



TESIS SS14-2501

**STUDI KONVERGENSI *AGEING* (PENUAAN) DI NEGARA-
NEGARA MEDITERANIA**

DIAN SAFRINA PUTRI
NRP 1312 201 902

PEMBIMBING
Laurence REBOUL
Marie-christine ROUBAUD
Nicolas PECH
Sebastien OLIVEAU

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2015



THESIS SS14-2501

**STUDY OF AGEING CONVERGENCE IN
MEDITERRANEAN COUNTRIES**

DIAN SAFRINA PUTRI
1312 201 902

SUPERVISORS
Laurence REBOUL
Marie-christine ROUBAUD
Nicolas PECH
Sebastien OLIVEAU

MAGISTER PROGRAM
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2015

STUDY OF AGEING CONVERGENCE IN MEDITERRANEAN COUNTRIES

This thesis is composed to fulfill the requirement to get Master degree
Magister Sains (M.Si)
in
Institut Teknologi Sepuluh Nopember

By:
DIAN SAFRINA PUTRI
NRP. 1312 201 902

Examination date : September, 19th 2014
Graduation period : March 2015

Approved by :

1. Laurence REBOUL (Supervisor)
2. Marie-christine ROUBAUD (Supervisor)
3. Nicolas PECH (Supervisor)
4. Sebastien OLIVEAU (Supervisor)



ABSTRACT

The ageing phenomenon became a much-discussed issue in demographic research lately. This study focuses on ageing in the area of Western Europe, especially in the Mediterranean countries. It is well established that people leave Northern Europe and other regions, making Southern Europe a destination for living. So this contributes to the greater presence of “old” people in Mediterranean countries. We want to explore whether there is a convergence of ageing in the Mediterranean countries. One of the convergence approaches is σ convergence. This approach views convergence as the reduction of the standard deviation in each period.

In this study, we aim at investigating territorial ageing behaviors over certain periods in order to identify the convergence of ageing. To measure the ageing in a territory, we adopt a common indicator of ageing: the proportion of people aged 65 years and over. We study the ageing behavior of four Mediterranean countries: Italy, France, Spain and Portugal that are divided into geographical units (in France we recognize this geographical unit as department) over three periods in 1990, 2000 and 2010. For that task, we apply several statistical methods, among those classification, Markov chain approach, regression model, and spatial statistics.

Classification method allows us to obtain five groups of territories with similar ageing behavior over time. A characterization of ageing trajectories within groups is given by the determination of territories' condition over time (young, rather young, moderate, rather old and old). We find that the median of each trajectory group increases over period, so they are heading to ageing.

A modeling by Markov chain allows us to describe ageing trajectories using five states obtained earlier. This approach confirmed that all states have the bigger probability going older. In the long run, Markov chain approach predicts that there will be no territory in the young state and almost all territories studied will tend to be in old state.

Regression methods lead to the same conclusion as above. Moreover, we observe that territories and their neighbors are becoming increasingly similar over time. So we can conclude that there is convergence of ageing in the Mediterranean based on this model.

To improve this study, it would be better to add the series of data in order not to lead to inaccurate conclusions. We can also apply other methods e.g. continuous-time Markov chain, space-time dynamic Markov chain, and generalized equation estimation model.

Keywords: Ageing, Mediterranean Countries, Convergence, Markov Chain

FOREWORD

Recently, the ageing issue became more and more popular in many journals in social science. The purpose in studying this topic differed depending on the institution concerned. For example, by knowing the trend of ageing, the government of a country can determine an appropriate policy to develop their country. To answer the questions about population issues in France, a population observatory named DemoMed established. DemoMed, stands for Observatoire Démographique de la Méditerranée, consists of the researchers on different disciplines: demographers, statistician, geographers, academic or experts that sited in Maison méditerranéenne des sciences de l'homme - MMSH (Aix-en-Provence). DemoMed aims to carry out scientific meeting and exchange between disciplines for development and knowledge diffusion in the Mediterranean population. In concrete, DemoMed works on demographic surveillance, future thinking about demography, methodological pooling, knowledge diffusion, scientific publication, and also an educational program (such as doctoral and post-doctoral).

I am a student of Master 2 in Aix-Marseille University, engage in a DemoMed's project called PEPS Interdisciplinaires titled "Aging convergence in the Mediterranean and temporal dynamic of Markov Chain". The purpose of this project is applying exploratory statistical method and modeling trajectory in different territories in term of ageing incorporating the spatial dimension. We want to measure the influence of neighborhood in the demographic phenomenon. I do an internship in Institut de Mathématiques de Marseille (I2M) that conducted from April to September 2014 and it is financed by PEPS project and I2M. I2M is a joint research unit CNRS / Aix-Marseille University / Central Marseille. It has about one hundred and thirty faculty members, thirty researchers CNRS, fifteen technical and administrative staff, sixty PhD students and twenty post-doctoral researchers. On doing this internship, I go to the Institut Mathématiques de Luminy (IML) which is a part of I2M before the restructuring.

Our team consists of several people who master demography and statistics. One day per week, we have a meeting to discuss the work I did. The first month of my internship, I read the literature review about ageing in the Mediterranean, Markov chain method, spatial statistics, etc. Then the next months, when I got the data that would be analyzed, I applied the method that I've been studied to the data. I did all programming use R, open source statistical software.

During this internship, I got a lot of experience in working in the demography domain which I have never been before during my undergraduate degree. In my opinion, this research integrates statistical methods in order to draw a valid conclusion, start from doing data exploratory until modeling the data by regression model. Other than that, I could have worked in a team containing people from different knowledge base that makes me better understand the work climate in terms of academic research in France.

I would thank to I2M and my team in PEPS project: Sebastien Oliveau, Nicolas Pech, Marie-Christine Roubaud, Laurence Reboul, and also Yoan Doignon; all of my lecturers in Aix-Marseille University and Mme Fabienne Picolet, and all my friends of Master 2 MASS. In addition, thanks to Indonesian government which has funded my study in France, postgraduate program of Institut Teknologi Sepuluh Nopember (ITS) with its assistance I can continue my study, and the lecturers of Statistics Department in ITS. The last, I want to thank to my family and my Indonesian friends that always pray for me and support me in achieving my dreams.

I realize that this report is not perfect, if there are some suggestions or want to discuss about this topic, I can be contacted by email dian.safrina.p@gmail.com.

TABLE OF CONTENTS

ABSTRACT	1
TABLE OF CONTENTS	2
FOREWORD	3
INTRODUCTION	5
CHAPTER 1: DATA AND METHODOLOGY	7
1. Data	7
1.1 Unit of Study	7
1.2 Variables.....	8
1.3 Database	8
2. Methodology of Analysis.....	9
2.1 Classification of Trajectory of Ageing.....	9
2.2 Markov Chain Approach	9
2.3 Regression Models	10
CHAPTER 2: ANALYSIS OF RESULT	11
1. Data Description	11
2. Classification of Territories in Function of Tajejectory of Ageing	12
2.1 Classification of Territories by Hierarchical Clustering.....	12
2.2 Discretization of Territories into Classes of Ageing.....	14
2.3 Characterization of Trajectory of Ageing based on Discretization of Variable	17
2.4 Trajectory of Ageing and Convergence	18
2.5 Distribution of Ageing Trajectories in Space	18
3. Markov Chain Approach.....	20
3.1 Using Data normalization	22
3.2 Estimation of Transition Matrix based on Data Normalization.....	27
4. Modeling of Ageing by Regression Models	30
4.1.Linear Regression Models	31
4.2.Linear Mixed Models.....	35
4.3.Relation of the Best Model with Ageing Convergence in Mediterranean.....	38
CHAPTER 3: CONCLUSION AND DISCUSSION	40
1. Conclusion	40
2. Discussion.....	41
BIBLIOGRAPHY	42
APPENDIX	44

BIBLIOGRAPHY

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- Casado-Diaz, M. A., Kaiser, C., & Warnes, A. M. (2004). Northern European Retired Residents in Nine Southern European Areas: Characteristics, Motivations and Adjustment. *Ageing and Society* , Vol 24: 353-381.
- Cnann, A., Laird, N. M., & Slasor, P. (1997). Tutorial in Biostatistics, Using The General Linear Mixed Model to Analyse Unbalanced Repeated Measures and Longitudinal Data. *Statistics in Medicine* , Vol 16: 2349-2380.
- Dlugosz, Z. (2011). Population Ageing in Europe. *Procedia Social and Behavioral Sciences* , Vol 19: 47-55.
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis: Third Edition*. USA: John Wiley & Sons, Inc.
- Fischer, M., & Getis, A. (2010). *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. New York: Springer, Heidelberg Dordrecht.
- Getis, A., & Aldstadt, J. (2004). Constructing the Spatial Weights Matrix Using a Local Statistics. *Geographical Analysis* , Vol 36 no 2.
- Griffith, D. A. (2009). Spatial Autocorrelation. *International Encyclopedia of Human Geography* , 308-316.
- Guyon, X., & Hardouin, C. (2002). Markov Chain Markov Field Dynamics: Models and Statistics. *Statistics A Journal of Theoretical and Applied Statistics* , 339-363.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis Seventh Edition*. USA: Prentice Hall.
- Leeuw, J. d., & Meijer, E. (2008). *Handbook of Multilevel Analysis*. New York: Springer Science+Business Media, LLC.
- Maindonald, J., & Braun, J. (2007). *Data Analysis and Graphics Using R: An Example-based Approach Second Edition*. Cambridge: Cambridge University Press.
- Meyn, S., & Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability Second Edition*. New York: Cambridge University Press.
- Pison, G. (2009). Population Ageing Will be Faster in the South than in the North. *Population & Societies* , No. 457.
- Rey, S. (2001). Spatial Empirics for Economic Growth and Convergence. *Geographical Analysis* , Vol 33: 195-214.

Seber, G. A., & Lee, A. J. (2003). *Linear Regression Analysis Second Edition*. New Jersey: John Wiley & Sons, Inc.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* , Vol 58: 236-244.

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton: Chapman & Hall/CRC.

Williams, A. M., King, R., & Warnes, T. (1997). A Place in the Sun: International Retirement Migration from Northern to Southern Europe. *European Urban and Regional Studies* , Vol 4 no 2: 115-134.

Williams, A. M., King, R., Warnes, A., & Patterson, G. (2000). Tourism Geographies: An International Journal of Tourism Space, Place and Environment. *Tourism and International Retirement Migration: New Forms of an Old Relationship in Southern Europe* , Vol 2: 28-49.

Wilson, C. (2011). Understanding Global Demographic Convergence since 1950. *Population and Development Review* , Vol 37 no 2: 375-388.

INTRODUCTION

Ageing phenomenon became a much-discussed issue in demographic research lately. The article written by (Pison, 2009) entitled “Population ageing will be faster in the South than in the North” gave a definition and some causes of population ageing. The author said that population ageing occurs when the old people’s proportion in a population increases and the young people’s proportion decreases. Population ageing is the result of the fertility decline and increasing life expectancy across the world. The main conclusion of this article is that ageing is just starting in many southern countries and it will be quicker than in the North.

One of population ageing measure is the demographic ageing index (I_{DA}). It is defined as the number of people aged 65 and over (old group) per 100 people under age 15 (young group). Both groups were roughly of the same size in the entire population of Europe in 2005. However, the index value was different for each country. For instance, it is known that Western Europe has a higher index than Eastern (for example Italy, Germany, Spain, and Greece). According to the population forecasting of (Dlugosz, 2011) between 2005 and 2045, ageing will continue to intensify in most European countries (especially Spain, Latvia, and Belarus), but in Scandinavia the ageing process has slowed down.

This study focuses on ageing in the area of Western Europe, especially in the Mediterranean countries. It is well established that people leave Northern Europe and other region, making Southern Europe as a prime destination to live such as Italy, Malta, Portugal, and Spain, see e.g. (Williams, King, & Warnes, A Place in the Sun: International Retirement Migration from Northern to Southern Europe, 1997), (Casado-Diaz, Kaiser, & Warnes, 2004), (Williams A. M., King, Warnes, & Patterson, 2000). We are interested in analyzing ageing process in Mediterranean countries by incorporating the spatial dimension, and also examining the convergence in ageing over a certain time.

More specifically, we want to explore whether the ageing in the Mediterranean countries is heading to the same direction. (Rey, 2001) presented σ convergence in his study to measure the convergence of US regional income over the 1924-1994 period. This approach defines convergence as the reduction of the standard deviation of the data over time. But this concept does not determine individual convergence while we want to examine individual convergence. Our study will examine the convergence based on the tendency of territories and their neighbors heading to the same behavior (in our case is ageing) over time.

The objectives that will be achieved in this study are:

1. Implement exploratory statistic methods to describe ageing in Mediterranean countries over time.
2. Modeling the trajectories of territories in terms of aging by incorporating the spatial dimension in order to explore convergence.

CHAPTER 1

DATA AND METHODOLOGY

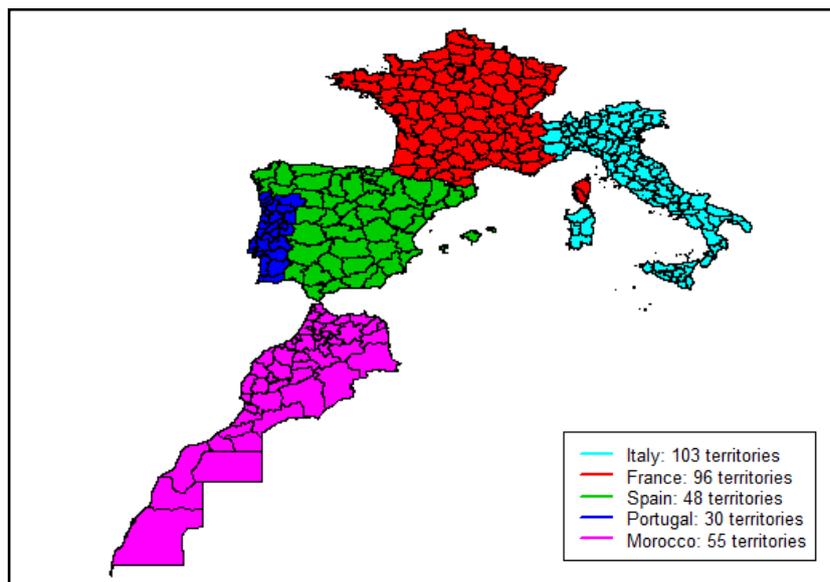
1. Data

In the explanation of the data, we will present unit of study, the variable which will be used and database before we treat it using various analysis.

1.1. Unit of Study

Based on Wikipedia Mediterranean countries are those that surround the Mediterranean Sea. The more precise definition is given by <http://www.fao.org>, in which the Mediterranean region is defined as countries bordering the Mediterranean Sea (plus Portugal) between about 27° to 47°N and 10°W to 37°E. Mediterranean countries are spread over three continents, Europe, Asia, and Africa. In this study, we focus on five countries: Italy, France, Spain, Portugal, and Morocco. They are divided by geographical unit, for example, statistics unit in France is divided by department. To simplify term of this geographical unit, hereinafter it will be called territory.

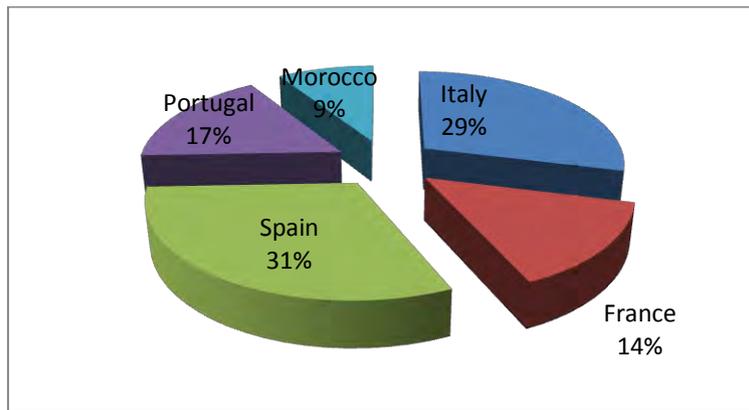
Figure 1. Map of Mediterranean Countries Used in Analysis



The total number of territories is 332 in all 5 countries, we observe following distribution of territories.

Figure 2 shows that the distribution of territories is unbalanced over countries. Two-thirds of the territories are located in Italy and Spain, both have a close number of territories which is also the case for France and Portugal (14 % and 17 % respectively). The country that has the fewest number of territories is Morocco, which is 9 % of the total number of territories.

Figure 2. Distribution of the number of territories in each country



1.2. Variables

To measure convergence of ageing, we need an indicator which represents the term of ageing. Many journals in demographic use the variable people aged 65 years and over as the indicator of ageing (see Pison, 2009; Dlugosz, 2011). This research will adopt this common indicator. This indicator is measured over four periods in our database, so the indicator of ageing can be written as follows.

x_{it} = proportion of people aged 65 years and over in the territory i for the year t
 where

i = Italy, Spain, France, Portugal, Morocco

j = 1982, 1990, 2000 and 2010

1.3. Database

The original database has 333 rows and 6 columns, presented in Table 1. The database contains the name of the territories in 5 Mediterranean countries, the proportion of people aged 65 years and above in year 1882, 1990, 2000 and 2010, and size of territories.

Table 1. Database PEPS: name, proportion, and total area each territory

NOM	P65_82	P65_90	P65_00	P65_10	SUPERFICIE
Bas-Rhin	0	11.96	13.48	15.1	4768.173
Haut-Rhin	0	12.04	14.15	15.87	3483.038
Dordogne	0	20.7	23.58	24.09	9065.082
Gironde	0	14.62	16.24	16.4	10074.24
Landes	0	18.27	20.76	20.64	9229.57
⋮					

There are 332 territories in these 5 countries, but we deleted Acores and Madeira which belonged to Portugal because these territories are far from the European mainland. So, the next analysis will use 330 territories across 5 countries. In addition, according to the

data preprocessing in Appendix II, we decided to not include Morocco in the analysis due to some reasons. First, Morocco has the smallest proportion of people aged 65 years and over. So we can say that the majority of people in Morocco is young. Second, in 2010, there is no data record for Morocco Finally, the total number of territories that will be analyzed is 275.

2. Methodology of Analysis

We use several statistical methods which will be introduced in this part. All the analyses are performed with R software, an open source statistical software which contains many packages that can be downloaded for free. We divide the analysis into three big parts:

2.1 Classification of Trajectory of Ageing

The methods that will be used are:

- a. Hierarchical clustering with Ward's aggregation method based on Euclidean distance to classify the trajectories of ageing. This allows us to see territories which tend to have similar behavior and those which are different. We use function "hclust" that are included in package stats..
- b. Classify the proportion of people aged 65 years and over into classes based on the global quintile. This method aims to characterize the groups of the above clustering of territories in terms of trajectories of ageing.
 - The territories that have proportion of people aged 65 years and over below first quintile are called as "young".
 - The territories that have proportion of people aged 65 years and over between first quintile and second quintile are called as "rather young".
 - The territories that have proportion of people aged 65 years and over between second quintile and third quintile are called as "moderate".
 - The territories that have proportion of people aged 65 years and over between third quintile and fourth quintile are called as "rather old"
 - The territories that have proportion of people aged 65 years and over more than fourth quintile are called as "old".
- c. Tests of spatial autocorrelation, more precisely join count test. It tests the dispersion of trajectories of ageing in space. It is also can be used to find out which groups are more spatially clustered than other groups.. In order to apply this method, we must define the neighbors of territories within group.

2.2 Markov Chain Approach

We want to model the convergence of ageing by discrete time homogeneous Markov chain. By this approach, we can estimate the transition matrix of the probabilities of a territory to move to the other state over time. To define the state of a territory, we use the discretization which has been defined before. To estimate transition matrix, we need to know the frequency of each class in 1990, 2000 and 2010.

$$f_1 = \frac{\text{among the territories in state 1 in 1990, the number of the territories in state 1 in 2000}}{\text{the number of territories in state 1 in 1990}}$$

$$f_2 = \frac{\text{among the territories in state 1 in 2000, the number of the territories in state 1 in 2010}}{\text{the number of territories in state 1 in 2000}}$$

$$\widehat{p}_{11} = \frac{n_1}{n_1 + n_2} \times f_1 + \frac{n_2}{n_1 + n_2} \times f_2$$

where n_1 and n_2 is the number of territories in state 1 in 1990 and 2000 respectively. Transition matrix describes the probability of a territory which is initially located in state i moves to state j . To compute the stationary probability of a state, we use package DTMCpack in R.

