Paula Cristina Pires Simões

Mestre em Probabilidades e Estatística

Econometric models for spatio-temporal count data

Dissertação para obtenção do Grau de Doutora em

Estatística e Gestão do Risco

Orientadora: Isabel Cristina Maciel Natário, Professora Auxiliar, Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa

Co-orientadoras: Maria Lucília Salema Carvalho , Professora Associada, Faculdade de Ciências da Universidade de Lisboa

Sandra Maria da Silva Figueiredo Aleixo, Professora Coordenadora, Instituto Superior de Engenharia de Lisboa

Júri

Presidente: Ricardo Luís Rosa Jardim Gonçalves
Arguentes: Maria Antónia da Conceição Abrantes Amaral Turkman
Giovani Loiola da Silva
Vogais: Manuel Leote Tavares Inglês Esquível
Marta Cristina Vieira Faias Mateus
Isabel Cristina Maciel Natário

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

March, 2018

Econometric models for spatio-temporal count data

Este documento foi gerado utilizando o processador (pdf)LaTeX, com base no template “unlthesis” [1] desenvolvido no Dep. Informática da FCT-NOVA [2].
[1] `https://github.com/joaomlourenco/unlthesis` [2] `http://www.di.fct.unl.pt`

To my grandmothers', in memoriam

ACKNOWLEDGEMENTS

First of all, I would like to thank my adviser Prof. Dra. Isabel Natário. It was a great pleasure to work and discuss with her. Thank you for supporting, encouraging and trusting me through these years. Secondly, I would like to also thank my co-advisors, Prof. Dra. Maria Lucília Carvalho and Prof. Dra. Sandra Aleixo for helpful discussions and valuable suggestions.

To Faculty of Science and Tecnology of Nova University of Lisbon (FCT-UNL) and to Higher Institute of Engineering of Lisbon (ISEL), by this thesis had been benefited from the regime of exception of doctoral fees, under the existing cooperation protocol between FCT-UNL and ISEL, that was indispensable and I am grateful.

To Centre of Mathematics and Applications (CMA) and Mathematics Department (DM) of FCT-UNL for all the constant support and resources as a researcher.

To my colleagues of ISEL and of CMA and, in addition, to the staff for all the help.

To my friends in the CMA for listening to me and providing perspective.

Last, but certainly not least, I would like to thank my children, Inês Sofia and João Pedro, and my husband, Dário, for their love, full support and much patience, whose encouragement and interest have helped me to move on.

This thesis is dedicated to my grandmothers', Emíla and Angélica, who had taught me to work hard for what I aspire to achieve and especially for always had believed in me.

ABSTRACT

To contribute to a better understanding of the fundamental process behind the spatial and temporal correlation as well as to describe the resulted dynamics, sometimes still less reflected in econometric models, is the aim of this study. It is intended to improve the development of the necessary economic analysis which allow to optimize management policies in the most diverse areas of activity, be it hospital, road or other. This work develops and applies econometric models for count data with dependencies in space and time. The existing models are often based on the Gaussian assumption, which is sometimes inadequate. It is interesting to extend it to other types of distributions, generalizing the applicability of the available models and accompanying this development with estimation methods that make them useful. Bayesian spatial autoregressive and hierarchical models are considered as alternatives to the aforementioned models, since they are a valid and flexible alternative in the modeling of spatial effects. Spatial and spatio-temporal versions of autoregressive Bayesian models are proposed, establishing the same mathematical framework for autoregressive and hierarchical models for counting data. This is an area still underdeveloped within econometrics, given the associated but necessary complexity, and it is essential to quantify the advantages and disadvantages of its use. For the proposed methodologies, it is considered its application and implementation, in several areas of activity with scientific and technological interest, namely in the health area. In this context, a study of hospital management data is carried out, specifically the calls for the national health care line, Saúde24, in order to the development of indicators for decision support, evaluation and implementation of management and government policies, as well as to the prediction of future behavior under different scenarios. Another application in the area of road safety is also considered.

Keywords: Spatial Econometrics, Count Data, Bayesian Hierarchical Models, Bayesian Autoregressive Models, Hospital Management.

RESUMO

Contribuir para melhor compreender o processo fundamental por detrás da correlação espacial e temporal bem como descrever as dinâmicas que daí resultam, por vezes ainda pouco refletidas nos modelos econométricos, é o objectivo deste estudo, com vista ao melhor desenvolvimento das necessárias análises económicas que visam melhorar políticas de gestão nas mais diversas áreas de atividade, seja no domínio hospitalar, rodoviário ou outras. Este trabalho desenvolve e aplica modelos econométricos para dados de contagem com dependências no espaço e no tempo. Os modelos já existentes assentam frequentemente no pressuposto simplificador de normalidade, o qual se revela, por vezes, pouco adequado, sendo interessante a sua extensão a outro tipo de distribuições, generalizando a aplicabilidade dos modelos disponíveis e fazendo acompanhar esse desenvolvimento com métodos de estimação que os tornem úteis. Consideram-se modelos Bayesianos espaciais autorregressivos e hierárquicos como alternativa aos modelos antes mencionados, uma vez que estes constituem uma alternativa válida e flexível na modelação dos efeitos espaciais. Propõem-se versões espaciais e espaço-temporais de modelos Bayesianos autorregressivos, estabelecendo-se uma mesma formulação matemática para os modelos autorregressivos e hierárquicos para dados de contagem. Esta é uma área ainda pouco desenvolvida no âmbito da econometria, dada a complexidade associada, mas necessária, sendo essencial quantificar as vantagens e desvantagens da sua utilização. Para as metodologias propostas, considera-se a sua aplicação e implementação, em diversas áreas de atividade com interesse científico e tecnológico, nomeadamente na área da saúde. Neste âmbito é efetuado um estudo de dados de gestão hospitalar, mais concretamente são analisadas as chamadas para a linha de cuidados de saúde nacional - Saúde24, com vista ao desenvolvimento de indicadores de apoio à decisão, avaliação e implementação de políticas de gestão e de governo, bem como previsão de comportamentos futuros, sob diferentes cenários. É ainda apresentada uma outra aplicação na área da segurança rodoviária.

Palavras-chave: Econometria Espacial, Dados de Contagem, Modelos Hierárquicos Bayesianos, Modelos Autorregressivos Bayesianos, Gestão Hospitalar.

CONTENTS

List of Figures **xvii**

List of Tables **xix**

Listings **xxi**

Acronyms **xxiii**

1 General Introduction **3**
- 1.1 Introduction 3
- 1.2 Spatial modelling in Econometrics 7
  - 1.2.1 Econometrics 7
  - 1.2.2 Spatial Data 9
  - 1.2.3 Spatial Econometrics 9
  - 1.2.4 Spatial Dependence 10
  - 1.2.5 The SAR and CAR models 12
  - 1.2.6 Spatial Econometric Classical Models for Continuous Data 14
  - 1.2.7 Lagrange Multipliers Tests 18
  - 1.2.8 Software for Spatial Econometric Analysis 20
- 1.3 Bayesian Inference 21
  - 1.3.1 Introduction 21
  - 1.3.2 Bayesian inference with Markov Chain Monte Carlo-MCMC . . . 24
  - 1.3.3 Bayesian Inference with Integrated Nested Laplace Approximation 28
  - 1.3.4 Software 30

2 Spatial Econometric Modelling Approaches for Count Data **31**
- 2.1 Traditional spatial econometric methods applied to count data 32
  - 2.1.1 From a Poisson log-linear model to a linear log model 32
- 2.2 Spatial models for count data - Bayesian hierarchical approach 33
  - 2.2.1 Hierarchical log-Poisson regression models 35
- 2.3 Model Selection 37
  - 2.3.1 Akaike Information Criterion (AIC) 37
  - 2.3.2 Deviance Information Criterion (DIC) 38

2.3.3 Watanabe-Akaike Information Criterion (WAIC) 38
2.4 An Application: A Spatial Econometric Analysis for Road Accidents in Lisbon 39
2.4.1 Introduction 39
2.4.2 Data 40
2.4.3 Spatial Correlation 42
2.4.4 Testing for Spatial Error and Spatial Lag Dependence 44
2.4.5 Hierarchical log-Poisson models 44
2.4.6 Main Conclusions 50

3 A Bayesian Spatial Autoregressive Model for Count Data **53**
3.1 Bayesian standard spatial lag model 53
3.2 A classical Poisson spatial lag model 55
3.3 A Bayesian Poisson spatial lag model 56
3.4 Application: A Spatial Econometric Analysis of the Calls to the Portuguese National Health Line 58
3.4.1 Introduction 58
3.4.2 The S24 Data 59
3.4.3 Non-spatial modelling: The log-Poisson regression model 60
3.4.4 Spatial correlation 64
3.4.5 Spatial modelling 64
3.4.6 Comparison of results 68
3.4.7 Main conclusions 70

4 Spatio-Temporal Bayesian Models for Count Data **73**
4.1 General log-Poisson regression model 73
4.1.1 Spatial component 74
4.1.2 Spatio-temporal component 75
4.2 Spatio-temporal Bayesian hierarchical log-Poisson model 76
4.2.1 Spatio-temporal Leroux linear model 77
4.2.2 Spatio-temporal Leroux dynamic model 77
4.3 Spatio-temporal Bayesian autoregressive log-Poisson model 78
4.3.1 Bayesian Poisson spatio-temporal lag linear model 79
4.3.2 Bayesian Poisson spatio-temporal lag dynamic model 80
4.4 Application: Spatio-Temporal Modelling of S24 Data for assessing hospital savings 81
4.4.1 Econometric analysis of S24 calls to assess hospital savings 81
4.4.2 Preliminary S24 data analysis 83
4.4.3 Spatial and temporal correlation 84
4.4.4 Spatio-temporal Bayesian econometric modelling of S24 85
4.4.5 The Savings Index 91

4.4.6 Main Conclusions 92

5 Discussion **95**

Bibliography **101**

A Appendices **107**

A.1 R-code used in R-INLA to implement the Bayesian Poisson spatial lag model 107

A.2 R-code used in R-INLA to implement spatio-temporal Bayesian hierarchical log-Poisson models 109

A.3 R-code used in R-INLA to implement Bayesian Poisson spatio-temporal lag models 110

LIST OF FIGURES

2.1 Number of all road accidents with victims in Lisbon's freguesias in 2007 (left) and maximum traffic in each freguesia in that year (right). 40

2.2 Histogram of all road accidents with victims in Lisbon's freguesias in 2007 (left) and corresponding boxplot (right). 41

2.3 Histogram of the transformed variable number of accidents per amount of traffic (left) and corresponding boxplot (right). 42

2.4 Moran's I Permutation Test Plot (left) and Moran's scatter plot (right), for lisboaqueen W matrix. 43

2.5 Estimated accident relative risk (left) and estimated unstructured random effects (right) for Model 1. 47

2.6 Estimated accident relative risk (left), estimated random effects (right) for Model 2. 48

2.7 Estimated accident relative risk (left) and estimated random effects (right) for Model 3. 49

2.8 Estimated accident relative risk (left), estimated random effects (right) for Model 4. 49

2.9 Estimated accident relative risk (left) and estimated random effects (right) for Model 5. 50

3.1 Calls to S24 by service in 2014 - Graph provided by DGS. 60

3.2 The Collection of Information in S24. 60

3.3 The S24 users age empirical distribution in 2014. 61

3.4 Number of TAE calls to S24 per 1000 inhabitants, in 2014. 61

3.5 Standard Call Rate to S24, in 2014. 62

3.6 Estimated random effects for BYM model(left) and for Leroux model (right), case 1. 67

3.7 Estimated random effects for BYM model (left) and for Leroux model (right), case 2. 67

3.8 Estimated random effects for the Poisson spatial lag model, case 1. 69

3.9 Estimated random effects for the Poisson spatial lag model, case 2. 69

4.1 Boxplot of the temporal trend in the rates of saving calls as a proportion of the total number of calls whose initial intention was hospital urgency (left) and rates for each municipality averaged over 2010-2016 (right). 82
4.2 SCR-2010 to 2016. 83
4.3 Scatterplot of the log-SCR and the selected covariates for all years. 85
4.4 Spatial evolution of the estimated posterior mean of the relative risks for the chosen model. 90
4.5 Estimated spatial effects for the ST.splm.dynamic model. 91
4.6 Estimated temporal effects for the ST.splm.dynamic model. 91
4.7 Estimated spatio-temporal interaction effects for years of 2010, 2012, 2014 and 2016 for the ST.splm.dynamic model. 92
4.8 The Savings Index from 2010 to 2016. 93

List of Tables

2.1 Covariates and their estimated coefficients for the log-Poisson model, for the 2007 data. 41

2.2 Summary measures of the several weight matrices considered. 42

2.3 Observed Moran's I and corresponding p-values for two sided test of spatial independence, both considering a randomization distribution for *I* and under normality. 43

2.4 Results of Lagrange Multipliers tests for spatial dependence and spatial lag dependence and of the Moran's test. 44

2.5 Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 1. 47

2.6 Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 2. 47

2.7 Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 3. 48

2.8 Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 4. 48

2.9 Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 5. 49

2.10 DIC measured for the 5 fitted models. 50

3.1 Covariates and their estimated coefficients for the quasi-Poisson log-regression model, case 1, for the S24 2014 data. 63

3.2 Covariates and their estimated coefficients for the quasi-Poisson log-regression model, case 2, for the S24 2014 data. 63

3.3 Parameter estimates (median, 2.5% and 97.5% quantiles) for the BYM hierarchical log-Poisson model, case 1, for the S24 2014 data. 65

3.4 Parameter estimates (median, 2.5% and 97.5% quantiles) for the BYM hierarchical log-Poisson model, case 2, for the S24 2014 data. 65

3.5 Parameter estimates (median, 2.5% and 97.5% quantiles) for the Leroux hierarchical log-Poisson model, case 1, for the S24 2014 data. 66

3.6 Parameter estimates (median, 2.5% and 97.5% quantiles) for the Leroux hierarchical log-Poisson model, case 2, for the S24 2014 data. 66

3.7 Parameter estimates (mean, 2.5% and 97.5% quantiles) for the Spatial lag Poisson model, case 1, for the S24 2014 data. 68

3.8 Parameter estimates (mean, 2.5% and 97.5% quantiles) for the Spatial lag Poisson model, case 2, for the S24 2014 data. 68

3.9 DIC and WAIC measured for the 3 models fitted for case 1. 70

3.10 DIC and WAIC measured for the 3 models fitted for case 2. 70

3.11 RRMSE measured for the 3 models fitted for case 1. 71

3.12 RRMSE measured for the 3 models fitted for case 2. 71

4.1 Covariates and their estimated coefficients (posterior mean and 95% credible intervals, CI) for the Poisson log-linear model with unstructured random effects. 84
4.2 Estimated Moran's I statistic and corresponding p-values for two sided test of spatial independence, considering TAE Saving Calls and the residuals of the log-Poisson regression model with unstructured effects. 86
4.3 Parameter estimates (mean and 95% credible intervals, CI) for the spatio-temporal ST.Leroux.linear model and ST.splm.linear model, for the 2010-2016 S24 data. σ^2 is the estimated variance of the spatial effects for the autoregressive model. 87
4.4 Parameter estimates (mean and 95% credible intervals) for the ST.Leroux.dynamic model and ST.splm.dynamic model, for the 2010-2016 S24 data. σ^2 is the estimated variance of the spatial effects for the autoregressive model. 88
4.5 DIC and WAIC measures for the 2 models fitted for the case of a linear trend. 90
4.6 DIC and WAIC measures for the 2 models fitted for the case of a dynamic trend. 90

LISTINGS

A.1 The slmINLA function 107
A.2 Bayesian Poisson spatial lag model-Case1 108
A.3 Bayesian Poisson spatial lag model-Case2 108
A.4 The Bayesian Hierarchical log-Poisson models-Spatio-temporal analysis . 109
A.5 The St.Leroux.linear model-Spatio-temporal analysis 109
A.6 The St.Leroux.dynamic model-Spatio-temporal analysis 110
A.7 The Bayesian Poisson spatial lag model- Spatio-temporal analysis 110
A.8 The ST.splm.linear Model- Spatio-temporal analysis 111
A.9 The ST.splm.dynamic Model- Spatio-temporal analysis 112

Acronyms

AIC Akaike information criteria.

AT Therapeutic Counseling.

BYM Besag-York-Mollié.

CAR Conditional Autoregressive.

DIC Deviance Information Criteria.

GLM Generalized Linear Models.

IGS General Health Information.

INLA Integrated Nested Laplace Approximations.

LM Lagrange Multiplier.

LMR Robust LM.

LSP Assistance in Public Health.

MCMC Markov Chain Monte Carlo.

S24 Portuguese national health line Saúde24.

SAR Simultaneous Autoregressive.

SEM Spatial Error Model.

SLM Spatial Lag Model.

TAE Triage, Counseling and Routing.

WAIC Watanabe-Akaike Information Criteria.

Notation

P( ) represents the Poisson Distribution

$\chi^2(r)$ represents a chi-squared distribution with parameter r

N(.,.) represents both a Normal and a multivariate Normal distribution

i.i.d. stands for independent and identically distributed

(a,b) open interval that excludes the endpoints a and b

tr trace operator for matrices

$\propto$ proportionality symbol

I_n the $n \times n$ identity matrix

log() denotes the natural logarithm

$\boldsymbol{y} = (y_1, \ldots, y_n)$ denotes a $n \times 1$ vector

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

The attention given to spatial data in Statistics, more specifically to spatial patterns, goes back to the pioneering results of Whittle in 1954 [105], followed by other classic articles such as Besag in 1974 [19], Moran and Besag in 1975 [20], Ord also in 1975 [69], and Ripley's book in 1981 [78]. The recognition of a spatial structure in the data led to the formulation of models and theories that could describe, model and predict the phenomena that depend on their location in space. These data can be classified into three main categories, namely, geostatistical data, areal data and point patterns data [3, 29]. In recent years, interest in spatial analysis in general and spatial econometrics in particular has been growing [12], much in the social sciences [45]. This increase is undoubtedly due to the high availability of growing volumes of geo-referenced data as well as to the development of easily manipulated technology to handle this type of geographical information [35, 44]. These factors potentiated new theoretical analysis perspectives of the geographical phenomena [22, 65].

Application of Bayesian methods in the adjustment of spatial models have increased in consideration, essentially due to the flexibility that this approach enables. Bayesian estimation has seen a boom in application with the development of computational methods and algorithms that use approximate methods or, more often, iterative simulation methods. In this context, Monte Carlo simulation methods via Markov chains (MCMC) have prevailed [36]. This estimation approach also provides formal solutions to a wide range of spatial econometric estimation problems. LeSage has greatly contributed to the diffusion of Bayesian techniques in spatial econometrics [60, 61, 62, 63]. Very recently another methodological approach for Bayesian estimation has been developed using some well known approximation methods (the Laplace approximation), known as Integrated Nested Laplace Approximations (INLA) [81], based on the idea of using deterministic approximations for the posterior marginal distributions of the parameters, in a computationally efficient way. Bivand, Goméz-Rubio and Rue have explored the use of INLA for Bayesian inference in some widely used models in spatial econometrics [23].

Despite the several studies and developments in spatial models, spatial econometric models and Bayesian inference, there is still a gap for handling count data with econometric models, assuming a Poisson distribution for counts. Certain types of spatial econometrics models for discrete data, such as the case of binary responses, have received more attention leading to the estimation, among others, of a spatial probit model [23, 43, 60]. One of the main objectives of this work is to improve the understanding of the fundamental process behind spatial and temporal data correlation, in order to better describe the dynamics that result from this in the econometric models for count data, as well as to contribute methodologically for this.

Spatial econometrics traditionally rely on autoregressive models, such as the Spatial Lag Model (SLM) or the Spatial Error Model (SEM), that assume the Gaussianity of the response variable, which does not hold for counts. Consequently, their usage for count data demands data transformation to meet the assumptions of the models. New possible modelling strategies for count data are investigated here in order to avoid that [87].

In spatial statistics, a hierarchical modelling approach is the natural way to handle areal count data better considered under the Bayesian paradigm [15]. This approach allows data to have any distribution, being for count data typically chosen the Poisson distribution, resulting on Bayesian Poisson hierarchical models. The spatial structure assumed for the risk of the phenomena that is being counted is included in the first level of the hierarchy, through a prior distribution of spatially structured random effects. In addition, some non-spatially structured random effects to account for risk variation, not as yet explained, can be considered. Bayesian inference for these models is then based on simulation methods, namely the Markov Chain Monte Carlo (MCMC) method [33] or based on approximation methods such as the Integrated Nested Laplace Approximation INLA [81].

For the hierarchical approach, spatial autocorrelation is accounted for in the disturbances and not in the observed responses, but a spatial autoregressive approach might also be considered. The fact that it is plausible to think that the risk of what is being counted in one area is related to the risk in the areas of its neighbourhood, driven by effects of important covariables that may certainly impact the risk in a neighbouring area, justifying the use of the autoregressive approach [62]. In this case, the response variable in a given area is most certainly a good predictor of the response variable in its neighbouring areas, meeting the different modelling strategy of the autoregressive models.

Both hierarchical and autoregressive approaches are not yet much explored for spatial econometric models for non-Gaussian data, but their development is surely more adequate for these, avoiding data transformation and corresponding to more realistic models.

Temporal effects can also be included similarly in both approaches, being possible to specify for temporal component different structures, both parametric and a nonparametrically [17, 52]. In this thesis it is proposed Bayesian Poisson spatio-temporal models for areal data, either for autoregressive and hierarchical approaches. It is further accommodated in the same mathematical framework the corresponding models derived for the two different modelling strategies. Despite this, models from both approaches differ greatly in terms of their interpretation and motivation, and are not usually considered to model the same data.

The application and implementation of the methodologies studied can be considered in several areas of activity with scientific and technological interest. Two main important social applications are explored in this work, one in road safety accidents with victims and another on hospital management context, more specifically the calls for the Portuguese national health line Saúde24 (S24).

In terms of the road safety application, a spatial econometric analysis for the number of road accidents with victims in the smallest administrative divisions of Lisbon is carry out. Considering, as a baseline, a log-Poisson model with the most significant environmental factors affecting road accident risk in Lisbon in 2007 [68], it is evaluated and modelled possible spatial dependence in that risk. Several neighbourhood structures are considered in this quest and spatial correlation in these data is investigated under different perspectives: considering the transformed data to satisfy the specificities of the application of classic spatial econometric techniques as well as considering the original data. Two different spatial econometric models are analysed for the transformed data. In all cases, no spatial correlation is detected. Given the discrete nature of the data, several Bayesian hierarchical log-Poisson models are then fitted, implementing a different approach and finding some evidences of spatial structure in these data [84, 85].

Urgency admissions is one of the most important factors regarding hospital costs, which can possibly be mitigated by the use of national health lines such as the Portuguese Saúde24 line (S24) [73]. For future development of decision support indicators in a hospital savings context, based on the economic impact of the use of S24 rather than hospital urgency services, the hospital management application here considered investigates spatial and spatio-temporal dependencies in the number of (saving) calls to S24 in each Portuguese municipality over the period 2010-2016. Different space-time structures are investigated under both hierarchical and autoregressive perspectives. Resorting to INLA methodology, the spatio-temporal structure is modelled through a set of autocorrelated random effects both in terms of Poisson Bayesian hierarchical models and within a spatio-temporal lag Poisson Bayesian model [86].

This thesis is organised as follows: first, chapter 1 reviews several theoretical aspects concerning spatial modelling in econometrics for continuous data and also important Bayesian analytical and computational background. It provides an overview of some spatial econometric approaches that are available to model "georeferenced" data as well as of Bayesian modelling, which are the basis for the new modelling directions for count data pursued in this thesis; then chapter 2 provides a description of several spatial econometric approaches to model Poisson count data, from the classical spatial econometrics models to the use of hierarchical Bayesian models. In this chapter it is also described the application of these methodologies to road safety area; chapter 3 details a new proposal of a Bayesian spatial autoregressive model for Poisson count data, combining insights from classical spatial econometrics and the analysis of spatial data. A spatial lag autoregressive component is incorporated in the model for counts, under the Bayesian paradigm and using INLA methodology, taking into consideration a standard spatial lag model, as recently developed within a new class of latent models defined in INLA [81] by Gómez-Rubio, Bivand and Rue in 2015 [42], and also taking into account a spatial autoregressive lag model of counts developed by Lambert, Brown and Florax in 2010 [53] under a classical perspective. This autoregressive proposal has been explored to implement econometric alternatives to model spatial structure in count data, resulting in a Bayesian Poisson spatial lag model, the object of the paper referenced as [86]. These methodologies are applied for modelling the number of calls to S24 in each municipality in 2014; chapter 4 is about spatio-time modelling, considering different space-time structures under both hierarchical and autoregressive perspectives. Resorting to INLA methodology and setting a common mathematical framework for both approaches, the spatio-temporal structure is modelled through a set of correlated random effects, both in terms of Bayesian Poisson hierarchical models and within a new Bayesian Poisson spatio-temporal lag model. In this chapter it is also presented the application study that investigates spatio-temporal dependencies in the number of (saving) calls to S24 in each Portuguese municipality over the period 2010-2016; finally, chapter 5 discusses some of the main achievements as well as some possibilities for future work.

1.2 Spatial modelling in Econometrics

The inclusion of spatial effects in econometric modelling evolved to form "one of the branches of econometrics" [7]. The definition and scope of spatial econometrics has expanded substantially over the last three decades, moving from the "margins of urban and regional modeling" to the mainstream of econometric methodology [7].
When sample data have a location component fundamental assumptions of traditional statistical methods are no longer guaranteed. Traditional econometrics has largely ignore this violation of the Gauss-Markov assumptions used in regression modeling [4]. There are alternative estimation approaches that can be used when dealing with spatial data samples [60]. An adequate alternative is to implement spatial econometric models that allow to assess the magnitude of the space influence by considering a specific weighting scheme in which relationships among spatial areas are specified [3]. The topology or spatial pattern of data is taken care by the choice of a spatial weights or contiguity matrix, commonly denoted by the letter W, and represents our comprehension of the spatial association among data in different spatial units [35]. Spatial econometrics is an appropriate area when dealing with data reflecting geographical events, which can accommodate spatial influences maintaining other factors or variables considered important to explain the phenomenons of interest [7].

1.2.1 Econometrics

Econometrics is based upon the development and application of statistical methods for studying and understanding economic events. It combines statistics with economic theory to analyze and test economic relationships. It is used in various fields of applied economics for estimating these relationships, testing economic theories, evaluating and implementing government and business policies as well as predicting future behaviors. Its main application is the forecasting of important macroeconomic variables, such as interest rates, inflation rates and gross domestic product. However, currently, the use of econometric methods goes beyond the study of economics, and it is also used in areas such as meteorology, biology, political science and education [107].

The type of analysis that can be performed in econometrics is conditioned by the nature of the data. The data sets can be classified essentially into three types, cross-sectional data, time series and panel data.

Cross-sectional data are multiple observations of several phenomena taken at a given point in time. Data on all units may not correspond precisely to the same time period. Cross-sectional data analysis gives importance to date values but not to their ordination. When one or more phenomenons are observed in two or more different moments in time, and then the observations are joined, data become classified as *pooled cross-sectional data*.

Time series data consist in observations of a variable or several variables along time. Unlike cross-sectional data, the chronological ordering of observations in a time series gives important information. Since past events can influence future events, time is considered an important dimension.

Panel data, also known as longitudinal data, consider a time series for each cross-sectional member in the data set. For example, one can collect information such as investment or financial data about the same set of firms over a fixed year time period. This type of information can also be collected on geographical units. When compared with pooled cross-sectional data, panel data differs by the fact that the same cross-sectional units, individuals, firms or countries, are followed over a given time period but no order is assigned to the observations.

The study and understanding of economic phenomena in econometrics is carried out resorting to different models and statistical techniques well established in the literature.

Simple and multiple regression is a main tool in econometrics [46]. For example, one may be interested in study the dependence of personal consumption expenditure on personal income. This kind of analysis may be important in estimating the marginal propensity to consume, that is, for a monetary unit of change in the real income, the change in average consumption expenditure [46].

There are also models that incorporate qualitative explanatory variables, called dummy variables. These variables that cannot be readily quantified, such as gender or religion, may influence the behavior of the dependent variable. It is also possible that the dependent variable in a regression model be qualitative itself. In some situations the variable is a “yes ” or “no” type, like a possession of an attribute. Some possible alternatives are the logit model or the probit model [107].

There are more then models containing a single equation relating a single dependent variable y and one or more explanatory variables, with a cause-effect relationship. In many situations this kind of relationship is not adequate, as for example if variable y is determined by explanatory variables, and some of these are, on the other hand, determined by y. For this case simultaneous equation models should be used.

To deal with statistical relationships among random variables in econometrics, correlation analysis is more used [46]. The main objective is to measure the strength or degree of association between two variables. For example, one can be interested in studying the correlation of personal consumption with personal income.

The term autocorrelation may be defined as “correlation between members of series of observations ordered in time, as in time series data, or in space, as in cross-sectional data” [46]. For example, in a time series regression of consumption expenditures data in the current period depends, among other things, on the consumption data of the previous period. For this situation autoregressive models are used.

Time series data are frequently and intensively used on empirical research in econometrics. To handle this type of data, assuming stationarity of the time series, ARIMA (Autoregressive Integrated Moving Average) modeling and extensions can be used for forecasting. Vector autoregression (VAR), another alternative, is also available. More details on the enounced methods are available in [46], for example.

The classical econometric methodologies assume a particular econometric model and investigate if it is appropriated to a given set of data typically under maximum likelihood inference. However, there are other alternatives, namely the ones that resort to Bayesian statistical inference, which can enhance some econometric techniques [14, 62].

1.2.2 Spatial Data

The availability of increasing volumes of geo-referenced data and a user friendly technology to manipulate these in geographic information systems, have been stimulating an increasing interest in spatial analysis [7, 35]. Data for which location attributes are taken into account cry for a spatial modelling approach. The recognition and incorporation of the spatial dimension can give more relevant results than an analysis that ignores it [29]. Observations for which the absolute location or relative position are explicitly taken into account are defined as spatial data. Such data are the subject of many research fields, such as epidemiology, econometrics, climatology, ecology, sociology, among others.

Analyses of Spatial data focus on detecting patterns, exploring and modelling relationships between data that form such patterns, in order to understand the processes responsible for them. Taking spatial patterns into account enables statistical analyses, for example, to emphasise the role of space as a potentially important explanatory variable of socioeconomic systems. Three main classes of spatial data can be distinguished: geostatistical or spatially continuous data, referring to observations associated with a continuous variation measure over space, taken at fixed sampling points; areal or lattice data, related to some measured attribute in partitions of the region of interest; and spatial point patterns, for which objects are the point locations where the events of interest have occurred [29].

1.2.3 Spatial Econometrics

Over the last three decades the interpretation and range of spatial econometrics developed gradually in the literature [7]. The definition provided by Anselin (1988) states that spatial econometrics is “the collection of spatial techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models” [3]. At that time, when comparing spatial econometrics to standard econometrics, a definition was given for spatial econometrics: “the specific spatial aspects of data and models in regional science that precludes a straightforward application of standard econometric methods”[3]. The referred spatial aspects may be classified into spatial dependence or spatial heterogeneity [7, 29].

Twenty years later this definition, whose subject and range were restricted to urban and regional modelling, has changed. The importance and application of spatial techniques registered an enormous growth in economics as well as in other mainstream sciences. According to Anselin (2006), spatial econometrics is now defined as a “subset of econometric methods that is concerned with spatial aspects present in cross-sectional and space

time observations" [5, 7].
When sample data have a location component associated, two settings can be considered: spatial autocorrelation between observations or spatial heterogeneity in relations. Under these, many fundamental assumptions of the classical statistical methods, namely, that data values are derived from independent observations or that exists a single relationship with constant variance across the sample data, are no longer guaranteed [60]. Spatial econometrics constitutes an adequate alternative that can be used when dealing with observations linked to geographic economic phenomena or events [35]. Variables related to location, distance and patterns are considered in model specification, estimation, diagnostic checking and prediction [60].
Similarly to what happens in any statistical modelling four important steps that define the modern spatial econometric methodology must be followed: model specification (which deals with the formal mathematical expression for spatial dependence and spatial heterogeneity in econometric models), estimation methods, testing and spatial prediction [7].

Geostatistical data, also termed field data, play an important role in environmental sciences (see, for example, [29] and references there in for more details), but are less important in spatial econometrics [35]. Areal data and spatial point processes are more used in spatial econometric analyses and their applications. In this work the focus is on areal data.

1.2.4 Spatial Dependence

Spatial association, also referred to as spatial autocorrelation, is present in situations where observations or spatial units are non-independent over space, that is, when nearby spatial units are associated in some way [29]. Such association can be identified in a number of ways, for example, using a scatter-plot where each observed value is plotted against the mean of observations in neighbouring areas - the Moran's scatter plot - or using a spatial autocorrelation statistic such as Moran's I or Geary's C. Moran's I is a measure of global spatial autocorrelation while Geary's C is more sensitive to local spatial autocorrelation [27].

Both these statistics require the choice of a spatial weights matrix, usually symmetric and denoted by the letter W (with elements w_{ij}, $i, j = 1, \ldots, n$, where n is the number of spatial units), that represents the topology or spatial arrangement of the data and our understanding of spatial association among all areas units [35]. Usually $w_{ii} = 0$, $i = 1, \ldots, n$, but for $i \neq j$ the association measure between area i and area j, w_{ij}, can be defined in many different ways, being the most usual the contiguity criterion between areas for which $w_{ij} = 1$ only if areas i and j share a common border and $w_{ij} = 0$ elsewhere [27].

Moran's scatter plot is a graph that allows to visually explore spatial autocorrelation. Considering a spatial weights matrix W, this graph has on the x axis the values of the variable of interest and on the y axis the weighted mean (by w_{ij}) of the variable values measured for the remaining spatial units. In the case that W is a contiguity matrix, the y axis corresponds to the average of the variable values of the neighbours of each spatial unit [27].

In terms of interpretation, a Moran's scatter plot depicting points essentially in the odd quadrants suggests the presence of a positive (direct) spatial correlation, with high or low values of the variable of interest tending to cluster in space; if the points are shown in the even quadrants that suggests the presence of a negative (inverse) spatial correlation, because locations tend to be surrounded by neighbours with very dissimilar values for the same variable; points around the origin indicates no spatial correlation.

