Masters Program in **Geospatial Technologies**

ANALYSIS AND SIMULATION OF SOCIAL UNREST IN EUROPE

Towards understanding social unrest in Europe

Elisabet Adeva Romero

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

UNIVERSIDADE NOVA ISEGI
UNIVERSITAT JAUME•I
Westfälische Wilhelms-Universität Münster

ANALYSIS AND SIMULATION OF SOCIAL UNREST IN EUROPE

Towards understanding social unrest in Europe

Supervisor:

Professor Ana Cristina Costa

Instituto Superior de Estatística e Gestão de Informação (ISEGI)

Universidade Nova de Lisboa,

Lisbon, Portugal.

Co-supervirors:

Prof. Dr. Edzer J. Pebesma.

Institute for Geoinformatics (iFGI),

Westfälische Wilhelms-Universität,

Muenster, Germany.

&

Prof. Dr. Jorge Mateu M

Department of Mathematics,

Universitat Jaume I (UJI)

Castellon, Spain.

February 2014

Dedication

To my mother

Acknowledgments

I am very grateful to my supervisor, Profª Doutora Professor Ana Cristina Marinho da Costa and co-supervisors Professor Edzer Pebesma and Jorge Mateu for taking time off a busy schedule to supervise me at every stage of this project.

I am very grateful to the European Commission for being part of the Erasmus Mundus scholarship program for trusting me, and foster my career and studies. Thanks to the three partner universities - New University of Lisbon Portugal, Institute for Geoinformatics Muenster and Universitat Jaume I, and Castellon Spain for this great experience.

I would like to express my very great appreciation to Kalev Leetaru and Philip A. Schrodt, for their research on GDELT: Global Data on Events, Location and Tone, which I have, enjoy working with and knowing deeper in this study, thus I ought to congratulate them as I believe it is a countless innovation, I a further development.

I am particularly grateful for the assistance given by Mark Padgham regarding to R software.

I sincerely thank all my colleagues that without knowing me have believed in me for their gentleness in my difficult times. I appreciate the words of support of my family and friends around the world, in special to my best friends, Olga Blanca and Elba Guillaumes for making me smile everyday even the distance.

My special thanks are extended to Neel for making me believe again, in the impossible.

And to the many others who spend with me their time and gave me free suggestion and comments, you all have improve indirectly this work, to those whom I have not mention, I mean no ungratefulness. Thank you to all.

Declaration of Originality

I declare that the submitted work is entirely my own and not belongs to any other person. All references, including citation of published and unpublished sources have been appropriately acknowledged in the work. I further declare that the work has not been submitted to any institution for assessment for any other purpose.

Lisbon,
February 2014

Elisabet Adeva

ABSTRACT

ANALYSIS AND SIMULATION OF SOCIAL UNREST IN EUROPE

Towards understanding social unrest in Europe

Protest of Europe from 2000 to 2010 where analyzed to foster understanding of the distribution and behaviour of those during the time mentioned.

The main object of this study was to discover the relation with variables available in Eurostat and discover the pattern of those protests around Europe. Ordinary Least Squared Method and Spatial point pattern analysis methods were implemented in the R software environment for this purpose. Overall, the variables selected do not define the protests behaviour but some of them are related and increase with the protests. Protest tend to increase occur mostly when other protest have happened. Protest tend to create hotspots within Europe, their location are mostly in urban areas and close to the borders with other European countries. Resulting models discovered that protest/event distribution does not imitate to a Poisson process, and that their behaviour can be better described by interaction between protests.

The final model chosen by best distance was analyzed for Europe and later for Germany, France, United Kingdom and Spain due to an unequal distribution of protests in Europe; a further temporal analysis was computed only for Spain. The set of models computed, showed that protest location are scattered within the European megalopolis, and reveals characteristics the attraction to some capitals so that clustered or hot spots pattern are observed.

This analysis is one of the first analyses prepared by the recently launched, Global Database of Events, Language, and Tone (GDELT), a big free online data base of over 250m events and 300 categories from riots and protests to diplomatic exchanges and peace appeals codified from world news sources. From this analysis we recommend that further models could be applied to compute dissemination and contagions over time within borders.

KEYWORDS

Protest

GDELT Global Database of Events, Language, and Tone

R

Spatial Autoregressive model

Spatial Point Pattern Analysis

Spatial distribution pattern

spatstat

ACRONYMS

- AI - Area Interaction
- CSR - Complete Spatial Randomness
- EUROSTAT - Statistical Office of the European Union
- GDP - gross domestic product
- GED - Geo-referenced Event Dataset
- GDELT - Global Database of Events, Language, and Tone
- ICP - Inhomogeneous Cluster Process
- IPP - Inhomogeneous Poisson Process
- OLS - Ordinary Least Square
- PPP - Point Pattern Process
- UCDP - Uppsala Conflict Data Program
- WEIS - World Event/Interaction Survey
- ICEWS - Integrated Conflict Early Warning System

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

CHAPTER 1: BACKGROUND

Now-a day, protests have been established as most well recognized medium for expressing civil views in conflicts as a non –violence manner. They are used for civilians as a practice of exercising democracy, change in policy leads to convenient growth, deterioration of the working class, unwanted arrest or cause of social unrest can lead to despair and trigger protest or inevitable revolution (Thomassen, 2012). Which means protest can be a crucial indicator to identify social or political unrest. What if we try to understand the dynamics of euro crisis by analyzing spatio-temporal happening of protest? What if we try to model crisis through protest? Is protest a potential indicator to predict crisis scenario? This study is targeted to study protest and the crisis in Europe in time and per country.

1.1 . Protests

The word protest has its origins from old French and Latin *protester* and *protestari.* The root of which in Latin if separated in two parts: *pro* – means implication of "forth, publicly" and *+test* means "to assert, to witness". In late Middle English to protest, was commonly considered as a verb, which concept by Oxford Dictionary was conceived as a description with the sense of 'making a solemn declaration'

Popularly the protest concept could be defined as a public demonstration, or a strong objection to an official policy or course of action, in conflict literature it is attributed to the use of individual right under democracy (Kornhauser, 2002) or conceived as a mechanism to change (Koopmans, 2004).

It classified as a non-violent action but has a connotation of conflict or fight because can be defined as a 'non-violent struggle', 'non-violent resistance', 'a direct action', 'a civil resistance', and 'a political defiance' (Zunes, 1999) or even a main indicator of domestic conflict (Braha,2012).

Protest is broadly not considered as a conflict but an engagements or prosecutor that aims to achieve political objectives. Study divided them differently (Tanter, 1966) separate them in General-Strikes; more than 1,000 against national government policies, Riots; clashes and violent actions caused by opposition to government policies, and anti-government demonstrations which are peaceful and try to rise voice against government authority or opposition towards government policies (Braha, 2012).

We will use in this study the GDELT Data base, Global Database of Events, Language, and Tone developed by the Kansas Event Data System. This database defines for Protests and subdivide them by action definition, CAMEO Code 141: "Demonstrate or rally and includes demonstrate for leadership change, demonstrate for policy change, demonstrate for rights, demonstrate for change in institutions, regime" and CAMEO Code 145 "Protest violently, riot and includes violent protest for leadership change, violent protest for policy change, violent protest for rights and violent protest for change in institutions, regime" (CAMEO Codebook, Event Codes).

We are paying attention in this study to the act of protest which we consider express tree elements: grievance, conviction of injustice and an action meant to provoke an ameliorative action towards an injust condition (Turner, 1969) related to the European financial crisis situation.

1.2 . Crisis

Crisis commonly defined as a time of intense difficulty or danger or similarly crisis is defined as "a time when a difficult or important decision must be made". The origin of the word crisis comes from Greek *krisis* 'decision', from *krinein* 'decide' and was used in the early 17th century as the general sense term to point to a 'decisive point' .

Now-a day, the financial crisis that started in 2008 in United States is no longer restricted to the financial sector, but has spread to global economies and has created an ongoing *class war* against the global capitalism (So, 2012) is the first time so big in the Western world.

There are different types of crisis, caused for different type of causes as natural resources, ethnicity, corruption, labour, (Rose-Ackerman, 1999; El-Mahdi, 2011) in this study we want to focus on economic indices, due to the current situation in Europe to define what type of causes could be affecting.

1.3 . Protests and European Crisis

There is a common believe of distrusts towards Global Institutions due the current crisis and how is been managed in the countries with more difficulties and less infrastructures in Europe. The most affected the so called “PIIGS” – Portugal, Ireland, Italy, Greece and Spain- have been prime beneficiaries and are now the most suffering countries due to the *economic union* and *political union* of Europe due to the *economic crisis* (Epitropoulos, 2012).

A broad explanation for the above fluctuation is called capitalism, which is as a global inventiveness that seeks profit maximization and capital growth wherever it can find it in the world. Irony is that if capitalism is improving the life of citizens, it is also making them poorer creating a class confrontation from those that do not have a present support towards that ones that are satisfy, and how the people not supported, are expressing disapproval publicly.

Some scholars defined the current financial crisis a crisis on democracy, the most studied factor that encouraged studies of this sort were that a series of economic condition where definitely linked to economic circumstances, leading to authoritarianism and a higher level of conduction of democracy (Gasiorowski, 2010).

States in crisis will potentially become *states of capitalism* according to Friedman this statement made in the early 1976, was define due to a continuously and successive need on benefit search instead of directing towards favor minorities with less benefits. This argument can be translate into the current situation setting the conviction towards an expected loss of parliamentary democracy, visibly described by the incapability of governments to meet intentions for both private and public sectors (Thomassen, 1990).

Therefore, under this circumstances economy is guilty for the decline of democracy, and therefore is compressible if demands coming from citizens in form of protests, are inevitably to occur and possible to rise. In other words, pacific and democratic action as protests increase, while demands of citizens and public sector in a recession time can no longer be met, it triggers to a decline of public/institution confidence, to an end of parliamentary democracy and therefore brings a 'decisive point' in time for an important decision to be made

1.4 . Why this study- Objective

I was motivated in this study because I am a direct affect person for the crisis situation in my country, as a Spanish citizen, student and future worker I had strong motivation in understanding the nature of the social unrest and the current disagreement within some part of the society.

My encounter with studies in the branch of international relations and peace-conflict research approach my interest into model non-conflict. Seen that conflict prediction studies in the recent years have been accepted as a science practice and is in an ongoing improvement (Nathaniel et al.2000) convinced me to choose this topic.

The main research questions of the study are:

- What is the average or density of events in the different European countries?
- Are the current grievances and trends in Europe caused by the current crisis and is there any relation with economic variables and unemployment?
- Does one protest intentionally take place close to another protests and do they influence each other?
- Is it possible to understand the dynamics of euro crisis by analyzing spatio-temporal happening of protest?
- Is it possible to model crisis through protest?
- Are protests a potential indicator to predict or state a ‘crisis’ / ‘limit’ scenario?

Specific objectives are as follows:

- Evaluate the influence of social variables in protests
- Is there any pattern of cluster or interaction
- Simulation of protests/events taking in account historic data
- Does one protest intentionally take place next to another?

1.5 . Available event data

Event data is becoming increasingly large subject in the conflict and political field, especially among studies that accentuate in social forecasting, however in special this year thanks, and conflict prediction is a *hot* theme with the innovation of GDELT- Global Database of Events, Language, and Tone database.

GDELT is considered as "big data", it contains a quarter of a billion georeferenced events capturing global behavior in more than 300 categories from 1979 to present with daily updates.

The strength of the GDELT database is due it is the first that enabled the functionality to record automatically and in real-time, online news from the entire world (Leetaru and Schrodt, 2012). This is a stepping-off on this field, as datasets related to the conflict literature, are mostly human coded. Historically, there are more than 10 projects of datasets related to conflicts, all of which are specific related:

Dataset	Focus	Geographical	Years	Geo-located?	Reference
ACLED	Conflict	Primarily Africa	1997–2010	Yes	Raleigh et al. 2010
EDACS	Violence	Failed states	1990–2009	Yes	Chojnacki et al. 2012
GTD	Terrorism	Global	1970–2010	Yes	START 2012
ICEWS	General	Asia; global	1998–2010	No	O'Brien 2010
KEDS	General	Primarily Middle East	1979–2011	No	Schrodt and Gerner 2010
KOSVED	One-sided violence	Selected states	varies by case	No	Schneider et al. 2012
MID3 Incidents	Conflict	Global	1993–2001	Yes; MID- LOC	COW 2007; Braithwaite 2010
NIRI	Violence	Northern Ireland	1968–1998	Yes	Sullivan et al. 2012
PITFWAD	One-sided violence	Global	1995–2012	Yes	PITF 2011
SCAD	Protest	Africa	1990–2010	Yes	Salehyan et al. 2012
SIGACTS	Violence	Afghanistan, Iraq	2004–2010	Yes	HSRP 2010; Linke et al. 2012
SPEED	General	Global	1946–2010	City	Nardulli 2011
UCDP-GED	Organized violence	Africa	1989–2010	Yes	Melander and Sundberg 2011
UCDP/PRIO ACD	Conflict	Global	1946–2011	Conflict site	Themnér and Wallensteen 2011; Hallberg 2012
Urban Violence	Urban disorder	Africa, Asia	1960–2009	Yes	Urdal and Hoelscher 2012
VRA	General	Global	1990–2004	No	King and Lowe 2006
WITS	Terrorism	Global	2004–2010	City	NCTC 2011
WARICC	Water-related conflict	Mediterranean, Middle East, Sahel	1997–2009	Yes	Bernauer et al. 2012

Table1: Event data projects
Source: Schrodt, 2012

Currently, the most known platform called Integrated Conflict Early Warning Systems (ICEWS) project includes the Penn State Event Data Project datasets to provide information in a daily level format and even in a regional scale (Yonamine, 2012).

Moreover, it is important to note this progress in conflict forecasting studies. There are two types of datasets, non-event data sets (MIDS, COW, ACLED, etc.) and the machine coded event datasets as WEIS and its derivatives, which use the CAMEO and IDEA (Brandt et al., 2013) coding system. The non-event data studies as in (Hegre et al. 2012) and (Goldstone et al. 2010) are human constructed criticism remark their need to rely on humans to select the observations. Building a collection *a priori* can trigger to mislay events. In the other hand, machine collections cannot be influenced by human interests, because automatically a broad observation will be included, collecting a huge magnitude of real-time news collection.

Nonetheless, the difficulty resides in the complexity and the need of technical skills to building datasets with GDELT, despite to select the correct sample inside a windows frame is more flexible in machine coded data event (Yonamine 2013) as shown in the Conflict in Afghanistan.

1.6 . Study area

Social and political forecasts have include both country-level assessments of political instability, civil wars, coups, etc. (inter alia, Beck, King and Zeng, 2000; Gold- stone et al., 2010; Hegre et al., 2012), as well as region and country specific predictions of social events of interest (Brandt, Freeman and Schrodt, 2011; Brandt *et al.* 2008; Brandt and Freeman, 2006; Goldstein and Pevehouse, 1997). Hense based on this studies this research has been motivated to work in countrywide and NUTS2 in Europe.

Europe is constantly under changes, one of the challenges the union is facing nowadays is the current financial crisis that is beating the southern countries. Social unrest has increasing in all countries, perhaps a cause of the introduction of the number of countries, in any case, is the first time that several countries rise their voice for same reason (Rucht, 2002).

Europe is a unique economic and political partnership of 28 European countries that over the time has been considered as a geographical region or for some researchers a continent, for others it is conceive much as an idea (Stråth, 2002) as it is a real geographic area, surrounded by oceans and seas and delimited for the Caucasus and the Urals in the East, has a high number of islands and rivers. According to the United Nations, Europe is slightly the 11% of the population of the world in 2007 while a century ago it has 25% the decrease is explained by the decrease of the nativity nearby 2.5 in almost all European countries.

Europe is also divided in 4 sub-divisions, most known as Northern Europe; formed by Ireland, Island, United Kingdom, Denmark, Lithuania, Lithonia, Latvia and some of the Scandinavian countries, the Easter Europe, starting from Poland, Hungary, Check Republic, Rumania, Moldavia until Russia, the Western Europe; France, Belgium, Netherlands, Germany, Luxemburg, Austria and Switzerland and the Southern Europe, Portugal, Spain, Italy, Greece, Croatia, Albania and Estonia and the Asian portion of the European countries as Georgia.

There are other sub regions in Europe that have risen by the different conceptions as the known PIIGS acronym that usually refers to the economies of Portugal, Ireland, Italy, Greece and Spain, has been change to the general southern economies of Europe and the current vulnerable economies.

Nonetheless, Europe is an idea fighting to make a unique reality which has been under changes over the years and tries to constructs a European identity expressing the strong unity of the different countries that form it. Europe was created after the Second World War, to foster economic cooperation, to trade and avoid conflicts, It changed into a political union developing a numerous institutions that work toward consensus regulations on all kinds of policies that include all the countries founding a big democratic net, that face the dynamics of the change of all the nations and economies taking part.

The estimation of population in Europe was around of 711 million in 2010 (Eurostat) is known for its excellence in education and its success on forming part of the most advances agreements on human rights and progressive policies leading to include people from all nationalities of the world in its huge system.

1.7. Structure of the thesis

This thesis is organized into four chapters, which follow a chronological course of ideas as follows:

Chapter 1- Introduction:

This chapter presents a background to the study, objectives and overview structure of this thesis.

Chapter 2- Materials and Methods:

Here, we present the software tools used for the thesis analysis, description of data used and analysis methods implemented in the study.

Chapter 3- Results of analysis:

In this chapter, we present the results obtained from statistical analysis of the protest data.

Chapter 4- Discussion, future research and conclusion:

We present discussion on the results of analysis in this chapter. Recommendations on the directions for future research, and general conclusions are also presented here.

1.8 Overview of methods used to support the analysis

The analysis has involved several steps. Firstly, we downloaded and process the dataset, choosing the coding, actors and actions that fall within the window of study of Europe. Process the data, selecting the codes and understanding the relations between actors was tented in order to understand protests related to the crisis situation in Europe.

Regarding the descriptive variables, we approached two spatial scales, primarily we approached the EU-NUTS2 socio-economic units, and subsequently we approached a country wise analysis. NUTS classification (*Nomenclature of territorial units for statistics*) is a hierarchical system for dividing EU. The aim of the first approach was to achieve a superior granularity of the study.

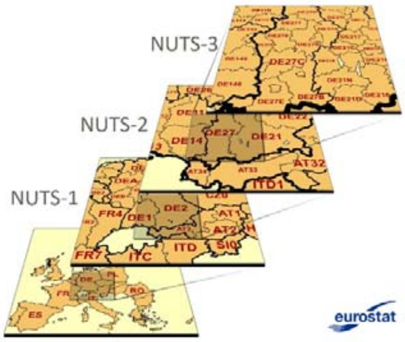


Figure 1: NUTS 2: basic regions for the application of regional policies

Computing NUTS2 analysis obligates to aggregate protests events by regions. It is worth to mention that due missing values and noise in the GDELT database, plus inconsistency in Eurostat database, while calculated we considered several reasons to finally reject. For that reason, not all the results are included in the study.

We computed in cases, an analysis for 10 years (2000-2010) yearly. The countries included in the analysis are the following: Belgium, Bulgaria, Czech Republic, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden and United Kingdom.

All together we want to compute spatial–temporal model that characterize most model of conflicts (Weiderman, 2010) and by linear model examine spatial dependencies-correlation related to the geo-located events.

1.8.1 Lattice approach

The first model approach is based on lattice approach. With the aim to find if unemployment rate and other variable related to austerity measures, and those affecting the increase of the number of protests with the porpoise to finally consider them as explanatory variables. Likewise, we choose economic, education and demographic variables seemly to estimate their influence and aim to discovery correlation with protests. To compute this objective we computed the Ordinary Least Squared Method.

The second model within the lattice approach included a temporal analysis of t-1 of Ordinary Least Square Regression Model, in which we operated protests of a past year and variables of the current year.

The overall process was computed using R environment, which I especially thanks to developers of stats package.

1.8.2 Point process analysis

This second part of the study concerns spatial analysis. Firstly we computed for Europe and then for each disjointed country. After results, we extend the analysis to a spatio-temporal analysis for one country, Spain.

Spatial point processes is a model used to compute random point pattern usually in 2 or 3 dimensions. In this study we will use the 2 dimension approach. Bivand in 2007 contributes to point pattern model defining it as "...a stochastic process in which we observe the locations of some events of interest within a bounded region or window *W*." (Bivand et al., 2007).

Spatial point process modeling is a leading study inside the field of spatial statistics that further explaining the spatial distribution and data pattern inside the database, drives to explain the reason of why the observed points precede from a particular pattern. This technique to model patterns can be compared with a regression models in classical (non-spatial) statistics, if there are covariance that can be included as determined content that influences a point location. There is a broad literature in different fields that use point patter analysis (of fields including forestry and plant ecology, epidemiology, seismology, wildlife ecology, geography and event conflict forecasting).

The core of spatial point processes modeling is first to affirm whether or not the points observed are in a point pattern random distribution. Therefore, in point process a first step is to compute, using the kernel function of intensity and evaluate the principle Complete Spatial Randomness (CSR) – where events are assumed to be randomly distributed or inhomogeneous, indeed, if would be no randomness any analysis will end here.

Point process models are essential to interpret the result of the different factors on point organization, but also is used to anticipate events location in regions where the point distribution is unknown by using only their location.

All this analysis was computed in R statistical software, to download database process and computes the first statistical analysis using STATS package, similarly for spatial point pattern models, in the packages spatstat (Baddeley and Turner 2005) and SPLANCS (Rowlinson and Diggle 1993).