Before calculating the transition matrix, we do normalizations and then we estimate the density curve of the data by kernel density estimator. In one case of normalization, we incorporate the influence of neighbor, so we must choose a definition of neighborhood. Here, we build spatial weight matrix which allows us to define the neighbor term. The package which we use such as maptools, spdep, plotrix, etc.

2.3 Regression Models

a. Linear regression model

We start our regression model with the simplest method: linear regression with one dependent variable (proportion of people aged 65 years and over) and multiple independent variables which will be explained in Chapter Analysis of Result.

b. Linear mixed model

In our study, we measure the proportion of people aged 65 years and over in 3 different periods, 1990, 2000, and 2010, so we have multiple measures for each territory (longitudinal data). Because of this situation, the independence assumption made in classical regression models may be difficult to be fulfilled. Therefore, we add a random effect for subject (in our study is territory).

For more details about the methods above see Appendix I and all programming which are conducted in R can be seen in Appendix V.

One of the methods that will be applied in this study is a Markov chains approach to estimate the probability of a territory to move from certain state to the other state. By knowing the probability, in the end, we can examine whether the territories in Mediterranean have the same behavior / characteristic in ageing. The other method carried out is linear regression to model the relationship between existing variables and the contribution of variables that serve as explanatory to the dependent variable. In this analysis, we can define which model that more appropriate to apply to our data.

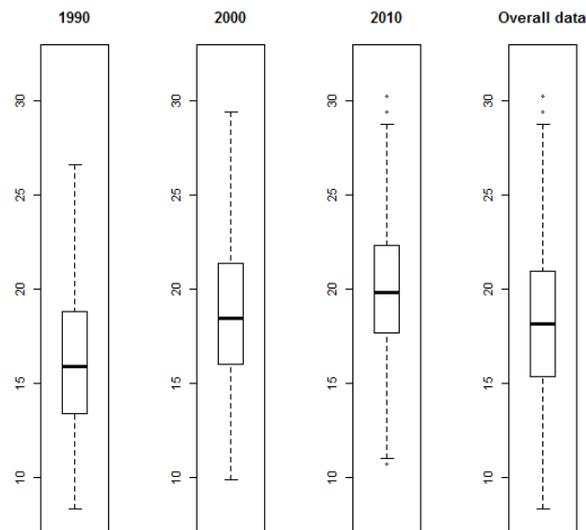
CHAPTER 2

ANALYSIS OF RESULT

1. Data Description

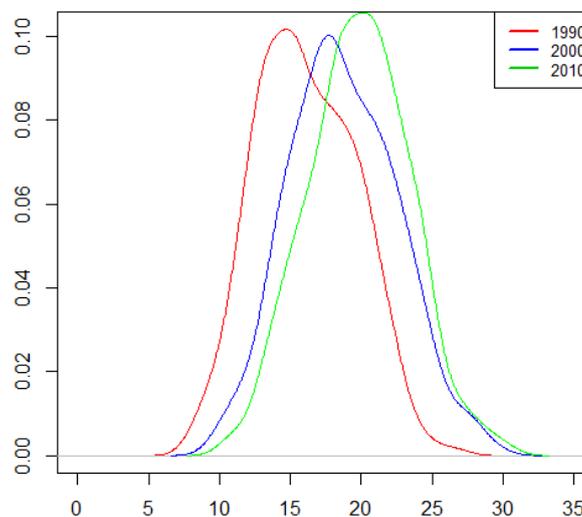
Before going to further analysis, we will present the data description. We have three variables representing proportion of people aged 65 year and over at three dates: 1990, 2000 and 2010. The distribution of this variable at each data is presented on Figure 3.

Figure 3. Boxplot of the indicator of ageing over period vs Box plot of the overall data



The boxplot shows that median increased, both from 1990 to 2000 and from 2000 to 2010. Tentatively, we can mention that all territories tend to age. But the enhancement happened between 2000 to 2010 is less than that between 1990 to 2000. The density curves below also confirm the augmentation of the proportion of population aged 65 years and above in the studied Mediterranean countries.

Figure 4. Density Curves over Three Periods



Overall data is obtained by combining the variable in 1990, 2000 and 2010 in one column. The boxplot of overall data presents that minimum proportion of people aged 65 years and above is 8.32% (Val d'Oise, France in 1990) and the maximum value is 30.23% (Pinhal Interior Sul, Portugal in 2010). There are 2 observations beyond the upper limit of box plot, we called them extreme values. These 2 observations is the proportion of people aged 65 years and over found in Pinhal Interior Sul in 2000 and 2010 respectively. The median of overall data is 18.17%, this means that half of the territories have proportion of people aged 65 years and over less than 18.17%.

For the next analysis, we want to examine whether all territories show the same behavior as shown by median.

2. Classification of Territories in Function of the Trajectory of Ageing

To be able to track the evolution of the territories, we want to group the territories that have similar trajectory. We define similarity by measuring the Euclidean distance between two territories for the three periods of data. Before calculating the distance, we need to standardize the proportion of people aged 65 years and over in order to avoid a misleading on conclusion caused by the inequality of variance in each period of data.

2.1 Classification of Territories by Hierarchical Clustering

One method which can be used to classify the territories is hierarchical clustering by Ward's method. In defining the optimum number of clusters, we often determine it intuitively based on dendrogram which has obtained. There is no specific way how to determine it, but here, we will compute the semi partial R-squared (SPRSQ) which can be taken as a tool to justify the optimum number of clusters. Based on the graph of semi-partial R-square, we determine the optimum number of classes is 5 (see Appendix IV).

The classification of the territories into five groups according to their ageing trajectory is shown on Figure 5. We see that the territories having similar behavior tend to be spatially clustered. For example, in the southwestern part of France and the middle of Italy, there is a grouping of "green" territory; in the northern France and southern Spain territories with almost the same behavior form a "red" group. But we don't know the characteristic of red, green and the other groups because we have no information about the territories which form those groups. In order to better understand the ageing trajectory of each group, we will define an ageing status of the territories by discretizing its proportion of people aged 65 years and over. A further analysis about spatial grouping will be presented in section 2.4

Figure 5. Group of the trajectory of ageing based on Ward method

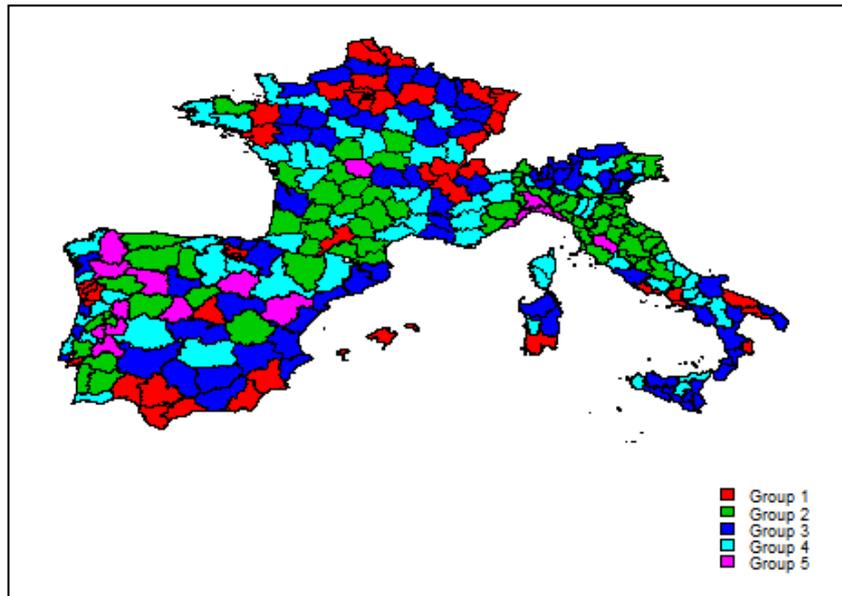


Figure 6. Boxplot of each group of the ageing trajectory

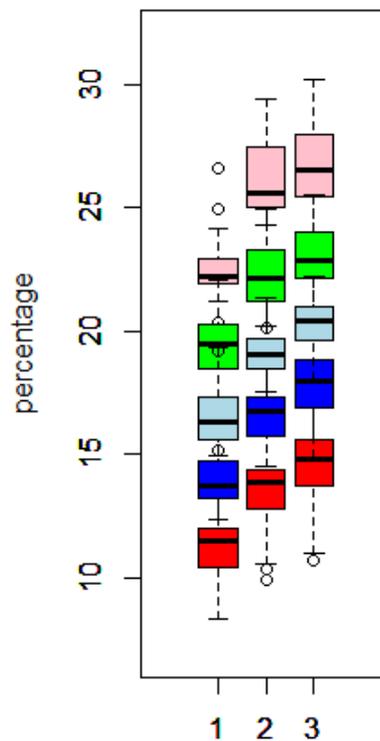
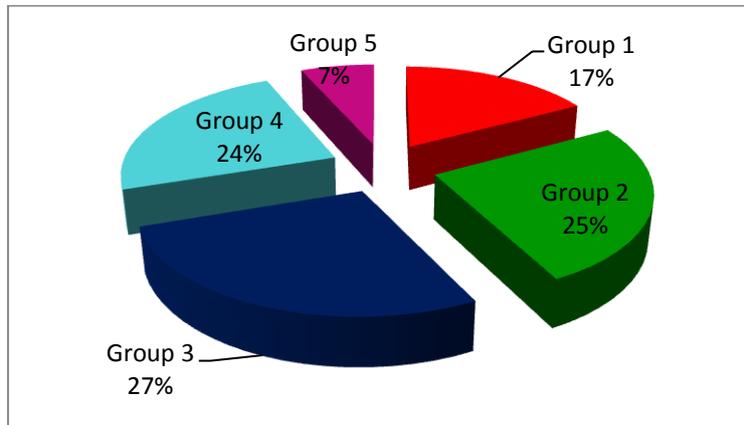


Figure 6 shows that proportion of people aged 65 years and over in each group increases over time, but the dispersion is not same between each boxplot. In order to figure out the behavior of each group, first, we have to define the state of a territory by discretizing its proportion in section 2.2. The further analysis about spatial grouping will be presented in section 2.4.

The distribution of the trajectory is not the same, we can see the repartition in Figure 7.

Figure 7. Percentage of territories in each group



About a quarter of the territories changed from a young state into moderate, from a rather old into old state, and from a rather young into the rather old state. The smallest percentage is the trajectory of territories that remaining in the old state.

2.2 Discretization of Territories into Classes of Ageing

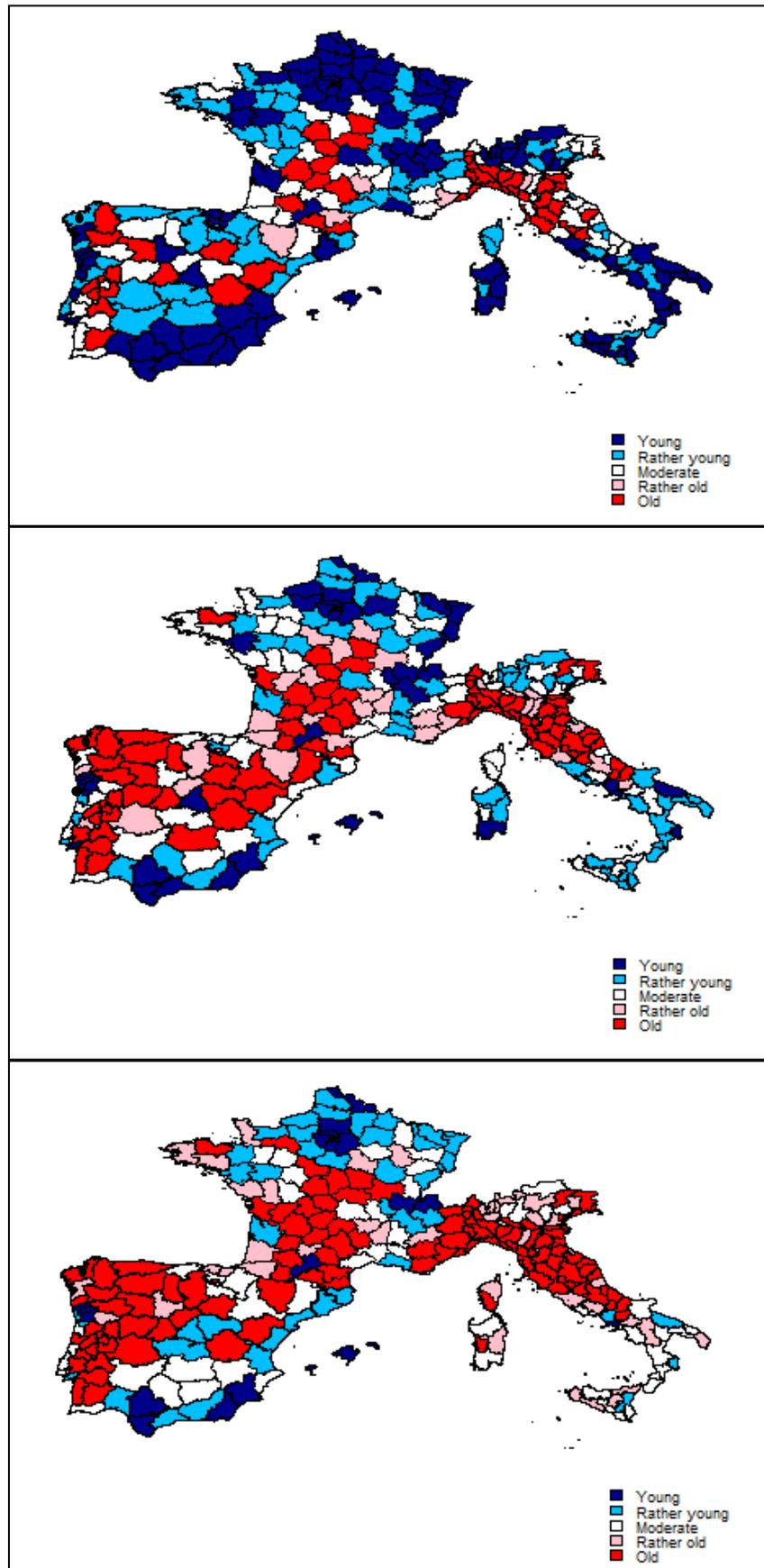
To discretize the proportion of the variable of ageing, we use the quintile of the overall data as the bounds of the classes. For the first step, we mix all data in one column and divide its value into quintile, then we discretize the proportion of people aged 65 years and over in each territory into five classes using rule in section 2.1 of Chapter 1.

Figure 8. Discretization of variables

Class 1	Class 2	Class 3	Class 4	Class 5
Q1	Q2	Q3	Q4	

We present the maps of the classified territories at each time in the Figure 9. The first map shows that the territories in North France, Southeast Italy (including Sicilia island), and South Spain are young in 1990. The old territories are located in some part of each country, for example, north and middle Portugal and eastern Spain. France has some old territories in the southwest area (Ariege, Lot), and also old territories spread in the Northwest of Italy. The moderate aged territories are spread over space in 1990 and 2000, but in 2010 there are clusters of moderate aged territories in northern Spain.

Figure 9. Class of territories in Mediterranean countries in 1990, 2000 and 2010

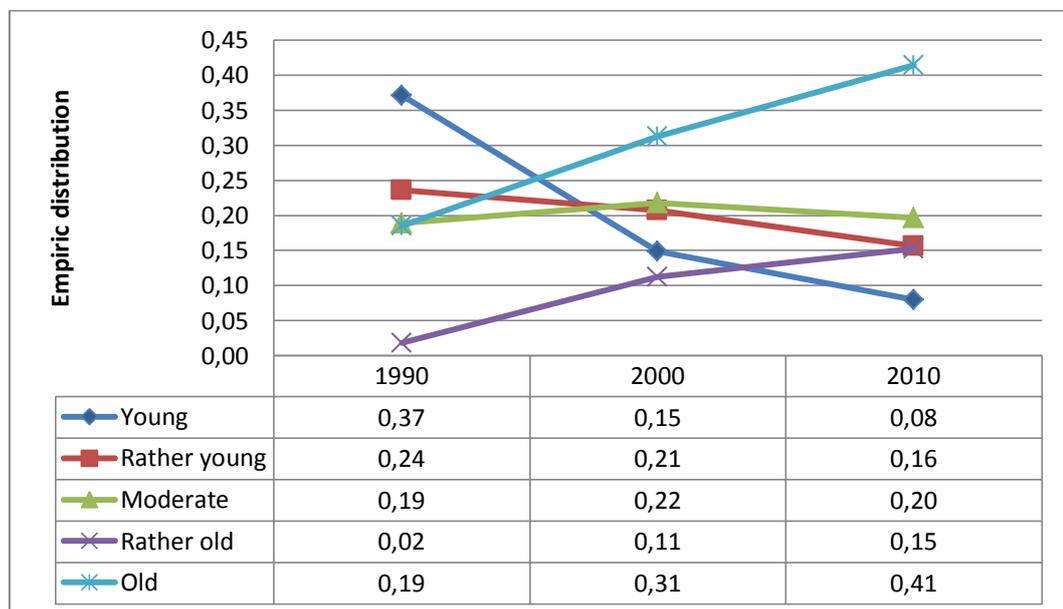


Based on Figure 9, in 2000, the territories that previously were in the young class changed into rather young class and also the number of old territories increased. Almost all of territories in Portugal were dominated by old class. The similar situation happened in Italy.

Therefore, we can observe the diffusion of ageing in the studied Mediterranean countries. In 1990, a high proportion of ageing occurred only in a few territories. But as time passes, the territories that are surrounding the old territories are also ageing. The interesting thing is that young territories didn't change directly to the old state, but there is a gradual change. For example, Galicia in Spain (indicated by a black dot) has a low proportion of people aged 65 years and over in 1990, but in 2000 and 2010 it classified as old class. Galicia has more than 21.650% of people aged 65 years and over in 2000 and 2010 (see Figure 9)

Beside that, we can notice that there is a spatial effect on ageing case. The young territories tend to spatially group with other young area, rather young also group with other rather young territories, and so on. Territories surrounding the old territories tend to behave as their neighbors. We can clearly see that there is grouping of old territories in the middle of Italy, North Spain, and Southwest France. Figure 10 below shows the evolution of each class in time.

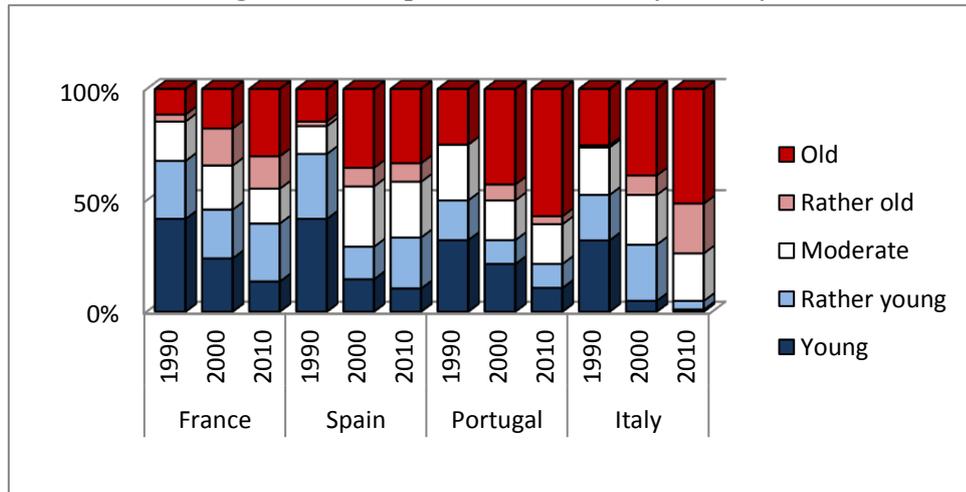
Figure 10. Class changes each year based on classification by quintile



We take the young and old classes as an example. In 1990 there are 37% of young territories. This rate decreases in 2000 to 15%, and finally reaches 8% in 2010. Old territories in 1990 amounted to 19% and they have a significant increase, so that in 2010 there is 41% of old territories. We can say that in the initial year (1990) the order from the largest to smallest percentage is young, rather young, moderate, old, and rather old. While in the final year (2010) the order is in contrast with 1990. The increase or decline of each state which occurred between 1990 and 2000 is greater than the increase or decrease

occurred between 2000 and 2010. This means that there is a global convergence of ageing. We can see the distribution of each country on Figure 11.

Figure 11. Comparison each class by country



In 1990, France and Spain had the largest proportion of young class, followed respectively by Portugal and Italy. But in 2010, the number of young territories decreased while the number old territories increased. This happened in all the four countries.

2.3 Characterization of Trajectory of Ageing based on Discretization of Variable

We can characterize the trajectory of ageing obtained in Figure 6 by the way of the classes obtained in Section 2.2. To make it easier in determining in what class a territory is, the following graphs is presented.