Moran's I statistics is one of the most used statistics to measure spatial association. This statistics can be used directly with the dependent variable of interest or with the residuals of a fitted model, and it is formally given by

$$I = \frac{n\sum_i \sum_j w_{ij}(y_i - \overline{y})(y_j - \overline{y})}{(\sum_i \sum_j w_{ij})\sum_i (y_i - \overline{y})^2} \tag{1.1}$$

representing y the quantity of interest.

Using I statistics, tests for the null hypotheses of spatial independence can be built under two different situations: using a randomized distribution of the statistics or Normal approximation. A significantly positive value of I indicates the presence of direct spatial correlation, a significantly negative value an inverse spatial correlation and when I is close to zero the absence of spatial correlation. Note that for relatively small values of n the I distribution may be far apart from the normal distribution approximation and the randomized test is preferred.

Another statistic used to measure the spatial association is the Geary's C statistic, given by:

$$c = \frac{(n-1)\sum_i \sum_j w_{ij}(y_i - y_j)^2}{2(\sum_i \sum_j w_{ij})\sum_i (y_i - \overline{y})^2} \tag{1.2}$$

This statistics is always positive having expected value one. Values of c less than the expected value indicate the presence of direct spatial association while otherwise the presence of inverse spatial association is signaled [27].

When spatial autocorrelation is identified, a specialized set of methods is needed [14, 60]. In order to capture dependencies across spatial units, spatially correlated variables can be introduced in the model specification [7].

The definition of the spatial weights matrix W, where the spatial topology of the spatial units is specified, is very important since estimation results may critically depend on the choice of this matrix. There are several approaches to define it, but they can essentially be classified into two main groups: spatial contiguity and distance based approaches. Typical types of neighbouring matrices for spatial contiguity approach are the linear, the rook, the bishop and the queen contiguity matrices, described below. For the distance approach there are, for example, the k-nearest neighbours or the critical cut-off neighbourhood matrices [60].

Contiguity Matrix: Represents a $n \times n$ symmetric matrix W, with elements $w_{ij} = 1$ when i and j are neighbours and 0 when they are not. By convention, the diagonal elements are set to zero. W is usually standardized so that all rows sum to one, $\widetilde{W} = (\widetilde{w}_{ij})_{n\times n}$ with $\widetilde{w}_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$, [60].

- **Linear contiguity:** Defines $w_{ij} = 1$ for regions that share a common edge to the immediate right or left (or up and down) to the region of interest;
- **Rook contiguity:** Two regions are considered neighbours if they share a common border, and for these $w_{ij} = 1$.
- **Bishop contiguity:** Defines $w_{ij} = 1$ for regions that share a common vertex;
- **Queen contiguity:** Regions that share a common border or a vertex are considered neighbours, and for these $w_{ij} = 1$.

Distance Approach: makes direct use of the latitude-longitude coordinates associated with spatial data observations for defining W [14].

- **Critical Cut-off Neighbourhood:** Two regions i and j are considered neighbours if $0 \leq d_{ij} < d^*$, with d_{ij} an appropriate distance adopted between regions, and d^* representing the critical cut-off or threshold distance, beyond which no direct spatial influence between spatial units is considered.
- **k-Nearest Neighbour:** Given the centroid distances from each spatial unit i to all units $j \neq i$ ranked as $d_{ij}(1) < d_{ij}(2) < \ldots < d_{ij}(n-1)$, for each $k = 1, 2, \ldots, n-1$, the set $N_k(i) = \{j(1), j(2), \ldots, j(k)\}$ contains the k closest units to i. Given k, the k-nearest neighbour matrix has elements of the form $w_{ij} = 1$, $j(i) \in N_k(i)$, $i \in 1, \ldots, n$, and is zero otherwise.

1.2.5 The SAR and CAR models

In order to capture dependencies across spatial units, through spatially correlated variables, the two most common approaches are the Simultaneous Autoregressive (SAR) specification and the Conditional Autoregressive (CAR) specification. These autoregressive specifications are frequently used to model spatial structure underlying areal data, and are known as areal or lattice models. The SAR models were first presented by Wittle in 1954 [105], the CAR models by Besag in 1974 [19] being among the most commonly used spatial statistical models. Both correspond to special cases of a general spatial process $\{y_i : i \in S\}$ for which a neighbouring structure is defined based on the shape of the area, formed by a countable set of locations, the indexing set S.

Choosing a matrix W for the neighbourhood structure, both models CAR and SAR incorporates spatial dependence into the model covariance structure as a function of W and a fixed unknown spatial autoregressive parameter [103].

In what follows consider $\{y_i : i \in S\}$ a Gaussian random process where the regions $\{S_1, \ldots, S_n\}$ constitute a partition of S, that is $S_1 \cup \ldots \cup S_n = S$ and $S_i \cap S_j = \emptyset$, $\forall i \neq j; i, j = 1, \ldots, n$.

This process $\boldsymbol{y} = (y_1, ..., y_n)^T$ can be modeled using a simultaneous autoregressive (SAR) model by

$$y_i = \sum_j b_{ij} y_j + \epsilon_i,\ i = 1, ..., n$$

with $E(y_i) = 0$, $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^T \sim N\left(0, \sigma^2 I_n\right)$, I_n the n dimensional identity matrix and $B = (b_{ij})_{n \times n}$ a matrix containing constants b_{ij}. B allows $\boldsymbol{y}$ to relate to itself, and is called the spatial dependence matrix. This model can be written as

$$\boldsymbol{y} = B\boldsymbol{y} + \boldsymbol{\epsilon}$$

and rewritten so that $\boldsymbol{y}$ only appears on the left side as

$$\boldsymbol{y} = (I_n - B)^{-1} \boldsymbol{\epsilon}.$$

Note that spatial units cannot depend on themselves, so matrix B must have zeros on the diagonal. Additionally, that $(I_n - B)^{-1}$ must exist.

Matrix B is responsible for the spatial dependence in the SAR model, because the error terms ϵ_i are correlated with $\{y_j : j \neq i\}$ the model is called "simultaneous", leading to the simultaneous autoregression of $\boldsymbol{y}$ on its neighbourhoods. The joint distribution of $\boldsymbol{y} = (y_1, ..., y_n)^T$ is then given by $\boldsymbol{y} \sim N(0, \Sigma_S)$, where

$$\Sigma_S = \sigma^2 (I_n - B)^{-1} ((I_n - B)^{-1})^T. \tag{1.3}$$

The covariance matrix must be positive definite, which is ensured by the fact $(I_n - B)^{-1}$ exists.

For SAR models the covariance matrix, Σ_S, must therefore comply with the following conditions: $(I_n - B)$ is nonsingular, and $b_{ii} = 0, \forall i$ [48].

The Conditional autoregressive (CAR) model is other possibility to model $\{y_i : i \in S\}$, in which each element of the random process is taken conditionally on the values of the neighbouring units, defined by

$$y_i | \boldsymbol{y}_{(-i)} \sim N(\sum_j c_{ij} y_j, \tau_i^2),\ i = 1, ..., n,$$

where $\boldsymbol{y}_{(-i)} = \{y_j : j \neq i\}$, $E(y_i) = 0$ and τ_i^2 is the conditional variance. $C = (c_{ij})_{n\times n}$ is the spatial dependence matrix and let D be the diagonal matrix with $d_{ii} = \tau_i^2$. The conditional variance often varies with unit i.

When $(I_n - C)^{-1}D$ is positive definite, the joint distribution of $\boldsymbol{y} = (y_1, ..., y_n)^T$ is a multivariate Normal distribution, $\boldsymbol{y} \sim N(0, \Sigma_C)$, with zero mean and variance-covariance matrix

$$\Sigma_C = (I_n - C)^{-1}D. \tag{1.4}$$

Σ_C must be symmetric requiring

$$\frac{c_{ij}}{d_{ii}} = \frac{c_{ji}}{d_{jj}}, \forall i, j.$$

For CAR models the covariance matrix Σ_C must therefore comply with the following conditions: $(I_n - C)$ has positive eigenvalues, and $\frac{c_{ij}}{d_{ii}} = \frac{c_{ji}}{d_{jj}}, \forall i, j$ [48].

Usually B and C are constructed with a single parameter that scales a defined neighbourhood matrix W, that is $B = \rho W$ or $C = \rho W$, with W as described in section 1.2.4 and pre-defined by the user. ρ is referred as the spatial correlation parameter or spatial autoregressive parameter. In order to satisfy the referred conditions on $(I_n - \rho W)$, for both CAR and SAR models, the restriction: $\frac{1}{\lambda_1} < \rho < \frac{1}{\lambda_N}$, with λ_1 the smallest eigenvalue and λ_N the largest eigenvalue, must hold [48].

If W is row standardized (recommended for internal consistent [28]), $B = \rho\widetilde{W}$ and $C = \rho\widetilde{W}$, the expected conditional means form an average rather than a sum. In this case the restriction for ρ becomes $\frac{1}{\lambda_1} < \rho < 1$. Usually $\frac{1}{\lambda_1} < -1$, due to irregularities for negative values near the lower bound and $-1 < \rho < 1$ is then considered [48, 103].

These definitions for SAR and CAR models are widely used for modelling irregular lattices on different research areas [103], as for example, in econometrics [9] or in disease mapping [94].

1.2.6 Spatial Econometric Classical Models for Continuous Data

This section summarizes some of the available spatial autoregressive econometric models that are used to model Gaussian spatial data and the corresponding classical inference. Spatial econometric models commonly employ SAR models and inference is typically carried out with the classical maximum likelihood method. For an exhaustive review on this topic see, for example, [7] or [60].

1.2.6.1 Spatial Autoregressive Model

Consider a vector $\boldsymbol{y} = (y_1, ..., y_n)$ of observations on n spatial units and W an $n \times n$ spatial contiguity matrix. A first-order spatial autoregressive model on the response, a SAR model, is given by

$$\begin{aligned} \boldsymbol{y} &= \rho W \boldsymbol{y} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{1.5}$$

Here variation on the response y is explained as a linear combination of the response variable in neighbouring units and no other explanatory variables. Parameter ρ is the autoregressive parameter. This model is frequently used for checking the existence of spatial correlation of residuals. The error term $\boldsymbol{\epsilon}$ is supposed to follow a Normal distribution with zero mean and variance-covariance matrix $\sigma^2 I_n$. σ^2 is a global variance parameter.

The ordinary least squares estimation is not appropriate here. It would result on a biased estimator $\widehat{\rho}$ of the spatial autoregressive parameter ρ, leading to inconsistent estimates. With,

$$\widehat{\rho} = (\boldsymbol{y}^T W^T W \boldsymbol{y})^{-1} \boldsymbol{y}^T W^T \boldsymbol{y}$$

one has,

$$\begin{aligned} E(\widehat{\rho}) &= E[(\boldsymbol{y}^T W^T W \boldsymbol{y})^{-1} \boldsymbol{y}^T W^T (\rho W \boldsymbol{y} + \boldsymbol{\epsilon})] = \\ &= \rho + E[(\boldsymbol{y}^T W^T W \boldsymbol{y})^{-1} \boldsymbol{y}^T W^T \boldsymbol{\epsilon}] \neq \rho. \end{aligned}$$

The possible spatial dependence between the observations in the vector $\boldsymbol{y}$ prevents the consistency of the least squares estimate of ρ according to Anselin (1988) [3].

Consequently, for estimating ρ in this model it is commonly used the maximum likelihood estimator obtained numerically from a "simplex univariate optimization routine". The correspondent likelihood function is [60, 62]

$$L(\rho, \sigma^2 | \boldsymbol{y}) = \frac{1}{(2\pi\sigma^2)^{(n/2)}} |I_n - \rho W| \exp\{-\frac{1}{2\sigma^2} (\boldsymbol{y} - \rho W \boldsymbol{y})^T (\boldsymbol{y} - \rho W \boldsymbol{y})\}.$$

To simplify the maximization problem, a concentrated log likelihood function is constructed eliminating the disturbance variance parameter [60], by considering $\widehat{\sigma}^2$ obtained first conditionally on ρ by maximizating the conditioned log-likelihood

$$\widehat{\sigma}^2 = \frac{1}{n} [(\boldsymbol{y} - \rho\, W \boldsymbol{y})^T (\boldsymbol{y} - \rho\, W \boldsymbol{y})]$$

in the previous likelihood function, which yields, conditioned

$$\log(L(\rho, \sigma^2 | \boldsymbol{y})) = -\frac{n}{2} \Big(\log(\pi \frac{2}{n}) + 1 \Big) - \frac{n}{2} \log \Big((\boldsymbol{y} - \rho W \boldsymbol{y})^T (\boldsymbol{y} - \rho W \boldsymbol{y}) \Big) + \log(|I_n - \rho W|).$$

Using $\widehat{\rho}$, the maximum of the previous expression, to estimate of ρ, then an estimate for the parameter σ^2 is provided by

$$\widehat{\sigma}^2 = \frac{1}{n} [(\boldsymbol{y} - \widehat{\rho} W \boldsymbol{y})^T (\boldsymbol{y} - \widehat{\rho} W \boldsymbol{y})].$$

Remember that within a SAR model a constrain is imposed on the parameter ρ. This parameter can assume viable values in the range $\frac{1}{\lambda_1} < \rho < \frac{1}{\lambda_N}$, with λ_1 the smallest eigenvalue and λ_N the largest eigenvalue of matrix W, restraining optimization search to values of ρ within this range [9].

1.2.6.2 Spatial Lag Model

An extension of the spatial autoregressive model is known as the spatial lag model (SLM), defined as

$$\begin{aligned} &\boldsymbol{y} = \rho W \boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_n) \end{aligned} \tag{1.6}$$

where X is a $n \times k$ matrix of explanatory variables and the vector of parameters $\boldsymbol{\beta}$ reflects the influence of these covariates on the $\boldsymbol{y}$ variation. This model is also named in the literature as "mixed regressive-autoregressive model" [60], because it combines the standard regression model with a spatially dependent variable model. It can be rewritten so that the response only appears on the left hand side as

$$\boldsymbol{y} = (I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\epsilon}) \Leftrightarrow$$
$$\boldsymbol{y} = (I_n - \rho W)^{-1}(X\boldsymbol{\beta}) + \boldsymbol{\epsilon}',$$

$$\boldsymbol{\epsilon}' \sim N(0, \Sigma),$$

with $\Sigma = \sigma^2 (I_n - \rho W)^{-1}((I_n - \rho W)^{-1})^T$ being the variance-covariance matrix a simultaneous autoregressive SAR specification [103]. As in the previous model, a maximum likelihood iterative estimation procedure is carried out in order to estimate/obtain the autoregressive parameter ρ that maximizes the likelihood function, and consequently allowing the estimation of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ [60, 62].

1.2.6.3 Spatial Error Model

The spatial error model (SEM), a regression model with spatial autocorrelation in the residuals, corresponding to a SAR model in this error terms, is defined by

$$\begin{aligned} &\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u} \\ &\boldsymbol{u} = \lambda W \boldsymbol{u} + \boldsymbol{\epsilon} \\ &\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_n) \end{aligned} \tag{1.7}$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of observations on the dependent variable, X is a $n \times k$ matrix of explanatory variables, for each observation with parameters' vector $\boldsymbol{\beta}$ reflecting the

influence of these variables on the variation of $\boldsymbol{y}$. W is a known $n \times n$ spatial contiguity matrix and λ is a spatial autocorrelation parameter of the error term $\boldsymbol{u}$. The error term $\boldsymbol{\epsilon}$ is assumed to follow a Normal distribution with zero mean and variance-covariance matrix $\sigma^2 I_n$.
This model can also be rewritten as

$$\boldsymbol{y} = X\boldsymbol{\beta} + (I_n - \lambda W)^{-1}\boldsymbol{\epsilon}, \Leftrightarrow$$

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}',$$

$$\boldsymbol{\epsilon}' \sim N(0, \Sigma),$$

with $\Sigma = \sigma^2(I_n - \lambda W)^{-1}(I_n - \lambda W^T)^{-1}$ a non-diagonal variance-covariance matrix for the error term.

As in the previous models, a maximum likelihood iterative estimation procedure is carried out that allows to estimate conditionally the value of λ. The values of the other parameters $\boldsymbol{\beta}$ and σ^2 are estimated as a function of the conditional maximum likelihood estimator of λ and of the observed data $\boldsymbol{y}$ and X [60, 62].

1.2.6.4 General Spatial Model

The most general form of a spatial autoregressive model, that includes both the spatial lag term and a spatially correlated error term, is

$$\begin{aligned} \boldsymbol{y} &= \rho W_1 \boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{u} \\ \boldsymbol{u} &= \lambda W_2 \boldsymbol{u} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{1.8}$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of observations on the dependent variable and X is an $n \times k$ matrix of explanatory variables. W_1 e W_2 are known $n \times n$ spatial weight matrices that define spatial relations between spatial units, using the contiguity or the distance based approach. ρ, $\boldsymbol{\beta}$, $\lambda, \boldsymbol{u}$ and $\boldsymbol{\epsilon}$ are defined as in the previous models.
The log likelihood function for this model is given by

$$\ln(L(\boldsymbol{\beta}, \lambda, \rho, \sigma^2 | \boldsymbol{y})) = C - \tfrac{n}{2}\ln(\sigma^2) + \ln(|A|) + \ln(|B|) - \tfrac{1}{2\sigma^2}(\boldsymbol{a}^T B^T B \boldsymbol{a}) \tag{1.9}$$

where C denotes an inessential constant, $\boldsymbol{a} = (A\boldsymbol{y} - X\boldsymbol{\beta})$, $A = (I_n - \rho W_1)$, $B = (I_n - \lambda W_2)$.

The log likelihood function for this model can be maximized through an optimization algorithm that allows to estimate conditionally on $\boldsymbol{\beta}$ and σ^2 the values of ρ and λ. Then values of the parameters $\boldsymbol{\beta}$ and σ^2 are estimated as a function of the maximum likelihood values of ρ, λ and the observed data $\boldsymbol{y}$ and X [60].

1.2.6.5 Spatial Durbin Model

The spatial Durbin model adds to a linear model a spatial lag on the dependent variable component as well as a spatial lag on the explanatory variables in matrix X component, being defined as

$$\begin{aligned} &\boldsymbol{y} = \rho W\boldsymbol{y} + X\boldsymbol{\beta}_1 + WX\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \\ &\boldsymbol{\epsilon} \ \sim \ N(0, \sigma^2 I_n) \end{aligned} \tag{1.10}$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of observations of the dependent variable, X corresponds to the $n \times k$ matrix of observed explanatory variables with parameters vector $\boldsymbol{\beta_1}$. W is a spatial weight matrix, and ρ represents the spatial lag parameter on the dependent variable. The matrix (WX) represents a spatial lag on the explanatory variables, with associated $k \times 1$ parameters vector $\boldsymbol{\beta_2}$.

The parameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ can be expressed as,

$$\begin{aligned} &\boldsymbol{\beta}_1 = (\widetilde{X}^T \widetilde{X})^{-1} \widetilde{X}^T \boldsymbol{y} \\ &\boldsymbol{\beta}_2 = (\widetilde{X}^T \widetilde{X})^{-1} \widetilde{X}^T W\boldsymbol{y}, \end{aligned} \tag{1.11}$$

and the log-likelihood function for this model is given by the following expression,

$$\ln(L) = C + \ln|I_n - \rho W| - \tfrac{n}{2}\ln(\boldsymbol{a}_1^T \boldsymbol{a}_1 - 2\rho \boldsymbol{a}_2^T \boldsymbol{a}_1 + \rho^2 \boldsymbol{a}_2^T \boldsymbol{a}_2) \tag{1.12}$$

where C denotes an inessential constant, $\boldsymbol{a}_1 = \boldsymbol{y} - \widetilde{X}\boldsymbol{\beta}_1$, $\boldsymbol{a}_2 = W\boldsymbol{y} - \widetilde{X}\boldsymbol{\beta}_2$, $\widetilde{X} = X^T \, W X$.

Given the value of ρ that maximizes conditionally the log-likelihood function (1.12), $\widehat{\rho}$, using,

$$\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta_1}} - \widehat{\rho}\widehat{\boldsymbol{\beta}}_2) \tag{1.13}$$

an estimate for σ^2 is obtained through

$$\widehat{\sigma}^2 = \frac{(\boldsymbol{y} - \widehat{\rho} W\boldsymbol{y} - \widetilde{X}\widehat{\boldsymbol{\beta}})^T (\boldsymbol{y} - \widehat{\rho} W\boldsymbol{y} - \widetilde{X}\widehat{\boldsymbol{\beta}})}{n}. \tag{1.14}$$

It should be noted that when this model is used columns of matrix $\widetilde{X}$ can suffer from severe collinearity problems, being necessary to take this possibility into account [60].

1.2.7 Lagrange Multipliers Tests

This section presents some of the tests that are used to decide which spatial autocorrelation model is more suitable for a specific situation, testing whether spatial autocorrelation is present in the residuals $\boldsymbol{e_R}$ of a Gaussian linear regression model. These tests are based on the log-likelihood of the SEM model or of the SLM model compared to the log-likelihood of the linear regression model [10].

Moran's I test for spatial autocorrelation is not the most appropriate approach for the choice of the best spatial autocorrelation model for a specific form of spatial dependence in data. For this purpose the Lagrange Multiplier (LM) tests have proven to be more adequate [11]. This is due to the fact that these tests are based on the difference between the log-likelihoods functions, from the SEM or SLM models, and the linear regression model. This difference takes a different form, whether the alternative hypothesis to the null hypothesis of non-existence of spatial autocorrelation, is related with a spatial error or related with a spatial lag model [60]. The obtained quantities are represented by statistics that, under the null hypothesis, are approximately qui-squared distributed [9, 10]. The LM statistics are defined as follows.

LM test for spatial error: the test statistics is

$$\mathrm{LM}_{\text{error test}} = \frac{(\boldsymbol{e_R}^T W \boldsymbol{e_R}/\hat{\sigma}^2)^2}{tr(W^T W + W^2)},$$

where W is the weight matrix and $tr(\cdot)$ represents the trace of the matrix.

LM test for spatial lag dependence: the test statistics is

$$\mathrm{LM}_{\text{lag test}} = \frac{\left(\boldsymbol{e_R}^T W \boldsymbol{y}/\hat{\sigma}^2\right)^2}{(WX\widehat{\boldsymbol{\beta}})^T M (WX\widehat{\boldsymbol{\beta}})/\hat{\sigma}^2 + tr(W^T W + W^2)},$$

where $M = I - X(X^T X)^{-1} X^T$.

Robust LM test for spatial error: This is a robust version of the LM (Robust LM (LMR)) test for spatial error robust to the presence of spatial lag dependence. The test statistics is

$$\mathrm{LMR}_{\text{error test}} = \mathrm{LM}_{\text{error test}} - \mathrm{LM}_{\text{lag test}}.$$

Robust LM test for spatial lag dependence: This is a robust version of the LM (LMR) test for spatial lag dependence robust to the presence of spatial error. The test statistics is

$$\mathrm{LMR}_{\text{lag test}} = \mathrm{LM}_{\text{lag test}} - \mathrm{LM}_{\text{error test}}.$$

LM test for spatial error and spatial lag dependence: The test statistics is

$$\mathrm{LM}_{\text{SARMA}} = \mathrm{LMR}_{\text{error test}} + \mathrm{LMR}_{\text{lag test}}.$$

Under the null hypothesis, the LM test statistics for spatial error, the LM test statistics for spatial lag dependence and its robust versions are approximately qui-squared distributed with one degree of freedom. The LM test statistics for spatial error and spatial lag dependence are approximately qui-squared distributed with 2 degree of freedom.

1.2.8 Software for Spatial Econometric Analysis

Statistical software plays a fundamental role in empirical studies. This section reviews some of the available software for spatial econometric analysis, in an areal data context.

In the past, difficulties associated with computer power, for the necessary routines, led to absence of dedicated software and consequently to slow diffusion of empirical studies in spatial econometric analysis. In recent years, this scenario has changed and several options for applying spatial econometrics methodologies for real cases are available to researchers. Currently, for spatial data analysis and modelling stands out the SpaceStat software, the GeoDa software, and some packages of the R software such as the `spdep` package for spatial regression analyses [24]. The main functions and utilities of these programs or packages are summarized next.

The Space Stat software was the first software written for spatial econometrics analysis, first released in 1991, and has been upgraded since then [14]. SpaceStat is a software for visualization, analysis, modeling and iterative exploration of spatio-temporal data. It allows to view the data of a given project, display those data in maps and graphs or perform statistics to evaluate spatial patterns. SpaceStat is a space-time information system, where time is a dimension of the data, rather than an attribute linked to each object in the data set. This software provides an extensive range of spatial and spatio-temporal statistical instruments, including explanatory spatial data analysis, spatial regression techniques and spatial econometrics analysis. In terms of its most powerful tools, are those related to the estimation and hypothesis testing of spatial regressions, where spatial dependence and spatial heterogeneity can be incorporated within the modelling framework [51].

GeoDa is a software designed to implement geospatial analyses, geovisualization, geosimulation and spatial process modeling [13, 14]. It provides techniques for exploratory data analyses on spatial data in the form of points or polygons, in geographical space. It has a user-friendly graphical interface based on a Windows environment, mainly focused on graphical tools and simple descriptive spatial analysis, such as spatial autocorrelation statistics. It allows for the evaluation of global and local spatial autocorrelation through Moran's I statistic or through Moran's scatter plot and has also available the analysis of spatial outliers. In terms of spatial weights matrices, this software enables the possibility of building matrices based on different neighbouring criteria by reading a digitalized map, through a *shape file*. It admites symmetric structures for spatial weights such as contiguity or distance-based weights, and also within other non-symetric structures, such as the k-nearest neighbours weighting scheme.

The R software made available to the scientific community as a free software, that allows the implementation of this methodology [95]. The `R-spdep` package was developed within the R programming language, providing software for spatial autocorrelation and

spatial regression analysis. It includes a collection of functions to create spatial weight matrix objects as polygon contiguities, among others, allowing their use in spatial data analysis [6, 24]. It has also several tests for spatial autocorrelation, like the global Moran's I test or and the local Geary's C test, as well as several functions to estimate spatial autoregressive models as the SLM model and the SEM model.
Some other techniques are available on this package, such as impact measures for spatial lag models, weighted and unweighed spatial regression models; semi-parametric and Moran's eigenvector spatial filtering; as well as generalized spatial two stage least squares models [6].

1.3 Bayesian Inference

Bayesian methods have had a huge development in the last decades and are now present in several research areas, in general, and in spatial econometric analyses, in particular. With the development of computational methods and computational algorithms which use approximate methods or, more often, iterative simulation methods, Bayesian Inference has become a reality. It stands out the Monte Carlo simulation methods via Markov chains (MCMC) [33], as well as another very recently approach developed using some well known approximation methods (the Laplace approximation) to do Bayesian inference known as INLA - Integrated Nested Laplace Approximation [81].

This section reviews these methods, starting with a brief presentation on the fundamentals of Bayesian inference according with the referencies [18, 70].

1.3.1 Introduction

The Bayesian paradigm is based on the subjectivist interpretation of probability: the probability of a certain event measures the degree of credibility assigned to it by a certain person, in possession of evidence.
Consider a model parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$, (where k can be one or more than one), $\boldsymbol{\theta} \in \Theta$. While for the classical approach the parameter is unknown but fixed, for the Bayesian approach the parameter is considered a scalar or a random vector, that being unknown is uncertain, and all uncertainty should be quantified in probabilistic terms [70]. The uncertainty about the values of the parameter that takes into account the current state of information, previous to data, is described by a prior distribution, which can be expressed in terms of the distribution probability function, $h(\boldsymbol{\theta})$.

Suppose that we have n observations $(y_1, y_2, ..., y_n)$, which have a probability distribution that depends on these k unknown quantities so that the probability distribution function of the vector $\boldsymbol{y} = (y_1, ..., y_n)$ depends on the vector $\boldsymbol{\theta}$ in a known way. The dependence of $\boldsymbol{y}$ can be expressed in terms of the probability mass or density function $f(\boldsymbol{y}|\boldsymbol{\theta})$. In what follows, with no loss of generality, continuous observations variables and continuous parameters are considered. The adaptation to the discrete case should be straightforward.

The prior information is then combined with the data model in the inferential process, in order to express the beliefs about $\boldsymbol{\theta}$ taking into account both prior beliefs and the data, giving rise to the posterior distribution [17, 57].

Consider the data as observations of a variable or random vector, y, with distribution function F , where F is unknown but belongs to a certain family $\digamma$, of distributions,

$$\digamma = \{f(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i|\theta) : \theta \in \Theta\}.$$

In terms of the statistical model, considering the prior distribution $h(\boldsymbol{\theta})$, through Bayes' Theorem it is possible to update the initial information on $\boldsymbol{\theta}$ with the sample data, resulting on the posterior information of $\boldsymbol{\theta}$, $h(\boldsymbol{\theta}|\boldsymbol{y})$, described as the posterior probability distribution of $\boldsymbol{\theta}$. All inference is then based on $h(\boldsymbol{\theta}|\boldsymbol{y})$ [70, 98].

Observing $(y_1, y_2, ..., y_n)$, one has

$$h(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})h(\boldsymbol{\theta})}{\int_{\Theta} \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})h(\boldsymbol{\theta})d(\boldsymbol{\theta})}, \ \boldsymbol{\theta} \in \Theta,$$

where $f(\boldsymbol{y}) = \int_{\Theta} \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})h(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal distribution of $(Y_1, Y_2, ..., Y_n)$. As the left hand side is a density for $\boldsymbol{\theta}$ and $f(\boldsymbol{y})$ is a constant, $h(\boldsymbol{\theta}|\boldsymbol{y}) \propto L(\boldsymbol{\theta}|\boldsymbol{y})h(\boldsymbol{\theta})$, with $L(\boldsymbol{\theta}|\boldsymbol{y}) \equiv f(\boldsymbol{y}|\boldsymbol{\theta})$ the likelihood function of $\boldsymbol{\theta}$.

Usually $f(\boldsymbol{y})$ is not possible to be obtain analytically and numerical methods must be used.

The use of prior information in Bayesian inference requires the specification of a prior distribution for the vector of interest $\boldsymbol{\theta}$. This distribution must represent, probabilistically, the existing knowledge about $\boldsymbol{\theta}$ before performing gathering evidence. Different forms of specifying the prior distribution can be used, namely, through subjective prior distributions, through conjugated prior distributions and through non-informative prior distributions. For more details of the enounced methods see, for example, reference [70]. Within a hierarchical formulation, Simpson *et al.* (2017) [88] have recently proposed a new specification for the prior distributions for the hyperparameters of the random effects, the penalised complexity (PC) priors. This formulation handles the random effects scaling and provides a way to define priors by taking the model structure into account, needed when different types of random effects are considered in the modelling [77].

The posterior distribution $h(\boldsymbol{\theta}|\boldsymbol{y})$ describes completely the current knowledge about $\boldsymbol{\theta}$, obtained from combining the prior information in $h(\boldsymbol{\theta})$ and the sampling information in data in $f(\boldsymbol{y}|\boldsymbol{\theta})$. It is of interest to resume this actualized information, since it is often necessary to address specific inferential questions on the parameter. A summarized description of $h(\boldsymbol{\theta}|\boldsymbol{y})$ should contain a graphical representation and quantitative summaries of the location, dispersion and shape of the distribution. The inferences about the non observable parameter $\boldsymbol{\theta}$ should be based on the posterior probabilities associated with different values of $\boldsymbol{\theta}$, and conditioned by the particular observed value of $\boldsymbol{y}$.

The point estimation of the parameters, in a Bayesian perspective, considers common descriptive measures of the posterior distribution. The most widely used estimates are the posterior mode $\left(\widehat{\boldsymbol{\theta}} : h(\widehat{\boldsymbol{\theta}}|\boldsymbol{y}) = \underset{\boldsymbol{\theta}\in\Theta}{arg\,max}\, h(\boldsymbol{\theta}|\boldsymbol{y}) = \underset{\boldsymbol{\theta}\in\Theta}{arg\,max}\,\{h(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{y})\}\right)$, the posterior mean $\left(\widehat{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\boldsymbol{y}] \text{ with } E[\theta_i|\boldsymbol{y}] = \int_{\Theta} \theta_i h(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta},\ i = 1,...,k\right)$ or the posterior median $\left(\widehat{\boldsymbol{\theta}} : P(\theta_i \geq \widehat{\theta}_i|\boldsymbol{y}) \geq 0.5\right)$ and $\left(P(\theta_i \leq \widehat{\theta}_i|\boldsymbol{y}) \geq 0.5,\ i = 1,...,k\right)$.

A more informative summary of $h(\boldsymbol{\theta}|\boldsymbol{y})$ is obtained by determining a region where each parameter, θ_i, has a high posterior probability ($\geq 1-\alpha$) of belonging to, or, better through, simultaneous credible regions as in [47]. Consider $\theta_i = \theta$, $i = 1,...,k$. $R(y)$ is a credible region φ for θ if $P(\theta \in R(y)|\boldsymbol{y}) = \int_{R(y)} h(\theta|\boldsymbol{y})d\theta \geq \varphi$ and a $(1-\alpha)_i \times 100\%$ Credible interval (CI) for θ, given $\boldsymbol{y}$, is

$$CI_{(1-\alpha)_i \times 100\%}(\theta) \equiv (\underline{\theta}(\boldsymbol{y}), \overline{\theta}(\boldsymbol{y})) \equiv (\underline{\theta}, \overline{\theta}),$$

such that $\int_{\underline{\theta}}^{\overline{\theta}} h(\theta|\boldsymbol{y})d\theta \geq (1-\alpha)_i$.

$R(y)$ is a credible region φ with highest posterior density (HPD) for θ if $R(y) = \{\theta : h(\theta|\boldsymbol{y}) \geq c_\varphi\}$, com $c_\varphi > 0$ the major constant such that $\int_{R(y)} h(\theta|\boldsymbol{y})d\theta \geq \varphi$ and a $(1-\alpha)\times 100\%$ Credible HPD interval (CI-HDP) for θ, given $\boldsymbol{y}$, is

$$CI - HPD_{(1-\alpha)_i \times 100\%}(\theta) \equiv (\theta', \theta'') \equiv \{\theta : h(\theta|\boldsymbol{y}) \geq k(\alpha)\},$$

such that $k(\alpha)$ is the largest real number satisfying $P(\theta' < \theta < \theta'') \geq (1-\alpha)_i$.

