

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

**Modelação de Dados Espaço-Temporais da COVID-19
em Portugal com Desagregação Temporal: Uma
Abordagem Estatística e Epidemiológica**

Maria da Conceição Dias Leal

Doutoramento em Matemática Aplicada e Modelação

Orientadora: Professora Doutora Teresa Oliveira

Coorientador: Professor Doutor Amílcar Oliveira

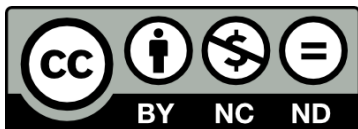
Coorientador: Professor Doutor Amitava Mukherjee

Outubro de 2024

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Esta obra está licenciada sob uma Licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 Internacional.

Pode partilhar esta obra, desde que atribua o devido crédito ao autor, mas não pode alterá-la de nenhuma forma, nem utilizá-la para fins comerciais.



CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Agradecimentos

A concretização de mais um sonho, de mais um projeto ou de mais um desafio nunca acontece se não for o resultado de um convergir de vontades.

A concretização do projeto aqui apresentado é o resultado disso mesmo. De uma recolha de vontades que se aliaram à minha!

À Professora Doutora Teresa A. Oliveira agradeço a vontade imensa de me fazer ir cada vez mais longe e o apoio nessa caminhada.

Ao Professor Doutor Amílcar Oliveira agradeço a vontade de ajudar, manifestada na disponibilidade com que sempre acolheu as minhas dúvidas.

Ao Professor Doutor Amitava Mukherjee agradeço a vontade de ajudar, manifestada na disponibilidade com que sempre acolheu as minhas dúvidas.

A todos os Professores e Colegas que caminharam comigo agradeço a vontade de partilharem os seus conhecimentos e as suas experiências.

Às minhas pessoas, que não preciso nomear, agradeço a vontade de partilharem o seu tempo e me apoiarem e estimularem em todas as etapas.

Aos meus pais agradeço a vontade de fazerem comigo, em todos os momentos e incondicionalmente, o caminho difícil e exigente, para mim e para eles, que culminou neste trabalho.

À minha pessoa especial agradeço a confiança, o apoio, o estímulo, os diálogos. O estar ..., sempre!

*Aos meus pais, sempre
Às minhas pessoas especiais...
À minha pessoa muito especial*





DECLARAÇÃO DE INTEGRIDADE

STATEMENT OF INTEGRITY

Declaro ter atuado com integridade na elaboração da presente dissertação/tese. Confirmando que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer outra forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Regulamento Disciplinar da Universidade Aberta, publicado no Diário da República, 2.ª série, n.º 215, de 6 de novembro de 2013.

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.

I further declare that I have fully acknowledged Disciplinary Regulations of the Universidade Aberta (regulation published in the official journal Diário da República, 2.ª série, N.º 215, de 6 de novembro de 2013).

Universidade Aberta, 31 de outubro de 2024

Nome completo/Full name: Maria da Conceição Dias Leal

Assinatura/Signature:

Assinado por: **Maria da Conceição Dias Leal**
Num. de identificação: 08201878
Data: 2024.10.31 20:20:19+00'00'

manuscrita ou digital / handwritten or digital

A monitorização estatística analisa dados recolhidos ao longo do tempo e/ou espaço, para detetar e quantificar alterações, anomalias ou tendências nos processos subjacentes. Esta tese desenvolve e aplica métodos estatísticos para monitorizar dados espaço-temporais, considerando as suas especificidades e desafios.

Usando como caso prático a COVID-19 em concelhos de Portugal Continental, o trabalho visa compreender a dinâmica de disseminação de uma epidemia, apoiar a tomada de decisão em Saúde Pública e contribuir com metodologias inovadoras em estatística.

A tese propõe três contributos principais: 1) desagregação temporal de dados, em que se obtém dados mais detalhados e confiáveis a partir de dados de diferentes fontes que os recolhem e tratam com critérios distintos, recorrendo a algoritmos que permitem infundir informação relevante do domínio; 2) modelação de dados, considerando efeitos espaciais, temporais e de interação *espaço × tempo*, com recurso ao ajuste de modelos hierárquicos bayesianos aos dados observados e desagregados, para análise e deteção de padrões temporais e espaciais; e 3) análise e deteção de fatores de risco e avaliação do impacto de eventos públicos e de covariáveis meteorológicas na disseminação de uma epidemia/pandemia, usando modelos lineares generalizados, visualização e testes de hipóteses. A tese explora dois casos relevantes com dados de COVID-19 de concelhos de Portugal Continental, criando formas mais sólidas de análise e monitorização dos dados.

A tese demonstra como a matemática e a epidemiologia se unem para enfrentar desafios da humanidade, como a pandemia COVID-19, avançando o conhecimento científico e melhorando as políticas públicas de saúde e outros processos.

Palavras-chave: COVID-19, Dados espaciais, Dados espaço-temporais, Desagregação de dados temporais, Epidemiologia, Modelação de dados espaço-temporais.

ABSTRACT

Statistical monitoring analyses data collected over time and/or space to detect and quantify changes, anomalies, or trends in underlying processes. This thesis develops and applies statistical methods for monitoring spatiotemporal data, considering their specificities and challenges.

Using COVID-19 in counties of Continental Portugal as a practical case, the work aims to understand the dynamics of epidemic spread, support decision-making in Public Health, and contribute with innovative statistical methodologies.

The thesis proposes three main contributions: 1) the temporal disaggregation of data, which obtains more detailed and reliable data, combining data from different sources that collect and treat them with different criteria, using algorithms that allow infusing relevant information from the domain; 2) data modelling, considering spatial, temporal, and *space* \times *time* interaction effects, using the adjustment of Hierarchical Bayesian Models to observed and disaggregated data for analysis and detection of temporal and spatial patterns; and 3) analysis and detection of risk factors and evaluation of the impact of public events and meteorological covariates on the spread of an epidemic/pandemic, using Generalized Linear Models, visualization, and hypothesis testing. The thesis explores two relevant cases with COVID-19 data from counties of Continental Portugal, creating more robust forms of analysis and monitoring of the data.

The thesis demonstrates how mathematics and epidemiology come together to address humanity's challenges, such as the COVID-19 pandemic, advancing scientific knowledge and improving public health policies and other processes.

The thesis demonstrates how mathematics and epidemiology come together to face challenges of humanity, such as the COVID-19 pandemic, advancing scientific knowledge and improving public health policies and other processes.

Keywords: COVID-19, Spatial data, Spatiotemporal data, Temporal data disaggregation, Epidemiology, Spatiotemporal data modelling.

Resumo	v
Abstract	vii
Índice	viii
Índice de tabelas.....	xi
Índice de figuras.....	xiv
NOMENCLATURA.....	xix
Capítulo 1.....	1
1. INTRODUÇÃO	2
1.1. Motivação	2
1.2. Estrutura da tese.....	5
Capítulo 2.....	8
2. ENQUADRAMENTO TEÓRICO.....	9
2.1. Modelos de crescimento	9
2.2. Modelos Lineares Generalizados (<i>Generalized Linear Models</i>).....	11
2.3. Processos e dados espaço-temporais.....	20
2.3.1. Processos espaciais	21
2.3.2. Processos espaço-temporais.....	24
2.3.3. Especificação de processos <i>lattice</i> e dados estatísticos associados (dados de área e dados <i>lattice</i>)	25
2.4. Métodos Bayesianos – modelos para dados espaciais e espaço-temporais.....	26
2.4.1. Inferência paramétrica Bayesiana: distribuição <i>a priori</i> e distribuição <i>a posteriori</i> 29	
2.4.2. Métodos computacionais de apoio à Inferência Bayesiana	37
2.4.3. INLA - Integrated Nested Laplace Approximations (AmaralTurkman et al., 2019; Blangiardo & Cameletti, 2015).....	41
2.5. Modelos Hierárquicos Bayesianos para dados contagem	45

2.6. Índices de comparação de modelos não aninhados baseados na Teoria da Informação	50
2.7. Avaliação da adequação de modelos.....	52
2.8. Testes à homogeneidade dos coeficientes dos modelos	53
2.9. Extração de dados por medição de gráfico de barras com distorção visual.....	55
2.10.Desagregação de Séries Temporais	57
2.10.1. Agregação de séries espaço-temporais	58
2.10.2. Desagregação (temporal) de séries temporais.....	60
2.10.3. Modelos de desagregação temporal	61
2.10.4. Métodos matemáticos de desagregação temporal.....	63
2.10.5. Métodos de regressão na desagregação de séries	67
2.10.6. Outros métodos	70
2.10.7. Desagregação de séries temporais com possibilidade de períodos sobrepostos	71
Capítulo 3.....	89
3. MODELAÇÃO DE DADOS ESPAÇO TEMPORAIS DE COVID-19 DA ÁREA METROLOPITANA DE LISBOA	90
3.1. Introdução	90
3.2. Dados de COVID-19 dos concelhos da Área Metropolitana de Lisboa.....	93
3.2.1. De onde vêm, o que são e como são	93
3.2.2. Dados demográficos da Área Metropolitana de Lisboa.....	94
3.3. Manifestação do 1.º de Maio	95
3.3.1. Caracterização dos dados	102
3.3.2. Estimação do <i>Reporting Delay</i>	107
3.3.3. Modelação e análise com modelos de regressão de Poisson	112
3.3.4. Discussão	127
3.4. Manifestação antirracismo.....	132
3.4.1. Caracterização do caso de estudo	132
3.4.2. Modelação, análise e discussão dos resultados.....	133
Capítulo 4.....	145
4. DESAGREGAÇÃO DE SÉRIES TEMPORAIS: DADOS DA COVID-19 EM PORTUGAL ...	146

4.1. Caracterização dos dados de COVID-19 disponibilizados pelas entidades públicas de saúde portuguesas.....	148
4.2. Comparação dos dados reais e dos dados desagregados dos dados reais dos concelhos do distrito do Porto, com o método do Mínimos Quadrados com pseudo-inversa de Moore-Penrose e com o método H-Fuse.....	150
Capítulo 5.....	163
5. MODELAÇÃO DE DADOS ESPAÇO-TEMPORAIS DE COVID-19 – CONCELHOS DE PORTUGAL CONTINENTAL	164
5.1. Introdução	164
5.2. Caracterização dos dados.....	166
5.3. Desagregação da série temporal nos períodos de 7 de julho a 26 de outubro de 2020 e de 27 de outubro de 2020 a 23 de março de 2021.....	170
5.3.1. Desagregação de dados pressupondo a distribuição uniforme	173
5.3.2. Desagregação recorrendo aos algoritmos dos Mínimos Quadrados e H-FUSE descritos na secção 3.4.....	175
5.4. Modelação espaço-temporal Bayesiana avançada de nível local por meio da estrutura multinível para estratégias de prevenção e controle de COVID-19.	180
5.4.1. Especificação do modelo	181
5.4.2. Resultados e discussão	185
Capítulo 6.....	202
6. CONCLUSÕES E CONSIDERAÇÕES FINAIS	203
REFERÊNCIAS BIBLIOGRÁFICAS.....	210

Tabela 3.1: Variáveis em análise.	106
Tabela 3.2: Critérios de bondade de ajustamento.	111
Tabela 3.3: Coeficientes dos modelos de Poisson com função de ligação de Identidade ajustados aos dados da AML e do resto do país.	114
Tabela 3.4: Coeficientes dos modelos de Poisson com função de ligação de <i>log</i> , ajustados aos dados da AML e do resto do país.....	116
Tabela 3.5: Coeficientes dos modelos $Po(Id)(S1)$, $Po(Id)(S3)$ e $Po(Id)$ (4.9), para cada concelho.119	
Tabela 3.6: <i>p</i> – values dos testes ANCOVA à homogeneidade dos coeficientes de <i>t</i> dos modelos $Po(Id)$ e $Po(Log)$ para os períodos S1 e S3 (hipóteses definidas em (3.12)).....	125
Tabela 3.7: <i>p</i> – values dos testes de Razão de Verossimilhança para comparação dos coeficientes de <i>t</i> dos modelos $Po(Id)$ e $Po(Log)$ (<i>pvalue</i> – pacote <i>parameters</i>) para os períodos S1 e S3 (hipóteses definidas em (3.13))	126
Tabela 3.8: Valores AIC para encontrar o modelo que melhor se ajusta aos dados para ambos os períodos, S1 e S3 (refletindo o crescimento linear ou exponencial).	129
Tabela 3.9: Resultados do teste exato de Fisher à independência entre a alteração do padrão de crescimento no número cumulativo de casos de COVID-19 e cada uma das cinco covariáveis.....	131
Tabela 3.10: Coeficientes dos modelos de Poisson com função de ligação de Identidade ajustados aos dados da AML e do resto do país dos períodos S1 (21/05/2020 a 11/06/2020) e S3(21/06/20 a 12/07/2020).	134
Tabela 3.11: Coeficientes dos modelos $Po(Id)$ (modelo de regressão de Poisson com função de ligação Identidade e com interação <i>Tempo</i> × <i>Período</i> , e $Po(Log)$ (modelo de regressão de Poisson com função de ligação Logaritmo e com interação <i>Tempo</i> × <i>Período</i> , para cada concelho. O coeficiente β_1 corresponde à taxa de crescimento do número de casos do período S1 e o coeficiente γ corresponde ao acréscimo de β_1 do período S1 para o período S3, ou seja, dá a variação na taxa de crescimento de um período para o outro. O coeficiente β_1 corresponde ao acréscimo do interseto do período S1 para o interseto do período S3.....	137
Tabela 3.12: <i>p</i> – values dos testes ANCOVA OneWay (Tipo III) à homogeneidade dos coeficientes de <i>t</i> dos modelos $Po(Id)$ e $Po(Log)$ para os períodos S1 e S3 (hipóteses definidas em 3.17).....	142

