional copula represents a clustering of K clusters and the copula model C describes the shape of the multivariate dependence structure among clusters (margins). Moreover, the copula parameter θ expresses the strength of the multivariate dependence. Consequently, each cluster can be viewed as the realisation of a random variable and it is identified by one (univariate) margin. Having K clusters means having K dependent margins and a copula makes possible to investigate this kind of dependence. Hence, objects in the same cluster are independent and identically distributed realisations from the same marginal distribution while objects across clusters, which can be called profiles, share an inter-cluster multivariate dependence structure, i.e. they have a mutually dependent relationship. Therefore, the CoClust aims to describe the within-cluster independence and the between-cluster dependence instead of the within-cluster homogeneity and the between-cluster separation, as the more traditional clustering approaches.

The starting point of the CoClust algorithm is a standard $N \times q$ data matrix in which N/K are the objects to be grouped in K groups and q are the segmentation variables. The basic idea behind the CoClust, along with a representation of how data are grouped and how the final profiles are identified, is represented in Figure 2. The main steps of the CoClust algorithm are represented in Figure 3 [refer to 26, 27, for more technical details] and described as follows:

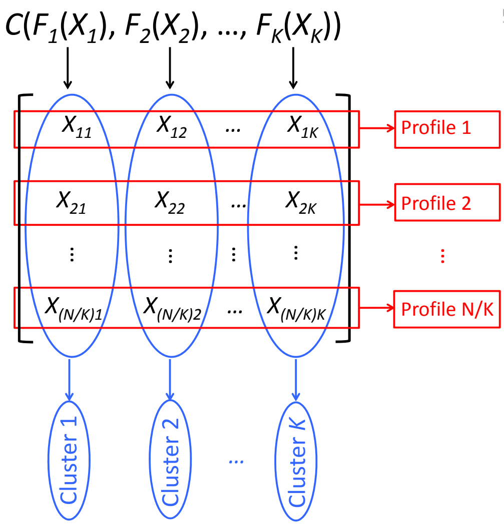


Figure 2: The basic idea of the CoClust algorithm.

1. for $k = 2, \ldots, K_{max}$, where $K_{max} \leq N$ is the maximum number of clusters to be tried:
   (a) select a subset of n_k k-plets of rows/profiles in the data matrix on the basis of the following multivariate measure of association based on pairwise Spearman's ρ correlation coefficient:

$$H(\Lambda_2|\Lambda_1) = \max_{i' \in \Lambda_2} \left\{ \psi_{i \in \Lambda_1} \left(\rho(\mathbf{x}_i, \mathbf{x}_{i'}) \right) \right\} \tag{6}$$

where Λ is a set of row index profiles such that $\Lambda = \Lambda_1 \cup \Lambda_2$, Λ_1 is the subset of profiles already selected to compose a k-plet, Λ_2 is the set of remaining candidates to complete a k-plet, $\mathbf{x}_i$ is the i-th profile, ψ is a selected function among the mean, the median or the maximum;

(b) fit the copula model on the n_k k-plets of profiles/rows through the maximum pseudo-likelihood estimation (for details see Section 2.1);

2. select the subset of n_k k-plets of rows/profiles, say n_K K-plets, that maximises the log-likelihood copula function; hence, the number of clusters K, i.e. the dimension of the copula, is automatically chosen;
3. select a K-plet using the measure in eq. (6) and estimate $K!$ copulas by using the observations already clustered and a permutation of those candidate to the allocation;
4. allocate the permutation of the selected K-plet to the clustering by assigning each observation to the corresponding cluster if it increases the log-likelihood of the copula fit, otherwise drop the entire K-plet of rows/profiles;
5. repeat steps 3. and 4. until all the observations are evaluated (either allocated or discarded).

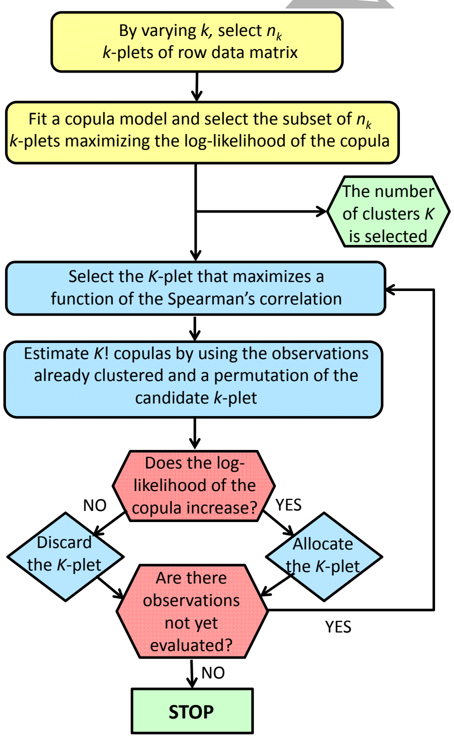


Figure 3: The CoClust algorithm procedure.

In summary, at the first two steps the algorithm selects the optimal number of clusters K while from the second step onwards, it evaluates a $K-$plet of rows at a time allocating the observations to the K clusters in a way that the complex dependence relationships among objects are represented by the K-dimensional copula function. Note that, the algorithm can be slow when the number of clusters is more than 6 and/or the sample size is big since the permutations of the selected $k-$plet have to be computed.

2.2.1. Selection of the number of clusters and the copula model

The CoClust algorithm selects automatically the number of clusters K on the basis of the log-likelihood of the copula function estimated on the subsets of k-plets allocated until a step predefined by the user. However, it is possible to select K post-clustering, that is, on the basis of the whole final clustering. In this case, the number of clusters can be selected by using an information criterion, such as the Bayesian information criterion (BIC, from now on) that, for a copula model m with single parameter, has the following expression:

$$\mathrm{BIC}_{K,m} = -2\log \Pi_{i=1}^{n} c_m \left\{ \hat{F}_1(X_{1i}), \ldots, \hat{F}_k(X_{ki}), \ldots, \hat{F}_K(X_{Ki}); \hat{\theta} \right\} + \log\left((N/K)q\right) \tag{7}$$

where $\hat{\theta}$ is in eq. (5) and $(N/K)q$ is the total number of observations allocated in each cluster (N/K q-dimensional vectors). According to [59], we can compute K as follows:

$$K = \arg\max_{k,m} \left[\frac{\mathrm{BIC}_{k,m} - \mathrm{BIC}_{k-1,m}}{\mathrm{BIC}_{k-1,m}} \right] \tag{8}$$

where m indicates a specific copula model and varies in a predefined set of models and $k \in \{2, \ldots, K_{max}\}$. The copula model used in CoClust is estimated through the two-stage inference for margins method in its semi-parametric version (see Section 2.1). The selected number of clusters K and copula model are the ones that maximize the reduction of the BIC.

2.3. Benefits of the CoClust over other algorithms

The main advantages of the CoClust with respect to more traditional clustering algorithms are as follows:

- it does not require a starting classification to be chosen;
- it does not require the number of clusters to be set a *priori*;
- it is able to capture multivariate and nonlinear dependence relationships underlying the observed data [see 26, 27, for details];
- it does not require the marginal probability distributions to be set as Gaussian;
- it is able to discard irrelevant observations [see 27, for details].