Figure 12. Median of each trajectory group over period

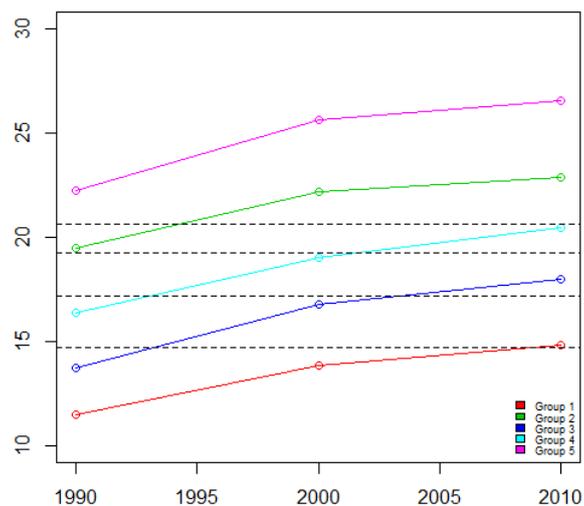


Figure 12 is median of group obtained by hierarchical clustering after we discretize the territories into five classes in Section 2.2. The black dotted lines show the first to fourth quintiles. In other words, the areas between the black dotted line are the classes of

territories. With the help of those black dotted lines, we can characterize the trajectory of ageing as follows.

- Red group: young territories remain young
- Green group: rather old territories change into old
- Dark blue group: young territories change into moderate
- Light blue group: rather young territories change into rather old
- Pink group: old territories remain old

2.4. Trajectory of Ageing and Convergence

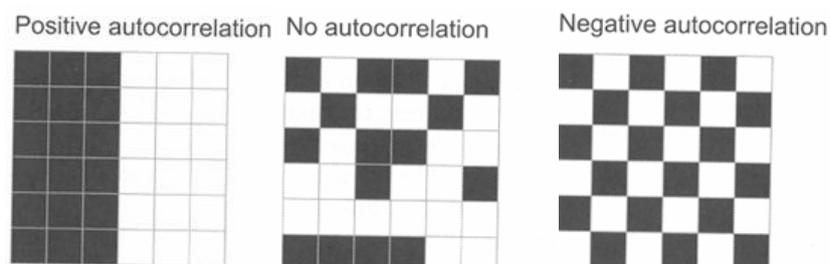
As we observe in Figure 12, the median of each group of territories increases over time, although there are groups which remain in the same age class (i.e. red and pink group). All the groups are heading to the same direction (ageing). So we can draw a conclusion that all territories converge in ageing, but not at the same rate. Overall, we note that:

- a. Age class changes generally occur gradually, from a class to the next one or the one following the next.
- b. The median increase between 1990 and 2000 is bigger than the median increase between 2000 and 2010.
- c. As it has been said before, there is a spatial clustering of ageing.

2.5. Distribution of Ageing Trajectories in Space

Based on Figure 6, some spatial patterns appear. For example, in Italy, the green group is concentrated in the middle area, while in the southwest there are groups of light blue and dark blue territories. We are interested in presenting the distribution of ageing trajectories in a map, whether they scattered, clustered, or do not have a spatial pattern (random in space). In order to conduct this analysis, we will apply a statistical test named joint count on the groups of classification 2.1. The principle of joint count is measuring the spatial similarity of examining territories sharing a border (therefore, neighborhood is here defined as contiguity). The pattern in Figure 13 presents example of spatially clustered or scattered territories.

Figure 13. Pattern of various signs of autocorrelation



Before defining contiguity, we must define the spatial matrix weight in order to know which order of contiguity that will give the maximum correlation. Here, we present the order of queen contiguity until order 30 in Figure 14.

Figure 14. Moran's I index generated by contiguity weight matrix

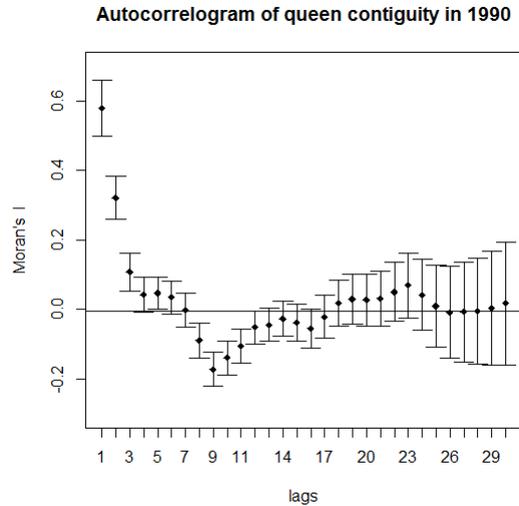


Figure 14 shows that the autocorrelogram for the 3 periods are almost the same (see Appendix 3), both the correlation value (Moran's I index) and the confidence interval of the Moran's I index. There is a great decrease between order 1 and order 2 and between order 2 and order 3. For the next 4 orders, there are no significance fluctuation, but from the 8th order until 17th order the correlations are negative, but they are still close to 0. This indicates a spatial random pattern with there is a dispersion tendency from 8th to 17th order. The highest correlation is at order 1, so we can use contiguity of order 1 to obtain join count statistic.

To apply the join count test on our data, we transform the 5 groups into pairwise combinations: for example group 1 and the remaining, group 2 and the remaining, and so on. Then we can compare which groups are more spatially clustered than the rest of the territories. Table 2 shows the p-value of joint count test applied to pairwise grouping as mentioned before.

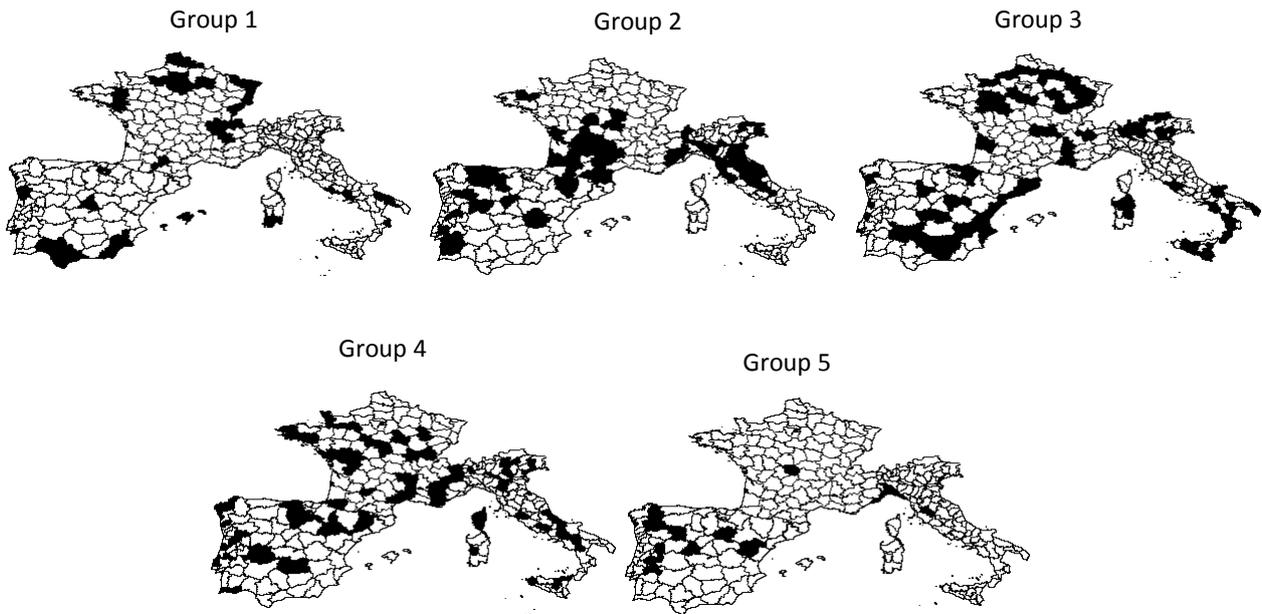
Table 2. P-value of joint count test

Group(i)	P-value based on queen contiguity	
	jc(i)	jc≠i
1	< 2.2e-16	5.614e-13
2	9.722e-13	2.713e-05
3	7.292e-07	4.827e-04
4	0.6895	0.7303
5	1.538e-05	0.1820

Alternative hypothesis: greater

The use of alternative hypothesis greater aimed at measuring the grouping/clustering, in the other words we want to test if there is a positive autocorrelation between territories. Based on the Table 2, all groups have a significant positive autocorrelation, except group 4. But the most clustered group are group 1 and 2, because they have a very small p-value. The maps below show the dispersion of the territories over each group.

Figure 15. Map of one group vs the rest



3. Markov Chain Approach

In order to refine the preceding analysis, we consider here a modelization by discrete time homogeneous markov chains whose states are defined as the classes of subsection 2.2. Namely, we have 5 states : young, rather young, moderate, rather old and old. We are interested in estimating the probability of the displacement of the territories from one state to another state. With the formula written in section 4 in chapter 1, we can estimate the transition matrix.

			1	2	3	4	5
P	=	1	0.43	0.48	0.09	0.00	0.00
		2	0.01	0.22	0.64	0.12	0.01
		3	0.00	0.04	0.17	0.70	0.10
		4	0.00	0.01	0.04	0.34	0.60
		5	0.00	0.00	0.00	0.04	0.96

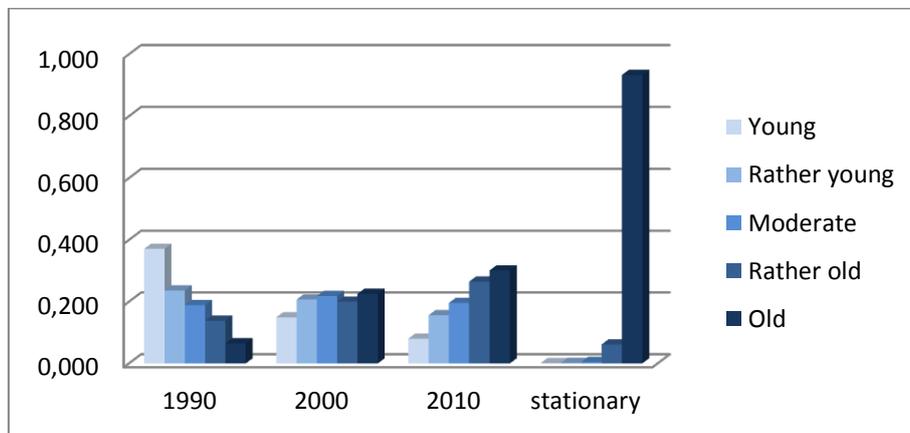
For example, if a territory is in state 1, the probability that it will stay in state 1 is 43% and the probability it will move to state 2 is 48%. We can see that, regardless of their starting

state all territories have a greater probability to move to the older state than to remain in the same state, except for state 5. So, based on the transition probability matrix above, we see that all territories tend to ageing. The stationary probability of each state in the long run is given by.

$$\mu = [0.000 \quad 0.001 \quad 0.004 \quad 0.061 \quad 0.933]$$

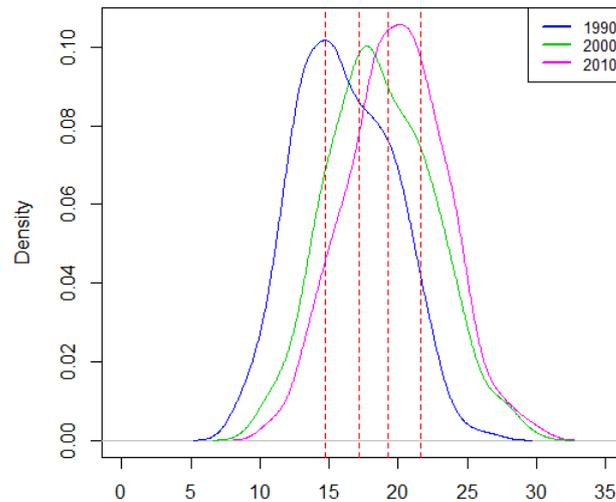
Therefore, in the long run, there will be no territory in state 1 and also the probability of a territory being in state 2, 3, or 4 is very small. This shows the convergence of ageing in the studied Mediterranean areas: in the future almost all territories will be in state 5 or have a big proportion of people aged 65 years and over. To make easier the comparison between the probabilities of each state in 1990, 2000, 2010, and in the long run, we can refer to Figure 16 below.

Figure 16. Comparison of the probability of each class and the stationary probability



In 1990, the young state dominated all territories in Mediterranean countries, followed by state 2 until state 5 gradually. In 2000, there were a decrease of the probabilities of state 1 and 2, while that of other states increased, particularly that state 5. The situation in 2010 was the opposite of that of 1990, where state 1 had the smallest probability and state 5 has the biggest one. The probability of each state in the long run is in line with the trend observed from 1990 to 2010, which is that all territories tend to ageing. This result strengthens the previous conclusion of subsection 2.4. But there is a problem in classification that will be shown in the Figure 17.

Figure 17. Density curve of people aged 65 years and over in each period



The red dotted lines show the quintile of the data, and the intervals between them define the state of the territories. If we look more closely, for example the rather old territories in 1990 can be considered as young territories in 2010, because we treat all territories in the same way. To avoid it, we will normalize the data based on several scenarios.

1.1 Using Data Normalization.

The objective of doing normalization is removing the effect of different periods. Inspired by (Rey, 2011), we will use three normalizations, as it is done by:

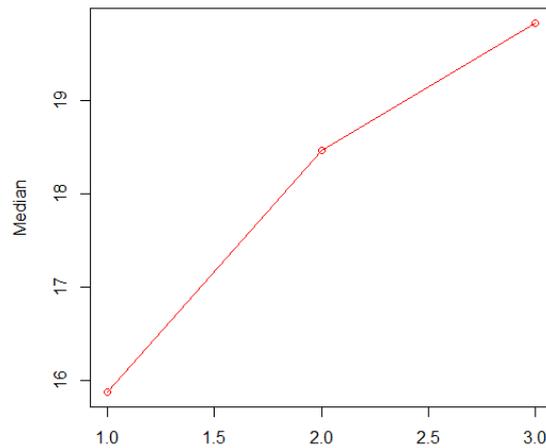
1. Normalizing each territory based on the median of all territories.
2. Normalizing each territory based on the median of the territories of each country
3. Normalizing each territory based on the median of the neighbors' territories.

The reason for using median in each normalization is that median is more robust (less sensitive to outliers in the data) than mean.

- a. Normalizing each territory based on the median of all territories (normalization 1)

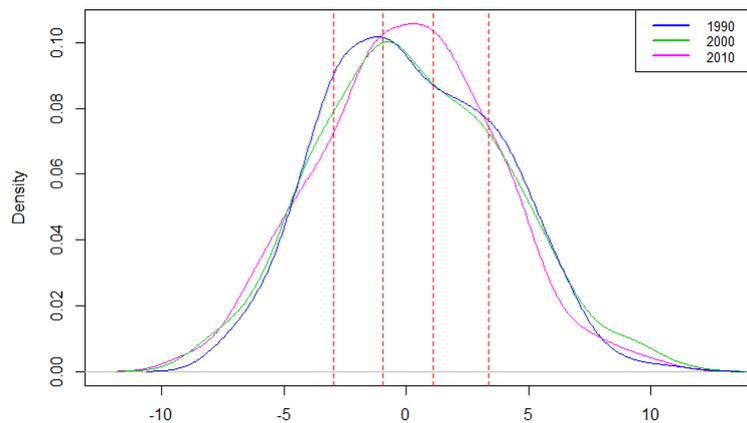
Normalizing each territory based on the median of all territories means that we reduced the proportion of people aged 65 years and over of each territory by the median indicator of all territories in the same year. By doing this normalization, we remove the effect of the period on the overall territories.

Figure 18. Median of overall territories each year



The overall territories in 1990 has the smallest median compared to that of 2000 and 2010. So, by reducing each territory by the overall median, we eliminate the effect of the period. The following graph shows the density curve of the data after reduced by the median.

Figure 19. Density curve of people aged 65 years and over in each period after normalization 1



After reducing by median each period (Normalization 1) and classify by the quantile of overall periods, the density curve of each period is located in the adjacent area. There is no more mis-classification, the old state in 1990 will be classified as old too in 2010. The density curves in 1990 and 2000 do not look like a Gaussian density, whereas the curve of 2010 seems look like a Gaussian distribution. Kurtosis and skewness of each density curves as well as Shapiro-Wilk test of normality are given in Table 3.

Table 3. Properties of density curve by normalization 1

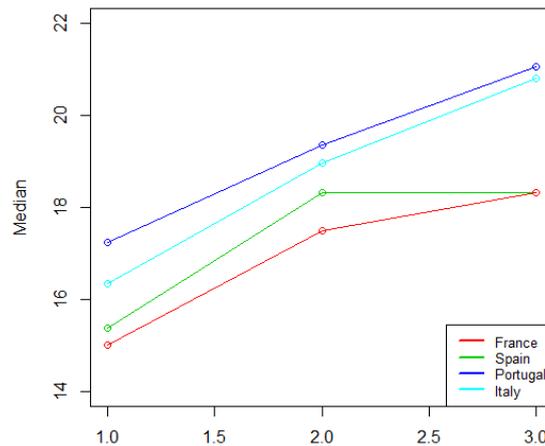
Year	Kurtosis	Skewness	P-value of Shapiro Test
1990	-0.534	0.177	0.058
2000	-0.326	0.185	0.337
2010	-0.177	0.044	0.887

Kurtosis of normal distribution is 0. Kurtosis of all curves is less than 0 which means that all curves are flatter than a normal distribution with a wider peak. Skewness of all curves is more than 0, it means that all curves have right skewed distribution or most values are concentrated on the left of the median, with extreme values to the right. Based on this result, the distribution are not Gaussian. Henceforth, we use Shapiro-Wilk test to examine the normality of each curve. We test whether the density curve normally distributed or not (H_0 : density curve follows the normal distribution). According to p-value of Shapiro-Wilk test, the distribution in 1990 is indeed not Gaussian, but there is no evidence of non Gaussianity for the distributions in 2000 and 2010.

b. Normalizing each territory based on the median of the country (normalization 2)

Normalizing each territory based on the median of the country conducted by reducing the proportion of people aged 65 years and over in each territory of each country by the median of the territories in this country.

Figure 20. Median of countries each year



France has the smallest median among the other countries and Portugal has the biggest one. It is clearly seen that Portugal, Italy and France were experiencing a convergence of ageing over the observation period, but not so with Spain. The territories in Spain rejuvenated between 2000 to 2010.

Using normalization 2 leads to the fact that data are detached from the effects of country ageing and time. Figure 21 shows the density of data for this normalization. All the density curves do not seem to follow the Gaussian distribution but rather a mixture of Gaussian distributions. The indicators presented in Table 4.

Figure 21. Density curve of people aged 65 years and over in each period after normalization 2

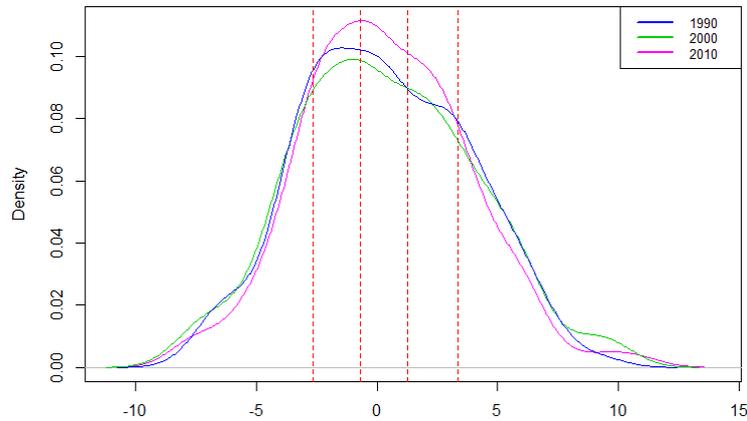


Table 4. Properties of density curve by normalization 2

Year	Kurtosis	Skewness	P-value of Shapiro Test
1990	-0.553	0.103	0.077
2000	-0.322	0.186	0.105
2010	0.115	0.213	0.283

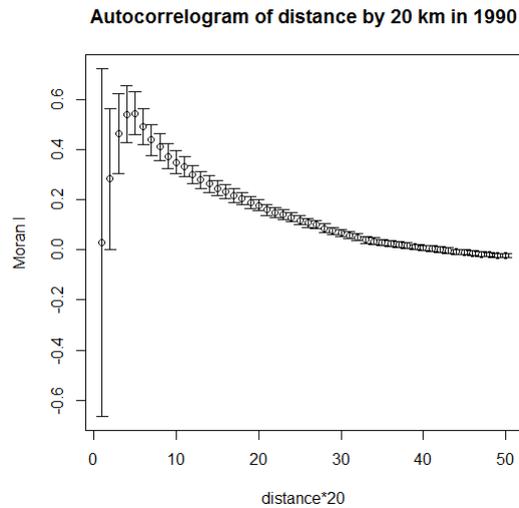
As for normalization 1, the curves are also than a normal density and have a wider peak. Based on the skewness values, all distributions are right skewed or most value are concentrated on the left of the median, with extreme values to the right. So we can conclude that, there is no Gaussian form. This is not confirmed by Shapiro-Wilk test, for which all the curves are considered as Gaussian densities (with $\alpha = 5\%$).

c. Normalizing each territory based on the median of the neighbors (normalization 3)

We normalize each territory based on the median of the neighbors by reducing the proportion of people aged more than 65 years in each territory by the median of its neighbors. Relevance of such a normalization is because we have seen in our previous analyses that a spatial pattern appears in the trajectory of ageing. So in this subsection, we will consider the influence of neighbors on the ageing of a territory. To involve neighbors in analysis, we need to define a neighborhood or equivalently a spatial weight matrix which takes into account the surrounding territories.