Tabela 3.13: p – values dos testes de Razão de Verosimilhança para comparação dos coeficientes de t dos modelos $Po(Id)$ e $Po(Log)$ ($pvalue$ – pacote $parameters$) para os períodos S1 e S3 (hipóteses definidas em (3.18)	144
Tabela 4.1: RMSE (Root Mean Square Error) dos diferentes modelos de desagregação, padronizado pelo desvio padrão,	156
Tabela 4.2: MAE (Mean Absolut Error) dos diferentes modelos de desagregação, padronizado pelo desvio padrão,	156
Tabela 5.1 Variáveis em análise.....	170
Tabela 5.2: Definição das variáveis usadas no modelo.....	184
Tabela 5.3: Modelo sem covariáveis ajustados aos dados. μ é a o intesepto, E é o número de casos esperado, u_j e v_j representam os efeitos espaciais estruturados e não estruturados, γ_t e ϕ_t são os efeitos temporais estruturados e não estruturados e δ_{jt} é o efeito de interação <i>espaço</i> \times <i>tempo</i>	185
Tabela 5.4: Os resultados de DIC, WAIC e Log-Likelihood marginal dos modelos apenas com efeitos temporais e espaciais (estruturados e não estruturados), ajustados com os dados observados e com os dados agregados semanalmente e desagregados pelos métodos LS e HF.	185
Tabela 5.5: μ é o intesepto do modelo, E é o número de casos esperado, $X_i, i = 1, 2, \dots, 29$ são as covariáveis e u_j e v_j representam os efeitos espaciais estruturados e não estruturados, γ_t e ϕ_t são os efeitos temporais estruturados e não estruturados e δ_{jt} é o efeito de interação <i>espaço</i> \times <i>tempo</i>	186
Tabela 5.6: Os resultados dos valores de DIC, WAIC e Log-Likelihood marginal do modelo geral ajustado com os dados observados e com os dados agregados à semana e desagregados pelos métodos LS e HF. Os modelos foram ajustados com todas as covariáveis, e com os efeitos temporais e espaciais.	186
Tabela 5.7: Matriz de correlações entre as variáveis meteorológicas usadas.....	187
Tabela 5.8: μ é o intesepto do modelo, E é o número de casos esperado, u_j e v_j representam os efeitos espaciais estruturados e não estruturados e γ_t e ϕ_t são os efeitos temporais estruturados e não estruturados.....	187
Tabela 5.9: Média posterior e intervalos de credibilidade (a 95%) dos parâmetros e hiperparâmetros do modelo $bym.rw.iid.M.2$, ajustado com o número diário de casos COVID-19 observados nos concelhos do distrito do Porto, entre 25/03/2020 e 30/06/20 e com as covariáveis meteorológicas com atraso de dois dias.	190
Tabela 5.10: Média posterior e intervalos de credibilidade (a 95%) dos efeitos fixos e dos hiperparâmetros do modelo $bym.rw.iid.M.2$, ajustado com o número diário de casos	

	COVID-19 desagregados dos dados agregados à semana com recurso ao algoritmo HF, nos concelhos do distrito do Porto, entre 25/03/2020 e 30/06/20 e com as variáveis meteorológicas com atraso de dois dias.	191
Tabela 5.11:	Média posterior e intervalos de credibilidade (a 95%) dos efeitos fixos e dos hiperparâmetros do modelo bym.rw.iid.M.2, ajustado com o número diário de casos COVID-19 desagregados dos dados agregados à semana com recurso ao algoritmo LS, nos concelhos do distrito do Porto, entre 25/03/2020 e 30/06/20 e com as variáveis meteorológicas com atraso de dois dias.	192
Tabela 5.12:	Média posterior e intervalos de credibilidade (a 95%) dos efeitos fixos e dos hiperparâmetros do modelo bym.rw.iid.M.2, ajustado com o número diário de casos COVID-19 dos concelhos de Portugal Continental, observados no período de 25/03/2020 a 4/07/2020 e desagregados pelo algoritmo H-Fuse no período de 05/07/2020 a 23/03/2021.	198

Figura 2.1: (a) Exemplo de medição de amostras; (b) Detalhe de gráfico de barras empilhadas com borrão visível.....	56
Figura 2.2: (a) Períodos temporais consecutivos com sobreposição; (b) Períodos temporais consecutivos com lacuna; (c) Períodos temporais consecutivos com conflito (o período menor tem um número maior de dados que o período maior).	71
Figura 3.1: Mapa da Área Metropolitana de Lisboa. Composição de vários mapas da entrada da Wikipedia AML.	94
Figura 3.2: Novos casos diários em cada uma das três principais regiões de Portugal, com indicação de 1 de Maio, quando a manifestação foi realizada em Lisboa.....	96
Figura 3.3: a) Modelo padrão de Richards com seis parâmetros, ajustado com número cumulativo de casos observado para o concelho de Lisboa entre 15/4 e 31/5; b) Modelo de Richards com seis parâmetros, ajustado com número cumulativo de casos observado para o concelho de Lisboa entre 15/4 e 31/5.....	97
Figura 3.4: Fluxograma que ilustra a metodologia implementada no artigo para a análise de dados dos concelhos da AML.	101
Figura 3.5: Evolução do número de casos por 100 000 habitantes observado no período de 15 de abril a 6 de junho de 2020 em cada um dos concelhos da AML. A curva a preto representa a evolução do número de casos por 100 000 habitantes na AML.....	102
Figura 3.6: Valores das covariáveis, por concelho: <i>Densidade populacional, Índice de envelhecimento, Distância do carro até ao local de demonstração em Lisboa, Tempo de viagem até ao local de demonstração em Lisboa, e Índice de mobilidade até Lisboa</i>	103
Figura 3.7: Representação gráfica da distribuição empírica do <i>Número cumulativo de casos</i> , em relação ao número de dias entre 15/4 e 6/6 de 2020.	104
Figura 3.8: Distribuição geográfica, por concelhos, de covariáveis, maximizando as semelhanças internas (Anselin et al., 2010).	105
Figura 3.9: Dados visuais fornecidos publicamente através de relatórios diários da DGS. (a) Amostra de gráfico dos relatórios diários (Relatório 52) (b) Amostra de pixels amplificados, mostrando distorção.....	109
Figura 3.10: Os gráficos de Shewhart para taxas de variação diárias do número de casos notificados (w) com início de sintomas em (a) 17 de Abril, (b) 27 de Abril, e (c) 6 de Maio.	110

Figura 3.11: As funções PDF e CDF ajustadas aos dados de atraso de comunicação (<i>reporting delay</i>).	111
Figura 3.12: Modelos de regressão Binomial Negativo/ Poisson. ajustados às séries temporais dos períodos S1 e S3. (a) Modelo Binomial Negativo com função de ligação Identidade para Portugal Continental sem os concelhos da AML; (b) Modelo com função de ligação Log para a AML. A hipótese de homogeneidade dos declives não foi rejeitada nos dois casos, com um nível de significância de 5%.	116
Figura 3.13: Modelos de regressão de Poisson, com função de ligação Identidade, ajustado com dados de $S1 \cup S3$, para cada concelho da Área Metropolitana de Lisboa ($i = 1, 2, \dots, 18$), para cada uma das subséries S1 e S3.....	120
Figura 3.14: Modelos de regressão Poisson com função de ligação Identidade, ajustados com dados de S1 e S3.....	121
Figura 3.15: Modelos de regressão de Poisson com função de ligação <i>log</i> , para a série $S1 \cup S3i$ e as séries $S1i$ e $S3i$, sendo i o i –ésimo concelho, $i = 1, 2, \dots, 18$	124
Figura 3.16: Modelos lineares (à esquerda) e exponenciais (à direita), ajustados a S1-extendido (3/24-5/6) e S3 (5/16-6/6).	128
Figura 3.17: (a) Grupo 1: Laranja-claro; Grupo 2: Laranja-escuro. (b) Concelhos com mudança de padrão: Azul-escuro; Concelhos sem mudança de padrão: Azul-claro.....	131
Figura 3.18: Demonstração “ <i>Black Lives Matter</i> ”: (a) Modelo de regressão de Poisson com função de ligação <i>Identidade</i> para Portugal Continental sem os concelhos da AML; (b) Modelo de regressão Binomial Negativa com função de ligação <i>Identidade</i> para a AML. O teste ANOVA rejeitou a hipótese de homogeneidade dos declives nos dois casos, com um nível de significância de 5%. O teste GLRT rejeitou a hipótese da igualdade dos declives dois casos contra a hipótese alternativa de o declive do modelo ajustado a S1 ser menor que o declive do modelo ajustado a S3.....	133
Figura 3.19: Modelos de regressão de Poisson, com função de ligação Identidade, ajustado com dados de $S1 \cup S3$, para cada concelho da Área Metropolitana de Lisboa ($i = 1, 2, \dots, 18$) e para cada uma das subséries S1 e S3.	138
Figura 3.20: Modelos de regressão Poisson com função de ligação Identidade, ajustados com dados de S1 e S3.....	139
Figura 3.21: Modelos de regressão de Poisson com função de ligação <i>log</i> , para a série $S1 \cup S3i$ e as séries $S1i$ e $S3i$, sendo i o i –ésimo concelho, $i = 1, 2, \dots, 18$	140
Figura 4.1: Concelhos do distrito do Porto – Portugal.	151
Figura 4.2: Dados de casos COVID-19 de 25/03/2020 a 30/06/2020, dos concelhos do distrito do Porto, agregados semanalmente.	151

Figura 4.3: Dados de casos COVID-19 de 25/03/2020 a 30/06/2020, dos concelhos do distrito do Porto, agregados pela soma dos últimos 14 dias e relatados semanalmente – sobreposição da segunda semana da quinzena anterior com a primeira semana da quinzena atual.	152
Figura 4.4: Dados diários observados nos 18 concelhos do Distrito do Porto de 25/03/2020 a 30/06/2020	153
Figura 4.5: Dados de frequência semanal desagregados com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.....	153
Figura 4.6: Dados de frequência semanal desagregados com recurso ao algoritmo H-Fuse.	154
Figura 4.7: Dados acumulados a 14 dias reportados com frequência semanal, desagregados com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.	154
Figura 4.8: Dados acumulados a 14 dias, repostados com frequência semanal, desagregados com recurso ao algoritmo H-Fuse.....	155
Figura 4.9: Mapas de risco dos dados observados, e os dados desagregados com os algoritmos LS e HF de dados agregados semanalmente, para os dias 30 e 31 de março, 1 e 2 de abril de 2020.....	159
Figura 4.10: Mapas de risco dos dados observados, e os dados desagregados com os algoritmos LS e HF de dados agregados acumulados a 14 dias, para os dias 30 e 31 de março, 1 e 2 de abril de 2020	160
Figura 4.11: Mapas de risco dos dados observados e os dados desagregados com os algoritmos LS e HF de dados agregados semanalmente, para dias compreendidos entre 15 de maio e 18 de maio de 2020.....	161
Figura 4.12: Mapas de risco dos dados observados e dos dados desagregados com os algoritmos LS e HF de dados agregados acumulados a 14 dias, para dias compreendidos entre 15 de maio e 18 de maio de 2020.....	162
Figura 5.1: Dados semanais de COVID-19, por concelho do distrito do Porto, observados no período de 6/7/2020 a 26/10/2020.....	171
Figura 5.2: Número de casos semanais de COVID-19, acumulados a 14 dias, por concelho do distrito do Porto, observados de 14/7/2020 a 26/10/2020.	171
Figura 5.3: Número de casos semanais de COVID-19, acumulados a 14 dias, por concelho do distrito do Porto, observados de 11/11/2020 a 23/03/2021.	172
Figura 5.4: Número de casos semanais de COVID-19, acumulados a 14 dias, por concelho do distrito do Porto, observados de 27/7/2020 a 23/03/2021.	172
Figura 5.5: Linha do tempo que exemplifica a sobreposição de quinzenas. A caixa lilás representa a quinzena $i - 1$ com $ad + 7i - 1$ novos casos diários estimados dessa quinzena e a caixa vermelha	