Nowadays copula has become a quite widespread tool in clustering context and in the literature two main approaches became popular. One approach belongs to the classical model-based clustering [33]. This assumes that data are generated by a finite mixture of probability distributions as follows:

$$f(x) = \sum_{k=1}^{K} \tau_k f_k(x) \tag{9}$$

where $f_k(x)$ is the density of an observation x from the k-th component that represents the k-th cluster and τ_k is the probability that an observation belongs to the k-th component. In general, a multivariate normal distribution with mean μ_k and covariance matrix Σ_k is assumed for each component. The parameters of the model in eq. (9) are estimated through an EM algorithm for a different number of clusters and covariance structures and the best model is selected by using the BIC. In the last decade some variants of this approach has been defined using the copula in the mixture. For example, [30] introduced a copula mixture model able to capture dependencies in the joint space of multiple views with the aim of performing a clustering of objects. More recently, [31, 32] proposed a clustering method based on mixture models, where each model in the linear combination is a copula model describing a cluster. With respect to the classical model-based approach, the latest approach is more flexible allowing the detection of linear and non-linear dependencies between objects.

Both the CoClust and the model-based clustering approach assume a probability model underlying the clustering and use an information criteria to select both the number of clusters and the model. However, they are based on a different concept of cluster and they aim at different clustering structures. In the CoClust *i*) a cluster is a set of independent observations, i.e. independent and identically distributed realisations of a univariate model, and not a set of dependent observations modelled through a multivariate model, *ii*) the whole clustering is modelled through one copula model and not through a mixture of copula models, and *iii*) the interest is on the among-group relationship, i.e. the multivariate dependence, and not on the within-group relationship, i.e. the independence. Thus, the main purpose of the CoClust is the identification of dependent groups in which the complex dependence among observations belonging to different groups can be uncovered. Therefore, the interpretation of the clustering is based on the within-group independence and the among-group dependence, i.e. on the relationships across clusters.

A more recent approach is due to [34] who developed a copula-based network clustering technique inspired to the CoClust. Here the aim is to find a partition of objects such that the ones belonging to the same cluster show a dependence structure. Differently from what the CoClust does, [34] look for clusters of dependent objects not assuming independence within clusters. Moreover, [34] do not use a sequential and forward evaluation of objects to be allocated but they directly look for the optimal partition of objects into clusters. Finally, in the copula-based network clustering a distribution function for each margin is not assumed but a cumulative distribution function is estimated for each object.

To sum up, the CoClust is the unique approach able to find *i*) dependent clusters of independent objects and *ii*) profiles of observations across clusters characterised by a complex multivariate dependence relationship.

3. Data

Annual data covering 1970 to 2011 from the 38 countries that constitute the continent of Europe (following the FAO list), plus Cyprus and Turkey, have been considered. Overall, these 40 countries constitute what we will consider from now on the set of EU countries. Average calories per capita per day from different food aggregates have been analysed in this study, since calories have been considered in the literature as a good approximation of food consumption useful to analyse changes over space and time [60]. Data have been obtained from the national food balance sheet of the FAO database [61]. The following 16 food aggregates have been analysed: (1) animal fats; (2) eggs; (3) fish and seafood; (4) meat; (5) milk (excluding butter); (6) other animal; (7) alcoholic beverages; (8) cereals (excluding beer); (9) fruits (excluding wine); (10) potatoes; (11) pulses; (12) sugar and sweeteners; (13) soyabeans; (14) vegetable oils; (15) vegetables; (16) other vegetables. Overall, these food aggregates make up the diet of any country included in the study and can be grouped into different aggregates depending on the research objectives. In this study, two different classifications have been considered: animal (1-6) vs. vegetables (7-16); Mediterranean (3, 8-11, 13-16) vs. non-Mediterranean (1, 2, 4-6, 12). Alcoholic beverages have been excluded from the Mediterranean/non-Mediterranean classification since it was impossible to separate the calories from healthy and unhealthy food items.

3.1. Healthiness of the EU diet

The average food consumption in EU countries was 3380 calories/capita/day in 1970 while it was 3534 calories/capita/day in 2011. The trend of the average EU food consumption, represented in Figure 4(a), shows a general increase over the years interrupted by a strong decrease in the early '90th. Comparing the average calories computed on all EU countries (black line figure 4(a)) with the average calories computed on the subgroup of countries in which the former Soviet Union/block and the former Yugoslavia are excluded (grey line figure 4(a)), it seems clear that the main cause of this decrease is the dissolution of the former Soviet Union/block (December, 1991) and of the former Yugoslavia (1991-1992).

As we can observe in Figure 4(b), in the EU countries the average proportion of calories derived from animal products is lower than the proportion derived from vegetable products in all years. Over time, three major trends in the proportion of calories derived from animal products can be observed, confirming what was found by [60] for the period 1970-1990.

To perform an in-depth analysis of the adherence of the EU diet to the MD, and therefore to a healthy diet, the MAI has been computed over time as described in section 2. Observing Figure 4(c), it is possible to note that the subgroup of EU countries where the Soviet block and the former Yugoslavia are excluded is characterised by a lower MAI, i.e. it is less adherent to the MD diet than the overall set of EU countries. Moreover, two major trends characterised the MAI: a downward trend starting in 1970 and ending in 1983; and an upward trend (the EU countries show an higher steepness when compared with the subgroup of EU countries selected) from 1984 to 2011. This last result can be partially described by the efforts made by EU governments to educate people towards the adoption of healthy diets and good lifestyle practices. Finally, it is particularly interesting to observe that the highest degree of adherence of the EU diet to the MD is observed in 2010-2011 and this value is close to the one observed 40 years before, in 1970.

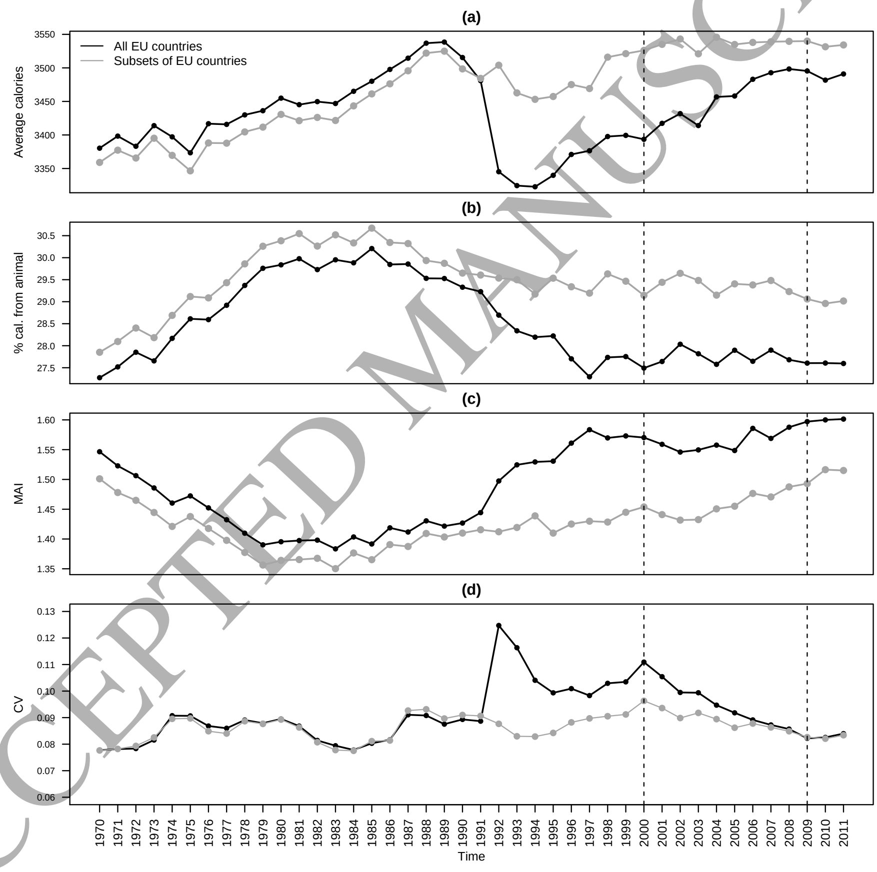


Figure 4: Evolution of average food consumption (daily calories), average proportions of calories from animal products (%), MAI and CV trends in EU countries.