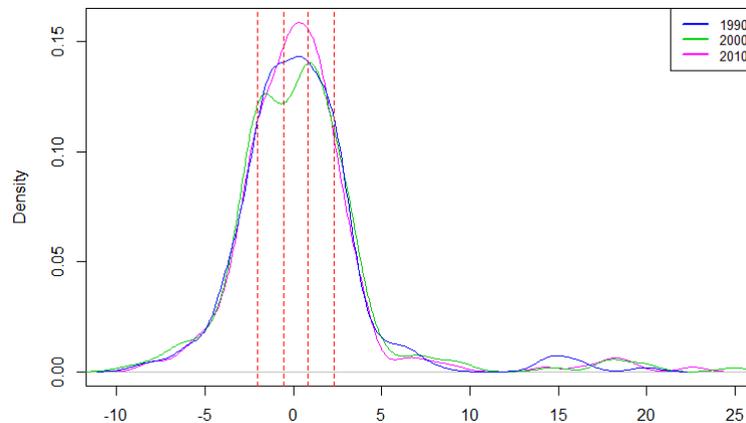
As we see in Figure 22, we traced Moran's I index with 20 km-interval until order 50. So we examined 1000 km distance in total. Figure 22 shows that the Moran's I index increased until it reaches the maximum value at point 5 and then it decreases until the last order. So we can say that 100 ($5 \cdot 20$) km is the best distance to be used to form the weight matrix based on the distance.

Figure 22. Moran's I index generated by distance weight matrix



To normalize the data by median of neighbors, we finally used the weight matrix based on distance, because we want to see the influence of the neighbors in a certain radius from a territory. We consider the neighbors of a territory are those which have the distance less than 100 km from the centroid of certain territory. After reducing by the median of its neighbors, we obtain the following density curve.

Figure 23. Density curve of people aged 65 years and over in each period after normalization 3



Based on the graph above, we can see that there are many outliers, in 1990, 2000, and 2010. If we consider the territories with a 7% of 65 years and more (after we reduce the proportion of people aged more than 65 years by the median of the neighbors) as an outlier, we find that there are 10 outlier territories, 9 of them are in Spain and the other in Portugal. It is clear that all the density curves above are not Gaussian. But it will be more convincing if we look at the following table.

Table 5. Properties of density curve by normalization 3

Year	Kurtosis	Skewness	P-value of Shapiro Test
1990	7.513	1.896	9.526e-16
2000	9.933	2.332	<2.20e-16
2010	10.805	2.513	<2.20e-16

1.2 Estimation of Transition Matrix based on Data Normalization

To estimate the transition matrix, Markov chain approach which has been explained in Chapter Data and Metodology. By that formula, we obtained the following transition matrix by each normalization.

a. Transition matrix by normalization 1

Following the transition matrix is obtained by applying the formula on the data that has been treated by normalization 1.

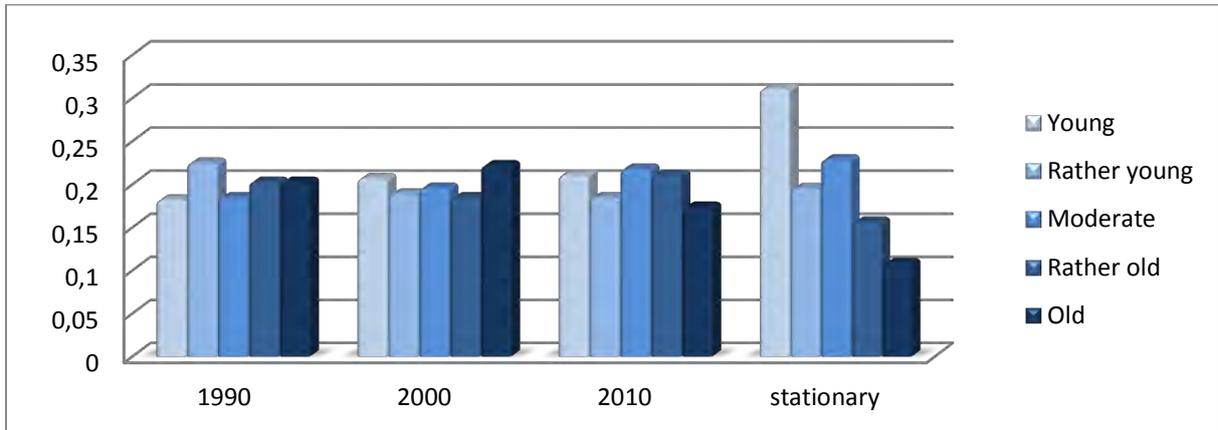
		1	2	3	4	5
P=	1	0.87	0.12	0.01	0.00	0.00
	2	0.14	0.68	0.18	0.00	0.00
	3	0.05	0.09	0.73	0.12	0.01
	4	0.01	0.03	0.15	0.70	0.11
	5	0.00	0.00	0.00	0.18	0.82

The result is different from the result before doing normalization. We observe that the probability of a territory remains in the same state is much larger than moving to the other states. The pattern which is appearing here is a territory can move to one previous state or one next state from the current state. It is also possible to move to two previous state or two next state, but the probability is less than the probability to move to one previous state or one next state. For example, the probability of territories remains in state 3 is 0.73. It can move to state 2 or 4 with probability 0.09 and 0.12 respectively. It can also move to state 1 or 5 with probability 0.05 and 0.01 (which is less than those in state 2 and 4). So, the reduction by median of all territories in each year will give the different result significantly. The main result is the majority of territories remain in their initial state. The stationary probability of a territory to be in certain class is as follows.

$$\mu = [0.311 \quad 0.195 \quad 0.229 \quad 0.156 \quad 0.108]$$

After we eliminate the influence of period and overall territories, we do not get the same conclusion as before. By normalization 2, in the long run the territories have an enough big probability to stay in their current state instead of ageing. So, period plays an important role in ageing convergence, if we remove the trend of period there will be no convergence of ageing. The following graphic shows the comparison of the probabilities.

Figure 24. Comparison of the probability of each class and the stationary probability by normalization 1



Based on the Figure 22, state 1 (young) increased over time, and in long run state 1 has the biggest probability. While the other states have a random fluctuation, so we can't conclude anything from this graph.

b. Transition matrix by normalization 2

The transition matrix obtained by normalization 2 is presented below.

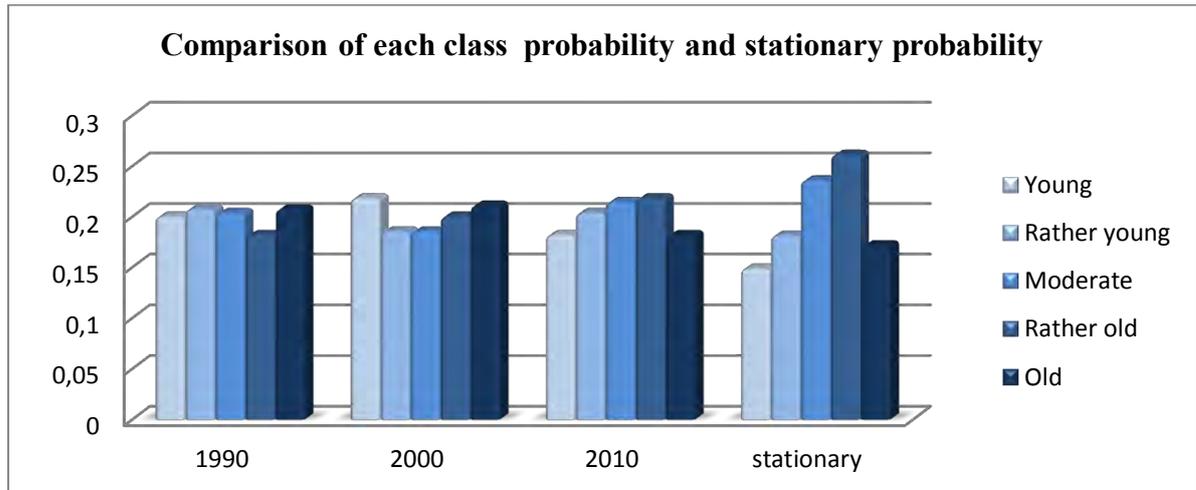
		1	2	3	4	5
	1	0.86	0.14	0.00	0.00	0.00
	2	0.08	0.72	0.19	0.00	0.00
P=	3	0.01	0.11	0.66	0.21	0.00
	4	0.01	0.01	0.17	0.70	0.10
	5	0.00	0.00	0.00	0.16	0.84

We get the same result with the two previous normalizations, which is the probability of staying in the same state is much bigger than the probability to move to the other state. State 4 and 5 has the possibility to go to all states, while state 1, 2 and 5 can move to certain states. The stationary probability of each class is shown in vector below.

$$\mu = [0.149 \quad 0.181 \quad 0.236 \quad 0.261 \quad 0.172]$$

To make the comparison clearer, we can see the following graph.

Figure 25. Comparison of the probability of each class and the stationary probability by normalization 2



In 1990, the probability of being in state 4 is the smallest among the other state, and the probability is almost same. The young state increased in 2000 and then decreased in 2010. In the long run, the biggest probability is being in rather old state , and the smallest is being in young state.

c. Transition matrix by normalization 3

The transition probability matrix of normalization 3 by median is as follows.

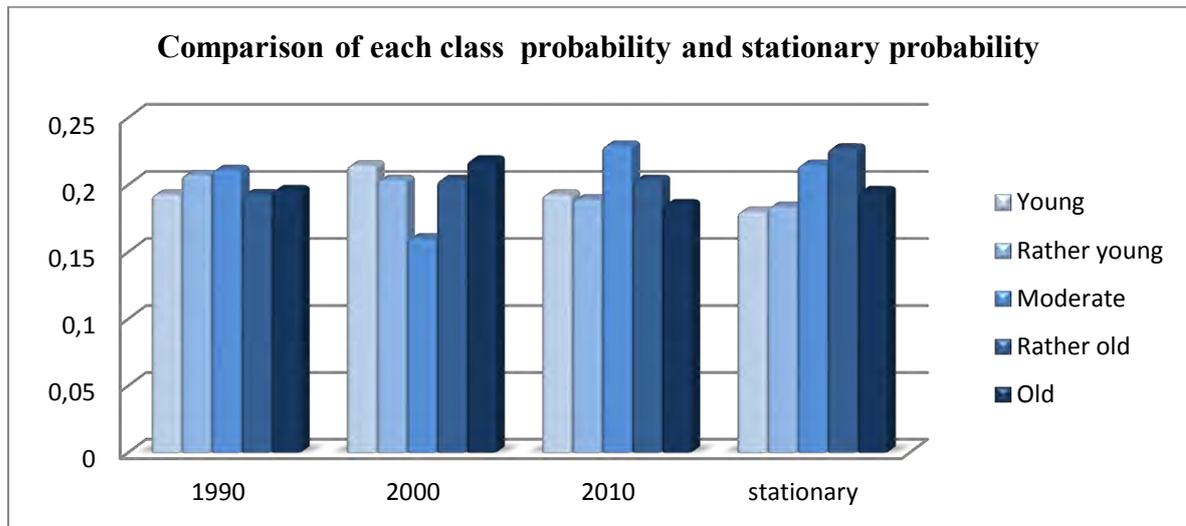
	1	2	3	4	5
1	0.79	0.20	0.02	0.00	0.00
2	0.21	0.62	0.16	0.01	0.00
P=3	0.00	0.14	0.64	0.20	0.03
4	0.00	0.02	0.19	0.66	0.13
5	0.00	0.00	0.01	0.17	0.82

We find the same conclusion with the previous normalization, that is the probability of a state remains in the initial state is bigger than the probability to move to the other state. The stationary probability of each class is as follows.

$$\mu = [0.180 \quad 0.184 \quad 0.215 \quad 0.227 \quad 0.195]$$

In the long run, the probability of the territories will be in a certain state is almost equal, about 20 %. So by eliminating the effect of neighbors, the territories in Mediterranean will not be heading to the same direction. They will be stay in each state with almost same probability.

Figure 26. Comparison of the probability of each class and the stationary probability by normalization 3



The result is very similar to the result of normalization 2, in 2010 state 3 has probability smallest among the other states. In the stationary distribution, the probability of being in state 1 and 2 is a roughly similar.

After treating data with different normalizations, we arrive at a conclusion that convergence of ageing will happen if we don't remove the effect of period from the analysis. As we observed, the convergence of ageing only occurs before normalizing data and not in normalizations. As a remark, the stationary probability is obtained when the transition matrix has been in a steady state. But we present the stationary probability in the previous analysis as a comparison with the actual transition matrix.

4. Modeling of Ageing by Regression Models

We are interested in modelling the influence of several variables to the proportion of people aged 65 years and over at period t . To realize that, we will build several regression models those have different explanatory variables in each model. The explanatory variables we are interested in are:

1. The proportion of people aged 65 years and over in period t (peps_y) as the dependent variable.
2. The proportion of people aged 65 years and over in period $t-1$ (peps_x)
3. Time= $\begin{cases} 0, & \text{shows the transition between 1990 and 2000} \\ 1, & \text{shows the transition between 2000 and 2010} \end{cases}$
4. Neighbor's influence in period t that is contained in variable we2. We apply queen contiguity in defining the territories as a neighbor because we want each territory had the neighbor. If we use the weighted matrix based on distance, the territories that have more than 100 km will not have neighbors.
5. Neighbor's influence in period $t-1$ that is contained in variable we1.
6. Interaction between time*peps_x

7. Interaction between time*we2

Before modelling those variables by regression method, we need to construct a data structure that organizes all those variables in a data set. The data set is shown below:

Table 6. Data structure used in linear regression analysis

Peps_y	Peps_x	Time	We1	We2	Name
P65_00	P65_90	0	We_90	We_00	Bas-Rhin
P65_00	P65_90	0	We_90	We_00	Haut-Rhin
⋮					
P65_00	P65_90	0	We_90	We_00	Bari
P65_10	P65_00	1	We_00	We_10	Bas_Rhin
P65_10	P65_00	1	We_00	We_10	Haut-Rhin
⋮					
P65_10	P65_00	1	We_00	We_10	Bari

Next we will model the dataset with linear regression and linear mixed model.

4.1 Linear Regression Models

The first method will be conducted to analyze the relationship between variables is a linear regression model. The estimation of parameters by ordinary least square estimation is summarized in Table 7.

The regression models in Table 7 don't include the neighbors of territories in the previous year, because we assume that the ageing of this year is much more affected by the ageing of previous year than the influence of the neighbors of the previous year.

Based on Table 7, all models have high R-squared (more than 90%) with almost all explanatory variables are significant.

Table 7. Parameter estimation of linear regression by ordinary least squares

Coefficient	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Intercept	3.221*	1.531*	0.391	0.981*	0.875*	0.904*	0.872*	0.899*
peps_x	0.963*	0.865*	0.945*	0.866*	0.978*	0.978*	0.914*	0.866*
Time	-1.420*	-1.349*	1.089*	-0.180	0.065			
we2		0.175*	0.167*	0.203*	0.112*	0.111*	0.169*	0.208*
Time*peps_x			-0.141*		-0.202*	-0.202*	-0.081*	
Time*we2				-0.061*	0.108*	0.111*		-0.070*
R ²	0.9126	0.9231	0.9278	0.9237	0.9288	0.9288	0.9269	0.9237
BIC	1688.331	1624.491	1596.032	1626.541	1594.694	1588.398	1596.505	1620.33

* significant at $\alpha = 0.05$

In the common use, the best model selected based on the smallest AIC or BIC. But in this study, we prefer to use BIC as the judgment in choosing the best model because we emphasize the interpretation of the model instead of the prediction. In this study, as we

can see in the table above, the smaller BIC of a model, the greater R-squared obtained. Here, model 6 meets the requirements as the best model based on its BIC. But we want to keep time as an explanatory variable, because we consider time that represents the transition is an important factor. Before selecting which model will be used in further analysis, we present the comparison of model 5 and 6 on Table 8.

Table 8. Comparison between model 5 and model 6 in model significance

Coefficient	Model 5 Estimate	P-value	Model 6	P-value
Intercept	0.875	0.020	0.904	0.001
peps_x	0.978	< 2e-16	0.978	< 2e-16
Time	0.065	0.907		
we2	0.112	6.08e-05	0.111	1.01e-05
Time*peps_x	-0.202	8.22e-10	-0.202	7.60e-10
Time*we2	0.108	0.00598	0.111	1.58e-04
R ²	0.9288		0.9288	
BIC	1594.694		1588.398	

As we see in the table above, the contribution of time in model 5 is small compared to the other variables. In addition, the p-value of time is almost 1, we can strongly stated that time is not significant. To see if the reduction of variable time was justified, we compare the simpler model (model 6) with model 5 using ANOVA.

Table 9. Comparison between model 5 and model 6 using ANOVA

Model	Res. DF	RSS	DF	Sum of Squares	F	P-value
5	544	539.79				
6	545	539.80	-1	-0.0135	0.0137	0.907

The purpose of doing this comparison is to test whether the coefficient for the time differs from zero, or it could also be said that the model with time is better than the model without time. Since our p-value of F test is much bigger than α (0.05), so we can say that the model with time is not significantly different with the model without time. Based on this result, the simpler model (model 6) would generally be more appropriate to explain the relationship between the dependent and independent variables. The equation of the selected model can be written as follows:

$$\widehat{peps}_y = 0.904 + 0.978 * peps_x + 0.111 * we2 - 0.202 * (time * peps_x) + 0.111 * (time * we2)$$

If time=0,

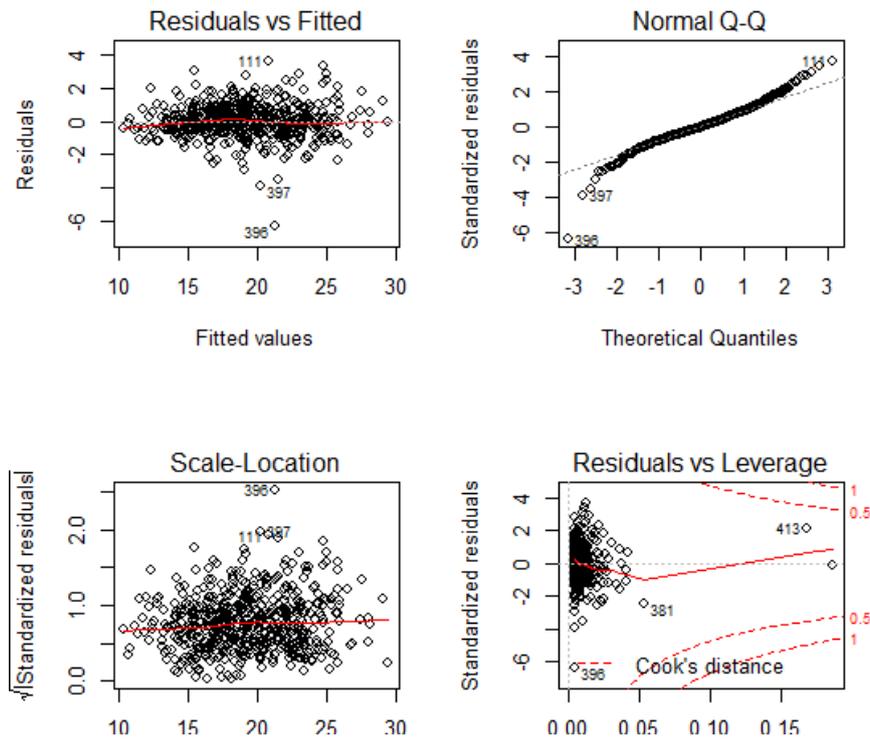
$$\widehat{peps}_y = . + . * peps_x + . * we$$

If time=1,

$$\widehat{peps}_y = . + . * peps_x + . * we$$

The coefficient for `peps_x` estimates the change in `peps_y` when `we2` is held constant. And so does `we2`. In the model with `time=0`, `peps_x` has a more important contribution than in model with `time=1`. But, the opposite happened on variable `we2`, where the coefficient is twice greater when `time=1`. In both models with `time=0` and `1`, the proportion of the previous year is much more contributing to the proportion in the certain territory in the current year than their neighbors in the same year. Then, we will evaluate the residuals of the model 6 related to the assumptions of linear regression.