representa a quinzena i com adi novos casos diários estimados da respetiva quinzena. x_{di} representa o número diário de novos casos estimados no dia d da semana i	174
Figura 5.6: Dados diários do período de 07/07/2020 a 26/10/2020, desagregados de dados de frequência semanal com recurso ao à distribuição uniforme $U(0,7)$	175
Figura 5.7: Dados diários, do período de 27/10/2020 a 23/03/2021, desagregados de dados acumulados a 14 dias, relatados semanalmente, com recurso à distribuição uniforme $U(0,14)$...	175
Figura 5.8: Cada segmento de reta corresponde a uma semana, ou seja, corresponde a sete períodos da escala fina diária.....	176
Figura 5.9: Dados diários do período de 07/07/2020 a 26/10/2020, desagregados de dados de frequência semanal com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.....	177
Figura 5.10: Dados diários do período de 07/07/2020 a 26/10/2020, desagregados de dados de frequência semanal com recurso ao algoritmo H-Fuse.....	177
Figura 5.11: Cada segmento de reta corresponde a 14 dias. Há sobreposição de uma semana em cada 14 dias e os relatórios não foram sempre disponibilizados a intervalos regulares.....	178
Figura 5.12: Dados diários, do período de 27/10/2020 a 23/03/2021, desagregados de dados acumulados a 14 dias, relatados semanalmente, com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.....	179
Figura 5.13: Dados diários, do período de 27/10/2020 a 23/03/2021, desagregados de dados acumulados a 14 dias, relatados semanalmente, com recurso ao algoritmo H-Fuse.....	179
Figura 5.14: Grafo direto acíclico que ilustra a estrutura hierárquica do modelo implementado ..	184
Figura 5.15: Distribuição posterior do coeficiente da covariável: A - Temperatura máxima média; B - Temperatura mínima média; C - Humidade máxima média; D – Humidade mínima média; E - Pluviosidade média.	190
Figura 5.16: Mapas de risco nos concelhos do distrito do Porto estimados pelos modelos iCAR.wr.iid ajustados com os dados observados, e os dados desagregados com os algoritmos LS e HF de dados agregados semanalmente, para os dias 30 e 31 de março, 1 e 2 de abril de 2020, com lag 2 para as covariáveis meteorológicas.	194
Figura 5.17: Mapas de risco nos concelhos do distrito do Porto estimados pelos modelos iCAR.wr.iid ajustados com os dados observados, e com os dados desagregados de dados agregados semanalmente, com os algoritmos LS e HF, para os dias 15,16, 17 e 18 de maio de 2020, com lag 2 para as covariáveis meteorológicas.	195
Figura 5.18: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de agosto de 2020.	196

Figura 5.19: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de novembro de 2020.....	196
Figura 5.20: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de janeiro de 2021.....	197
Figura 5.21: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de fevereiro de 2021.	197
Figura 5.22: Mapa da distribuição do risco para (A) $SIR > 1$ (n° de casos observados $> n^\circ$ de casos esperados) e para $SIR \leq 1$ (n° de casos observados $\leq n^\circ$ de casos esperados) e (B) risco estimado pelo modelo, em 10/08/2020.....	200
Figura 5.23: Mapa da distribuição do risco para (A) $SIR > 1$ (n° observado de casos $> n^\circ$ esperado de casos) e para $SIR \leq 1$ (n° de casos observados $\leq n^\circ$ de casos esperados) e (B) risco estimado pelo modelo, em 10/11/2020.....	200
Figura 5.24: Mapa da distribuição do risco para (A) $SIR > 1$ (n° de casos observados $> n^\circ$ de casos esperados) e para $SIR \leq 1$ (n° de casos observados $\leq n^\circ$ de casos esperados) e (B) risco estimado pelo modelo, em 10/01/2021.....	201
Figura 5.25: Mapa da distribuição do risco para (A) $SIR > 1$ (n° de casos observado $> n^\circ$ de casos esperado) e para $SIR \leq 1$ (n° de casos observado $\leq n^\circ$ de casos esperado) e (B) risco estimado pelo modelo, em 10/02/2021.....	201

NOMENCLATURA

DGS Direção Geral de Saúde

DSSG PT Data Science for Social Good Portugal

AML Área Metropolitana de Lisboa

COVID-19 Coronavirus Disease 19

SARS-CoV-2 Severe acute respiratory syndrome coronavirus 2

SIR Risco esperado= N° casos observados/ N° Casos esperados

Po(Id) Modelo de Poisson com função de ligação Identidade

Po(Log) Modelo de Poisson com função de ligação Logaritmo

MHB Modelos Hierárquicos Bayesianos

MLG Modelos Lineares Generalizados

MGL Modelos Gaussianos Latentes

GLM Generalized Generalized Linear Models

BYM Modelo Besag-York-Mollié

INLA Integrated Nested Bayesian Approximations

GMRF Gaussian Markov Random Field

iCAR Modelo autorregressivo condicional intrínseco (Intrinsic Conditional Autoregressive)

CAR Modelo condicional autorregressivo (Conditional Autoregressive)

M0 Modelo só com efeitos espacial e temporal – dados observados

M0.HF Modelo só com efeitos espacial e temporal – dados desagregados com o algoritmo HF

M0.LS Modelo só com efeitos espacial e temporal – dados desagregados com o algoritmo LS

MG Modelo com todas as covariáveis e com efeitos espacial e temporal, estruturados e não estruturados – dados observados

MG.HF Modelo com todas as covariáveis e com efeitos espacial e temporal, estruturados e não estruturados – dados desagregados com HF

MG.LS Modelo com todas as covariáveis e com efeitos espacial e temporal, estruturados e não estruturados – dados desagregados com LS

iid.M (IID) Modelo com covariáveis e com efeito temporal não estruturado

rw2.M (RW) Modelo com covariáveis e com efeito temporal estruturado – dados observados

besag.M (CAR) Modelo com covariáveis e com efeito espacial não estruturado

bym.M (iCAR) Modelo com covariáveis e com efeitos espaciais estruturados e não estruturados

bym.rw.iid.M(iCAR) Modelo com covariáveis e com efeitos espaciais e temporais estruturados e não estruturados e com interação tipo I

iCAR.rw.iid.M2.2.HF Modelo com covariáveis e com efeitos espaciais e temporais estruturados e não estruturados e com interação tipo I. Covariáveis meteorológicas com *lag* 2 e com dados desagregados de agregação semanal com o algoritmo HF

iCAR.rw.iid.M2.2.LS Modelo com covariáveis e com efeitos espaciais e temporais estruturados e não estruturados e com interação tipo I. Covariáveis meteorológicas com *lag* 2 e com dados desagregados de agregação semanal com o algoritmo LS

iCAR.rw.iid.M2.7.LS.Modelo com covariáveis e com efeitos espaciais e temporais estruturados e não estruturados e com interação tipo I. Covariáveis meteorológicas com lag 7 e modelo ajustado com dados desagregados com o algoritmo LS a partir dos dados com agregação semanal

iCAR.rw.iid.M2.2.LS.14 Modelo com covariáveis e com efeitos espaciais e temporais estruturados e não estruturados e com interação tipo I. Covariáveis meteorológicas com lag 2 e modelo ajustado com dados desagregados com o algoritmo LS a partir dos dados acumulados a 14 dias

iCAR.rw.iid.M2.14.HF.14 Modelo com covariáveis e com efeitos espaciais e temporais estruturados e não estruturados e com interação tipo I. Covariáveis meteorológicas com lag 14 e modelo ajustado com dados desagregados com o algoritmo HF a partir dos dados acumulados a 14 dias

Diario.Unif N° diário de casos observados de COVID-19

x.HF N° diário de casos de COVID-19 desagregados com o algoritmo HF

x.LS N° diário de casos de COVID-19 desagregados com o algoritmo LS

DuMMY Variável *dummy* com codificação dos S1 e S3 com 0 e 1(capítulo 3)

tM.mean.2 Temperatura Máxima média com lag 2

tm.mean.2 Temperatura mínima média com lag 2

hM.mean.2 Humidade Máxima media com lag 2

hm.mean.2 Humidade mínima média com lag 2

pr.mean.2 Precipitação media com lag 2

tM.mean.7 Temperatura Máxima média com lag 7

tm.mean.7 Temperatura mínima média com lag 7

hM.mean.7 Humidade Máxima media com lag 7

hm.mean.7 Humidade mínima média com lag 7

pr.mean.7 Precipitação media com lag 7

tM.mean.14 Temperattura Máxima média com lag 14

tm.mean.14 Temperattura mínima média com lag14

hM.mean.14 Humidade Máxima media com lag 14

hm.mean.14 Humidade mínima média com lag 14

pr.mean.14 Precipitação media com lag 14

Popdensity Densidade populacional (estimativa 2019)

GMM.T.2018 Ganho Médio Mensal (2018)

IDI.2019 Índice de Dependência de Idosos - 2019

Mais.65 População com mais de 65 anos

Region_str Variável codificadora da região para o efeito espacial estruturado

Region_unstr Variável codificadora da região para o efeito espacial não estruturado

Date_str Variável codificadora da região para o efeito temporal estruturado

Date_unstr Variável codificadora da região para o efeito temporal não estruturado

region-date Variável codificadora da região para o efeito de interação espaço \times tempo

v_j Efeitos aleatórios espaciais não estruturados no concelho j

u_j Efeitos aleatórios espaciais estruturados no concelho j

φ_t Efeitos temporais não estruturados no dia t

γ_t Efeitos temporais não estruturados no dia t

δ_{jt} Termo de interação no dia t e no concelho j

$\tau_{((\text{Region_str (iid component)})}$ Precisão do efeito espacial não estruturado

$\tau_{(\text{Region_unstr (spatial component)})}$ Precisão do efeito espacial estruturado

$\tau_{(\text{Date_str})}$ Precisão do efeito temporal não estruturado

$\tau_{(\text{Date_unstr})}$ Precisão do efeito temporal estruturado

$\tau_{(\text{region_date})}$ Precisão do efeito de interação espaço \times tempo

\ln Função logaritmo na base e

RMSE Root Mean Square Error

RMSE.S.LS Root Mean Square Error dos modelos ajustados com dados semanais desagregados com o algoritmo LS.

RMSE.S.HF Root Mean Square Error dos modelos ajustados com dados semanais desagregados com o algoritmo H-Fuse.

RMSE.14.LS Root Mean Square Error dos modelos ajustados com dados acumulados a 14 dias, desagregados com o algoritmo LS.

RMSE.14.HF Root Mean Square Error dos modelos ajustados com dados acumulados a 14 dias, desagregados com o algoritmo HF.

MAE Mean Absolut Error

MAE.S.LS Mean Absolut Error dos modelos ajustados com dados semanais desagregados com o algoritmo LS.

MAE.S.HF Mean Absolut Error dos modelos ajustados com dados semanais desagregados com o algoritmo HF.

RMSE.14.LS Mean Absolut Error dos modelos ajustados com dados acumulados a 14 dias, desagregados com o algoritmo LS.

RMSE.14.HF Mean Absolut Error dos modelos ajustados com dados acumulados a 14 dias, desagregados com o algoritmo HF.

AIC Akaike Information Criterion

BIC Bayes Information Criterion

DIC Deviance Information Criterion

WAIC Widely Applicable Information Criterion

GLRT Teste de razão de verosimilhança generalizada

MLE Estimativas de máxima verosimilhança

H-Fuse Algoritmo de desagregação de séries temporais H-Fuse

HF Algoritmo de desagregação de séries temporais H-Fuse

LS Algoritmo de desagregação de séries temporais com o método dos Mínimos Quadrados com Pseudo-Inversa de Penrose

S1 Período de 15 de abril a 6 de maio de 2020

S2 Período de 7 de maio a 15 de maio de 2020

S3 Período de 16 de maio a 6 de junho de 2020

CAPÍTULO 1

INTRODUÇÃO

1. INTRODUÇÃO

1.1. Motivação

A pandemia COVID-19, causada pelo novo coronavírus SARS-CoV-2, é um dos maiores desafios sanitários, sociais e económicos da história recente. Desde o seu início na China, em dezembro de 2019, até princípio de dezembro de 2023, a COVID-19 já infetou mais de 300 milhões de pessoas e causou mais de 6 milhões de mortes em todo o mundo, segundo a Organização Mundial da Saúde (OMS)¹. Em Portugal, o primeiro caso confirmado foi registado em 2 de março de 2020, e desde então o país enfrentou várias ondas epidémicas, com diferentes graus de intensidade e distribuição geográfica, tendo impactos significativos na saúde, na economia e na sociedade. Até o fim de novembro de 2023, Portugal contabilizava mais de 5,5 milhões de casos e mais de 28 mil óbitos por COVID-19 (DGS)².