In what concerns the test of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \Theta$, doing it in a Bayesian approach it is from a conceptual point of view, usually simpler than in a classical context.

Considering $P(\Theta_0) = \int_{\Theta_0} h(\theta)d\theta$ and $P(\Theta_1) = \int_{\Theta_1} h(\theta)d\theta$, the prior odds chance of H_0 against H_1 is given as

$$O(H_0, H_1) = \frac{P(\Theta_0)}{P(\Theta_1)}.$$

After observing $\boldsymbol{y}$, considering $P(\Theta_0|\boldsymbol{y}) = \int_{\Theta_0} h(\theta|\boldsymbol{y})d\theta$ and $P(\Theta_1|\boldsymbol{y}) = \int_{\Theta_1} h(\theta|\boldsymbol{y})d\theta$, the posterior odds chance of H_0 against H_1 is given as

$$O(H_0, H_1|\boldsymbol{y}) = \frac{P(\Theta_0|\boldsymbol{y})}{P(\Theta_1|\boldsymbol{y})}.$$

The hypothesis test decision is made by comparing the two odds chances in order to measure the influence of data in the change of relative credibility of the two hypothesis, defined by **Bayes factor**:

$$B(\boldsymbol{y}) = \frac{O(H_0,H_1|\boldsymbol{x})}{O(H_0,H_1)},$$

H_0 should not be rejected for large values of the Bayes factor, because a very large factor revels a strong tendency in data in favor of an hypothesis against the other.

The inferences made on the parameters of the postulated model allows to predict new data. Specifically one can predict the behavior of Y_{n+1} after observing $(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n)$, using

$$f(y_{n+1}|\boldsymbol{y}) = \int_{\Theta} f(y_{n+1}|\theta)h(\theta|\boldsymbol{y})d\theta.$$

In conclusion, we can say that Bayesian inference is mainly done through the evaluation and description of the posterior distribution of the parameters of interest, using several ways to summarize the available information. The posterior distribution can be summarized in terms of the expected value of some parameter function, in case of θ be a scalar,

$$E[g(\theta)|\boldsymbol{y}] = \int g(\theta)h(\theta|\boldsymbol{y})d\theta,$$

or by marginal posterior distributions, in case $\boldsymbol{\theta}$ is multidimensional, $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$. Thus it is frequently necessary to calculate such integrals according to the posterior distribution.

Therefore the integration of functions, often complex and multidimensional, is extremely important in Bayesian inference. Exact inference will only be possible if these integrals can be calculated analytically, otherwise approximations must be used. When one arrives to a posterior distribution $h(\boldsymbol{\theta}|\boldsymbol{y}) \propto h(\boldsymbol{\theta})f(\boldsymbol{y}|\boldsymbol{\theta})$, often there is not an easy way of finding the integrating constant and one must resort to some numerical techniques, such as numerical or simulation methods [57].

In this context the simulation methods of Monte Carlo are an appropriate alternative, whose functionality is achieved by the application of the Law of Large Numbers [2, 33, 36].

1.3.2 Bayesian inference with Markov Chain Monte Carlo-MCMC

A MCMC method is based on the Monte Carlo integration using Markov chains. Monte Carlo integration main idea is based on expressing an integral that we want to calculate as an expected value, being the problem of calculating an integral transformed into another problem of calculating an expected value. Monte Carlo integration draw samples from the distribution of interest (the posterior distribution) and then takes samples averages to approximate expectations. When the distribution of interest is not available other proposed distribution is used and the corresponding sample values are corrected to be accepted as values of the distribution of interest. This is the basis of non-iterative methods [2, 80].

MCMC is an alternative to these methods for which sampled values are generated independently. The MCMC approach keeps the idea of obtaining a sample from the posterior distribution, and calculate samples averages. However, by using an iterative simulation technique based on Markov chains, it obtains generated values that are not independent. MCMC draws these samples by running a cleverly constructed Markov chain for a long time [49, 80].

A Markov Chain is a discrete time stochastic process where given the present state, past and future states are independent [33]. The main ideia of MCMC methods is to simulate a Markov chain $\{\theta^{(t)} : t \in T\}$ for some set T in the sample space of θ that converges to a stationary distribution, $\pi(\boldsymbol{\theta})$, which is the distribution of interest, in our case, the posterior distribution of $\boldsymbol{\theta}$ known up to a constant.

There are several ways of constructing these chains, special cases of the general framework of Metropolis *et al* (1953) and Hastings (1970), as the Gibbs Sampler.

MCMC method requires that the Markov chain to use is: homogeneous (the state transition probabilities are invariant), irreducible (any state can be reached from any other state in a finite number of iterations) and aperiodic (no absorbing states).

Let $\pi(\boldsymbol{\theta})$ be the distribution of interest from which we can not sample directly. Consider the realizations $\{\boldsymbol{\theta}^{(t)}, t = 0, 1, ...\}$ of a Markov chain obtained from a simulation scheme which has π as the equilibrium distribution, under the stated conditions, and a real-valued function g. With $\boldsymbol{\theta}^{(t)}$ converging to a random vector $\boldsymbol{\theta}$ distributed as $\pi(\boldsymbol{\theta})$, after a sufficiently large burn-in period of m iterations, we have that

$$\frac{1}{n-m}\sum_{t=m+1}^{n} g(\boldsymbol{\theta}^{(t)}) \underset{n\to\infty}{\rightarrow} E_{\pi}[g(\boldsymbol{\theta})],$$

(almost everywhere) is an MCMC estimate of $E_{\pi}[g(\boldsymbol{\theta})]$. This result is known as the Ergodic Theorem, which ensures the convergence to the require expectation [36].

The initial values influence the chain behaviour, but it gradually forgets these initial values and during the process they should be discarded - burn-in period.

The most used algorithms for implementing this method are the Metropolis-Hastings algorithm and the Gibbs Sampler, producing the desired Markov chains.

The objective of Metropolis-Hastings algorithm is to simulate from a particular distribution. This algorithm starts with an initial value $\boldsymbol{\theta}$ and specifies a rule for simulating the next value in the sequence $\boldsymbol{\theta}'$ given the previous. This rule is based on a proposal density $q(.|\boldsymbol{\theta})$ from which is simulated a candidate value. The proposal distribution may depend on the chain current value. For example, it could be a Normal distribution centered in $\boldsymbol{\theta}$. Then an acceptance probability is computed, indicating the probability of that candidate value to be accepted as the next value in the sequence. This correction mechanism is responsible for the chain convergency to the equilibrium distribution [33, 36].

The Gibbs Sampler considers a joint distribution $\pi(\boldsymbol{\theta})$ from which the main focus is to sample from. For example a posterior distribution, $\pi(\boldsymbol{\theta}|y)$, where $\boldsymbol{\theta}$ is a vector of parameters. The full conditional distribution is the distribution of the *ith* component of $\boldsymbol{\theta}$ conditioned on all the remaining components. It is derived from the joint distribution and consists in

$$\pi(\theta_i|\boldsymbol{\theta}_{-i}) = \frac{\pi(\boldsymbol{\theta})}{\int \pi(\boldsymbol{\theta})d(\theta_i)},$$

where $\boldsymbol{\theta}_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_p)$.

If the full conditional distribution for all parameter are known, the Gibbs sampler could be used to sample from the joint distribution, where the state transitions are made according to these distributions [49]. An advantage of Gibbs sampler method is that the chain always moves to a new value, there is not rejection. One disadvantage is that we have to know all the full conditional distributions. If the full conditional distributions are known and can be sampled from, then the Gibbs sampling proceeds [36].

Monte Carlo and MCMC sampling methods can not be seen as models and they do not bring anymore information that already is in data $\boldsymbol{y}$ and in $h(\boldsymbol{\theta})$. They constitute a "simple way" of looking to the posterior distribution of $h(\boldsymbol{\theta}|\boldsymbol{y})$ and help to describe it, in order to make inference about the parameter $\boldsymbol{\theta}$ [36, 49, 80].

In terms of the algorithm convergence one must be aware that the sample size should be large enough. Markov chain convergency for the target distribution should be motorized and evaluated. Some important issues related with the Markov chain convergency must be addressed. It is necessary to ensure that the equilibrium condition for the generated chain was reached. Through graphical ou numerical diagnostics it is possible to decide if the chain has sufficiently explored the entire posterior distribution, assessing convergence. The Geweke method [41] or the approach suggested by Gelman and Rubin [37], for example, allow to carry out this analysis.

The Gelman and Rubin approach suggests comparing the behavior of several generated chains in terms of the variance of some summary statistics as a way of monitoring convergence of a MCMC chain, as in the the one-way analysis of variance (ANOVA): the between-sample and within-sample variances.

Let υ be a scalar summary statistics that estimates some parameter of the distribution of interest. Consider the generated values of the k chains $\{C_{ij} : 1 \leq i \leq k, 1 \leq j \leq m\}$ of length m. Compute $\{\upsilon_{im} = \upsilon(C_{i1}, ..., C_{im})\}$ for each chain. If the chains are converging to the distribution of interest, as $m \to \infty$, the sampling distributions of the considered statistics should converge to the same distribution [79].

To estimate an upper bound and a lower bound for the variance of the summary statistic v, $Var(v)$, this approach uses the between-sequence variance and the within-sequence variance of v, that converges to the variance v from above and below, respectively, as the chain converges to the distribution of interest.

The between-sequence variance is given by

$$B = \tfrac{m}{k-1} \textstyle\sum_{i=1}^{k} (\overline{v}_{i.} - \overline{v}_{..})^2. \tag{1.15}$$

where $\overline{v}_{i.} = \frac{1}{m}\sum_{i=1}^{m}(v_{ij})$ and $\overline{v}_{..} = (\frac{1}{mk})\sum_{i=1}^{k}\sum_{j=1}^{m} v_{ij}.$.

The sample variance, within the ith sequence, is

$$s_i^2 = \frac{1}{m}\sum_{j=1}^{m}(v_{ij} - \overline{v}_{i.})^2.$$

The estimate of the within sample variance is

$$W = \tfrac{1}{k}\textstyle\sum_{i=1}^{k} s_i^2. \tag{1.16}$$

The between-sequence and within-sample estimates of the variance are combined to estimate an upper bound for $Var(v)$, given by

$$\widehat{Var}(v) = \tfrac{m-1}{m} W + \tfrac{1}{m} B. \tag{1.17}$$

which is an unbiased estimator of $Var(v)$ if the chains can be considered as random samples from the distribution of interest [79].

The Gelman-Rubin statistic is the estimated potential scale reduction,

$$\sqrt{\widehat{R}} = \sqrt{\tfrac{\widehat{Var}(v)}{W}} \tag{1.18}$$

$\sqrt{\widehat{R}}$ should be closer to one if the chains have approximately converged to the distribution of interest. It is suggested that $\widehat{R}$ should be less than 1.1 or 1.2 [37].

In terms of the Geweke method, this is based on the application of the usual techniques of time series to monitor convergence of simulated sequences by MCMC. Considering the MCMC simulated values $\{\boldsymbol{\theta}^{(t)}, t = 0, 1, \ldots\}$ and being v a function of $\boldsymbol{\theta}$, $v(\boldsymbol{\theta}^{(t)})$ defines a temporal series. Having a sufficiently large number of iterations M, the mean of the first m_f iterations, as well as the mean of the m_l last iterations, $v_f = \frac{1}{m_f}\sum v(\boldsymbol{\theta}^{(t)})$ and $v_l = \frac{1}{m_l}\sum v(\boldsymbol{\theta}^{(t)})$ are calculated.

If the chain converges then the referred means should be similar. Considering $\frac{m_f}{M}$ and $\frac{m_l}{M}$ fixed,

$$\frac{(v_f - v_l)}{\sqrt{\frac{s_f^2}{m_f} + \frac{s_l^2}{m_l}}}$$

is asymptotic Normal(0,1) distributed ($M \to \infty$), with s_f^2 and s_l^2 independent estimates of the asymptotic variances of v_f and v_l, respectively.

So, in terms of the convergence diagnostic for the samples proposed by Geweke method, values within the interval $(-1.96, 1.96)$ for this statistic are indicative of convergence. For more details of the enounced methods see, for example, references [79, 98].

1.3.3 Bayesian Inference with Integrated Nested Laplace Approximation

Recently Rue *et al.* [81] have developed an approximate method, known as the Integrated Nested Laplace Approximation (INLA), that allows to estimate the marginal posterior distribution of the parameters of interest in a Bayesian model, being particularly efficient in the estimation of latent Gaussian models and capable of providing accurate and fast results [25, 81]. It is quit general in the type of model that it can fit, allowing for great automation of the inferential process. Nowadays this method is implemented in the package R-INLA of the R software. Next it is described the INLA methodology and the corresponding context in which it should be applied, according to [25] and [67].

Consider n observed values of the response variable, $\boldsymbol{y} = (y_1, \ldots, y_n)$, that are assumed to be distributed according to one of the distributions in the exponential family, with mean parameter μ_i, related to a linear predictor through a link function $g(.)$:

$$g(\mu_i) = \eta_i.$$

This linear predictor η_i is defined as:

$$\eta_i = \beta_0 + \textstyle\sum_{k=1}^{K} \beta_k X_{ki} + \sum_{j=1}^{J} f_j(z_{ji}) \tag{1.19}$$

where β_0 is a scalar that represents the intercept, $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_K\}$ correspond to linear effects of chosen covariates $\boldsymbol{X} = (X_1, \ldots, X_K)$ on the response, and $\boldsymbol{f} = \{f_1(.), \ldots, f_J(.)\}$ one non-linear effects, functions of variables $\boldsymbol{z} = (z_1, \ldots, z_J)$. The vector of latent effects, $\boldsymbol{u} = (\beta_0, \boldsymbol{\beta}, \boldsymbol{f})$ forms a Gaussian Markov Random Field (GMRF) with precision matrix $Q(\boldsymbol{\theta}_2)$, $\pi(\boldsymbol{u}|\boldsymbol{\theta}_2) \equiv N(0, Q^{-1}(\boldsymbol{\theta}_2))$ where $\boldsymbol{\theta}_2$ is a vector of hyperparameters. The distribution of $\boldsymbol{y}$ will depend on a number of parameters $\boldsymbol{\theta_1}$.

This class of models is very flexible, the terms $f_j(.)$ can assume many different forms as non-linear effects of covariates, seasonal effects, temporal or spatial random effects, covering generalized linear models, hierarchical models, spatial and spatio-temporal models.

Let $\pi(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\theta_1})$ be the conditional density function of $\boldsymbol{y}$. Assuming conditional independence given $\boldsymbol{u}$ and $\boldsymbol{\theta_1}$, the distribution of the n observations is given by

$$\pi(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\theta_1}) = \textstyle\prod_{i=1}^{n} \pi(y_i|\boldsymbol{u}, \boldsymbol{\theta_1}). \tag{1.20}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ be a single vector of parameters with prior density function $\pi(\boldsymbol{\theta})$. The posterior distribution of the latent effects $\boldsymbol{u}$ and of the parameters $\boldsymbol{\theta}$, with precision matrix $Q(\boldsymbol{\theta})$, is given by

$$\begin{aligned}\pi(\boldsymbol{u}, \boldsymbol{\theta}|y) &\propto \pi(\boldsymbol{\theta}) \times \pi(\boldsymbol{u}|\boldsymbol{\theta}) \times \pi(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) \times \pi(\boldsymbol{u}|\boldsymbol{\theta}) \times \textstyle\prod_{i=1}^n \pi(y_i|u_i, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) \times |Q(\boldsymbol{\theta})|^{\frac{1}{2}} \exp(-\tfrac{1}{2}\boldsymbol{u}^T Q(\boldsymbol{\theta})\boldsymbol{u}) \times \textstyle\prod_{i=1}^n \exp(\log(\pi(y_i|u_i, \boldsymbol{\theta}))) \\ &\propto \pi(\boldsymbol{\theta}) \times |Q(\boldsymbol{\theta})|^{\frac{1}{2}} \exp[(-\tfrac{1}{2}\boldsymbol{u}^T Q(\boldsymbol{\theta})\boldsymbol{u}) + \textstyle\sum_{i=1}^n \log(\pi(y_i|u_i, \boldsymbol{\theta}))],\end{aligned} \tag{1.21}$$

corresponding to the product of the likelihood (1.20), of the GMRF prior density function for $\boldsymbol{u}$ and the parameter prior distribution $\pi(\boldsymbol{\theta})$.

INLA approach does not estimate the posterior marginal distributions of the latent effects $\pi(u_i|\boldsymbol{y})$ and the hyperparameters $\pi(\theta_k|\boldsymbol{y})$, given by

$$\pi(u_i|\boldsymbol{y}) = \int \pi(u_i|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta},$$

$$\pi(\theta_k|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-k},$$

but rather the whole posterior distribution, by constructing "nested approximations", numerical approximations based on the Laplace approximation method. This method allows one to approximate density functions by the first terms of Taylor series expansion of the log of the densities,

$$\widetilde{\pi}(u_i|\boldsymbol{y}) = \int \widetilde{\pi}(u_i|\boldsymbol{\theta}, \boldsymbol{y})\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta},$$

$$\widetilde{\pi}(\theta_k|\boldsymbol{y}) = \int \widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-k},$$

where $\widetilde{\pi}$ corresponds to the approximate density function. The proposed Laplace approximation for $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is then given by,

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\theta})\pi(\boldsymbol{u}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\widetilde{\pi}(\boldsymbol{u}|\boldsymbol{\theta},\boldsymbol{y})}\Big|_{\boldsymbol{u}=\boldsymbol{u}^*(\theta)}, \tag{1.22}$$

where $\widetilde{\pi}(\boldsymbol{u}|\boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation, given by the Laplace approximation method, of the complete conditional distribution of $\boldsymbol{u}$, $\pi(\boldsymbol{u}|\boldsymbol{\theta}, \boldsymbol{y})$ and $\boldsymbol{u}^*(\boldsymbol{\theta})$ is the mode for a given $\boldsymbol{\theta}$.

The INLA approximation of $\pi(u_i|\boldsymbol{y})$ follows three main steps:

i Computation of an approximation to the posterior distribution of the hyperparameters, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, as in (1.22);

ii New use of the Laplace approximation to obtain $\pi(u_i|\boldsymbol{\theta},\boldsymbol{y})$. For example, rewriting the vector of parameters as $u = (u_i, \boldsymbol{u}_{-i})$,

$$\widetilde{\pi}(u_i|\boldsymbol{\theta},\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{u},\boldsymbol{\theta}|\boldsymbol{y})}{\widetilde{\pi}(\boldsymbol{u}_{-i}|u_i,\boldsymbol{\theta},\boldsymbol{y})}|_{\boldsymbol{u}_{-i}=\boldsymbol{u}^*_{-i}(u_i,\boldsymbol{\theta})}$$

where $\widetilde{\pi}(u_{-i}|u_i,\boldsymbol{\theta},\boldsymbol{y})$ is the Laplace Gaussian approximation to $\pi(u_{-i}|u_i,\boldsymbol{\theta},\boldsymbol{y})$ and $\boldsymbol{u}^*_{-i}(u_i,\boldsymbol{\theta})$ is its mode;

iii Using the previous steps and a numerical integration,

$$\widetilde{\pi}(u_i|\boldsymbol{y}) = \int \widetilde{\pi}(u_i|\boldsymbol{\theta},\boldsymbol{y})\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$$

can be solved through a finite weighted sum:

$$\widetilde{\pi}(u_i|\boldsymbol{y}) \approx \sum_m \widetilde{\pi}(u_i|\boldsymbol{\theta}^m,\boldsymbol{y})\widetilde{\pi}(\boldsymbol{\theta}^m|\boldsymbol{y})\Delta_m$$

considering that $\boldsymbol{\theta}$ has m relevant elements $\{\boldsymbol{\theta}^m\}$, with a corresponding set of weights Δ_m, where m small.

1.3.4 Software

Bayesian inference has become a reality with the development of computational methods and computational algorithms. It has been largely used in spatial statistics in recent years, mainly due to the availability of these computational methods for fitting spatial models.

The R software has been a huge contribution for this development. It made available to the scientific community, as a free software, that allows the implementation of this methodology. Also noteworthy are another five specific softwares for fitting Bayesian models, for general purpose, using Markov chain Monte Carlo methods, namely, Jags [71], OpenBugs [92], BayesX [16], Stan [97] and Nimble [101] softwares. These can also be used through connection to R. In terms of monitoring chains convergence, that can be done using the R-packages CODA [72] and BOA [89, 98].

There are also specialized spatial modelling packages developed under software R that implements MCMC for more complex Bayesian models (which will in the meantime be presented) such as CARBayes [55].

The R-INLA package offers an interface to INLA methodology being adequate for estimating a large number of the most common models. Simultaneously with the INLA methodology development, their authors have been developing a set of R-functions (R-INLA) to implement the method. Initially in simple modelling settings, they have greatly developed since then, such that nowadays a huge number of models are already covered and readily available to be fitted in R-INLA. However it is also possible for the user to develop new models [82].

CHAPTER 2

SPATIAL ECONOMETRIC MODELLING APPROACHES FOR COUNT DATA

Understanding geographical variation in discrete data, that evidence some sort of spatial association is less developed compared to the standard methods for continuous data. Two different modelling approaches are available to study spatial patterns in count data.

One, that is more used in econometrics, is the employment of classical autoregressive econometric models, as the ones presented in section 1.2.6. These models were originally designed for continuous data, thus demanding count data transformation to meet the model's assumptions [14]. However there are some important exceptions, essentially for certain types of discrete spatial data, such as the case of a binary outcome with Bernoulli distribution. For this, the spatial autoregressive lag specification has been extended through the spatial probit model or spatial tobit models [60, 62].

Another approach, more used in statistics, is the use of hierarchical Bayesian models, where data can be modeled as having any distribution, being Poisson the natural choice for count data. In spatial statistics, hierarchical modelling is the natural way to handle areal count data to account for data overdispersion as it is well established in the literature [15, 29]. Poisson log-linear models are typically used for the analyses where the linear predictor includes important factors for the phenomenon explanation. For that, non-observable random effects can be added to the effects of existing covariates in modelling extra variation that might exist in counts. This chapter provides an overview of the considered approaches to model Poisson count data.

2.1 Traditional spatial econometric methods applied to count data

For modeling count data using spatial econometric models for continuous data, by means of a spatial lag model (SLM) or a spatial error model (SEM) or even with both through the general spatial model as referred in section 1.2.6, it is necessary to transform the count dependent variable into an approximately gaussian distributed variable. This is also the case for making use of the Lagrange Multipliers tests described in section 1.2.7, which relies on a Gaussian linear model [60].

2.1.1 From a Poisson log-linear model to a linear log model

Considering the number of counts observed in each spatial unit, the transformation most commonly used is the log transformation. If it is necessary to convert the dependent variable into a rate variable the inclusion of an offset may be considered. A Gaussian linear model can then be fitted to the transformed data. The conversion into a rate variable allows comparisons between the results obtained through a Poisson log-linear model (fitted to counts before the covariates had also been transformed as needed) and the results from the application of classical spatial econometrics models, as exemplified below [84].

Consider the Poisson log-linear model

$$y \sim \text{Poisson}(\text{offset.var} \times \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)),$$

where $X_1, ..., X_k$ represents the considered covariates with parameters $\beta_1, ..., \beta_k$ and offset.var is an offset variable, for which

$$\begin{aligned} & E[y] = \mu_y = \text{offset.var} \times \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k) \Leftrightarrow \\ & \Leftrightarrow \log\left(E\left[y\right]\right) = \log(\text{offset.var}) + \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k. \end{aligned}$$

Consider now the usual log transformation to be performed to the Poisson data in order to achieve symmetry

$$w = \begin{cases} \log(y), & y > 0 \\ 0, & y = 0 \end{cases}$$

Focus on the first branch of w definition. Here:

$$\begin{aligned} w &= \log(y) = \log(\mu_y) + \log\left(1 + \frac{(y-\mu_y)}{\mu_y}\right) \approx \text{(2nd order Taylor expansion of the log)} \\ &\approx \log(\mu_y) + \frac{(y-\mu_y)}{\mu_y} - \frac{(y-\mu_y)^2}{2\mu_y^2} \end{aligned}$$

So,

$$\begin{aligned}
E[w] &= E[\log(y)] \approx \log(\mu_y) + E\left[\frac{(y-\mu_y)}{\mu_y}\right] - E\left[\frac{(y-\mu_y)^2}{2\mu_y^2}\right] = \\
&= \log(\mu_y) + \frac{1}{\mu_y}(E[y]-\mu_y) - \frac{1}{2\mu_y^2}(V[y]) = \log(\mu_y) - \frac{1}{2\mu_y} = \\
&= \log(\text{offset.var} \times \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)) - \frac{1}{2\text{offset.var} \times \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)} = \\
&= \log(\text{offset.var}) + (\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k) - \frac{\exp(-\beta_0 - \beta_1 X_1 - \ldots - \beta_k X_k)}{2\text{offset.var}} \approx \\
&\approx \text{(1st order Taylor expansion of the exponencial)} \\
&\approx \log(\text{offset.var}) + (\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k) - \frac{1 - \beta_0 - \beta_1 X_1 - \ldots - \beta_k X_k}{2\text{offset.var}} \\
&= \log(\text{offset.var}) + \underbrace{\left\{\beta_0\left(1 + \frac{1}{2\text{offset.var}}\right) - \frac{1}{2\text{offset.var}}\right\}}_{\text{Intercept}} + \beta_1 X_1\left(1 + \frac{1}{2\text{offset.var}}\right) + \\
&+ \ldots + \beta_k X_k\left(1 + \frac{1}{2\text{offset.var}}\right).
\end{aligned}$$

Consequently, and because in the particular case where it can be considered that $P(y=0)$ is very small so that $E[w]$ can be well approximated by the value above, the coefficients of the linear model are approximately the same as the coefficients of the Poisson log-linear model, provided that the values of the covariates $X_1, X_2, \ldots, X_n$ are multiplied by $\left(1 + \frac{1}{2\text{offset.var}}\right)$.

2.2 Spatial models for count data - Bayesian hierarchical approach

In order to use the traditional spatial econometric models for continuous data to model count data, defined into spatial units of a lattice, it is necessary to transform the discrete dependent variable to meet the required assumptions [60]. However, there are some alternative models which can be applied directly to count data wherein the spatial dependency structure is defined conditionally.

Part of the spatial autocorrelation can be accommodated by including known covariate risk factors in a generalized linear regression model.

The Generalized Linear Models (GLM) introduced by Nelder and Wedderburn (1972) have been playing an increasingly important role in statistical analysis, due to the large number of models that they encompass and facility of analysis associated with the rapid computer development, in responding to situations which are not properly explained by the normal linear model [99].

In a GLM the outcome $\boldsymbol{y}$ is assumed to be distributed as a member of the exponential family of distributions, with mean parameter μ, such as the Poisson distribution.

A family of distributions is said to belong to the exponential family if its probability or density function $f(\boldsymbol{y}|\boldsymbol{\theta})$ it can be expressed as

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = h(\boldsymbol{y})c(\boldsymbol{\theta})exp(\textstyle\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(\boldsymbol{y})),$$

here $h(\boldsymbol{y}) \geq 0$, $t_1(\boldsymbol{y}),\ldots,t_k(\boldsymbol{y})$ are real-valued functions of $\boldsymbol{y}$, not dependent on $\boldsymbol{\theta}$, $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}),\ldots,w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-value parameter $\boldsymbol{\theta}$, not dependent on $\boldsymbol{y}$.

Consider a vector $\boldsymbol{y} = (y_1,\ldots,y_n)$ of observations and a vector of covariates $\boldsymbol{X}^T = (X_1,..,X_k)$ with parameters $\beta_1,\ldots,\beta_k$. The relationship between the mean of the ith observation on the dependent variable, μ_i, and a linear predictor on the vector of covariates of the ith observation, $\boldsymbol{X}_i$, $i = 1,\ldots,n$ defines the systematic component of the GLM model and is established through a link function $g(.)$:

$$g(\mu_i) = \eta_i = \boldsymbol{X}_i^T\beta.$$

Suppose that y_i is Poisson (μ_i) distribuited, $i = 1,\ldots,n$. A possible and more common link function is the logarithmic function, resulting into the general log-Poisson regression model defined as

$$y_i|\boldsymbol{\beta} \backsim \text{Poisson}(\mu_i)$$
$$\log(\mu_i) = \eta_i = \boldsymbol{X}_i^T\beta.$$

It is common that some spatial structure remains in the residuals of this fitted model, even after accounting for these covariate effects. For modelling the residual autocorrelation, the most common approach is to expand the linear predictor with a set of spatially correlated random effects, in terms of a generalized linear model with random effects [15]. The generalized linear models with random effects are a different way of modelling the outcome $\boldsymbol{y}$ considering covariates and random effects, either spatially structured or not, to account for spatial autocorrelation in the analysis of spatial data [66].

The referred spatial random effects are usually modeled by a conditional autoregressive (CAR) model [21], which induces *a priori* spatial autocorrelation through the contiguity structure of the spatial units. Different CAR prior distributions commonly used for modelling spatial autocorrelation have been establish in the literature: from the Besag, York and Mollié (BYM) proposal [21], to the alternatives developed by Leroux, Lei and Breslow [58] and Stern and Cressie [93], where each model is a special case of a Gaussian Markov Random Field (GMRF).

The general model is a generalized linear mixed model for spatial areal unit data, a hierarchical model where the responses $\boldsymbol{y}$ are assumed to be Poisson distributed, better handled under the Bayesian paradigm.

The next subsection describes and explains different Bayesian hierarchical models for Poisson count data.

2.2.1 Hierarchical log-Poisson regression models

Considering a spatial domain divided into n spatial units (or areas), let $\boldsymbol{y} = (y_1, ..., y_n)$ and $\boldsymbol{e} = (e_1, ..., e_n)$ represent, respectively, the number of observed and expected cases of the phenomena that is being counted in each spatial unit, the latter obtained by some standardization procedure. The counts y_i are assumed to be Poisson distributed with expected value $E(y_i) = \mu_i = e_i \, \theta_i$, where θ_i is the relative risk in area i. Let $\boldsymbol{X}_i^T = (X_{i1}, ..., X_{ik})$ denote a set of k covariates measured in spatial unit i, for $i = 1, ..., n$, the first of which, X_{i1} corresponds to a intercept term and being $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)$ be the corresponding regression coefficients.

The general hierarchical log-Poisson regression model is defined as

$$y_i | \eta_i \sim \text{Poisson}(e_i \theta_i), \tag{2.1}$$

where $\eta_i = \log(\theta_i)$, $i = 1, ..., n$, are the log relative risks, [31, 55]. Note that $\log(e_i)$ enter as known offsets in the model. In (2.1) the log relative risks are decomposed into the effects of covariates plus some random effects that are able to account for possible over-dispersion

$$\eta_i = \log(\theta_i) = \boldsymbol{X}_i^T \beta + u_i. \tag{2.2}$$

Under formulation (2.2), it is possible to have a spatial hierarchical log-Poisson model if u_i includes spatially autocorrelated random effects ε_i, modelled by a conditional autoregressive (CAR) prior distribution.

Therefore, when spatial autocorrelation is detected in data, the spatial structure can be considered through a global CAR prior. From the existing possibilities, the ones to be used in this thesis are the Besag-York-Mollié and the Leroux models. The CAR specification defines prior conditional distributions of the spatial random effects ε, where the distribution of ε_i conditioned on $\varepsilon_{-i} = (\varepsilon_1, ..., \varepsilon_{i-1}, \varepsilon_{i+1}, ..., \varepsilon_n)$ is dependent only on the ε_j that are neighbours, according to the chosen spatial structure. CAR prior is then specified as a set of n univariate full conditional distributions, $f(\varepsilon_i | \varepsilon_{-i})$, for $i = 1, ..., n$, rather than via the multivariate specification [19].

It is necessary to establish a neighbouring criterion, by considering a symmetric non-negative weight matrix (an adjacency matrix in this work) W with elements w_{ij}, $i, j = 1, ..., n$, where n is the number of spatial units. Here it is considered the contiguity criterion between areas for which $w_{ij} = 1$ only if areas i and j share a common border and $w_{ij} = 0$ elsewhere [27].

For the regression coefficients $\beta_j,\ j = 1,\ldots,k$ Normal$(\mu_\beta, \sigma_\beta^2)$ prior distributions are considered.

2.2.1.1 Besag-York-Mollié (BYM) Model

The BYM model [21] comprises two sets of random effects, spatially correlated ε_i and unstructured γ_i random effects, that is $u_i = \varepsilon_i + \gamma_i$ in (2.2). The unstructured random effects partially account for possible effects of over-dispersion and are implemented with the exchangeable prior

$$\gamma_i \sim \mathrm{N}(0, \sigma^2), \tag{2.3}$$

with an an Inverse-Gamma prior distribution assigned to the variance parameter

$$\sigma^2 \sim \text{Inverse-Gamma}(a, b). \tag{2.4}$$

For the spatial random effects a CAR prior is proposed, where the conditional expectation of each effect is given as the average of the random effects in its neighbouring areas, while the conditional variance is inversely proportional to the number of neighbours. Thus the more areas that are close to area i and have similar values to ε_i, results in reducing uncertainty. The prior distribution of the random effects is given by,

$$\varepsilon_i | \varepsilon_{-i} \sim \mathrm{N}\left(\frac{\sum_{j=1}^{n} w_{ij}\varepsilon_j}{\sum_{j=1}^{n} w_{ij}}, \frac{\sigma_B^2}{\sum_{j=1}^{n} w_{ij}}\right), \tag{2.5}$$

with an Inverse-Gamma prior distribution assigned to σ_B^2,

$$\sigma_B^2 \sim \text{Inverse-Gamma}(a, b). \tag{2.6}$$

This model accommodates both weak and strong spatial autocorrelation. The spatial structure is split into strongly spatial correlated variation and independent spatial variation.