Figure 27. Diagnostic plots for residuals based on model 6



Based on the 4 graphs above, several important things can be observed.

- The first is the plot between residuals and the fitted value that indicates the variability of residuals. It seems that the residuals and the fitted value are uncorrelated, so we can say that the residuals satisfy identical assumption (homoscedastic).
- The second plot indicates that the residuals non-normally distributed, they don't scatter following the straight line. To ensure the normality of residual, Shapiro test was performed and it yields $p\text{-value} = 9.78e-11$ that means the residuals is non-normal.
- The fourth graph (residuals vs leverage plot) shows the way of checking on extreme values. The presence of extreme values might influence on parameter estimation. A data points with high leverage has a potential to move the regression line. Regression line represents the regression equation that contains the coefficient of the explanatory variables. High leverage points will make our parameter estimation inaccurate and it can be misleading our conclusion. We can find several influential points which are 381, 396, and 413. But not all influential points considered as outliers. By conducting

Bonferroni test for outliers in linear models, there are 2 territories that are considered as outlier: Guadalajara and Toledo in time = 1. In the same graph, we found Cook's distance that indicates the actual influence of each point on the slope of the regression line. Cook's distance can be interpreted as a measure of how different the regression coefficients (including the intercept) would be if the particular observation is left out of the analysis altogether. Any observation for which the Cook's distance is close to one or more, or that is substantially larger than other Cook's distances, calls for investigation (Maindonald & Braun, 2007).

After checking model, then we create a residual map in Figure 28.

Figure 28. Residual Map of Linear Regression Model between 1990 and 2000

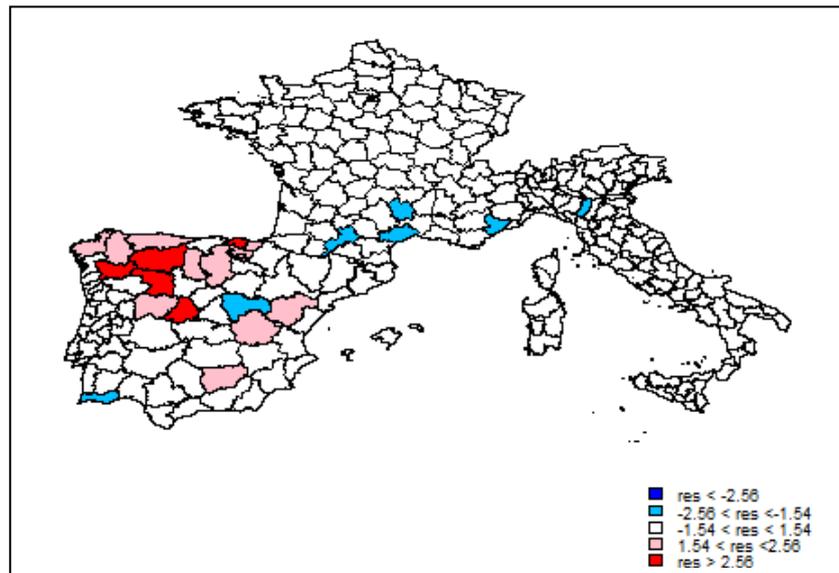
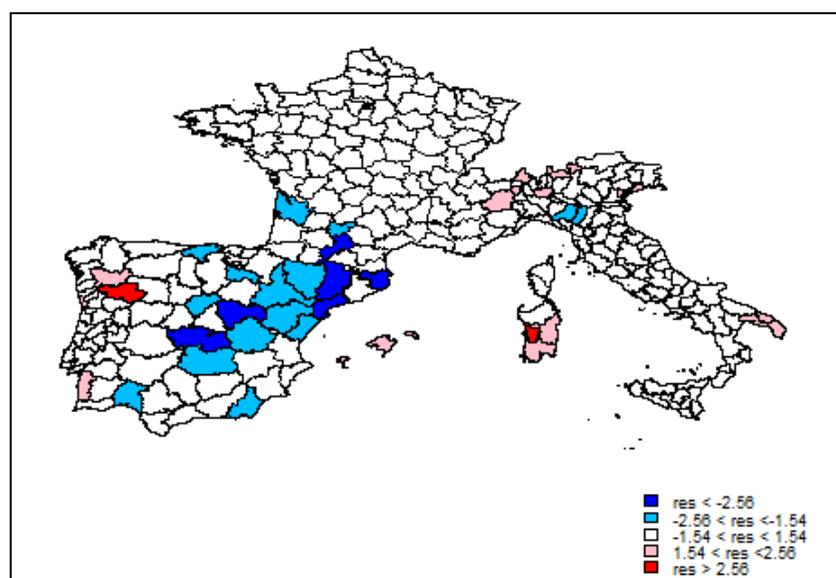


Figure 29. Residual Map of Linear Regression Model between 2000 and 2010



Based on the residual map in Figure 28, Spain has some territories that have residuals more than 1.54 that grouping in the southeast and northern part. It means that those territories' predictive values are older than their actual values. In Portugal, there is a territory that has predictive value less than their actual value between 1990 and 2000, it is Algarve which is located in the most southern Portugal.

Figure 29 shows that there is the grouping of residuals which less than -1.54 in Spain. There is declining of residuals in Spain, some territories have residuals less than -1.54 between 2000 and 2010. The opposite occurs in Italy, based on model 6, some territories in Italy have residuals more than 1.54.

Based on the 8 models we have, there are some conclusions as follows:

1. In some models (especially models with independent variables more than 2), time variable is not significant.
2. There is multicollinearity in some independent variables of some models ($VIF > 10$). But if we do a further analysis, we find that there are several models that have high VIF. VIF is an indicator of the presence of multicollinearity in a model. Variable time, time*peps_x and time*we2 in model 3, 4, 5 and 6 have VIF value more than 10
3. The assumption of normality of residual is not satisfied, so the test which is performed by relying on normality assumption is inaccurate.

Because of the reasons above, we will do another approach in analyzing the data set which known as linear mixed model.

4.2 Linear Mixed Models

The main difference between linear mixed model and regression classic is how we treat the explanatory variable. In the linear mixed model, there are two types of variables, they are fixed variable and random variable. We use the territories as random effect, the remaining variables as fixed effects. The parameter estimations are summarized in the following table.

Table 10. Parameter estimation of linear mixed model by maximum likelihood

Coefficient	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Intercept	3.237*	1.352*	-0.061	0.639	0.363	0.558	0.557	0.685*
Peps_x	0.962*	0.844*	0.912*	0.841*	0.940*	0.940*	0.881*	0.841*
Time	-1.417*	-1.323*	1.335*	0.099	0.423			
We2		0.202*	0.219*	0.243*	0.173*	0.162*	0.214*	0.240*
Time*peps_x			-0.152*		-0.209*	-0.207*	-0.079*	
Time*we2				-0.073*	0.098*	0.118*		-0.068*
BIC	1694.389	1624.633	1578.633	1623.503	1575.973	1570.507	1586.35	1617.23

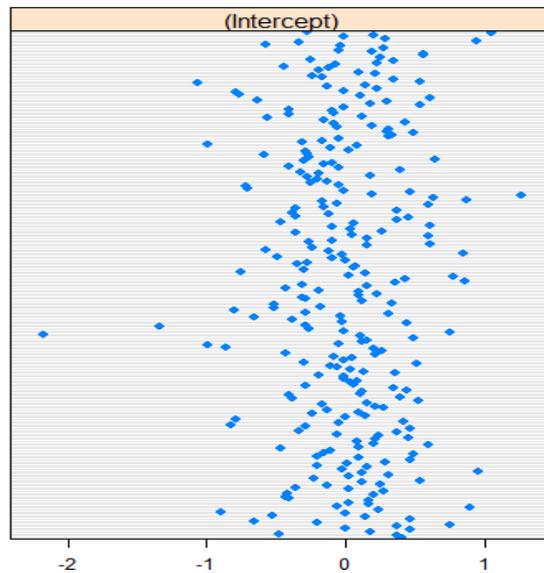
* significant at $\alpha = 0.05$

The same with the previous analysis in linear regression, we choose the best model based on BIC. We also perform a comparison between model 5 and model 6, and the result is

there is no significant difference between these models. So, model 6 is chosen as the best model and will be used in the further analysis.

In model 6, we have territory as a random effect. It means that the intercept of model is different for each territory. The standard deviation of the random effect is 0.578 and the residual's is 0.814. Based on it, we can say that the variability of the dependent variable explained by territories is less than explained by residuals. The coefficients / intercept of random effect are different in each territory that presented in the following graph. The intercepts varied between -2 to 1.

Figure 30. Intercept of random effect



The equation of the linear mixed model can be written as follows.

$$Y_{ti} = \underbrace{\beta_1 X_{ti}^{(1)} + \beta_2 X_{ti}^{(2)} + \beta_3 X_{ti}^{(3)} + \dots + \beta_p X_{ti}^{(p)}}_{\text{Fixed effect}} + \underbrace{u_{1i} Z_{ti}^{(1)} + \dots + u_{qi} Z_{ti}^{(q)}}_{\text{Random effect}} + \varepsilon_{ti}$$

For example, we take Paris to modeled.

$$\hat{Y}_{ti} = 0.94 * peps_x + 0.162 * we2 - 0.20 * (time * peps_x) + 0.118 * (time * we2) - 0.138$$

$$\text{If time}=0, \text{ then } \hat{Y}_{ti} = 0.94 * peps_x + 0.162 * we2 - 0.138$$

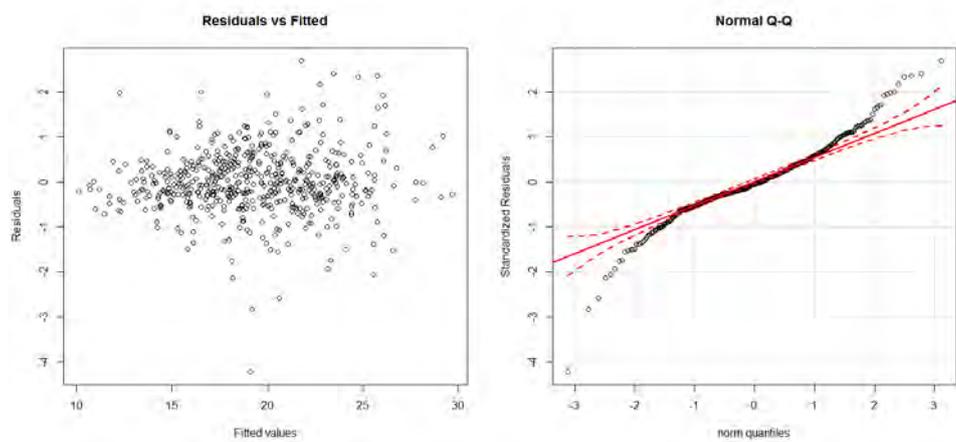
$$\text{If time}=1, \text{ then } \hat{Y}_{ti} = 0.74 * peps_x + 0.280 * we2 - 0.138$$

The model above lead the similar result and interpretation with the model obtained in linear regression, except the intercept because we allow the intercept varies for each random effect (territory).

The residual plot of model 6 is presented in Figure 31. Residual plot generated by this model seems to have a particular pattern. The residuals tend to form a bell pattern, which

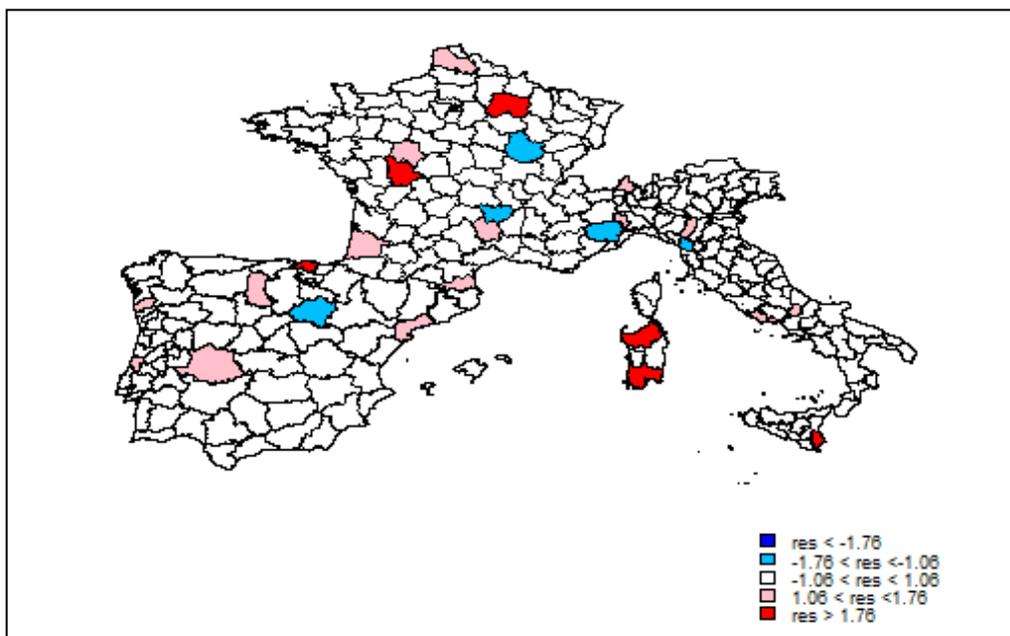
is sharp in the left and getting wider to the right. It means the variability of residuals gets larger along with larger fitted values. This indicates that the variance of residuals is not constant.

Figure 31. Diagnostic plots for residuals



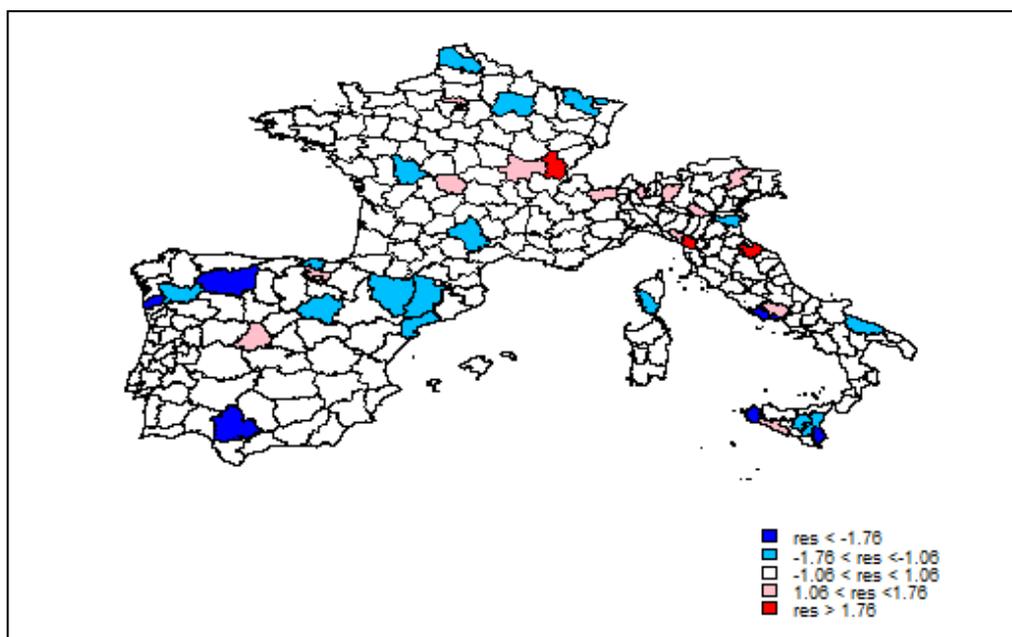
For the second graphic, it seems that the residuals don't follow the straight line and also there are many territories that have residuals beyond the confidence intervals. This indicates that residuals don't satisfy the normality assumption. The Shapiro test of residuals gives p-value less than α (0.05), so we can conclude that residuals have non-normal distribution. After checking model, then we create a residual map on the following picture.

Figure 32. Residual map of linear mixed model between 1990 and 2000



We obtain quite different result with the linear regression model between 1990 and 2000. The standard deviation of residuals is declined, consequently we have a range between classes is shorter than they are in linear regression, as well as between 2000 and 2010. Based on the diminution of residuals, we can say that linear mixed method produces better predicted values than linear regression. The linear mixed model is best fit to the data and it is explaining the data better than the linear regression model. As we can see in the map above, most of the territories in Mediterranean country are appropriate modeled by linear mixed model because the variation of residuals is small (between -1.06 and 1.06). But, there are some territories that have residuals vary more than 1.76, they are Avila, Cuenca, Leon, Teruel, Vizcaya and Zamora.

Figure 33. Residual map of linear mixed model between 2000 and 2010



The map above shows the residuals dispersion between 2000 and 2010 using linear mixed method. The distribution of classes is more varied than the earlier map. There are more territories that have residuals less or more than 1.76, but we can observe that some territories rejuvenate between 2000 and 2010, such as Girona, Guadalajara, Lleida, Segovia, Teruel, Toledo, etc.

4.3 Relation of the Best Model with Ageing Convergence in Mediterranean

Recall the coefficient of variable $peps_x$ and $we2$ in the linear mixed model as follows.

Table 11. Coefficient of variable $peps_x$ and $we2$

Time	$peps_x$	$we2$
0 (from 1990 to 2000)	0.94	0.162
1 (from 2000 to 2010)	0.74	0.280

From 1990 to 2000, the proportion of people aged 65 years and over in the current period is very similar to the proportion in the previous period because the proportion in the previous period has a very big coefficient, it closes to 1. Contribution of the neighbors in determining the proportion of the current period is only 16.2%. But from 2000 to 2010, the influence of the proportion in the previous period decline and the contribution of the neighbors increase. Automatically, the correlation between the territories and their neighbors will also increase over time. Based on this phenomena, we can observe that as time passed, the territories will have the more similar behavior with the neighbors. So we can conclude that there is convergence of ageing in the Mediterranean based on this model.

CHAPTER 3

CONCLUSION AND DISCUSSION

1. Conclusion

Based on the analysis which we have been performed in Chapter 2 Analysis of result, we can summarize the conclusion as follows:

- a. Our first analysis is exploring the proportion of people aged 65 years and over using median. The median shows that ageing in studied Mediterranean countries converge or all territories tend to ageing over time.

Apart from that, we examine the territories that maybe have similar behavior of ageing over time, but not in general manner. In the next analysis, we group the territories by hierarchical clustering method and we obtained the following group of trajectories:

- Group 1: young territories remain young
- Group 2: rather old territories change into old
- Group 3: young territories change into moderate
- Group 4: rather young territories change into rather old
- Group 5: old territories remain old

The median of each group of territories increases over time, although there are groups which remain in the same age class (i.e. red and pink group). All the groups are heading to the same direction (ageing). So we can draw a conclusion that all territories converge in ageing, but not at the same rate.

We continued the analysis of the Markov chain approach to obtain the probability of each territory to move to another state or to stay in the current state. Our result confirmed that all states have a bigger probability to move to the older state than to remain in the same state. In the long run, we found that there will be no territory in young state and the probability of a territory will be in rather young, moderate and rather old is very small. In the long run, the Markov chain approach predicts in the future almost all territories in Mediterranean will tend to be in old state or have a very big proportion of people aged 65 years and over.

- b. After modeling the proportion of people aged 65 years and above with regression model, we decided that linear mixed model is explaining the data better than the linear regression model because we found the variance of the residuals decreased. Based on this linear mixed model we observed that the correlation between the territories and their neighbors increase over time. As the result, the territories will have the more similar behavior with the neighbors as time passed. So we can conclude that there is convergence of ageing in the studied Mediterranean countries based on this model.