Num cenário desta natureza, torna-se clara a importância de monitorizar e analisar os dados da COVID-19, tanto para compreender a dinâmica da doença, como para orientar as medidas de prevenção, controlo e mitigação e avaliar a sua eficácia e adequação. De facto, qualquer questão de saúde pública atrai a atenção do público pelo impacto significativo na economia, nos transportes ou na interação social e a pandemia da COVID-19 veio destacar a relevância da vigilância neste âmbito e a importância da sua eficácia na deteção de possíveis mudanças nas taxas de infeção da doença, bem como quando e onde ocorrem, para agir (Haridy et al., 2021). Para tal, é necessário dispor de dados fiáveis, atualizados e desagregados, que permitam caracterizar a situação epidemiológica a diferentes níveis geográficos e temporais. No entanto, a análise dos dados da COVID-19 apresenta diversos desafios, como a qualidade, a disponibilidade e a granularidade dos dados, a heterogeneidade e a complexidade dos fatores envolvidos na transmissão, a variabilidade temporal e espacial da pandemia, e a influência de eventos específicos, de fatores ambientais e comportamentais que podem alterar o padrão de disseminação da doença.

Neste contexto, esta tese tem como objetivo geral desenvolver e aplicar métodos estatísticos para a monitorização de dados de um processo pandémico/epidémico, em particular, dos dados espaço-temporais da COVID-19 em concelhos de Portugal continental,

¹ <https://covid19.who.int/> - Consultado em 4 de dezembro de 2023.

² <https://covid19.min-saude.pt/numero-de-novos-casos-e-obitos-por-dia/> - consultado em 4 de dezembro 2023

tendo em conta as suas especificidades e desafios. A motivação para este trabalho surgiu da necessidade de compreender melhor a dinâmica de disseminação da COVID-19 em Portugal, bem como de apoiar a tomada de decisão em saúde pública, e contribuir com metodologias no âmbito da estatística com potencial para serem aplicadas noutras situações de risco.

A monitorização estatística é uma área de investigação que se dedica à análise de dados recolhidos ao longo do tempo e/ou do espaço, com o intuito de detetar e quantificar alterações, anomalias ou tendências nos processos subjacentes. Apesar de ser amplamente utilizada em contextos industriais, a sua aplicação em contextos de saúde ainda é incipiente e desafiante, devido à complexidade e à incerteza dos fenómenos envolvidos (Woodall et al., 2010; Woodall & Montgomery, 2014). Em saúde pública a monitorização é implementada, por exemplo, para avaliar a qualidade dos serviços de saúde; para monitorizar o estado de saúde da população; para analisar o impacto de intervenções de saúde pública, comparando o perfil de indicadores de saúde antes e depois da intervenção, ou entre grupos expostos e não expostos, para verificar se houve mudanças significativas nos resultados esperados. Para a monitorização da incidência da COVID-19 é difícil definir o perfil de referência e atualizá-lo periodicamente de acordo com as mudanças no cenário epidemiológico e as medidas de prevenção e controle, pois a doença é nova e está em constante evolução, com o surgimento de novas variantes, vacinas e tratamentos. Além disso, os dados epidemiológicos podem estar sujeitos a erros, atrasos, subnotificações ou inconsistências, que comprometem a sua qualidade e confiabilidade.

Apesar dos dados de COVID-19 disponíveis aquando da realização da investigação desta tese serem referentes a um período de tempo relativamente curto e, portanto, não permitirem estimar o padrão longitudinal regular das taxas de incidência espacial e/ou temporal ou do número de casos, correspondente a um perfil de referência que deve ser estimado a partir de um conjunto de dados em controlo estatístico (Qiu, 2019), nesta tese foi concebida uma metodologia de monitorização, baseada essencialmente em modelos de regressão de Poisson, adequados para dados de contagem, visualização e testes de hipóteses, e que está descrita no capítulo 3. Foi ainda considerada a monitorização do comportamento espacial e temporal da pandemia, através da modelação do número de casos diários ou do número cumulativo de casos diários, recorrendo a Modelos Hierárquicos Bayesianos (MHB).

Na investigação desenvolvida procuraram-se casos de estudo relevantes, tendo surgido dois, em particular, que se revelaram de especial interesse. Um deles visa detetar o efeito de

eventos públicos, com ajuntamentos de pessoas, na disseminação da COVID-19, e os eventos analisados foram a Manifestação do 1.º de Maio de 2020, em Lisboa e a Manifestação Antirracismo realizada em junho de 2020, também em Lisboa. No segundo caso de estudo procurou-se detetar a influência de covariáveis meteorológicas – temperatura, humidade e pluviosidade, na disseminação da COVID-19 nos concelhos de Portugal Continental, considerando ainda covariáveis demográficas - *densidade populacional*, *índice de dependência do idosos e população com mais de 65 anos* e socioeconómica – *ganho médio mensal*, que serviram de confundidores. Este último caso de estudo motivou a investigação e exploração de algoritmos de desagregação temporal de dados que permitissem ultrapassar as dificuldades encontradas com os dados disponibilizados pelas autoridades de saúde portuguesas e que se encontram descritos no capítulo 2, secção 2.10 e aplicados e avaliados nos capítulos 4 e 5. Os resultados estão descritos nos capítulos 3, 4 e 5. Para cada caso, foram exploradas formas de modelar e analisar dados que proporcionem abordagens mais sólidas para os decisores.

Os dados de COVID-19 usados nesta tese foram disponibilizados agregados, em diferentes métricas e granularidades temporais, como descrito no capítulo 5, secção 5.2, o que causa limitações e possíveis vieses de interpretação que é preciso ter em conta na análise dos resultados. A dificuldade em obter dados de qualidade, capazes de proporcionarem resultados fiáveis é recorrente, especialmente na Epidemiologia. É importante investir em esforços de colaboração e comunicação entre as diferentes fontes de dados, na partilha de conhecimentos, de metodologias, no recurso a melhores práticas, para que os dados sejam, à partida, o mais fiáveis possível. Enquanto este esforço não se torna efetivo, é essencial investir na pesquisa de novas metodologias que permitam contornar as dificuldades encontradas. Neste sentido, estudaram-se métodos de extração de dados descritos no capítulo 3 e algoritmos de desagregação temporal capazes de lidar com problemas nos dados e que podem ser aplicadas noutras situações para além das estudadas.

Assim, esta tese, propõe abordagens inovadoras e adaptadas à monitorização estatística da COVID-19 em Portugal, com base nas seguintes metodologias principais: desagregação temporal de dados, modelação para análise e deteção de padrões temporais e espaciais e de fatores de risco com potencial para contribuir para a disseminação da doença.

A primeira metodologia focou-se na desagregação dos dados diários de casos de COVID-19 dos concelhos de Portugal Continental, agrupados em diferentes métricas e

granularidades, conforme divulgados pela Direção-Geral da Saúde (DGS). Utilizaram-se algoritmos de desagregação temporal, baseados em séries temporais, para obter uma série temporal homogénea e de granularidade mais fina, neste caso, de casos de COVID-19 por concelho, a qual é fundamental para as análises subsequentes.

A segunda envolveu a modelação de dados de contagem, número diário de casos de COVID-19 – cumulativos ou não, com recurso a modelos de Regressão de Poisson, Bayesianos ou não. Os modelos foram usados para a deteção do impacto de eventos públicos na propagação da COVID-19 nos concelhos onde ocorrem e visou quantificar o efeito de eventos públicos, como manifestações, festivais ou jogos de futebol, na propagação de uma epidemia/pandemia nos concelhos envolvidos, comparando os períodos antes e depois desses eventos. Modelos Hierárquicos Bayesianos, incorporando efeitos fixos – covariáveis meteorológicas, demográficas e socioeconómicas, e efeitos aleatórios espaciais e temporais, permitiram estimar parâmetros epidemiológicos de interesse, como a taxa de incidência e a sua distribuição espacial. O objetivo foi desenvolver uma metodologia própria e original que permitisse a identificação de padrões e tendências espaciais e temporais. Esta metodologia foi aplicada aos dados da COVID-19 em Portugal, para além de simultaneamente avaliar o impacto de variáveis meteorológicas na transmissão do SARS-CoV-2.

As metodologias propostas nesta tese são inovadoras e relevantes, pois permitem obter informações mais detalhadas e confiáveis sobre a evolução da COVID-19 em Portugal, contribuindo para o aprimoramento das políticas públicas de saúde e para o avanço do conhecimento científico na área da Epidemiologia e da Bioestatística. Além disso, são flexíveis e adaptáveis, podendo ser aplicadas a outros contextos e situações que envolvam dados de saúde pública ou dados de outras áreas do conhecimento.

1.2. Estrutura da tese

Esta tese está estruturada em seis capítulos. Para além do presente capítulo, está organizada da forma que se segue.

No Capítulo 2, é feito o enquadramento teórico dos Modelos Lineares Generalizados (MLG), com particular ênfase nos modelos de regressão de Poisson e Binomial Negativa. É feita a caracterização dos Modelos Hierárquicos Bayesianos e de métodos para os ajustar e analisar, em particular, a abordagem INLA – *Integrated Nested Bayesian Approximations* (Rue et al., 2009). Adicionalmente, apresentam-se ferramentas de extração visual de dados, para lidar com a dificuldade em obter dados de COVID-19 de concelhos de Portugal

Continental disponibilizados através de gráficos. A metodologia desenvolvida tem potencial para ser usada em problemas similares. São descritos alguns métodos de desagregação temporal de séries temporais. É ainda feito um breve enquadramento teórico dos dados espaço-temporais.

No Capítulo 3, apresenta-se uma metodologia baseada na modelação de Regressão de Poisson e testes estatísticos para extrair conclusões estatisticamente sustentadas sobre possíveis efeitos de um evento de grande dimensão no padrão de contágio de uma epidemia/pandemia, usando como casos práticos a Manifestação do 1.º de Maio, em 2020, na Área Metropolitana de Lisboa e a Manifestação Antirracismo, que ocorreu na mesma região em junho de 2020.

No Capítulo 4 são aplicados dois algoritmos de desagregação de séries temporais: o algoritmo dos Mínimos Quadrados (LS) e o algoritmo H-Fuse (HF) (Liu et al., 2017), com dados dos concelhos do distrito do Porto, disponibilizados com frequência diária, de 25 de março de 2020 a 30 de junho de 2020. Estes dados são agregados em frequência semanal e acumulados a 14 dias e reportado com frequência semanal, e novamente desagregados em frequência diária. Os resultados são comparados visualmente com os dados observados, e a bondade do ajustamento aos dados observados é avaliada com recurso ao *Root Mean Square Error (RMSE)* e ao *Mean Absolut Error (MAE)*. Na computação respetiva foi usado o *software R*.

No Capítulo 5 os dados espaço-temporais de COVID-19 dos concelhos do distrito do Porto, de 25 de março de 2020 a 30 de junho de 2020, são modelados com recurso a Modelos Hierárquicos Bayesianos, para caracterizar a distribuição espacial e temporal do risco de contrair COVID-19 através de mapas de risco. Os mapas são comparados com os mapas de risco obtido pela razão entre o número cumulativo de casos acumulados observados e o número cumulativo de casos esperado (SIR). Os dados dos concelhos de Portugal Continental, observados de 25 de março de 2020 a 4 de julho de 2020 e os dados desagregados de 7 de julho de 2020 a 23 de março de 2021, são modelados recorrendo a MHB e são usados mapas de distribuição de risco estimado pelo modelo, que são comparados com os mapas de risco SIR. Além de se considerar os efeitos espaciais e temporais sobre a dinâmica de propagação da COVID-19, também se estuda o efeito de covariáveis meteorológicas (*temperatura, humidade e pluviosidade*), demográficas (*densidade populacional, índice de dependência de idosos e população com mais de 65*

anos) e socioeconómicas (*Ganho médio mensal*). Primeiro, são explorados diferentes modelos com os dados de 25 de março de 2020 a 30 de junho de 2020, observados e desagregados, para avaliar a precisão do método de desagregação e o modelo mais adequado a ser usado na modelação do risco de 25 de maio de 2020 a 23 de março de 2021, porque no período de 7 de julho de 2020 a 23 de março de 2021 já não se dispõe de dados diários observados, mas apenas dados desagregados.

O Capítulo 6 resume as conclusões e apresenta algumas considerações futuras.

CAPÍTULO 2

ENQUADRAMENTO TEÓRICO

2. ENQUADRAMENTO TEÓRICO

No presente capítulo apresentam-se os fundamentos teóricos que sustentam as modelações e análises de dados exploradas nos capítulos 3, 4 e 5.

É com recurso aos Modelos Lineares Generalizados (MLG) – Poisson e Binomial Negativa, que no capítulo 3 se desenvolve uma metodologia inovadora para monitorizar os dados e identificar eventuais mudanças nas dinâmicas de propagação de uma epidemia/pandemia na sequência da realização de eventos públicos que implicam o ajuntamento de multidões.