3.2. Convergence among EU dietary structures

The evolution of the CV for the EU diet, computed as described in section 2, is plotted in Figure 4(d). The period 1975-1984 is characterised by a reduction in the CV in food consumption [this behaviour is defined gamma convergence by 9] among EU countries, while in 1985-1992 we observe an increase in the CV ending with a strong peak. As observed for the average calories time series, the peak in the CV trend can also be attributed to the particular political situation experienced by the former Soviet Union/block and the former Yugoslavia in the 1990s. From 1992 to 2000 a period of instability is registered while from 2000 to 2009 we observe an overall tendency of gamma convergence. Finally, in 2011 a slight increase in the CV is observed and this can be considered as a sign that the trend is about to change towards a divergence in food consumption patterns among EU countries. It is interesting to note that, as for the MAI, the CV values observed in 1970 and in 2010-2011 are similar, meaning that the current EU situation regarding both homogenisation in food consumption and adherence of the EU diet to the MD is close to the EU situation observed 40 years ago.

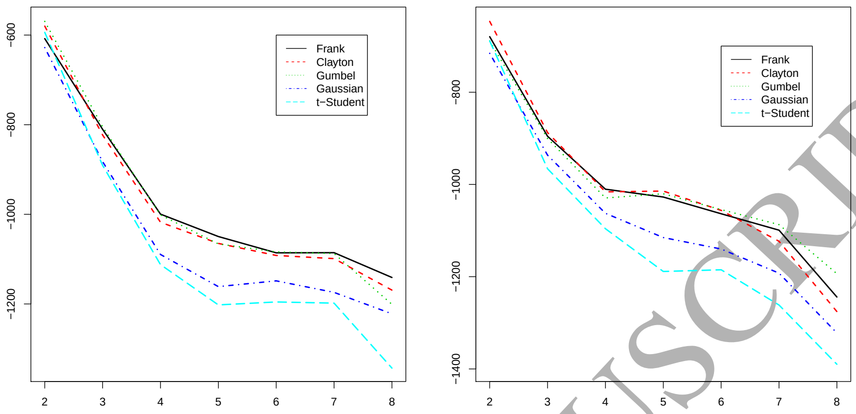
In the following Section 4, the discussion will focus on the most recent and long gamma convergence period in food consumption observed across EU countries, i.e. the period from 2000 to 2009. It is important to observe that also the MAI increase during this period, meaning that EU countries are converging towards a healthier diet.

4. Results of clustering analysis

The CoClust algorithm has been run separately on data collected in 2000 and 2009 to have an in-depth understanding of the food dietary characteristics across EU countries during the identified convergence period. The algorithm has been implemented in the R package `CoClust` which is available on CRAN [62]. The number of clusters (`dimset`) has been set to vary from 2 to 8. The sample size of the set for selecting the number of clusters (`noc`) was 2 and the function in ψ in eq. (6) was the median. As for the estimation task, we used the empirical cumulative distribution function for margins (`method.ma="empirical"`) and the maximum pseudo-likelihood estimation method for the copula (`method.ma="ml"`). Moreover, the algorithm has been run for all the Elliptical and the Archimedean families of copula models setting the degrees of freedom of the t-copula equal to 2 and 4.

In both years, the number of clusters K and the most suitable copula model have been selected by using the BIC as in eq. (8). Figure (5) shows the value of the BIC for any partition from 2 to 8 clusters and for 5 different copula models. In both years, when $k = 2$ the Gaussian copula is the best copula model, while for $k > 2$ the $t-$Student copula with 2 degrees of freedom is the copula model that allows us to obtain smaller values of the BIC. The lines of the $t-$Student show the typical elbow when $k = 5$ indicating that the decrement of the BIC is maximum. Hence, the selected number of clusters in each year is $K = 5$ and the most suitable copula is the t-Student with 2 degrees of freedom.

Figure 5: BIC value by varying copula models (y-axis) and number of clusters (x-axis) for the two years under investigation.



(a) 2000 (b) 2009

In both years, the CoClust algorithm allocates all countries, meaning that 8 profiles of 5 countries each have been identified. It is important to remember that the dietary structure of countries in the same cluster are independent and identically distributed while the dietary structure of countries in the same profile are dependent, i.e. countries in the same profile have a mutual multivariate structure of dependence. The 8 profiles obtained in each year are shown and compared in Table 1. Looking at the two-way table, it is possible to observe that profiles are made up of two parts: one part, called static from now on, comprises the countries characterized by common changes in dietary structure such that any country does not change profile from 2000 to 2009 (groups of countries located on the main diagonal of Table 1); the other part, called dynamic from now on, comprises countries with a dietary structure dependent on different countries in different years (single country or groups of countries located outside the main diagonal of Table 1). Summing up, it is possible to identify 10 static aggregates of countries (for the sake of simplicity labelled SAs from now on). The remaining 11 countries, i.e. Hungary (H), Turkey (TR), Serbia-Montenegro (S_M), Czech Republic (CZ), Italy (I), Poland (PL), the United Kingdom (UK), Latvia (LV), Croatia (HR), Malta (M) and the Republic of Moldova (MD), constitute the dynamic part, moving from one profile to another and likely to embrace different diet compositions over time.

Figure 6 maps each element of Table 1, i.e. SAs and each country of the dynamic part, providing a spatial visualisation of the clustering results. It is interesting to note that, even though the geographical parameter is not always a valid criterion for grouping countries, some SAs (1, 2, 3, 8, and 9) are made up by neighbouring countries.

Since the set of countries that make up the profiles changes over time, it is

Table 1: EU countries profiles 2000-2009.

2009 \ 2000	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5	Profile 6	Profile 7	Profile 8
Profile 1	Portugal Finland Norway Iceland Sweden (SA1)							
Profile 2		France Austria Switzerland Germany (SA2)						Croatia
Profile 3			Greece Slovenia Albania FYROM* (SA3)			Italy		
Profile 4				Denmark Belgium Slovakia (SA4)		Poland		Malta
Profile 5				Serbia-Montenegro	Luxembourg Cyprus Spain (SA5)		Latvia	
Profile 6		Hungary		Czech Republic	Ireland Bulgaria (SA6)	UK		
Profile 7						the Netherlands Romania (SA7)	Russian Federation Ukraine (SA8)	Republic of Moldova
Profile 8			Turkey				Belarus Estonia (SA9)	Lithuania Bosnia-Herzegovina (SA10)

*The Former Yugoslav Republic of Macedonia.