2. Discussion

This research could be better if we can perform several improvements and analyses below:

- a. To study about the convergence of ageing, actually we need more series of data in order not to lead to inaccurate conclusion. As a comparison, (Rey, 2001) conducted his research about US regional income over the 1924-1994 period, (Wilson, 2011) studied about global demographic convergence since 1950 until 2005.
- b. Perform another Markov chain approach, e.g. a continuous-time Markov chain that allows us to analyze the change of the process at any time along a continuous interval instead of a discrete time. We also can apply space-time dynamic Markov chain which is incorporating the spatial dimension in the Markov chain analysis.
- c. Our regression model performed based on only one variable: proportion of people aged 65 years and over in the Mediterranean. In my opinion, it would be better if there is more than one variable which is used to explain the proportion of ageing in the current period, for example fertility, mortality rate, life expectancy, etc.
- d. Using other approaches to estimate the regression parameter, e.g. generalized equation estimation (GEE) and because of the presence of spatial pattern, we can also applied Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), etc on our data.

APPENDIX

I. Literature Review

1. Kernel Density Estimator

Consider we have random quantity X that has probability density function f , the probability associated with X is given by the formula:

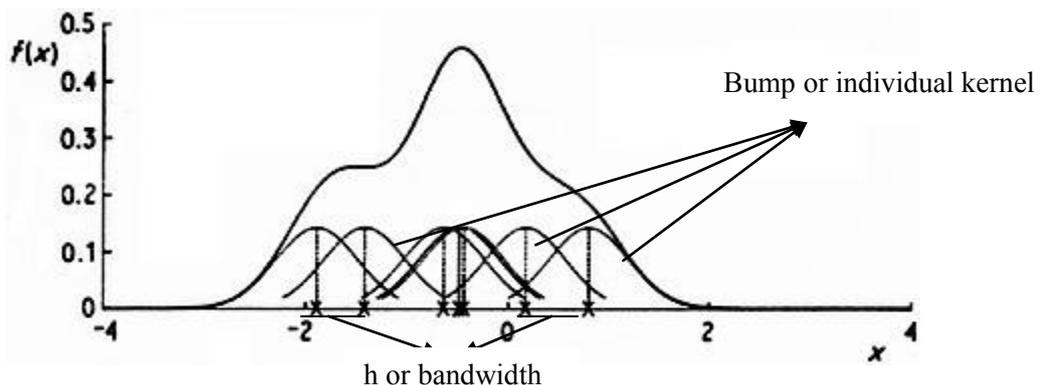
$$P(a < X < b) = \int_a^b f(d)dx \text{ for all } a < b$$

Suppose we have a set of observation which is assumed to be a sample from an unknown probability density function. Density estimation is the construction of an estimate of the density function of the observed data. There are two approaches to estimate density function, they are parametric and non parametric approach. The parametric approach assumes that the data are drawn from a known parametric family of distribution, for example normal distribution. The density of the data can be estimated by finding estimates μ and σ^2 from the data and then substituting the estimates into the normal density formula. Different from parametric approach, non parametric approach is less rigid assumptions and the data allowed to speak for themselves in determining the estimate of density function f .

The oldest and well known density estimator is the histogram. If we want a smooth estimator, the histogram is less suitable to be used because of its discontinuity. The density estimator used in R is a kernel estimator is defined by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

where K usually, but not always, will be a symmetric probability density function which determines the shape of the “bumps”. And h is the smoothing parameter or in R it is usually called bandwidth that determines the bump’s width. For more details, see the following picture (Silverman, 1986).



2. Hierarchical Clustering

In exploring the data of the proportion of people aged 65 years and over, we are interested in grouping the territories that have a similar behavior in ageing within a period of ten years

during the 20 years. To do that, we need a classification method which can group the territories in order to get some groups which have a similar behavior within them. Cluster analysis is one of many grouping technique that can minimize the homogeneity within a group and maximize the homogeneity between groups. So, we will get the territories which tend to behave the same in one group.

To perform cluster analysis, we require a measure of similarity between two territories. The most common technique used is measuring distance between two territories. We have recognized some formulas to calculate distance, e.g. Euclidean distance, Manhattan distance, Mahalanobis distance, etc. In this analysis, we need an actual distance between all pairs of territories, so we will apply Euclidean distance here. The formula of Euclidean distance of three data periods can be written as follows:

$$d = \sqrt{\sum_{i=1}^3 [x_i(t) - x_j(t)]^2}$$

An important remark is this distance measure is sensitive to different scales of measure. In general, variables with larger dispositions, for example standard deviation, have more impact on the similarity value (Hair, Black, Babin, & Anderson, 2009). To avoid to the misleading on conclusion, standardization will be accomplished here. The standardization formula which performed in R is:

$$z_i = \frac{x_i - \bar{x}}{\sqrt{\text{var}(x)}}$$

After finding a method to estimate the similarity, the next important step is defining how to form clusters. The procedure we choose must be able to group the territories that are most similar (based on a specified distance) into a group. A well known procedure to form clusters is hierarchical clustering. Hierarchical procedure classifies observations into a hierarchy or tree like structure. According to Hair, et al, two types hierarchical procedure is agglomerative and divisive. Agglomerative method treats each single observation as cluster and then they successively joined until remain only one cluster. Divisive method is the contrary of agglomerative method (Hair, Black, Babin, & Anderson, 2009). This step will produce a tree diagram, or we usually called it as a dendrogram.

In performing hierarchical clustering, we need a clustering algorithm that is used for defining how similarity is defined to form a cluster. There are many algorithms in agglomerative methods: single-linkage, complete-linkage, Ward's method, etc. In this study, we will run Ward's method. The aim of Ward's method is to form each possible number of groups $n, n-1, \dots, 1$, in a manner that would minimize the loss associated with each grouping (Ward, 1963).

The loss here can be represented by the sum of squares. At each step, the two clusters combined are those that minimize the increase in the total sum of squares across all variables in all clusters (Hair, Black, Babin, & Anderson, 2009). In addition, Ward suggests using his method in large scale number of observations ($n > 100$).

Now, we turn to how determine the optimal number of group / cluster. There is no specific way how we choose the optimal group. Hair, et al in their book propose some approaches, for example percentage changes in the heterogeneity (used by SPSS), root mean square standard

deviation (RMSSTD) and Cubic Clustering Criterion (CCC used by SAS). But here, we will use approaches named semi partial R-squared (SPRSQ)..

3. Spatial Autocorrelation

(Fischer & Getis, 2010) give a simple definition about spatial autocorrelation that it represents the relationship between nearby spatial units, as seen on maps, where each unit coded with a realization of a single variable. The other statistician defined spatial autocorrelation as self correlation, or values within a given variable are correlated, resulting in the variable being correlated with itself (Griffith, 2009).

This concept is inspired by Tobler’s first law: “Everything is related to everything else, but near things are more related than distant things”. So, in this case, naturally the proportion of people aged 65 years and over of a territory is more affected by the surrounding territories than the distant territories. We often called the surrounding territories by “neighbors”. To measure the spatial autocorrelation, we use the Moran’s I statistic that has been widely known. Given the proportion of territory i is symbolized by y_i and the proportion of its neighbors (for example, in territory j) is symbolized by y_j , the Moran’s I index is written as follows:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij} (y_i - \bar{y})(y_j - \bar{y}) / \sum_{i=1}^n \sum_{j=1}^n c_{ij}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 / n} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 / n}}$$

Where n is the number of territories and c_{ij} is a spatial weights – in many sources, it is often called W_{ij} – whose value is 1 for neighbors and 0 otherwise. To simplify the computation of Moran’s I index, usually W_{ij} is standardized in row. In the other words, if we sum the row component of matrix W_{ij} , it will resulting 1 ($\sum_{j=1}^n w_{ij} = 1, i = 1, \dots, n$). The Moran’s I statistic ranges from -1 to 1. There are three possibilities of I value:

- a. If I less than 0, we can say that a negative spatial autocorrelation occurred. This indicates that the nearby or neighboring areas are unlike. For example, high proportion territories tend to be geographically neighbors of low proportion territories, low proportion territories tend to be geographically neighbors of high proportion territories.
- b. If I more than 0, a positive spatial autocorrelation occurred. This is the contrary of negative spatial autocorrelation: the neighboring areas are more alike.
- c. If I is 0 or close to 0, it indicates a random spatial pattern.

The magnitude of I depends on the choice of W matrix. (Kooijman, 1976) in (Fischer & Getis, 2010) and (Griffith, 2009) proposed to choose W in order to maximize Moran’s coefficient. (Anselin, 1988) defined W matrix as the formal expression of spatial dependence between observations. There are many W matrix based on the determination of the neighbors. In this study, we will present two types of W matrix: weight matrix based on the contiguity and weight matrix based on the distance

Further explanation about these W matrices as follows (Getis & Aldstadt, 2004)

- a. Weight matrix based on the contiguity

There are 2 types of contiguity: rook and queen. Rook contiguity is defined as the four neighbors of each cell in the cardinal directions are considered as neighbors, so we give the value 1. And the other cells are not considered as neighbors. While queen contiguity is defined as the eight neighbors of each cell in all directions are given the value 1, and 0 for the other cells.

b. Weight matrix based on radial distance

If we considered distance as an important criterion of spatial influence, we can form a W matrix based on a certain distance. Suppose d_{ij} is the distance between the centroid of territory i and j, d is a determined distance, the W matrix is as follows:

$$w_{ij} = \begin{cases} 1, & \text{if } 0 \leq d_{ij} \leq d \\ 0, & d_{ij} > d \end{cases}$$

4. Markov Chain: Estimation of Transition Matrix

Most of study in probability relies on the assumption of independence between observations, which means the earlier events do not affect the next event. We take a coin toss as an example. We can say for sure, the side which will appear on the next toss is not affected by the previous toss. But there are some events in nature which the event today is probably influenced by previous day, for example the weather conditions. Modern probability theory studies chance of process for which the knowledge of earlier events affects predictions for future events.

In 1907, A. A. Markov introduced a stochastic process called Markov chain. Stochastic process is a mathematical model which evolves over a period of time in a probabilistic manner. Markov chain can be defined as: the outcome of an experiment (or event) depends only on the outcome of the previous experiments (or event). The process $\mathbf{X} = (x_0, x_1, \dots)$ is a Markov chain if every n, and any sequence of states $\{s_0, s_1, \dots, s_n\}$,

$$\begin{aligned} P_{\mu}(x_0 = s_0, x_1 = s_1, x_2 = s_2, \dots, x_n = s_n) \\ = \mu(s_0)P_{s_0}(x_1 = s_1)P_{x_1}(x_2 = s_2) \dots P_{x_{n-1}}(x_n = s_n) \end{aligned}$$

or equivalently, in terms of the conditional probabilities of the process \mathbf{X} ,

$$P_{\mu}(x_{n+1} = s_{n+1} | x_n = s_n, \dots, x_0 = s_0) = P(s_n, s_{n+1})$$

The probability μ is called the initial distribution of the chain. The process \mathbf{X} is a time-homogeneous Markov chain if the probabilities $P_{s_j}(x_1 = s_{j+1})$ depend only on the values of s_j, s_{j+1} and are independent of the time points j (Meyn & Tweedie, 2009).

The process starts in one of these states and moves successively from one state to another state. We called this movement as step and we can calculate the probability of the step. If the chain is in step s_i then it goes to state s_j in the next step with a probability p_{ij} , this probability does not depend upon which state the chain was in before the state s_i . We know the probability p_{ij} as transition probability. The chain can remain in the current state after one step with probability p_{ii} . The transition probabilities between the existing states can be stored in a matrix called transition matrix. The matrix $P = \{P(s, y), s, y \in X\}$ is called a Markov transition matrix if

$$P(s, y) \geq 0, \quad \sum_{z \in S} P(s, z) = 1, \quad s, y \in S$$

If the chain have done n step, the n-step transition matrix $P^n = \{P^n(s, y), s, y \in S\}$ by defining $P^0 = I$,

$$P^n(s, z) = \sum_{y \in S} P(s, y) P^{n-1}(y, z) \text{ (Meyn \& Tweedie, 2009)}$$

5. Linear Regression Model

The purpose of linear regression is to construct mathematical models which explain relationships that may exist between variables (Seber & Lee, 2003). In regression analysis, there are independent and dependent variables, that may be more than one variable. The general equation of linear regression can be written in the form

$$Y = X\beta + \varepsilon$$

where

- Y is an $(n \times 1)$ vector of observations,
- X is an $(n \times p)$ matrix known form,
- β is a $(p \times 1)$ vector of parameters,
- ε is an $(n \times 1)$ vector of errors or residuals

and where $E(\varepsilon) = \mathbf{0}$, $V(\varepsilon) = I\sigma^2$, so the elements of ε are uncorrelated. Parameters β are estimated by ordinary least square (OLS) method, which is conducted by minimize $\varepsilon \varepsilon$. The solution of the equation above written as (Draper & Smith, 1998)

$$b = (X'X)^{-1}X'Y$$

After obtaining the estimation of parameters, we must check the residuals. The residuals in linear regression must fulfill assumptions: identic, independent, and normally distributed in order to make tests that depend on the assumption of normality does not lead to the wrong conclusion.

6. Linear Mixed Model

West, et al in their book “Linear Mixed Models A practical Guide Using Statistical Software” give an easy-understand definition about linear mixed model. They explain linear mixed models (LMMs) as statistical models for continuous variables with normally distributed residuals, but may not be independent or have constance variance (West, Welch, & Galecki, 2007). The data sets may be appropriate analyzed by LMMs are (Cnann, Laird, & Slasor, 1997):

- a. Clustered (or nested) data, for example, students in a classroom, or experimental design with random blocks.
- b. Longitudinal or repeated-measures data, which the subjects are measured repeatedly. We can say that these repeated data are nested within subjects.
- c. Multivariate and correlated data.
- d. Growth and dose-response curve data

Linear mixed model has two types of independent variables, which are fixed variable and random variable. Fixed-effect parameters explain the relationships of the explanatory variables (independent variables) to the response variable (dependent variables) for an entire population, while random-effect parameters related to the subjects (or territories in our study) within a population. This allows us to resolve the non-independence by putting a different origin for each territory. So, LMM has linear parameters and its independent variables contain a mix of fixed and random effects.

We can also think our longitudinal data as multilevel data sets. This concept is introduced in hierarchical linear modeling literature by Raudenbush and Bryk in 2002. The second level is the territories, whereas the proportion of people aged 65 years and over in each period as level 1. We can formulate our case in the following equation:

$$Y_{ti} = \beta_1 X_{ti}^{(1)} + \beta_2 X_{ti}^{(2)} + \beta_3 X_{ti}^{(3)} + \dots + \beta_p X_{ti}^{(p)} \quad \} \text{fixed}$$

$$+ u_{1i} Z_{ti}^{(1)} + \dots + u_{qi} Z_{ti}^{(q)} + \varepsilon_{ti} \quad \} \text{random}$$

Y_{ti} : the measure of the continuous response (independent) variable Y taken on time t in territory i

β_1, \dots, β_p are coefficients of fixed-effect variables.

$X^{(1)}, \dots, X^{(p)}$: dependent variables considered as fixed effects

u_1, \dots, u_q : coefficients of random-effect variables

$X^{(1)}, \dots, X^{(q)}$: dependent variables considered as random effects

Or we can reformulate the equation above in a matrix form as follows:

$$Y_i = X_i \beta + Z_i u_i + \varepsilon_i$$

$$u_i \sim V(\mathbf{0}, D)$$

$$\varepsilon_i \sim N(\mathbf{0}, R_i)$$

where

$$Y_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \\ \vdots \\ Y_{ti} \end{pmatrix} \quad X_i = \begin{pmatrix} X_{1i}^{(1)} & X_{1i}^{(2)} & \dots & X_{1i}^{(p)} \\ X_{2i}^{(1)} & X_{2i}^{(2)} & \dots & X_{2i}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ti}^{(1)} & X_{ti}^{(2)} & \dots & X_{ti}^{(p)} \end{pmatrix} \quad Z_i = \begin{pmatrix} Z_{1i}^{(1)} & Z_{1i}^{(2)} & \dots & Z_{1i}^{(p)} \\ Z_{2i}^{(1)} & Z_{2i}^{(2)} & \dots & Z_{2i}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{ti}^{(1)} & Z_{ti}^{(2)} & \dots & Z_{ti}^{(p)} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \quad u_i = \begin{pmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{qi} \end{pmatrix} \quad \varepsilon_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{ti} \end{pmatrix}$$

where

$$u_i \sim N \left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} Var(u_{1i}) & cov(u_{1i}, u_{2i}) & \dots & cov(u_{1i}, u_{qi}) \\ cov(u_{1i}, u_{2i}) & Var(u_{2i}) & \dots & cov(u_{2i}, u_{qi}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(u_{1i}, u_{qi}) & cov(u_{2i}, u_{qi}) & \dots & Var(u_{qi}) \end{pmatrix} \right)$$

$$\varepsilon_i \sim N \left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \text{Var}(\varepsilon_{1i}) & \text{cov}(\varepsilon_{1i}, \varepsilon_{2i}) & \dots & \text{cov}(\varepsilon_{1i}, \varepsilon_{ti}) \\ \text{cov}(\varepsilon_{1i}, \varepsilon_{2i}) & \text{Var}(\varepsilon_{2i}) & \dots & \text{cov}(\varepsilon_{2i}, \varepsilon_{ti}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\varepsilon_{1i}, \varepsilon_{ti}) & \text{cov}(\varepsilon_{2i}, \varepsilon_{ti}) & \dots & \text{Var}(\varepsilon_{ti}) \end{pmatrix} \right)$$

In a model with intercept, the first column of X matrix would be equal to 1 for all observations.

To estimate the parameter of linear mixed model, we use the methods that have widely known: maximum likelihood (ML) estimation and restricted or residual maximum likelihood (REML) estimation. ML estimation is a method used for estimating unknown parameters by optimizing a likelihood function of distributional assumptions. While, REML was introduced by Patterson and Thompson in 1971 is a method of estimating variance components in the context unbalanced incomplete block designs.

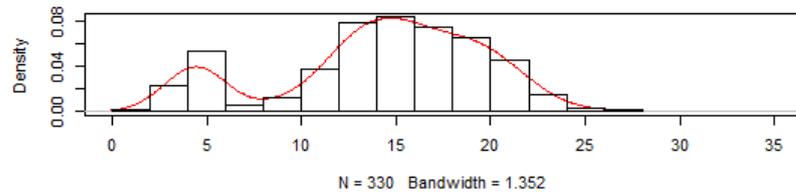
To make clearer our comprehension about REML, suppose we have linear model $y_i = \mu + \varepsilon_i$, with $\varepsilon_i \sim N(0, \sigma^2)$. The maximum likelihood estimator of μ is the sample mean and for σ^2 is the standard deviation of sample (if the sample number is small, usually it is divided by $n-1$ to get unbiased variance). There is a method that permits to adjust bias which allows us to use the log-likelihood. We can compute the likelihood of deviations of the mean or the residuals. These residuals have a singular multivariate normal distribution, and the maximum likelihood estimate of the variance turns out to be precisely the bias-adjusted estimate (Leeuw & Meijer, 2008).

II. Preprocessing Data

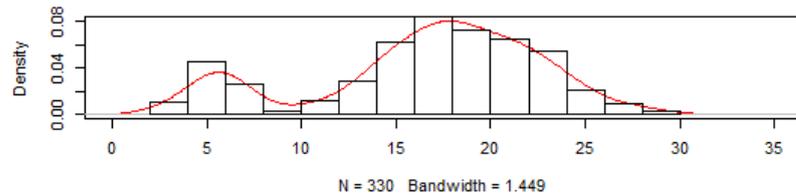
Before modeling the data with the Markov chain approach, we will see the distribution of data either by country, year, and overall. In 1990 and 2000, the histogram has 2 modes that can be considered as Gaussian form. The little Gaussian is the distribution of Morocco's that has a range between 0 – 8% with mode 5%, while the big Gaussian is the distribution of those 4 Mediterranean countries in Europe that has the proportion of people aged over 65 years between 8 – 26% in 1990. There is an enhancement of the proportion of people aged over 65 years in 2000. We can see that the curve moved to the right and the mode of the big Gaussian increased to approximately 17%.

In 2010, there are no data of people aged over 65 years in Morocco, so we have just a single Gaussian curve ranged between 10 – 30% with the mode 20%. Based on the 3 curves above, we can conclude that in each 10 years there is an increase in ageing of the population aged over 65 years in Mediterranean countries.