2.2.1.2 Leroux, Lei and Breslow Model

The previous model requires two random effects to be estimated for each data point, whereas only their sum is identifiable from data. To get through this, Leroux, Lei and Breslow [58] proposed an alternative CAR prior distribution for modelling spatial autocorrelation, using a single set of random effects for modelling varying strengths of spatial autocorrelation, that is $u_i = \varepsilon_i$ in (2.2). The prior distribution of the random effects is given by

$$\begin{aligned}
&\varepsilon_i | \varepsilon_{-i} \sim \mathrm{N}\left(\frac{\rho\sum_{j=1}^{n} w_{ij}\varepsilon_j}{\rho\sum_{j=1}^{n} w_{ij} + 1 - \rho}, \frac{\sigma_L^2}{\rho\sum_{j=1}^{n} w_{ij} + 1 - \rho}\right) \\
&\sigma_L^2 \sim \text{Inverse-Gamma}(a, b), \\
&\rho \sim \text{Uniform}(0, 1),
\end{aligned} \tag{2.7}$$

where ρ is a spatial dependence parameter, with value zero in case of independence and values near one for strong spatial autocorrelation. An uniform prior distribution on the unit interval is specified for this parameter, ρ, and an Inverse-Gamma prior distribution is adopted for the variance of the random effects, σ_L^2. This model formulation makes a compromise between unstructured and structured variation using ρ as a mixing parameter.

The CAR prior distributions defined for these models enforce a single global level of spatial smoothing for the set of random effects, which for the Leroux model is controlled by ρ.

The inference for these methods is based on MCMC methods or based on approximation methods as the INLA.

2.3 Model Selection

Bayesian models can be evaluated and compared by measuring their performance through their predictive accuracy. This can be estimated using cross-validation which requires training sets to re-fit the models, which is less convenient, or information criteria, that use functions of the deviance. Given the data $\boldsymbol{y} = (y_1, \ldots, y_n)$ with $L(\boldsymbol{\theta}|\boldsymbol{y}) \equiv f(\boldsymbol{y}|\boldsymbol{\theta})$ as likelihood function, the deviance of the model is defined as

$$D(\boldsymbol{\theta}) = -2\log(f(\boldsymbol{y}|\boldsymbol{\theta})),$$

where $\boldsymbol{\theta}$ corresponds to the parameters of the likelihood. The deviance of a model measures the variability linked to the likelihood [25]. For the information criteria approach, measures like the Akaike information criteria (AIC), [1], the Deviance Information Criteria (DIC), [25, 91] and the Watanabe-Akaike Information Criteria (WAIC), [38, 104] are the most used. DIC is a generalization of AIC, developed especially for Bayesian model comparison [91], and WAIC can be seen as an improvement over the DIC [104]. The preferred model will be the one with lower values for the considered criteria. These criteria are defined bellow.

2.3.1 Akaike Information Criterion (AIC)

The AIC measure of predictive accuracy is composed by two components, one associated with the quality of the adjustment of the model and another associated with its complexity, quantified by two times the parametric dimension, given by,

$$AIC = -2\log f(\boldsymbol{y}|\widehat{\boldsymbol{\theta}}) + 2p,$$

where $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$ and p the number of parameters.

2.3.2 Deviance Information Criterion (DIC)

In Bayesian formulation, the deviance is a random variable using the posterior mean of the deviance $\overline{D} = E_{\boldsymbol{\theta}|\boldsymbol{y}}(D(\boldsymbol{\theta}))$ as a measure of fit. Replacing in AIC measure the maximum likelihood estimate of $\boldsymbol{\theta}$ by its posterior mean and replacing p with a data based bias correction, another measure of the predictive accuracy is,

$$DIC = \overline{D} + p_D,$$

where p_D, the effective numbers of parameters, is given by

$$p_D = E_{\boldsymbol{\theta}|\boldsymbol{y}}[D(\boldsymbol{\theta})] - D(E_{\boldsymbol{\theta}|\boldsymbol{y}}[\boldsymbol{\theta}]) = \overline{D} - D(\overline{\boldsymbol{\theta}}),$$

where $D(\overline{\theta})$ is the deviance computed on the posterior mean of the parameters.

The DIC measure is also composed by two components, one for quantifying the model fit (measured through the posterior expectation of the deviance) and the other for evaluating the model complexity (measured through the effective number of parameters).

Note that DIC depends of a only data dependent function that can be omitted when the models to compare are based on the same sampling model, although the model for $\boldsymbol{y}$ may differ on the parametric structured [98].

2.3.3 Watanabe-Akaike Information Criterion (WAIC)

This measure of predictive accuracy introduced by Watanabe (2010) is given by,

$$WAIC = -2\sum_{i=1}^{n} \log E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[f(y_i|\boldsymbol{\theta})\right] + 2p_W,$$

with two different proposals for the effective number of parameters, p_W. One possibility uses the variance of individual terms in the log predictive density summed over the n data points, calculated through the posterior variance of the log predictive density for each data point y_i,

$$p_{W_1} = \sum_{i=1}^{n} Var_{\boldsymbol{\theta}|\boldsymbol{y}}(\log(f(y_i|\boldsymbol{\theta})).$$

Another possibility for p_W is the one similar to that used to construct p_D,

$$p_{W_2} = -2\sum_{i=1}^{n}\Big(\log(E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[f(y_i|\boldsymbol{\theta})\right]) - E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[\log(f(y_i|\boldsymbol{\theta}))\right]\Big).$$

It can be used either p_{W_1} or p_{W_2} as a bias correction in WAIC.

There is still some disagreement on which one of the criteria should be used. For example AIC does not perform well on settings with strong prior information; DIC can produce negative estimates of the effective number of parameters, it is based on a point estimate, when the posterior distribution is not well summarized by its mean, provides nonsensical results; WAIC uses the posterior distribution rather then a point estimate and it is invariant to re-parametrization, being referred to as "fully bayesian". However, WAIC depends on data partition that might raise difficulties for structured models [38]. Nevertheless, according to recent studies "WAIC has various advantages over simpler estimates of predictive error such as AIC and DIC" but, because it requires an additional computational effort, it is less used in practice [102]. Given the above, in this thesis we focus on WAIC and DIC measures to compare models. Actually, DIC is the predictive measure most used in Bayesian applications and WAIC has been shown to be more stable and particularly helpful with hierarchical and mixture structures, in which the number of parameters increases with sample size although when working with point estimates, it is not the most appropriate approach [38].

2.4 An Application: A Spatial Econometric Analysis for Road Accidents in Lisbon

Lisbon, the capital city of Portugal, has one of the Europe's highest number of road accidents per inhabitant [34]. In the year of 2007, for example, 2149 road accidents with victims were registered in an universe of about 565 000 inhabitants and an area of 85Km^2 (Project SACRA, Spatial Analysis of Child Road Accidents, PTDC/TRA/66161/2006, and Instituto Nacional de Estatística). This application presents a spatial econometric analysis for the number of road accidents with victims in the smallest administrative divisions of Lisbon, considering as a baseline a log-Poisson model for environmental factors, and is the object of the papers [84, 85].

2.4.1 Introduction

The occurrence of road accidents is often conditioned by many factors, from those intrinsic to the accident as the driver's specificities, type of vehicles or weather conditions, to road specific variables including road physical state and traffic signs, as well as traffic conditions. However, frequently, external factors that also contribute for the happening are not taken into account, such as all the environmental surroundings of the incident. For example, if we are talking about a city neighbourhood with many children, as expected around school grounds, or with many elderly as expected around hospitals or medical centers, it may constitute a place of higher accident risk. Additionally, if not accounted for, these environmental factors may induce some degree of spatial dependence in the risk of accident.

Considering the number of road accidents in Lisbon, in each of its 53 smallest administrative division of Lisbon city, called *freguesia*, Nunes in 2011 [68] studied the most significant environmental factors affecting road accident risk in Lisbon in 2007, but has not made an attempt to incorporate or evaluate the possible spatial dependence in that risk, for these aggregated data setting.

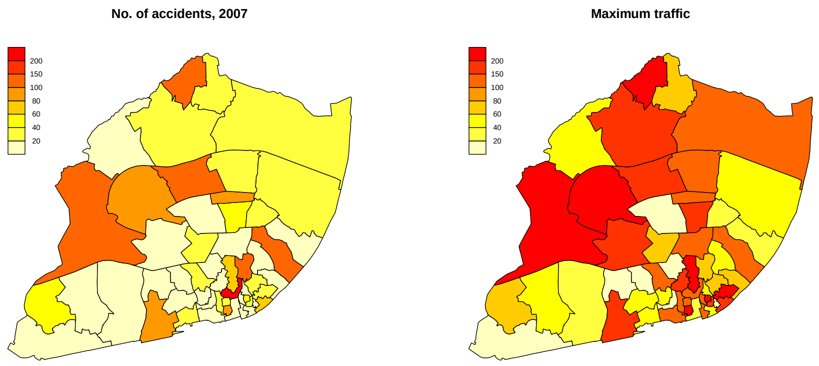

Figure 2.1: Number of all road accidents with victims in Lisbon's freguesias in 2007 (left) and maximum traffic in each freguesia in that year (right).

Through this study, standard spatial econometrics techniques are used to look for spatial dependence in the number of accidents, determining common association measures, previously to the inclusion of the environmental factors, and afterwards. Several neighborhood structures are considered in this quest. The number of accidents in each *freguesia* was also analysed in a hierarchical Bayesian setting for count data, wherein the spatial dependency structure is defined conditional, as part of the hierarchy.

2.4.2 Data

The data available is a comprehensive data set of all road accidents with victims in the city of Lisbon for the years of 2004 to 2007. These data have been geo-referenced and, under project SACRA (Spatial Analysis of Child Road Accidents, PTDC/TRA/66161/2006), they have been further organized into the smallest city administrative divisions, called *freguesias*. Data for several environmental factors have also been gathered and selected in order to better estimate risk. For more details on data and covariate detailed description see [68].

This work focuses on the 2007 data. The panel on the left of Figure 2.1 depicts the map of all the registered accidents by freguesia in that year.

Approximately 41 accidents in average per *freguesia* have occurred in 2007, 50% of the *freguesias* have between 12 and 55 accidents, with a median value of 23, and a positive asymmetric distribution of the number of accidents per *freguesia* is evident - see the data histogram and boxplot, respectively, in the left and in the right panel of Figure 2.2. There are a couple of *freguesias* with a number of accidents oddly higher then the remaining, higher than 120, but that most probably is associated with the dimension of the *freguesia* and corresponding traffic, which has to be considered in the modelling.

The number of accidents in each *freguesia* was previously modelled through a log-Poisson model [68]. Because the number of accidents is naturally proportional to the

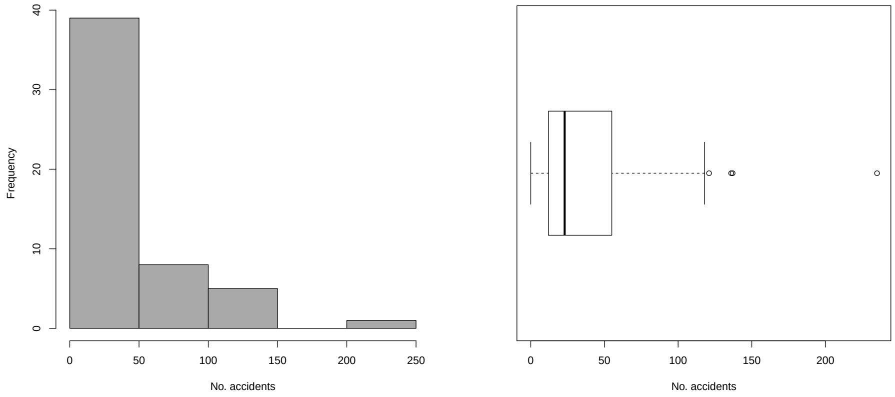

Figure 2.2: Histogram of all road accidents with victims in Lisbon's freguesias in 2007 (left) and corresponding boxplot (right).

amount of traffic, this information is included in the model as an offset so that, in fact, what is modelled is the rate of accidents per traffic unit. The information used on this refers to the maximum number of cars circulating in a certain road section, per hour, in the morning rush hour and is depicted in the right panel of Figure 2.1.

The most significant factors were found to be the *number of hospitals and health centers per freguesia*, the *number of schools per freguesia*, the *proportion of elderly residents*, the *proportion of resident population using car*, the *proportion of resident population working in the same freguesia*, the *proportion of resident population with no school education* and the *monthly housing charges*. Table 2.1 depicts the estimated coefficients of the considered log-Poisson regression model.

Variable	Id	Coefficients
The proportion of resident population **using car**	x_1	-0.0086
The proportion of resident population **working** in the same freguesia	x_2	3.667
The proportion of resident population with **no school education**	x_3	7.544
The number of **hospitals and health centers**	x_4	0.039
The number of **schools**	x_5	0.144
The monthly **housing charges**	x_6	0.0019
The proportion of **elderly** residents	x_7	-2.272
Intercept		-2.805

Table 2.1: Covariates and their estimated coefficients for the log-Poisson model, for the 2007 data.

From these factors, the number of schools and the proportion of resident population working in the same freguesia are the ones most strongly positive correlated with the number of accidents (0.8), followed by the proportion of resident population with no school education (0.56). The proportion of resident population using car (0.46) and the number of hospitals and health centers (0.49), reveals a moderated correlation.

To proceed with the evaluation of spatial dependence for this data, by making use of

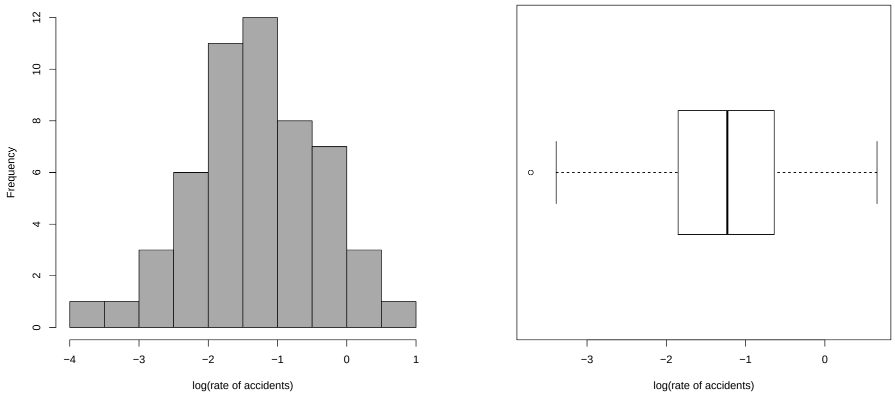

Figure 2.3: Histogram of the transformed variable number of accidents per amount of traffic (left) and corresponding boxplot (right).

Weight matrices	% non-zero weights	No. Neighbours
lisboaqueen	9.8	5.1
lisboaqueen2	17.2	9.1
lisboa5neigh	9.43	5
lisboa.min.Dist	33.3	17.7
lisboa.2min.Dist	65.7	34.8

Table 2.2: Summary measures of the several weight matrices considered.

the tests described in Section 1.2.7 which rely on a Gaussian linear model, it was considered the log transformation of the number of accidents per traffic unit, whose histogram and respective boxplot are presented in Figure 2.3 are quite symmetric. A Gaussian linear model was fitted to the transformed data, but in order to allow comparisons with the Poisson log-linear model fitted before, the covariates had also to be transformed by a factor of $\left(1+\frac{1}{2\times\text{traffic}}\right)$, considering in this particular application that $P(y=0)$ is very small (there are always accidents!) - according to Section 2.1.1.

2.4.3 Spatial Correlation

Spatial dependencies on data were investigated. Given that all spatial analysis are conditional on the choice of the spatial weights matrix, several weight matrices were considered according to the neighbour structure as described in section 1.2.4 - first and second order queen neighbourhood (*lisboaqueen* and *lisboaqueen2*), 5 nearest neighbours (*lisboa5neigh*) - given that 5 is the average number of contiguity neighbours - and distance criteria (*lisboa.min.Dist*, *lisboa.2min.Dist*) - corresponding to the minimum distance necessary to meet some neighbour and twice that distance. The rook neighbourhood structure was not considered because it basically was coincided with the queen one. Table 2.2 summarizes the percentage of non-zero weights and the average number of neighbours corresponding to each matrix.

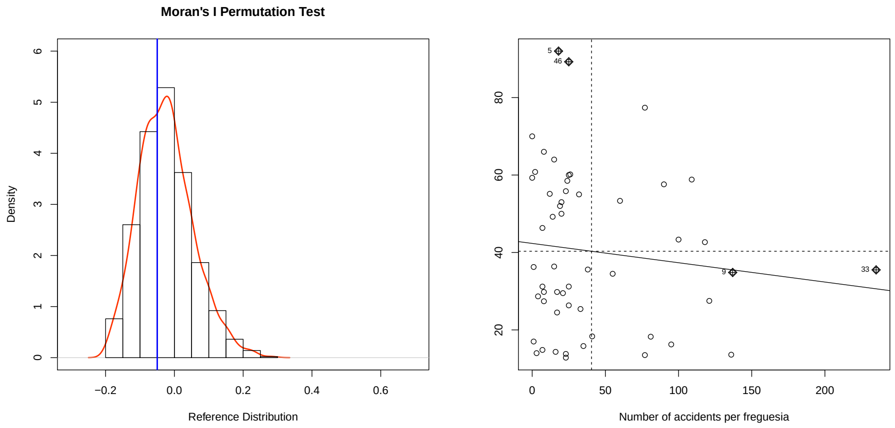

Figure 2.4: Moran's I Permutation Test Plot (left) and Moran's scatter plot (right), for lisboaqueen W matrix.

Weight matrices	Randomization	Normality
lisboaqueen	$I = -0.019$ $(p = 0.9989)$	$I = -0.019$ $(p = 0.9989)$
lisboaqueen2	$I = 0.042$ $(p = 0.3091)$	$I = 0.042$ $(p = 0.3096)$
lisboa5neigh	$I = -0.03$ $(p = 0.8849)$	$I = -0.03$ $(p = 0.885)$
lisboa.min.Dist	$I = 0.042$ $(p = 0.3504)$	$I = 0.042$ $(p = 0.3509)$
lisboa.2min.Dist	$I = -0.017$ $(p = 0.9434)$	$I = -0.017$ $(p = 0.9435)$

Table 2.3: Observed Moran's I and corresponding p-values for two sided test of spatial independence, both considering a randomization distribution for I and under normality.

All the results presented bellow were computed using package spdep [24] of R-project software, according with [6].

Considering first the *lisboaqueen* W matrix, using Moran's I statistic (1.1), under normality ($I = -0.050$, $p = 0.71$) or considering a randomization distribution of the statistic ($I = -0.050$, $p = 0.70$), resulted in a clear non-rejection of the spatial independence hypothesis. Left panel of Figure 2.4 depicts the Moran's I Permutation Test Plot, based on 999 Monte Carlo simulations of the permutation statistics distribution. Here it can be seen that the observed value of the statistic (in blue) is right in the middle of the estimated statistics distribution, leading to the non-rejection of the null hypothesis. This is further confirmed in the right side of Figure 2.4, where the Moran's scatter plot is presented.

For all the other W matrices considered the results were similar and are summarized in Table 2.3.

Although spatial autocorrelation was not found in data, we further investigated the spatial autocorrelation on the residuals of the log-Poisson model fitted in Nunes (2011) [68] and none was found: using a randomization distribution of the statistic, $I = -0.01$

($p = 0.91$), for *lisboaqueen*, $I = -0.10$ ($p = 0.19$) for *lisboaqueen2*, $I = -0.03$ ($p = 0.93$) for *lisboa5neigh*, $I = 0.04$ ($p = 0.35$) for *lisboa.min.Dist* and $I = -0.002$ ($p = 0.49$) for *lisboa.2min.Dist*.

2.4.4 Testing for Spatial Error and Spatial Lag Dependence

The Lagrange Multipliers tests are now used to look for spatial autocorrelation, based on the Gaussian linear model fitted to the log number of accidents per traffic unit, including the transformed covariates according to what is explained in the end of Subsection 2.4.2. The results are summarized in Table 2.4 and indicate the absence of spatial error dependence and spatial lag dependence. Additionally, the Moran's test for spatial dependence of the residuals of the Gaussian linear model is also presented for comparison purposes, but with the same results.

Test \ W matrix	*lisboaqueen*	*lisboaqueen2*	*lisboa5neig*	*lisboaminTrh*	*lisboa2minTrh*
Moran's test	0.014 (p=0.33)	0.028 (p=0.23)	0.004 (p=0.38)	0.059 (p=0.11)	-0.005 (p=0.30)
LM error test	0.025 (p=0.87)	0.012 (p=0.91)	0.067 (p=0.80)	0.053 (p=0.82)	0.078 (p=0.96)
LMR error test	0.177 (p=0.67)	0.003 (p=0.96)	0.222 (p=0.64)	0.048 (p=0.83)	0.225 (p=0.89)
LM lag test	0.002 (p=0.96)	0.566 (p=0.45)	0.363 (p=0.55)	0.927 (p=0.34)	0.929 (p=0.63)
LMR lag test	0.684 (p=0.41)	0.013 (p=0.91)	1.020 (p=0.31)	0.347 (p=0.56)	1.030 (p=0.60)
LM SARMA test	0.019 (p=0.89)	0.334 (p=0.56)	0.002 (p=0.96)	0.318 (p=0.57)	0.337 (p=0.85)

Table 2.4: Results of Lagrange Multipliers tests for spatial dependence and spatial lag dependence and of the Moran's test.

Spatial autocorrelation was not found and this was a bit unexpected. The available tests for spatial autocorrelation heavily depend on a linear model assumption, which is frequently not appropriated. In the next section the alternative of modelling count data with a hierarchical log-Poisson model is considered, including spatial random effects in the risk, in a Bayesian setting.

All the results presented above were computed using package `spdep` [24] of R-project software, according with [6].

2.4.5 Hierarchical log-Poisson models

Several hierarchical log-Poisson models were fitted to the number of accidents in each *freguesia* per traffic unit (included as an offset), in a Bayesian setting, where random effects were included to model both unstructured $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)$ and spatially structured random effects $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)$. A contiguity specification for neighbours was used. The random effects structure is imposed in the prior distribution of the effects as described before. The estimates were obtained via Markov Chain Monte Carlo (MCMC) methods, implemented in R-package `CARBayes` [55]. 8000 MCMC samples were obtained from running a few chains for 100 000 samples, discarding 20 000 burn-in samples and thinning by 10 samples, in order to reduce autocorrelation. For the regression coefficients, β_j weak informative Normal(0,1000) prior distributions are considered. An Inverse-Gamma prior ($a = b = 0.001$, by default however other values can be chosen by the user) is considered for the variance parameter, for both of the unstructured random effects and for the spatially structured random effects. Five models were considered:

- **Model 1:** Log-linear model including the covariates previously considered to be relevant and only unstructured random effects $\boldsymbol{\gamma}$;

$$\begin{aligned} \eta_i &= \boldsymbol{X}_i^T \beta + \gamma_i \\ \gamma_i &\sim \mathrm{N}(0, \sigma^2), \\ \\ \sigma^2 &\sim \text{Inverse-Gamma}(0.001, 0.001), \\ \beta_j &\sim \mathrm{N}(0, 1000) \end{aligned} \tag{2.8}$$

- **Model 2:** Log-linear model with only unstructured $\boldsymbol{\gamma}$ and spatially structured random effects $\boldsymbol{\varepsilon}$ modeled through a BYM CAR prior distribution; no covariates were included;

$$\begin{aligned} \eta_i &= \beta_1 + \gamma_i + \varepsilon_i \\ \gamma_i &\sim \mathrm{N}(0, \sigma^2), \\ \\ \sigma^2 &\sim \text{Inverse-Gamma}(0.001, 0.001), \\ \varepsilon_i | \varepsilon_{-i} &\sim \mathrm{N}\left(\frac{\sum_{j=1}^n w_{ij} \varepsilon_j}{\sum_{j=1}^n w_{ij}}, \frac{\sigma_B^2}{\sum_{j=1}^n w_{ij}} \right) \\ \\ \sigma_B^2 &\sim \text{Inverse-Gamma}(0.001, 0.001) \end{aligned} \tag{2.9}$$

- **Model 3:** Log-linear model with only spatially structured random effects $\boldsymbol{\varepsilon}$ modeled through Leroux CAR prior distribution; no covariates were included;

$$\begin{aligned} \eta_i &= \beta_1 + \varepsilon_i \\ \varepsilon_i | \varepsilon_{-i} &\sim \mathrm{N}\left(\frac{\rho \sum_{j=1}^n w_{ij} \varepsilon_j}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho}, \frac{\sigma_L^2}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho} \right) \\ \\ \sigma_L^2 &\sim \text{Inverse-Gamma}(0.001, 0.001), \\ \\ \rho &\sim \text{Uniform}(0, 1) \end{aligned} \tag{2.10}$$

- **Model 4:** Log-linear model with covariates plus unstructured $\boldsymbol{\gamma}$ and spatially structured random effects $\boldsymbol{\varepsilon}$ modeled through BYM CAR prior distribution;

$$
\begin{aligned}
&\eta_i = \boldsymbol{X}_i^T \beta + \gamma_i + \varepsilon_i \\
&\gamma_i \sim \mathrm{N}(0, \sigma^2), \\
\\
&\sigma^2 \sim \text{Inverse-Gamma}(0.001, 0.001), \\
&\varepsilon_i | \varepsilon_{-i} \sim \mathrm{N}\left(\frac{\sum_{j=1}^n w_{ij} \varepsilon_j}{\sum_{j=1}^n w_{ij}}, \frac{\sigma_B^2}{\sum_{j=1}^n w_{ij}} \right) \\
\\
&\sigma_B^2 \sim \text{Inverse-Gamma}(0.001, 0.001)
\end{aligned}
\tag{2.11}
$$

- **Model 5:** Log-linear model with covariates spatially structured random effects $\boldsymbol{\varepsilon}$ modeled through Leroux CAR prior distribution;

$$
\begin{aligned}
&\eta_i = \boldsymbol{X}_i^T \beta + \varepsilon_i \\
&\varepsilon_i | \varepsilon_{-i} \sim \mathrm{N}\left(\frac{\rho \sum_{j=1}^n w_{ij} \varepsilon_j}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho}, \frac{\sigma_L^2}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho} \right) \\
\\
&\sigma_L^2 \sim \text{Inverse-Gamma}(0.001, 0.001), \\
\\
&\rho \sim \text{Uniform}(0, 1)
\end{aligned}
\tag{2.12}
$$

In general, for most parameters Metropolis-Hastings algorithm acceptance rates were about 60%. MCMC output convergence was assessed through visual inspection of the samples traces, auto-correlation function plots and by the application of the statistics of Gelman-Rubin [37] and of the Raftery and Lewis method [75], implemented in `R`-package `CODA` [72].

Model 1 This model only includes, besides the linear predictor including the covariates *number of hospitals and health centers per freguesia*, the *number of schools per freguesia*, the *proportion of elderly residents*, the *proportion of resident population using car*, the *proportion of resident population working in the same freguesia*, the *proportion of resident population with no school education* and the *monthly housing charges*, unstructured random effects. The results are displayed in Table 2.5.
None of the covariates turned out to be significant, being the data variability captured by the unstructured random effects, indicating that possibly other covariates should be used instead of the ones detected above - Figure 2.5.

Model 2: This model only includes non-structured and spatially structured (BYM CAR prior) random effects not considering covariates. The results are displayed in Table 2.6. The estimates of the relative risks as well as the random effects seem to indicate that there is some spatial variability beneath these data - Figure 2.6.

Variable	Id	median	2.5%	97.5%
Number of **hospitals** and **health centers**	x_1	0.0201	−0.1129	0.1478
Number of **schools** per freguesia	x_2	0.1153	−0.0048	0.2685
Proportion of resident **elderly**	x_3	−2.2357	−8.7354	3.6572
Proportion of resident population **using car**	x_4	−0.0042	−0.0360	0.0266
Proportion of the **resident population** working	x_5	7.6235	−2.9897	18.1601
Proportion of resident population with **no school education**	x_6	12.2284	−6.5278	32.9899
Monthly **housing charges**	x_7	0.0031	−0.0023	0.0089
Intercept		−4.0726	−9.3728	0.8840
σ^2		0.4023	0.2336	0.7432

Table 2.5: Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 1.

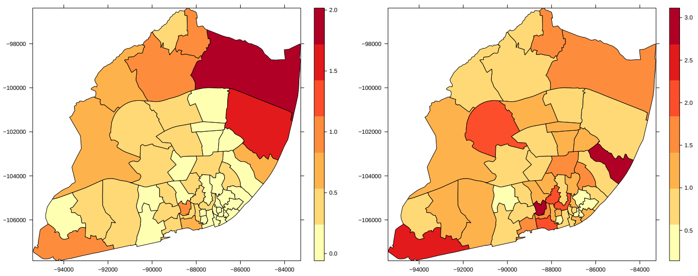

Figure 2.5: Estimated accident relative risk (left) and estimated unstructured random effects (right) for Model 1.

Variable	median	2.5%	97.5%
Intercept	−1.3103	−1.3941	−1.2306
σ_B^2	1.7751	0.4255	3.3452
σ^2	0.0742	0.0005	0.5929

Table 2.6: Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 2.

Model 3: The model fitted here only includes random effects (Leroux CAR prior) not considering covariates. The results are displayed in Table 2.7.
As with the previous model, the estimates of the relative risks as well as the random effects seem to indicate that there is some spatial variability beneath these data, Figure 2.7, which is partially confirmed by an estimated value of ρ of 0.55.

Model 4: This model expands model 2 by adding the covariates to the non-structured and spatially structured (BYM CAR prior) random effects. The results are displayed in Table 2.8.
None of the covariates revealed to be significant, but the random effects display some

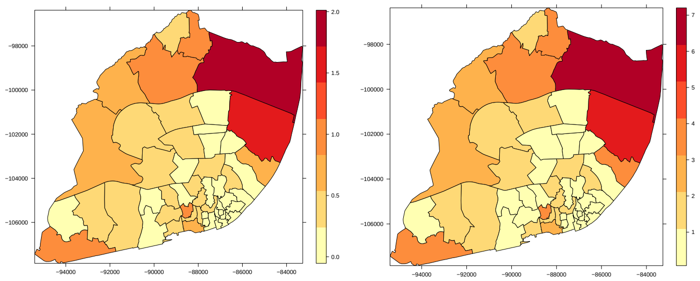

Figure 2.6: Estimated accident relative risk (left), estimated random effects (right) for Model 2.

Variable	median	2.5%	97.5%
Intercept	−1.2977	−1.3826	−1.2185
σ_L^2	1.5647	0.8614	2.6877
ρ	0.550	0.2000	0.8800

Table 2.7: Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 3.

patterns - Figure 2.8.

Variable	Id	median	2.5%	97.5%
Number of **hospitals** and **health centers**	x_1	0.0124	−0.1403	0.3321
Number of **schools** per freguesia	x_2	0.1192	−0.1785	0.2459
Proportion of resident **elderly**	x_3	−3.6337	−15.1426	4.1448
Proportion of resident population **using car**	x_4	−0.0166	−0.0788	0.0305
Proportion of the **resident population** working	x_5	6.5891	−9.5722	29.9274
Proportion of resident population with **no school education**	x_6	5.5374	−18.0464	27.1404
Monthly **housing charges**	x_7	0.0028	−0.0036	0.0094
Intercept		−2.5661	−9.1312	4.1977
σ_B^2		0.6172	0.0008	3.3573
σ^2		0.4013	0.0014	2.5876

Table 2.8: Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 4.

Model 5: This model expands model 3 by adding the covariates to the random effects (Leroux CAR prior). The results are displayed in Table 2.9. None of the covariates revealed to be significant, but the random effects display some patterns - Figure 2.9, with an estimated value of ρ of 0.41.

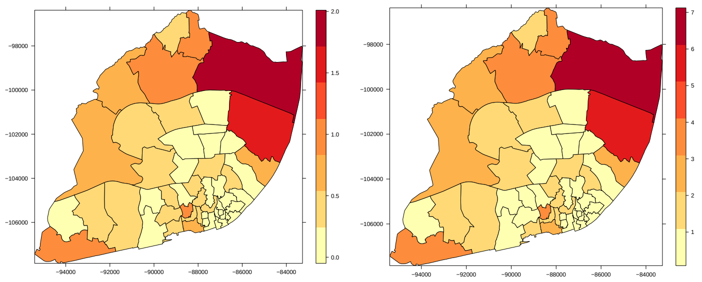

Figure 2.7: Estimated accident relative risk (left) and estimated random effects (right) for Model 3.

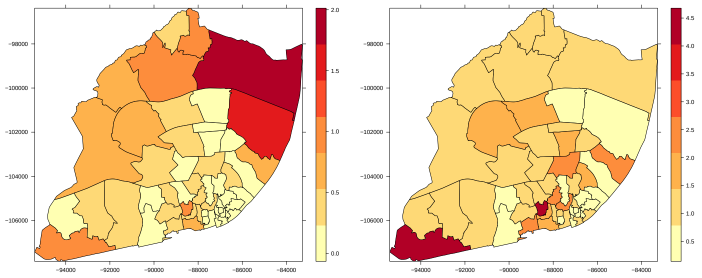

Figure 2.8: Estimated accident relative risk (left), estimated random effects (right) for Model 4.

Variable	Id	median	2.5%	97.5%
Number of **hospitals** and **health centers**	x_1	−0.0140	−0.1683	0.1132
Number of **schools** per freguesia	x_2	0.1418	−0.0095	0.4035
Proportion of resident **elderly**	x_3	−1.9614	−7.3186	5.8148
Proportion of resident population **using car**	x_4	7.1437	−6.6707	24.8661
Proportion of the **resident population** working	x_5	7.1437	−6.6707	24.8661
Proportion of resident population with **no school education**	x_6	6.9096	−12.8641	28.6583
Monthly **housing charges**	x_7	0.0021	−0.0058	0.0095
Intercept		−2.8269	−10.7065	2.4805
σ_L^2		0.9922	0.4152	2.5934
ρ		0.4100	0.0598	0.8700

Table 2.9: Parameter estimates (median, 2.5% and 97.5% quantiles) for Model 5.

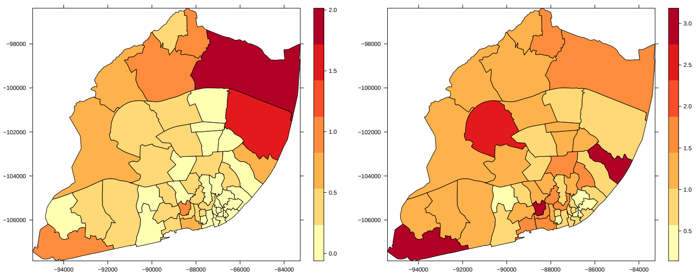

Figure 2.9: Estimated accident relative risk (left) and estimated random effects (right) for Model 5.