CHAPTER 2: MATERIALS AND METHODS.

This chapter describes the methodology followed in analyzing the impact of unemployment and other variable and the distribution and simulation of protest event data. It presents the study area, sources of data and the tools used to analyze the data.

2.1 Problem and area of study

In 1999, most countries in the European Union adopted the euro as a common currency. This union allowed poorer countries like Spain, Portugal, Italy and Ireland to borrow money at a low rate of interest and get rich as other northern countries as Germany.

The strategy mentioned, summed policies allowing building a huge bubble of debt and housing. It was collectively accepted in countries of the south of Europe, Greece and Spain, hence sponsored the idea of achieving a large public-welfare state. Unfortunately, the output was not what was expected after 20 years of strategy, Greece was the first country to broke, with a high corruption and leaving the country with a superficial development.

Grievances regarding this situation haven't change since the starting of the first wave of crisis in 2000; indeed they have grown in each country, creating large European citizen complaint in the contemporaneous history of Europe. The problem is bringing chaos, the sum of sudden cuts on social policies and health, the amount of privatizations recommended by the named 'austerity measures' as lead the default states and of. In consequence, investors are more concerned about the risk and indeed consequently borrowing risk in the region increase. After all this years, not only citizens of Greece but also Germans and all over Europe start to believe that this situation is contagious and not easy to solve.

Europe that has 711 million habitants based in Eurostat source in 2007, has felt one of the biggest protests of its time if unifying all southern countries and Ireland, and some minorities of other countries. In May 15th in Spain 15,000 people gathered in “Puerta del Sol”, in May 17th 30 cities around Spain including Barcelona and Valencia, joined this protests. According to Britain's The Guardian, "tens of thousands" camped in the floor of the Spanish capital in Madrid and around the country on the night of 19–20 May. Similarly, in Greece, the pacific protests have transformed into riots, attacking bank and firing state infrastructures, as a form of demonstration after the dead of a 15 year student in Athens in 2008, other type of protest did already happened before as the interruption of students in the public TV before the crisis started.

Outside Greece, as a solidarity riots and in some places clashes with local police happened, in around 70 cities around Europe; London, Brussels, Rome, Dublin, Paris, Berlin, Copenhagen, Madrid, The Hague, Barcelona, Frankfurt, Bordeaux, Seville, Cologne, Nicosia (Cyprus), Amsterdam to mention some of the places that contributed.

2.2 Materials

2.2.1 Analysis tools.

In this study we have used more than one tool. An important tool used mainly to download the data, process and do statistics was R Software in different versions in which we use different packages within the R environment. The principal package was spatstat. Event protest data was also processed using Microsoft Office Excel but was mainly for the explanatory variables where it mostly used. ArcGIS was used to have a preliminary overview of the data for NUTS2 and country in a layers format.

2.2.2 Event data: GDELT

Global Database of Events, Language, and Tone database comes as an sum of columns and rows where each observations includes a series of complex attributes under 3 dimensions: Action, Actor and Temporal-Dimension.

An event is read from online sources, where machine learning systems convert the transitive sentences to natural language sentences, in order that *events* are indirect or indirect interactions between elements of “set of actors” or “set of actions”, associated to a time and space. (Schrodt, 2013)

Event and actor ontology, compound of simple verbs and nouns phrases dictionaries, that are automated and define the EOI (Events Of Interest).
The GDELT database thus has [who-did what-to whom, ***when***] structure.

Date	Source	Target	Action	Latitude	Longitude
201240512	GOV	CVL	140	41.39484	2.175179

Table 2: Example of a GDELT Event Data Event

Actions include protests, bombings, speeches, peace agreements and a myriad of others.The interest of this study is to take protest events, thus we must convert raw event datasets to a more usable format (Yanomine, 2012). In order to indicate the specific actions that occur between actors an action typology is.
GDELT events are under the WEIS Conflict data framework inside the CAMEO Code also used for IDEA event data studies.

14: PROTEST
140: Engage in political dissent, not specified below
141: Demonstrate or rally, not specified below

- 1411: Demonstrate for leadership change
- 1412: Demonstrate for policy change
- 1413: Demonstrate for rights
- 1414: Demonstrate for change in institutions, regime

142: Conduct hunger strike, not specified below

- 1421: Conduct hunger strike for leadership change
- 1422: Conduct hunger strike for policy change
- 1423: Conduct hunger strike for rights
- 1424: Conduct hunger strike for change in institutions, regime

- 1443: Obstruct passage to demand rights
- 1444: Obstruct passage to demand change in institutions, regime

143: Conduct strike or boycott, not specified below

- 1431: Conduct strike or boycott for leadership change
- 1432: Conduct strike or boycott for policy change
- 1433: Conduct strike or boycott for rights
- 1434: Conduct strike or boycott for change in institutions, regime

144: Obstruct passage, block, not specified below

- 1441: Obstruct passage to demand leadership change
- 1442: Obstruct passage to demand policy change

145: Protest violently, riot, not specified below

- 1451: Engage in violent protest for leadership change
- 1452: Engage in violent protest for policy change
- 1453: Engage in violent protest for rights
- 1454: Engage in violent protest for change in institutions, regime

The CAMEO Code has 20 “categories”, or classes of events, called as Event Codes and containing sub and sub-sub categories. In our case of study we proceed to process Actions 14 named PROTEST which includes also subsets.

In a second step to process the event historic database, we selected all the events inside the window of observation (Europe) which Actors ware European countries and which Actions occur in Europe [when]. There are two Actor s in each event [who-did what-to whom], as it is considered that sometimes there is a direction or two parties taking part, thus we find two different columns, one for Actor1 and other for Action2.

The actors usually are countries, but can also be secondary role codes, global or other action codes. Countries are defined in 2-Digits form based on the NATO country codes, the name of the country in English and also the region name and code, can be found in a separated column. Moreover, when we are this Actor can be further explained with attributes that specifies the type of event. In the following tables are some of the codes attributes we selected:

Country Codes	
Primary Role codes	
COP	Police forces, officers
GOV	Government, the executive, government parties, coalitions parties
OPP	Political opposition, opposition parties, individuals, activists
Secondary Role Codes	
BUS	Businessmen, companies, etc..
CVL	Civilian individual or group
EDU	Education, educators, school, students
ELI	Elites, former government officials or celebrities
LAB	Labor, workers, unions
LEG	Legislature, parliaments, assemblies, Lawmakers
MED	Media journalist, newspapers, television stations

Table 3: Primary and secondary codes added to country codes for each Action in GDELT

Global	
IGOGOE	Group of Eight (G8)
IGOSCE	Council of Security and Cooperation in Europe (OSCE)
IGOIMF	International Monetary Fund (IMF)
IGOUNWBK	World Bank
IGOWTO	World Trade Organization (WTO)
Other Action codes	
IGOEEC	European Union
EUR	Europe
IGOEEC	European Union

Table 4: International/Transnational Actors

GDELT database is a Geo-referenced Event Dataset (GED) offering the geo-location of each event recorded in columns as “longitude” and “latitude”. Moreover, the names of the locations are also recorded thanks to the automatic

TABARI Code system, the observations recorded preserve the local names (if mentioned in the new) and countries and translates them into the 2-Digit NATO code system. To end, the temporal-spatial dimension is found in the first column with a MM/DD/YYYY structure of the moment the new was published.

In this study we went for region and after assessing the missing values and noise, we decided to take country-wise of event codes, selecting the columns of the country codes called GeoAction_ConuntryCode and expressed in longitude and latitude.

The database is an ongoing project that now is online available there has been several improvements, some sources have been unclouded in the past times, for that reasons is compulsory to normalize the data for the all average of protests recorded since they started. Luckily, all the information is free and available online in a blog posted by **Patrick Brandt**
(http://blog.gdelt.org/2013/09/28/normalizing-gdelt-protest-data/)**.**

To download the entire database, process and visualize event data we used R software, the GDELTools package and Excel. The database is a XXX size with contain XX points for 1979-2012 with daily updates of news from the entire world.

2.2.3 Dependent, explanatory variables or predictor variables

Eurostat is a Directorate-general of the European Commission. Its main responsibilities are to provide statistical information to the institutions of the European Union and to promote the harmonization of statistical methods across its member states and candidates for accession as well as EFTA (European Free Trade Association) countries. The organizations which cooperate with Eurostat are summarized under the concept of the European statistical System. The Eurostat statistical work is structured into Themes and Sub-themes from witch I have selected variables related to economy and finance, population and social conditions, industry trade and service and science and technology.
The list of the variables selected is the following:

a. *Government deficit and debt*
b. *Taxes on production and imports*
c. *Total genera/central/local government expenditure/revenue/Net lending(+)/net borrowing(-)*
d. *International investment position*
e. *National GDP and main components*
f. *Final consumption aggregates*
g. *Labour productivity*
h. *Population density*
i. *Crimes recorded by the police*
j. *Social protection expenditure*
k. *House price index.*
l. *Unemployment rates by sex, age and highest level of education attained*
m. *Unemployment rates by sex, age and highest level of education attained (%)*
n. *Expenditure on education as % of GDP or public expenditure*
o. *Number of Student in Tertiary education (1 000)*
p. *Income, saving and net lending/ borrowing*

It has been mentioned before that economic crisis leads to an increasing gab to produce a good government performance, to see in what extend is this gab we have selected indicators of government performance for all countries: total government expenditure and revenue general, central and local (*a,b,c,d,e*) can affect directly the performance of the governance of a country.

According to "modernization theory" the level per capital income (l,*m,p*), the extent of literacy and education (*o,n*), the degree of urbanization (*g*), and the quality and extent of communication media (not considered as GDELT is built from media) has also a role into this transition of democracy to authoritarianism (Gasiorowski ,2010).

A first approach to achieve a better resolution/scale in the analysis was to employ data from NUTS2 from each country, but the consistency in the database for the needed 10 years encouraged to dismiss this solution and choose a country wise database for the previous mention variables.

Based on the availability of the data, a second approach was to take country level database for year for all the countries taking in consideration. This last approach was the one decided to use based on the availability of the database. An attempt to acquire monthly data and quarterly data was dismissed as the amount of data was not available for all the variables.

Details on the datasets constructed on the explanatory variables can be found on Appendix A.

2.3 Method chart

Figure X shows a flow chart of the overall methodology used to analyze the protests event data. Each component of the methodology is detailed below.

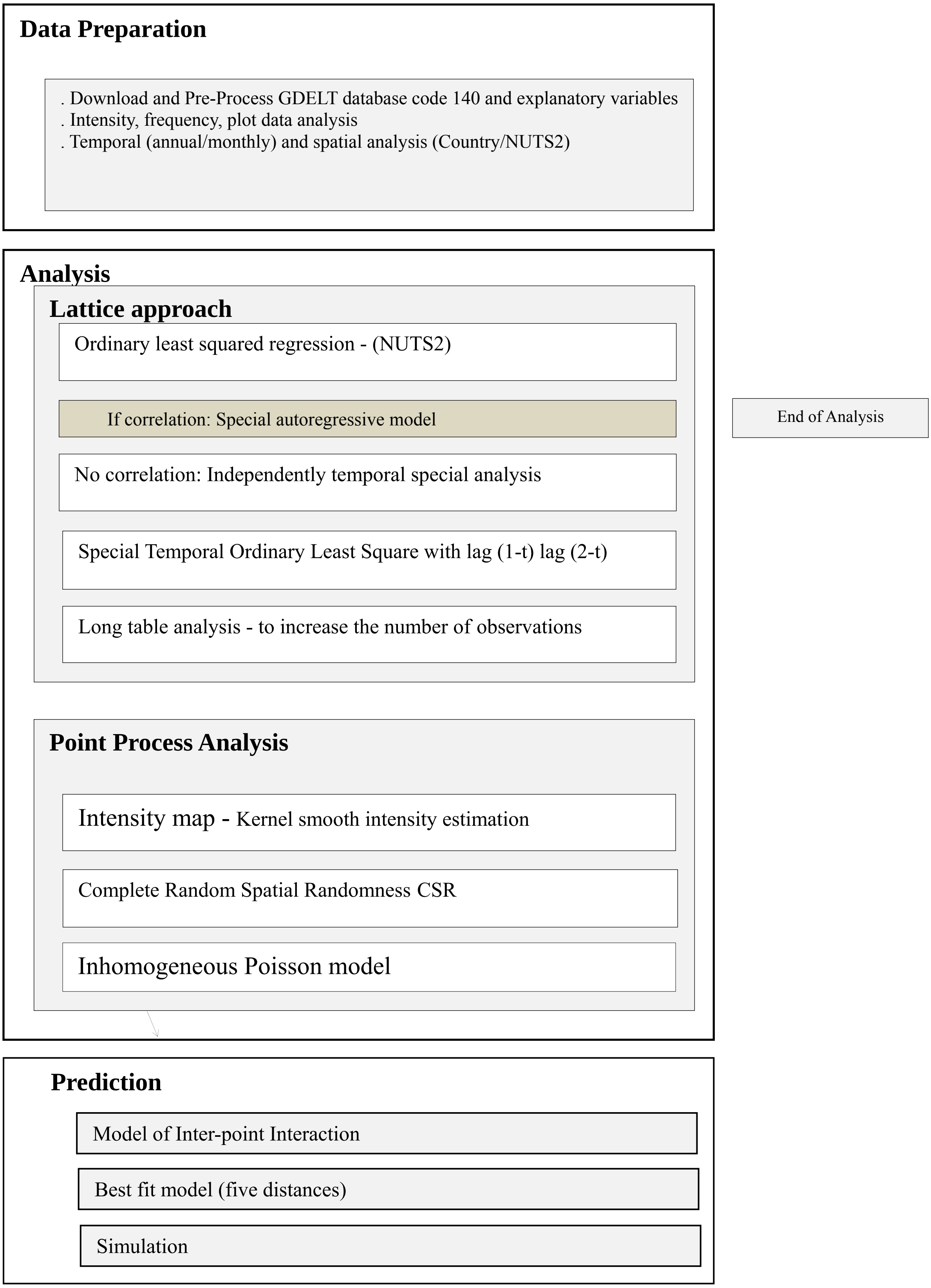


Figure 2: Flow chart of overall methodology enforced to analyze protest event data

2.3.1 Explanatory Analysis

a. Data pre-processing

GDELT database is available online to download for free in.csv or excel format and a blog to download it into R Software, moreover and in April 2013 was the publication of a package called GDEL Tools in R. The GDELT database has been growing since 1979 to present, several sources have been added in 2005, and thus the situation requires normalizing the conjunct of data before we can do any further of analysis.

The CAMEO codebook is a method to gives order to the recorded news. Our study focus in protests, which is translated in CAMEO Code as number 14 and for deeper to specification we selected a political deviation selecting only code 140, 145 and 141, which description is the following:

14: PROTEST
140: Engage in political dissent, not specified below
141: Demonstrate or rally, not specified below

- 1411: Demonstrate for leadership change
- 1412: Demonstrate for policy change
- 1413: Demonstrate for rights
- 1414: Demonstrate for change in institutions, regime

145: Protest violently, riot, not specified below

- 1451: Engage in violent protest for leadership change
- 1452: Engage in violent protest for policy change
- 1453: Engage in violent protest for rights
- 1454: Engage in violent protest for change in institutions, regime

Once we selected the Action_code, we need to select the area of study. The European countries and the window can be selected by choosing names, codes and long/lat units of the action observed.

Note that, since GDET database is recorded and organized as a table, with rows and columns, we have to select the necessary rows to build or own dataset. Another dimension of the dataset, that we have to define are the Actors1 and Actor2, so we selected the name of the countries, their codes (NATO codes) or regions.

a. Descriptive analysis with variable

Least Squared Regression Method is used to fit linear models, is frequently used to calculate the slope and interception of the best line through a set of data arguments. It is also use to carry out regression, single section analysis of variance between two datasets and to analyze the covariance (Chambers, 1992)

We choose Least Square Regression because is a common technique to compare data in most models of conflict (Wiedmann, 2010). Defines the relationship between two methods (1), the test and the reference method, and is based on two arguments, the slope and intercept of this line, providing an estimation of the constant error.

$$y = a_{y.x} + b_{y.x}\, x \quad (1)$$

Equation x: Equation of Ordinary Squared Regression Methodology

Where *x* is the independent variable (protests) y is the dependent variable (variables from Eurostat), $b_{y.x}$ is the slope and $a_{y.x}$ is the interception of the line.

The method also finds residuals, resulting of the difference between the *y* values observed and those predicted by least-square model. More closely, after computed the line by least-square regression, it minimizes the sum of squares of each point distance that is found between the observed data points and the line.

In principle, what we try is to fit the data minimizing the residuals, meaning the error from the observed data and the real data. The methodology to fit these errors in the variables is the least square.

b. Ordinary Square Method for t-1

We believe that social unrests-protests-riots are a consequence of an injustice or a disagreement caused for what a majority conceive it as an injustice policy.
The ambitious component of the hypothesis is to find out how to quantify its influence. We know that unemployment, austerity measure caused by higher debt from the central governments or the break of the banks has affect societies and people in the real life of the citizens, as we followed in international news in the last years, mostly in the southern countries of Europe.
Besides discussing if the measures are proper or not, we want to see if this measures are a cause.

2.3.2 First order characteristics of protest events location and spatial

distribution.

First-order provides a general idea of the spatial distribution of our point/event data. Firstly, we study the distribution since intensity is the average density, there for it will determine if the data is constant ('uniform' or 'homogeneous') or inhomogeneous. This is a mandatory step in order to start with first order characteristics; it is the first measurement to learn more about our data and its distribution concerning our boundary or concrete area of study (Bivand et al. 2007).

Our hypothesis is that every event/ protest is independent from other events in space since each event has some characteristics that are the cause. In any case, we need to prove that our assumption is true, for that reason we need to discover if there is a pattern.

During this process, the main objective is to discover the pattern and mostly to decode if the intensity of the point process differs from place to place, in which case, the pattern intensity may be inhomogeneous. If the output is inhomogeneous the events are not related to each other and they location depends on their own coordinates.

The observations we are considering are defined within a defined region of study or window, we accurately selected those points inside the window, and therefore any point that is outside that boundary is not taken in account in the study (Baddeley 2008).

The distribution is given by the kernel intensity estimation function (1), where if there is concentration or the intensity function will be satisfying.

$$\mathbb{E}[N(\mathbf{X} \cap B)] = \int_B \lambda(u)\,\mathrm{d}u \quad (1)$$

In deep, Kernel-smoothing intensity was calculated by an algorithm in R that uses a regular grid and then uses the fast Fourier to calculate to approximate to the kernel and uses linear approximation to evaluate the density (Scott, 1992).

The statistical properties of a kernel are defined as *sig^2 (K) = int(t^2 K(t)* *dt)* that *=1* and *R(K) = int(K^2(t) dt)* (Scott, 1992).

The function defined in an area an estimation of the total number of observations fall in the same areas, and uses the measures it accordingly to the given location ("long/lat") of each observation.

Non parametrical technique is the one suspected in this research, as we work with data naturally inhomogeneous (Diggle, 2003).

2.3.3 Inhomogeneous Point Process

b. Second-order characteristics in event protests spatial distribution.

After a first-order analysis, we went for a second-order characteristics in order to provide information about a possible event-point to event-point interaction. The first-order distribution exclusively provides information about the event pattern distribution, where we can find trends in the data, but on the other hand, second-order properties offer the best spatial tool to evaluate a point pattern interaction and explain if there is a relation between each point (Illian et al. 2008).

Second order uses Ripley's K function to test for Complete Spatial Randomness of the protests events. There are three reasons for testing Complete Spatial Randomness or inhomogeneous point processes:

(1) CSR is used when data is random to verify its randomness

(2) CSR is a start in order to explore the distribution of our data

(3) CSR has to be rejected to continue a point pattern modeling as it found clustering

Therefore, to prove our data is random and pursue these three condition we run the kernel density which estimates with the given kernel and bandwidth 50km the distribution of the protests, considering all the points and therefore will determine trends.

Adjusting the model for the spatial inhomogeneity of the data takes in consideration modification of the K function, due to in this case each point will be weighted by $\mathbf{w}_i = 1/\lambda(\mathbf{x}_i)$.

Moreover, K function considers different distances from point to point relationship, second-order properties are especially interesting because are used to include points that are far away. Ripley function, takes for each point further distances than the nearest neighbors, thus each point computes a large distance.

The concept behind K function, by Ripley (1977), considers first the average number of points within a distance *r* from a regular point. Then λK(r) marks the mean number of points in a disc around the radius *r* from the centroid of the point *x* selected, then this is computed in n times and in *N* number of points that are lcoated in the window *W* of study. As a result we have an equation where *n i (r) = N(b(x i , r)\{x i })* meaning that the total number of points *N* that occur within a distance r from the selected point x i , that excludes x i itself,

$$K_{inhom,pots}(r) = \pi r^2$$

Equation x: Inhomogeneous K function

estimates λK(r) distances. means the expected number of events within a distance *r* of an random event-point i and λ is the intensity or the mean number of points per unit area.

When computed the K function it is firstly compared under Complete Spatial Randomness to find the intensity of the points and the spatial distribution, then computed 99 simulations (recommended in point process modeling).

Based on point process model we fitted the model and out if this model we computed 99 simulations and figured an envelope to define the upper limit of the envelope and the lower limit of the envelope. If the model for each country relies inside the envelope we will consider that the model is good and thus, the protests are independent from interaction between each other, if the lines of the model rely outside the envelope the model will not serve to describe the inhomogeneity.