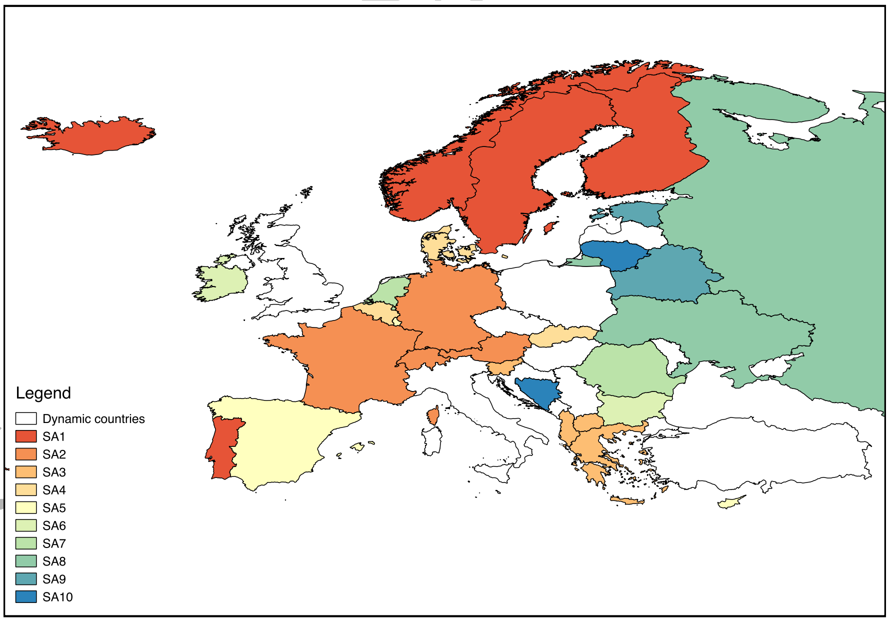


Figure 6: Geographical distribution of static and dynamic aggregates of countries.

not meaningful to study the evolution of profile diets. Therefore, the following analysis will focus on diet evolutions of SAs and each country that belongs to the dynamic part of the profiles.

4.1. Who converges towards a healthy diet?

Diet compositions per SAs and single countries belonging to the dynamic part over time are reported in Tables A.1 and A.2 in Appendix A.

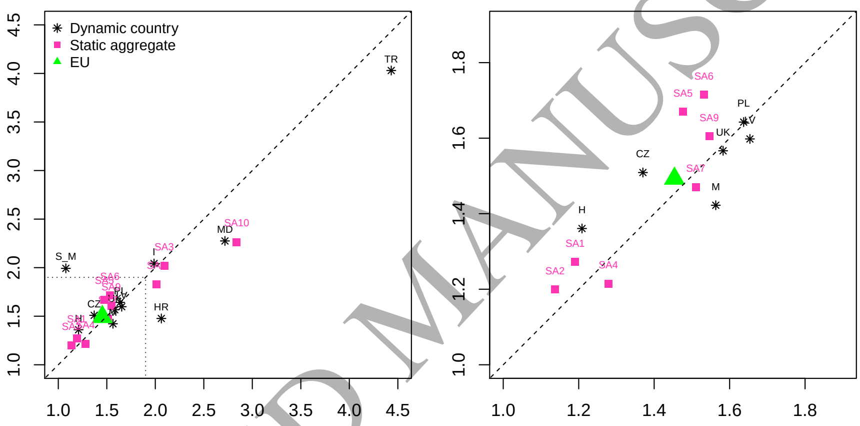
To evaluate how healthy the diets of SAs and single dynamic countries are and how each diet has evolved over time, the MAI has been computed and represented in Figure 7. Single countries and SAs above the main diagonal have experienced an increase in MAI and the higher the vertical distance to the diagonal, the healthier the diet over time. Tables 2 and 3 offer more information on changes in MAI respectively for SAs and countries that belong to the dynamic aggregate of countries. In particular, the proportions of calories from Mediterranean and non-Mediterranean aggregates have been computed and tests for proportions have been calculated to identify significant changes over time. Moreover, percentage changes in MAI (%$\triangle$MAI) have been calculated and included at the end of Tables 2 and 3.

Among the SAs, the SA7 shows the lowest, in absolute value, percentage change in MAI. Furthermore, SA7 is the only aggregate of countries characterised by non-significant changes in the proportions of calories from both Mediterranean and non-Mediterranean aggregates, meaning that this aggregate of countries does not significantly change its diet towards either a healthier or unhealthier diet. All the remaining SAs experienced a significant change in their diet towards either a healthier diet (SA1, SA2, SA5, SA6, and SA9) or an unhealthier diet (SA3, SA4, SA8, and SA10). In particular, SA5 and SA6 experienced the highest percentage increase of the MAI (respectively 13% and 12%), mainly attributable to an increase in the proportion of calories from vegetable oils, cereals (excluding beer), and pulses. On the other hand, SA10 shows the highest deterioration of its diet (-20%), mainly due to an increase in the proportion of calories from meat, sugar and sweeteners.

From the analysis of the dynamic part (Table 3), it is possible to recognise a group of countries that did not experience significant movements over time towards or away from the MD, namely Italy, Latvia, Poland, and the United Kingdom. On the contrary, Czech Republic, Hungary, and Serbia and Montenegro changed their diets towards a healthier diet while the remaining countries experienced a decrease in MAI, i.e. their diets became less healthy over time. Serbia and Montenegro show the highest percentage increase of the MAI (85%) followed by Hungary (13%). In particular, Serbia and Montenegro experienced an increase in the proportion of calories from cereals (excluding beer), fruits (excluding wine), pulses and other vegetables while Hungary moved from a diet highly characterised by calories from animal products, in particular from meat, to a diet characterised by a higher proportion of calories from cereals (excluding beer) and vegetable oils. In contrast, reducing the proportion of calories from potatoes, other vegetables and cereals (excluding beer), Croatia moved from a vegetables-oriented diet towards an animal-oriented diet and its decrement in

MAI is the highest observed among countries that belong to the dynamic part. Even if in both 2000 and 2009 Turkey was characterised by the highest proportion of calories from Mediterranean aggregates and therefore by the highest MAI values, its diet became less healthy. In particular, the Turkish diet moved from a fruit-pulses oriented diet towards a cereals-vegetables-oriented diet, reducing the proportion of calories from Mediterranean aggregates. Finally, Malta and the Republic of Moldova were used to follow a diet rich in cereals and vegetables, but in 2009 they joined different groups of countries both characterised by more animal products-oriented diets, i.e. animal fats and meat.

Figure 7: MAI value per country aggregate in 2000 (x-axis) and in 2009 (y-axis) for all aggregates (a) and for a subset of aggregates characterised by MAI values smaller than 1.9 in both years (b).



(a) (b)

Table 2: Mediterranean and Non-Mediterranean proportions over time and SAs.

	SA1	SA2	SA3	SA4	SA5	SA6	SA7	SA8	SA9	SA10
Mediterranean aggregates										
2000	51.84	50.17	65.29	52.91	55.21	56.17	57.48	64.66	58.19	70.46
2009	53.53	51.51	64.93	52.16	58.51	58.93	56.83	61.30	57.17	64.85
$p-$value χ^2_1	0.002***	0.022**	0.544	0.278	<0.001***	0.001***	0.448	<0.001***	0.235	<0.001***
Non-Mediterranean aggregates										
2000	43.54	44.12	31.21	41.38	37.39	36.66	38.05	32.12	37.63	24.81
2009	42.07	42.95	32.17	42.94	35.04	34.37	38.65	33.55	35.62	28.66
$p-$value χ^2_1	0.006***	0.043**	0.095*	0.023**	0.001***	0.005***	0.482	0.084*	0.015**	<0.001***
%△MAI	6.88	5.44	-3.55	-5.00	13.09	11.93	-2.64	-9.22	3.81	-20.32

Notes: $^{***}p-$value$\leqslant 0.01$, $^{**}p-$value$\leqslant 0.05$, $^{*}p-$value$\leqslant 0.1$

5. Discussion and conclusions

In the global action plan 2013-2020, the World Health Organization (WHO) has defined specific actions for the development or strengthening of food and

Table 3: Mediterranean and Non-Mediterranean proportions over time and countries that belong to the dynamic part.