Os Modelos Hierárquicos Bayesianos apresentam-se cada vez mais como a melhor opção para a modelação de dados espaciais e espaço/temporais. Estudar/monitorizar a dinâmica espacial e temporal de uma epidemia/pandemia é um contributo fundamental para que as autoridades possam tomar medidas de controlo de forma fundamentada. Nesta perspetiva, estes modelos são usados no capítulo 5 com dados de COVID-19 dos concelhos do distrito do Porto e de Portugal Continental.

A dificuldade, reconhecida, em obter dados adequados para a modelação justifica a importância do contributo que desta tese com a metodologia inovadora apresentada para a extração visual de dados a partir de gráficos e a metodologia de desagregação temporal de séries temporais. No primeiro caso, a partir da medida dos pixels de uma barra de um gráfico é obtida uma estimativa da frequência absoluta da respetiva classe. Esta metodologia é ilustrada num gráfico de barras, mas pode ser generalizada a outros tipos de gráficos. No segundo caso, a desagregação de dados agregados para uma granularidade temporal mais fina é aplicada a dados de COVID-19, no capítulo 4, com recurso ao R. No presente capítulo é feito o respetivo enquadramento teórico.

2.1. Modelos de crescimento

Ao longo do tempo, desde que a pandemia COVID-19 foi declarada em Wuhan, província de Hubei, na China, em fevereiro de 2020, foram adotados muitos modelos baseados apenas em premissas matemáticas. No entanto, existem muitos fatores não quantificáveis, como a ocorrência de grandes eventos, as mudanças nas políticas de transporte, as estratégias de saúde pública ou as características sociodemográficas da população, que podem afetar substancialmente as previsões da evolução da curva epidemiológica e cuja influência na

dinâmica da doença tem sido investigada (Kraemer et al., 2020; Lalwani et al., 2020, 2020; Prem et al., 2020; Tirachini & Cats, 2020).

Uma estratégia comum na modelação do número cumulativo de casos de COVID-19 para prever a evolução da pandemia socorre-se de uma curva de crescimento em forma de S (sigmoide) para assim se observar graficamente o comportamento previsível da curva nos dias seguintes (Bellomo et al., 2020). Essas curvas são geralmente baseadas em funções de distribuição cumulativa bem conhecidas, como aquelas correspondentes às distribuições de Gompertz, logística, log-normal e Gumbel que são casos particulares da curva sigmoide. Como caso particular das curvas de crescimento, a curva de Richards tradicional (Richards, 1959) assume que o número cumulativo de casos da doença no tempo t é indicado pela expressão:

$$C(t) = \frac{A}{[1 + m \exp(-k(x - i))]^{\frac{1}{m}}} \quad (2.1)$$

Esta parametrização da curva de Richards apresenta uma estrutura adequada para modelação de doenças infecciosas (Hsieh, 2009; Luo et al., 2021). Wang et al. (2012) fornecem interpretações biológicas para todos os parâmetros neste modelo e introduzem uma restrição para abordar o problema de *overfitting* observado em algumas pesquisas existentes (o modelo ajusta-se muito bem aos dados observados, mas é pobre a prever novos dados). Segundo Wang et al.(2012), na fórmula (2.1), A é o número cumulativo de casos no final do período epidémico em análise, k é a taxa de crescimento *per capita* do número cumulativo de casos, m é o parâmetro de forma (expoente usado para captar o desvio da simetria da dinâmica em forma de S do modelo logístico simples; determina a proporção do número cumulativo final de casos (A) no qual o ponto de inflexão da curva ocorre (Freitas, 2005)); e i é o *lag phase* da trajetória. Este modelo tem sido usado na modelação para previsão em tempo real de surtos infecciosos (Alaimo Di Loro et al., 2021; Hsieh, 2009; Hsieh et al., 2004, 2010; S. Y. Lee et al., 2020; Wang et al., 2012; K. Wu et al., 2020).

Quando a evolução da uma epidemia revela a existência de mais que uma sigmoide sucessiva, torna-se adequado um modelo duplo de Richards, definido pela combinação de dois modelos de Richards usual.

$$C(t) = \frac{A}{[1 + m \exp(-k(x - i))]^{\frac{1}{m}}} + \frac{A'}{[1 + m' \exp(-k'(x - i'))]^{\frac{1}{m'}}} \quad (2.2)$$

O pacote `flexparmacurve` do R (Oswald et al., 2012) implementa este modelo, proporcionando um vasto leque de versões que dependem do número de parâmetros que são estimados ou fixados na média do conjunto de dados ou num valor especificado pelo utilizador. Especificando os parâmetros da segunda curva em $m' = 1$, $m' = 0$ ou $m' = -0.3$, obtêm-se as curvas Logística, de Gompertz e von Bertalanffy, respetivamente.

Os Modelos Lineares Generalizados são modelos muito versáteis que são adequados para modelar os dados de contagem da evolução de uma pandemia, nomeadamente, o modelo de regressão de Poisson e o modelo Binomial Negativo.

2.2. Modelos Lineares Generalizados (*Generalized Linear Models*)

Os Modelos Lineares Generalizados (MLG) (ou GLM – *Generalized Linear Models*) (Nelder & Wedderburn, 1972) constituem uma generalização dos Modelos Normais de Regressão, abrindo o leque de opções para a distribuição da variável resposta a qualquer distribuição numa classe designada família exponencial de distribuições. Estes modelos permitem uma maior flexibilidade para a relação entre o valor médio da resposta e a combinação linear de variáveis preditoras do que a que a relação definida pelo modelo linear normal.

O modelo geral da família de funções exponenciais é da forma

$$f(y|\theta, \phi) = \exp \left\{ \frac{(y\theta - b(\theta))}{a(\phi)} + c(y, \phi) \right\} \quad (2.3)$$

onde θ e ϕ são parâmetros escalares e $a(\cdot)$, $b(\cdot)$ e $c(\cdot, \cdot)$ funções reais conhecidas.

θ é a forma canónica do parâmetro de localização (parâmetro natural) e ϕ é um parâmetro de dispersão, regra geral conhecido. Caso seja desconhecido, a distribuição pode fazer parte da família exponencial biparamétrica ou não. Admite-se que a função $b(\cdot)$ é diferenciável, que $a(\phi) > 0$ e que o suporte da distribuição (domínio onde está definida) não depende dos parâmetros.

Prova-se, usando a função *Score*: $S(\theta) = \frac{\partial[\ln(\theta; \phi, Y)]}{\partial \theta}$, que:

$$E(Y) = b'(\theta) \quad (2.4)$$

$$var(Y) = a(\phi) b''(\theta) \quad (2.5)$$

Note-se que $var(Y)$ é o produto de uma função que depende apenas de θ , $b''(\theta)$, que se designa por *função de variância*, por uma outra função que depende apenas de ϕ , $a(\phi)$. A função de variância, normalmente, designa-se por $V(\mu)$.

A função de variância desempenha um papel importante na família exponencial, uma vez que caracteriza a distribuição, na medida em que, dada a função de variância tem-se uma classe de distribuições correspondentes, e vice-versa. Esta propriedade permite a comparação de distribuições através de testes simples para a função de variância.

Em muitas situações $a(\phi)$ toma a forma $\frac{\phi}{\omega}$, sendo ω uma constante conhecida. Neste caso, a variância de Y é o produto do parâmetro de dispersão por uma função do valor médio,

$$var(Y) = \phi b''(\theta) = \phi \frac{d\mu}{d\theta} \quad (2.6)$$

Nestas condições, o modelo (2.3) toma a forma

$$f(y|\theta, \phi, \omega) = \exp \left\{ \frac{\omega}{\phi} (y\theta - b(\theta)) + c(y, \phi, \omega) \right\} \quad (2.7)$$

A família exponencial de distribuições é vasta e inclui algumas das mais importantes e conhecidas distribuições, contínuas e discretas.

Seja Y uma variável aleatória que representa a variável resposta, quantitativa contínua ou discreta (na forma de contagem), ou qualitativa (fator) e seja x um vetor de variáveis explicativas, designadas também de covariáveis, que se acredita explicam parte da variabilidade observada em Y . As covariáveis podem ser deterministas ou aleatórias, de natureza quantitativa – contínuas ou discretas, ou qualitativa – ordinais ou nominais.

Assume-se que se tem dados da forma (y_i, x_i) , com $i = 1, 2, \dots, n$, que resultam da realização de (Y, x) em n indivíduos ou unidades experimentais, sendo as componentes y_i do vetor Y independentes.

Os MLG são caracterizados por uma estrutura constituída por três elementos: uma componente aleatória, um preditor linear e uma função de ligação (McCullagh & Nelder, 2019). Estes dois últimos elementos permitem definir a componente sistemática do modelo.

i) Componente aleatória

Dado um vetor de covariáveis $x = (x_1, x_2, \dots, x_p)$ e as variáveis $Y_i, i = 1, 2, \dots, n$, nesta componente especifica-se a distribuição condicional da variável resposta Y_i , para a i -ésima das n observações amostradas independentemente, dados os valores das variáveis explicativas no modelo. Na sua formulação original (Nelder & Wedderburn, 1972), a distribuição de Y_i é membro de uma família exponencial (2.3), (Gaussiana, Binomial, Poisson, Gama ou famílias de distribuições gaussianas inversas), mas trabalhos subsequentes estenderam os MGL a famílias exponenciais multivariadas, a algumas famílias não

exponenciais, como a distribuição Binomial Negativa de dois parâmetros e a algumas situações em que a distribuição de Y_i não é completamente especificada.

ii) Componente sistemática (ou estrutural)

Existe um regressor linear η_i tal que

$$\eta_i = z_i^T \beta \tag{2.8}$$

onde β é um vetor de parâmetros, de dimensão p .

z_i é um vetor de especificação de covariáveis, de dimensão p , função do vetor de covariáveis $x = (x_1, x_2, \dots, x_p)$. Regra geral, $z_i = (1, x_1, x_2, \dots, x_{p-1})^T$. As covariáveis podem ser quantitativas, transformações de covariáveis quantitativas, regressores polinomiais, regressores *dummy*, interações e outras. Quando há covariáveis qualitativas, estas, em geral, têm de ser devidamente codificadas à custa de uma variável binária muda (variável *dummy*) e cada uma destas variáveis tem de ser incluída no vetor z_i .

iii) Função de ligação

Existe uma função l , monótona e diferenciável, designada função de ligação, que permite relacionar as covariáveis x_i com o valor esperado μ_i , tal que

$$l(\mu_i) = \eta_i \tag{2.9}$$

A função de ligação escolhida depende do tipo de dados que estão envolvidos no problema particular em estudo, nomeadamente no que diz respeito à distribuição de probabilidade que caracteriza a variável resposta.

Há um conjunto de funções de ligação que são frequentemente consideradas, havendo especial interesse no caso em que o preditor linear coincide com o parâmetro canónico. Neste caso, $\theta_i = \eta_i$, o que implica que $\mu_i = z_i^T \beta$ e a função de ligação designa-se por *função de ligação canónica*. Uma função de ligação diz-se canónica se transforma o valor esperado da resposta no parâmetro canónico:

$$l(E(y_i)) = \mu_i \tag{2.10}$$

As funções de ligação canónica são úteis porque simplificam de forma assinalável o estudo do modelo. De certa forma, uma função de ligação canónica é a função de ligação “natural” para o respetivo tipo de distribuição da variável resposta. A utilização desta função na função de verosimilhança garante a concavidade desta última e, portanto, garante a unicidade da estimativa, quando existe.

Existe uma função de ligação canónica para cada função distribuição da resposta. Em particular, alguns exemplos:

$$\begin{aligned}
 \eta &= \mu \quad (\textit{Identidade} - \text{distribuição Normal}) \\
 \eta &= \ln\left(\frac{\mu}{1-\mu}\right) \quad (\textit{Logit} - \text{distribuição Binomial}) \\
 \eta &= \ln(\mu) \quad (\textit{Log} - \text{distribuição de Poisson}) \\
 \eta &= \frac{1}{\mu} \quad (\textit{Inversa} - \text{distribuição Gamma}) \\
 \eta &= \sqrt{\mu} \quad (\textit{Raiz quadrada} - \text{Distribuição de Poisson})
 \end{aligned} \tag{2.11}$$

Como a função de ligação é invertível, também se pode escrever

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(z_i^T \beta) \tag{2.12}$$

e, assim, o MGL pode ser pensado como um modelo linear para uma transformação da resposta esperada ou como um modelo de regressão não linear para a resposta. A ligação inversa $g^{-1}(\cdot)$ também é chamada de função média.

A seleção de uma função de ligação que torne linear a regressão de Y nos X e que seja uma ligação promissora, removerá as restrições no intervalo da resposta esperada. Numa resposta Y de contagem, que toma valores inteiros não negativos e, portanto, μ_i é um valor esperado em $[0, +\infty[$, a ligação logaritmo mapeia μ_i em toda a reta real. No entanto, a escolha da função de ligação não é inteiramente determinada pelo intervalo da variável resposta.