When create an envelope we look at different indicators in order to explain the results, the empirical line (black line) which is computed with the real data and the theoretical line (red line). In the case of having the black line close to the theoretical line, means that the model fits. If the black line is over the envelope, suggests that there exist clustering (protests gathering), while in the contrary case it suggest heterogeneity (protests dispersion) on the spatial distribution. If both lines are closer to each other, it means that the model we computed is worthily, both real data and expected data are the matching.

We used spatstat package to compute the K function, the 99 envelopes, select the best fit the model and plot the simulation.

2.3.4 Area Interaction fitting model and simulation.

The Models were fitted to estimate the intensity function and possibly describe the best of point pattern model.

1. Modeling for a homogeneous Poisson Process (HPP)

This model assumes that the data is constant in a study area *W (window)*, where the mean intensity is λ|W|, for the Poisson distribution, and therefore describes isotropy and stationary (Bivand et al. 2007) meaning that there is not spatial interaction between points and furthermore, explains that the number of events *n* occurring within the *W* (*window*) are uniform. (Baddeley, 2008; Bivand et al., 2007).

2. Modeling for an Inhomogeneous Poisson Process (IPP)

Protests are hard to simulate as they can occur under any circumstances and their behaviour is not linear, they can change under any circumstance. If the model computed by IPP fits correctly, the assumption of inhomogeneity is confirmed, meaning that protests are independent from interaction between points.

We used spatstat functionalities as Maximum Pseudo likelihood and polynomial to compute a wide range of sequence of distances. The Maximum Psudo Likelihood Estimation Method was used to estimate the coefficients for IPP, to find the best model, using this function we found the optimal distance, and using this distance we fitted the best model.

Polynomial function was also used to capture the spatial trend of the point distribution. Polynomial function is defined as a conditional intensity function by

2nd order polynomial function (Baddeley 2008).

In this section we expect to see a fit model, which explain the distribution or trends of the event/protests. If the model doesn't fit, meaning the lines are not inside the envelopes, we'll conclude that a better model can explain the trend or the spatial pattern distribution of the events.

3. Area Interaction model

Area interaction model was computed in order to discover if there is a spatial pattern between the points. We will calculate this model, due Inhomogeneous Poisson process model did not fit the data. Therefore, Area Interaction attempts to capture the higher order properties as inter-point interaction, merely if the model before was inadequate.

The assumption of Area Interaction is that the distribution of the protests is influenced by the same distribution of the protests, and assumes that interaction is influencing the distribution of the events.

The model imprisons a radius for each location, first a default distance and then we set five different distances with an interval of 50km each. The analysis is made to find the probability of finding a point base on inter-point interaction within different discs. Thanks to this computation we can understand if the data is clustered or is dispersed.

In order to perform the simulations, we used the command areainter in R software. The envelopes are a summary statistic than when plotted in envelope. . (Baddeley, 2014). We used R and the spatstat functions for to compute this analysis.

2.3.5 Predictions from the fitted models.

I used the fitted models to evaluate the trend of the location of the future protests and so predict them in the area of study. It would be important on conflict science to verify where else protest could occur, given the fitted model, using historical event-point data with locations we obtain information about the unknown protests locations, taking in account different distances. This is provided by the function predict.ppm in spatstat (Baddeley and Turner, 2000).

This function fits the observed point pattern, including spatial trend, interpoint interaction (Baddeley and Turner, 2000) and can also compute the dependence on covariates but in this study it was not computed.

The model have fitted the point process model computed the interpoint interaction of the observed point pattern, considering the best distance for each point observed and computed the starting from 1000 m until 250000 m, and compute the function by 50000m, creating a sequence 50 km until 250 km.

CHAPTER 3: RESULTS OF ANALYSIS.

3.1 Explanatory Analysis

We computed an overall explanatory analysis regarding the protests, in order to understand the nature of the GDELT data base. In Fig 2 we can observe the countries with more number of events since the beginning in 1979. United Kingdom, France, Germany and Greece, Spain, Ireland and Italy have between 20000 to 60000 observations.

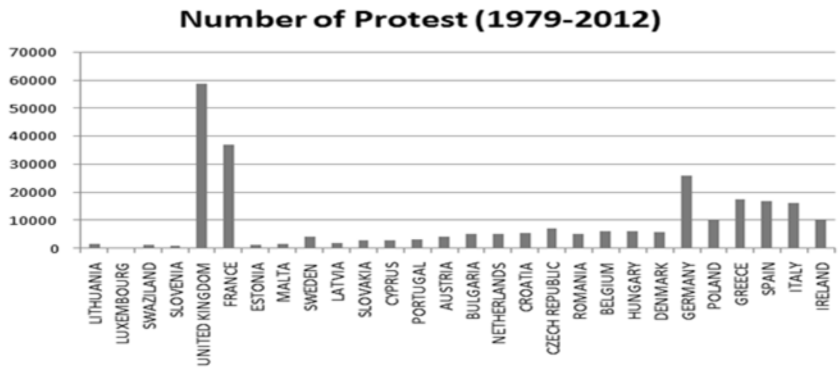


Figure 3: Protests events found in the data base, since its start.

Temporal analysis during those countries displays a clear raising trend of protests which can be observed in the following Figure, for the above countries mentioned.

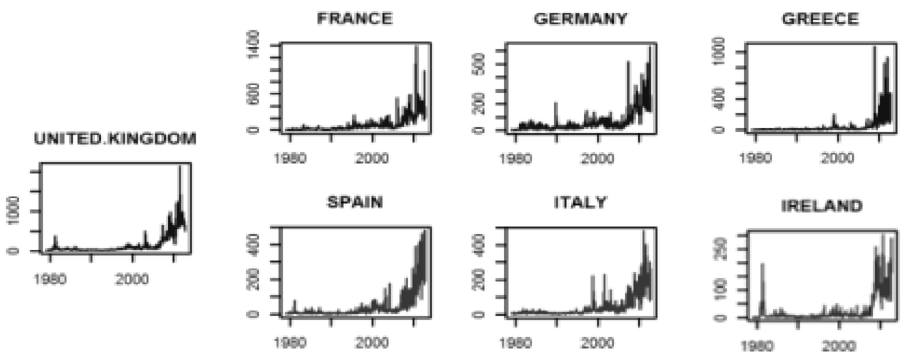


Figure 4: Trend of protest since 1980 to 2010of the countries of the study.

Considering acutely the Actors involved in the protests events during the period studied (2000-2010) we observe a clear amount of events related to Government (GOV) and to Opposition (OPP), followed by education (EDU) and police (POL).

The higher counts are in United Kingdom and Greece counts with a not significant amount of events related to government compared with other countries, in despite of being the code with higher numbers. Police and opposition are the second code for all the countries.

In the other hand, education, labor (LAB) and civil (CVL) are also noticeable in all the countries. United Kingdome (UK) differentiate from the patters described before, because has a high rate of observations related to business (BUS).

Codes	AGR	BUS	COP	CVL	EDU	GOV	JUD	LAB	LEG	MED	MIL	NGO	OPP	(blank)	Total Events
UK	40	213	489	155	613	831	44	371	138	62	105	13	847	5117	9207
FR	59	84	318	66	430	492	40	440	58	29	46	9	428	3824	6422
GM	36	62	247	53	130	292	18	164	23	38	34		364	3037	4578
IT	16	19	143	100	127	228	20	95	43	33	15	3	257	2077	3215
SP	27	20	71	51	76	190	14	98	66	30	10	3	163	1930	2778
EZ	12	3	66	34	49	114	10	34	19	11	4	1	70	1123	1590

Table 5: Actions codes related to the countries of the study

SEP	separatist groups	LAB	Labor
AGR	Agriculture: individuals and groups	LEG	parliaments, assemblies
BUS	businessmen, companies	MED	Media
COP	Police forces, officers	MIL	Military
CVL	Civilian	MNC	Multi-national corporations
EDU	educators, schools, students	NGO	Non-Government Organizations
GOV	Government	OPP	Political opposition
JUD	judges, courts	RAD	Extremist
		REB	Rebels

Table 6: Legend of the observed action codes

In order to understand in deep the nature of the relation between Actors coded in the protest, we attempt to describe their relation computing a spider diagram. We selected the actors with more than 100 mentions shown in Table 6 and plot them. The first spider diagrams consider all the actors, and then we computed another spider diagram erasing the actors with more mentions.

Government is mostly related to labor and to police being the actors with more mentions, about 300 mentions. It is followed by opposition; which obviously fits with the nature of the protests.

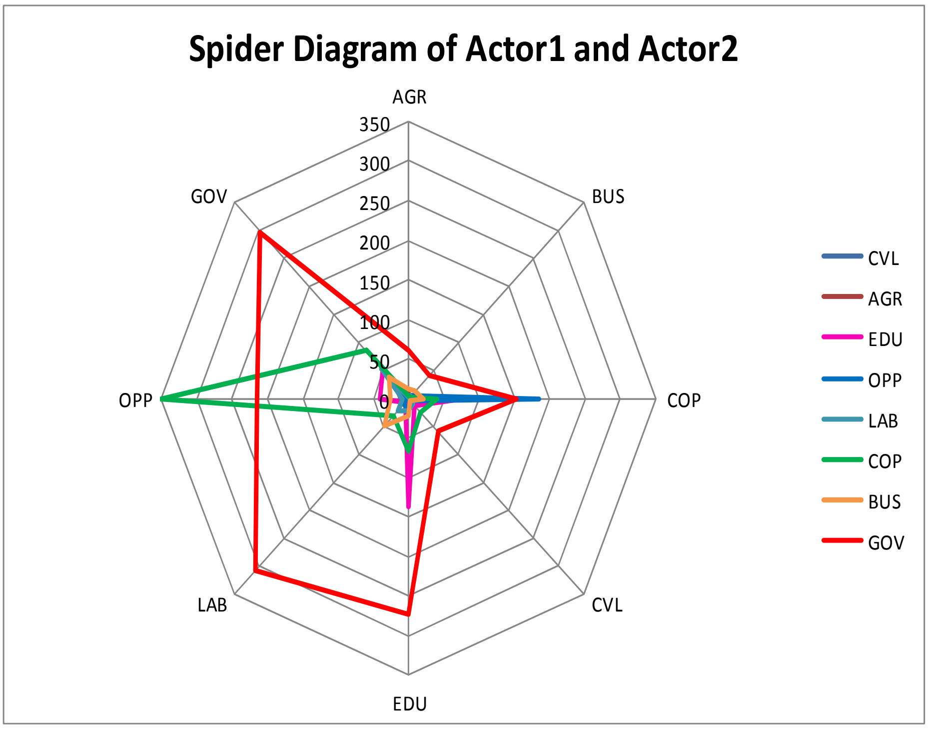


Table 7: Spider diagram of Action1 and Acton2 in protest events

After erasing the most mentioned actors, we see how education and police have a high relation of around 100 to 50 mentions. We can also notice a sign in labor and in police, perhaps related to news where police was involved, which could also be the same new coded twice.

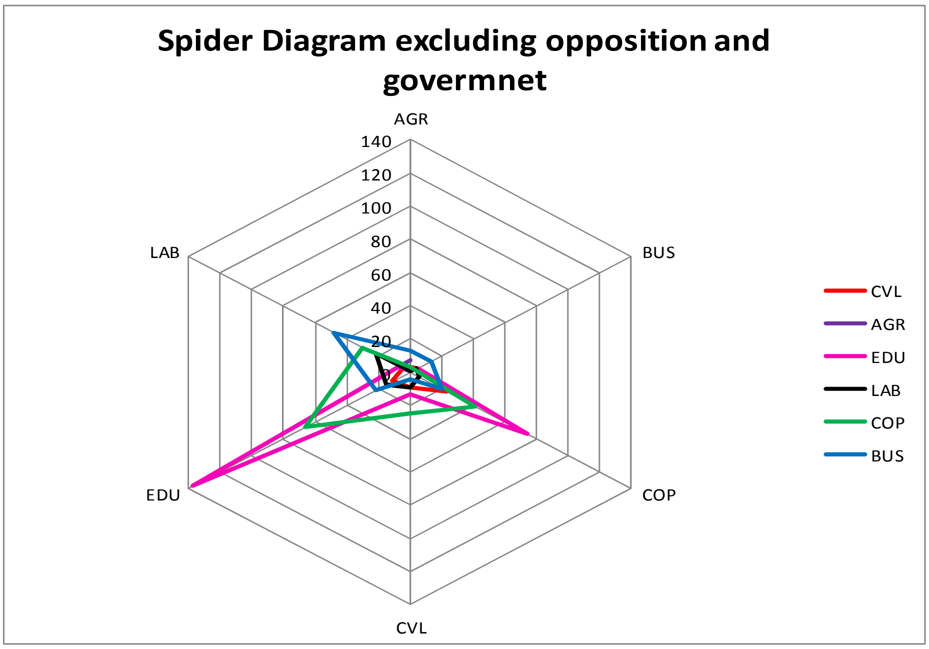


Table 8: Spider diagram of Action1 and Acton2 without GOV and OPP

a. Regression analysis for NUTS2

The output of the following table has no correlation with the actual protests. We can conclude that the model fails to explain the actual protests, because the proportion of variation of the dependent variable explained by all indices is very low (Adjusted R-squared=0.0948).

	Min	**1Q**	**Median**	**3Q Max**
	-181.88	-62.80	-28.41	17.82 1644.02
	Estimate	St. Error	T value	Pr(>ItI)
(Intercept)	-2.174e+02	5.051e+01	-4.304	2.10e-05
unemployment_00_10	2.314e+00	2.923e+00	0.792	0.4291
People in education_00_10	2.705+00	1.720e+00	1.573	0.1165
e_00_10	6.686e-02	2.740e-02	2.440	0.0151
hdu_00_10	-3.850e-02	2.963e-02	-1.299	0.1945
ppi_00_10	1.462e+00	2.958e-01	4.926	1.22e-06
rop_00_10	-6.014e-01	5.958e-01	-1.009	0.3134
Young population 00_10	3.135e-03	7.087e-03	0.442	0.6584
Young not working not studying 00_10	2.946e+00	1.689e+00	1.744	0.0819
Residual standard error: 167.7 on 409 degrees of freedom (2574 observation deleted due the missingness)				
Multiple R-squared: 0.1122, Adjusted R-squared 0.0948				
F-statistics: 6.459 on 8 and 409 DF, p-value: 6.32e-08				

b. Ordinary Least Squares

We computed another database country wide in which we put the year's one under each then, to increment the number of observations and then computed the Ordinary Least Square Regression.

Table 9; the description of regression variables is listed in Annex 1).

Residuals:				
Min	1Q	Median	3Q	Max
-2923.6	-521.5	-71.1	394.5	8991.4
Coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.824e+03	1.118e+03	1.632	0.104325
Unemployment grow	-9.082e+00	4.177e+01	-0.217	0.828104

(difference)					
Unemployment ratio	-3.558e+01	3.430e+01	-1.037	0.300885	
Gdp per capita	-1.530e-01	5.174e-02	-2.956	0.003498	**
Products consumed	3.423e-01	1.005e-01	3.407	0.000798	***
Labour production	-6.861e+00	3.273e+01	-0.210	0.834159	
labour	-6.459e+01	3.665e+01	-1.763	0.079538	.
Gov Finances- lending (+)/net borrowing (-)	-1.125e+02	3.742e+01	-3.006	0.002991	**
investment	2.713e-03	6.634e-04	4.090	6.31e-05	***
gov2_00_10	-1.260e+01	2.820e+01	-0.447	0.655421	
gov4_00_10	-2.242e+00	2.597e+01	-0.086	0.931294	
National Debt	-2.301e+00	8.885e+00	-0.259	0.795975	
Num. of students at 3rd (university)	-2.279e-01	3.047e-01	-0.748	0.455415	
earn_00_10	-2.360e-01	5.611e-02	-4.206	3.95e-05	***
Taxes per earnings	-5.084e+01	1.869e+01	2.721	0.007099	**
Number of policeman	1.150e-02	3.430e-03	3.354	0.000958	***
ext_00_10	3.722e-03	1.618e-03	2.300	0.022482	*
gba_00_10	3.857e+00	9.059e-01	4.257	3.21e-05	***

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

Residual standard error: 1319 on 195 degrees of freedom
(40 observations deleted due to missingness)

Multiple R-squared: 0.5085, Adjusted R-squared: 0.4657
F-statistic: 11.87 on 17 and 195 DF, p-value: < 2.2e-16

All variables are averages for 2000 to 2010 period of study.

There is a 10% statistical significance of gov*_00_10* which is the average of the general government revenue, expenditure and main aggregates indicator of net lending (+)/net borrowing (-).

The situation of having, that could affect them and do not agree on, and so they want to stop it or change it.

c. Ordinary Least Square Method for lag-1 Results

We created a database with the average of protest for each year from 2000 to 2010 and added the variables we explained before in the materials chapter.

We computed Least Squared Regression for lag-1, taking the protests for the last years and using the data from that current year. The result was a correlation of 100% significance with last year protests, meaning that last year increases the probability to other protest to occur.

A possible explanation of this result can be that, social unrest could be considered as a sign where citizens act pacifically in order to rise the voice towards a policy that is been discussed or will be apply for the government (Braha, 2011). To confirm this statement under our study, I suggest identifying the dates concerning a policy change using other Cameo Codes.

We can conclude with the statement that protests have an accumulative property which the influence falls to the past protests or increases, as previous protests encourage other protest to happen (Koopmans, 2004). This statement as been founds in our study with the correlation with last past events.

3.2 First-order characteristics of protest site distribution

The map in Fig. 3 explains the distribution of the protest events during the years 2000 to 2010 all the European countries considered in this study. By visual analysis events are dispersed with some hotspots in the cities or urban areas, as for example Paris and Berlin.

The higher amounts of protests occur in United Kingdom, France, and Germany. A possible explanation can be that most of news sources included in GDELT are in English, which can be disapproved because one of the sources is French. Another explanation that needs deeper analysis is to know the amount of newspapers in each country. Also, it would be important to understand why a country such as Germany, which is usually perceived as not being very active protesting, is essentially protesting more continuously than other countries. Spain for example, has fewer protests more disperse but perhaps the tone is higher.

Figures in this page are the events some of which are overlapped.

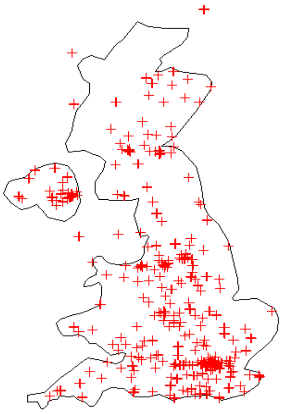


Figure 5a: United Kingdom protest events distribution after removing overlapped points from 2000-2010

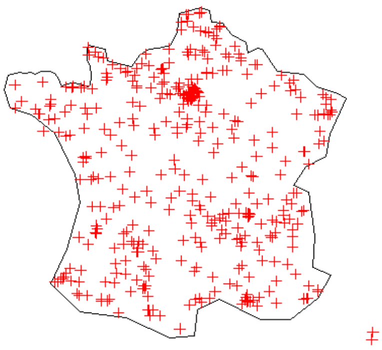


Figure 5b: France protest events distribution after removing overlapped points from 2000-2010

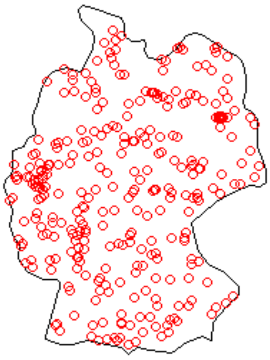


Figure 5c: Germany protest events distribution after removing overlapped points from 2000-2010

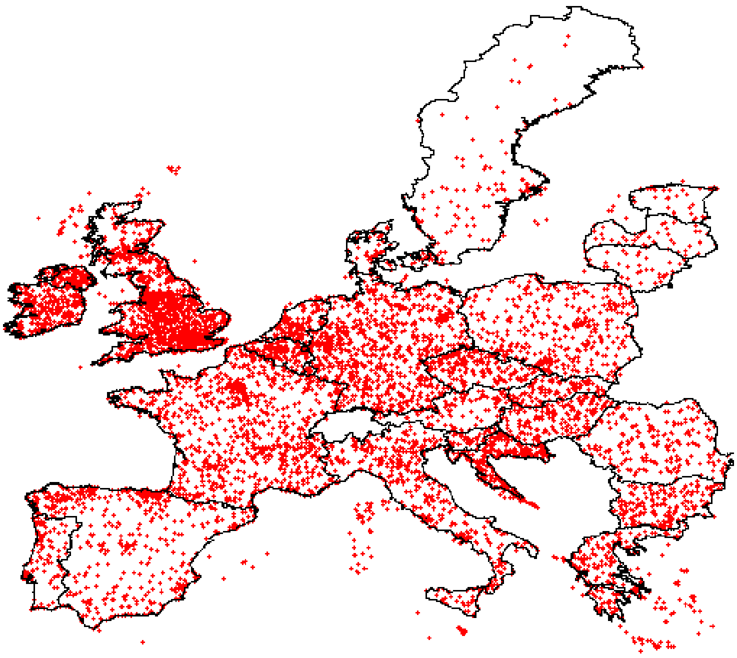


Figure 5d: Europe and protest events distribution after removing overlapped points 2000-2010

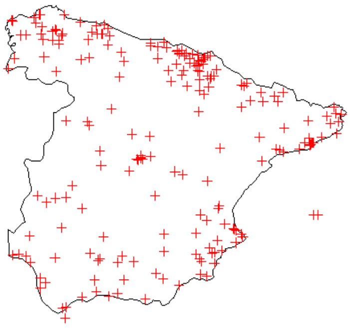


Figure 5e: Spanish protest events distribution after removing overlapped points from 2000-2010

3.3 Testing for Complete Spatial Randomness in protest events distribution.

In order to test the Complete Spatial Randomness we computed map intensity in R tools and commands. The results show a clear random distribution of protests, with some hot spots in some areas as in the central area of Europe enclosing Belgium, France and Germany until Greece and Italy. The dispersion of the protest is probably influenced for the amount of protests in the United Kingdom.