	CZ	H	HR	I	LV	M	PL	S_M	TR	UK	MD
Mediterranean aggregates											
2000	52.86	51.48	62.59	64.03	59.22	59.38	59.73	49.23	81.12	58.31	71.54
2009	55.26	54.17	56.76	65.24	57.20	56.94	59.21	63.19	79.60	58.39	67.24
$p-$value χ^2_1	0.050*	0.029**	<0.001***	0.283	0.101	0.040**	0.663	<0.001***	0.097*	0.965	0.001***
Non-Mediterranean aggregates											
2000	38.58	42.59	30.35	32.22	35.81	37.98	36.48	45.67	18.30	36.84	26.32
2009	36.62	39.81	38.43	31.94	35.79	40.03	36.05	31.71	19.75	37.27	29.56
$p-$value χ^2_1	0.101	0.022**	<0.001***	0.809	1.000	0.080*	0.713	<0.001***	0.108	0.722	0.008***
%△MAI	10.13	12.58	-28.39	2.79	-3.38	-9.01	0.33	84.89	-9.07	-1.02	-16.29

*N*otes: $^{***}p-$value$\leqslant$ 0.01, $^{**}p-$value$\leqslant$ 0.05, $^{*}p-$value$\leqslant$ 0.1

nutrition policies and measures, and the implementation of recommendations and strategies to monitor and control dietary intake all over the world. In line with the WHO plan, the exploration of food consumption evolution across EU countries becomes a crucial public health objective that allows policy-makers to prevent unhealthy diets.

To this aim, the convergence of EU country diets towards a healthy diet, i.e. the Mediterranean diet (MD), have been analysed over a period of 10 years. The adherence of each country diet towards the MD has been measured by the Mediterranean Adequacy Index (MAI), the convergence in food consumption has been detected by means of the Coefficient of Variation (CV), while the CoClust has been adopted to identify profiles of countries sharing common dependency structures. In contrast with more classical clustering techniques, the CoClust algorithm allows the identification of sets of EU countries, called profiles, characterised by complex associations in their food consumption patterns. As discussed in Section 2.3, the CoClust aims at identifying dependent groups of independent countries differently from what the mixture-based clustering techniques do. Hence, the relationships across clusters of the final clustering make it possible the identification of countries that share a certain dependence structure.

The CoClust algorithm has been run using data on the proportion of calories from 16 different food aggregates collected by the Food and Agriculture Organization of the United Nations (FAO) on 40 EU countries. The clustering analysis has been performed separately on data observed in 2000 and 2009 since these years represent the beginning and the end of the more recent and long gamma convergence period in EU, as identified by the CV. In each year, a 5-dimensional t-Student copula model has been selected and all countries have been allocated to one cluster. Therefore, 8 profiles, each of which made up by 5 countries characterised by a multivariate dependence structure in their food consumption, have been detected and further analysed.

Within the 8 identified profiles, 10 different sets of countries stable over time regarding countries composition (the SAs), have been highlighted. Among the SAs, sets like the Nordic countries, the Western EU countries and the Balkans have been identified confirming the findings of [22, 60], although a different set of EU countries has been considered. Moreover, 11 countries, that belong over time to different profiles, have been identified. Most of the time, these countries changed their diets towards the (un)healthier diet of the SA that belongs to the profile towards which the country was going.

While the univariate descriptive analyses, jointly provided by the MAI and the CV, showed that from 2000 to 2009 EU countries experienced a convergence towards a common healthier food dietary structure, the multivariate explorative analyses, provided by CoClust, suggested a different EU food dietary picture. Diets of EU countries are inevitably becoming more and more similar thanks to the adoption of common public policies (as for instance those regarding organic, local products and FBDG), multinational market strategies (with the creation of EU brands) and the internationalisation of food distribution. However, dietary differences within the EU countries still exist, confirming the findings of [23], and, maybe linked to migration and globalisation issues, some countries, either individually or as a group, changed their dietary structure over years towards a (un)healthier diet as represented in Figure 8.

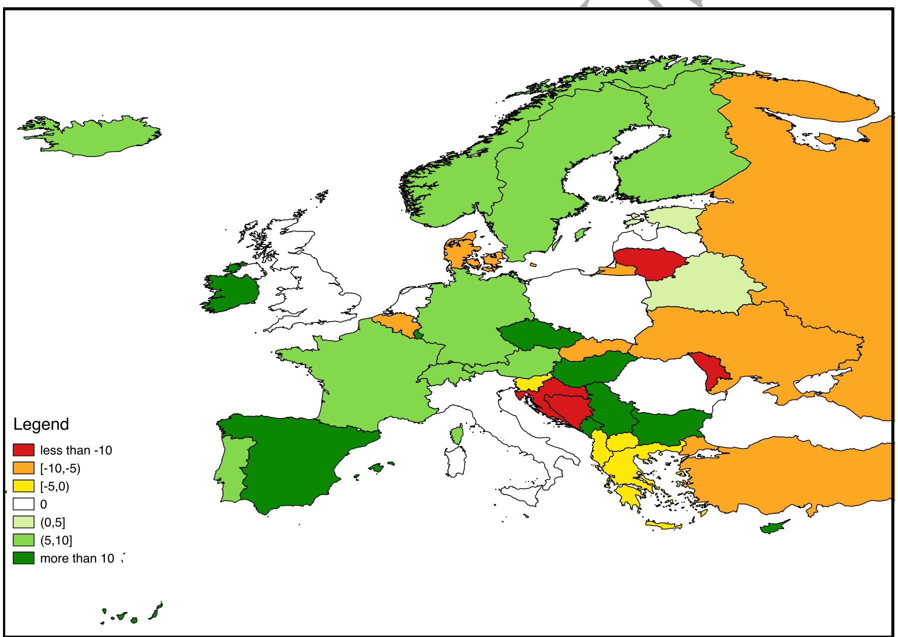


Figure 8: Geographical distribution of the percentage changes in MAI among European countries.

In particular, it is important to underline that SA10 (made up of Lithuania and Bosnia-Herzegovina), SA8 (made up of the Russian Federation and Ukraine), Malta, Republic of Moldova and Croatia experienced a worrisome increase in the consumption of high-calories and nutrient-poor foods (high in fats and sweeteners) that will lead to an increase in obesity and diet-related chronic disease. This results partially confirm the findings of [21] who, comparing the 1960 with the 1998, found a general tendency of Mediterranean countries to embrace a diet rich in fats, typical of the northern countries. Conversely, Serbia and Montenegro is going towards a modern healthy diet [as defined by 19] rich in vegetables and fruits. Among the countries that did not experience a significant change in their food composition, it is worth noting that SA7 (the Netherlands and Romania) and Italy present respectively the lowest and the highest MAI values in both years analysed. In particular, the Netherlands and Romania might introduce new or more powerful and persuasive food policies that encourage people to follow a healthier diet with lower consumption of meat and milk. On the other hand, Italy, together with SA3 (made up by Greece, Slovenia, Albania and FYROM), seem to be worthy ancestors of the Greek peasant farmers of the 1950s, from which the Mediterranean diet originates, embracing varied and healthy diets rich in cereals (excluding beer), fruits (excluding wine), vegetables and vegetable oils (especially Italy).