Por vezes é usada a função de ligação *Identidade* para dados de contagem. No entanto, é preciso cuidado pois poder-se-á obter $\mu < 0$. Este problema não ocorre se a função de ligação for a função logaritmo (\ln na base e – modelo *loglinear*) ou a função *raiz quadrada*. A função \ln é a função de ligação usual nos modelos de regressão de Poisson. No entanto, uma distribuição de Poisson tem variância igual à média. Quando este pressuposto é violado e a variância é maior que a média (há sobredispersão nos dados), então assume-se que a variância é proporcional à média e usa-se o modelo Binomial Negativo para dados de contagem.

Modelo de Regressão de Poisson

O modelo de Regressão de Poisson tem as três componentes que definem um modelo MGL (Dobson & Barnett, 2018; Hardin & Hilbe, 2018) definidas como se segue:

1. A *distribuição de probabilidade* de Y , sendo Y a variável resposta.

$$E(Y) = \mu ; \quad Y \sim Po(\mu) \quad (2.13)$$

A função massa de probabilidade é dada por:

$$p(y|\mu) = \frac{\mu^y \exp(-\mu)}{y!} \quad (2.14)$$

2. Um *preditor linear*:

$$\eta = X^T \beta \quad (2.15)$$

3. Uma *função de ligação* $g(\cdot)$, diferenciável e monótona, que relaciona o preditor linear com a média da variável resposta da seguinte forma:

$$g(\mu) = \eta \quad (2.16)$$

Na sua forma usual, a regressão de Poisson usa a função de ligação \ln (logaritmo na base e), por ser a que tem melhores propriedades matemáticas, embora possam ser usadas as funções identidade ou a função raiz quadrada.

A função de ligação logarítmica, designada por função de ligação canónica, tem a forma seguinte:

$$\ln(\mu) = X^T \beta \Leftrightarrow \mu = \exp(X^T \beta) \quad (2.17)$$

Como já referido, a função de ligação *Identidade* é menos comum, pois a média de uma variável aleatória de Poisson deve ser positiva, condição que esta função não garante. As funções de ligação \ln e *raiz quadrada*, no entanto, garantem este pressuposto. No modelo de regressão de Poisson também se supõe que a média e a variância são iguais.

A interpretação dos coeficientes difere de acordo com a função de ligação (Hardin & Hilbe, 2018). No caso da função de ligação *Identidade*, os coeficientes são interpretados como o declive da linha de regressão linear: taxa de variação – neste caso, o modelo coincide com o modelo linear usual. No caso da função de ligação logarítmica, o coeficiente associado a um preditor X é a mudança esperada no resultado de Y , na escala logarítmica, por unidade de mudança de X . Assim, mantendo todas as outras variáveis no modelo constantes, aumentando X em 1 unidade (ou passando de 1 nível para o próximo), a taxa de Y é multiplicada por $\exp^\beta (= e^\beta)$.

Se Y tem distribuição de Poisson e t é índice de tempo, espaço ou uma variável de outra natureza (p.e., população) a taxa de ocorrência do evento e o respetivo valor esperado são definidos, respetivamente, por:

$$(1) \quad \frac{Y}{t}$$

$$(2) \quad E\left(\frac{Y}{t}\right) = \frac{1}{t}E(Y) = \frac{\mu}{t} \quad (2.18)$$

Assim, a relação de μ com o preditor linear pode ser definida pela função de ligação canónica:

$$\ln\left(\frac{\mu}{t}\right) = X^T \beta \Leftrightarrow \ln(\mu) - \ln(t) = \exp(X^T \beta) \quad (2.19)$$

que é equivalente a:

$$\ln(\mu) = \ln(t) + \exp(X^T \beta) \quad (2.20)$$

$\ln(t)$ é referido como um *offset* (deslocamento) e é um termo de ajustamento, necessário quando as unidades de observação são diferentes em alguma dimensão (p.e., diferentes populações, diferentes tamanhos geográficos) e o resultado é proporcional a essa dimensão.

MLG Binomial Negativo (NB2)

O modelo Binomial Negativo pode ser utilizado como alternativa ao modelo de Poisson, quando a variância amostral é muito superior à média amostral. Quando isso acontece, diz-se que há superdispersão nos dados. Neste caso, pode-se supor que:

$$y|\mu, \phi \sim \text{Binomial Negativo}(\mu, \phi) \quad (2.21)$$

Uma abordagem possível, seguindo Agresti, 2015 (2015) ou McCullagh & Nelder (2019), consiste em adotar um modelo Poisson $Po(\mu)$ para a contagem Y , mas supor que a contagem esperada μ^* é em si uma variável aleatória não observável que tem distribuição *Gama* com média μ e parâmetro de escala constante ϕ , implicando que o parâmetro de forma *Gama* seja $\psi = \frac{\mu}{\phi}$. Então, a contagem observada Y , por período de tempo ou de espaço, segue uma distribuição binomial negativa cuja função massa de probabilidade é definida por:

$$p(y|\mu, \phi) = \frac{\Gamma(y + \phi)}{\Gamma(\phi) \cdot y!} \cdot \frac{\mu^y \cdot \phi^\phi}{(\mu + \phi)^{y+\phi}} \quad (2.22)$$

onde μ é a média condicional, ϕ^{-1} é um parâmetro de dispersão e $\Gamma(\cdot)$ é a função *Gama*.

Com ϕ conhecido, este modelo é um membro de uma família exponencial, apropriado para variáveis discretas, com parâmetro natural $\ln\left[\frac{\mu}{\mu+\phi}\right]$.

De facto, a expressão (2.22) pode ser reescrita da seguinte forma:

$$\begin{aligned}
 p(y|\mu, \phi) &= \frac{\Gamma(y + \phi)}{\Gamma(\phi) \cdot y!} \cdot \left(\frac{\mu}{\mu + \phi}\right)^y \cdot \left(\frac{\phi}{\mu + \phi}\right)^\phi & (2.23) \\
 &= \exp \left\{ y \ln \left(\frac{\mu}{\mu + \phi}\right) + \phi \ln \left(\frac{\phi}{\mu + \phi}\right) \right. \\
 &\quad \left. + \ln \left(\frac{\Gamma(y + \phi)}{\Gamma(\phi) \cdot y!}\right) \right\} \\
 &= \exp \left\{ \phi \left[\frac{y}{\phi} \ln \left(\frac{\mu}{\mu + \phi}\right) + \ln \left(\frac{\phi}{\mu + \phi}\right) \right] \right. \\
 &\quad \left. + \ln \left(\frac{\Gamma(y + \phi)}{\Gamma(\phi) \cdot y!}\right) \right\}
 \end{aligned}$$

Então $\frac{y}{\phi}$ pertence à família exponencial (2.7) sendo $\theta = \ln \left[\frac{\mu}{\mu + \phi} \right]$, $b(\theta) = -\ln(1 - e^\theta)$, $\omega = 1$ e $c(y, \phi, \omega) = \ln \left(\frac{\Gamma(y + \phi)}{\Gamma(\phi) \cdot y!} \right)$.

O valor esperado e a variância são respetivamente:

$$\begin{aligned}
 E(Y) &= \mu \\
 Var(Y) &= \mu + \phi^{-1} \mu^2 & (2.24)
 \end{aligned}$$

Note-se que se $\frac{1}{\phi} \rightarrow 0$, o modelo converge para o modelo de Poisson (Cameron & Trivedi, 2014). É usual substituir em (2.22) ϕ por $\frac{1}{\alpha}$, sendo α o parâmetro de heterogeneidade.

O modelo Binomial Negativo, com esta variância quadrática, designado por NB2 (Cameron & Trivedi, 1986), é o modelo binomial negativo MLG.

Este modelo, usualmente, é especificado como o modelo de Poisson, ou seja, com a mesma função de ligação ($\ln(\mu) = X^T \beta$) e o mesmo preditor linear, $\eta = X^T \beta$.

Também neste caso é considerado o termo *offset* descrito no modelo de Poisson.

Para testar a superdispersão num modelo de Poisson calcula-se a soma dos quadrados dos resíduos padronizados, $\sum_{i=1}^n z_i^2$, e compara-se o valor obtido com a distribuição χ_{n-k}^2 , que é o que se esperaria sob o modelo (usando $n - k$ graus de liberdade para ter em conta a estimativa de k coeficientes de regressão). A distribuição χ_{n-k}^2 , tem valor médio $n - k$, e então a razão,

$$\text{Superdispersão estimada} = \frac{1}{n-k} \sum_{i=1}^n z_i^2. & (2.25)$$

é a estatística do teste χ_{n-k}^2 para a superdispersão nos dados em comparação com o modelo ajustado (Gelman & Hill, 2006).

Estimação dos parâmetros

a) Modelo de regressão de Poisson

Os coeficientes dos modelos de regressão de Poisson são estimados por Máxima-Verossimilhança, a partir de uma amostra aleatória (X_t, Y_t) , $t = 1, 2, \dots, n$.

Dadas n observações independentes Y_t , $t = 1, 2, \dots, n$, a função de Log-verossimilhança é dada por:

$$\mathcal{L}(\beta) = \ln L(\beta) = \sum_{t=1}^n \{Y_t X_t^T \beta - \exp(X_t^T \beta) - \ln(Y_t!)\} \quad (2.26)$$

Os estimadores de Máxima Verossimilhança dos coeficientes do modelo de Poisson, $\hat{\beta}_p$, são a solução das condições de primeira ordem:

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \Leftrightarrow \sum_{t=1}^n (Y_t - \exp(X_t^T \beta)) X_t = 0 \quad (2.27)$$

e que satisfazem as condições de segunda ordem, garantindo que a matriz Hessiana da função de log-verossimilhança é definida negativa. As condições de segunda ordem são dadas por:

$$\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta^T} = - \sum_{t=1}^n (\exp(X_t^T \beta)) X_t X_t^T < 0 \quad (2.28)$$

Assim, assegura-se que o estimador $\hat{\beta}_p$ corresponde a um máximo da função de log-verossimilhança.

O método padrão para cálculo de $\hat{\beta}_p$ é o método iterativo de Newton-Raphson. A convergência é garantida, porque a função de Log-verossimilhança é globalmente côncava. Na prática, muitas vezes são necessárias menos de dez iterações (Cameron & Trivedi, 1990).

A função *Deviance*, que compara o modelo saturado com o modelo ajustado, é definida por:

$$D(Y, \mu) = 2 \sum_{t=1}^n \left\{ Y_t \ln \left(\frac{Y_t}{\hat{\mu}_t} \right) - (Y_t - \hat{\mu}_t) \right\} \quad (2.29)$$

onde $\hat{\mu}$ é o vetor de valores previstos pelo modelo e Y é o vetor de valores observados.

b) Modelo de regressão Binomial Negativa

Os coeficientes do modelo regressão Binomial Negativa são estimados por Máxima Verossimilhança (Lord et al., 2012).

Dadas n observações independentes Y_t , com $t = 1, 2, \dots, n$, após alguns rearranjos, a função de Log-verossimilhança é dada por

$$\begin{aligned}\mathcal{L}(\mu; Y, \phi) &= \ln L(\mu; Y, \phi) \\ &= \sum_{t=1}^n \left\{ Y_t \ln \left(\frac{\mu}{\mu + \phi} \right) - \phi \ln \left(\frac{\mu + \phi}{\phi} \right) \right. \\ &\quad \left. + \ln(\Gamma(Y_t + \phi)) - \ln(Y_t!) - \ln(\Gamma(\phi)) \right\}\end{aligned}\tag{2.30}$$

A Log-verosimilhança do modelo Binomial Negativo, parametrizada em termos de β , pode ser expressa como:

$$\begin{aligned}\mathcal{L}(\beta; Y, \phi) &= \ln L(\beta; Y, \phi) \\ &= \sum_{t=1}^n \left\{ Y_t \ln \left(\frac{\exp(X_t^T \beta)}{\exp(X_t^T \beta) + \phi} \right) - \phi \ln \left(\frac{\exp(X_t^T \beta) + \phi}{\phi} \right) \right. \\ &\quad \left. + \ln(\Gamma(Y_t + \phi)) - \ln(Y_t!) - \ln(\Gamma(\phi)) \right\}\end{aligned}\tag{2.31}$$

Os estimadores de Máxima Verosimilhança dos parâmetros do modelo Binomial Negativo, $\hat{\phi}_{BN}$ e $\hat{\beta}_{BN}$, sendo $\mu = \exp(X^T \beta)$, são as soluções das condições de primeira ordem (Cameron & Trivedi, 2014):

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \Leftrightarrow \sum_{t=1}^n \frac{(Y_t - \mu_t) X_t}{1 + \phi^{-1} \mu_t} = 0\tag{2.32}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \phi} = 0 \Leftrightarrow \sum_{t=1}^n \left\{ \phi^2 \left(\ln \left(\frac{\phi + \mu_t}{\phi} \right) + \psi(Y_t + \phi) - \psi(\phi) \right) \right. \\ \left. + \frac{Y_t - \mu_t}{\phi^{-1} + \phi^{-2} \mu_t} \right\} = 0\end{aligned}\tag{2.33}$$

onde ψ é a função digama, derivada do logaritmo da função *Gama*, $\ln(\Gamma(\cdot))$.