2.4.5.1 Results comparison

Models can be compared by means of the *Deviance Information Criterion* (DIC) measure [91], see table 2.10. From here the model with smaller DIC (preferred model) is the one including the covariates and the unstructured and the spatially structured random effects (BYM CAR prior - Model 4). The inclusion of covariates in the models with only random effects improved estimation in both cases, but spatial random effects are still necessary, as they seem to dominate estimation - see how random effects maps resemble. Model 1 only including covariates and unstructured random effect performs worse. This indicates that most probably there are some relevant covariates that are not yet being considered in the model.

Model	DIC	pD
1	363.00	48.19
2	358.17	47.04
3	359.33	47.74
4	355.63	45.11
5	358.15	46.32

Table 2.10: DIC measured for the 5 fitted models.

2.4.6 Main Conclusions

Within the scope of the spatial econometric methods for road accidents with victims in Lisbon, spatial correlation was investigated for data alone and for the residuals of the log-linear model without and with spatial-autocorrelated and spatial-lagged terms, considering transformed data to meet the specificities of the application of these techniques.

In all the cases, spatial-correlation was not found and the addition of spatial structure to the linear model did not improve estimation [84].

This study further comprehends the analysis of only of those accidents that were considered severe - because someone has died as a consequence of the accident within 30 days of its happening. As the results were similar, the corresponding analysis was omitted here.

Given the ongoing analysis and the discrete nature of data, several hierarchical log-Poisson models were further fitted, in a Bayesian setting, implementing a different approach and finding some evidences of spatial structure in data [85].

The hierarchical log-Poisson regression model described before can be well estimated with the INLA procedure, using R-INLA [25], what was further done (not shown), obtaining similar results to the corresponding analysis.

CHAPTER 3

A BAYESIAN SPATIAL AUTOREGRESSIVE MODEL FOR COUNT DATA

Spatial patterns can be modelled differently through autoregressive models, very common in spatial econometrics literature, in which spatial dependence is included in a way such that the value of one observation is dependent on the value of its neighbouring observations [23]. This approach is also valid for count data when it is plausible to think that the space relation between these counts is driven by the effects of covariates, whose values in one area may impact the counts in that area is neighbourhood, even if those variables are not considered in the model [62].

The majority of the classical spatial autoregressive econometric models assume a continuous response variable. However there are alternatives for modelling counts that are explored here.

3.1 Bayesian standard spatial lag model

Consider the first-order spatial autoregressive model on the response with covariates, also known as SLM, presented in section 1.2.6.2, given by

$$\begin{aligned} \boldsymbol{y} &= \rho W \boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \mathrm{N}(0, \sigma^2 I_n). \end{aligned} \tag{3.1}$$

As referred there, spatial contiguity matrix W is usually row-standardized. This model explains the variation on the response $\boldsymbol{y}$ as a linear combination of the response in neighbouring units and some explanatory variables. Parameter ρ is the autoregressive parameter and parameters in vector $\boldsymbol{\beta}$ reflect the influence of the covariates values X on the $\boldsymbol{y}$ variation over the spatial domain. The error term $\boldsymbol{\epsilon}$ is assumed to follow a Normal distribution with zero mean and variance-covariance matrix $\sigma^2 I_n$, where σ^2 is a global variance parameter [7, 60].

The methodology INLA, [81], provides an alternative to the simulation methods for doing Bayesian inference, being based on numerical approximation techniques. It is quite broad in application, just requiring the models to be written in a special but quite general framework, as described in section 1.3.3, where a function g of the expected value of the response variable $\boldsymbol{\mu} = E[\boldsymbol{y}]$ is decomposed into

$$g(\mu_i) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ki} + \sum_{j=1}^{J} f_j \tag{3.2}$$

where f_j are random effects. The vector of latent effects $(\beta_0, \boldsymbol{\beta}, \boldsymbol{f})$ forms a Gaussian Markov Random Field (GMRF).

In practice, there are still models that are not implemented in INLA and in the software, which led Gómez-Rubio, Bivan and Rue, [42], to have recently implemented in R-INLA a new class of models which includes the standard spatial lag model (3.1).

For this particular case of Gaussian models, the spatial lag model (3.1) can be rewritten as

$$\boldsymbol{y} = (I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\epsilon}).$$

The authors implement the expression

$$\boldsymbol{u} = (I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\epsilon})$$

as a random effect that includes, besides the non-linear effect, the intercept and the linear effects of the chosen covariates,

$$g(\mu_i) = u_i = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ki} + \sum_{j=1}^{J} f_{ji},\ i = 1, \ldots, n,$$

where $\boldsymbol{\epsilon}$, related to f_j's, is assumed Normal distributed,

$$\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2 I_n),$$

and where ρ, β and X, W are defined as in the (3.1).

For this model, the prior distributions considered for the vector of parameters $\boldsymbol{\beta}$, to the spatial autoregressive parameter ρ and the precision error term $\tau = \frac{1}{\sigma^2}$, are:

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathrm{N}(0, Q), \\ \mathrm{logit}(\rho) &\sim \mathrm{N}(a, b), \\ \tau &\sim \mathrm{Gamma}(c, d), \end{aligned} \tag{3.3}$$

with Q a precision matrix (that has to be specified).

In the development of this Bayesian spatial lag model, a Gaussian distribution was considered for the response variable $\boldsymbol{y}$, but it is possible to extend this to other distributions due to the broad INLA methodology model's formulation (1.19). The case of a binary response, leading to the estimation of a spatial probit model is proposed in [43] and exemplified in [23]. The case of a Poisson response variable, suitable for counts is proposed and developed in section 3.3.

3.2 A classical Poisson spatial lag model

In this subsection a spatial autoregressive lag model of counts developed by Lambert, Brown and Florax in 2010 [53] under a classical inference framework is described. The spatial autoregressive count model suggested by these authors was motivated by their previous work on estimating temporally lagged count processes. These processes are time series y_t, $t = 0, 1, \ldots,$ with,

$$E(y_t) = \mu_t = \exp(\beta X_t) y_{t-1}^{\rho}. \tag{3.4}$$

It specifies a multiplicative relation between a predetermined count and future outcomes [53]. Their autoregressive model for spatial lagged means, for count responses, specifies a multiplicative relationship between the mean μ_i of the Poisson response y_i in each area and all the means μ_j of the response in its neighbours, similarly to the multiplicative time series model for count data [53].

Consider the non-spatial log-Poisson regression model for a vector of counts $\boldsymbol{y} = (y_1, \ldots, y_n)$ assumed to be Poisson distributed with expected value $E(y_i) = \mu_i$,

$$f(y_i) = \frac{\mu^{y_i} \exp(-\mu_i)}{y_i!}. \tag{3.5}$$

Being $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iK})$ a set of covariates with associated parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ the response expected values is decomposed as

$$E(y_i) = \mu_i = \sum_{k=1}^{K} \exp(\beta_k X_{ik}) = \exp(\boldsymbol{X}_i^T \boldsymbol{\beta}),\ i = 1, \ldots, n. \tag{3.6}$$

Inspired by the model (3.4), Lambert, Brown and Florax developed a spatial autoregressive count process that lays in the specification of the expected mean of counts at location i as a function of its j neighbours, given by,

$$\mu_i = E(y_i) = \exp(\boldsymbol{X}_i^T \boldsymbol{\beta}). \textstyle\prod_{j \neq i}^{n} E(y_j)^{\rho w_{ij}}, \tag{3.7}$$

where w_{ij} are the elements of a weight matrix W and ρ is a spatial autocorrelation parameter. This specification has a multiplicative autoregressive component $\prod_{j \neq i}^{n} (E(y_j))^{\rho w_{ij}}$, added to the non-spatial log-Poisson regression model (3.6). Including that in the exponential part leads to the structural model, written in terms of the predictor $\eta_i = \log(\mu_i)$, as follows

$$\begin{aligned} \mu_i &= \exp(\boldsymbol{X}_i^T \boldsymbol{\beta} + \log(\textstyle\prod_{j=1}^{N} (\mu_j)^{\rho w_{ij}})) \Leftrightarrow \\ \mu_i &= \exp(\boldsymbol{X}_i^T \boldsymbol{\beta} + \rho \textstyle\sum_{j \neq i} w_{ij} \log(\mu_j)) \Leftrightarrow \qquad (3.8) \\ \eta_i &= \boldsymbol{X}_i^T \boldsymbol{\beta} + \rho \textstyle\sum_{j \neq i} w_{ij} \eta_j. \end{aligned}$$

Expressing (3.8) in matrix notation, including all spatial units, leads to the reduced form of the conditional log-mean function

$$\eta = (I_n - \rho W)^{-1}(X\boldsymbol{\beta}), \qquad (3.9)$$

where $(I_n - \rho W)^{-1}$ is called the spatial multiplier term. Inference is done by usual maximum likelihood [53].

3.3 A Bayesian Poisson spatial lag model

Consider a vector of counts $\boldsymbol{y} = (y_1, \ldots, y_n)$ assumed to be Poisson distributed with expected value $E(y_i) = \mu_i$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iK})$ a set of covariates with parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$.

The Lambert, Brown and Florax previous model, can be seen as a GLM model with spatially structured random effects (with log as link function), defined through the relationship

$$\eta_i = \log(\mu_i) = \boldsymbol{X}_i^T \boldsymbol{\beta} + \varepsilon_i, \qquad (3.10)$$

with

$$\varepsilon_i = \rho \sum_{j \neq i} w_{ij} \eta_j. \qquad (3.11)$$

These random effects ε_i include a spatial lag term on the log mean, resulting in a Poisson spatial autoregressive lag model. ρ represents the spatial autoregressive parameter, for a considered weight or adjacency matrix W.

It is proposed now that the spatial lag autoregressive component (3.11) is incorporated in a model for counts, as described in the classical Poisson spatial lag model, in section 3.2, being afterwards the estimation done under the Bayesian paradigm. For this the Bayesian standart spatial model in section 3.1 is adapted, being INLA methodology used for doing inference, under formulation (3.2)

$$\begin{aligned} \boldsymbol{y} &\sim \text{Poisson} \\ \boldsymbol{\mu} &= E[\boldsymbol{y}] \\ \log(\boldsymbol{\mu}) &= \boldsymbol{\eta} = (I_n - \rho W)^{-1}(X\boldsymbol{\beta}). \end{aligned}$$

This construction allows a Bayesian spatial lag model for a Poisson response, considering η as a random effect in the linear predictor, borrowed from the classical spatial lag Poisson model from the previous section 3.1, having

$$\boldsymbol{u} = \boldsymbol{\eta} = (I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\epsilon}),$$

where $\boldsymbol{\epsilon}$ is assumed normal distributed, $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2 I_n)$.

This results in a Bayesian Poisson spatial lag model, an alternative to do Bayesian inference for spatial autoregressive econometric models for count data.

For this model, the prior distributions assigned to the spatial autoregressive parameter ρ and to the precision error term τ are chosen as

$$\begin{aligned} \operatorname{logit}(\rho) &\sim \mathrm{N}(a, b), \\ \tau &\sim \mathrm{Gamma}(c; d). \end{aligned} \tag{3.12}$$

A Normal prior is assigned for the vector of parameters $\boldsymbol{\beta}$ with precision matrix Q,

$$\boldsymbol{\beta} \sim \mathrm{N}(0, Q). \tag{3.13}$$

Different prior distribution can be specified as well as other hyper-parameters.

Note that an offset can be used as a correction factor in the model specification, considering $E(y_i) = \mu_i = e_i \theta_i$, where e_i represents the number of expected cases of what is being measured in each spatial unit $i = 1, \ldots, n$ and θ_i is the relative risk in area i. With this Bayesian Poisson spatial lag model, the risk of what is being counted in one area is related to the risk of what is being counted in the areas of its neighbourhood, driven by effects of important covariables on explaining the phenomenon in one area. The use of this modelling strategy of the autoregressive models, allows to evaluate if the risk of the phenomena that is being counted in a given location may be simultaneously determined by the risk in neighbouring locations. This way of modelling spatial structure for areal data does not ignore the discrete nature of data whenever it applies, incorporating it in the model. In this case, the response variable in a given area is a good predictor of the response variable in its neighbourhood areas, addressing a *global* spatial autocorrelation arising from dependence between counts.

The R-code used in R-INLA to implement this Bayesian Poisson spatial lag model, using the "slpm" function [42, 86], can be found in appendix A.1.

3.4 Application: A Spatial Econometric Analysis of the Calls to the Portuguese National Health Line

One of the most relevant factors regarding hospital costs in the Portuguese health care system are urgency admissions, consuming large financial and human resources. It is possible that a considerable part of the admissions corresponds to non-urgent cases that could be handled by primary health care services, namely the family doctor, or in a self-care basis eventually assisted by a remote nursing service. This helps to understand why the Portuguese hospital urgency service became one of the most important worries of the Portuguese Health Ministry over the last years. The Portuguese national health line, Saúde24 (S24) service directs users to the most appropriate institutions of the public health service or offers counsels on self-care measures. It is hoped that its use mitigates the unnecessary urgent care in hospitals and that the reached savings can be channeled towards other needy areas. This study aims to describe and evaluate the use of S24 by analysing the number of calls received, at a municipal level, under two different spatial econometric approaches. This analysis is important for future development of decision support indicators in a hospital context, based on the economic impact of the use of this health line rather than on the criterion of hospital urgency.

3.4.1 Introduction

An initiative to improve accessibility to health care and to rationalize the use of existing resources was carried out by the Portuguese Health Ministry through the creation of a national health line, S24, in April 2007 [73]. These objectives are accomplished by the S24 service which directs users to the most appropriate institutions of the national public health service or by counseling self-care home measures.

The location attribute for S24 data is an important source of information to describe its use, which leads to analyse the number of calls to S24 at a municipal level. As space is an important feature of these data, ignoring it results in a poorer analysis [11, 29].

To model the number of calls to S24, in each municipality, with spatial models, given the discrete nature of data (counts), an alternative is to use a hierarchical Bayesian model with covariates [15]. For the hierarchical approach spatial autocorrelation is accounted for in the disturbances and not in the observed responses, as happens with spatial autoregressive approaches. The latter is a different modelling strategy common in spatial econometrics literature that may also be considered for these data. It is plausible to think that the number of calls to S24 in one municipality is related to the number of calls in the municipalities of its neighbourhood, driven by effects of covariates such as the number of hospitals in one municipality, which may certainly have an impact on the number of calls to S24 in a neighbouring municipality, or others not considered in the modelling [62]. Hierarchical and autoregressive modelling perspectives have already been used to model the same data sets [23, 43, 74].

This analysis begins with the use of standard spatial econometric techniques to look for spatial dependence in the number of calls to S24 in each municipality, considering a neighbourhood contiguity structure, as well as in the residuals of a baseline log-Poisson regression model with covariates. The number of calls is further analysed, on one hand, through different hierarchical log-Poisson models, and on the other hand, through a Poisson spatial lag model, implementing different econometric approaches to model spatial structure in data. The results of this study are intended to be used in the near future in co-operation with the Portuguese Directorate-General of Health to analyse, test, implement and predict consequences of different government management policies at hospital level under distinct scenarios. The savings from the correct use the S24 will avoid unnecessary urgent care in hospitals that can then be channeled towards other needy areas.

3.4.2 The S24 Data

The data considered in this study were provided by the Support Unit of the Call Center of the National Health Service of the Portuguese Directorate-General of Health (DGS). It is a comprehensive data set of the calls recorded by the S24 health line in the year 2014 and includes information such as user's gender, residence, age, call's day of the week, together with the health problem specification.

The S24 has two call centres and offers various services such as: Triage, Counseling and Routing (TAE); Therapeutic Counseling (AT) to clarify issues relating to medication; Assistance in Public Health (LSP) in specific topics such as flu, heat, poisoning etc.; General Health Information (IGS), such as the location of public health units, pharmacies, among others. The S24 service is provided by qualified nurses, trained to give the best advice or, when appropriate, to assist citizens in solving the situation by themselves. The service is available to the beneficiaries of all different kinds of health sub-systems. The S24 incorporates approximately 300 nurses and 16 clinical supervisors.

After removing inappropriate calls - see Figure 3.1- most of the calls answered by S24, approximately 92%, are catalogued as TAE. For those, the description of the health problem and the original intention of the user about how to solve it (go to an urgency room, for example) are recorded, and then a decision algorithm follows. The final disposition given by this algorithm, jointly with the evaluation of the nurse, falls in one of two possibilities: emergent or not emergent situation. The non emergent situation calls are the ones analyzed in this study - see Figure 3.2.

This study focuses on the number of TAE calls to S24 in 2014 at a municipality level, in Continental Portugal. For this year, 50% of the users were aged between 4 and 46, with a median of 26 years and a range of 111 years. Elderly users are less than 13% - see Figure 3.3. The distribution of the number of TAE calls to S24, by municipality in 2014, is mapped in Figure 3.4. The average raw call rate by municipality is 32 per 1000 inhabitants.

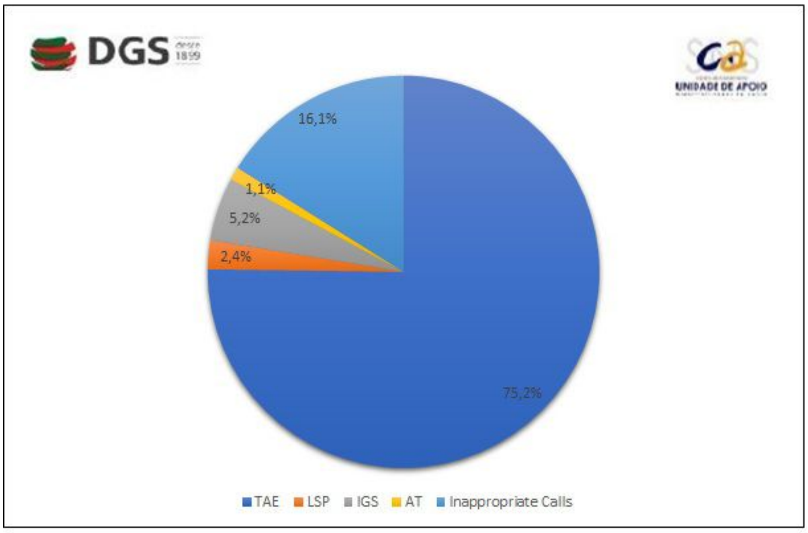

Figure 3.1: Calls to S24 by service in 2014 - Graph provided by DGS.

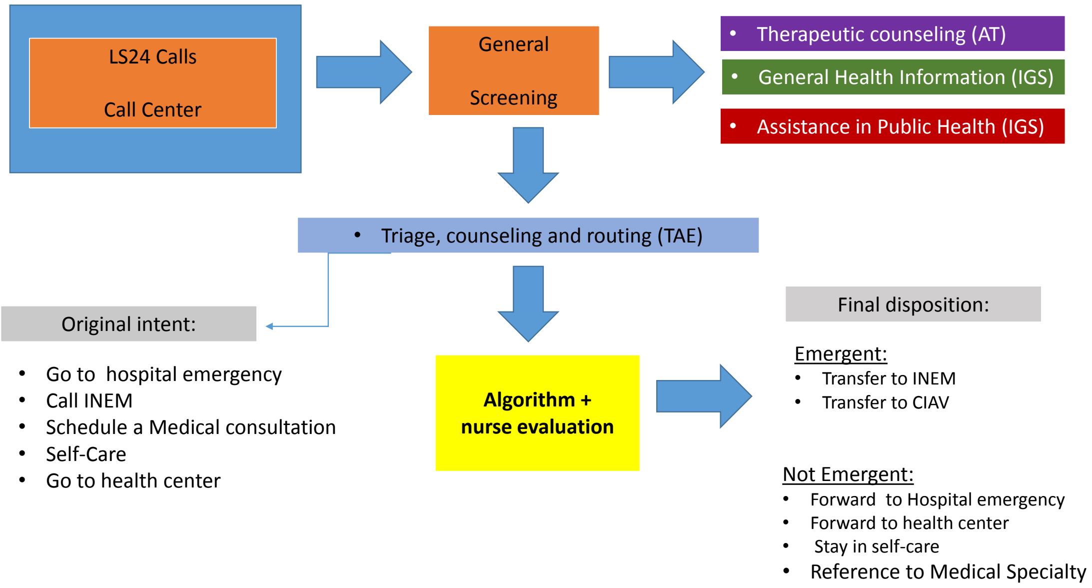

Figure 3.2: The Collection of Information in S24.

3.4.3 Non-spatial modelling: The log-Poisson regression model

The number of TAE calls to S24 in each of the 278 municipalities of Continental Portugal was first modelled via a log-Poisson regression model before considering the need of a spatial analysis.

An indirect standardization of these numbers has been carried out, applied to the

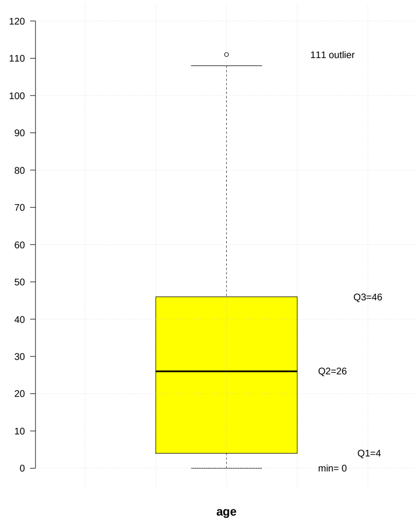

Figure 3.3: The S24 users age empirical distribution in 2014.

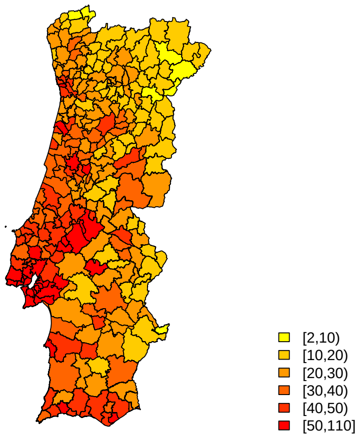

Figure 3.4: Number of TAE calls to S24 per 1000 inhabitants, in 2014.

resident population of each municipality in terms of age groups, namely 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79 and +80. This method considers standard age rates

$$\varphi_j = \frac{\sum_i y_{ij}}{\sum_i n_{ij}}, j = 1, ..., 9,$$

with y_{ij} the number of cases (calls) and n_{ij} the at risk population (resident population), in municipality i and age group j, with $i = 1, ..., 278$, with $j = 1, ..., 9$, in order to obtain $e_i = \sum_i n_{ij}\varphi_j$, $i = 1, ..., 278$, the expected number of calls in each municipality, that is included in the model as an offset. So, in fact, what is modelled is the relative call risk,

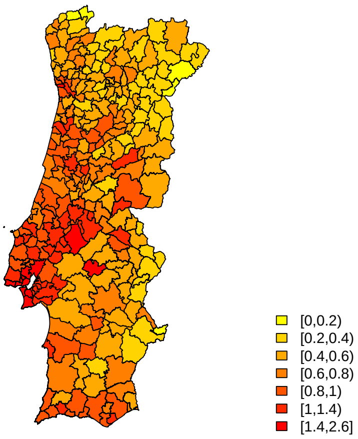

Figure 3.5: Standard Call Rate to S24, in 2014.

which can be roughly estimated by the Standard Call Rate (SCR), mapped in Figure 3.5. This ratio is calculated between the observed number of cases and the expected number of cases, allowing comparisons across different populations,

$$SCR_i = \frac{y_i}{e_i}, i = 1, \dots, 278.$$

The resident population of each municipality, in terms of age groups, was obtained from Census 2011 data and adjusted for subsequent years [32].

Demographic and socio-economic information, development indicators as well as characteristics of the Portuguese health system at the municipal level, were investigated as possible covariates for modelling the TAE call counts, in order to understand if the inclusion of certain covariates obviated the need for a spatial model. Using the Stepwise methodology [76] for selecting covariates, under different scenarios, the two best sets of the most significant explanatory variables are:

> Case 1: The *average number of years of schooling*, the *proportion of elderly residents*, the *unemployment rate*, the *rurality index*, the *number of hospitals and health centres* per 1000 inhabitants and the *proportion of women*, in each municipality (AIC: 29530);
>
> Case 2: The *monthly average income*, the *proportion of children*, the *unemployment rate*, the *rurality index*, the *number of hospital and health centres* (both per 1000 inhabitants), and the *proportion of women*, in each municipality (AIC: 36980).

From these variables, the average number of years of schooling and the monthly average income are the ones that show a stronger positive correlation with the response variable (0.67 and 0.61, respectively), followed by the proportion of children (0.49). The rurality index and the proportion of elderly residents are negatively correlated with the response (-0.45 and -0.35, respectively).

Over-dispersion in these Poisson data is expected, since it is suspected that space to be an important feature for their modelling. If this over-dispersion is ignored, the standard errors of the covariate effects are underestimated, resulting in an incorrect assessment of the significance of individual regression parameters. So, instead, it has been opted to fit a quasi-Poisson model to account for the over-dispersion, realizing that the significant covariates under this approach were in fact different from the ones of the Poisson model (although the estimated effects are, of course, the same).

Tables 3.1 and 3.2 depict the estimated coefficients of the considered quasi-Poisson log-regression models for these analyses, with

$$\log(\theta_i) = \beta_0 + \sum_{j=1}^{7} \beta_j X_{ij},\ i = 1, \ldots, 278,$$

where θ_i is the relative risk in the ith municipality. For case 1, the unemployment rate turned out to be not significant after all, and for case 2, the same happened with the rurality index, the number of hospital and health centres.

Variable	Id	Coefficients	p-values
Average number of years of **schooling**	x_1	0.322	<2e-16
Proportion of **elderly** residents	x_2	4.456	9.52e-13
Unemployment rate	x_3	−0.743	0.3156
Rurality index	x_4	−0.741	4.10e-06
Number of **hospitals**	x_5	−3.822	6.68e-06
Number of **health centres**	x_6	−1.289	0.0437
Proportion of **women**	x_7	−5.509	0.0661
Intercept		−0.288	0.8390

Table 3.1: Covariates and their estimated coefficients for the quasi-Poisson log-regression model, case 1, for the S24 2014 data.

Variable	Id	Coefficients	p-values
The monthly average **income**	x_1	0.001	5.97e-16
Proportion of **children**	x_2	5.727	0.0004
Unemployment rate	x_3	−1.679	0.0391
Rurality index	x_4	−0.212	0.1884
Number of **hospitals**	x_5	−0.398	0.6504
Number of **health centres**	x_6	−0.902	0.1702
Proportion of **women**	x_7	11.810	0.0003
Intercept		−7.930	7.64e-06

Table 3.2: Covariates and their estimated coefficients for the quasi-Poisson log-regression model, case 2, for the S24 2014 data.

Package `stats` of R-project software was used to obtain the results presented in this section [96].

3.4.4 Spatial correlation

In this subsection, standard spatial techniques are used to look for spatial dependence in the number of TAE calls, considering a contiguity neighbourhood structure, and also in the residuals of the log-Poisson regression models fitted before.

For the considered contiguity neighbourhood structure, the first order queen neighbourhood, there are 1.9% non-zero weights and the average number of neighbours is 5.3. Taking the corresponding queen neighbourhood matrix, and using Moran's I statistics (1.1) both under normality ($I = 0.6182$, $p =< 2.2\text{e-}16$) or considering a randomized distribution of the statistics ($I = 0.6182$, $p =< 2.2\text{e-}16$), resulted in a clear rejection of the spatial independence hypothesis of the number of TAE calls, suggesting that there is a positive spatial correlation among these.

The spatial autocorrelation in the residuals of the log-Poisson regression models fitted in section 3.4.3 was further investigated, using a randomized distribution of the statistic and a two sided test, having $I = 0.1513$ ($p = 1.102\text{e-}05$) for case 1 and $I = 0.2702$ ($p = 2.276\text{e-}14$) for case 2. The results suggest a high positive spatial autocorrelation in the residuals. With spatially correlated residuals, the fitted models may be providing biased estimates of the parameters, leading to incorrect interpretations and misleading conclusions [60]. It is then clear that space is an important feature of these data and that must be considered in the modelling.

Package `spdep` [24] of `R`-project software was used to obtain the results presented in this section according to [6].

3.4.5 Spatial modelling

3.4.5.1 Spatial hierarchical log-Poisson regression model

In order to capture and model data spatial variability, the number of TAE calls in each municipality is now analysed through different spatial hierarchical log-Poisson regression models. The residual autocorrelation of the log-Poisson regression model considered before can be explained, in a Bayesian setting, adding to the model's predictor a set of spatially structured $\boldsymbol{\varepsilon}$ random effects, considering the contiguity neighbourhood structure mentioned before. Additional unstructured random effects $\boldsymbol{\gamma}$ can be considered, if needed. The prior distributions of the random effects define their structure, as described in section 2.2. Two models were considered differing on the way the random effects are included, the BYM and the Leroux models.

The estimates were obtained via Markov Chain Monte Carlo (MCMC) method, implemented in `R`-package `CARBayes` [55]. A few MCMC run of 1 000 000 iterations were made, discarding 50 000 burn-in iterations and thinning by 100, in order to reduce autocorrelation, resulting in 9500 sample points.

In general for most parameters acceptance rates of the Metropolis-Hastings algorithm were about 40%. MCMC output convergence was assessed through visual inspection of the samples traces, auto-correlation function plots and by the application of the Geweke method [41] available in R-package `CARBayes` [55].

BYM model The Besag-York-Mollié (BYM) model, as described in section 2.2, for both cases 1 and 2 as in sub-section 3.4.3, is a log-Poisson regression model with the covariates considered before plus unstructured ($\boldsymbol{\gamma}$) and spatially structured random effects ($\boldsymbol{\varepsilon}$), for which a CAR prior is chosen. The main parameter estimates are summarised in Tables 3.3 and 3.4.

For case 1 only one of the covariates, the average number of years of schooling, showed to be significant, whereas in case 2 it was the monthly average income. The estimated random effects, given by $\exp(u_i) = \exp(\varepsilon_i + \gamma_i)$, still display some patterns for both cases - left panels of Figure 3.6 and Figure 3.7.

Variable	Id	median	2.5%	97.5%
Average number of years of **schooling**	x_1	0.1931	0.0062	3.5386
Proportion of **elderly** residents	x_2	0.4840	−1.6883	2.6921
Unemployment rate	x_3	1.8680	−1.5338	4.7276
Rurality index	x_4	−0.0930	−0.4980	0.3126
Number of **hospitals**	x_5	−0.2549	−1.9709	1.4916
Number of **health centres**	x_6	0.3777	−1.1932	1.8507
Proportion of **women**	x_7	−1.2479	−10.1867	7.7633
Intercept		−1.3506	−6.4541	3.5386
σ_B^2		0.2268	0.1440	0.3494
σ^2		0.0326	0.0140	0.0592

Table 3.3: Parameter estimates (median, 2.5% and 97.5% quantiles) for the BYM hierarchical log-Poisson model, case 1, for the S24 2014 data.

Variable	Id	median	2.5%	97.5%
The monthly average **income**	x_1	0.001	0.0	0.0021
Proportion of **children**	x_2	1.7348	−2.1687	6.3659
Unemployment rate	x_3	1.9042	−1.5689	5.4343
Rurality index	x_4	−0.1838	−0.5460	0.2105
Number of **hospitals**	x_5	−0.2282	−2.0447	1.3542
Number of **health centres**	x_6	−0.0613	−1.3820	1.3585
Proportion of **women**	x_7	0.6238	−8.9309	9.4824
Intercept		−1.9574	−7.2782	3.4761
σ_B^2		0.2443	0.1594	0.3690
σ^2		0.0313	0.0149	0.0540

Table 3.4: Parameter estimates (median, 2.5% and 97.5% quantiles) for the BYM hierarchical log-Poisson model, case 2, for the S24 2014 data.

Leroux model The Leroux model, as described in section 2.2, is a log-Poisson regression model with the covariates previously considered and the random effects for which the Leroux CAR prior is chosen. The main parameter estimates are displayed in Table 3.5 for case 1 and in Table 3.6 for case 2.

Here, for the first case, only one of the initial covariates showed to be significant, the average number of years of schooling. The estimates of the random effects, given by $\exp(\varepsilon_i)$, seem to indicate that there still is spatial variability in these data - right panel of Figure 3.6, which is strongly confirmed by an estimated value of ρ of 0.90.

Considering this model for the second case, also only one of the initial covariates was significant, the monthly average income. This model has an estimated value of ρ of 0.89, and the estimates of the random effects seem to indicate that there still is spatial variability - right panel of Figure 3.7.

Variable	Id	median	2.5%	97.5%
Average number of years of **schooling**	x_1	0.1897	0.0141	0.3187
Proportion of **elderly** residents	x_2	0.8586	−1.3888	2.8322
Unemployment rate	x_3	2.1413	−0.7070	4.8630
Rurality index	x_4	−0.1129	−0.4562	0.2337
Number of **hospitals**	x_5	−0.2606	−1.7421	1.1745
Number of **health centres**	x_6	0.3458	−0.8881	1.6717
Proportion of **women**	x_7	−1.9482	−9.8121	6.2431
Intercept		−1.1342	−5.1422	3.1621
σ_L^2		0.3492	0.2829	0.4494
ρ		0.9059	0.7008	0.9888

Table 3.5: Parameter estimates (median, 2.5% and 97.5% quantiles) for the Leroux hierarchical log-Poisson model, case 1, for the S24 2014 data.

Variable	Id	median	2.5%	97.5%
The monthly average **income**	x_1	0.001	0.0001	0.0019
Proportion of **children**	x_2	1.8345	−1.9861	5.7947
Unemployment rate	x_3	1.6751	−1.3984	4.7253
Rurality index	x_4	−0.2145	−0.5562	0.0639
Number of **hospitals**	x_5	−0.3568	−1.8623	1.1085
Number of **health centres**	x_6	−0.0069	−1.2701	1.2484
Proportion of **women**	x_7	0.5789	−7.0199	8.6312
Intercept		−1.9231	−6.5246	2.4100
σ_L^2		0.3581	0.2911	0.4508
ρ		0.8936	0.6879	0.9855

Table 3.6: Parameter estimates (median, 2.5% and 97.5% quantiles) for the Leroux hierarchical log-Poisson model, case 2, for the S24 2014 data.

3.4.5.2 The Bayesian Poisson spatial lag model

A modelling alternative is to account for spatial autocorrelation in the observed responses instead of the disturbances, as before, using an autoregressive perspective. This approach may also be considered for these data.