Finally, as it has been observed (see Figure 6), the geographical proximity among countries does not imply either a common food dietary or the convergence to a common diet over years but, looking at Figure 8, it seems that this is a relevant criterion in understanding times and modalities by which common guidelines and policies are implemented among EU countries. Therefore, the findings of this research can help governments and policy makers encourage the adoption of common policies across EU countries where similar trends dietary are identified.

Acknowledgment

The authors would like to thank professor Tim Lloyd for his valuable comments and suggestions which helped us to improve the paper. The first author acknowledges the support of Free University of Bozen-Bolzano, Faculty of Economics and Management, via the project "Correlation, Clustering, and Copulas: Detecting association in multivariate data" (COCCO).

References

[1] T. Kastner, M. J. I. Rivas, W. Koch, S. Nonhebel, Global changes in diets and the consequences for land requirements for food, PNAS 109 (2012) 6868–6872.

[2] A. Naska, M.-A. Berg, C. Cuadrado, H. Freisling, K. Gedrich, M. Gregorič, C. Kelleher, E. Leskova, M. Nelson, L. Pace, A.-M. Remaut, S. Rodrigues, W. Sekula, M. Sjöstrom, K. Trygg, A. Turrini, J. L. Volatier, G. Zajkas, A. Trichopoulou, Food balance sheet and household budget survey dietary data and mortality patterns in Europe, British Journal of Nutrition 102 (2009) 166–171.

[3] H. Sengul, S. Sengul, Food consumption and economic development in Turkey and European Union countries, Applied Economics 38 (2006) 2421–2431.

[4] B. Traill, Convergence in US and EU food systems? the case of food consumption, in: L. Caswell, R. Cotterill (Eds.), Strategy and Policy in the Food System: Emerging Issues. Part Three: Are food marketing Systems in North America and the European Union Converging?, Food marketing policy center, University of Connecticut and University of Massachusetts, Amherst, MA, 1997, Ch. 11, pp. 141–145.

[5] J. Schmidhuber, W. B. Traill, The changing structure of diets in the European Union in relation to healthy eating guidelines, Public Health Nutrition 9 (5) (2006) 584–595.

[6] R. Capone, H. El Bilali, P. Debs, G. Cardone, N. Driouech, Mediterranean food consumption patterns sustainability: setting up a common ground for future research and action, American Journal of Nutrition and Food Science 1 (2014) 37–52.

[7] H. de Ruiter, T. Kastner, S. Nonhebel, European dietary patterns and their associated land use: Variation between and within countries, Food Policy 44 (2014) 158–166.

[8] A. Tukker, R. A. Goldbohm, A. de Koning, M. Verheijden, R. Kleijn, O. Wolf, I. Pérez-Domínguez, J. M. Rueda-Cantuche, Environmental impacts of changes to healthier diets in Europe, Ecological Economics 70 (2011) 1776–1788.

[9] G. H. Wan, Convergence in food consumption in rural China: evidence from household survey data, China Economic Review 16 (2005) 90–102.

[10] S. Bleich, J. Jones-Smith, J. Wolfson, X. Zhu, M. Story, The complex relationship between diet and health, Health Affairs 34 (11) (2015) 1813–1820.

[11] A. Regmi, L. Unnevehr, Convergence or Divergence in Food Demand: A Comparison of Trends in the EU and North America, Paper presented at the 11th Congress of the EAAE, The Future of Rural Europe in the Global Agri-Food System, Copenhagen, Denmark, Aug. 24-27, 2005.

[12] R. J. Barro, X. Sala-i Martin, Convergence, The Journal of Political Economy 100 (1992) 223–251.

[13] O. S. I. Kalawole, Convergence and determinants of change in nutrient supply, British Food Journal 117 (2015) 2880–2898.

[14] S. Lyons, K. Mayor, R. S. J. Tol, Convergence of consumption patterns during macroeconomic transition: A model of demand in Ireland and the OECD, Economic Modelling 26 (2009) 702–714.

[15] K. Elsner, M. Hartmann, Convergence of food consumption patterns between Eastern and Western Europe, Discussion Paper No.13, Institute of Agricultural Development in Central and Eastern Europe, Germany, 1998.

[16] R. Herrmann, C. Roder, Does food consumption converge internationally? measurement, empirical tests and determinants, European Review of Agricultural Economics 22 (1995) 400–412.

[17] J. M. Connor, North america as a precursor of changes in Western European food-purchasing patterns, European Review of Agricultural Economics 21 (1994) 155–173.

[18] M. Pérez-Ortiz, M. de la Paz-Marín, P. A. Gutiérrez, C. Hervás-Martínez, Classification of EU countries' progress towards sustainable development based on ordinal regression techniques, Knowledge-Based Systems 66 (2014) 178–189.

[19] A. E. Healy, Convergence or difference? Western European household food expenditure, British Food Journal 116 (2014) 792–804.

[20] C. Lazaroua, M. Karaolisb, A.-L. Matalasa, D. B. Panagiotakos, Dietary patterns analysis using data mining method. An application to data from the CYKIDS study, Computer methods and programs in biomedicine 108 (2012) 706–714.

[21] R. Balanza, P. Garcia-Lorda, C. Perez-Rodrigo, J. Aranceta, M. Bullo Bonet, J. Salas-Salvado, Trends in food availability determined by the Food and Agriculture Organization's food balance sheets in Mediterranean Europe in comparison with other European areas, Public Health Nutrition 10 (2007) 168–176.

[22] D. A. Petrovici, C. Ritson, M. Ness, Exploring disparities and similarities in European food consumption patterns, Cahiers d'économie et sociologie rurales 75 (2005) 24–49.

[23] J. M. Gil, A. Gracia, Regional food consumption patterns in the EU, Économies et Sociétiés 24 (2000) 39–48.

[24] D. Blandford, Changes in food consumption patterns in the OECD area, European Review of Agricultural Economics 11 (1984) 43–65.

[25] A. Alberti-Fidanza, F. Fidanza, M. P. Chiuchiu, G. Verducci, D. Fruttini, Dietary studies on two rural Italian population groups of the seven countries study. 3. trend of food and nutrient intake from 1960 to 1991, European Journal of Clinical Nutrition 53 (1999) 854–860.

[26] F. M. L. Di Lascio, S. Giannerini, A copula-based algorithm for discovering patterns of dependent observations, Journal of Classification 29(1) (2012) 50–75.

[27] F. M. L. Di Lascio, S. Giannerini, Clustering dependent observations with copula functions, Statistical Papers (2016) 1–17doi:10.1007/s00362-016-0822-3.

[28] R. B. Nelsen, Introduction to copulas, New York: Springer, 2006.

[29] A. Sklar, Fonctions de répartition à n dimensions et leures marges, Publications de l'Institut de Statistique de L'Université de Paris 8 (1959) 229–231.

[30] M. Rey, V. Roth, Copula mixture model for dependency-seeking clustering, in: J. Langford, J. Pineau (Eds.), Proceedings of the 29th International Conference on Machine Learning (ICML), 2012.