Além disso, para garantir que estes estimadores correspondem a um máximo da função de Log-verosimilhança, é necessário verificar as condições de segunda ordem, assegurando que a matriz Hessiana é definida negativa.

As condições de segunda ordem são dadas pelas segundas derivadas da função de Log-verosimilhança em relação a β_{BN} e ϕ_{BN} :

Para β

$$\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta^T} = - \sum_{t=1}^n \frac{\mu_t (1 + \phi^{-1} Y_t)}{(1 + \phi^{-1} \mu_t)^2} X_t X_t^T < 0\tag{2.34}$$

Este termo representa a matriz de informação observada, e o facto desta matriz ser definida negativa assegura que a estimativa de β corresponde a um máximo da função de Log-verosimilhança.

Para ϕ

$$\frac{\partial^2 \mathcal{L}}{\partial \phi^2} = - \sum_{t=1}^n \left[\psi'(Y_t + \phi) - \psi'(\phi) - \frac{(Y_t - \mu_t)^2}{(\phi + \mu_t)^2} \right] < 0 \quad (2.35)$$

Este termo é negativo o que garante que a estimativa de ϕ corresponde a um máximo.

Condição mista entre β_{BN} e ϕ_{BN} :

Além disso, a derivada cruzada deve ser considerada:

$$\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \phi} = - \sum_{t=1}^n \left[\frac{\mu_t(1 + \phi^{-1}Y_t)}{(1 + \phi^{-1}\mu_t)^2} X_t \right] < 0 \quad (2.36)$$

A função *Deviance* é definida por:

$$D(Y, \mu, \phi) = 2 \sum_{t=1}^n \left\{ Y_t \ln \left(\frac{y_t}{\hat{\mu}_t} \right) - (Y_t + \phi) \ln \left(\frac{\phi + y_t}{\phi + \hat{\mu}_t} \right) \right\} \quad (2.37)$$

2.3. Processos e dados espaço-temporais

“Everything is related to everything else, but near things are more related than distant things.”³

(Tobler, 1970)

Compreender a distribuição temporal, espacial ou espaço-temporal de dados provenientes de fenómenos ocorridos no tempo, no espaço ou no espaço-tempo é um grande desafio em diversas áreas do conhecimento como saúde, ambiente, geologia, agronomia, indústria, entre tantas outras. A Estatística Espacial tem um papel fundamental neste âmbito.

Como noutros ramos da Estatística, a Estatística Espacial tem como objetivo principal aprender com os dados. A forma como neste contexto se abordam os dados e as questões depende do cenário temporal, espacial ou espaço-temporal com que se ocupa.

Os avanços em sistemas relacionados com a informação espacial, como o Sistema de Informação Geográfica (SIG) e o Sistema de Posicionamento Global (GPS), bem como os

³ “Todas as coisas estão relacionadas com todas as outras, mas coisas próximas estão mais relacionadas do que coisas distantes.” (trad. Aragão e Dos Santos (2019))

desenvolvimentos em metrologia e o avanço nos procedimentos computacionais vieram permitir e facilitar a representação do espaço com dados associados a processos ou fenómenos que ocorrem nestes sistemas, resultando daqui o interesse crescente em ajustar, analisar e monitorizar modelos para dados espacialmente referenciados.

A análise de dados espaciais ou espaço-temporais é o conjunto de técnicas estatísticas e modelos que usam explicitamente a referência espacial ou espaço-temporal de cada dado ou objeto que é especificado no sistema em estudo (Haining, 2003). A aplicação de métodos de análise espacial ou espaço-temporal exige a assunção de pressupostos e a utilização de dados que descrevam as relações ou interações espaciais entre os dados. Os resultados da análise dependem da distribuição espacial de valores ou da configuração da estrutura espacial do sistema sob investigação.

Os dados espaciais surgem numa grande diversidade de áreas e de aplicações e por isso há uma grande diversidade de tipos, estruturas e cenários de dados espaciais. N. Cressie & Moores (2021) enquadram os dados espaciais em função da natureza do domínio espacial de interesse em que um dado fenómeno é estudado, domínio este que, juntamente com a quantificação da incerteza, define o processo espacial aleatório subjacente.

2.3.1. Processos espaciais

Processos aleatórios univariados

Seja $D \subset \mathbb{R}^d$ um domínio de interesse e $s \in D$ um dado genérico de localização no espaço Euclideano de dimensão d , \mathbb{R}^d . Tipicamente $d = 1, 2$, ou 3 .

O processo espacial aleatório, de dimensão d , é definido por

$$Y \equiv \{Y(s) : s \in D \subset \mathbb{R}^d\} \tag{2.38}$$

onde que $Y(s)$ é o valor do atributo aleatório Y na localização s .

O dado espacial $Z(s) = [z(s_1), z(s_2), \dots, z(s_n)]'$ é uma realização finita da variável aleatória $Z(s)$ de Y , que representa o atributo a ser observado. s_i , com $i = 1, \dots, n$, indica a unidade espacial i , observada no domínio de interesse D .

Dependendo das diferentes propriedades do conjunto D , o processo espacial é um processo geoestatístico, um processo reticulado, ou de área (*lattice*), ou um processo pontual (Banerjee et al., 2014; Blangiardo & Cameletti, 2015; N. A. C. Cressie, 1993; N. Cressie & Moores, 2021; Haining, 2003).

A classificação do processo espacial depende das características dos elementos de D , ou seja, da forma como são indexadas as observações do processo Y .

Se s varia continuamente em D , então o processo é geoestatístico ou de dados referenciados em pontos e é contínuo. D é um conjunto conhecido de área (volume) positiva(o), e é amostrado num subconjunto de n localizações fixas de \mathbb{R}^d .

Se D é um subconjunto discreto fixo, dividido num número finito de áreas, de formato regular ou irregular, com limites bem definidos e em que os dados são agregados por área, então o processo é reticulado, em treliça (*lattice*) ou de dados regionais ou de área (*Areal Data*).

Quando D é um conjunto de localizações aleatórias (D é um processo aleatório) e $Y(s_i)$ é um conjunto aleatório para cada localização s_i , então tem-se um processo pontual ou de padrões de pontos espaciais.

O objetivo da análise de dados geoestatísticos é modelar, identificar e separar variações em pequena e grande escala, prever em locais não observados e reconstruir a superfície contínua com base num conjunto de dados observados. Com a análise de dados de área pretende-se modelar o padrão espacial de dados e estabelecer relações com outras variáveis. Com a análise de padrões de pontos pretende-se estudar a distribuição espacial dos eventos estudados (padrão aleatório, aglomerado ou regularmente distribuído (N. A. C. Cressie, 1993). É a localização espacial do fenómeno de interesse que se pretende modelar.

Designando por $[A,B]$ o modelo de probabilidade conjunta das variáveis aleatórias A e B , $[A|B]$ o modelo de probabilidade condicional de A dado B e $[A]$ o modelo de probabilidade marginal de A , então para o processo Y e os dados Z (\cdot), segundo Cressie e Moores (2021), definem-se do seguinte modo:

$$[Y, D] = [Y | D] [D] \text{ é o modelo do processo estocástico} \quad (2.39)$$

$$[Z | Y, D] \text{ é o modelo dos dados} \quad (2.40)$$

Sendo ainda $[\theta]$ o modelo de probabilidade que captura a incerteza dos parâmetros do modelo (ou o modelo de probabilidade *a priori*), a incerteza total de um problema da estatística espacial pode ser expressa através da medida de probabilidade conjunta seguinte:

Modelo Hierárquico Bayesiano

$$[Z, Y, D, \theta] = [Z | Y, D, \theta] [Y | D, \theta] [D | \theta] [\theta] \quad (2.41)$$

Se substituirmos θ na igualdade (2.39) por uma estimativa eficiente $\hat{\theta}$, a partir de Z , obtém-se o Modelo Hierárquico Empírico.

Nos processos geoestatísticos a incerteza está no processo Y , que é indexado continuamente em D ; nos processos reticulados a incerteza está em Y que é indexado num número contável de localizações espaciais; e nos processos pontuais a incerteza está nas localizações espaciais.

Para os processos geoestatístico e reticulado, uma vez que D é fixo e conhecido, pode-se deixar cair D e (2.39) toma a forma

$$[Z, Y, \theta] = [Z | Y, \theta] [Y | \theta] [\theta] \quad (2.42)$$

isto é,

$$\begin{aligned} & [\text{Dados}, \text{Processo}, \text{Parâmetros}] \\ &= [\text{Dados} | \text{Processo}, \text{Parâmetros}] \\ &\times [\text{Processo} | \text{Parâmetros}] \times [\text{Parâmetros}] \end{aligned} \quad (2.43)$$

Processos aleatórios multivariados

Para que os modelos se tornem representações mais realistas de um mundo complexo, há a necessidade de expressar interações entre múltiplos processos. Os processos aleatórios múltiplos respondem a esta complexidade:

$$Y = \{Y(s) : s \in D^k\} \quad (2.44)$$

onde o vetor k -dimensional $Y(s) \equiv (Y_1(s), \dots, Y_k(s))^T$ representa os múltiplos processos na localização genérica $s \in D$.

A abordagem de modelação estatística hierárquica unidimensional pode ser adotada para construir processos espaciais multivariados, com qualquer um dos tipos de processos univariados nas suas componentes, o que permite definir espaços aleatórios multivariados híbridos, mas usando um processo geoestatístico nuclear para controlar a dependência espacial desses processos univariados. N. Cressie & Moores (2021) fazem uma revisão de vários modelos de processos espaciais: geoestatísticos, de área (*lattice*) e de pontos espaciais, comumente usados, incluindo modelos multivariados. Discutem a discretização de processos geoestatísticos, por aproximação a processos *lattice* regulares e irregulares e generalizam os modelos de processos espaciais para modelos de processos espaço-temporais. Sain e Cressie (2007) desenvolvem modelos de campos aleatórios de Markov para dados *lattice* multivariados, com especial atenção a modelos que incorporam formas gerais de correlações

espaciais e correlações cruzadas entre variáveis, em diferentes locais. Aplicam a metodologia com um modelo hierárquico que incorpora o modelo que propõem e usam inferência Bayesiana e o Método de Monte Carlo via Cadeias de Markov na análise de um problema de avaliação da equidade ambiental, nomeadamente, para avaliar a relação entre a distribuição de grupos demográficos e a localização de instalações tóxicas. Sain et al. (2011) introduzem um modelo estatístico hierárquico espacial, projetado principalmente para a análise da saída de modelos climáticos regionais, com base num conjunto simples (condições iniciais perturbadas) o que permite apresentar projeções probabilísticas de mudanças climáticas regionais baseadas no conjunto usado. Esta abordagem hierárquica é multivariada, no centro da qual está uma representação de um campo aleatório multivariado de Markov, que permite a modelação flexível de dependências espaciais multivariadas, incluindo as dependências cruzadas entre variáveis. Aplicam o modelo a um conjunto resultante de uma experiência de modelo climático regional e usam-no para prever as mudanças sazonais na temperatura e pluviosidade a longo prazo. Martinez-Beneito (2013) introduz uma abordagem ao mapeamento multivariado de doenças que engloba diversas propostas de modelos baseados em campos aleatórios de Markov Gaussianos, disponíveis na literatura. A estrutura proposta que começa com uma identidade simples, reformulando produtos de Kronecker de matrizes de covariância como produtos de matrizes simples, permite combinar estruturas de dependência espacial e diferentes relações entre doenças com esforço computacional limitado. Kim et al. (2001) propõem um método de Bayes bivariado para estimar as taxas de mortalidade de uma doença para uma determinada população, usando informações adicionais de uma segunda doença e este procedimento é comparado a um procedimento Bayesiano Hierárquico univariado que usa informações de apenas uma doença.

2.3.2. Processos espaço-temporais

Há uma extensão natural de dados espaciais para dados de espaço-tempo, em que o índice temporal pode ser um ponto (um instante no tempo) ou um intervalo (um período temporal).

Na configuração espaço-temporal o processo aleatório é definido por

$$\{Y(s, t), s \in D \text{ e } t \in T\}, \text{ com } D \subset \mathbb{R}^d \text{ e } T \subset \mathbb{R} \quad (2.45)$$

Na maioria das aplicações $D \subset \mathbb{Z}^2$ e $T \subset \mathbb{Z}^+$ (Shermann, 2011). Em particular, cada realização de Y corresponde a uma variável aleatória $Z(s, t) = [z(s_i, t_j)]'$, $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, T$. $z(s_i, t_j)$ representa i -ésima observação feita no local s_i no tempo t_j da variável de interesse $Z(s, t)$ de Y . A configuração espacial é a configuração espaço-temporal para $t = 1$.