Here, the TAE number of calls in each municipality is then analysed through the Bayesian Poisson spatial lag model where a spatial autocorrelation lag is incorporated in the econometric model of counts. The estimates were obtained via INLA methodology in R-package `R-INLA`, according to the R-code presented in appendix A.1. The prior

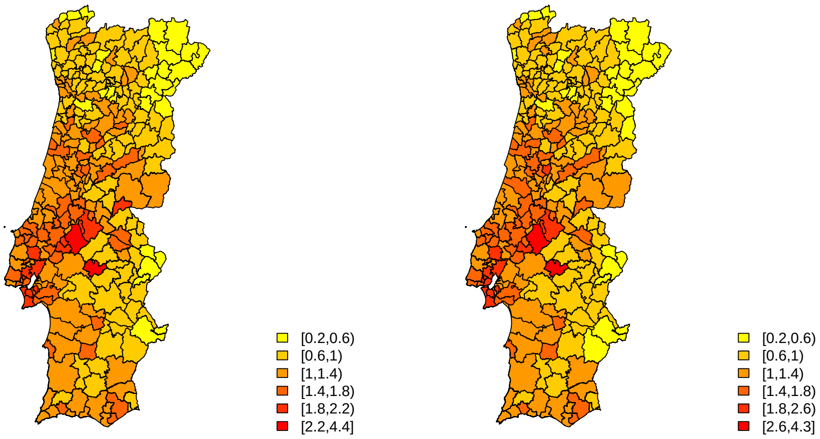

Figure 3.6: Estimated random effects for BYM model(left) and for Leroux model (right), case 1.

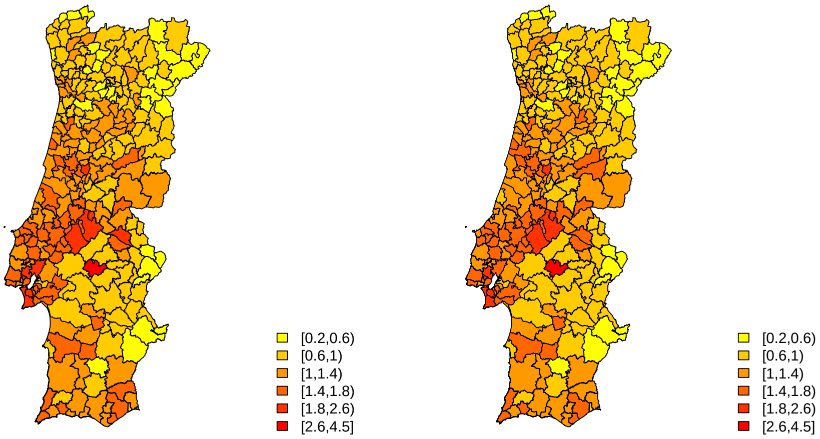

Figure 3.7: Estimated random effects for BYM model (left) and for Leroux model (right), case 2.

distributions assigned to the spatial autoregressive parameter ρ and to the precision error term τ are, by default, $\text{logit}(\rho) \sim N(0, 10)$ and $\tau \sim \text{Gamma}(1; 5 \times 10^{-5})$, however other values can be chosen by the user.

Poisson spatial lag model This is the Bayesian Poisson spatial lag autoregressive model with the covariates initially considered significant. Table 3.7 and Table 3.8 summarize the main parameter estimates for case 1 and case 2, respectively.

For case 1 only one of the previous covariates revealed to be significant, the average number of years of schooling. This model has an estimated value of ρ of 0.852. As for the second case, only the monthly average income is significant. This second model has an estimated value of ρ of 0.859. The estimated random effects, given by $\exp(\varepsilon_i)$, still display some patterns for both cases - Figure 3.8 for case 1 and Figure 3.9 for case 2.

Variable	Id	mean	2.5%	97.5%
Average number of years of **schooling**	x_1	0.179	0.121	0.237
Proportion of **elderly** residents	x_2	0.591	−0.288	1.473
Unemployment rate	x_3	0.605	−0.851	2.048
Rurality index	x_4	−0.005	−0.209	0.197
Number of **hospitals**	x_5	−0.179	−1.300	0.941
Number of **health centres**	x_6	0.173	−0.316	0.660
Proportion of **women**	x_7	−0.919	−4.702	2.848
Intercept		−1.071	−3.012	0.868
σ^2		0.070	0.568	0.085
ρ		0.852	0.806	0.893

Table 3.7: Parameter estimates (mean, 2.5% and 97.5% quantiles) for the Spatial lag Poisson model, case 1, for the S24 2014 data.

Variable	Id	mean	2.5%	97.5%
The monthly average **income**	x_1	0.001	0.000	0.001
Proportion of **children**	x_2	0.972	−1.232	3.190
Unemployment rate	x_3	0.065	−1.408	1.520
Rurality index	x_4	−0.098	−0.291	0.094
Number of **hospitals**	x_5	0.151	−0.989	1.290
Number of **health centres**	x_6	−0.095	−0.560	0.369
Proportion of **women**	x_7	0.307	−3.570	4.182
Intercept		−1.015	−3.180	1.141
σ^2		0.074	0.061	0.089
ρ		0.859	0.813	0.90

Table 3.8: Parameter estimates (mean, 2.5% and 97.5% quantiles) for the Spatial lag Poisson model, case 2, for the S24 2014 data.

3.4.6 Comparison of results

In the various spatial fits the covariates considered important for explaining the number of calls and the corresponding effects were the same. These fits were further compared by means of their predictive accuracy, using the *Deviance Information Criterion* (DIC)

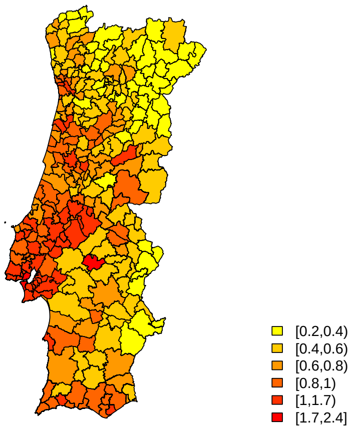

Figure 3.8: Estimated random effects for the Poisson spatial lag model, case 1.

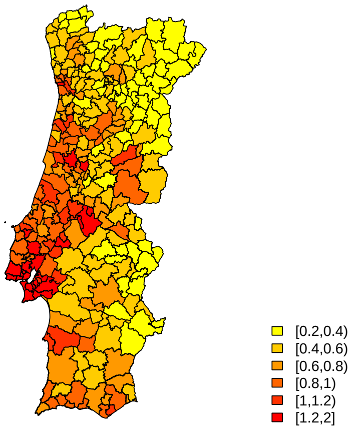

Figure 3.9: Estimated random effects for the Poisson spatial lag model, case 2.

measure and the *Watanabe-Akaike Information Criterion* (WAIC) measure. See Table 3.9 for case 1 and Table 3.10 for case 2. The *Relative Root Mean Square Error* (RRMSE) was also considered to measure goodness-of-fit:

$$RRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \widehat{y}_i)^2}{\widehat{y}_i^2}}.$$

Results are displayed in tables 3.11 and 3.12.

In terms of spatial hierarchical log-Poisson regression models, the model with smaller DIC (preferred model) is the one including the covariates and the spatially structured random effects through Leroux CAR prior. This was confirmed by the RRMSE values. For the sake of comparison, the fit measures for the baseline log-Poisson regression model without random effects, fitted by MCMC, are further displayed in the first line of the tables. The log-Poisson regression model was also fitted, including only covariates and unstructured random effects (results not shown here), which performed worse, indicating that spatial random effects are indeed necessary in the models. This might indicate that there are possibly some relevant covariates that are not yet being included in the model. There is a spatial asymmetry that is not explained by the variables. Similar conclusions were reached when the autoregressive perspective was considered in terms of the Bayesian Poisson spatial lag model.

In order to compare both hierarchical and autoregressive model fits, WAIC measure was used, as it is more appropriate for comparing different model structures. The autoregressive model reveals better performance, according to this measure. As for the RRMSE values, they are very similar although they are somewhat smaller for the hierarchical models.

Model	DIC	pD	WAIC	pW
Baseline Model MCMC	2816.9	287.2	2830.8	198.5
BYM model	2801.1	275.2	2773.8	179.9
Leroux model	2788.6	267.16	2744.5	157.2
Poisson spatial lag model	2778.63	261.96	2717.6	144.9

Table 3.9: DIC and WAIC measured for the 3 models fitted for case 1.

Model	DIC	pD	WAIC	pW
Baseline Model MCMC	2816.9	287.2	2830.8	198.5
BYM model	2800.0	275.6	2770.7	169.5
Leroux model	2793.2	272.4	2743.0	156.7
Poisson spatial lag model	2777.3	262.6	2714.1	143.8

Table 3.10: DIC and WAIC measured for the 3 models fitted for case 2.

3.4.7 Main conclusions

This application study combines insights from classical spatial econometrics and the analysis of spatial data in order to handle spatial count data, both in a spatial hierarchical and

Model	RRMSE
Baseline Model	0.588
BYM model	0.029
Leroux model	0.028
Poisson spatial lag model	0.037

Table 3.11: RRMSE measured for the 3 models fitted for case 1.

Model	RRMSE
Baseline Model	0.469
BYM model	0.025
Leroux model	0.026
Poisson spatial lag model	0.034

Table 3.12: RRMSE measured for the 3 models fitted for case 2.

in a spatial autoregressive perspectives. The approach applied here allows the limitations of the classical econometrics methods to be circumvented.

Within the scope of the spatial econometric methods and also resorting to Bayesian hierarchical and autoregressive methodology, their application to study of the number of TAE calls to the national health line S24 revealed spatial-correlation and the addition of spatial structure in the models improved estimation.

In this study, the count data were first analysed with a log-Poisson regression model and then the inclusion of spatial random effects in a hierarchical Bayesian setting proved to be relevant, as expected, being the preferred model the one including the covariates and the spatially structured random effects through Leroux CAR prior distribution. However the modelling may perhaps be improved by considering some other more adequate covariates. Aditionally, a Bayesian Poisson spatial lag model was developed and implemented, an alternative to do Bayesian inference for spatial econometric models for count data. Similar conclusions were drawn when both the hierarchical and the autoregressive perspectives were considered.

The average number of years of schooling for case 1 of the analysis and the average monthly income for case 2 stand out as being important in explaining the use of S24. The spatial component for both cases was quite relevant, which was confirmed by the high values of the estimates of the spatial autocorrelation parameter.

It is intended to proceed with this application study of S24 data set in order to be able to describe and evaluate in which municipalities the use of S24 should be encouraged, as well as detecting those regions that most contribute to the economic success of the good use of the line for future assessment of hospital savings [50].

Additionally, this analysis will be extended to include data available for the years between 2010 and 2016, fitting some spatio-temporal models [30] under an econometric approach and developing and implementing the temporal effects on Bayesian hierarchical models [25, 56], or on Bayesian autoregressive models [25] for count data.

CHAPTER 4

SPATIO-TEMPORAL BAYESIAN MODELS FOR COUNT DATA

This chapter describes and explains different Bayesian spatio-temporal models for Poisson count data, that account for spatial and temporal correlations through random effects. For spatial correlation, the models presented here use either hierarchical conditional autoregressive prior distributions [15, 59] or consider an autoregressive approach in space [23]. They further include time effects modelled simply through a parametric linear trend model [17] or through a nonparametric dynamic trend model [25, 52]. Different space-time structures are investigated under both hierarchical and autoregressive perspectives, resorting to Integrated Nested Laplace Approximation (INLA) methodology, resulting on Bayesian hierarchical Poisson models and a novel Bayesian Poisson spatio-temporal lag model, that is newly developed in this chapter. A common mathematical framework for both approaches, is persuaded and detailed.

4.1 General log-Poisson regression model

Considering the study region divided into a set of n spatial units let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})$ and $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{nt})$ represent, respectively, the number of observed and expected cases of what is being counted in each spatial unit $i = 1, \ldots, n$ and recorded for $t = 1, \ldots, T$ consecutive time periods. The counts y_{it} are assumed to be Poisson distributed with expected value $E(y_{it}) = \mu_{it} = e_{it}\theta_{it}$, where θ_{it} is the relative risk in area i and time t.

The general log-Poisson regression model is defined as:

$$y_{it}|\eta_{it} \sim \text{Poisson}(e_{it}\theta_{it}), \tag{4.1}$$

where $\eta_{it} = \log(\theta_{it})$, $i = 1, \ldots, n$, $t = 1, \ldots, T$, are the log relative risks. Note that $\log(e_{it})$ enter as known offsets in the model. The log relative risks are decomposed into the effects of covariates $\boldsymbol{X}_{it} = (X_{it1}, \ldots, X_{itk})$ measured for spatial unit i and time period t, corresponding

X_{it1} to the intercept term and being $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ the associated regression parameters, plus some random effects that are able to account for possible over-dispersion

$$\eta_{it} = \log(\theta_{it}) = \mathbf{X}_{it}^T \boldsymbol{\beta} + u_{it}. \tag{4.2}$$

The u_{it} term is a latent component for spatial unit i and time period t which may capture remaining space, time and spatio-temporal effects or trends in the data, after accounting for the covariates. Different configurations of u_{it} correspond to different models.

For the regression coefficients β_j weak informative Normal$(\mu_\beta, \sigma_\beta^2)$ prior distributions are considered.

4.1.1 Spatial component

Under formulation (4.2), it is possible to have a hierarchical log-Poisson model if u_{it} includes spatially autocorrelated random effects ε_i, modelled by a conditional autoregressive prior distribution. Alternatively, an autoregressive lag Poisson model is obtained if u_{it} includes instead a spatial lag term ε_{it} on the log relative risks. These choices are detailed below. Note that in each approach the same notation ε is used for different things, identified by the corresponding approach.

4.1.1.1 Bayesian hierarchical log-Poisson model

Bayesian hierarchical models for count data frequently account space correlation in data by using Conditional Autoregressive (CAR) prior distributions for spatial random effects [15]. From the existing possibilities the one adopted here is the CAR prior distribution proposed by Leroux *et al.* [59], being necessary to establish a neighbouring criterion between spatial units, through a weight or adjacency matrix W [29].

For the set of spatial random effects $(\varepsilon_1, \ldots, \varepsilon_n)$ the referred CAR prior distribution follows:

$$\varepsilon_i | \varepsilon_{-i} \sim \mathrm{N}\left(\frac{\rho_S \sum_{j=1}^n w_{ij} \varepsilon_j}{\rho_S \sum_{j=1}^n w_{ij} + 1 - \rho_S}, \frac{\sigma_S^2}{\rho_S \sum_{j=1}^n w_{ij} + 1 - \rho_S} \right), \tag{4.3}$$

where ρ_S represents the spatial dependence parameter, which takes the value zero in case of independence and values near one for strong spatial correlation. A non informative uniform prior distribution on the unit interval is specified for this parameter. An inverse-gamma prior distribution is specified for the variance of the random effects σ_S^2, [25].

4.1.1.2 Bayesian autoregressive log-Poisson model

In the formulation of the Bayesian Poisson spatial lag model, proposed as an alternative spatial autoregressive econometric model for count data [86] and adapted here for spatio-temporal data, spatial dependencies are accounted in the observed responses through a spatial lag term ε_{it} on the log relative risks, to be included in u_{it} and given by

$$\varepsilon_{it} \quad = \quad \rho \sum_{j \neq i} w_{ij} \eta_{jt}, \tag{4.4}$$

where η_{it} is the log relative risk for unit area i and time t and ρ is the spatial autoregressive parameter. For the logit transformation of parameter ρ, $logit(\rho)$, a Normal prior distribution is considered. The log-risks η_{jt} will be modelled also with temporal and spatio-temporal effects, justifying the dependency of the quantities ε_{it} on time.

It is also considered here the weight or adjacency matrix W defined previously by the contiguity criterion between areas [29].

4.1.2 Spatio-temporal component

Temporal effects may also added into the u_{it} component, in a similar way for both hierarchical and autoregressive approaches. These effects are going to be considered in two different ways, under a simplicity argument. First adopting a parametric trend for the temporal component as proposed by Bernardelli *et al.* [17], that includes in u_{it} a separate linear trend for each area. Secondly considering a nonparametric dynamic trend as proposed by Knorr-Held [52] for the temporal component, that includes in u_{it} a temporally structured temporal effect and an unstructured space-time interaction effect.

4.1.2.1 Linear trend model

This model incorporates the spatio-temporal variation additively in u_{it} in (4.2) as a separate linear trend in time for each area:

$$(\alpha + \delta_i)t^*, \tag{4.5}$$

where t^* corresponds to the time normalized to belong to an unit interval. The spatial unit i has then its own linear time trend with a spatially varying intercept $(\beta_1 + \varepsilon_i)$ for the hierarchical approach and $(\beta_1 + \varepsilon_{it})$ for the autoregressive approach, and a spatially varying slope $(\alpha + \delta_i)$ for both approaches. The differential trend δ_i represents the difference between the global trend α and the spatial unit specific trend. If $\delta_i > 0$ the trend for the corresponding spatial unit is steeper than the mean trend while it $\delta_i < 0$ is the other way around.

This specification is similar to the spatially varying linear time trend model proposed by Bernardinelli *et al.* [17] for spatio-temporal hierarchical models. It considers spatio-temporal interactions δ_i differentiated for each spatial unit.

It is assumed for each element of $\boldsymbol{\delta} = (\delta_1, ..., \delta_n)$ and for the overall slope parameter α independent Gaussian prior distributions,

$$\begin{aligned} \delta_i &\sim \mathrm{N}(0, \sigma^2_{slo}) \\ \alpha &\sim \mathrm{N}(\mu_\alpha, \sigma^2_\alpha) \end{aligned} \tag{4.6}$$

For σ^2_{slo} an inverse-gamma prior distribution is considered [25]. The corresponding hyperparameters $(a, b, \mu_\alpha, \sigma^2_\alpha)$ are chosen in this work such that the resulting prior distributions are weakly informative.

The assumption of linearity of the δ_i component can be relaxed, using instead a dynamic nonparametric formulation of the temporal component as follows.

4.1.2.2 Dynamic trend model

This model is similar to the simplest form of the one proposed by Knorr-Held [52] for spatio-temporal hierarchical models (although there are other more complex alternatives that could also have been explored), and it incorporates additively in u_{it} in (4.2) temporal main effects δ_t as well as space-time interactions γ_{it}:

$$\delta_t + \gamma_{it} \tag{4.7}$$

Within this model, spatio-temporal variation is decomposed into three components: an overall spatial effect common to all time periods ε_i for the hierarchical approach and specific year spatial effects ε_{it} for the autoregressive approach; an overall temporal trend common to all spatial units δ_t modelled dynamically by a random walk of order 1, [25]; an unstructured space-time random effect.

Denoting $\boldsymbol{\delta} = (\delta_1, ..., \delta_T)$ the set of temporal random effects, they are modelled through a random walk of order 1:

$$\delta_t - \delta_{t+1} \sim \mathrm{N}\left(0, \sigma^2_T\right) \tag{4.8}$$

For the independent space-time interactions $\boldsymbol{\gamma} = (\gamma_{11}, ..., \gamma_{nT})$ Gaussian prior distributions are specified,

$$\gamma_{it} \sim \mathrm{N}\left(0, \sigma^2_I\right) \tag{4.9}$$

For σ^2_I an inverse-gamma prior distribution is considered and for $\log(\sigma^{-2}_T)$ a log-gamma prior distribution is considered.

4.2 Spatio-temporal Bayesian hierarchical log-Poisson model

Spatio-temporal extensions of the Bayesian hierarchical log-Poisson model are then possible, defined as the spatio-temporal Leroux linear model (ST.Leroux.linear) when using a linear temporal trend formulation and as the spatio-temporal Leroux dynamic model (ST.Leroux.dynamic) for the dynamic temporal trend specification.

4.2.1 Spatio-temporal Leroux linear model

The model ST.Leroux.linear includes a spatial effects component ε_i, a main linear trend α which represents the global time effect and a differential time trend δ_i. It is given by,

$$
\begin{aligned}
& y_{it}|\eta_{it} \sim \text{Poisson}(e_{it}\theta_{it}) \\
& \eta_{it} = \log(\theta_{it}) = \boldsymbol{X}_{it}^T\boldsymbol{\beta} + u_{it} \\
& u_{it} = \varepsilon_i + (\alpha + \delta_i)t^* \\
& \\
& \beta_j \sim \text{N}(\mu_\beta, \sigma_\beta^2) \\
& \\
& \varepsilon_i|\varepsilon_{-i} \sim \text{N}\left(\frac{\rho_S \sum_{j=1}^n w_{ij}\varepsilon_j}{\rho_S \sum_{j=1}^n w_{ij} + 1 - \rho_S}, \frac{\sigma_S^2}{\rho_S \sum_{j=1}^n w_{ij} + 1 - \rho_S}\right) \\
& \rho_S \sim \text{Uniform}(0,1) \\
& \sigma_S^2 \sim \text{Inverse-Gamma}(a,b) \\
& \\
& \alpha \sim \text{N}(\mu_\alpha, \sigma_\alpha^2) \\
& \delta_i \sim \text{N}(0, \sigma_{slo}^2) \\
& \sigma_{slo}^2 \sim \text{Inverse-Gamma}(c,d)
\end{aligned}
\tag{4.10}
$$

4.2.2 Spatio-temporal Leroux dynamic model

The model ST.Leroux.dynamic decomposes spatio-temporal variation into three components: spatial effects ε_i, time structured effects δ_t and unstructured interaction effects γ_{it}.

It is given by,

$$
\begin{aligned}
&y_{it}|\eta_{it} \sim \text{Poisson}(e_{it}\theta_{it}) \\
&\eta_{it} = \log(\theta_{it}) = \boldsymbol{X}_{it}^T\boldsymbol{\beta} + u_{it} \\
&u_{it} = \varepsilon_i + \delta_t + \gamma_{it} \\
\\
&\beta_j \sim \text{N}(\mu_\beta, \sigma_\beta^2) \\
\\
&\varepsilon_i|\varepsilon_{-i} \sim \text{N}\left(\frac{\rho_S\sum_{j=1}^n w_{ij}\varepsilon_j}{\rho_S\sum_{j=1}^n w_{ij}+1-\rho_S}, \frac{\sigma_S^2}{\rho_S\sum_{j=1}^n w_{ij}+1-\rho_S}\right) \\
&\rho_S \sim \text{Uniform}(0,1) \\
&\sigma_S^2 \sim \text{Inverse-Gamma}(a,b) \\
\\
&\delta_t - \delta_{t+1} \sim \text{N}\left(0, \sigma_T^2\right) \\
&\log(\sigma_T^{-2}) \sim \text{Log-Gamma}(e,f) \\
\\
&\gamma_{it} \sim \text{N}\left(0, \sigma_I^2\right) \\
&\sigma_I^2 \sim \text{Inverse-Gamma}(c,d)
\end{aligned}
\tag{4.11}
$$

These Bayesian spatio-temporal hierarchical log-Poisson models are estimated under INLA methodology in R-INLA. The spatial component is implemented in R-INLA using the "generic1" function in order to implement the Leroux *et al.* [59] CAR prior distribution for the spatially random effects [42, 100].

The R-code used in R-INLA to implement these spatio-temporal Bayesian hierarchical log-Poisson models, can be found in Appendix A.2

4.3 Spatio-temporal Bayesian autoregressive log-Poisson model

The Bayesian Poisson spatial lag model, a spatial autoregressive econometric model for count data [86], is extended here to allow the inclusion of a temporal component in u_{it} in (4.2) through (4.5) or (4.7).

Analogously to [86], and given the exposed in section 4.1, we are able to write the predictor η_{it} of the spatio-temporal autoregressive in space log-Poisson model as:

$$\eta_{it} = \log(\theta_{it}) = \boldsymbol{X}_{it}^T\boldsymbol{\beta} + \rho\sum_{j\neq i} w_{ij}\eta_{jt} + \delta_t + \gamma_{it}, \tag{4.12}$$

considering, without loss of generality, that δ_t represents a temporal effect and γ_{it} a space-time interaction effect. For each time t this can be expressed in matrix notation as:

$$\eta_t = \boldsymbol{X}_t^T \boldsymbol{\beta} + \rho W \eta_t + \delta_t 1_{n\times 1} + \gamma_t, \tag{4.13}$$

where η_t denotes the vector of log-risks η_{it} for year t, $\boldsymbol{X}_t^T$ represents the observed covariates in year t (including a first column of 1's for the intercept), $1_{n\times 1}$ is a column of n 1's and γ_t denotes vector of the space-time interaction effects γ_{it} for year t.

Equivalently this can be expressed as:

$$\begin{aligned} \eta_t &= (I_n - \rho W)^{-1} \boldsymbol{X}_t^T \boldsymbol{\beta} + (I_n - \rho W)^{-1} \delta_t 1_{n\times 1} + (I_n - \rho W)^{-1} \gamma_t \quad \Leftrightarrow \\ \eta_t &= (I_n - \rho W)^{-1} \boldsymbol{X}_t^T \boldsymbol{\beta} + \delta_t^* + \gamma_t^*, \end{aligned} \tag{4.14}$$

where

$$\begin{aligned} \delta_t^* &= (I_n - \rho W)^{-1} \delta_t 1_{n\times 1} \quad \Leftrightarrow \\ \delta_t 1_{n\times 1} &= (I_n - \rho W) \delta_t^* \end{aligned} \tag{4.15}$$

and

$$\begin{aligned} \gamma_t^* &= (I_n - \rho W)^{-1} \gamma_t \quad \Leftrightarrow \\ \gamma_t &= (I_n - \rho W) \gamma_t^* \end{aligned} \tag{4.16}$$

Expression for η_t in (4.14) is then given as a sum of three random effects, $x_1 = (I_n - \rho W)^{-1} \boldsymbol{X}_t^T \boldsymbol{\beta}$, $x_2 = \delta_t^*$, and $x_3 = \gamma_t^*$, allowing its estimation through INLA methodology, which requires models to be written under such a formulation, given the exposed in section 3.3 [86]. Temporal and spatio-temporal effects are considered in this transformed version (δ_t^* and γ_{it}^*), being related to the untransformed effects δ_t and γ_{it}, respectively, by a factor of $(1-\rho)$. This relation can be used to recover the original effects.

When the choice of the temporal component is a linear trend as proposed by Bernardelli *et al.* [17], the resulting model is the Bayesian Poisson spatio-temporal lag linear model (ST.splm.linear), with spatially varying linear time trends. When considering a simple dynamic nonparametric formulation for the temporal component and an unstructured space-time interaction as proposed by Knorr-Held [52], that results into a the Bayesian Poisson spatio-temporal lag dynamic model (ST.splm.dynamic).

4.3.1 Bayesian Poisson spatio-temporal lag linear model

The ST.splm.linear formulation includes an autoregressive spatial component ε_{it}, a main linear trend α^* which represents the global time effect and a differential time trend δ_i^*.

$$
\begin{aligned}
& y_{it}|\eta_{it} \sim \text{Poisson}(e_{it}\theta_{it}) \\
& \eta_{it} = \log(\theta_{it}) = \boldsymbol{X}_{it}^{T}\beta + u_{it} \\
& u_{it} = \varepsilon_{it} + (\alpha^{*} + \delta_{i}^{*})t^{*} \\
& \\
& \beta_j \sim \text{N}(\mu_\beta, \sigma_\beta^2) \\
& \\
& \varepsilon_{it} = \rho \textstyle\sum_{j \neq i} w_{ij}\eta_{jt} \\
& \text{logit}(\rho) \sim \text{N}(a, b) \\
& \\
& \alpha^{*} \sim \text{N}(\mu_{\alpha^{*}}, \sigma_{\alpha^{*}}^{2}) \\
& \delta_{i}^{*} \sim \text{N}(0, \sigma_{slo}^{2}) \\
& \sigma_{slo}^{2} \sim \text{Inverse-Gamma}(c, d)
\end{aligned}
\tag{4.17}
$$

4.3.2 Bayesian Poisson spatio-temporal lag dynamic model

The ST.splm.dynamic model decomposes spatio-temporal variation into three components: autoregressive spatial effects ε_{it}, time structured effects δ_t^* and unstructured interaction effects γ_{it}^*.

$$
\begin{aligned}
& y_{it}|\eta_{it} \sim \text{Poisson}(e_{it}\theta_{it}) \\
& \eta_{it} = \log(\theta_{it}) = \boldsymbol{X}_{it}^{T}\beta + u_{it} \\
& u_{it} = \varepsilon_{i} + \delta_{t}^{*} + \gamma_{it}^{*} \\
& \\
& \beta_j \sim \text{N}(\mu_\beta, \sigma_\beta^2) \\
& \\
& \varepsilon_{it} = \rho \textstyle\sum_{j \neq i} w_{ij}\eta_{jt} \\
& \text{logit}(\rho) \sim \text{N}(a, b) \\
& \\
& \delta_{t}^{*} - \delta_{t+1}^{*} \sim \text{N}\left(0, \sigma_{T}^{2}\right) \\
& \log(\sigma_{T}^{-2}) \sim \text{Log-Gamma}(e, f) \\
& \\
& \gamma_{it}^{*} \sim \text{N}\left(0, \sigma_{I}^{2}\right) \\
& \sigma_{I}^{2} \sim \text{Inverse-Gamma}(c, d)
\end{aligned}
\tag{4.18}
$$

The Bayesian Poisson spatial lag model is implemented in R-INLA using the "slpm" function [42, 86], and the R-code used in R-INLA to implement the spatio-temporal lag Poisson models, ST.splmINLA.linear and ST.splmINLA.dynamic, can be found in Appendix A.3.

The models derived from the two different modelling strategies for count data, the hierarchical and the autoregressive ones, were accommodate in the same mathematical framework given by,

$$\begin{gathered} y_{it}|\eta_{it} \sim \text{Poisson}(e_{it}\theta_{it}), \\ \eta_{it} = \log(\theta_{it}) = \boldsymbol{X}_{it}^T\boldsymbol{\beta} + u_{it}, \text{i=1,...,n, t=1,...,T} \end{gathered} \tag{4.19}$$

with different representations of u_{it} corresponding to different model specifications. Within this common frame other possibilities can be also investigated for spatio-temporal modelling on the hierarchical and autoregressive approaches. For example, time effects can be further explored resorting to other possibilities for spatio-temporal modelling, as for example considering temporal effects prior distributed according to the time analogous of Leroux et al. [59] CAR distribution or by additionally considering a set of structured space-time interactions [56].

4.4 Application: Spatio-Temporal Modelling of S24 Data for assessing hospital savings

Urgency admissions is one of the most important factors regarding hospital costs, which can possibly be mitigated by the use of national health lines such as the Portuguese Saúde24 line (S24). The assessment of that premise is the motivation of this application study.

4.4.1 Econometric analysis of S24 calls to assess hospital savings

This study focuses on the number of TAE calls to S24 at a municipality level in Continental Portugal annually over the period 2010-2016. The interest is evaluating the economic impact of the use of this health line concerning the reached savings by avoid unnecessary urgent care in hospitals. For that what is relevant are the TAE calls whose initial intention of the user was to go to hospital urgency but, according with the final health line disposition, that is not necessary. These are the calls analyzes in this work, described as TAE savings calls.

Considering the rate of savings calls as a proportion of the total number of calls whose initial intention was the hospital urgency, its temporal trend can be visualized through the boxplot shown in the left panel of Figure 4.1. The plot shows a slight decrease, with median rate values between 65% in 2016 and 77% in 2010. In the right panel of Figure 4.1 one can observe the average spatial pattern over time of this rate.

Taking into account that the savings achieved with S24 are proportional to the number of savings calls, this work investigates and models spatial and temporal dependencies in the number of TAE savings calls in the Portuguese municipalities for the time period from 2010 to 2016. This will allow, for example, to describe and evaluate in which

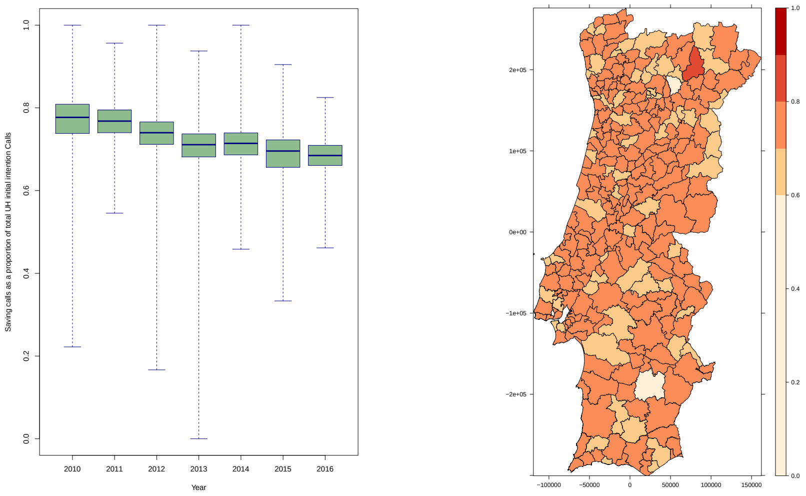

Figure 4.1: Boxplot of the temporal trend in the rates of saving calls as a proportion of the total number of calls whose initial intention was hospital urgency (left) and rates for each municipality averaged over 2010-2016 (right).

municipalities the use of S24 must be potentiated and highlight those regions that most contribute to the economic success of the use of the health line. Given this, the data considered in this study are the number of TAE savings calls that were registered in each of the 278 municipalities of Continental Portugal from 2010 to 2016 inclusive.

Within this context, the expected savings can be given by the following simple measure:

$$S_t = \sum_{i=1}^{278} y_{it} p_t - C_t,$$

where y_{it} are the number of savings calls in each municipality $i = 1, \ldots, 278$ and year $t = 1, \ldots, 7$; p_t corresponds to the price of a simple medical appointment in the urgency, which is set every year in a decree-law by the Ministry of Health; C_t are costs associated with annual health line maintenance, provided by Portuguese Directorate-General of Health.

It should be emphasized that this is a simplistic measure which does not consider collateral costs as, for example, those caused by unnecessary urgency increased contagion risk in the population or those caused by increasing the number of patients in the urgency, requiring more human and financial resources. The savings from the correct use the S24, avoiding urgent care in a hospital, can then be channeled for other needed areas.