[31] V. Arakelian, D. Karlis, Clustering dependencies via mixtures of copulas, Comm. Statist. Simulation Comput. 43 (7) (2014) 1644–1661.

[32] I. Kosmidis, D. Karlis, Model-based clustering using copulas with applications, Statistics and Computing 26 (5) (2016) 1079–1099.

[33] C. Fraley, A. E. Raftery, How many clusters? which clustering method? answers via model–based cluster analysis, The Computer Journal 41 (8) (1998) 578–588.

[34] A. Chessa, I. Crimaldi, M. Riccaboni, L. Trapin, Cluster analysis of weighted bipartite networks: A new copula-based approach, PLoS ONE 9 (10) (2014) e109507.

[35] F. Durante, S. Saminger-Platz, P. Sarkoci, On representations of 2-increasing binary aggregation functions, Information Sciences 178 (23) (2008) 4534–4541.

[36] J. Arias-Garca, H. D. Meyer, B. D. Baets, Multivariate upper semilinear copulas, Information Sciences 360 (2016) 289–300.

[37] H. D. de Mello Junior, L. Mart, A. V. A. da Cruz, M. M. R. Vellasco, Evolutionary algorithms and elliptical copulas applied to continuous optimization problems, Information Sciences 369 (2016) 419–440.

[38] F. Durante, J. A. Rodrguez-Lallena, M. beda Flores, New constructions of diagonal patchwork copulas, Information Sciences 179 (19) (2009) 3383–3391.

[39] F. Durante, J. Fernndez-Snchez, J. J. Quesada-Molina, M. beda Flores, Diagonal plane sections of trivariate copulas, Information Sciences 333 (2016) 81–87.

[40] S. M. Bartram, Y.-H. Wang, European financial market dependence: An industry analysis, Journal of Banking and Finance 59 (2015) 146–163.

[41] T. Berger, Forecasting based on decomposed financial return series: A wavelet analysis, Journal of Forecasting 35 (5) (2016) 419–433.

[42] F. Durante, E. Foscolo, P. Jaworski, H. Wang, Connectedness measures of spatial contagion in the banking and insurance sector, Advances in Intelligent Systems and Computing 315 (2015) 217–224.

[43] S. Ling, M. McAleer, H. Tong, Frontiers in time series and financial econometrics: An overview, Journal of Econometrics 189 (2) (2015) 245–250.

[44] A. Min, C. Czado, SCOMDY models based on pair-copula constructions with application to exchange rates, Computational Statistics and Data Analysis 76 (2014) 523–535.

[45] S. J. H. Shahzad, R. R. Kumar, S. Ali, S. Ameer, Interdependence between Greece and other European stock markets: A comparison of wavelet and VMD copula, and the portfolio implications, Physica A: Statistical Mechanics and its Applications 47 (2016) 8–33.

[46] G.-J. Wang, C. Xie, Tail dependence structure of the foreign exchange market: A network view, Expert Systems with Applications 46 (2016) 164–179.

[47] Y. Weng, P. Gong, Modeling spatial and temporal dependencies among global stock markets, Expert Systems with Applications 43 (2016) 175–185.

[48] J. V. Pérez-Rodríguez, F. Ledesma-Rodríguez, M. Santana-Gallego, Testing dependence between GDP and tourism's growth rates, Tourism Management 48 (2015) 268–282.

[49] J. Tang, S. Sriboonchitta, V. Ramos, W.-K. Wong, Modelling dependence between tourism demand and exchange rate using the copula-based GARCH model, Current Issues in Tourism 19 (9) (1959) 876–894.

[50] L. Zhu, C. Lim, W. Xie, Y. Wu, Modelling tourist flow association for tourism demand forecasting, Current Issues in Tourism In press (2016) 1–15.

[51] J. C. Reboredo, Do food and oil prices co-move?, Energy policy 49 (2012) 456–467.

[52] J. C. Reboredo, M. Ugando, US dollar exchange rate and food price dependence: implications for portfolio risk management, North American Journal of Economics and Finance 30 (2014) 72–89.

[53] S. Y. Yen, B.-H. Lin, Quasi-maximum likelihood estimation of a censored equation system with a copula approach: meat consumption by u.s. individuals, Agricultural Economics 39 (2008) 207–217.

[54] R. Brakers, I. Van Keilegom, Flexible modeling based on copulas in nonparametric median regression, Journal of Multivariate Analysis 100 (2008) 1270–1281.

[55] B. S. Everitt, S. Landau, M. Leese, Cluster Analysis, 4th Edition, Hodder Arnold, London, 2001.

[56] P. K. Trivedi, D. M. Zimmer, Copula Modeling: An Introduction for Practitioners, Vol. 1, Foundations and Trends in Econometrics, 2005.

[57] H. Joe, J. Xu, The estimation method of inference functions for margins for multivariate models, Tech. rep., Department of Statistics, University of British Columbia (1996).

[58] C. Genest, K. Ghoudi, L. P. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika 82 (1995) 543–552.

[59] A. E. Raftery, D. Nema, Variable selection for model-based clustering, Journal of the American Statistical Association 101 (473) (2006) 168–178.

[60] J. M. Gil, A. Gracia, L. P. Y. Pérez, Food consumption and economic development in the European Union, European Review of Agricultural Economics 22 (1995) 385–399.

[61] FAOSTAT, Food supply on FAOSTAT webpage, FAOSTAT, 2016, http://faostat3.fao.org/download/FB/CC/E (accessed July, 2016).

[62] F. M. L. Di Lascio, S. Giannerini, Coclust: copula based cluster analysis. R package version 0.3-1, http://CRAN.R-project.org/package=CoClust.

Appendix

A. Table

Table A.1: Diet composition of static aggregates (SAs) in 2000 and 2009 (percentage).