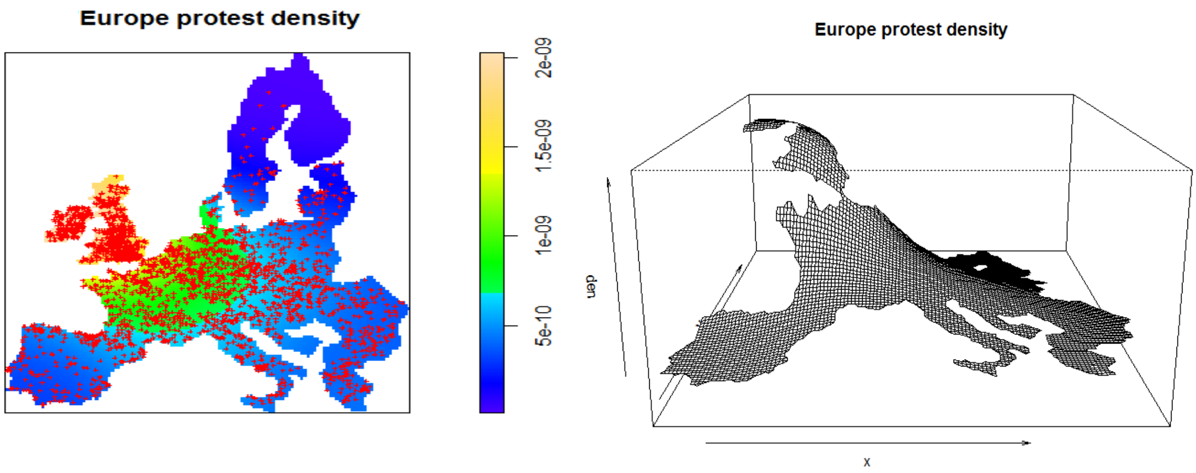


Table 10: Intensity map protest/events in Europe
from 2000 to 2010, as yellow the color as higher the intensity

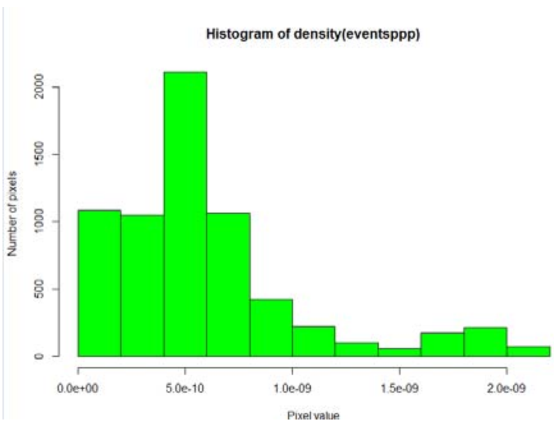


Table 11: Histogram of intensity protest/events in Europe
from 2000 to 2010.

This results gives is a probe of the randomness of the data, in which case we can consider to continue how analysis. After seeing we decided to go for each country as the intensity map for Europe presented an important influence of United Kingdom.

3.4 Inhomogeneous Poisson Point Process

1. Modeling for an Inhomogeneous Poisson Process (IPP).

We plotted the inhomogeneous point data, which state that the trend is mostly until certain location clustered but in the same time is not homogenous, the intensity is not constant or may vary from location to location has we saw in the intensity map before.

Similarly the description for each country (Spain, United Kingdom, France and Germany) in the following figures is presented. The following figures display graph for Inhomogeneous Poisson Process for those countries, taking in consideration all points of the event protests location. Spain and the United Kingdom graphs below, demonstrate the occurrence of overlapping events at the same location 0 m. Both have a similar graph; presenting an extraordinary cluster or overlapped points at 0m and dispersion until 150000m and 80000m respectively. Nonetheless, Spain differs from United Kingdom in two means, intensity and distribution. Spain has a 10% higher intensity of protests than UK, and distribution; there is a smooth stage at 50000m of overlapped points. Both graph present complete dispersion at certain as they diverge from the center and become more distant.

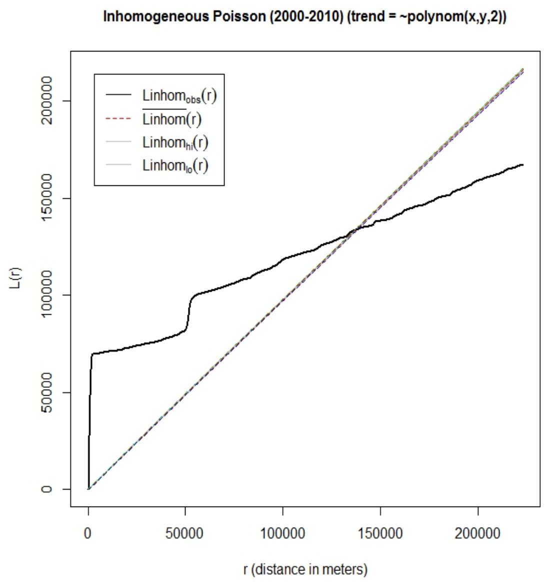


Figure 6: Inhomogeneous Poisson process for all protest events Spain (2000-2010)

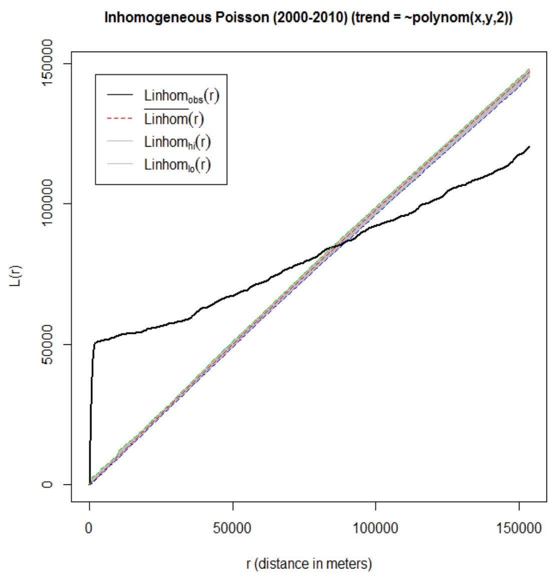


Figure 7: Inhomogeneous Poisson process for all protest events United Kingdom (2000-2010)

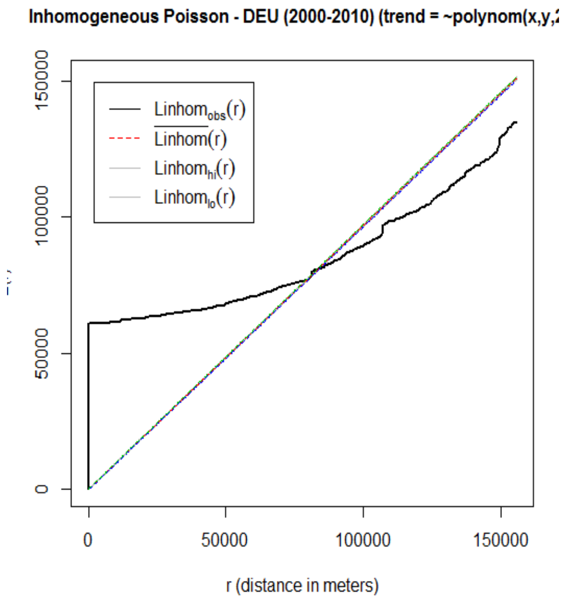


Figure 8: Inhomogeneous Poisson process for all protest events France (2000-2010)

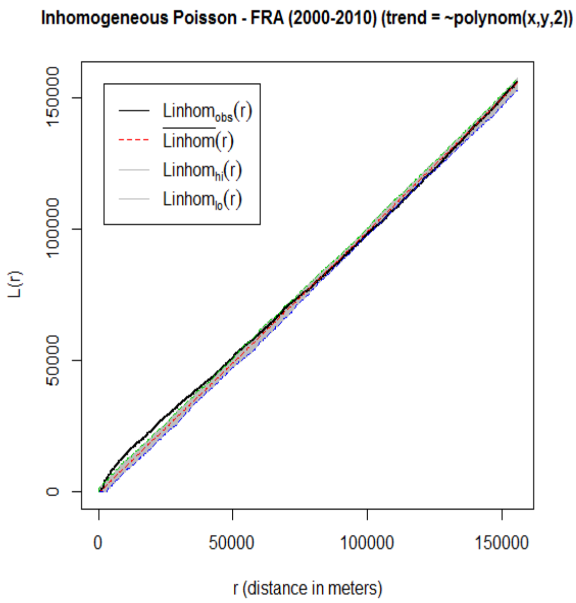


Figure 9: Inhomogeneous Poisson process for all protest events Germany (2000-2010)

Germany tracks similar silhouette to those presented previously, similar to United Kingdom on the dispersion point at 8000m and intensity.

Accordingly to the results, and after the presented results of the empirical line was found outside the envelope, there is a clear potential improvement of the model, and therefore this model is not enough to explain the distribution of the events/protests, as there incites to discover a spatial trend. Necessarily, we decided that the model requires removing the overlapped points and the results are shown in the following Figures. The estimator we use to calculate to estimate inhomogeneous Point Process was k function as explained in other sections.

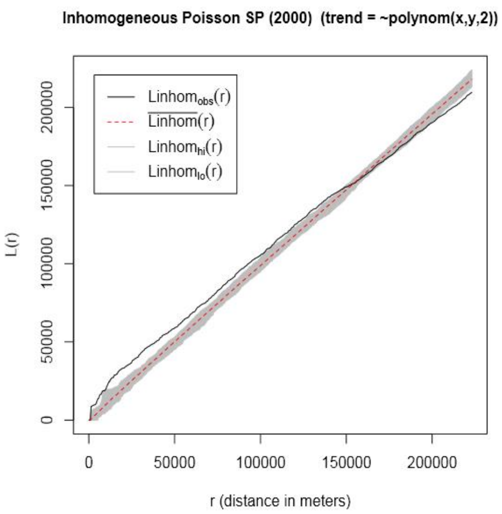


Figure 10 Inhomogeneous Poisson process for all protest events Spain (2000-2010)

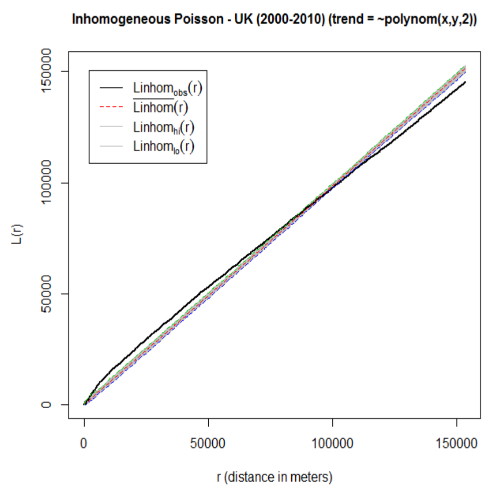


Figure 11: Inhomogeneous Poisson process for all protest events United Kingdom (2000-2010)

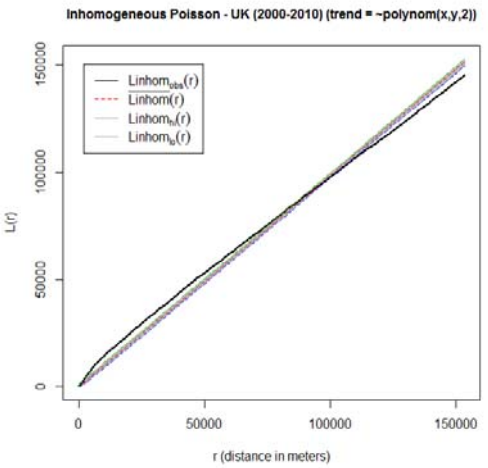


Figure 12: Inhomogeneous Poisson all protest events France (2000-2010)

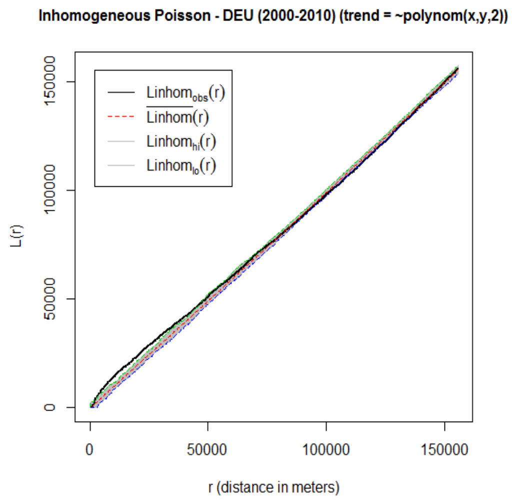


Figure 13: Inhomogeneous Poisson process for all protest events Germany (2000-2010)

3.5 Area Interaction fitting and simulation

In this section we display four models computing the influence of the interaction from the location of protest. All the countries are well fitted and describe a good model as they are inside the simulations but, only Germany is not well defined, that explain the situation on which Germany doesn't have a good prediction. France shows a Germany the data is not clustered at all. France cluster until xxx, inside the envelope, Spain also next to the theoretical line.

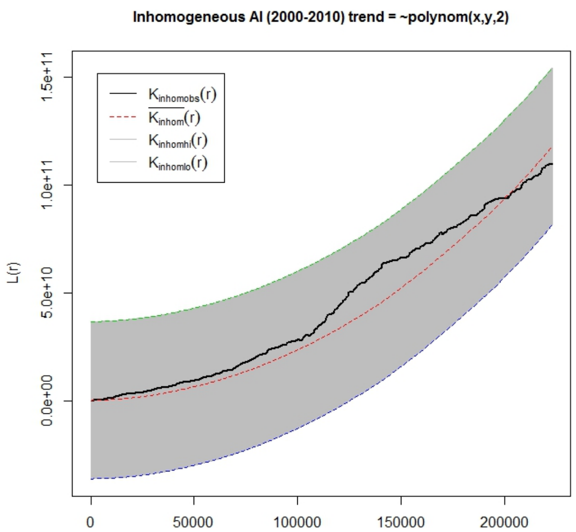


Figure 14: Inhomogeneous Area Interaction process for all protest /events France (2000-2010)

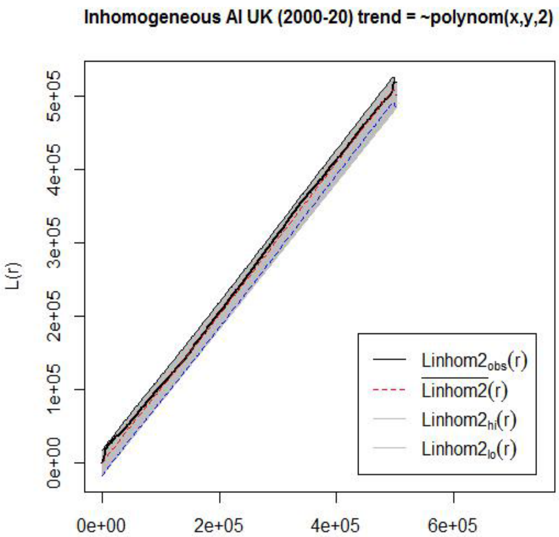


Figure 15: Inhomogeneous Area Interaction process for all protest/events Germany (2000-2010)

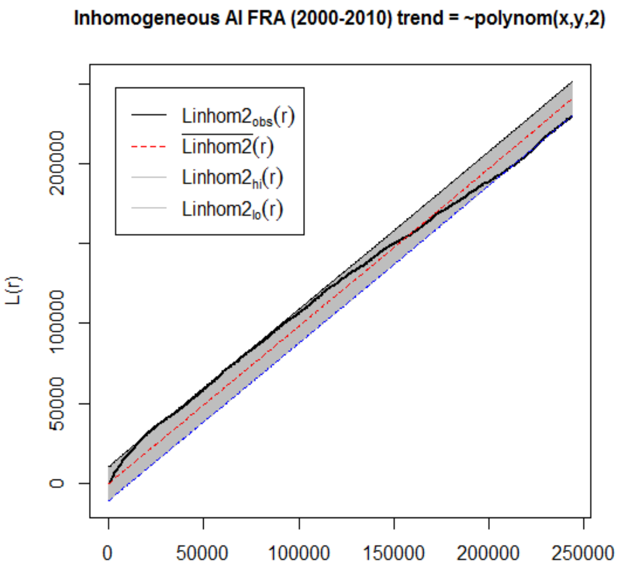


Figure 16: Inhomogeneous Area Interaction process for all protest /events France (2000-2010)

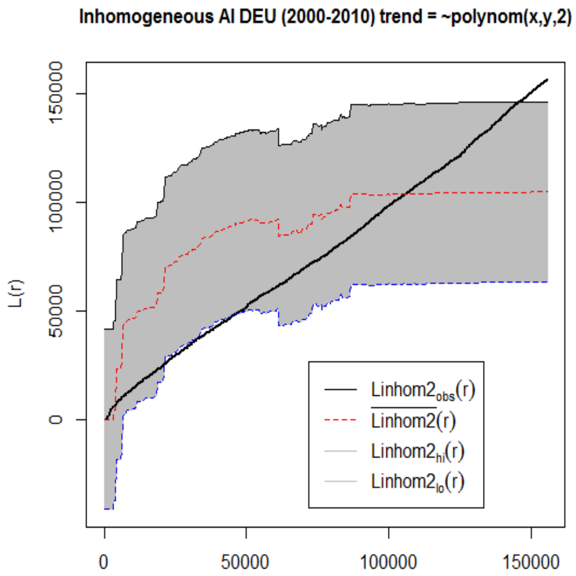
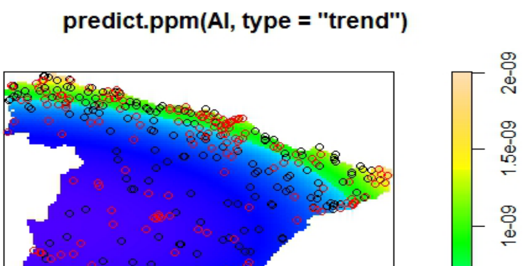
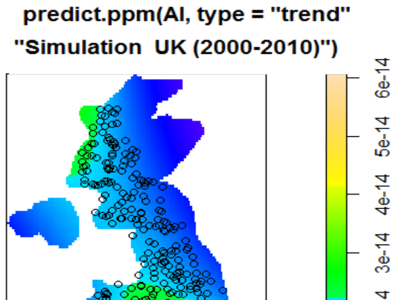


Figure 17: Inhomogeneous Area Interaction process for all protest/events Germany (2000-2010)

We computed the simulation, for Spain, United Kingdom, France and Germany. The legend in the figure explains the intensity of the protests, yellow-white color is for high intensity and dark-blue is for low intensity. The results are an average of protests/events for the period of study. Spain, present the light areas of higher amount of protests in the north and north-east side, as well as in the south. United Kingdom has clear clustering in the south-east in the urban area of London and surroundings as well as in the north-west of Scotland, the intensity is lower between this two areas. Similarly, France has a clear concentration of protests in the border with Belgium and the capital Paris. Germany has the higher amount of protests and all are mostly concentrated in the east part of the country, touching the Netherlands. Spain and Germany have also a hot spot in their capitals, but we can conclude that the amount of protests in these areas is not higher enough which in this case explains other factors and not the location of the capital is creating this amount of events.