4.4.2 Preliminary S24 data analysis

The number of TAE savings calls are registered in each of the 278 municipalities of Continental Portugal from 2010 to 2016 inclusive. Naturally, these numbers depend on the size and demographic structure of the population, in terms of age and gender characteristics of the underlying population, which requires adjustments in terms of (indirect) standardization. The observed y_{it} and the resulting expected number e_{it} of saving calls in each municipality $i = 1, ..., 278$ and year $t = 1, .., 7$ can then be used to roughly estimate the relative call risk, θ_{it}, through the Standard Call Rate (SCR) in municipality i and year t, $\text{SCR}_{it} = \frac{y_{it}}{e_{it}}$, mapped in Figure 4.2.

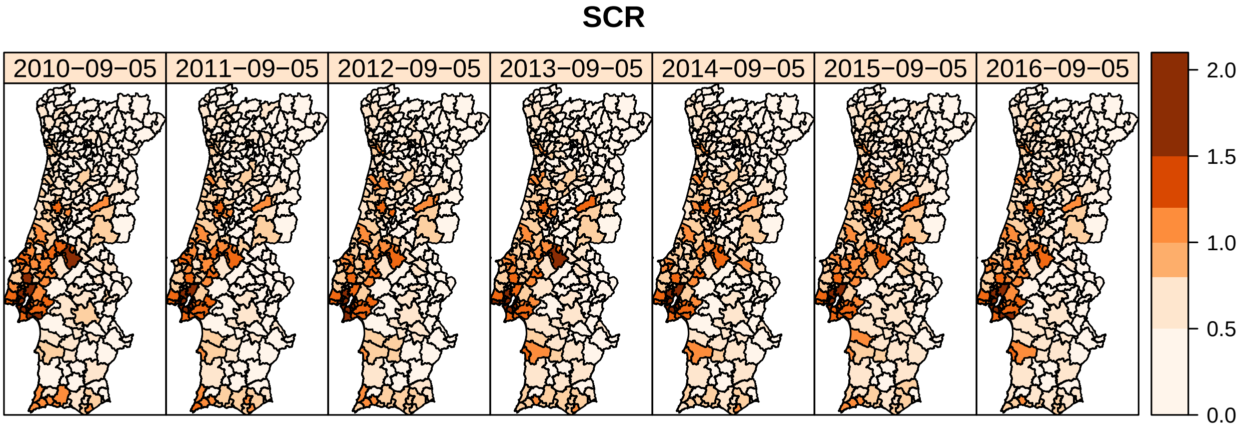

Figure 4.2: SCR-2010 to 2016.

Several demographic, socio-economic, development and health indicators were investigated as possible covariates for modelling the TAE savings calls counts using a simple log-Poisson model. The expected number of saving calls is included in the model as an offset. Bayesian covariable selection using indicator auxiliary variables under different scenarios [40] was performed. The most significant set of explanatory variables turned out to be the *average number of years of schooling* (x_1), the *proportion of active population* (x_2), the *rurality index* (x_3), the *number of hospitals* (x_4) and the *number of health centers per 1000 inhabitants* in each municipality (x_5). This was implemented in R-package `R2WinBUGS` [39] that run WinBugs [64].

It was further fitted a log-Poisson model with unstructured random effects to account for possible over-dispersion, that is expected in these Poisson data. This confirmed the significance of some of the previous set of covariates, two, the *number of hospitals* and the *number of health centers* that were not significant. Nevertheless these covariates were kept in the model for their intrinsic econometric interest. Table 4.1 depicts the corresponding estimated coefficients of this overdispersed Poisson log-linear model. The estimates were obtained via INLA methodology in R-package `R-INLA` [82].

Variable	**Id**	**Mean (CI)**
Average number of years of **schooling**	x_1	0.39 (0.36 ; 0.42)
Proportion of **active population**	x_2	−1.91 (-2.53 ; -1.29)
Rurality index	x_3	−0.27 (-0.38 ; -0.16)
Number of **hospitals**	x_4	−0.11 (-0.73 ; 0.50)
Number of **health centers**	x_5	−0.46 (-0.75 ; 0.17)
Intercept		−1.85 (-2.30 ; 2.40)

Table 4.1: Covariates and their estimated coefficients (posterior mean and 95% credible intervals, CI) for the Poisson log-linear model with unstructured random effects.

Taking into account that the covariates are related in the model to the natural log of SCR, figure 4.3 shows the relationships between the selected variables for all years under analysis and the natural log of SCR. This scatterplot shows a positive relationship between the log and the average number of years of schooling, suggesting that increased levels of schooling are connected to an increased risk of saving calls. It is also noted a slightly negative relationship between the log-SCR and the number of hospitals, suggesting that an increase in the number of hospitals have a decreasing impact on the saving calls risk.

4.4.3 Spatial and temporal correlation

This investigation begins by first looking for spatial dependence in the number of TAE savings calls to S24 per 1000 inhabitants, by municipality and by year. Then it proceeds by looking for spatial dependence in the residuals of the log-Poisson model with covariates and unstructured random effects fitted before for each year. For this, Moran's I statistic is considered [11], with the contiguity neighbourhood spatial structure in matrix W described before, being a permutation test conducted for each year of data separately.

The results are displayed in Table 4.2 and were obtained with R-project package `spdep`. They suggests that there is positive spatial correlation among of the number of TAE saving calls, from 2010 to 2016, and also a strong evidence of unexplained spatial correlation in each year, after accounting for the covariate effects.

The residual temporal correlation could be assessed similarly considering a temporal neighbourhood structure but with only 7 time periods the resulting estimates would not be reliable [8, 56].

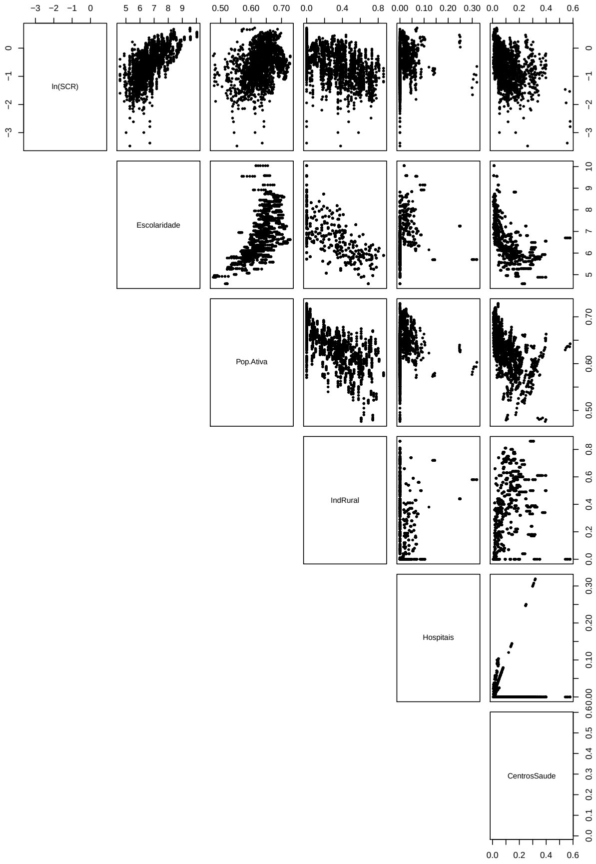

Figure 4.3: Scatterplot of the log-SCR and the selected covariates for all years.

Given the above, the assumption of independence is not valid for these data and spatial and eventually spatio-temporal effects should be considered in the model.

4.4.4 Spatio-temporal Bayesian econometric modelling of S24

For modelling the number of TAE savings calls in the Portuguese municipalities between 2010 and 2016, different space-time structures are now considered under both hierarchical and autoregressive perspectives.

In order to compare models, it was chosen the measures of the WAIC (Watanabe-Akaike information criteria) and the DIC (Deviance information criteria), as DIC is the predictive mesure most used in Bayesian applications [91] and WAIC revealed to be more stable and particulary helpful with hierarchical and mixture structures [38].

Year	**Moran's test for TAE Saving Calls**	**Moran's test for the residuals of the log-Poisson model**
2010	$I = 0.588$ $(p < 2.2e-16)$	$I = 5.723$ $(p = 1.1e-8)$
2011	$I = 0.584$ $(p < 2.2e-16)$	$I = 6.141$ $(p = 8.2e-10)$
2012	$I = 0.548$ $(p < 2.2e-16)$	$I = 5.230$ $(p = 1.7e-7)$
2013	$I = 0.590$ $(p < 2.2e-16)$	$I = 3.951$ $(p = 7.8e-5)$
2014	$I = 0.579$ $(p < 2.2e-16)$	$I = 4.346$ $(p = 1.4e-5)$
2015	$I = 0.579$ $(p < 2.2e-16)$	$I = 4.164$ $(p = 3.1e-5)$
2016	$I = 0.570$ $(p < 2.2e-16)$	$I = 4.462$ $(p = 8.1e-6)$

Table 4.2: Estimated Moran's I statistic and corresponding p-values for two sided test of spatial independence, considering TAE Saving Calls and the residuals of the log-Poisson regression model with unstructured effects.

4.4.4.1 Spatio-temporal Bayesian hierarchical log-Poisson modelling of S24 data

This formulation models spatial dependence by including in a log-Poisson model spatial random effects hierarchically and assuming a parametric trend for the temporal component through the ST.Leroux.linear model or a nonparametric dynamic space-time model with ST.Leroux.dynamic model. The estimates for this approach were obtained via INLA methodology, based on the generic1 latent class to carry out the Leroux *et al.* CAR prior distribution, in R-package `R-INLA`, according to the R-code presented in appendix A.2. The results are summarized in tables 4.3 and 4.4.

For the regression coefficients β_j weak informative Normal(0,1000) prior distributions were considered. A non informative uniform prior distribution on the unit interval was specified for parameter ρ_S. An weak informative inverse-gamma$(a = 1, b = 0.01)$ prior distribution was specified for the variance of the random effects σ_S^2.

In the ST.Leroux.linear model, for the σ_{slo}^2 an inverse-gamma prior distribution is considered [25]. The corresponding hyperparameters $(a, b, \mu_\alpha, \sigma_\alpha^2)$ were chosen in this work such that the resulting prior distributions were weakly informative. It is assumed here that $a = 1, b = 0.01$, $\mu_\alpha = 0$ and $\sigma_\alpha^2 = 1000$.

In the ST.Leroux.dynamic model, for the independent space-time interactions $\gamma = (\gamma_{11}, ..., \gamma_{nT})$ Gaussian prior distributions are specified. For σ_I^2 a weak informative inverse-gamma prior distribution was considered $(a = 1, b = 0.01)$ and for $\log(\sigma_T^{-2})$ a weak informative log-gamma prior distribution was considered $(a = 1, b = 0.00005)$.

It stands out from ST.Leroux.linear model the fact that only two of the initial covariates revealed to be significant, the average number of years of schooling and the rurality index. The main linear trend α was estimated as 0.06, being significant. In terms of ST.Leroux.dynamic model also only two of the initial covariates revealed to be significant, the average number of years of schooling and the proportion of active population. The spatial dependence parameter ρ_S exhibits a relatively high estimated value of 0.96.

		Model	
		ST.Leroux.linear	**ST.splm.linear**
Variable	**Id**	**Mean** **(CI)**	**Mean** **(CI)**
Average number of years of **schooling**	x_1	0.24 (0.18;0.3)	0.41 (0.14;0.62)
Proportion of **active population**	x_2	−0.14 (-0.97;0.69)	−3.89 (-7.66;-0.02)
Rurality index	x_3	−0.22 (-0.44;-0.004)	0.17 (-0.53;0.84)
Number of **hospitals**	x_4	0.32 (-0.18;1.06)	0.18 (-4.01;4.33)
Number of **health centers**	x_5	0.20 (-0.25;0.65)	0.61 (-2.31;1.00)
Intercept		−2.02 (-2.77;-1.27)	−1.11 (-3.95;1.68)
α / α^*		0.06 (0.04;0.09)	0.06 (0.05;0.09)
ρ_S / ρ		0.94 (0.82;0.99)	0.56 (0.55;0.56)
σ_S^2 / σ^2		0.34 (0.29;0.43)	0.08 (0.07;0.10)
σ_{slo}^2		0.04 (0.03;0.05)	0.04 (0.03;0.05)

Table 4.3: Parameter estimates (mean and 95% credible intervals, CI) for the spatio-temporal ST.Leroux.linear model and ST.splm.linear model, for the 2010-2016 S24 data. σ^2 is the estimated variance of the spatial effects for the autoregressive model.

4.4.4.2 Spatio-temporal Bayesian autoregressive log-Poisson modelling of S24 data

The number of TAE savings calls in each municipality for the period 2010-2016 is analysed here through the Bayesian Poisson spatial lag model with temporal effects. The first model considered incorporates a spatial autocorrelation lag component in the econometric model of counts and a parametric trend for the temporal component through the ST.splm.linear model. Then a nonparametric dynamic space-time model is also implemented within the Bayesian Poisson spatial lag model with temporally structured effects

		Model	
		ST.Leroux.dynamic	**ST.splm.dynamic**
Variable	**Id**	**Mean (CI)**	**Mean (CI)**
Average number of years of **schooling**	x_1	0.19 (0.13;0.26)	0.14 (0.10;0.19)
Proportion of **active population** residents	x_2	1.82 (0.95;2.69)	−0.11 (-0.88;0.63)
Rurality index	x_3	−0.16 (-0.37;0.06)	0.11 (-0.05;0.26)
Number of **hospitals**	x_4	0.06 (-0.68;0.79)	0.54 (-0.42;1.49)
Number of **health centers**	x_5	0.14 (-0.39;0.68)	−0.17 (-0.55;0.21)
Intercept		−2.97 (-3.78;-2.15)	−1.00 (-1.61;-0.39)
σ_S^2 / σ^2		0.34 (0.29;0.41)	0.08 (0.07;0.09)
ρ_S / ρ		0.96 (0.89;0.99)	0.56 (0.55;0.57)
σ_T^2		1.69e-04 (6.17e-05;7.67e-04)	1.17e-04 (3.90e-05;5.74e-04)
σ_I^2		0.005 (0.004;0.006)	0.005 (0.004;0.006)

Table 4.4: Parameter estimates (mean and 95% credible intervals) for the ST.Leroux.dynamic model and ST.splm.dynamic model, for the 2010-2016 S24 data. σ^2 is the estimated variance of the spatial effects for the autoregressive model.

and allowing an unstructured interaction between space and time.

The estimates were obtained in R-INLA, using "slpm" function, readjusted for the spatio-temporal linear and dynamic cases. Main results are displayed in Tables 4.3 and 4.4 and correspond to formulation (4.14), as well as the others results to come. The estimated temporal and spatio-temporal effects of the original formulation were retrieved in an ad hoc manner by the existing relation between parameters δ_t and δ_t^*, γ_{ti} and γ_{ti}^*, being quite similar to each other.

For the regression coefficients β_j weak informative Normal$(0,1000)$ prior distributions were considered. For the logit transformation of parameter ρ, logit(ρ), a Normal(0,10) weak informative prior distribution was considered.

In the ST.splm.linear model, it is assumed for each element of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ and for the overall slope parameter α independent Gaussian prior distributions. For σ_{slo}^2 an inverse-gamma prior distribution was considered [25]. The corresponding hyperparameters $(a, b, \mu_\alpha, \sigma_\alpha^2)$ are chosen in this work such that the resulting prior distributions are weakly informative. It is assumed here that $a = 1, b = 0.01$, $\mu_\alpha = 0$ and $\sigma_\alpha^2 = 1000$.

In the ST.splm.dynamic model, for the independent space-time interactions $\boldsymbol{\gamma} = (\gamma_{11}, ..., \gamma_{nT})$ Normal prior distributions were specified. For σ_I^2 a weak informative inverse-gamma prior distribution were considered $(a = 1, b = 0.01)$ and for $\log(\sigma_T^{-2})$ a weak informative log-gamma prior distribution was considered $(a = 1, b = 0.00005)$.

In relation to the ST.splm.linear model, only two of the initial covariates revealed to be significant, the average number of years of schooling and proportion of active population residents. This model has an estimated value of the spatial dependence parameter ρ of 0.56. The main linear trend α, which represents the global time effect, is estimated as 0.06, being significant, and the estimated value of σ_{slo}^2 is 0.04. For the ST.splm.dynamic formulation only one of the initial covariates is now significant, the average number of years of schooling, the estimated values of σ_I^2 and σ_T^2 are 0.005 and $1.17e{-}04$, respectively.

4.4.4.3 Model comparison for TAE saving calls

In order to compare all the fitted models, from both hierarchical and autoregressive approaches, the DIC and WAIC measures were used. For comparing different model structures WAIC measure was considered, not showing much difference between the two approaches. However, WAIC was a bit smaller for the autoregressive approach indicating a slight better performance. Within each approach, in terms of the different temporal structures, DIC and WAIC were concordant and smaller for the dynamic parametrization. See Table 4.5 for the models with the linear trend and Table 4.6 for the models with a dynamic trend.

In what concerns to each of the different temporal structures considered, one can say that within the parametric trend formulation, similar estimates were obtained for the effects with the same type of structure for the two different hierarchical and autoregressive approaches; the main linear trend was significant and displays an estimated value of 0.06 for both perspectives, although the significance of the inicial set of covariates differed from the simple log-Poisson regression model. When considering the dynamic nonparametric formulation, only one of the initial set of covariates revealed to be significant, the average number of years of schooling.

The dynamic parametrization for the autoregressive and for the hierarchical approach improved the model fit (smaller DIC and WAIC) suggesting that this parametrization for time is more appropriate to S24 data, slightly better in terms of WAIC for the autoregressive approach. In what follows the main results for the latter model are summarized.

ST.splm.dynamic model This fitted model has an estimated value of the spatial dependence parameter ρ of 0.56, similar to the value obtained for the linear case. The estimated posterior mean of the relative risks $\theta_{it} = \exp(\eta_{it})$ were calculated as in chapter 7 of [25] and plotted for the available 7 years, 2010 to 2016. They are displayed in Figure 4.4,

Model	DIC	pD	WAIC	pW
ST.Leroux.linear	16994	463	17476	776
ST.splm.linear	16997	463	17474	773

Table 4.5: DIC and WAIC measures for the 2 models fitted for the case of a linear trend.

Model	DIC	pD	WAIC	pW
ST.Leroux.dynamic	16382	997	16373	759
ST.splm.dynamic	16387	1007.9	16370	760

Table 4.6: DIC and WAIC measures for the 2 models fitted for the case of a dynamic trend.

allowing to observe the slight changes in the estimated spatial pattern over time. The estimated posterior mean of the spatial main effect, $\zeta_i = \exp(\varepsilon_i)$, and the estimated posterior temporal trend $\xi_t = \exp(\delta_t^*)$, are depicted in Figures 4.5 and 4.6 respectively, suggesting a mild increase in the time effect. The estimated posterior mean for the space-time interaction for the chosen model, for the years 2010, 2012, 2014 and 2016, are displayed in Figure 4.7. The considered space-time interaction effects did not seem to be of added value in explaining the TAE savings calls risk.

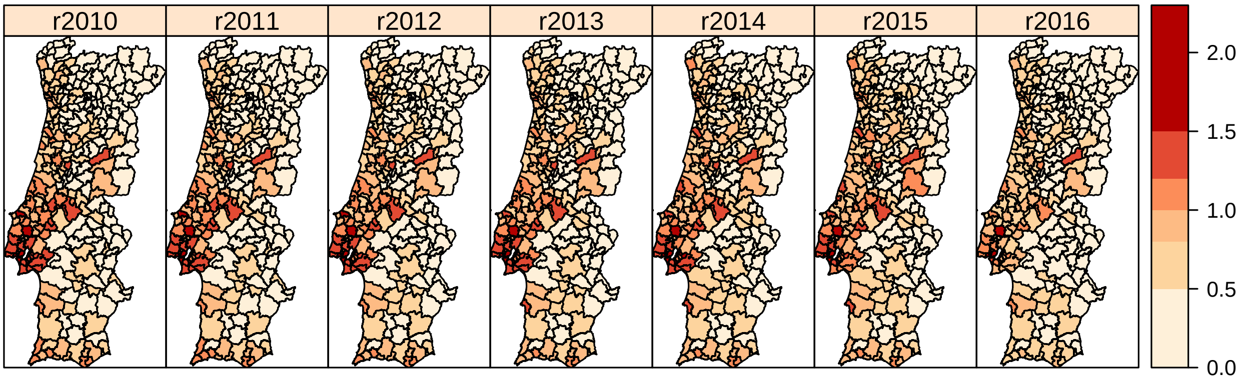

Figure 4.4: Spatial evolution of the estimated posterior mean of the relative risks for the chosen model.

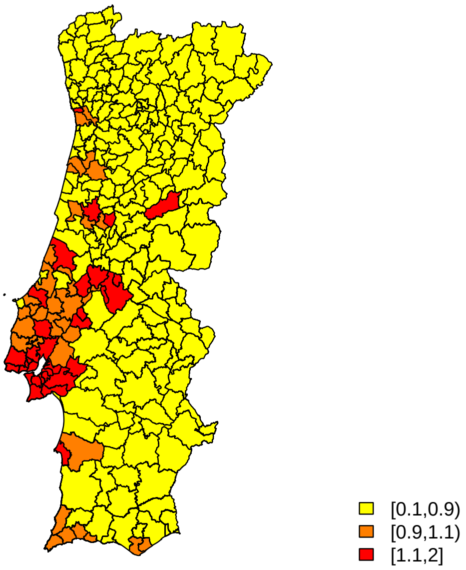

Figure 4.5: Estimated spatial effects for the ST.splm.dynamic model.

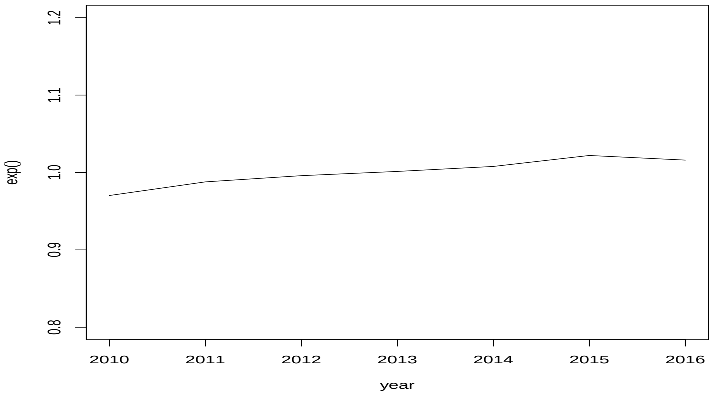

Figure 4.6: Estimated temporal effects for the ST.splm.dynamic model.

4.4.5 The Savings Index

The S24 data analysis can be extended, exploring the econometric analysis of the Savings Index, obtained for each municipality $i = 1, \ldots, 278$ and year $t = 1, \ldots, 7$:

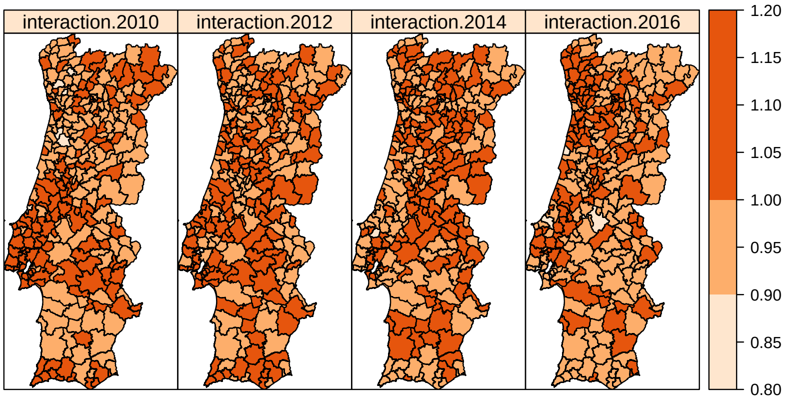

Figure 4.7: Estimated spatio-temporal interaction effects for years of 2010, 2012, 2014 and 2016 for the ST.splm.dynamic model.

$$IS_{it} = \frac{(y_{it} p_t - C_{it})}{U_{it}}$$

where y_{it} are the number of savings calls, p_t corresponds to the price of a simpler medical appointment in the urgency, C_{it} costs associated with annual health line maintenance in charge of municipality i and U_{it} are the total number of calls whose initial intention is urgency, mapped in Figure 4.8.

The municipalities of *Odivelas*, *Lisboa*, *Vila Franca de Xira*, *Loures*, *Moita*, *Setúbal*, *Cascais*, *Amadora*, *Almada* and *Oeiras*, stand out as the ones where the savings assessment are more accentuated, on the other hand municipalities from northern interior should be considered for implement management policies in order to improve the impact of S24 in the solving of non-urgent emerging situations.

4.4.6 Main Conclusions

The estimation results obtained for the S24 data are concordant in terms of the different hierarchical and autoregressive approaches, being slightly better for the latter, which is not perhaps the first choice in terms of the methodology for this type of data. A spatial structure was evident from the analysis as well as a slight temporal trend. The considered space-time interaction effects did not seem to be of added value.

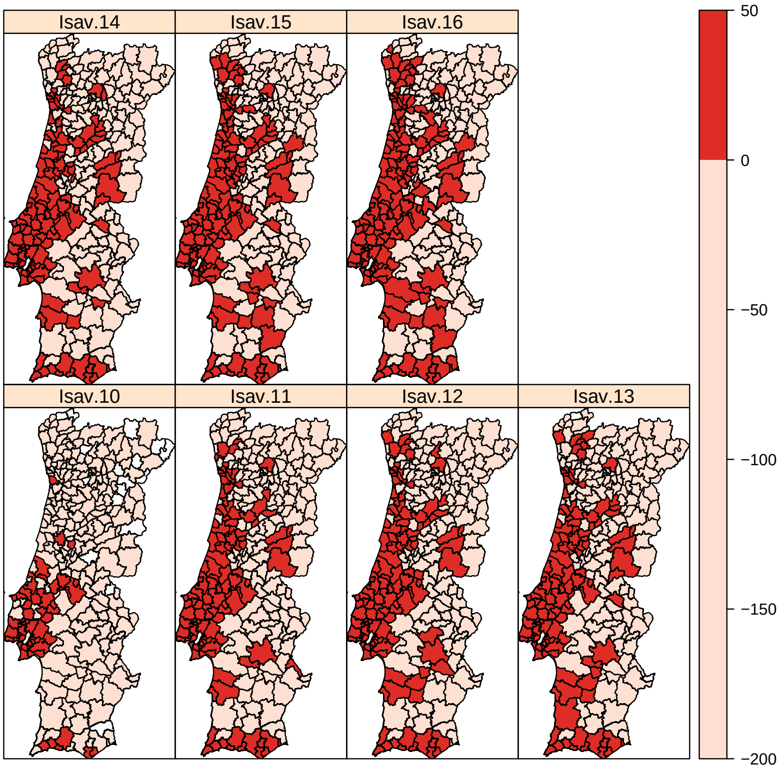

Figure 4.8: The Savings Index from 2010 to 2016.

The temporal correlation could not be assessed as well as the spatial correlation, due to the short time series of seven time periods, although was considered that, it was important to include the temporal component and allowing an interaction term between time and space. The small number of time periods was also the reason for not considering more complex structures for space-time interaction effects. An alternative would have been considering a different Leroux *et al.* (2000), [59], CAR prior distribution for the spatial effects for each year, resulting in differentiated autocorrelation parameter for each year, and evaluating afterwards the time evolution of these estimates [56].

In terms of the savings index, one can say that over the last seven years, there has been a more comprehensive spatial effectiveness of the S24 line in solving the non-urgent emergency situations, that could be handled by primary health care services or in a self care basis. The savings index allows to detect which municipalities most contribute to the economic success of the S24 line and, at the same time, shows in which regions the use of the health line should be encouraged.

The results of this study contribute to understand one of the major factors regarding hospital costs of Portuguese health care system, the non-urgent emergency situations, responsible for major efficacy losses of the hospital systems, and at the same time helps to realize that this phenomenon have been mitigated through the creation, promotion and use of S24, knowing that presently there are still about 15% of non-urgent cases per year in Portuguese Hospital Care System.

CHAPTER 5

DISCUSSION

Count data for estimating risk of events are typically modelled through Poisson regressions that classically would not consider space or time. However when justified these should be taken into account. Considering count data in areas or spatial units - *lattice data* [15, 19, 29] the classical Poisson regression models can further accommodate a spatial component, in a hierarchical or in an autoregressive way, better accomplished under the Bayesian paradigm. Additionally, time effects can also be incorporated, possibly varying over space and capturing dynamic evolutions [25, 30, 54].

The main achievements of this thesis are model developments for spatial and spatio-temporal Poisson count data, particularly for the class of autoregressive models, and the establishment of a mathematical common framework for autoregressive and hierarchical modelling approaches [83, 86, 87]. These have been applied in an unprecedented analysis of the number of calls and saving calls to the S24 [83, 86]. Data on road accidents were also modelling considering space [84, 85].

In spatial statistics, hierarchical modelling is the natural way to handle areal count data to account for data overdispersion as it is well established in the literature [15, 25, 29]. For that, non-observable random effects can be added to the effect of existing covariates in modelling extra variation that might exist in counts. These counts are allowed to follow any distribution, the Poisson distribution being the most common choice. In this thesis, in order to explain spatial and temporal correlations, hierarchical Bayesian models with covariates, unstructured random effects and spatially and temporally structured random effects were considered [85, 86].

The hierarchical approach accounts for spatial autocorrelation in the disturbances and not in the observed responses, but it might also be interesting to consider a spatial autoregressive approach. This idea is motivated by the fact that it is plausible to think that the risk in one area is related to the risk in the areas of its neighbourhood, driven by effects of important covariables either included or not in the model, [62]. Additionally, very often the covariates considered display themselves high spatial dependency. Consequently the response variable in a given spatial unit is most certainly a good predictor of the response variable in its neighbouring areas. This is a different modelling strategy common in spatial econometrics literature that is also considered in this work.

Traditional spatial econometric models, such as the spatial autoregressive model (SLM) and the spatial error model (SEM), rely on the Gaussian assumption of the distribution of the response variable [62], which does not hold for count data. Consequently their usage for this type of data, such as the number of accidents or the number of calls, demands data transformation to meet the assumptions of the models. In order to avoid that, this thesis investigated new possible spatial and spatio-temporal modelling strategies for handling count data, assuming a Poisson distribution for those. In the scope of these spatial autoregressive models there are alternatives for modelling counts that have been explored [86]. More specifically: considering a standard spatial lag model, recently developed within a new class of latent models defined in Integrated Nested Laplace Approximations (INLA) [81], by Gómez-Rubio, Bivand and Rue in 2015 [42]; considering a spatial autoregressive lag model of counts, developed by Lambert, Brown and Florax in 2010 [53], under a classical perspective; a spatial lag autoregressive component was incorporated in the model for counts, under a Bayesian paradigm and using INLA methodology.

This resulted in a Poisson spatial lag model, an alternative to do Bayesian inference in spatial econometric models for count data, [86]. In this thesis, this was further elaborated into a more complex spatio-temporal autoregressive model. The lag model with temporal and spatio-temporal interaction effects has not yet been implemented before in a Poisson response Bayesian setting. Combining insights from classical spatial econometrics and from the analysis of spatial-temporal data, novel Bayesian Poisson spatio-temporal lag models were presented.

One should bear in mind that there are different underlying theoretical motivations for using each of these different types of models, although hierarchical and autoregressive modelling perspectives have already been used to model the same data sets [23, 43, 74]. In this thesis temporal and spatio-temporal effects were implemented both within a Bayesian Poisson spatial lag model and a Bayesian Poisson hierarchical model, under INLA methodology.

In this quest different space-time structures were considered under two different modelling perspectives for count data mentioned before, hierarchical and autoregressive approaches [15, 62]. Temporal effects could be included similarly in both approaches, being specified for the temporal component both a parametric and a nonparametric structure in this study. When the purpose of the analysis was to estimate which areas are exhibiting increasing or decreasing linear trends in the response over time a parametric structure was considered with the model proposed by Bernardelli *et al.* (1995) [17], however if time trends and spatial patterns are to be estimated overall region wide, a dynamic nonparametric formulation is more adequate, and the propose of Knorr-Held (2000) was considered [52] in its simplest form, although there are other more complex alternatives that could have also be explored. It was considered a linear interaction but a non-linear interaction could also be taken into account.

Whereas the hierarchical approach is the natural modelling way to handle areal count data, the fact that it is plausible to justify the use of an autoregressive approach [62]. The benefits of the autoregressive approach over the classical hierarchical approach for count data are mainly concerned with econometric interpretation, since the response variable in a given area constitutes a good predictor of the response variable in its neighbourhood areas. In this thesis we framed two different modelling approaches, one that is more used in econometrics, the autoregressive, and another that is more used in statistics, the hierarchical, in the same mathematical context.

The models derived from the two different modelling strategies for count data, the hierarchical and the autoregressive ones, were accommodate in the same mathematical framework given by,

$$\begin{gathered} y_{it}|\eta_{it} \sim \text{Poisson}(e_{it}\theta_{it})), \\ \eta_{it} = \log(\theta_{it}) = \boldsymbol{X}_{it}^T\beta + u_{it}, \text{ i=1,...,n, t=1,...,T} \end{gathered} \tag{5.1}$$

with different representations of u_{it} corresponding to different model specifications. Within this common frame other possibilities can be also investigated for spatio-temporal modelling on the hierarchical and autoregressive approaches. For example, time effects can be further explored resorting to other possibilities for spatio-temporal modelling, as for example considering temporal effects prior distributed according to the time analogous of Leroux et al. [59] CAR distribution or by additionally considering a set of structured space-time interactions [56].

In this work, it was made a first application to road accidents in Lisbon [84, 85]. Within classical spatial econometrics methods, spatial-correlation was not found and the addition of spatial structure in the model did not improve estimation. During this analysis it was realized that the tests available for spatial autocorrelation heavily depend on a linear model assumption, which frequently is not appropriated. It is possibly needed a generalization of the tests for a vaster class of models as, for example, the generalized linear models in general and for spatial autoregressive models in particular [66]. So an alternative was implemented by modelling count data with a log-Poisson hierarchical model, including spatial random effects in the risk, in a Bayesian setting, implementing a different approach and finding some evidences of spatial structure in data. For these data, these effects revealed to be relevant, as expected, although the modelling was not yet achieved in a completely satisfactory way. This might be due to the degree of data aggregation that was used, which was the smallest administrative division of Lisboa. This could be explored in future work considering these data geo-referenced.