Food categories	SA1	SA2	SA3	SA4	SA5	SA6	SA7	SA8	SA9	SA10
2000										
Animal fats	5.94	7.56	2.38	9.67	1.78	3.75	2.97	3.03	4.14	2.41
Eggs	1.07	1.30	1.15	1.35	1.28	0.97	1.70	1.32	1.38	0.94
Fish and seafood	2.99	0.98	0.50	0.98	1.64	0.74	0.78	1.02	0.87	1.13
Meat	11.90	12.50	8.04	7.49	14.20	9.71	10.02	5.72	8.45	5.21
Milk (excluding butter)	11.12	9.31	9.92	7.89	9.00	10.53	11.39	7.58	8.76	8.10
Other animals	1.43	1.67	1.55	1.94	1.74	1.76	1.96	1.86	1.73	1.30
Animal calories	34.45	33.33	23.54	29.31	29.65	27.46	28.82	20.54	25.34	19.08
Alcoholic beverages	4.62	5.71	3.50	5.71	7.39	7.17	4.47	3.22	4.18	4.73
Cereals (excluding beer)	24.77	22.65	33.59	23.87	21.93	30.10	27.67	37.20	29.17	44.72
Fruits (excluding wine)	3.13	3.34	5.13	2.56	4.18	2.19	2.93	1.40	1.90	1.78
Potatoes	3.46	3.24	3.35	4.74	2.82	3.88	5.04	7.11	8.54	5.83
Pulses	0.45	0.34	1.16	0.48	0.88	0.74	0.41	0.55	0.19	1.13
Sugar and sweeteners	12.08	11.77	8.17	13.05	9.39	9.94	10.02	12.61	13.17	6.86
Soyabeans	0.00	0.09	0.05	0.00	0.00	0.00	0.00	0.00	0.03	0.00
Vegetable oils	8.85	11.48	10.02	10.51	13.30	10.39	11.03	6.98	5.66	4.43
Vegetables	1.74	1.78	3.36	2.24	2.56	2.03	2.54	2.02	1.53	2.32
Other vegetables	6.44	6.27	8.11	7.53	7.91	6.10	7.07	8.37	10.29	9.13
Vegetable calories	65.55	66.67	76.46	70.69	70.35	72.54	71.18	79.46	74.66	80.92
2009										
Animal fats	5.76	7.07	3.69	9.84	1.87	4.04	2.67	2.77	3.46	2.67
Eggs	1.08	1.29	0.96	1.56	1.24	1.05	1.24	1.66	1.47	0.98
Fish and seafood	2.92	1.14	0.60	1.21	1.74	0.71	0.77	1.24	0.91	1.75
Meat	12.27	11.73	8.01	7.96	12.84	9.39	9.59	6.89	8.95	7.07
Milk (excluding butter)	11.58	8.88	10.62	8.21	9.00	8.48	13.54	7.70	8.04	8.63
Other animals	1.39	1.59	1.34	1.79	1.63	1.56	1.47	2.02	2.05	1.34
Animal calories	35.00	31.70	25.22	30.56	28.32	25.23	29.28	22.28	24.89	22.44
Alcoholic beverages	4.40	5.54	2.90	4.90	6.45	6.70	4.52	5.16	7.21	6.49
Cereals (excluding beer)	24.92	22.76	30.92	25.07	24.06	31.89	28.09	31.91	26.54	36.76
Fruits (excluding wine)	3.45	3.29	5.13	2.88	3.39	2.36	3.09	1.86	2.57	2.67
Potatoes	3.05	2.76	2.97	3.68	2.39	3.13	4.69	6.42	7.38	4.65
Pulses	0.58	0.30	1.37	0.42	1.26	1.00	0.51	0.43	0.03	1.73
Sugar and sweeteners	9.99	12.39	7.55	13.58	8.45	9.85	10.14	12.50	11.64	7.98
Soyabeans	0.00	0.16	0.04	0.00	0.00	0.01	0.00	0.00	0.01	0.17
Vegetable oils	9.57	13.35	11.40	9.81	14.84	11.36	9.27	9.24	7.38	4.96
Vegetables	2.01	1.86	3.42	2.48	2.53	1.71	2.79	2.33	2.57	2.94
Other vegetables	7.02	5.88	9.07	6.61	8.30	6.76	7.63	7.87	9.78	9.21
Vegetable calories	65.00	68.30	74.78	69.44	71.68	74.77	70.72	77.72	75.11	77.56

Table A.2: Diet composition of countries belonging to the dynamic aggregates in 2000 and 2009 (percentage).

Food categories	CZ	H	HR	I	LV	M	PL	S_M	TR	UK	MD
2000											
Animal fats	4.78	9.54	3.17	4.21	6.48	5.31	6.07	9.19	0.98	3.78	1.30
Eggs	1.91	1.78	1.32	1.25	1.22	1.18	1.10	0.86	1.16	0.98	0.95
Fish and seafood	0.67	0.24	0.50	1.12	0.89	1.60	0.78	0.11	0.40	0.95	0.34
Meat	9.08	11.59	5.52	10.33	5.82	7.76	9.40	16.70	2.39	12.11	3.73
Milk (excluding butter)	8.02	6.61	8.55	7.15	10.16	8.77	6.79	9.12	5.04	9.06	9.87
Other animals	2.33	2.05	1.50	1.59	1.61	1.43	1.32	1.33	1.23	1.20	1.33
Animal Calories	26.78	31.80	20.56	25.64	26.18	26.06	25.45	37.31	11.20	28.08	17.52
Alcoholic beverages	8.56	5.93	7.05	3.75	4.97	2.64	3.79	5.10	0.58	4.84	2.13
Cereals (excluding beer)	24.39	24.42	28.07	30.22	29.99	31.18	32.03	26.64	44.84	22.91	55.31
Fruits (excluding wine)	2.54	2.79	3.60	4.71	2.40	3.23	1.64	3.12	3.75	2.61	2.63
Potatoes	4.36	3.47	6.88	1.98	7.14	3.82	6.58	2.26	3.32	6.01	3.77
Pulses	0.51	0.77	0.78	1.33	0.00	1.41	0.54	1.69	3.32	1.47	0.04
Sugar and sweeteners	12.47	11.03	10.30	7.70	10.52	13.52	11.81	8.47	7.50	9.71	9.14
Soyabeans	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Vegetable oils	12.07	11.77	10.08	16.54	7.83	5.26	7.97	7.90	11.78	12.95	3.16
Vegetables	1.57	2.61	2.60	3.02	1.58	3.43	2.31	2.30	3.88	1.77	2.06
Other vegetables	6.69	5.42	10.08	5.12	9.40	9.45	7.89	5.21	9.82	9.63	4.23
Vegetable calories	73.22	68.20	79.44	74.36	73.82	73.94	74.55	62.69	88.80	71.92	82.48
2009											
Animal fats	5.80	8.67	5.23	4.08	6.42	4.28	6.35	1.27	1.16	3.11	4.08
Eggs	1.67	1.72	1.27	1.31	1.53	1.48	1.19	0.74	0.98	1.09	1.19
Fish and seafood	0.60	0.33	1.12	1.28	1.78	1.76	1.00	0.36	0.39	1.01	0.95
Meat	9.39	10.51	8.68	10.34	8.20	9.01	10.23	10.04	2.59	12.09	4.66
Milk (excluding butter)	7.29	6.87	8.80	7.17	9.30	7.67	5.45	8.39	5.89	9.39	10.51
Other animals	2.07	1.90	1.45	1.55	2.06	1.71	1.38	1.62	1.06	1.31	1.50
Animal Calories	26.82	29.99	26.55	25.73	29.29	25.92	25.60	22.41	12.07	27.99	22.89
Alcoholic beverages	8.13	6.02	4.81	2.83	7.01	3.02	4.75	5.10	0.65	4.34	3.20
Cereals (excluding beer)	25.24	26.23	27.61	30.34	25.67	29.90	32.06	33.28	42.45	24.72	40.75
Fruits (excluding wine)	2.27	2.92	4.14	5.28	1.64	3.05	1.90	5.55	4.14	3.49	1.53
Potatoes	3.39	2.80	3.42	1.84	5.82	2.46	5.83	2.42	2.53	5.02	3.71
Pulses	0.78	0.69	0.30	1.33	0.00	1.37	0.43	2.86	2.77	0.74	0.37
Sugar and sweeteners	10.40	10.15	13.00	7.49	8.28	15.87	11.45	9.65	8.07	10.29	7.62
Soyabeans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00
Vegetable oils	14.47	13.82	9.89	17.09	11.11	6.49	8.52	8.19	14.53	12.61	11.70
Vegetables	1.49	2.35	2.30	2.88	2.37	3.89	2.28	2.90	3.67	1.83	2.62
Other vegetables	7.01	5.03	7.98	5.20	8.79	8.01	7.19	7.63	9.05	8.98	5.61
Vegetable calories	73.18	70.01	73.45	74.27	70.71	74.08	74.40	77.59	87.93	72.01	77.11