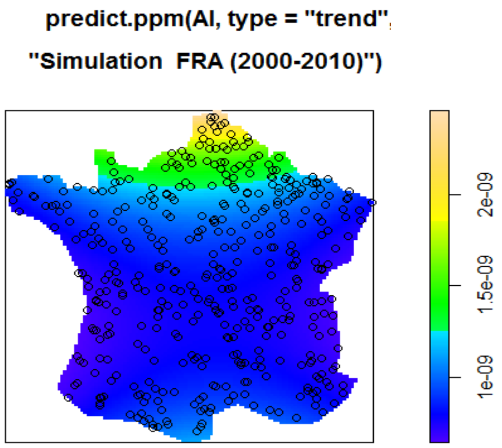


Figure 20: Simulation of protests events in France, based on data from 2000-2010

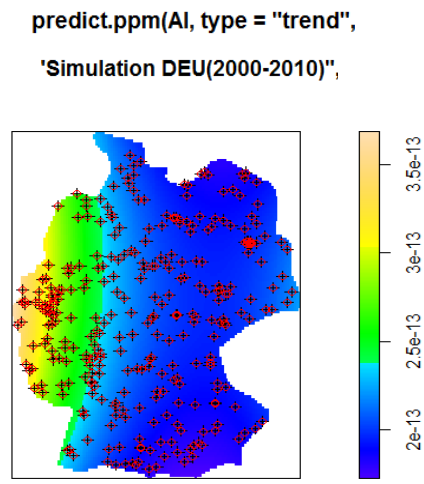
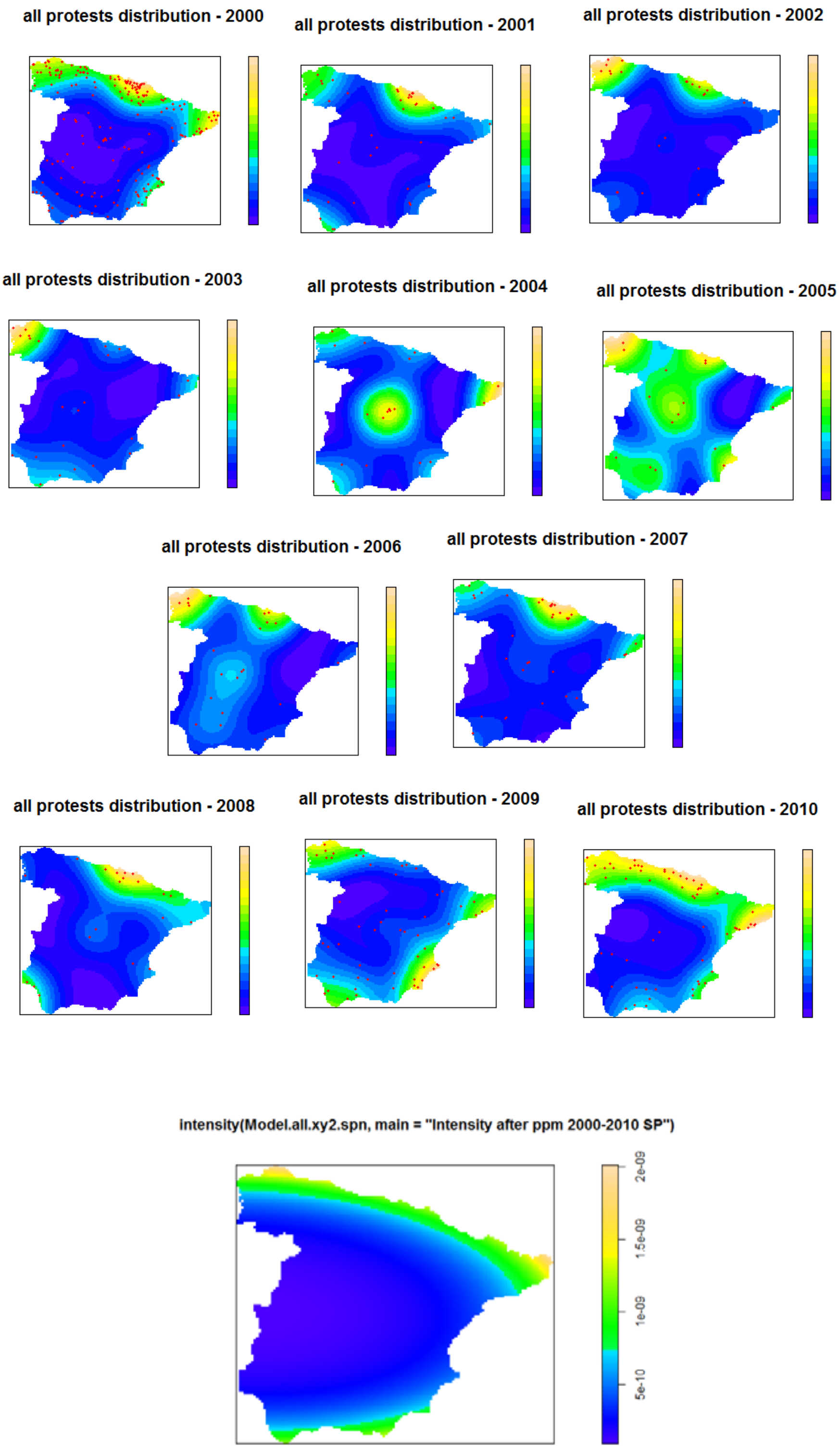


Figure 21: Simulation of protests events in Germany based on data from 2000-2010, in black historical events and red simulated protest.

3.6 Testing for Complete Spatial Randomness in protest events distribution in Spain.

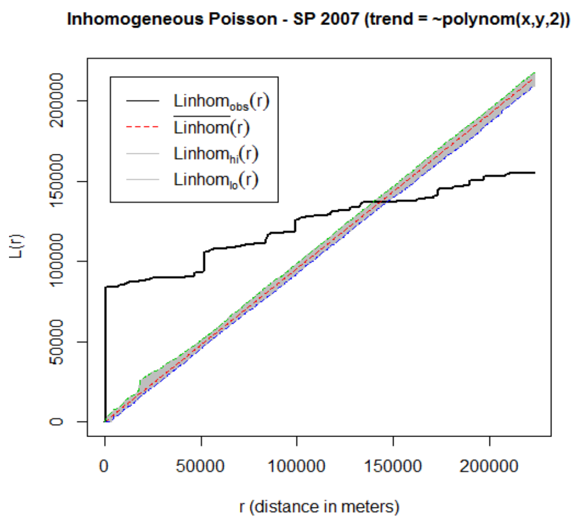
In order to test the Complete Spatial Randomness we computed map intensity in R tools and commands, only for one country, to avoid the influence of United Kingdom, in the case of the Europe. Before starting the analysis, we took each year and evaluate the randomness of the events. As we can see in Figure X, protests vary from time to time, concentrating in mostly in the peripheries, as Galicia, Bask Country and Catalonia and the capital, Madrid. All this areas have the same features, high demographic density and high income or GDP.



Figures 22: Intensity maps for each year from 2000 to 2010 for Spain.
The red points are the protests locations and the last map is represents the average of protests from the overall period.

In order to give more importance to the primary idea before mentioned, we computed the Inhomogeneous Poisson process with all the data protests and we saw clearly that protest are overlapped. The arguments to define Inhomogeneity are by understanding the output of the Figure 25. In that Figure we see there is a big amount of points in the same location, defined by the straight line at 0m and abrupt trend of steps along the prolongation of the line. A possible justification is the amount of missing points or insufficient number of points, where there are potential locations but there is not enough data, that is why the escalation of the black line.

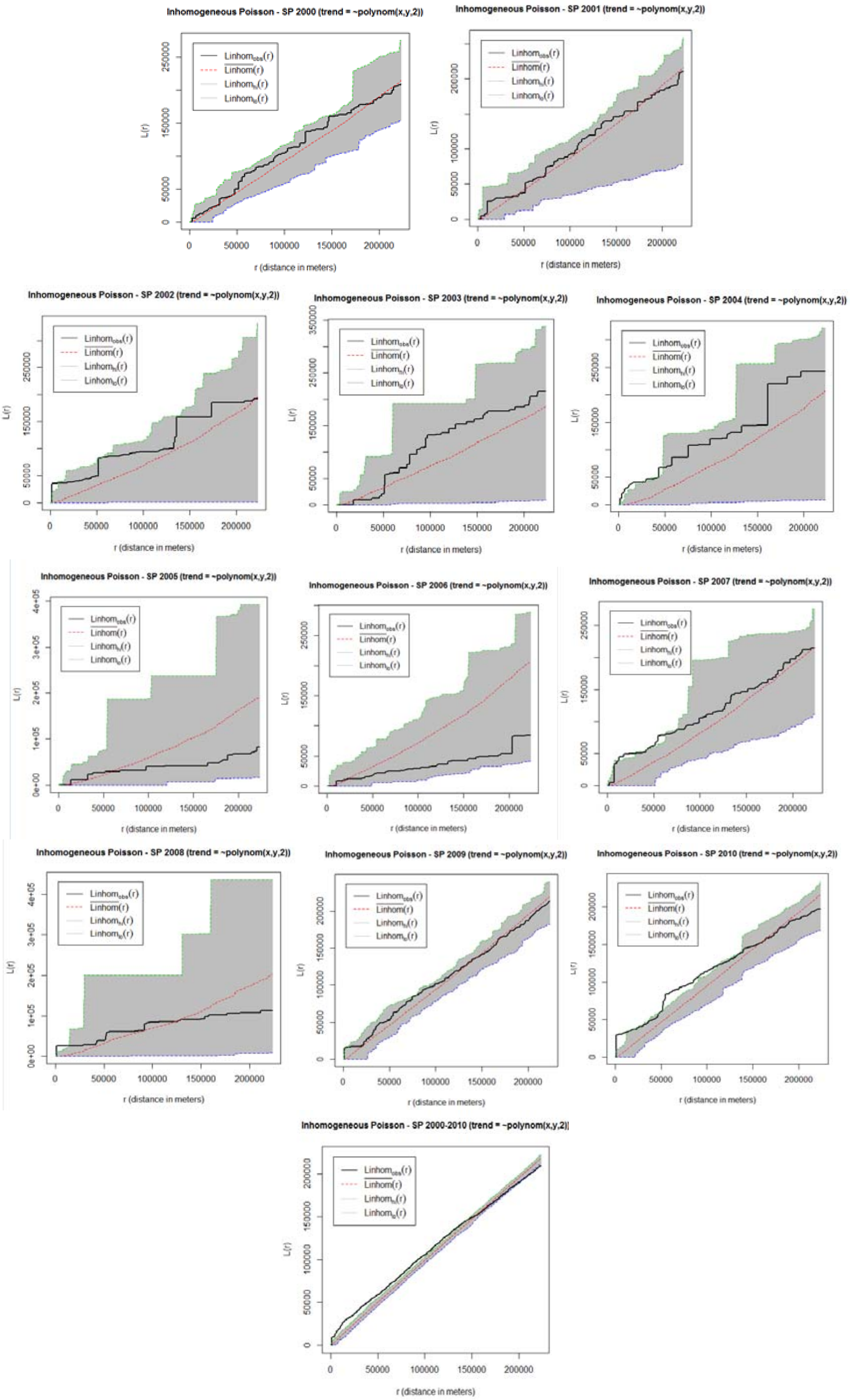
We can explain the model by observing at the trend of the empirical line. If it is under the theoretical line is red color, means that there is no interaction and if it is over it, means that the data is clustered.



Figures 23: Sample of Inhomogeneous Poisson Analysis for 2007 in Spain,
Whithout removing overlapped points.

The reason before mentioned, encourages to remove the overlapped points and compute again the identical analysis; the output is visualized the following Figures for each year, from 2000 to 2010.

We compute inhomogeneous Poisson for each year, to discover if there is a spatial tends in the events.

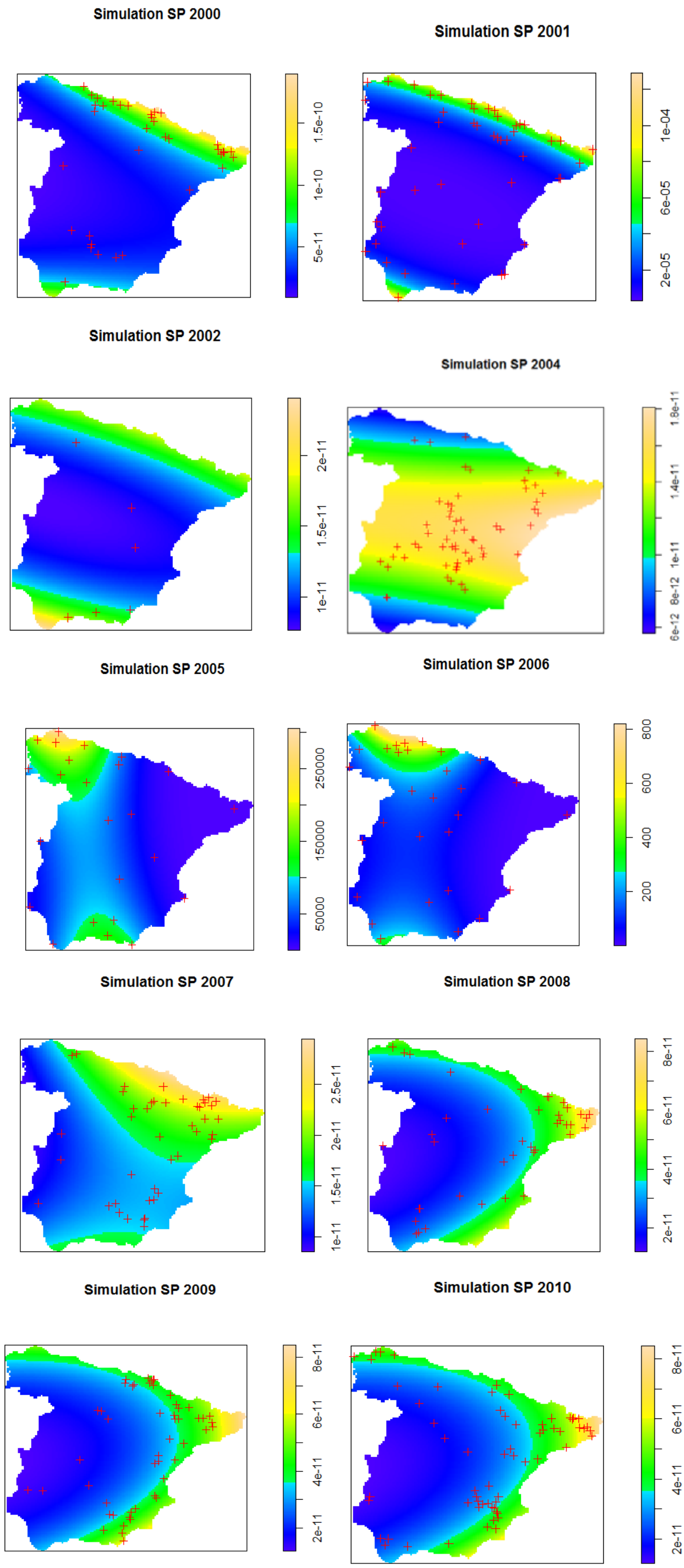


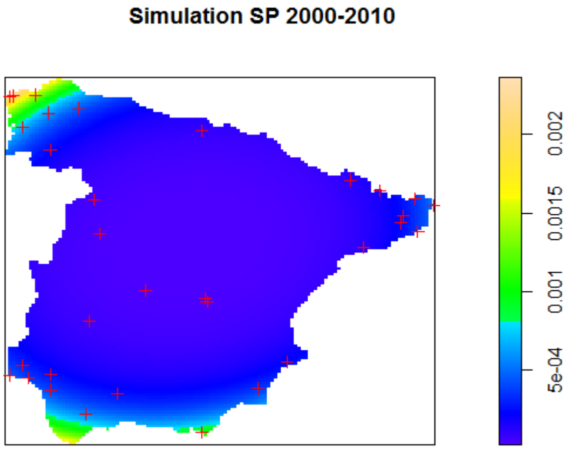
Figures 24: Inhomogeneous Poisson Analysis for each year from 2000 to 2010 for Spain.

The analysis of Inhomogeneous Poisson varies for every year and all of them are inside the envelope, which concludes that the model fits and therefore the location of the protest is not depending on the interaction of other events. In despite of, we see some years as 2000 and 2001 indicating a good fitted model, where we find both lines close to each other. Other models indicate dispersion of the envelopes towards a longer distance, meaning that there are not enough points in those areas. Also, we found in the last years as 2009 and 2010 smaller envelopes and both lines again closer. In overall, the model computed for all years together, from 2000 to 2010, show how the empirical line is outside the envelope and there is clear defined cluster until 150km.

Mostly all the models are fitting inside the envelopes which mean that the model computed has a 100% of fidelity in all cases. Spain has similar figures for all years, in which all graphs show empirical lines and theoretical lines inside the envelope computed. The main variance is the trend of the empirical line suffers abrupt variances and doesn't show a smooth tendency, probably owed to missing data.

Regarding the simulation for Spain, we computed, all years from 2000 to 2010 to see the distribution of the protests in time in space. We observe different patterns for each year, which displays a pattern of distribution that would need further research in order to give a fitting explanation. This would be a start for a further research that could be applicable for each country of study.
We could try to explain the pattern we obtained, in vast lines. Spanish crisis was in progress in the beginning of 2000, which regarding the graph we see a north and south higher distribution in the north and south. Spain has a characteristic distribution of protests, wealth and population density. The north of Spain has the highest income and GDP, but also the capital, Madrid which is in the center. Regarding population density and the South has high rate of youth people as the capital, but it has one of the higher rates of unemployment, which could explain the protest intensity. In the other hand, when a very important protest happens, usually occur in capital or big cities. It would be brave, and in the same time humble because it needs references and further study, to confirm the possibility to term the intensity found in the North-East (Catalonia) in 2008 to 2010 is a starting of the current *sectionsim* movement in Catalonia, due the austerity measures and a bad management of history.


Simulation SP 2000
1.5e-10
1e-10
5e-11
Simulation SP 2001
1e-04
6e-05
2e-05
Simulation SP 2002
2e-11
1.5e-11
1e-11
Simulation SP 2004
1.8e-11
1.4e-11
1e-11
8e-12
6e-12
Simulation SP 2005
250000
150000
50000
Simulation SP 2006
800
600
400
200
Simulation SP 2007
2.5e-11
2e-11
1.5e-11
1e-11
Simulation SP 2008
8e-11
6e-11
4e-11
2e-11
Simulation SP 2009
8e-11
6e-11
4e-11
2e-11
Simulation SP 2010
8e-11
6e-11
4e-11
2e-11



Figures 25: Simulation maps for each year from 2000 to 2010 for Spain.
The red points are the location of the protests.

CHAPTER 4: DISCUSSION, RECOMMENDATIONS AND CONCLUSIONS

4.1 Results overview and limitations of the study

The main objective of this study was to evaluate how economic, education and demographic variables, as unemployment could affect to increase the level of protests in Europe. We wanted to compute a simulation for Europe using protests events

In this project we have computed different types of analysis. Our first analysis was the lattice approach, in which we have worked in different resolutions. In order to asses a good resolution of analysis, to have a significant output, we started selecting the variables from Eurostat in a base NUTS2 regions database.

After computing this analysis, we conclude choosing a NUTS2 region for this type of analysis is not the best option. The reasons why this is not the best option is because the variables are disorganized and there is a high degree of incompleteness, which makes difficult to compare southern European countries with northern countries. One of the reasons for this incompleteness is because a wide range of NUTS2 in Europe is empty or as the case of Greece can be falsified [1] variables are empty for certain years. Other countries as Italy and Spain do not have the same procedure and publishes variables in a different region definition, as are geographical divisions (e.g north, south, east, west) which create any comparison complicated.

Based on the argumentation before, we decided to repeat the analyses in country wise bases and founded completeness, thus this is the best practice. Additionally a good practice was to implement the variables selected from Eurostat in a country wise as a long table, where we could observe each year under another.

Similarly, in the GDELT database, we face incompleteness, in situations where observations within Protest events 145, ActionGeo_CountryCode was empty. In this situation, the GDETL database is automatize and sets as a location the centroid of the county mentioned, having then another motive to use a country wise approach.

The GDELT data base suffers of noise, for example, in Inhomogeneous Poisson

Process graphs for all data presents a distortion of 50000m in Spain. We consider it as an inaccuracy in the geo-location tool as it is most likely to not define the boundary of the urban areas.

Least-square regression analysis is the appropriate technique use in a model where the independent variable *x (protests),* is measured without errors and the dependent variable *y* is random, for this reason we consider that the GDELT data base is not good enough to compute quantitative analysis but as it is constant development , we consider that this is a punctual opinion.

4.2 Conclusions:

During the lattice approach we found difficult to create a model that will explain the causes of the appearance of protest, as we are dealing with very different type of countries. After this analysis we considered that each country should be studied separately for better accuracy.

On the other hand, we faced some difficulties to select the amount of historical data (time scale) we need for a simulation. The issues regarding the efficiency over the forecasting zones (scale) have been discussed over the literature. Important questions as how larger conflict datasets are build, thus the sample size and the methods of estimation are still unclear.

The complexity remains here; the literature doesn't clarify how to quantify and qualify the scale and temporal dimension of the historical data. Building a new database is an important task as the approach can influence the result. Moreover, the selection of regions and aggregation facts/observations (other events) are in current treatment in the branch of international studies and political forecasting (Brandt et al., 2013).

The objective of the study is not forecasting events as in social and political science, even the popularity of applying quantitative studies it still not trusted due the unpredictability of social behaviour events. Instead an objective is to achieve a further understanding of how people participate in collective protest because of long-standing social, economic, and political stress, and because others have recently done so (Braha, 2012).

Forecasting must considers until what extend past tendencies will continue in a future. In social behaviour or political forecasting, as social unrest events, the linearity of the trend can change abruptly. To evaluate the reasons and when these change of dynamics or the critical points are a complex task that a system still cannot predict (Doran. 2002). Even tough, the developments in this field are based on knowledge of past behavior and acknowledge to the last technology development it will be better-quality.

Regarding the decision of performance of Least Squared Regression Analysis, total government expenditure and revenue, even the correlation, might be off interest, because in any economic crisis, governments are force, meaning no-choice to bring their growing deficits under control, thus these leads countries to narrow strategies towards policies that favor the macro economy.

It is assumed that relationship between different indexes affect protests, but taking only labor market status and/or use of military violence as descriptive variables is courageous. In fact in our study it was negative or not enough important.

Crime and poor societies are correlated with protests, studies point out that both can intensification together, as protests can convert to serious disorders depending on the government reaction and crimes can claim ideological meanings (Turner, 1969). Other weakness as lack of democracy in government institutions, international pressure and difference between elite population and the working class, ethnicity, all of them can be potentially describing the creation of revolutions (Goldstone, 2001; Braha, 2012).

In other words, it is assumed that social, economic, and political tensions accumulates through the country as an inter-state conflict, before they lead to sudden explosion of unrest (Dowe et al. 2001, Glurr, 1970, Braha, 2012)

On the other side, it's being hard to include governance indices for institutions, authority characteristics of states and country regime trends, as those variables are binominals (0,1) and consider a big scale. An example is the Polity IV Project database form the Center for Systematic Peace and Societal-System Research Inc, are world rankings making difficult to register minor changes in e.g in European countries because they are not noticeable compared with African countries.