In what concerns to the second practical application [86], the analysis of the number of TAE calls to S24 at municipality level, in Continental Portugal, annually over the period 2010-2016, this was the first time these data have been modelled. Spatio-temporal autocorrelation was taken into account, since observations from geographically close spatial units and temporally close time periods tend to have more similar values, requiring the selection of a neighbourhood structure that specified the relations between regions to identify spatial association among them [11, 29].

Within the scope of the spatial econometric methods and also resorting to Bayesian hierarchical and autoregressive methodology, their application to study the number of TAE calls to the national health line, S24, for 2014, revealed spatial-correlation and the addition of spatial structure in the models improved estimation. In this study, count data were first analysed with a log-Poisson regression model. Then the inclusion of spatial random effects in a hierarchical Bayesian setting proved to be relevant, as expected, although the modelling may perhaps be improved by considering some other more adequate covariates. Aditionally, the recent alternative for doing Bayesian inference using INLA [81] for spatial econometric models for count data, the Bayesian Poisson spatial lag model, was also fitted to this data.

In terms of spatial hierarchical log-Poisson regression models, the preferred model was the one including the covariates and the spatially structured random effects through Leroux CAR prior. Similar conclusions were drawn when both the hierarchical and the autoregressive perspectives were considered, however the autoregressive model reveals better performance.

The average number of years of schooling and the average monthly income stands out as being important in explaining the use of S24. Also, the spatial component was quite relevant, which was confirmed by the high values of the estimates of the spatial autocorrelation parameter. This study further comprehended the analysis for each for the remaining available years, but as the results were similar, the corresponding analysis was omitted here. In order to be able to describe and evaluate in which municipalities the use of S24 should be encouraged, as well as detecting those regions that most contribute to the economic success of the good use of the line for future assessment of hospital savings [50], the analysis has been extended to include available data from years between 2010 and 2016. Fitting some spatio-temporal models [30] under an econometric approach, and developing and implementing the temporal effects on Bayesian hierarchical models [15, 56] and on Bayesian autoregressive models [25], for count data, a spatio-temporal Bayesian econometric approach for processing count data was carried out. The estimation results obtained for the S24 data are concordant in terms of the different hierarchical and autoregressive approaches, being slightly better for the latter, which is not perhaps the natural choice in terms of the methodology for this type of data. A spatial structure was evident from the analysis as well as a slight temporal trend. The considered space-time interaction effects did not seem to be of added value.

The significance of the inicial set of covariates differed from the simple log-Poisson regression model, the average number of years of schooling, the rurality index and the proportion of active population residents, stand out as being important in explaining the impact of S24 in the solving of non-urgent emerging situations mitigating the unnecessary urgent care in hospitals.

Considering different temporal structures considered, one can say that within the linear trend formulation similar estimates were obtained for the effects with the same type of structure for the two different hierarchical and autoregressive approaches. The main linear trend, which represents the global time effect, was significant, and displays the same estimated value for both perspectives. This allows to estimate which municipalities are exhibiting increasing or decreasing linear trends in the TAE savings calls.

When considering the dynamic formulation only one of the initial set of covariates reveal to be significant: the average number of years of schooling. This formulation improved the model fit, suggesting that it is more appropriate to these data.The results indicated that, one can say that the number of TAE savings calls in one municipality is somehow related to the number of calls in municipalities of its neighbourhoods, driven by the effects of the covariates.

The temporal correlation could not be assessed as well as the spatial correlation, due to the short time series of seven time periods, although, it was considered that it was important to include the temporal component and allowing an interaction term between time and space. The small number of time periods was also the reason for not considering more complex structures for space-time interaction effects. An alternative would have been to consider a different [59] CAR prior distribution for the spatial effects for each year, resulting in differentiated autocorrelation parameter for each year, and evaluating afterwards the time evolution of these estimates [56].

A savings index was considered, through which it was to realized that the non-urgent emergency situations phenomenon, responsible for major efficacy losses of the hospital systems, have been mitigated over the last seven years, through the creation, promotion and use of S24, with a more comprehensive spatially effectiveness of the S24 line in solving situations that could be handled by primary health care services or in a self care basis, avoiding unnecessary urgent care in hospitals.

This analysis could now to be extended by an econometric analysis of the savings obtained by the use of S24, in order to reach a more detailed health econometric analysis, developing an economic understanding of the advantages for the health system and learning about the political and economic factors that influence health policies at a global, national, regional and local levels. The main reason for the support of public health research is the common understanding that new knowledge leads to more effective health care, expecting that the upper mentioned econometric analysis contributes, in general, to an improvement by helping to reduce the per capita costs of the health care, through management policies [90].

For the future, it is also intended to proceed with the study of S24 data set, maybe considering some other possible relevant covariates and to carry out analyses under other possible scenarios. For example, through the development of a spatial-temporal Durbin model for count data, in a Bayesian setting, implementing a spatial lag on the count dependent variable as well as a spatial lag on the explanatory variables.

In this work the hyperparameters of the hierarchical spatial and spatio-temporal models were chosen so that the corresponding prior distributions were weakly informative. However, for these models, Simpson *el al.* (2017) [88], have recently proposed a penalised complexity (PC) type of prior distributions, an alternative that overcomes some of the problems associated with the inclusion of different type of random effects in a model, by taking the model structure into account. However these PC priors can not always be used as, for example, for Leroux model [59], as Leroux CAR prior makes a compromise between unstructured and structure variation using ρ as a mixing parameter. Consequently, it cannot be scaled because by construction scaling would depend on the value os ρ [77].

Another possibility is modelling which factors influence the individual decision of keeping or changing the individual decision of going to a hospital emergency after the call to S24, through a binary regression [26, 106].

Also the use of spatial point patterns models for these type of data is a possibility, since these constitute a valid alternative for understanding temporal and spatial effects. Under this perspective, the phenomenon is described through occurrences identified as points located in space, over time, with the objective of studying their spatial and temporal distribution and testing hypotheses about the observed pattern. It is intended to use the INLA method for spatio-temporal modeling on this perspective.

It is expected that analyses like the one considered and suggested here in terms of the data of road accidents in Lisbon or the S24 data contribute, in general, to the improvement of management policies in several areas of activity, the road safety and hospital domain in this case, or in others such as education, consumption, employment or demography.

BIBLIOGRAPHY

[1] H. Akaike. “Information theory and an extension of the maximum likelihood principle”. In: *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[2] J. Albert. *Bayesian Computation with R, Second edition*. Springer, 2009.

[3] L. Anselin. *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht, 1988.

[4] L. Anselin. “Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis”. In: *Journal of Regional Science* 30 (1990), pp. 185–207.

[5] L. Anselin. “Spatial econometrics”. In: *Econometric Theory, In Mills, T. and Patterson, K., editors, Palgrave Handbook of Econometrics: Volume 1*. Springer, 2006, 901—969.

[6] L. Anselin. *Spatial Regression Analysis in R - A Workbook*. University of ILLinois, Urbana-Champaign: Center for Spatially Integrated Social Sciences, 2007.

[7] L. Anselin. “Thirty Years of Spatial Econometrics”. In: *Papers in Regional Science* 89 (2010), pp. 3–25.

[8] L. Anselin and A. K. Bera. “Spatial dependence in Linear Regression Models with an Introduction to Spatial Econometrics”. In: *Statistics Textbooks and Monographs* 155 (1998), pp. 237–290.

[9] L. Anselin and R. Florax. “Small sample properties of tests for spatial dependence in regression models: Some further results”. In: *New directions in spatial econometrics*. Springer, 1995, pp. 21–74.

[10] L. Anselin and S. Rey. “Properties of tests for spatial dependency in linear regression models”. In: *Geographical Analysis* 23 (1991), pp. 112–131.

[11] L. Anselin, A. Bera, R. Florax, and M. Yoon. “Simple Diagnostic Tests for Spatial Dependence”. In: *Regional Science and Urban Economics* 26 (1996), pp. 77–104.

[12] L. Anselin, R. Florax, and S. Rey. “Econometrics for spatial models: recent advances”. In: *Advances in spatial econometrics*. Springer, 2004, pp. 1–25.

[13] L. Anselin, I. Syabri, and Y. Kho. “GeoDa: an introduction to spatial data analysis”. In: *Geographical analysis* 38.1 (2006), pp. 5–22.

[14] G. Arbia. *Spatial Econometrics, Statistical Foundations and Aplications to Regional Convergence*. Heidelberg and Berlin: Springer, 2006.

[15] S. Banerjee, B. Carlin, and A. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman and Hall/CRC, 2004.

[16] C. Belitz, A. Brezger, T. Kneib, S. Lang, and N. Umlauf. *BayesX: Software for Bayesian inference in structured additive regression models. Version 2.1*. 2015.

[17] L. Bernardinelli, D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini. “Bayesian analysis of space-time variation in disease risk”. In: *Statistics in Medicine* 14.21-22 (1995), pp. 2433–2443.

[18] J. Bernardo and A. Smith. *Bayesian Theory*. Wiley, 1994.

[19] J. Besag. “Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion)”. In: *Journal of the Royal Statistical Society B* 36, No.2 (1974), pp. 192–236.

[20] J. Besag and P. Moran. “On the estimation and testing of spatial interaction in Gaussian lattice processes”. In: *Biometrika* 62.3 (1975), pp. 555–562.

[21] J. Besag, J. York, and A. Mollié. “Bayesian Image Restoration with Two Applications in Spatial Statistics (with discussion)”. In: *Annals of the Institute of Statistical Mathematics* 43 (1991), pp. 1–59.

[22] R. Bivand. “Implementing representations of space in economic geography”. In: *Journal of Regional Science* 48.1 (2008), pp. 1–27.

[23] R. Bivand, V. Goméz-Rubio, and H. Rue. “Approximate Bayesian inference for spatial econometric models”. In: *Spatial Statistics* 9 (2014), pp. 146–165.

[24] R. Bivand, L. Anselin, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. Dormann, S. Dray, R. Halbersma, N. Lewin-Koh, et al. *spdep: Spatial dependence: weighting schemes, statistics and models*. 2014.

[25] M. Blangiardo and M. Cameletti. *Spatial and Spatial-temporal Bayesian Models with R-INLA*. Wiley, 2015.

[26] A. C. Cameron and P. K. Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.

[27] M. L. Carvalho and I. Natário. *Análise de Dados Espaciais*. Sociedade Portuguesa de Estatística, 2008.

[28] D. Clayton. “Bayesian methods for mapping disease risk”. In: *Geographical and environmental epidemiology: methods for small-area studies* (1992), pp. 205–220.

[29] N. Cressie. *Statistics for Spatial Data*. Jonh Wiley & Sons,Inc., 1993.

[30] N. Cressie and C. Wikle. *Statistics for Spatio-Temporal Data*. Jonh Wiley & Sons,Inc., 2011.

[31] S. Dass, C. Lim, and T. Maiti. *Experiences with Aproximate Bayes Inference for the Poisson-CAR Model*. Tech. rep. Techinal Report RM679. Michigan State University: Department of Statistics and Probability, 2010.

[32] *Database of contemporary Portugal, organized by Fundação Francisco Manuel dos Santos*. https://http://www.pordata.pt/, Accessed on 3 April 2016.

[33] A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001.

[34] *European Transport Safety Council. 5th Road Safety PIN Report. Road Safety Target Outcome: 100,000 fewer deaths since 2001*. Accessed in May 2012. Brussels: http://www.etsc.eu/documents/pin/report.pdf, 2011.

[35] M. Fischer. *Spatial Analysis and Geocomputation*. Vienna, Austria: Springer, 2006.

[36] D. Gamerman and H. Lopes. *Markov Chain Monte Carlo - Stocastic Simulation for Bayesian Inference*. Chapman and Hall, CRC, 2006.

[37] A. Gelman and D. Rubin. "Inference from Iterative Simulation Using Multiple Sequences (with discussion)". In: *Statistical Science* 7, No.4 (1992), pp. 457–511.

[38] A. Gelman, J. Hwang, and A. Vehtari. "Understanding predictive information criteria for Bayesian models". In: *J. Stat Comput* 24 (2014), pp. 997–1016.

[39] A. Gelman, S. Sturtz, U. Ligges, G. Gorjan, and J. Kerman. *Package 'R2WinBUGS'*. 2015.

[40] E. I. George and R. E. McCulloch. "Approaches for Bayesian variable selection". In: *Statistica sinica* (1997), pp. 339–373.

[41] J. Geweke. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Vol. 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN USA, 1991.

[42] V. Goméz-Rubio, R. Bivand, and H. Rue. "A new latent class to fit spatial econometrics models with Integrated Nested Laplace Approximations". In: *Spatial Statistics: Emmerging Patterns-Part2* 27 (2015), pp. 116–118.

[43] V. Goméz-Rubio, R. Bivand, and H. Rue. *Estimating Spatial Econometrics Models with Integrated Nested Laplace Approximations*. Tech. rep. Technical Report-Preprint to Elsevier, 2015.

[44] M. Goodchild, R. Haining, and S. Wise. "Integrating GIS and spatial data analysis: problems and possibilities". In: *International journal of geographical information systems* 6.5 (1992), pp. 407–423.

[45] M. Goodchild, L. Anselin, R. Appelbaum, and B. Harthorn. "Toward spatially integrated social science". In: *International Regional Science Review* 23.2 (2000), pp. 139–159.

[46] N. Gujarati. *Basic Econometrics*. McGrawHill, 1995.

[47] L. Held. "Simultaneous posterior probability statements from Monte Carlo output". In: *Journal of Computational and Graphical Statistics* 13.1 (2004), pp. 20–35.

[48] J. M. V. Hoef, E. M. Hanks, and M. B. Hooten. "On the Relationship between Conditional (CAR) and Simultaneous (SAR) Autoregressive Models". In: *arXiv preprint arXiv:1710.07000* (2017).

[49] P. Hoff. *A first course in Bayesian statistical methods*. Springer, 2009.

[50] D. Hughes and A. McGuire. "Stochastic demand, production responses and hospital costs". In: *ournal of Health Economics* 22 (2003), pp. 999–1010.

[51] G. Jacquez, P. Goovaerts, A. Kaufmann, and R. Rommel. "SpaceStat 4.0 user manual: software for the space-time analysis of dynamic complex systems". In: *Ann Arbor: BioMedware. Google Scholar* (2014).

[52] L. Knorr-Held. "Bayesian modelling of inseparable space-time variation in disease risk". In: *Statistics in medicine* 19 (2000), pp. 2555–2567.

[53] D. Lambert, J. Brown, and R. Florax. "A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application". In: *Regional Science and Urban Economics* 40 (2010), pp. 241–252.

[54] A. Lawson. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. CRC Press, Boca Raton, 2008.

[55] D. Lee. “CARBayes: an R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors”. In: *Journal of Statistical Software* 55, Issue 13 (2013), pp. 1–24.

[56] D. Lee, A. Rushworth, and G. Napier. *CARBayesST: An R Package for Spatio-temporal Areal Unit Modelling with Conditional Autoregressive Priors*. 2013.

[57] P. Lee. *Bayesian Statistics, An Introduction*. Wiley, 2012.

[58] B. Leroux, X. Lei, and N. Breslow. “Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence. In ME Halloran, D Berry (eds.)” In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials, Springer-Verlag, NewYork* (1999), pp. 135–178.

[59] B. Leroux, X. Lei, and N. Breslow. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Chapter Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence. Verlag, New York: Springer, 2000, pp. 179–191.

[60] J. LeSage. *The Theory and Practice of Spatial Econometrics*. University of Toledo, Toledo, Ohio, 1999.

[61] J. LeSage. “Software for Bayesian cross section and panel spatial model comparison”. In: *Journal of Geographical Systems* 17.4 (2015), pp. 297–310.

[62] J. LeSage and R. Pace. *Introduction to Spatial Econometrics*. Boca Raton: CRC Press, 2009.

[63] J. LeSage. “Spatial econometric panel data model specification: A Bayesian approach”. In: *Spatial Statistics* 9 (2014), pp. 122–145.

[64] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. “The BUGS project: evolution, critique and future directions”. In: *Statistics in medicine* 28.25 (2009), pp. 3049–3067.

[65] C. F. Manski. *Economic analysis of social interactions*. Tech. rep. National bureau of economic research, 2000.

[66] P. McCullagh and J. Nelder. *Generalized Linear Models, 2nd edition*. Boca Raton: Chapman and Hall, CRC Press, 1989.

[67] I. Natário. “Métodos Computacionais: INLA, Integrated Nested Laplace Approximation.” In: *Boletim da Sociedade Portuguesa de Estatística* Outono de 2013 (2013), pp. 52–56.

[68] A. Nunes. “Modelação Espacial de Acidentes Rodoviários na Cidade de Lisboa”. MA thesis. FCT-UNL, 2011.

[69] K. Ord. “Estimation methods for models of spatial interaction”. In: *Journal of the American Statistical Association* 70.349 (1975), pp. 120–126.

[70] C. Paulino, M. A. Turkman, and B. Murteira. *Estatística Bayesiana*. Calouste Gulbenkian Foundation, 2003.

[71] M. Plummer et al. “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling”. In: *Proceedings of the 3rd international workshop on distributed statistical computing*. Vol. 124. Vienna, Austria. 2003, p. 125.

[72] M. Plummer, N. Best, K. Cowles, and K. Vines. “CODA: Convergence Diagnosis and Output Analysis for MCMC”. In: *R News* 6, No.1 (2006), pp. 7–11.

[73] *Portal of the National Portuguese Health Service.* https://www.dgs.pt/paginas-de-sistema/saude-de-a-a-z/saude-24.aspx, Accessed on 3 September 2015.

[74] M. Quddus. "Modelling area-wide count outcomes with spatial correlation and heterogeneity - An analysis of London crash data". In: *Accident Analysis and Prevention* 40 (2008), pp. 1486–1497.

[75] A. Raftery and S. Lewis. "How Many Iterations in the Gibbs Sampler?" In: *in Bernardo, J.M., Berger, J., Dawid, A.P., Smith, A.M.F., Eds: Bayesian Statistics. Oxford University Press, Oxford* (1992), pp. 763–773.

[76] J. Rawlings, S. Pantula, and D. Dickey. *Applied Regression Analysis: A Research Tool, Second Edition*. Springer, 1998.

[77] A. Riebler, S. H. Sørbye, D. Simpson, and H. Rue. "An intuitive Bayesian spatial model for disease mapping that accounts for scaling". In: *Statistical methods in medical research* 25.4 (2016), pp. 1145–1165.

[78] B. Ripley. *Spatial Statistics*. Cambridge University Press, 1981.

[79] M. L. Rizzo. *Statistical computing with R*. CRC Press, 2007.

[80] C. Robert and G. Casella. *Introduction Monte Carlo Methods with R*. Springer, 2010.

[81] H. Rue, S. Martino, and N. Chopin. "Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion)". In: *Journal of the Royal Statistical Society, Series B* 71 (2009), pp. 319–392.

[82] H. Rue, S. Martino, and F. Lindgren. "The r-inla project". In: *R-INLA http://www. r-inla. org* (2012).

[83] P. Simões, M. L. Carvalho, S. Aleixo, S. Gomes, and I. Natário. "Spatio-temporal modelling the number of calls to a national health line for assessing hospital savings". In: *Submitted* ().

[84] P. Simões, S. Shrubsall, and I. Natário. "A Spatial Econometrics Analysis for Road Accidents in Lisbon". In: *Proceedings ICCSA2014, Springer* LNCS 8581, Part III (2014), pp. 269–286.

[85] P. Simões, S. Shrubsall, and I. Natário. "A Spatial Econometrics Analysis for Road Accidents in Lisbon *(extended version)*". In: *Int. J. Business Intelligence and Data Mining* 10, Issue 2 (2015), pp. 152–173.

[86] P. Simões, M. L. Carvalho, S. Aleixo, S. Gomes, and I. Natário. "A Spatial Econometric Analysis of the Calls to The Portuguese National Health Line". In: *Econometrics* 5,24, MDPI journals (2017).

[87] P. Simões and I. Natário. "Spatial econometric approaches for count data: an overview and new directions". In: *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 10.1 (2016), pp. 348–356.

[88] D. Simpson, H. Rue, A. Riebler, T. G. Martins, S. H. Sørbye, et al. "Penalising model component complexity: A principled, practical approach to constructing priors". In: *Statistical science* 32.1 (2017), pp. 1–28.

[89] B. Smith. "Bayesian output analysis program (BOA) for MCMC". In: *R package version* 1.5 (2004).

[90] B. Sobolev and A. Levy. *Comparative Effectiveness Research in Health Services*. Springer, 2016.

[91] D. Spiegelhalter, N. Best, B. Carlin, and A. Van der Linde. “Bayesian Measures of Model Complexity and Fit”. In: *Journal of the Royal Statistical Society B* 64, No.4 (2002), pp. 583–639.

[92] D. Spiegelhalter, A Thomas, N. Best, and D Lunn. *WinBUGS version 1.4 user manual. MRC Biostatistics Unit, Cambridge, UK.* 2003.

[97] Stan, Development, Team, et al. “Stan: A C++ library for probability and sampling”. In: *Online: http://mc-stan. org* (2014).

[93] H. Stern and N. Cressie. “Inference for Extremes in Disease Mapping”. In: *In A.B. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.F. Viel, R. Bertollini (Eds.): Disease Mapping and Risk Assessment for Public Health, John Wiley & Sons* (1999), pp. 63–84.

[94] H. Stern and N. Cressie. “Posterior predictive model checks for disease mapping models”. In: *Statistics in medicine* 19.17-18 (2000), pp. 2377–2397.

[95] R Team. “R development core team”. In: *RA Lang Environ Stat Comput* 55 (2013), pp. 275–286.

[96] R. C. Team and contributors worldwide. *The R Stats Package.* R package version 3.5.0 — URL http://CRAN.R-project.org/. 2013.

[98] M. A. Turkman and C. Paulino. *Estatística Bayesiana Computacional.* Sociedade Portuguesa de Estatística, 2015.

[99] M. A. Turkman and G. Silva. *Modelos Lineares Generalizados- da teoria á prática.* Sociedade Portuguesa de Estatística, 2000.

[100] M. Ugarte, A. Adin, T. Goicoa, and A. Militino. “On fitting spatio-temporal disease mapping models using approximate Bayesian inference”. In: *Statistical methods in medical research* 23.6 (2014), pp. 507–530.

[101] P. de Valpine, D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. “Programming with models: writing statistical algorithms for general model structures with NIMBLE”. In: *Journal of Computational and Graphical Statistics* 26.2 (2017), pp. 403–413.

[102] A. Vehtari, A. Gelman, and J. Gabry. “Pratical Bayesian model evaluation using leave-one-out cross-validation and WAIC.” In: *J. Stat Comput* (1999), pp. 1–20.

[103] M. M. Wall. “A close look at the spatial structure implied by the CAR and SAR models”. In: *Journal of statistical planning and inference* 121.2 (2004), pp. 311–324.

[104] S. Watanabe. “Asymptotic equivalence of Bayes cross-validation and widely applicable information criterion in singular learning theory”. In: *Journal of Machine Learning Research* 11 (2010), pp. 3571–3594.

[105] P. Whittle. “On stationary process in the plane”. In: *Biometrika* 41 (1954), pp. 434–449.

[106] J. Wooldridge. *Econometric analysis of cross section and panel data.* MIT press, 2010.

[107] J. Wooldrige. *Introductory Econometrics, A Modern Approach.* Thomson South Western, 2006.

APPENDIX A

APPENDICES

A.1 R-code used in R-INLA to implement the Bayesian Poisson spatial lag model

Start with the R-Code for implement the slmINLA fuction, available in `http://www.math.ntnu.no/inla/r-inla.org/doc/latent/slm.pdf`.

Listing A.1: The slmINLA function

```

## Index for the latent model
BD3$idx

## Define adjacency using a row-standardised matrix
concelhos_listw<-nb2listw(concelhos_nb)
W <- as(as_dgRMatrix_listw(concelhos_listw), "CsparseMatrix")

## Model definition
f1 <- n.chamadas ~Escolaridade + Perc.idosos+TxDesemp+IndRural+
  +Hospitais1000+CentrosSaude1000+ PropMulheres

f2 <- n.chamadas ~Perc.criancas + TxDesemp+Rendim+IndRural +
  +Hospitais1000+CentrosSaude1000+ PropMulheres

#Covariate matrix###
mmatrix1 <- model.matrix(f1,BD3)
mmatrix2<- model.matrix(f2,BD3)

## Zero-variance for error term
zero.variance = list(prec=list(initial = 25, fixed=TRUE))

## Compute eigenvalues for slm model, used to obtain rho.min and
## rho.max,
e = eigenw(concelhos_listw)
re.idx = which(abs(Im(e)) < 1e-6)
rho.max = 1/max(Re(e[re.idx]))
rho.min = 1/min(Re(e[re.idx]))
rho = mean(c(rho.min, rho.max))
```

```


## Precision matrix for beta coeffients' prior
betaprec <- .0001
Q.beta = Diagonal(n=ncol(mmatrix), betaprec)


## Priors on the hyperparameters
hyper = list(
  prec = list(
    prior = "loggamma",
    param = c(0.01, 0.01)),
  rho = list(
    initial=0,
    prior = "logitbeta",
    param = c(1,1)))
```

Next, there is the R-code for fitting the Bayesian Poisson spatial lag model, case 1 and 2 of analysis:

Listing A.2: Bayesian Poisson spatial lag model-Case1

```

## slpmINLA Model
## Fit model  for SCR case1
slmm1 <- inla( n.chamadas ~ -1 +
                 f(idx, model="slm",
                   args.slm=list(
                     rho.min = rho.min,
                     rho.max = rho.max,
                     W=W,
                     X=mmatrix1,
                     Q.beta=Q.beta),
                   hyper=hyper),
               data=BD3, family="poisson",
               control.family = list(hyper=zero.variance),
               control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
               offset=log(e_st)
)

summary(slmm1)

## Summary of the coefficients (at the end of the vector of random effects)
n <- nrow(BD3)
slmm1$summary.random$idx[n+1:ncol(mmatrix1),]
```

Listing A.3: Bayesian Poisson spatial lag model-Case2

```
slmm2 <- inla( n.chamadas ~ -1 +
                 f(idx, model="slm",
                   args.slm=list(
                     rho.min = rho.min,
                     rho.max = rho.max,
                     W=W,
                     X=mmatrix2,
                     Q.beta=Q.beta),
                   hyper=hyper),
               data=BD3, family="poisson",
```

```
               control.family = list(hyper=zero.variance),
               control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
               offset=log(e_st)
)
summary(slmm2)

## Summary of the coefficients (at the end of the vector of random effects)
n <- nrow(BD3)
slmm2$summary.random$idx[n+1:ncol(mmatrix2),]
```

A.2 R-code used in R-INLA to implement spatio-temporal Bayesian hierarchical log-Poisson models

First it is displayed here the R-Code that which is common to both models, the ST.Leroux.linear model and the ST.Leroux.dynamic model.

Listing A.4: The Bayesian Hierarchical log-Poisson models-Spatio-temporal analysis

```

## Index's for the models
BD5$ID<-as.numeric(rep(as.character(1:nrow(BD5.10))))
BD5$ID1<-as.numeric(rep(as.character(1:nrow(BD5.10))))


BD5$ID.Ano<-as.factor(BD5$Ano)
levels(BD5$ID.Ano)
levels(BD5$ID.Ano)<-c("1","2","3","4","5","6","7")
BD5$ID.Ano<-as.numeric(BD5$ID.Ano)
BD5$ID.Ano1<-BD5$ID.Ano


BD5$Ano.cent<-((BD5$Ano-mean(BD5$Ano))/7)
BD5$ID.Area.Ano<-as.numeric(1:1946)


#Spacetime object

times<-as.Date(as.character(2010:2016), format="%Y")
dados.S24<-STFDF(as(concelhos, "SpatialPolygons"),
               times,
               BD5)
names(dados.S24@data)

#Matrix ICAR-precision of instrinsic CAR specification

ICARmatrix <- Diagonal(nrow(concelhos_adj), apply(concelhos_adj, 1, sum))
- concelhos_adj
Cmatrix <- Diagonal(nrow(concelhos_adj), 1) - ICARmatrix

#\lambda.max
max(eigen(Cmatrix)$values)
```

Next, there is the R-code for fitting the Bayesian Hierarchical log-Poisson models in terms of the ST.Leroux.linear model and the ST.Leroux.dynamic Model.

Listing A.5: The St.Leroux.linear model-Spatio-temporal analysis

```

#### ST.Leroux.Linear:
#################################
m.leroux1.center = inla(n.chamadas.poup ~ 1 + Escolaridade +
                          + Perc.pop.Act+IndRural +
                          +Hospitais1000+CentrosSaude1000 +
                          f(ID, model = "generic1", Cmatrix = Cmatrix)+
                          +f(ID1,Ano.cent, model="iid")+Ano.cent,
                        data = as.data.frame(dados.S24),
                        E = dados.S24$e_st2, family ="poisson",
                        control.predictor = list(compute = TRUE),
                        control.compute = list(dic = TRUE, waic = TRUE))

summary(m.leroux1.center)
round(m.leroux1.center$summary.fixed[,1:5],3)
round(m.leroux1.center$summary.hyperpar[,1:5],3)
```

Listing A.6: The St.Leroux.dynamic model-Spatio-temporal analysis

```

#### ST.Leroux.dynamic:
#################################
m.leroux2  = inla(n.chamadas.poup ~ 1 + Escolaridade +
                   +Perc.pop.Act+IndRural +
                   +Hospitais1000+CentrosSaude1000 +
                   f(ID, model = "generic1", Cmatrix = Cmatrix)+
                   +f(ID.Ano, model="rw1")+f(ID.Area.Ano, model="iid"),
                 data = as.data.frame(dados.S24)),
                 E = dados.S24)$e_st2, family ="poisson",
                 control.predictor = list(compute = TRUE),
                 control.compute = list(dic = TRUE, waic = TRUE))

summary(m.leroux2)
round(m.leroux2$summary.fixed[,1:5],3)
round(m.leroux2$summary.hyperpar[,1:5],3)
```

A.3 R-code used in R-INLA to implement Bayesian Poisson spatio-temporal lag models

First it is displayed here the R-Code that is available in [86] to fit the slpmINLA, which is common to both models, the ST.splm.linear Model and the ST.splm.dynamic Model, adapted for the spatio-temporal case:

Listing A.7: The Bayesian Poisson spatial lag model- Spatio-temporal analysis

```
## Index's for the models
BD5$ID
BD5$ID1

BD5$Ano
BD5$Ano.cent


#Model Definition
f1 <- n.chamadas.poup ~Escolaridade + Perc.pop.Act+IndRural+
  +Hospitais1000+CentrosSaude1000
```

```

#Covariate matrix + intercept###
mmatrix.slm <- model.matrix(f1,BD5)

library("spdep")
#W
W.concelhos2 <- as(nb2mat(concelhos_nb, style = "W"), "Matrix")

W.concelhos.temp <- cBind (W.concelhos2,W.concelhos2,W.concelhos2,
W.concelhos2,W.concelhos2,W.concelhos2,W.concelhos2)
dim(W.concelhos.temp)

W.concelhos.temp2<- rBind
(W.concelhos.temp,W.concelhos.temp,W.concelhos.temp,W.concelhos.temp,
W.concelhos.temp,W.concelhos.temp,W.concelhos.temp)
dim(W.concelhos.temp2)

## Zero-variance for error term
zero.variance = list(prec=list(initial = 25, fixed=TRUE))

#Q
Q.beta = Diagonal(n = ncol(mmatrix.slm), x = 0.001)

#Range of rho
## Compute eigenvalues for SLM model, used to obtain rho.min and
## rho.max
e = eigenw(concelhos_listw)
re.idx = which(abs(Im(e)) < 1e-6)
rho.max = 1/max(Re(e[re.idx]))
rho.min = 1/min(Re(e[re.idx]))
rho = mean(c(rho.min, rho.max))


#Arguments for 'slm'
args.slm = list(
  rho.min = rho.min ,
  rho.max = rho.max,
  W=W.concelhos.temp2,
  X = mmatrix.slm,
  Q.beta = Q.beta
)
#Prior on rho
hyper.slm = list(
  prec = list(
    prior = "loggamma", param = c(0.01, 0.01)),
  rho = list(initial=0, prior = "logitbeta", param = c(1,1))
)
```

Next, there is the R-code for fitting the Bayesian Poisson spatial lag model with temporal effects in terms of the ST.splm.linear model and the ST.splm.dynamic Model.

Listing A.8: The ST.splm.linear Model- Spatio-temporal analysis

```
#### ST.splm.linear Model:
#################################
m.slm1.center <- inla( n.chamadas.poup ~ -1 +
 f(ID, model = "slm",args.slm = args.slm, hyper = hyper.slm)+
 f(ID1,Ano.cent, model="iid",constr = TRUE)+Ano.cent,
  data = as.data.frame(BD5), family = "poisson",
  offset=log(e_st2),
   control.predictor = list(compute = TRUE),
```

```
   control.compute = list(dic = TRUE, waic = TRUE)
)
summary(m.slm1.center)
round(m.slm1.center$summary.hyperpar[,1:5],2)
round(m.slm1$summary.fixed[,1:5],3)

## Summary of the coefficients (at the end of the vector of random effects)

n <- nrow(BD5)
round(m.slm1.center$summary.random$ID[n+1:ncol(mmatrix.slm),],3)
```

Listing A.9: The ST.splm.dynamic Model- Spatio-temporal analysis

```
#### ST.splm.dynamic Model:
#####################################
m.slm2.1 <- inla( n.chamadas.poup ~ -1 +
   f(ID, model = "slm",args.slm = args.slm, hyper = hyper.slm)+
   f(ID.Ano, model="rw1")+f(ID.Area.Ano, model="iid"),
   data = as.data.frame(BD5), family = "poisson",
    offset=log(e_st2),
     control.predictor = list(compute = TRUE),
     control.compute = list(dic = TRUE, waic = TRUE),
     control.results = list(return.marginals.predictor=TRUE)
)
summary(m.slm2.1)
round(m.slm2.1$summary.hyperpar[,1:5],2)

## Summary of the coefficients (at the end of the vector of random effects)
n <- nrow(BD5)
round(m.slm2.1$summary.random$ID[n+1:ncol(mmatrix.slm),],3)
```