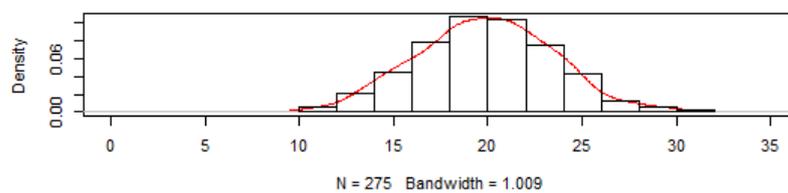
Distribution of People Aged Over 65 Years in Mediterranean Countries in 1990



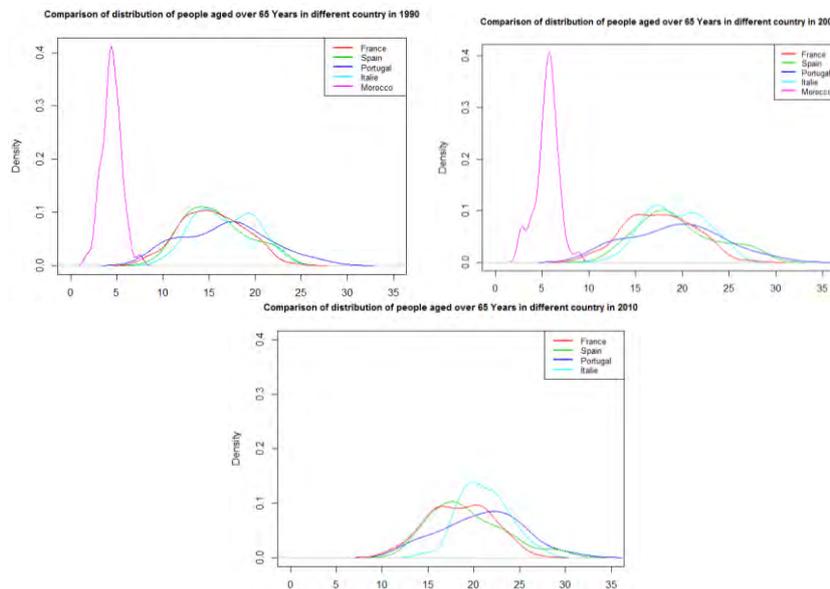
Distribution of People Aged Over 65 Years in Mediterranean Countries in 2000



Distribution of People Aged Over 65 Years in Mediterranean Countries in 2010

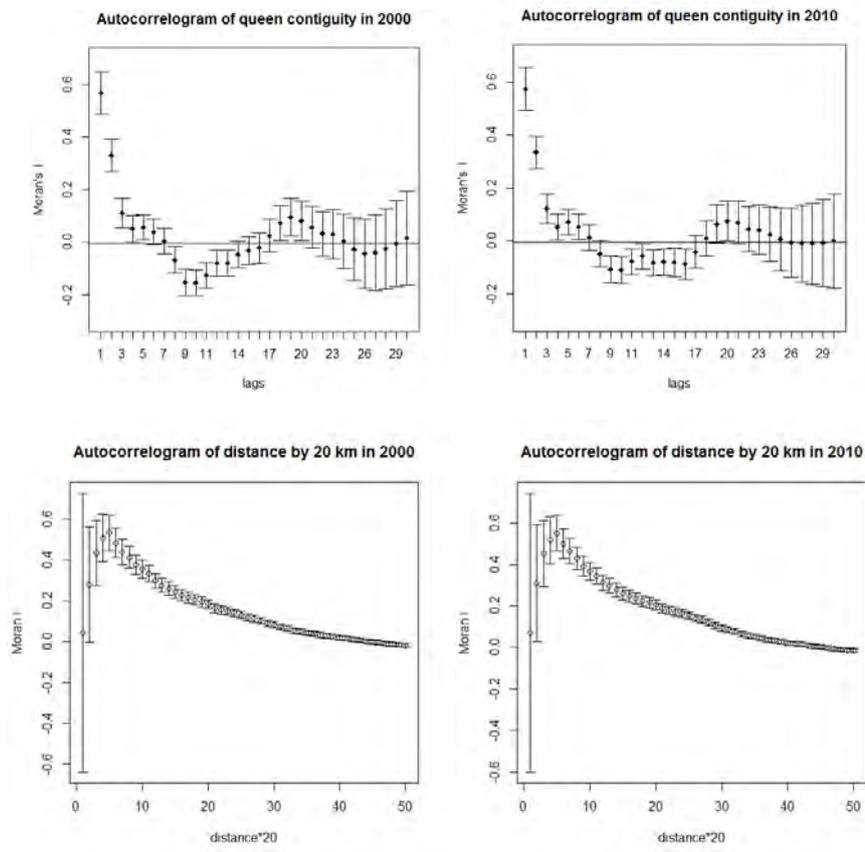


After seeing globally, it's shown the distribution for each country in each year and then we can compare the distribution of ageing in Mediterranean countries.

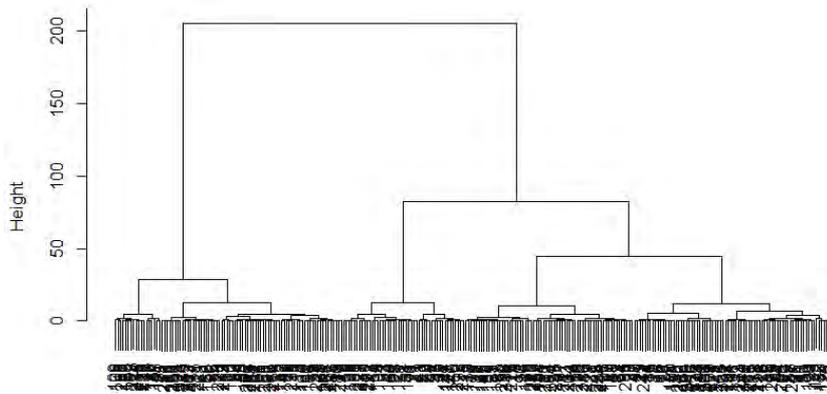


Morocco has the smallest proportion of people aged over 65 years than the other countries. So we can say that the majority of people in Morocco is young. In 4 other countries, the interval of the proportion of people aged over 65 years are almost the same, but Portugal has the longest one. Similar to the case for all countries, the proportion of people aged over 65 years for each country also increased in 20 years. In 2010, there is no data record for Morocco, so it is decided to not include Morocco in further analysis due to the reasons above. The total number of territories that will be analyzed is 275.

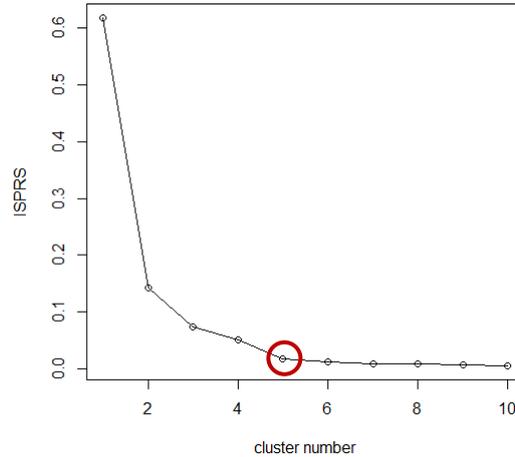
III. Spatial Weight Matrices based on Queen Contiguity and Distance



IV. Hierarchical Clustering



In defining the optimum number of clusters, we often determine it intuitively based on dendrogram which has obtained. There is no specific way how to determine it, but here, we will compute the semi partial R-squared (SPRSQ) which can be taken as a tool to justify the optimum number of clusters. SPRSQ measures the decrease in the proportion of variance accounted for caused by combining two clusters.



We choose the optimum number of classes based on the number of points that have the biggest difference to the point afterward. Based on the graph of semi-partial R-square, we determine the optimum number of classes is 5.

V. Programming in R

1. Reading data

```
library(maptools)
library(coda)
library(deldir)
library(sp)
library(spdep)
peps=read.dbf("E:/drive j3rd/Matkul/S2/SEMESTRE 4 STAGE/BD/Ensemble_regionpays.dbf")
peps1=peps[-173:-174,]
peps2=peps1[-276:-330,]
cpeps=readShapeSpatial("E:/drive j3rd/Matkul/S2/SEMESTRE 4
STAGE/BD/Ensemble_region.shp")
cpeps1=cpeps[cpeps$NOM!=c("Açores","Madeira"),]
cpeps2=cpeps1[peps1$PAYS!="Maroc",]
coords=coordinates(cpeps1)
coords2=coordinates(cpeps2)
```

2. Estimate the density plot using Kernel estimator

```
par (mfrow=c(3,1))
plot(density(peps2$P65_90),col="red",xlim=c(0,35),main="Distribution of People Aged Over 65
Years in Mediterranean Countries in 1990",add=T)
hist(peps2$P65_90,prob=T, xlab="Percentage",add=T)
plot(density(peps2$P65_00),col="red",xlim=c(0,35),main="Distribution of People Aged Over 65
Years in Mediterranean Countries in 2000",add=T)
hist(peps2$P65_00,prob=T, xlab="Percentage",add=T)
plot(density(peps2$P65_10),col="red",xlim=c(0,35),main="Distribution of People Aged Over 65
Years in Mediterranean Countries in 2010",add=T)
hist(peps2$P65_10,prob=T,breaks=14,xlab="Percentage",add=T)
```

```
#Territory grouping by its country#
```

```
peps2_F=peps2[1:96,]
peps2_E=peps2[97:144,]
peps2_P=peps2[145:172,]
peps2_I=peps2[173:275,]
peps2_M=peps2[276:330,]
```

```
#Density plot each country each year#
plot(density(peps2_M$P65_90),type="l",col=6,xlim=c(0,35),main="Comparison of distribution
of people aged over 65 Years in different country in 1990", lines = rep(c("Morocco",
"Spain", "Portugal", "Italie", "France")),cex.main=0.9)
lines(density(peps2_ESP65_90),col=3)
lines(density(peps2_PSP65_90),col=4)
lines(density(peps2_ISP65_90),col=5)
lines(density(peps2_FSP65_90),col=2)
legend("topright",c("France", "Spain", "Portugal", "Italie", "Morocco"),lty=c(1,1,1,1,1),lwd=c(2,2,2,
2,2),col=c(2,3,4,5,6),cex=0.8)
```

```
plot(density(peps2_M$P65_00),type="l",col=6,xlim=c(0,35),main="Comparison of distribution
of people aged over 65 Years in different country in 2000", lines = rep(c("Morocco",
"Spain", "Portugal", "Italie", "France")),cex.main=0.9)
lines(density(peps2_ESP65_00),col=3)
lines(density(peps2_PSP65_00),col=4)
lines(density(peps2_ISP65_00),col=5)
lines(density(peps2_FSP65_00),col=2)
legend("topright",c("France", "Spain", "Portugal", "Italie", "Morocco"),lty=c(1,1,1,1,1),lwd=c(2,2,2,
2,2),col=c(2,3,4,5,6),cex=0.8)
```

```
plot(density(peps2_ISP65_10),type="l",col=5,xlim=c(0,35),ylim=c(0,0.4),main="Comparison of
distribution of people aged over 65 Years in different country in 2010", lines =
rep(c("Spain", "Portugal", "Italie", "France")),cex.main=0.9)
lines(density(peps2_PSP65_10),col=4)
lines(density(peps2_ESP65_10),col=3)
lines(density(peps2_FSP65_10),col=2)
legend("topright",c("France", "Spain", "Portugal", "Italie"),lty=c(1,1,1,1),lwd=c(2,2,2,2),col=c(2,3,
4,5),cex=0.8)
```

3. Create trajectories of ageing by hierarchical clustering method

```
#evolution of state based on hierarchial clustering#
d=dist(scale(peps2[,3:5]),method="euclidean")
fit=hclust(d, method="ward.D")
plot(fit) # display dendrogram

#define number of class based on sprs#
P65_90=scale(peps2$P65_90)*sqrt(length(peps2$P65_90)/(length(peps2$P65_90)-1))
P65_00=scale(peps2$P65_00)*sqrt(length(peps2$P65_00)/(length(peps2$P65_00)-1))
P65_10=scale(peps2$P65_10)*sqrt(length(peps2$P65_10)/(length(peps2$P65_10)-1))
mat=cbind(P65_90,P65_00,P65_10)
d=dist(mat)
cl=hclust(d^2,method="ward")
dib=cl$height/(2*275)
sum(dib)
sprs=dib/sum(dib)
isprs=sprs[length(sprs):1]
plot(isprs,type="o", main="Semi-Partial R-sq",xlab="iteration number")

#there are 5 classes based on sprs#
group=cutree(fit, k=5) # cut tree into 5 clusters
# draw dendrogram with red borders around the 4 clusters
rect.hclust(fit, k=5, border="red")
```

```

cols[group==1]=2
cols[group==2]=3
cols[group==3]=4
cols[group==4]=5
cols[group==5]=6
plot(cpeps2,col=cols)
legend("bottomright",legend=c("Group 1","Group 2","Group 3","Group 4","Group
5"),fill=c(2,3,4,5,6),bty="n",y.intersp=0.8,cex=0.6)

#Plot of median of each trajectories group in each period#
peps2$group=group
pepsg=peps2[order(group), ]
temp=aggregate(pepsg[,c(3:5,8)],FUN=median,by=list(pepsg$group))
plot(c(1990,2000,2010),temp[1,2:4],type="o",col=2,ylim=c(10,30))
lines(c(1990,2000,2010),temp[2,2:4],type="o",col=3)
lines(c(1990,2000,2010),temp[3,2:4],type="o",col=4)
lines(c(1990,2000,2010),temp[4,2:4],type="o",col=5)
lines(c(1990,2000,2010),temp[5,2:4],type="o",col=6)
abline(h=14.722,lty=2,col=1)
abline(h=17.186,lty=2,col=1)
abline(h=19.26,lty=2,col=1)
abline(h=20.65,lty=2,col=1)
lines(c(1990,2000,2010),c(mean(peps2$P65_90),mean(peps2$P65_00),mean(peps2$P65_10)),typ
e="o",col=1)
legend("bottomright",legend=c("Group 1","Group 2","Group 3","Group 4","Group
5"),fill=c(2,3,4,5,6),bty="n",y.intersp=0.8,cex=0.6)

#Classification of territories into classes by quintil of overall data and the map#
#combining 3 periodes of time in one vector
P65=c(peps2$P65_90,peps2$P65_00,peps2$P65_10)
summary(P65)
q=quantile(P65, probs = c(0.2,0.4,0.6,0.8,1))
#BOXPLOT
boxplot(P65)
par(mfrow=c(1,3))
boxplot(peps2$P65_90)
boxplot(peps2$P65_00)
boxplot(peps2$P65_10)

#Classify P65 based on overall quintil#
class90=ifelse(peps2$P65_90 <= q[1],1,ifelse((peps2$P65_90 <= q[2] & peps2$P65_90 >
q[1]),2,ifelse ((peps2$P65_90 <= q[3] & peps2$P65_90 > q[2]),3,ifelse ((peps2$P65_10 <= q[4]
& peps2$P65_10 > q[3]),4,5))))
cols[class90==1]="blue4"
cols[class90==2]="deepskyblue1"
cols[class90==3]="white"
cols[class90==4]="pink"
cols[class90==5]="red"
plot(cpeps2,col=cols,main="Classification of territories in 1990")
legend("bottomright",legend=c("Young","Rather young","Moderate","Rather
old","Old"),fill=c("blue4","deepskyblue1","white","pink","red"),bty="n",y.intersp=0.8,cex=0.6)

```

```

class00=ifelse(peps2$P65_00 <= q[1],1,ifelse((peps2$P65_00 <= q[2] & peps2$P65_00 >
q[1]),2,ifelse ((peps2$P65_00 <= q[3] & peps2$P65_00 > q[2]),3,ifelse ((peps2$P65_10 <= q[4]
& peps2$P65_10 > q[3]),4,5))))
cols[class00==1]="blue4"
cols[class00==2]="deepskyblue1"
cols[class00==3]="white"
cols[class00==4]="pink"
cols[class00==5]="red"
plot(cpeps2,col=cols,main="Classification of territories in 2000")
legend("bottomright",legend=c("Young","Rather young","Moderate","Rather
old","Old"),fill=c("blue4","deepskyblue1","white","pink","red"),bty="n",y.intersp=0.8,cex=0.6)

```

```

class10=ifelse(peps2$P65_10 <= q[1],1,ifelse((peps2$P65_10 <= q[2] & peps2$P65_10 >
q[1]),2,ifelse ((peps2$P65_10 <= q[3] & peps2$P65_10 > q[2]),3,ifelse ((peps2$P65_10 <= q[4]
& peps2$P65_10 > q[3]),4,5))))
cols[class10==1]="blue4"
cols[class10==2]="deepskyblue1"
cols[class10==3]="white"
cols[class10==4]="pink"
cols[class10==5]="red"
plot(cpeps2,col=cols,main="Classification of territories in 2010")
legend("bottomright",legend=c("Young","Rather young","Moderate","Rather
old","Old"),fill=c("blue4","deepskyblue1","white","pink","red"),bty="n",y.intersp=0.8,cex=0.6)

```

###JOINTCOUNT TEST based on queen contiguity###

```

nbq=poly2nb(cpeps2,queen=T)
cclistw2=nb2listw(nbq,style='W',zero.policy=T)
##GROUP 1 vs OTHERS##
A=ifelse(groupn1==1,1,2)
joincount.test(as.factor(A),cclistw2,zero.policy=T,alternative="greater")
cols[A==1]="black"
cols[A==2]="white"
plot(cpeps2,col=cols)
##GROUP 2 vs OTHERS##
B=ifelse(groupn1==2,1,2)
joincount.test(as.factor(B),cclistw2,zero.policy=T,alternative="greater")
cols[B==1]="black"
cols[B==2]="white"
plot(cpeps2,col=cols)
##GROUP 3 vs OTHERS##
C=ifelse(groupn1==3,1,2)
joincount.test(as.factor(C),cclistw2,zero.policy=T,alternative="greater")
cols[C==1]="black"
cols[C==2]="white"
plot(cpeps2,col=cols)
##GROUP 4 vs OTHERS##
D=ifelse(groupn1==4,1,2)
joincount.test(as.factor(D),cclistw2,zero.policy=T,alternative="greater")
cols[D==1]="black"
cols[D==2]="white"
plot(cpeps2,col=cols)
##GROUP 5 vs OTHERS##
E=ifelse(groupn1==5,1,2)
joincount.test(as.factor(E),wdd,zero.policy=T,alternative="greater")

```

```
cols[E==1]="black"
cols[E==2]="white"
plot(cpeps2,col=cols)
```

4. Moran's plot based on different spatial weight matrices

```
#Spatial weight matrix based on queen contiguity#
nbq=poly2nb(cpeps2,queen=T)
plot(sp.correlogram(nbq, peps2$P65_90, order = 30, method = "I",randomisation=F, style =
"W",zero.policy=T),main="Autocorrelogram of queen contiguity in 1990",ylim=c(-0.3,0.7))
plot(sp.correlogram(nbq, peps2$P65_00, order = 30, method = "I",randomisation=F, style =
"W",zero.policy=T),main="Autocorrelogram of queen contiguity in 2000",ylim=c(-0.3,0.7))
plot(sp.correlogram(nbq, peps2$P65_10, order = 30, method = "I",randomisation=F, style =
"W",zero.policy=T),main="Autocorrelogram of queen contiguity in 2010",ylim=c(-0.3,0.7))

#Spatial weight matrix based on radial distance#
#For data in 1990#
di=function(d)
{
dd=dnearneigh(coords2,0,d,longlat=T)
we=nb2listw(dd,style="W", zero.policy=T)
mor=moran.test(peps2$P65_90,we, zero.policy=T)
mor$estimate
}

library(plotrix)
mo=numeric(length(50))
lo=numeric(length(50))
up=numeric(length(50))
for(i in 1:50)
{
mo[i]=di(20*i)[1]
up[i]=mo[i]+1.96*sqrt(di(20*i)[3])
lo[i]=mo[i]-1.96*sqrt(di(20*i)[3])
}
plotCI(mo,ui=up,li=lo, xlab="distance*20",ylab="Moran I",main="Autocorrelogram of distance
by 20 km in 1990")

#For data in 2000#
di=function(d)
{
dd=dnearneigh(coords2,0,d,longlat=T)
we=nb2listw(dd,style="W", zero.policy=T)
mor=moran.test(peps2$P65_00,we, zero.policy=T)
mor$estimate
}
library(plotrix)
mo=numeric(length(50))
lo=numeric(length(50))
up=numeric(length(50))
for(i in 1:50)
{
mo[i]=di(20*i)[1]
up[i]=mo[i]+1.96*sqrt(di(20*i)[3])
lo[i]=mo[i]-1.96*sqrt(di(20*i)[3])
}
```

```

}
plotCI(mo,ui=up,li=lo, xlab="distance*20",ylab="Moran I",main="Autocorrelogram of distance
by 20 km in 2000")

#For data in 2010#
di=function(d)
{
dd=dnearneigh(coords2,0,d,longlat=T)
we=nb2listw(dd,style="W", zero.policy=T)
mor=moran.test(peps2$P65_10,we, zero.policy=T)
mor$estimate
}
library(plotrix)
mo=numeric(length(50))
lo=numeric(length(50))
up=numeric(length(50))
for(i in 1:50)
{
mo[i]=di(20*i)[1]
up[i]=mo[i]+1.96*sqrt(di(20*i)[3])
lo[i]=mo[i]-1.96*sqrt(di(20*i)[3])
}
plotCI(mo,ui=up,li=lo, xlab="distance*20",ylab="Moran I",main="Autocorrelogram of distance
by 20 km in 2010")