Indexes of administration variables (number of government employees, number of policemen, income) are of difficult access, Europe doesn't have a database with such *Institution* variables regarding the institutions, with which we could consider to evaluate in deep the political structures or some of the accessible ones are indexes with no access to the raw data.

The simulations driven in this analysis take as a final assumption that area interaction method is good enough to estimate the location of future protest/events. The result presents an average 250 km which is more likely to belong to the cities or urban areas. We can assume that urban areas tend to poses higher difficulties respect to services, poverty, and human well-being which are indivisibly linked with social unrest (Gurr, 1970; Dowe, 2001 and Bai, 2005).

This conclusion appears to be intuitive and thus there is need to increase the resolution to the world area or bigger regional areas, in order to obtain more efficient results, which is to get a general overview of real-time social sciences in the earth.

GDELT database is a very interesting initiative, is the first news database automatized, for all the world and daily and 24h free available. After this analysis, we conclude that GDELT database should be [i]used as a tool to define worldwide trends or bigger areas instead of an accurate precision in regions, because data is chaotic regarding geo locations. For instance, some events have empty values and thus if the name is explicit is given the centroid of the country, which can confuse results.

Area interaction explained that point to point location is a good model to simulate future protest/events. Even though, it is still unclear the cause or explanation of the protests and its location, perhaps in Spain, where the fact that the capital is in the centroid of the country, influence to smooth the results. There is, nonetheless an higher intensity in the north part of Spain. An affirmative explanation could be a relation with gdp and income, *which carries high qualify inhabitants who we could assume to be more conscious and higher civil engaged culture*. The higher amount of protests, result on the same areas are still in dispute after the Civil War, as those areas have a historical culture of grievance towards the government, the amount of protest could be related to that subject. The south is also considerate in Spain as a broad visited university area with higher amount of young population.

Richard Rosecrance concludes that through time there is a tendency for international instability to be associated with domestic insecurity of elites (Rosecrance, 193, p.304)

4.3 Further Analysis

Relative to the type of data and the processes to study political science, international conflict data in general have unusual characteristics. They are based on thousands of and whether the amount of dyads should be included or dismissed, or if only actors in the conflict should be considered, all nations, or only some group is in the literature no clear.

Regarding the window of study, using Europe as a unique region is a challenge, the lack of data for each country, emptiness or the use of different nomination of its geographic regions made it complicated, even the effort to harmonize and join statistics in EU all data is acquired for each country. The variables we selected to build a model are good starting to understand the environment of each country, but for a further analysis there is a need for further selection on economic variables.

After this analysis, we conclude that in complicated cases of studies as conflict forecasting/politics, should include countries with similar features together, therefore the analysis should be deepen country by country. In despite of, each country is different, and so each country should be studied separately, there are facts that social unrest contagion is governed by the same mechanisms to individual countries and geographic regions (Stanley, 2000).

We suggest further analysis within GDELT database functionalities. Find correlation between Action and Actors, to improve the accuracy of the observation we want to describe as the dynamics of protests occur as a response of sudden and punctual events (Dowe et al. 2001), a suggestion would be to combine news coded as *appeal* and *threat* (CAMEO Code 02 and 13) could introduced an interesting descriptive factors (Turner, 1969).

We could select events by the tone in further studies, in the following graph we show the events that have a tone of more than 10 mentions on the sources; this could help us to define the most important events.

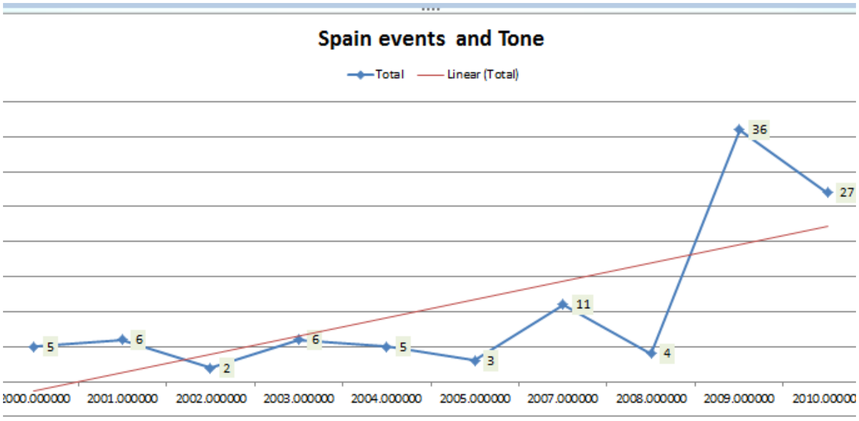


Table 13: Spain event numbers and tone

Spatial analysis can be improved, despite the computed analysis in this project has been satisfactory enough. We have observed that protests are close to the most populated places, which sometimes rely close to boundaries (e.g Germany & France) we accordingly we conclude that a further analysis regarding contamination of protest from country to country would be interesting.

To include different forms of social or communication networks influence protests (González-Bailón, 2011) and has not been included in this study. This will imply understanding a new technological factor which influences a faster spreading of civil unrest and a new way of organization (Braha, 2012).

The new database it should be considered as a prodigious innovation due to the increase of sources and observations recorded, it is the first database to include protests events as an observation and it still on constant development (Schrodt, 2013) for those interested in what it has been called a *social science of earth observation studies*.

Diggle, P. J. (2003). Statistical analysis of spatial point patterns. Arnold, London

Baddeley, Adrian. Analysing spatial point patterns in R. Technical report, CSIRO, 2010. Version 4. Available at www. csiro. au/resources/pf16h. html, 2008.

Baddeley, Adrian, and Rolf Turner. "Spatstat: an R package for analyzing spatial point patterns." *Journal of statistical software* “12.6 (2005): 1-42.

Baddeley, A., Diggle, P.J., Hardegen, A., Lawrence, T., Milne, R.K. and Nair, G. (2014) On tests of spatial pattern based on simulation envelopes. *Ecological Monographs*, to appear.

Bannon, Ian, and Paul Collier, eds. *Natural Resources and Violent Conflict: Options and Actions.* World Bank-free PDF, 2003.

Beck, Nathaniel, Gary King, and Langche Zeng. "Improving quantitative studies of international conflict: A conjecture." *American Political Science Review*(2000): 21-35.

Braha, Dan. "Global civil unrest: contagion, self-organization, and prediction."PloS one 7.10 (2012): e48596.

McGranahan, Gordon, and Xuemei Bai. "Ecosystems and Human Well-Being: Volume 1: Current State and Trends: Urban Systems Ecosystems and Human Well-Being: Volume 1: Current State and Trends: Urban Systems."

Brandt, Patrick T., John R. Freeman, and Philip A. Schrodt. "Real time, time series forecasting of inter-and intra-state political conflict." *Conflict Management and Peace Science* 28.1 (2011): 41-64.

Brandt, Patrick T., John R. Freeman, Tse-min Lin and Phillip A. Schrodt. 2012. “Evaluating
Forecasts of Political Conflict Dynamics.” Manuscript.

Brandt, Patrick T., Michael Colaresi and John R. Freeman. 2008. “The Dynamics of Reci-
procity, Accountability, and Credibility.” Journal of Conflict Resolution 52:343–374.

Brandt, Patrick T., et al. "Forecasting Conflict in the Cross-Straits: Long Term and Short Term Predictions." (2013).

Chambers, J. M. (1992) *Linear models.* Chapter 4 of *Statistical Models in S* eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Cincotta, Richard P., Robert Engelman, and Daniele Anastasion. *The security demographic: Population and civil conflict after the Cold War*. POPULATION ACTION INTERNATIONAL-SECURITY DEMOGRAPHIC WASHINGTON DC, 2003.

Crozier, Michel, Samuel P. Huntington, and Joji Watanuki. The crisis of democracy. Vol. 70. New York: New York University Press, 1975.

Doran, Charles F. "Why forecasts fail: the limits and potential of forecasting in international relations and economics." *International Studies Review* 1.2 (1999): 11-41.

Dowe D, Haupt H-G, Langewiesche D, Sperber J, editors (2001) Europe in 1848: revolution and reform. New York: Berghahn Books. 976 p.

Goldstone, Jack A., Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik,
Monty G. Marshall, Jay Ulfelder and Mark Woodward. 2010. "A Global Model for Fore-
casting Political Instability." American Journal of Political Science 54(1) :190-208

Goldstone, Jack A., Toward a Fourth Generation of Revolutionary Theory (January 1, 2001). Annual Review of Political Science, Vol. 4, pp. 139-187, 2001. Available at SSRN: http://ssrn.com/abstract=1531902 or http://dx.doi.org/10.2139/ssrn.1531902

González-Bailón, Sandra, et al. "The dynamics of protest recruitment through an online network." *Scientific reports* 1 (2011).

Cramer, Christopher. "Unemployment and participation in violence." (2011).

Gurr TR (1970) Why men rebel. Princeton: Princeton University Press. 438 p.

Hegre, H˚avard, Joakim Karlsen, H˚avard Mokleiv Nyg˚ard, H˚avard Strand and Henrik Urdal.
2012. "Predicting Armed Conflict, 2010–2050." International Studies Quarterly 1:1–21

Kornhauser, Marjorie E. "Legitimacy and the right of revolution: The role of tax protests and anti-tax rhetoric in America." *Buff. L.* Rev. 50 (2002): 819.

Koopmans, Ruud. "Protest in time and space: The evolution of waves of contention." The Blackwell companion to social movements (2004): 21.

Leetaru, Kalev and Schrodt, Philip. (2013). GDELT: Global Data on Events, Language, and Tone, 1979-2012. *International Studies Association Annual Conference*, April 2013. San Diego, CA.

Ripley, B. D. (1977). "Modelling Spatial Patterns." J. Royal Statist. Soc. Ser. 39: 172-212.

Rucht, Dieter. "Chapitre 2. The EU as a Target of Political Mobilisation: Is there a Europeanisation of Conflict?" *Académique* (2002): 163-194.

Scott, D. W. (1992) *Multivariate Density Estimation. Theory, Practice and Visualization*. New York: Wiley.

Stråth, Bo. "A European Identity To the Historical Limits of a Concept."*European Journal of Social Theory* 5.4 (2002): 387-401.

So, Alvin Y. "The Global Capitalist Crisis and." *Beyond the Global Capitalist Crisis: The World Economy in Transition* (2012): 123.

Stanley, H. E., et al. "Scale invariance and universality: organizing principles in complex systems." *Physica A: Statistical Mechanics and its Applications* 281.1 (2000): 60-68.

Tanter R (1966) Dimensions of conflict behavior within and between nations, 1958–60. J Conflict Resolut 10: 41–64

Turner, Ralph H. "The public perception of protest." *American Sociological Review* (1969): 815-831.

T. Kariya and H. Kurata (2004). *Generalized Least Squares*. Wiley.

Thomassen, Jacques JA. "Economic crisis, dissatisfaction and protest." (1990): 103-134.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. "The perils of policy by p-value: Predicting civil conflicts." *Journal of Peace Research* 47.4 (2010): 363-375.

Weidmann, Nils B., and Michael D. Ward. "Predicting conflict in space and time." Journal of Conflict Resolution 54.6 (2010): 883-901.

Yonamine, James E. *A nuanced study of political conflict using the Global Datasets of Events Location and Tone (GDELT) dataset*. Diss.THE PENNSYLVANIA STATE UNIVERSITY, 2014.

Yonamine, Using Jay. 2013. “Predicting Future Levels of Violence in Afghanistan GDELT.” http://jayyonamine.com/wp-content/uploads/2013/04/ Predicting-Future-Levels-of-Violence-in-Afghanistan-Districts-using-GDELT. pdf.

Zunes, Stephen. "The role of non-violent action in the downfall of apartheid." *The Journal of Modern African Studies* 37.1 (1999): 137-169.

Appendix 1: Download and normalization of GDELT database in R

```
#.....................................................................
....
# Title: Download_Data_SP.R
#
# Author: Adeva, Elisabet
#
# Date: January 26th 2014
#
## Topic: Assessing economic variables and simulation of social unrest in
Europe.

# ************************* RELOAD DATA AND START AGAIN
****************************
setwd("/home/elisabet/Documents/eliLINUXmint/Documents/Thesis/GDELT.1979-
2012.reduced/GDELT.1979-2012.reduced")
load ("subset19792011_sp14.dat")

recodes <- c(1, 4, 6:12)
for (i in recodes) {
  dat.sp [, i] <- as.numeric (paste (dat.sp [, i]))}

# Map of the country:
plot (dat.sp$ActionGeo_Lon, dat.sp$ActionGeo_Lat, pch=1, col="red",
     # xlim=c(07,20), ylim=c(37,48))  #thi is Italy

# Map of whole world:
plot (dat.sp$ActionGeo_Lon, dat.sp$ActionGeo_Lat, pch=1, col="red")
library(maps)
map (add=TRUE)

years <- sort (unique (floor (dat.sp$Day / 10000)))
nyears=length(years)
EV <- sort (unique(dat.sp$EventCode)) # the codes
ncodes <- length (EV)

# Aggregate into all events per year:
num.events.sp <- array (NA, dim=c(nyears, ncodes))
for (i in 1: nyears){
  for (j in 1:ncodes) {
    indx <- which (floor (dat.sp$Day / 10000) == years [i] &
dat.sp$EventCode == EV [j])
    num.events.sp [i, j] <- length (indx)
  }
}
num.events.sp <- data.frame (num.events.sp)
names (num.events.sp) <- paste (EV)
rownames (num.events.sp) <- paste (years)

# Aggregate into only some events 141 and 145 per year:
num.events.sp <- array (NA, dim=c(nyears, ncodes))
for (i in 1: nyears){
  for (j in 1:ncodes) {
    indx <- which (floor (dat.sp$Day / 10000) == years [i] &
dat.sp$EventCode == EV [2,6])
    num.events.sp [i, j] <- length (indx)
  }
}
num.events.sp <- data.frame (num.events.sp)
names (num.events.sp) <- paste (EV)
rownames (num.events.sp) <- paste (years)


# ********************* AGGREGATION INTO MONTHLY COUNTS
***************************
years <- sort (unique (floor (dat.sp$Day / 10000)))
nyears=length(years)
nmonths <- nyears * 12
EV <- sort (unique(dat.sp$EventCode)) # the codes
ncodes <- length (EV)
```

```
# Aggregate into events per months:
num.events.sp.mo <- data.frame (array (NA, dim=c(nmonths, ncodes + 3)))
names (num.events.sp.mo) <- c ("year", "month", "indx", paste (EV))
for (i in 1: nyears){
  for (j in 1:12) {
    # make a list of the months from the dates, which are yyyymmdd. First
remove
    # the year by dividing by 10,000 and getting the remainder (= "%%"):
    month <- dat.sp$Day %% 10000
    # That is then mmdd, so then remove the day by:
    month <- floor (month / 100)
    tindx <- (i - 1) * 12 + j
    num.events.sp.mo$year [tindx] <- floor (dat.sp$Day / 10000)
    num.events.sp.mo$month [tindx] <- j
    num.events.sp.mo$indx [tindx] <- tindx
    for (k in 1:ncodes) {
      indx <- which (floor (dat.sp$Day / 10000) == years [i] & month == j &
                       dat.sp$EventCode == EV [k])
      num.events.sp.mo [tindx, k + 3] <- length (indx)
    }
  }
  rownames (num.events.sp.mo) <- paste (years)
  num.events.sp.mo


  # Load and pre-process normalisation data
  library (GDELTtools)
  data (NormEventCountsData)
  norm.yc <- NormEventCountsData$yearly.country
  # norm.yc is then the normalisation dat.spa for each year and each country
  norm.yc <- norm.yc [norm.yc$country == "IT",] # NATO country code no like
in norm.yc that ISO Codes are used
  indx <- order (norm.yc$year)
  norm.yc <- norm.yc [indx,]
  EVt <- norm.yc$total

  # annual unemployment country
  setwd ("/home/elisabet/Documents/Thesis/Countries/EU/with_crisis/SPN")
  unempl.sp <- read.csv (file = "SPN.csv", header = TRUE)
  unempl.sp <- data.frame (unempl.sp)
  names (unempl.sp) <- c ("Year", "unempl")

  #monthly unemployment country
  setwd("/media/elisabet/OS/Users/Elisabet/Desktop/Eli_Data/monthly")
  unempl.sp.mo <- read.csv (file = "unemply_.csv", header = FALSE)
  unempl.sp.mo <- data.frame (unempl.sp.mo)
  names(unempl.sp.mo) <- ("","unempl")

  #annual unemployment nuts2
  setwd
("/home/elisabet/Documents/eliLINUXmint/Documents/Thesis/Countries/EU/SPN/NU
TS2")
  unempl.sp.nuts <- read.csv(file = "sp_NUTS2_unemp.csv", head = TRUE)
  unempl.sp.nuts <- data.frame (unempl.sp.nuts)
  names(unempl.sp.nuts) <- c ("","")


  # Convert all three data sets (GDELT = dat.sp, normalisation = norm.yc,
and
  # unemployment = unempl ) to have same number of rows for the same years.
  years <- 2000:2011 #change the years related to the unemployment data
  dat.sp$year <- floor (dat.sp$Day / 10000)
  indx.dat.sp <- which (as.numeric (rownames (num.events.sp)) %in% years)
  indx.norm <- which (norm.yc$year %in% years)
  indx.unempl <- which (unempl.sp$Year %in% years)

  # Then reduce all data just to the common years:
  num.events.sp <- num.events.sp [indx.dat.sp, ]
  norm.yc <- norm.yc [indx.norm, ]
  unempl.sp <- unempl.sp [indx.unempl, ]
```

```
  # Then normalise the numbers of events,
  #thus all the norm.unempl and ita.events have the same amount of
years/rows
  num.events.sp <- num.events.sp / norm.yc$total
  num.events.sp$total.events <- rowSums (num.events.sp)



# # ***************************** PLOTTING EACH
EVENT****************************
  # Time series:
  ncols <- ncol (num.events.sp) - 1
  ylims <- range (num.events.sp [1:ncols])
  # First plot time series of numbers of events
  par (mar=c(2,2,0.5,0.5))
  plot (years, num.events.sp [,1], "l", col=1, ylab="Number of Events in
SP", ylim=ylims)
  for (i in 1:ncols) {
    lines (years, num.events.sp [,i], col=i)
    points (years, num.events.sp [, i], col=i, pch=19)
  }
  legend ("topleft", names (num.events.sp) [1:ncols], lwd=1, col=1:ncols,
bty="n")

  # Then related unemployment to numbers of events. There are 10 types of
events, so
  # arrange these in a 3-by-4 grid
  par (mfrow=c(3,4))
  for (i in 1:length (EV)) { # length (EV) = number of categories of events
    plot (unempl.sp$unempl, num.events.sp [,i], pch=1, col=i,
          xlab="Unemployment", ylab="Number of events")
    # Then calculate a linear regression:
    mod <- lm (num.events.sp [,i] ~ unempl.sp$unempl)
    mod.predict <- predict.lm (mod)
    lines (unempl.sp$unempl, mod.predict, col=i)
    r2 <- formatC (summary (mod)$r.squared, format="f", digits=4)
    title (main=paste ("Event#", EV [i], ": R2 = ", r2, sep=""))
  }

  # *************************** PLOTTING TOTAL
EVENT**********************************
  # Time series year:
  ylims <- range (num.events.sp$total.events)
  # First plot time series of numbers of events
  #par (mar=c(2,2,0.5,0.5))
  plot (years, num.events.sp$total.events, "l", col=1, ylab="Total Number of
Events in SP", ylim=ylims)
  par (new=TRUE)
  plot (years, unempl.sp$unempl, "l", col="blue", xlab="", ylab="",
xaxt="n", yaxt="n", frame=FALSE)
  axis (pretty (unempl.sp$unempl), side=4)
  legend("topright", lwd=1, col=c("black", "blue"), legend=c("#events",
"unemployment"))


  # Time series monthly:
  ylims <- range (num.events.sp.mo$total.events)
  # First plot time series of numbers of events
  #par (mar=c(2,2,0.5,0.5))
  plot (years, num.events.sp.mo$total.events, "l", col=1, ylab="Total Number
of Events in SP", ylim=ylims)
  par (new=TRUE)
  plot (years, unempl.sp.mo$unempl, "l", col="blue", xlab="", ylab="",
xaxt="n", yaxt="n", frame=FALSE)
  axis (pretty (unempl.sp.mo$unempl), side=4)
  legend("topright", lwd=1, col=c("black", "blue"), legend=c("#events",
"unemployment"))


  # Autoregression analysis yearly:
  ar.sp <- ar (num.events.sp$total.events, unempl.sp$unempl)
```

```
  plot (unempl.sp$unempl, num.events.sp$total.events)
  par (ps=10) # point size= size of font
  text (unempl.sp$unempl, num.events.sp$total.events, labels=years,
pos=3,col="blue")
  mod <- lm (num.events.sp$total.events ~ unempl.sp$unempl)
  mod.predict <- predict.lm (mod)
  lines (unempl.sp$unempl, mod.predict, col="red", lwd=2)
  r2 <- formatC (summary (mod)$r.squared, format="f", digits=4)
  title (main=paste ("R2 = ", r2, sep=""))


  # Autoregression analysis yearly:
  ar.sp <- ar (num.events.sp.mo$total.events, unempl.sp.mo$unempl)

  plot (unempl.sp.mo$unempl, num.events.sp.mo$total.events)
  par (ps=10) # point size= size of font
  text (unempl.sp.mo$unempl, num.events.sp$total.events, labels=years,
pos=3,col="blue")
  mod <- lm (num.events.sp$total.events ~ unempl.sp$unempl)
  mod.predict <- predict.lm (mod)
  lines (unempl.sp$unempl, mod.predict, col="red", lwd=2)
  r2 <- formatC (summary (mod)$r.squared, format="f", digits=4)
  title (main=paste ("R2 = ", r2, sep=""))
```

Appendix 2: R Script for NUTS2 analysis.

```
#.......................................................................
....
# Title: NUTS2_Analysis_Thesis.R
#
# Author: Adeva, Elisabet
#
# Date: January 26th 2014
#
## Topic: Assessing economic variables and simulation of social unrest in
Europe.

reading the data
data_1=read.csv("v_comb.csv", head=T, na.strings = "NA")

# variables
p_00_10=data_1$p_00_10
u_00_10=data_1$u_00_10
edu_00_10=data_1$edu_00_10
e_00_10=data_1$e_00_10
hdu_00_10=data_1$hdu_00_10
ppi_00_10=data_1$ppi_00_10
rop_00_10=data_1$rop_00_10
y_00_10=data_1$y_00_10
ypnie_00_10=data_1$ypnie_00_10

# creating the regression formula with dependent and independent variables
formula_st_long=p_00_10 ~
u_00_10+edu_00_10+e_00_10+hdu_00_10+ppi_00_10+rop_00_10+y_00_10+ypnie_00_10
# run the OLS model with the above formula
mod.lm_st_long <- lm(formula_st_long, data = data_1)
# read the model output
summary(mod.lm_st_long, Nagelkerke=TRUE)


reading the data
data_1=read.csv("v_comb.csv", head=T, na.strings = "NA")

# variables
p_10=data_1$p_00_10
u_00_10=data_1$u_00_10
edu_00_10=data_1$edu_00_10
e_00_10=data_1$e_00_10
hdu_00_10=data_1$hdu_00_10
ppi_00_10=data_1$ppi_00_10
rop_00_10=data_1$rop_00_10
y_00_10=data_1$y_00_10
ypnie_00_10=data_1$ypnie_00_10

# creating the regression formula with dependent and independent variables
formula_st_long=p_00_10 ~
u_00_10+edu_00_10+e_00_10+hdu_00_10+ppi_00_10+rop_00_10+y_00_10+ypnie_00_10
# run the OLS model with the above formula
mod.lm_st_long <- lm(formula_st_long, data = data_1)
# read the model output
summary(mod.lm_st_long, Nagelkerke=TRUE)
```

Appendix 3: R Script for analysis.

```
#.......................................................................
....
# Title: Eco_Analysis_Thesis.R
#
# Author: Adeva, Elisabet
#
# Date: January 26th 2014
#
# Topic: Assessing economic variables and simulation of social unrest in
Europe.
#.......................................................................
....
# Covariates: unemployment, General government gross, General government
gross debt, inflation rate, Gross domestic product at market prices, GDP per
capita, Final consumption expenditure , Labour productivity, Export market
shares, International investment, Students in 3rd education, Expenditure on
education , Population density ,Old-age-dependency ratio, Inactive
population, Temporary employees, Annual net earnings, Temporary employees,
Social security funds , Tax rate, Police officers, Crimes recorded by the
police, Share of imports by Member State, public budget spending related to
R & D, Annual net earnings, Social protection expenditure
#.......................................................................
....

#####################last model long table eco
database###############################

data_1=read.csv("C:\\Users\\Elisabet\\Documents\\Thesis_update_27thDec\\data
\\eco_data\\eu_ecodata_longtable_until2000.csv", head=T, na.strings = "NA")

data_1
str(data_1)

#Explanatory variables (see attachment for further explanation)
geo=data_1$geo_code
p_=data_1$p_
deb_00_10=data_1$debt_            # % of GDP Debt
deb_t00_10=data_1$debt_t          # t-1 year % of GDP Debt
deb_d00_10=data_1$debt_d          # Difference in years in % of GDP Debt
unempl_00_10=data_1$unempl        # Unemployment rate %
unempl_t00_10=data_1$unempl_t     # t-1 year Unemployment rate %
unempl_d00_10=data_1$unempl_d     # Difference on unemployment rate %
gdp_00_10=data_1$gdp_             # General government gross
gdpcap_00_10=data_1$gdp_cap       # General government gross per capita
consump_00_10=data_1$consum_gdp   # Euro per inhabitant-Final consumption
expenditure
consum_00_10=data_1$consum        # Percentage of GDP-Final consumption
expenditure
lab_prod_00_10=data_1$lab_prod_eu   #Labour productivity
lab_00_10=data_1$lab_prod
prc_hicp_00_10=data_1$prc_hicp    #current prices
share_00_10=data_1$share           #Export market shares
invest_00_10=data_1$invest         # International investment
infl_00_10=data_1$inflat           #Inflation rate
inclend_00_10=data_1$inc_lend      #Income, saving and net lending/
borrowing
tax_00_10=data_1$tax_gov           #Main national accounts tax aggregates
gov_00_10=data_1$gov_gen_main     #General government, Net lending(+)/net
borrowing(-)
gov1_00_10=data_1$gov_gen_exp      #General government, Total expenditure
gov2_00_10=data_1$gov_gen_rev     #General government, Total revenue
gov3_00_10=data_1$gov_cen_main    # Central government, Net lending(+)/net
borrowing(-)
gov4_00_10=data_1$gov_cen_exp        # Central government, Total expenditure
gov5_00_10=data_1$gov_cen_rev        # Central government, Total revenue
gov6_00_10=data_1$gov_loc_main    #Local government, Net lending(+)/net
borrowing(-)
gov7_00_10=data_1$gov_loc_exp      # Local government, Total expenditure
gov8_00_10=data_1$gov_loc_rev        # Local government, Total revenue
gov9_00_10=data_1$gov_soc_main #Social security funds, Net lending(+)/net
```

```
borrowing(-)
gov10_00_10=data_1$gov_soc_exp  # Social security funds, Total Gener Govern expenditure
gov11_00_10=data_1$gov_soc_rev   # Social security funds, Ttal Gener Govern revenue
edu3rd00_10=data_1$edu_3    #students in 3rd education
edu_exp00_10=data_1$edu_exp        #Expenditure on education(%GDP/public expenditure)
educ_invest00_10=data_1$educ_invest
part5600_10=data_1$part56
pdens_00_10=data_1$pdens           # Population density
oldp_00_10=data_1$old_dep          # Old-age-dependency ratio
inactive00_10=data_1$inactive      # Inactive population
temp00_10=data_1$temp              # Temporary employees (1 000)
earn_00_10=data_1$earn_snk_net     # Annual net earnings
ern_tax00_10=data_1$ern_tx         # Tax rate
poli_00_10=data_1$poli             # Police officers
crim_00_10=data_1$crime            # Crimes recorded by the police
ext_00_10=data_1$ext_trade         # Share of imports by Member State (%)
gba_00_10=data_1$gba_nabste        # all public budget spending related to R & D
spr_00_10=data_1$exp_publ          # Social protection expenditure

##Multivariable Linear Regression Model
formula=p_~
unempl_d00_10+unempl_t00_10+gdpcap_00_10+consum_00_10+lab_prod_00_10+lab_00_10+gov_00_10+invest_00_10+gov2_00_10+gov4_00_10+deb_d00_10+edu3rd00_10+earn_00_10+ern_tax00_10+poli_00_10+ext_00_10+gba_00_10

mod.lm<- lm(formula, data = data_1)
summary(mod.lm, Nagelkerke=TRUE)

#No correlation variables
+unempl_00_10
+gov10_00_10
+consump_00_10
+deb_00_10
+share_00_10
+deb_t00_10
+tax_00_10
+gdp_00_10
+gov1_00_10
+infl_00_10 # no correlation, perhaps try with the difference as I did with the unempl
+gov3_00_10
+gov5_00_10
+gov6_00_10
+gov7_00_10
+gov8_00_10
+gov9_00_10
+temp00_10
+prc_hicp_00_10
+gov11_00_10
+inclend_00_10
+educ_invest00_10
+edu_exp00_10
+exp_3rd00_10
+part5600_10
+oldp_00_10
+inactive00_10
+crim_00_10
+pdens_00_10
```

Appendix 4: R Script for ppm analysis.

```
#.......................................................................
....
#
# Title: ppm_Analysis_Thesis.R
#
# Author: Adeva, Elisabet
#
# Date: January 26th 2014
#
# Topic Assessing economic variables and simulation of social unrest in
Europe.
################## point process analysis for Europe 2000-
2010##########################

library(splancs)
library(spatstat)
library(stpp)
library(maptools)

#reading the boundary polygon
bound_owin=
readShapePoly("/media/elisabet/MYLINUXLIVE/Thesis_update_27thDec/data/Shapef
iles/eu_region_generalized_etrs89larea.shp")
proj4string(bound_owin)=CRS("+init=epsg:3034")  # ETRS 1989 Larea EPSG:3034

#reading the protest data
events=read.csv("/media/elisabet/MYLINUXLIVE/Thesis_update_27thDec/data/prot
est_data/eu_data_with_real_coordinates.csv",header=T)
CoorXY=events[,60:61]
coordinates(events)<-CoorXY
x<-events$X
y<-events$Y

#making the polywindow region
polyowinRegion=as(bound_owin,"owin")
source("/media/elisabet/MYLINUXLIVE/Thesis_update_27thDec/others/kinhom2.txt
")

# making the point process class combining the protest and spain boundary
eventsppp=ppp(x=events$lon_m_etrs,y=events$lat_m_etrs,window=polyowinRegion)

# removing the duplicate points
eventsppp=unique(eventsppp)
summary(eventsppp)
QRegion=quadscheme(data= eventsppp,
dummy=list(x=events$lon_m_etrs,y=events$lat_m_etrs))
QRegion

# MODEL 1: Inhomogeneous Poisson (2000-2010) (trend = ~polynom(x,y,2))

# running the point process model
Model.all.xy2.eu=ppm(eventsppp,~polynom(x,y,2),Poisson())
# creating the envelop from 99 simulation
Alldata_Po_xy2.eu=envelope(Model.all.xy2.eu,Linhom,nsim=99, global=F)
# plotting the ppm with simulation envelop
plot(Alldata_Po_xy2.eu,ylab="L(r)",xlab="r (distance in
meters)",cex.lab=1,cex.axis=1,cex.main=1,main="Inhomogeneous Poisson EU
(2000-2010) (trend = ~polynom(x,y,2))");
polygon(c(Alldata_Po_xy2.eu$r,rev(Alldata_Po_xy2.eu$r)),c(Alldata_Po_xy2.eu$
lo,rev(Alldata_Po_xy2.eu$hi)),col="grey",border="grey");
lines(Alldata_Po_xy2.eu$r,Alldata_Po_xy2.eu$obs,lwd=2);
lines(Alldata_Po_xy2.eu$r,Alldata_Po_xy2.eu$mmean,col=2,lty=2,lwd=1);
lines(Alldata_Po_xy2.eu$r,Alldata_Po_xy2.eu$lo,col=4,lty=2,lwd=1);
lines(Alldata_Po_xy2.eu$r,Alldata_Po_xy2.eu$hi,col=3,lty=2,lwd=1)


# MODEL 2 : Area-Interection Model

# creating the sequence for testing the best bit
s = data.frame(r=seq(1000,250000, by=50000))
```

```
ratioAIa <- profilepl(s,AreaInter,eventsppp,~polynom(x,y,2),rbord=0.05)
# choosing the best fit model
AI<-ratioAIa$fit
# crating the simulation envelop for the fitted model from 9 simulaiton
Model_AI_envelopes=envelope(AI,fun=Linhom2,global=T,nrank=2,nsim=99,correcti
on="border",control=list(expand=1))
# plotting the AI model with envelop
plot(Model_AI_envelopes,ylab="L(r)",xlab="r",cex.lab=1,cex.axis=1,cex.main=1
, main="Inhomogeneous AI (2000-2010) trend = ~polynom(x,y,2)");
polygon(c(Model_AI_envelopes$r,rev(Model_AI_envelopes$r)),c(Model_AI_envelop
es$lo,rev(Model_AI_envelopes$hi)),col="grey",border="grey");
lines(Model_AI_envelopes$r,Model_AI_envelopes$obs,lwd=2);
lines(Model_AI_envelopes$r,Model_AI_envelopes$mmean,col=2,lty=2,lwd=1);
lines(Model_AI_envelopes$r,Model_AI_envelopes$lo,col=4,lty=2,lwd=1);
lines(Model_AI_envelopes$r,Model_AI_envelopes$hi,col=3,lty=2,lwd=1)

# making the prediction map
plot(predict.ppm(AI, type="trend"))
# calculating the AIC value of the fitted model (explains the strength of
the model)
extractAIC(AI)
# simulating points and plot
sim.all<-rmh(AI)
plot(sim.all)

########################Analysis for United Kingdom 2000-
2010############################

library(splancs)
library(spatstat)
library(stpp)
library(maptools)
library(lattice)

#1
events=read.csv("C:\\Users\\Elisabet\\Desktop\\Thesis_update_27thDec\\data\\
protest_data\\eu_data_with_real_coordinates_file_used_for_eu.csv",header=T)
events=subset(events, ActionGeo_CountryCode %in% "UK")
events=subset(events, Year %in% 2000 & ActionGeo_CountryCode %in% "UK")
names(events)
events$ActionGeo_CountryCode
events

#2
bound_owin=
readShapePoly("C:\\Users\\Elisabet\\Desktop\\Thesis_update_27thDec\\data\\Sh
apefiles\\boundaries\\deu_uk_fra\\deu_uk_fra\\gbr_region_etrs89larea_v01.shp
")
proj4string(bound_owin)=CRS("+init=epsg:3034")  # ETRS 1989 Larea EPSG:3034
plot(bound_owin)

#3
####here I have to add the subset for a particular year
CoorXY=events[,60:61] # look for lon_m_etrs in the file and write the
possition here
coordinates(events)<-CoorXY
x<-events$lon_m_etrs
y<-events$lat_m_etrs
plot(events,add=T,col="red")

#4
polyowinRegion=as(bound_owin,"owin")
source("C:\\Users\\Elisabet\\Documents\\Thesis_update_27thDec\\others\\kinho
m2.txt")

#5
events=ppp(x=events$lon_m_etrs,y=events$lat_m_etrs,window=polyowinRegion)
events=unique(events)
summary(events)
QRegion=quadscheme(data=events,
dummy=list(x=events$lon_m_etrs,y=events$lat_m_etrs))
QRegion
```

```
plot(events,add=T,main="Protests in United Kingdom (2000-
2010)",cex=0.3,pch=3,mark.col=TRUE,mark.col="red")
jpeg(file =
"C:\\Users\\Elisabet\\Documents\\R\\plots\\uk\\events20002010.jpg")
dev.off()

# MODEL 1: Inhomogeneous Poisson UK (2000-2010) (trend = ~polynom(x,y,2))
Model.all.xy2.uk=ppm(events,~polynom(x,y,2),Poisson())
Alldata_Po_xy2.uk=envelope(Model.all.xy2.uk,Linhom,nsim=99, global=F)

plot(Alldata_Po_xy2.uk,ylab="L(r)",xlab="r (distance in
meters)",cex.lab=1,cex.axis=1,cex.main=1,main="Inhomogeneous Poisson - UK
(2000-2010) (trend = ~polynom(x,y,2))");
polygon(c(Alldata_Po_xy2.uk$r,rev(Alldata_Po_xy2.uk$r)),c(Alldata_Po_xy2.uk$
lo,rev(Alldata_Po_xy2.uk$hi)),col="grey",border="grey");
lines(Alldata_Po_xy2.uk$r,Alldata_Po_xy2.uk$obs,lwd=2);
lines(Alldata_Po_xy2.uk$r,Alldata_Po_xy2.uk$mmean,col=2,lty=2,lwd=1);
lines(Alldata_Po_xy2.uk$r,Alldata_Po_xy2.uk$lo,col=4,lty=2,lwd=1);
lines(Alldata_Po_xy2.uk$r,Alldata_Po_xy2.uk$hi,col=3,lty=2,lwd=1)
jpeg(file = "C:\\Users\\Elisabet\\Documents\\R\\plots\\uk\\M120002010.jpg")
dev.off()


# MODEL 2: Area-Interection Model

# creating the sequence for testing the best bit
s = data.frame(r=seq(1000,250000, by=50000)) #1km to 250km, by 50km
ratioAIa <- profilepl(s,AreaInter,events,~polynom(x,y,2),rbord=0.05)
# choosing the best fit model
AI<-ratioAIa$fit
# crating the simulation envelop for the fitted model from 99 simulaiton
Model_AI_envelopes=envelope(AI,fun=Linhom2,global=T,nrank=2,nsim=99,correcti
on="border",control=list(expand=1))
# plotting the AI model with envelop
plot(Model_AI_envelopes,ylab="L(r)",xlab="r",cex.lab=1,cex.axis=1,cex.main=1
, main="Inhomogeneous AI UK (2000-2010) trend = ~polynom(x,y,2)");
polygon(c(Model_AI_envelopes$r,rev(Model_AI_envelopes$r)),c(Model_AI_envelop
es$lo,rev(Model_AI_envelopes$hi)),col="grey",border="grey");
lines(Model_AI_envelopes$r,Model_AI_envelopes$obs,lwd=2);
lines(Model_AI_envelopes$r,Model_AI_envelopes$mmean,col=2,lty=2,lwd=1);
lines(Model_AI_envelopes$r,Model_AI_envelopes$lo,col=4,lty=2,lwd=1);
lines(Model_AI_envelopes$r,Model_AI_envelopes$hi,)
jpeg(file = "C:\\Users\\Elisabet\\Documents\\R\\plots\\uk\\M220002010.jpg")
dev.off()

# making the prediction map
plot(predict.ppm,(AI, type="trend",main="Simulation  UK (2000-2010)"))
#”trend”,“cif”,“lambda”,“se”,“SE”,  col=“red”,cex=0.5
# calculating the AIC value of the fitted model (explains the strength of
the model)
extractAIC(AI)
# simulating points and plot
sim.all<-rmh(AI)
plot(sim.all main="Simulation UK (2000-2010)") # pch=3,
mark.col=5,col='sienna', col='sienna'

#################Analysis for France 2000-2010###########################
library(splancs)
library(spatstat)
library(stpp)
library(maptools)
library(lattice)

#1 Read point data in R and create a subset
events=read.csv("C:\\Users\\Elisabet\\Desktop\\Thesis_update_27thDec\\data\\
protest_data\\eu_data_with_real_coordinates_file_used_for_eu.csv",header=T)
events=subset(events, ActionGeo_CountryCode %in% "FR")
events=subset(events, Year %in% 2000 & ActionGeo_CountryCode %in% "FR")

#2
bound_owin=
readShapePoly("C:\\Users\\Elisabet\\Desktop\\Thesis_update_27thDec\\data\\Sh
```

```
apefiles\\boundaries\\deu_uk_fra\\deu_uk_fra\\fra_region_etrs89larea_v01.shp
")
proj4string(bound_owin)=CRS("+init=epsg:3034")  # ETRS 1989 Larea EPSG:3034
plot(bound_owin)

#3 Creation of point data into class “ppp”
#### and here I have to add the subset for a particular year
CoorXY=events[,60:61] # look for lon_m_etrs in the file and write the
possition here
coordinates(events)<-CoorXY
x<-events$lon_m_etrs
y<-events$lat_m_etrs
plot(events,add=T,col="red")
jpeg(file =
"C:\\Users\\Elisabet\\Documents\\R\\plots\\fra\\events20002010.jpg")
dev.off()

#4 creation of a window
polyowinRegion=as(bound_owin,"owin")
source("C:\\Users\\Elisabet\\Documents\\Thesis_update_27thDec\\others\\kinho
m2.txt")

#5 point pattern analysis
events=ppp(x=events$lon_m_etrs,y=events$lat_m_etrs,window=polyowinRegion)
events=unique(events)
summary(events)
QRegion=quadscheme(data=events,
dummy=list(x=events$lon_m_etrs,y=events$lat_m_etrs))
QRegion
plot(events,add=T,main="Protests in France(2000-2010)")
#cex=0.3,pch=3,mark.col=TRUE,mark.col="red"
jpeg(file =
"C:\\Users\\Elisabet\\Documents\\R\\plots\\fra\\events20002010.jpg")
dev.off()

# MODEL 1: Inhomogeneous Poisson FRA (2000-2010) (trend = ~polynom(x,y,2))
Model.all.xy2.fra=ppm(events,~polynom(x,y,2),Poisson())
Alldata_Po_xy2.fra=envelope(Model.all.xy2.fra,Linhom,nsim=99, global=F)

plot(Alldata_Po_xy2.fra,ylab="L(r)",xlab="r (distance in
meters)",cex.lab=1,cex.axis=1,cex.main=1,main="Inhomogeneous Poisson - FRA
(2000-2010) (trend = ~polynom(x,y,2))");
polygon(c(Alldata_Po_xy2.fra$r,rev(Alldata_Po_xy2.fra$r)),c(Alldata_Po_xy2.f
ra$lo,rev(Alldata_Po_xy2.fra$hi)),col="grey",border="grey");
lines(Alldata_Po_xy2.fra$r,Alldata_Po_xy2.fra$obs,lwd=2);
lines(Alldata_Po_xy2.fra$r,Alldata_Po_xy2.fra$mmean,col=2,lty=2,lwd=1);
lines(Alldata_Po_xy2.fra$r,Alldata_Po_xy2.fra$lo,col=4,lty=2,lwd=1);
lines(Alldata_Po_xy2.fra$r,Alldata_Po_xy2.fra$hi,col=3,lty=2,lwd=1)
jpeg(file = "C:\\Users\\Elisabet\\Documents\\R\\plots\\fra\\M120002010.jpg")
dev.off()


# MODEL 2: Area-Interection Model

# creating the sequence for testing the best bit
s = data.frame(r=seq(1000,250000, by=50000)) #1km to 250km, by 50km
ratioAIa <- profilepl(s,AreaInter,events,~polynom(x,y,2),rbord=0.05)
# choosing the best fit model
AI<-ratioAIa$fit
# crating the simulation envelop for the fitted model from 99 simulaiton
Model_AI_envelopes=envelope(AI,fun=Linhom2,global=T,nrank=2,nsim=99,correcti
on="border",control=list(expand=1))
# plotting the AI model with envelop
plot(Model_AI_envelopes,ylab="L(r)",xlab="r",cex.lab=1,cex.axis=1,cex.main=1
, main="Inhomogeneous AI FRA (2000-2010) trend = ~polynom(x,y,2)");
polygon(c(Model_AI_envelopes$r,rev(Model_AI_envelopes$r)),c(Model_AI_envelop
es$lo,rev(Model_AI_envelopes$hi)),col="grey",border="grey");
lines(Model_AI_envelopes$r,Model_AI_envelopes$obs,lwd=2);
lines(Model_AI_envelopes$r,Model_AI_envelopes$mmean,col=2,lty=2,lwd=1);
lines(Model_AI_envelopes$r,Model_AI_envelopes$lo,col=4,lty=2,lwd=1);
lines(Model_AI_envelopes$r,Model_AI_envelopes$hi,)
jpeg(file = "C:\\Users\\Elisabet\\Documents\\R\\plots\\uk\\M220002010.jpg")
```

```
dev.off()

# making the prediction map
plot(predict.ppm(AI, type="trend",main="Simulation  FRA (2000-2010)"))
#”trend”,“cif”,“lambda”,“se”,“SE”,  col=“red”,cex=0.5
# calculating the AIC value of the fitted model (explains the strength of
the model)
extractAIC(AI)
# simulating points and plot
sim.all<-rmh(AI)
plot(sim.all, main="Simulation FRA (2000-2010)",add=T,mark.col=5 ) # pch=3,
mark.col=5,col='sienna
```

Source: Eurostat

Data from 2000 to 2010

NOTE:	Greece and Ireland where excluded due to missing data for the required years. Croatia and Norwey are/where not a member state of the European Union (EU). Swiss data is also missing.