```

5. Estimation of transition matrix

```

#####NORMALISATION 1#####
#@#@#@#@DATA REDUCED BY GLOBAL MEDIAN#@#@#@#@#
n1a90=peps2$P65_90-median(peps2$P65_90)
n1a00=peps2$P65_00-median(peps2$P65_00)
n1a10=peps2$P65_10-median(peps2$P65_10)
median(peps2$P65_90)
median(peps2$P65_00)
median(peps2$P65_10)
plot(c(15.88,18.47,19.83),type="o",col=2,ylab="Median",xlab="",main="Median of all territories
each year")

par(mfrow=c(1,4))
boxplot(n1a90,xlab="1990",main="Normalization 1 by median")
boxplot(n1a00,xlab="2000")
boxplot(n1a10,xlab="2010")
boxplot(c(n1a90,n1a00,n1a10),xlab="Total")

#kurtosis, skewness, shapiro-wilk test#
kurtosis(n1a90)
kurtosis(n1a00)
kurtosis(n1a10)
skewness(n1a90)
skewness(n1a00)
skewness(n1a10)
shapiro.test(n1a90)
shapiro.test(n1a00)
shapiro.test(n1a10)

```

```

####Density Plot####
plot(density(n1a10),type="l",col=6,cex.main=0.9,main="Comparison of distribution of people
aged over 65 Years in each year after normalization 1 by median")
lines(density(n1a00),col=3)
lines(density(n1a90),col=4)
legend("topright",c("1990","2000","2010"),lty=c(1,1,1),lwd=c(2,2,2),col=c(4,3,6),cex=0.8)
abline(v=-2.974,lty=2,col=2)
abline(v=-0.94,lty=2,col=2)
abline(v=1.092,lty=2,col=2)
abline(v=3.386,lty=2,col=2)

####QQ-Plot####
par(mfrow=c(2,2))
qqnorm(scale(n1a90),main="1990")
abline(0,1,col=2)
qqnorm(scale(n1a00),main="2000")
abline(0,1,col=2)
qqnorm(scale(n1a10),main="2010")
abline(0,1,col=2)

#####CLASS BASED ON QUANTILE ALL YEAR#####
quantile(c(n1a90,n1a00,n1a10), probs = c(0.2,0.4,0.6,0.8,1))
classn1a90=ifelse(n1a90 <= -2.974,1,ifelse((n1a90 <= -0.94 & n1a90 > -2.974),2,ifelse ((n1a90
<= 1.092 & n1a90 > -0.94),3,ifelse ((n1a90 <= 3.386 & n1a90 > 1.092),4,5))))
classn1a00=ifelse(n1a00 <= -2.974,1,ifelse((n1a00 <= -0.94 & n1a00 > -2.974),2,ifelse ((n1a00
<= 1.092 & n1a00 > -0.94),3,ifelse ((n1a00 <= 3.386 & n1a00 > 1.092),4,5))))
classn1a10=ifelse(n1a10 <= -2.974,1,ifelse((n1a10 <= -0.94 & n1a10 > -2.974),2,ifelse ((n1a10
<= 1.092 & n1a10 > -0.94),3,ifelse ((n1a10 <= 3.386 & n1a10 > 1.092),4,5))))

print(table(classn1a90,classn1a00))
print(table(classn1a00,classn1a10))
table(classn1a90)/275
table(classn1a00)/275
table(classn1a10)/275

statdistr(matrix(c(0.87,0.14,0.05,0.01,0,0.12,0.68,0.09,0.03,0,0.01,0.18,0.73,0.15,0,0,0,0.12,0.7,0.
18,0,0,0.01,0.11,0.82),ncol=5))

#####NORMALISATION 2#####
peps2_F=peps2[1:96,]
peps2_E=peps2[97:144,]
peps2_P=peps2[145:172,]
peps2_I=peps2[173:275,]
#####DATA REDUCED BY MEDIAN OF EACH COUNTRY#####
n2a90=c(peps2_F$P65_90-median(peps2_F$P65_90),peps2_E$P65_90-
median(peps2_E$P65_90),peps2_P$P65_90-median(peps2_P$P65_90),peps2_I$P65_90-
median(peps2_I$P65_90))
n2a00=c(peps2_F$P65_00-median(peps2_F$P65_00),peps2_E$P65_00-
median(peps2_E$P65_00),peps2_P$P65_00-median(peps2_P$P65_00),peps2_I$P65_00-
median(peps2_I$P65_00))
n2a10=c(peps2_F$P65_10-median(peps2_F$P65_10),peps2_E$P65_10-
median(peps2_E$P65_10),peps2_P$P65_10-median(peps2_P$P65_10),peps2_I$P65_10-
median(peps2_I$P65_10))

```

```

plot(c(median(peps2_I$P65_90),median(peps2_I$P65_00),median(peps2_I$P65_10)),type="o",c
ol=5,ylab="Median",xlab="",ylim=c(14,22),main="Median of each country in each year")
lines(c(median(peps2_E$P65_90),median(peps2_E$P65_00),median(peps2_E$P65_10)),type="o"
,col=3)
lines(c(median(peps2_P$P65_90),median(peps2_P$P65_00),median(peps2_P$P65_10)),type="o"
,col=4)
lines(c(median(peps2_F$P65_90),median(peps2_F$P65_00),median(peps2_F$P65_10)),type="o"
,col=2)
legend("bottomright",c("France","Spain","Portugal","Italy"),lty=c(1,1,1,1),lwd=c(2,2,2,2),col=c(2
,3,4,5),cex=0.8)

```

```

par(mfrow=c(1,4))
boxplot(n2a90,xlab="1990",main="Normalization 2 by median")
boxplot(n2a00,xlab="2000")
boxplot(n2a10,xlab="2010")
boxplot(c(n2a90,n2a00,n2a10),xlab="Total")

```

```

#kurtosis, skewness, shapiro-wilk test#

```

```

kurtosis(n2a90)
kurtosis(n2a00)
kurtosis(n2a10)
skewness(n2a90)
skewness(n2a00)
skewness(n2a10)
shapiro.test(n2a90)
shapiro.test(n2a00)
shapiro.test(n2a10)

```

```

####Density Plot####

```

```

plot(density(n2a10),type="l",col=6,cex.main=0.9,main="Comparison of distribution of people
aged over 65 Years in each year after normalization 2 by median")
lines(density(n2a00),col=3)
lines(density(n2a90),col=4)
legend("topright",c("1990","2000","2010"),lty=c(1,1,1),lwd=c(2,2,2),col=c(4,3,6),cex=0.8)
abline(v=-2.638,lty=2,col=2)
abline(v=-0.7,lty=2,col=2)
abline(v=1.252,lty=2,col=2)
abline(v=3.372,lty=2,col=2)

```

```

####QQ-Plot####

```

```

qqnorm(scale(n2a90))
abline(0,1,col=2)
qqnorm(scale(n2a00))
abline(0,1,col=2)
qqnorm(scale(n2a10))
abline(0,1,col=2)

```

```

#####CLASS BASED ON QUANTILE ALL YEAR#####

```

```

quantile(c(n2a90,n2a00,n2a10), probs = c(0.2,0.4,0.6,0.8,1))

```

```

classn2a90=ifelse(n2a90 <= -2.638,1,ifelse((n2a90 <= -0.7 & n2a90 > -2.638),2,ifelse ((n2a90 <=
1.252 & n2a90 > -0.7),3,ifelse ((n2a90 <= 3.372 & n2a90 > 1.252),4,5)))

```

```

classn2a00=ifelse(n2a00 <= -2.638,1,ifelse((n2a00 <= -0.7 & n2a00 > -2.638),2,ifelse ((n2a00 <=
1.252 & n2a00 > -0.7),3,ifelse ((n2a00 <= 3.372 & n2a00 > 1.252),4,5))))
classn2a10=ifelse(n2a10 <= -2.638,1,ifelse((n2a10 <= -0.7 & n2a10 > -2.638),2,ifelse ((n2a10 <=
1.252 & n2a10 > -0.7),3,ifelse ((n2a10 <= 3.372 & n2a10 > 1.252),4,5))))

print(table(classn2a90,classn2a00))
print(table(classn2a00,classn2a10))
statdistr(matrix(c(0.86,0.08,0.01,0.01,0,0.14,0.72,0.11,0.01,0,0,0.19,0.66,0.17,0,0,0,0.21,0.7,0.16,
0,0,0,0.1,0.84),ncol=5))
table(classn2a90)/275
table(classn2a00)/275
table(classn2a10)/275

#####NORMALISATION 3#####
####MEDIAN OF NEIGHBORS####
dd=dnearneigh(coords2,0,100,longlat=T,row.names=row.names(coords2))
wdd=nb2listw(dd,style="W", zero.policy=T)
mn90=numeric(length(wdd$neighbours))
for(i in 1:length(wdd$neighbours))
{
  mn90[i]=median(peps2$P65_90[wdd$neighbours[[i]])]
}
mn00=numeric(length(wdd$neighbours))
for(i in 1:length(wdd$neighbours))
{
  mn00[i]=median(peps2$P65_00[wdd$neighbours[[i]])]
}
mn10=numeric(length(wdd$neighbours))
for(i in 1:length(wdd$neighbours))
{
  mn10[i]=median(peps2$P65_10[wdd$neighbours[[i]])]
}

mn90[is.na(mn90)]=0
mn00[is.na(mn00)]=0
mn10[is.na(mn10)]=0

#DATA REDUCED BY MEDIAN OF ITS NEIGHBORS
n3a90=peps2$P65_90-mn90
n3a00=peps2$P65_00-mn00
n3a10=peps2$P65_10-mn10
plot(density(mn10),col=6)
lines (density(mn00),col=3)
lines (density(mn90),col=4)
legend("topright",c("median in 1990","median in 2000","median in
2010"),lty=c(1,1,1),lwd=c(2,2,2),col=c(4,3,6),cex=0.8)

par(mfrow=c(1,4))
boxplot(n3a90,xlab="1990",main="Normalization 3 by median")
boxplot(n3a00,xlab="2000")
boxplot(n3a10,xlab="2010")
boxplot(c(n3a90,n3a00,n3a10),xlab="Total")

```

```

#kurtosis, skewness, shapiro-wilk test#
kurtosis(n3a90)
kurtosis(n3a00)
kurtosis(n3a10)
skewness(n3a90)
skewness(n3a00)
skewness(n3a10)
shapiro.test(n3a90)
shapiro.test(n3a00)
shapiro.test(n3a10)

####Density Plot####
plot(density(n3a10),type="l",col=6,cex.main=0.9,main="Comparison of distribution of people
aged over 65 Years in each year after normalization 3 by median")
lines(density(n3a00),col=3)
lines(density(n3a90),col=4)
legend("topright",c("1990", "2000", "2010"),lty=c(1,1,1),lwd=c(2,2,2),col=c(4,3,6),cex=0.8)
abline(v=-2.056,lty=2,col=2)
abline(v=-0.546,lty=2,col=2)
abline(v=0.804,lty=2,col=2)
abline(v=2.286,lty=2,col=2)
peps2[which(n3a10>7),]

####QQ-Plot####
qqnorm(scale(n3a90))
abline(0,1,col=2)
qqnorm(scale(n3a00))
abline(0,1,col=2)
qqnorm(scale(n3a10))
abline(0,1,col=2)

#####CLASS BASED ON QUANTILE ALL YEAR#####
quantile(c(n3a90,n3a00,n3a10), probs = c(0.2,0.4,0.6,0.8,1))

classn3a90=ifelse(n3a90 <= -2.056,1,ifelse((n3a90 <= -0.546 & n3a90 > -2.056),2,ifelse ((n3a90
<= 0.804 & n3a90 > -0.546),3,ifelse ((n3a90 <= 2.286 & n3a90 > 0.804),4,5))))
classn3a00=ifelse(n3a00 <= -2.056,1,ifelse((n3a00 <= -0.546 & n3a00 > -2.056),2,ifelse ((n3a00
<= 0.804 & n3a00 > -0.546),3,ifelse ((n3a00 <= 2.286 & n3a00 > 0.804),4,5))))
classn3a10=ifelse(n3a10 <= -2.056,1,ifelse((n3a10 <= -0.546 & n3a10 > -2.056),2,ifelse ((n3a10
<= 0.804 & n3a10 > -0.546),3,ifelse ((n3a10 <= 2.286 & n3a10 > 0.804),4,5))))

print(table(classn3a90,classn3a00))
print(table(classn3a00,classn3a10))
statdistr(matrix(c(0.79,0.21,0,0,0,0.2,0.62,0.14,0.02,0,0.02,0.16,0.64,0.19,0.01,0,0.01,0.02,0.66,0.
17,0,0,0.03,0.13,0.82),ncol=5))
table(classn3a90)/275
table(classn3a00)/275
table(classn3a10)/275

```

6. Linear Regression Model

```

#####regression with queen contiguity#####
nbq=poly2nb(cpeps2,queen=T)
wenb=nb2listw(nbq,style="W", zero.policy=T)
wynb90=lag.listw(wenb,peps2$P65_90,zero.policy=T)

```

```

wynb00=lag.listw(wenb,peps2$P65_00,zero.policy=T)
wynb10=lag.listw(wenb,peps2$P65_10,zero.policy=T)
#Create a table contained variables that will be modeled#
peps_y=c(peps2$P65_00,peps2$P65_10)
peps_x=c(peps2$P65_90,peps2$P65_00)
time=c(rep(0,275),rep(1,275))
we2=c(wynb00,wynb10)
we1=c(wynb90,wynb00)
inter1=time*peps_x
inter2=time*we2
tabf=cbind(peps_y,peps_x,time,we1,we2,inter1,inter2)
tabff=as.data.frame(tabf)
tabff$Nom=peps2$NOM
tab1=tabff[order(tabff$Nom),]
mod1=lm(peps_y~time+peps_x)
summary(mod1)
mod2=lm(peps_y~time+peps_x+we2)
summary(mod2)
mod3=lm(peps_y~time+peps_x+we2+inter1)
summary(mod3)
mod4=lm(peps_y~time+peps_x+we2+inter2)
summary(mod4)
mod5=lm(peps_y~time+peps_x+we2+inter1+inter2)
summary(mod5)
mod6=lm(peps_y~peps_x+we2+inter1+inter2)
summary(mod6)
mod7=lm(peps_y~peps_x+we2+inter1)
summary(mod7)
mod8=lm(peps_y~peps_x+we2+inter2)
summary(mod8)
mod6$res
boxplot(mod6$res~time)
par(mfrow=c(2,2))
plot(mod6)

#Residual's map based on model 6#
tabff$res=mod6$res
a=mean(tabff$res)-2.5*sd(tabff$res)
b=mean(tabff$res)-1.5*sd(tabff$res)
c=mean(tabff$res)+1.5*sd(tabff$res)
d=mean(tabff$res)+2.5*sd(tabff$res)
##For TIME=0##
tab0=subset(tabff,time==0)
class0=ifelse(tab0$res <= a,1,ifelse((tab0$res <= b & tab0$res > a),2,ifelse ((tab0$res <= c &
tab0$res > b),3,ifelse ((tab0$res <= d & tab0$res > c),4,5)))
cols[class0==1]="blue"
cols[class0==2]="deepskyblue1"
cols[class0==3]="white"
cols[class0==4]="pink"
cols[class0==5]="red"
plot(cpeps2,col=cols)
legend("bottomright",legend=c("res < -2.56","-2.56 < res <-1.54","-1.54 < res < 1.54","1.54 < res
<2.56","res >
2.56"),fill=c("blue","deepskyblue1","white","pink","red"),bty="n",y.intersp=0.8,cex=0.6)

```

```

##For TIME=1##
tab1=subset(tabff,time==1)
class1=ifelse(tab1$res <= a,1,ifelse((tab1$res <= b & tab1$res > a),2,ifelse ((tab1$res <= c &
tab1$res > b),3,ifelse ((tab1$res <= d & tab1$res > c),4,5))))
cols[class1==1]="blue"
cols[class1==2]="deepskyblue1"
cols[class1==3]="white"
cols[class1==4]="pink"
cols[class1==5]="red"
plot(cpeps2,col=cols)
legend("bottomright",legend=c("res < -2.56","-2.56 < res <-1.54","-1.54 < res < 1.54","1.54 < res
<2.56","res >
2.56"),fill=c("blue","deepskyblue1","white","pink","red"),bty="n",y.intersp=0.8,cex=0.6)

```

```

#Comparing Model 5 and Model 6 using ANOVA#
anova(mod4,mod6)

```

7. Linear Mixed Model

```

Library(nlme)
mo1=lme(peps_y~time+peps_x,random=~1|Nom,data=tab1,method="ML")
summary(mo1)
mo2=lme(peps_y~time+peps_x+we2,random=~1|Nom,data=tab1,method="ML")
summary(mo2)
mo3=lme(peps_y~time+peps_x+we2+inter1,random=~1|Nom,data=tab1,method="ML")
summary(mo3)
mo4=lme(peps_y~time+peps_x+we2+inter2,random=~1|Nom,data=tab1,method="ML")
summary(mo4)
mo5=lme(peps_y~time+peps_x+we2+inter1+inter2,random=~1|Nom,data=tab1,method="ML")
summary(mod5)
mo6=lme(peps_y~peps_x+we2+inter1+inter2,random=~1|Nom,data=tab1,method="ML")
summary(mo6)
mo7=lme(peps_y~peps_x+we2+inter1,random=~1|Nom,data=tab1,method="ML")
summary(mo7)
mo8=lme(peps_y~peps_x+we2+inter2,random=~1|Nom,data=tab1,method="ML")
summary(mo8)

```

```

#Diagnostic residual plot#
tab1$fit=fitted(mo6)
tab1$res=tab1$peps_y-tab1$fit
sd(tab1$res)
tab1$stdres=tab1$res/sd(tab1$res)
par(mfrow=c(1,2))
plot(tab1$fit,tab1$res,main="Residuals vs Fitted",ylab="Residuals",xlab="Fitted values")
qqPlot(tab1$res, main="Normal Q-Q",ylab="Standardized Residuals")

```

```

#Residual map base on model 6#
a=mean(tab1$res)-2.5*sd(tab1$res)
b=mean(tab1$res)-1.5*sd(tab1$res)
c=mean(tab1$res)+1.5*sd(tab1$res)
d=mean(tab1$res)+2.5*sd(tab1$res)
#For time=0#
tab0=subset(tab1,time==0)

```

```

class0=ifelse(tab0$res <= a,1,ifelse((tab0$res <= b & tab0$res > a),2,ifelse ((tab0$res <= c &
tab0$res > b),3,ifelse ((tab0$res <= d & tab0$res > c),4,5))))
cols[class0==1]="blue"
cols[class0==2]="deepskyblue1"
cols[class0==3]="white"
cols[class0==4]="pink"
cols[class0==5]="red"
plot(cpeps2,col=cols)
legend("bottomright",legend=c("res < -1.76","-1.76 < res <-1.06","-1.06 < res < 1.06","1.06 < res
<1.76","res >
1.76"),fill=c("blue","deepskyblue1","white","pink","red"),bty="n",y.intersp=0.8,cex=0.6)
tab0$class0=class0
tab0$Nom[which(tab0$class0=="5")]

#For time=1#
tab_b=subset(tab1,time==1)
class1=ifelse(tab_b$res <= a,1,ifelse((tab_b$res <= b & tab_b$res > a),2,ifelse ((tab_b$res <= c &
tab_b$res > b),3,ifelse ((tab_b$res <= d & tab_b$res > c),4,5))))
cols[class1==1]="blue"
cols[class1==2]="deepskyblue1"
cols[class1==3]="white"
cols[class1==4]="pink"
cols[class1==5]="red"
plot(cpeps2,col=cols)
legend("bottomright",legend=c("res < -1.76","-1.76 < res <-1.06","-1.06 < res < 1.06","1.06 < res
<1.76","res >
1.76"),fill=c("blue","deepskyblue1","white","pink","red"),bty="n",y.intersp=0.8,cex=0.6)
tab_b$class1=class1
tab_b$Nom[which(tab_b$class1=="5")]
#Comparing Model 5 and Model 6 using ANOVA#
anova(mo4,mo6)

```