GOVERMENT STATISTICS - ECONOMY AND FINANCE

debt_	
General government gross debt	
% of GDP and million EUR	
PC_GDP	
infl_	
HICP - inflation rate	
Annual average rate of change (%)	
gdp_	
GDP and main components - Current prices	
UNIT	Euro per inhabitant
INDIC_NA	Gross domestic product at market prices
gdp_cap	
GDP per capita - annual Data	
INDIC_NA	Nominal Gross Domestic Product per capita
UNIT	Euro per inhabitant
balance	
GDP and main components - Current prices	
INDIC_NA	External balance of goods and services
UNIT	Euro per inhabitant
consum	
Final consumption aggregates - Current prices [nama_fcs_c]	
UNIT	Euro per inhabitant
INDIC_NA	Final consumption expenditure
consum_gdp	
Final consumption aggregates - Current prices [nama_fcs_c]	
UNIT	Percentage of GDP
INDIC_NA	Final consumption expenditure
lab_prod	
Labour productivity - annual data [nama_aux_lp]	
INDIC_NA	Real labour productivity per person employed
UNIT	Percentage change on previous period
lab_prod_eu	
Labour productivity - annual data	

[nama_aux_lp]

INDIC_NA	Real labour productivity per hour worked
UNIT	Euro per hour worked

PRICES

HICP

HICP (2005 = 100) - annual data (average index and rate of change) [prc_hicp_aind]

INFOTYPE	Annual average index
COICOP	All-items HICP

prc_hicp

HICP (2005 = 100) - annual data (average index and rate of change

INFOTYPE	Annual average index
COICOP	All-items HICP

Balance of payments - International transactions

share

Export market shares [bop_q_exmash]

UNIT	Percentage of world total
POST	Current account, Goods
FLOW	Credit
PARTNER	All countries of the world

invest

International investment position - annual data [bop_ext_intpos]

CURRENCY	Million euro (from 1.1.1999)/Million ECU (up to 31.12.1998)
PARTNER	All countries of the world
FINPOS	Net positi on
FIN_TYP	International investment position: Total

(inclend) inc_lend

Income, saving and net lending/ borrowing - Current prices

UNIT	Euro per inhabitant
INDIC_NA	Gross domestic product at market prices

tax_gov

Main national accounts tax aggregates [gov_a_tax_ag]

SECTOR	General government; institutions of the EU
percentage of GDP	
INDIC_NA	Total receipts from taxes and social contributions (including imputed social contributions) after deduction of amounts assessed but unlikely to be collected

tax_pi

Taxes on production and imports less subsidies
At current prices

Annual Goverment Finance Statistics

(gov)gov_gen_main	
Government revenue, expenditure and main aggregates [gov_a_main]	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S13 - General government
INDIC_NA	B9 - Net lending (+) /net borrowing (-)
(gov1) gov_gen_exp	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S13 - General government
INDIC_NA	TE - Total general government expenditure
(gov2) gov_gen_rev	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S13 - General government
INDIC_NA	TR - Total general government revenue
(gov3) gov_cen_main	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1311 - Central government
INDIC_NA	B9 - Net lending (+) /net borrowing (-)
(gov4) gov_cen_exp	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1311 - Central government
INDIC_NA	TE - Total general government expenditure
(gov5) gov_cen_rev	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1311 - Central government
INDIC_NA	TR - Total general government revenue
(gov6) gov_loc_main	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1313 - Local government
INDIC_NA	B9 - Net lending (+) /net borrowing (-)
(gov7) gov_loc_exp	
UNIT	PC_GDP - Percentage

	of GDP
SECTOR	S1313 - Local government
INDIC_NA	TE - Total general government expenditure
(gov8) gov_loc_rev	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1313 - Local government
INDIC_NA	TR - Total general government revenue
(gov9) gov_soc_main	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1314 - Social security funds
INDIC_NA	B9 - Net lending (+) /net borrowing (-)
(gov10) gov_soc_exp	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1314 - Social security funds
INDIC_NA	TE - Total general government expenditure
(gov11) gov_soc_rev	
UNIT	PC_GDP - Percentage of GDP
SECTOR	S1314 - Social security funds
INDIC_NA	TR - Total general government revenue

EDUCATION

edu_3rd

Students

Tertiary education (1 000)

low_edu

% Persons with low educational attainment, by age group

Y25-64:From 25 to 64 years

Precentage of persons aged 25 to 64 with an education level ISCED (International Standard Classification of Education) of 2 or less. The ISCED levels 0-2 are: pre-primary, primary and lower secondary education.

educ_exp

Expenditure on education as % of GDP or public expenditure

INDIC_ED	Total public expenditure on education as % of GDP, for all levels of education combined

unempl_edu

Unemployment rates by sex, age and highest level of education attained (%)

SEX	Total
AGE	From 15 to 64 years
ISCED97	All ISCED 1997 levels

early

SEX	Total
WSTATUS	Population
UNIT	Percentage

DEMOGRAPHY

pdens

Population density
Inhabitants per km²

old_dep

%

Inactive population by sex, age and nationality (1 000)

AGE	From 15 to 64 years

Temporary employees by sex, age and highest level of education attained (1 000)

AGE	From 15 to 64 years

Tax rate

ECASE	Single person without children, 50% of AW

GOVERMNET

Police officers

mil_

source

crim_gen

Crimes recorded by the police

UNIT	Number
CRIM	Total

spr_expend

Social protection expenditure	Total expenditure
Expenditure: main results	Euro per inhabitant
SPDEPS	
UNIT	

Living conditions and welfare

Expenditure: main results [spr_exp_sum]

UNIT	Euro per inhabitant

ext_trade

House price index - deflated - 1 year % change

AVR

Households - availability of

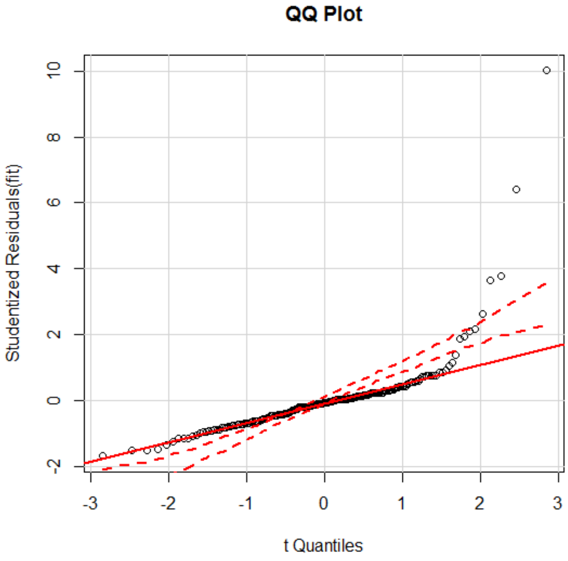
computers	
INDIC_IS	Households having access to, via one of its members, a Personal computer
HHTYP	Total

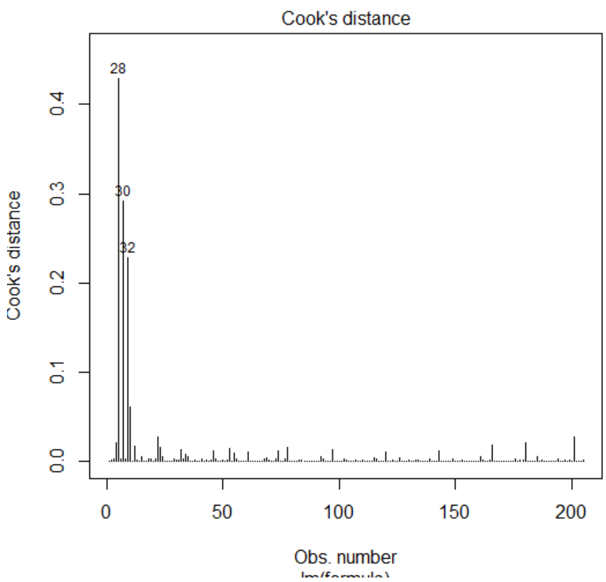
Elisabet Adeva
28th February, Germany

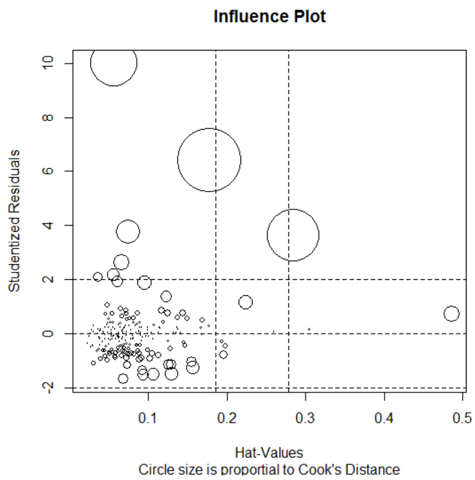
Formula

Outliers

	rstudent	unadjusted p-value	Bonferonni p
32	10.008904	3.7151e-19	7.6160e-17
28	6.412700	1.1377e-09	2.3323e-07
33	3.762235	2.2521e-04	4.6168e-02

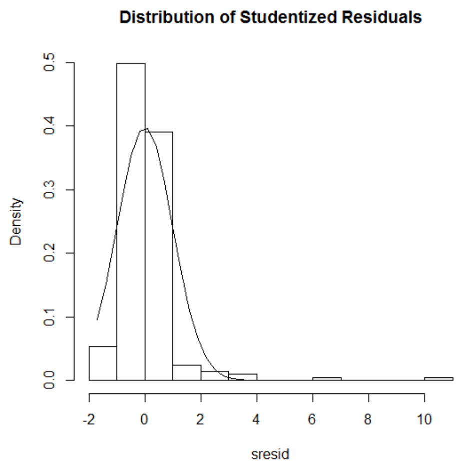






outlierTest(fit)

	rstudent	unadjusted p-value	Bonferonni p
32	10.008904	3.7151e-19	7.6160e-17
28	6.412700	1.1377e-09	2.3323e-07
33	3.762235	2.2521e-04	4.6168e-02

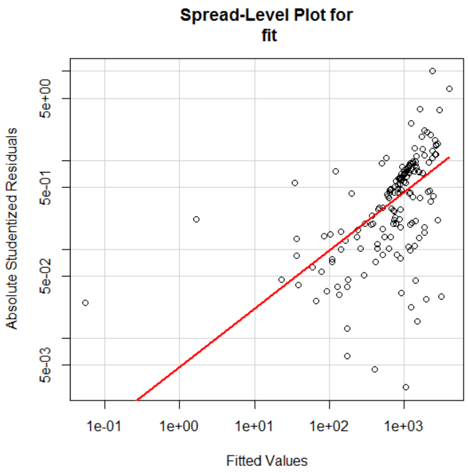


Non-normality

Non-constant Variance Score Test

Variance formula: ~ fitted.values

Chisquare = 285.9831 Df = 1 p = 3.731008e-64



Non-constant Error Variance

lag Autocorrelation D-W Statistic p-value
1 0.1851067 1.62722 0.018
Alternative hypothesis: rho != 0

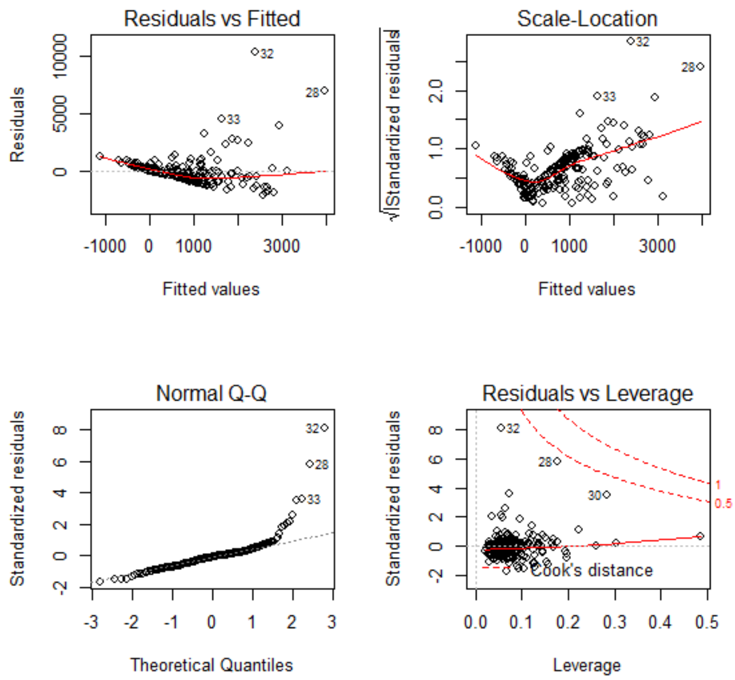
sis of Variance Table

Response: p_

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
deb_00_10	1	50816722	50816722	29.5311	1.691e-07	***
deb_t00_10	1	37322835	37322835	21.6894	6.048e-06	***
deb_d00_10	1	1819293	1819293	1.0572	0.30517	
gdp_00_10	1	2858217	2858217	1.6610	0.19905	
consump_00_10	1	7620895	7620895	4.4287	0.03667	*
consum_00_10	1	4635338	4635338	2.6937	0.10242	
lab_prod_00_10	1	6559157	6559157	3.8117	0.05238	.
lab_00_10	1	10676684	10676684	6.2045	0.01361	*
prc_hicp_00_10	1	704404	704404	0.4093	0.52308	
share_00_10	1	13069021	13069021	7.5948	0.00643	**
invest_00_10	1	6851704	6851704	3.9817	0.04744	*
infl_00_10	1	1101244	1101244	0.6400	0.42473	
tax_00_10	1	80791	80791	0.0470	0.82869	
gov_00_10	1	4071335	4071335	2.3660	0.12569	
gov1_00_10	1	1357018	1357018	0.7886	0.37566	
gov2_00_10	1	140233	140233	0.0815	0.77560	
Residuals	188	323508140	1720788			

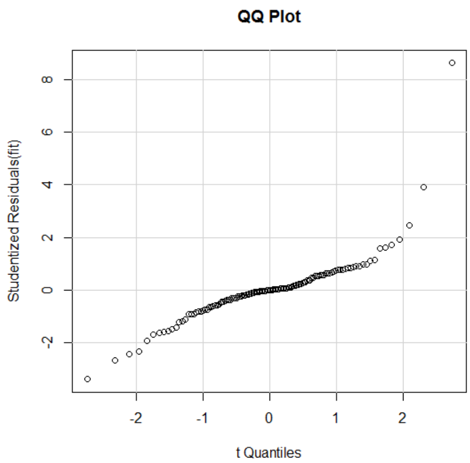
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

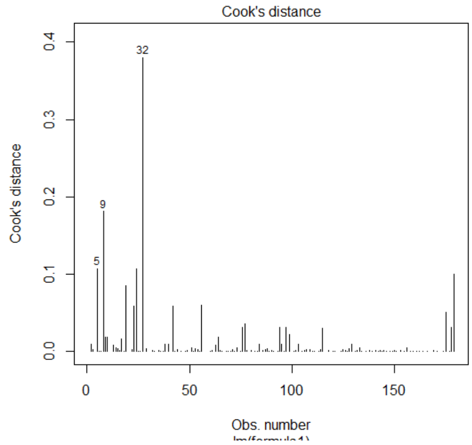
layout(matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
plot(fit)



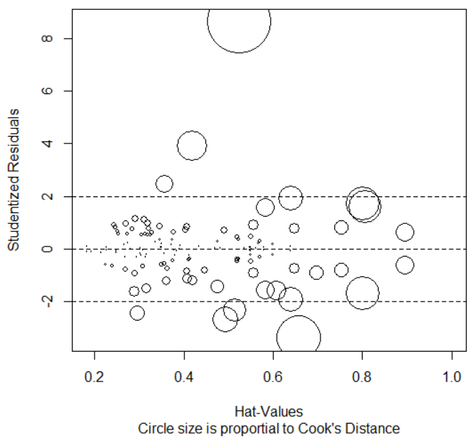
Formula1

	rstudent	unadjusted p-value	Bonferonni p
32	8.631152	1.1298e-12	1.4349e-10
22	3.909882	2.0888e-04	2.6527e-02

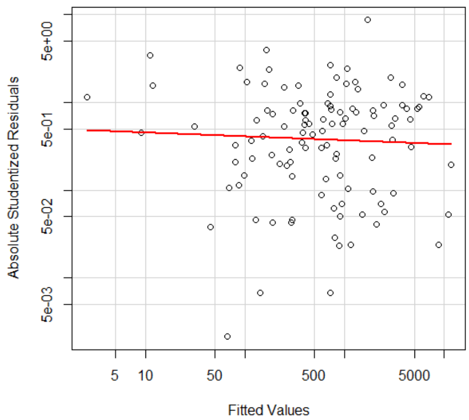


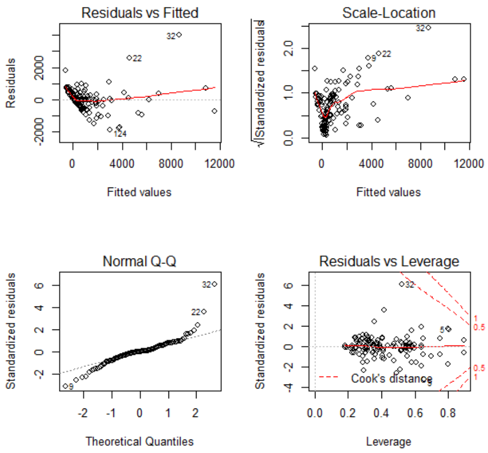


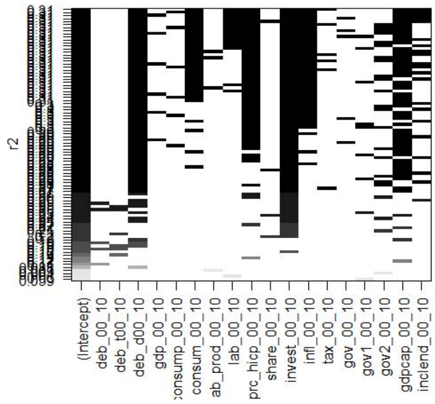
Influence Plot



Spread-Level Plot for fit




Residuals vs Fitted
Scale-Location
Normal Q-Q
Residuals vs Leverage
Residuals
Fitted values
Standardized residuals
Theoretical Quantiles
Leverage
Cook's distance


r2
(Intercept)
deb_00_10
deb_t00_10
deb_d00_10
gdp_00_10
consum_00_10
lab_00_10
prc_hicp_00_10
share_00_10
invest_00_10
infl_00_10
tax_00_10
gov_00_10
gov1_00_10
gov2_00_10
gdpcap_00_10
inclend_00_10

Masters
Program
in **Geospatial**
Technologies

ANALYSIS AND SIMULATION OF SOCIAL UNREST IN EUROPE

Towards understanding social unrest in Europe

Elisabet Adeva Romero

Dissertation submitted in partial fulfilment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*

UNIVERSIDADE NOVA
UNIVERSITAT JAUME·I
WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER

Masters Program in **Geospatial Technologies**

Supported by:

Education and Culture

ERASMUS MUNDUS