Se $T = \{1, 2, \dots\}$ então o processo espaço-temporal torna-se num processo espacial de séries temporais, $\{Y(s; 1), Y(s; 2), \dots: s \in D\}$.

Se se definir $Y_j(s) \equiv Y(s; j)$, para $j = 1, \dots, k$, então o processo espaço-temporal resultante pode ser representado como um processo espacial multivariado dado por (2.44), pelo que a abordagem para modelar a dependência dos processos pode ser a mesma que a que é usada para os processos espaciais multivariados.

2.3.3. Especificação de processos *lattice* e dados estatísticos associados (dados de área e dados *lattice*)

Num processo *lattice*, D é uma coleção contável de locais espaciais de \mathbb{R}^d , regular (*lattice*) ou não (*área*), onde os dados são observados. A coleção D de tais locais é chamada de rede que é depois complementada com informações de vizinhança. Do ponto de vista matemático, D é uma coleção de nodos numa grelha, pixels ou pequenas áreas e das respetivas localizações espaciais:

$$D = \{s_1, s_2, \dots, s_n\} \tag{2.46}$$

Para especificar o processo é necessário definir uma matriz de vizinhanças mediante determinado critério, por exemplo, adjacência ou proporcional ao inverso da distância euclidiana, e especificar uma função de probabilidade conjunta válida.

As probabilidades de todos os $s_i \in D$, condicionadas pelos valores do atributo Y nos s_i vizinhos (2.46) devem resultar numa função de probabilidade conjunta válida (N. Cressie & Moores, 2021).

$$\{[Y(s_i)|Y(D \setminus \{s_i\})], s_i \in D\} = \{[Y(s_i)|Y(V(s_i))], s_i \in D\} \tag{2.47}$$

onde $V(s_i) \subset D \setminus \{s_i\}$ é o conjunto de elementos diferentes de s_i e $Y = \{Y(s_i), i = 1, 2, \dots, n\}$

A dependência estatística espacial neste tipo de dados é modelada pela matriz de vizinhanças.

Os processos *lattice* mais comuns são séries temporais e Campos Aleatórios de Markov.

2.4. Métodos Bayesianos – modelos para dados espaciais e espaço-temporais

Nesta tese serão tratados apenas dados de área e, portanto, processos *lattice*.

Uma única característica dos dados espaciais ou espaço-temporais é que os diferencia de outros dados estudados na literatura clássica de mineração de dados é a presença de dependências entre as medições, induzidas pelas dimensões espaço e tempo (Atluri et al., 2018). Muitos dos métodos estatísticos utilizados na análise destes dados são baseados na suposição de que as instâncias de dados (áreas) são independentes e identicamente distribuídas (i.i.d.), pressupostos que, regra geral, não se verificam. De facto, neste tipo de dados as instâncias são estruturalmente relacionadas entre si no contexto de espaço e tempo e apresentam propriedades variadas em diferentes regiões espaciais e períodos de tempo (Atluri et al., 2018). Ignorar estas dependências durante a análise de dados pode levar a uma baixa precisão e dificuldade na interpretação dos resultados (Dormann, 2007), pelo que a modelação espaço-temporal as deve incorporar como mostram Banerjee & Dey (2005) e Zhou & Hanson (2018).

Dados epidemiológicos são frequentemente caracterizados por uma estrutura espacial e/ou temporal que precisa ser tida em consideração no processo inferencial (Dormann, 2007). Nos modelos para dados de área, a dependência espacial pode ser considerada através de efeitos aleatórios baseados na estrutura espacial de vizinhanças. No entanto, a inclusão de efeitos aleatórios num modelo com dados não normais pode induzir alguma complexidade no modelo, e a maneira mais eficaz de lidar com isso é por meio da modelação hierárquica Bayesiana. Esta abordagem é, neste contexto e com estas considerações, particularmente eficaz (Dunson, 2001) e tem sido usada em várias aplicações epidemiológicas.

Seguem-se alguns exemplos ilustrativos.

A. B. Lawson e Zhou (2005) examinam a necessidade de modelos epidémicos descritivos no espaço, no tempo e de modelos espaço-temporais para a dinâmica epidémica. Discutem a utilização de modelos Bayesianos para o estudo da disseminação de doenças e aplicam-nos a um surto de febre aftosa no Reino Unido.

Jewell et al. (2009) propõem uma metodologia totalmente Bayesiana para realizar inferência e previsão *online* para epidemias em populações estruturadas. A metodologia é baseada em Métodos de Monte Carlo via Cadeias de Markov (*Markov Chain Monte Carlo* - MCMC e MCMC adaptativo) para estimação de parâmetros, previsão em epidemias e

avaliação *online* do risco de infeções atualmente não observadas e é ilustrada usando dois estudos complementares: uma análise da epidemia de febre aftosa do Reino Unido, em 2001 e a modelação do risco potencial de uma possível futura epidemia de gripe aviária para a indústria avícola do Reino Unido.

Chen et al. (2022) usam um Modelo Hierárquico Bayesiano para estudar a subnotificação de COVID-19 nos Estados Unidos, ao nível dos estados, a partir do fim de abril de 2020. O modelo examina os efeitos de diferentes covariáveis na subnotificação e nas taxas de incidência e considera a dependência espacial. Além da subnotificação (falsos negativos), exploram o impacto da supernotificação (falsos positivos).

Dayaratna et al. (2022) utilizam um modelo linear generalizado de Poisson acoplado a uma estrutura espacial autorregressiva para estudar o impacto de fatores económicos, demográficos e relacionados à mobilidade na transmissão do COVID-19 em 2020. Ilustram a importância de incorporar a análise espacial para compreender os fatores que influenciam a proliferação da doença.

Nazia et al. (2022) apresentam uma revisão das abordagens metodológicas usadas para identificar as variações espaciais e espaço-temporais da COVID-19 e os fatores socioeconómicos, demográficos e climáticos de tais variações. Concluem que os modelos Bayesianos dos estudos encontrados na literatura incorporaram vários efeitos espaciais, temporais e espaço-temporais nos esquemas de modelação e que a revisão destacou a necessidade de uma modelação espaço-temporal Bayesiana avançada ao nível local, através da estrutura multinível, para estratégias de prevenção e controle da COVID-19.

Richardson & Best (2003) descrevem Modelos Hierárquicos Bayesianos e ilustram a sua utilização em estudos epidemiológicos dos efeitos do meio ambiente na saúde.

Coly et al. (2021) propuseram e testaram 60 Modelos Hierárquicos Bayesianos em 400 conjuntos de dados simulados e dados reais de tuberculose bovina. Com a análise implementada mostraram a relevância dos processos CAR (*Conditional AutoRegressive*) para lidar com a estrutura de risco.

Refiram-se ainda alguns exemplos de aplicações desta abordagem noutras áreas.

Na ecologia, Clark (2004) aplica a estrutura hierárquica Bayesiana a dois problemas da ecologia: estimar a fecundidade de sementes de carvalho e o processo migratório de aves. N. Cressie et al. (2009) discutem a modelação estatística hierárquica Bayesiana como uma maneira poderosa de abordar a análise ecológica na presença de incertezas inevitáveis, mas

quantificáveis, mesmo quando questões práticas exigem compromissos pragmáticos. de Rivera et al. (2019) especificam um modelo espacial Bayesiano para estudar a distribuição de espécies, em particular, para estudar o problema de distribuição espacial de espécies de anfíbios num Parque Nacional em Espanha.

Nas ciências ambientais, Wikle (2003) fornece uma visão geral das abordagens de modelações hierárquicas aplicadas aos processos ambientais, considerando que a modelação hierárquica fornece uma estrutura capaz de simplificar sistemas ambientais complicados, vinculando a incerteza de forma coerente. Discutem alguns temas relacionados, em particular, a importância de cientistas da área das ciências do ambiente e estatísticos trabalharem de forma colaborativa na construção detalhada do modelo, uma vez que a essência de um bom modelo hierárquico é uma forte informação prévia que os cientistas devem ser capazes de transmitir e o estatístico deve ser capaz de traduzir em modelos estatísticos razoáveis. Clark e Gelfand (2006) reúnem um conjunto de contribuições com aplicações específicas, com foco nas ciências ambientais, como previsão da biodiversidade, circulação atmosférica, qualidade do ar, para fornecer uma visão geral não técnica dos métodos hierárquicos de Bayes e Monte Carlo via Cadeia de Markov (MCMC). Sharkey & Winter (2019) descrevem um modelo espacial de valores extremos com a estrutura de modelação hierárquica Bayesiana, considerando a dependência espacial e temporal nos dados ao realizar inferência nos parâmetros do modelo. Aplicam o modelo a dados de precipitação na Grã-Bretanha. Britten et al. (2021) ilustram a utilidade da abordagem hierárquica Bayesiana, através da sua aplicação a um grande conjunto de dados ambientais. Ajustam um Modelo Hierárquico Bayesiano de escala global para descrever o perfil vertical do fluxo de carbono orgânico com profundidade.

Os modelos espaço-temporais para dados de contagem são referentes a problemas em que a variável assume valores inteiros não negativos, que surgem da contagem de ocorrências de um evento numa unidade de área geográfica e numa determinada unidade de tempo (Anderson & Ryan, 2017). As observações dizem respeito a um conjunto de unidades de áreas contíguas e não sobrepostas, por períodos consecutivos. Na modelação podem ainda ser usadas séries de covariáveis - discretas, contínuas ou fatores, medidas para cada unidade de área e/ou tempo. Essas covariáveis podem ser comuns no tempo ou na área (Ibañez et al., 2021).

Estes modelos são aplicados numa grande variedade de problemas que surgem em diferentes contextos, como a produção agrícola, a captura de pesca, erupções vulcânicas, contagem de crimes, etc., e também na epidemiologia, como ilustram alguns dos exemplos referidos acima. Os modelos espaço-temporais para dados de área são usados para estimar o efeito de um fator de risco ou uma resposta, para identificar aglomerados (*clusters*) de áreas adjacentes com padrões de resposta semelhantes ou prever observações futuras, para cada unidade de área, fornecendo um conjunto de estimativas para toda a região coberta pelas unidades de área. Também pode interessar identificar a extensão espacial dos efeitos das covariáveis.

2.4.1. Inferência paramétrica Bayesiana: distribuição *a priori* e distribuição *a posteriori*

A utilização de métodos Bayesianos na análise estatística tem sofrido um grande incremento nos últimos tempos, gozando de ampla aplicação científica e prática, nomeadamente com aplicação crescente na ecologia, na biologia evolutiva, na climatologia e na epidemiologia e em Saúde Pública em geral (A. B. Lawson, 2018). Isto deve-se muito ao desenvolvimento de algoritmos computacionais rápidos, que facilitam a implementação de métodos Monte de Carlo via Cadeias de Markov para amostragem da distribuição *a posteriori*, e de *software* como o WinBUGS ou o R-INLA que fornecem uma plataforma de amostragem e/ou análise de modelos complexos.

A modelação neste contexto pode ser vista como uma extensão da abordagem frequencista, assumindo que os próprios parâmetros do modelo são variáveis aleatórias, às quais se atribui uma distribuição conhecida, a ‘*distribuição a priori*’. Essa distribuição *a priori* é então combinada com a verosimilhança tradicional para se obter, para os parâmetros de interesse, a distribuição *a posteriori*, na qual se baseia a inferência estatística. Este processo induz estrutura hierárquica na modelação Bayesiana.

A inferência Bayesiana é uma abordagem estatística na qual todas as formas de incerteza são expressas em termos de probabilidade. A incerteza está nos parâmetros do modelo e não nos dados, como acontece na inferência frequencista. Esta abordagem capta melhor a incerteza, seja nos parâmetros, seja na especificação do modelo. No entanto, tem uma carga de subjetividade na escolha da distribuição *a priori* dos parâmetros que aconselha algum cuidado.

A inferência Bayesiana é o processo de produção de inferência estatística sob o ponto de vista Bayesiano, isto é, um paradigma estatístico/probabilístico em que um conhecimento prévio, modelado por uma distribuição de probabilidade, é atualizado sempre que uma nova observação, cuja incerteza é modelada por outra distribuição de probabilidade, é registada.

A ideia geral que rege o paradigma Bayesiano está sustentado no Teorema de Bayes que expressa a relação entre o conhecimento atualizado (*a posteriori*), o conhecimento prévio (*a priori*) e o conhecimento que resulta da observação (a verosimilhança).

$p(Y = y|\theta)$, doravante representado por $p(y|\theta)$, designa a função de verosimilhança e especifica o modelo de distribuição de $y = (y_1, \dots, y_n)$, dado um vetor de parâmetros aleatórios $\theta = (\theta_1, \dots, \theta_k)$. Observe-se que $p(\cdot)$ indica a função de probabilidade ou a função de densidade de uma variável aleatória Y , conforme esta seja discreta ou contínua. Assume-se que os dados são uma amostra aleatória da população em estudo, levando a que a variabilidade de y dependa apenas da seleção da amostra, uma vez que a incerteza resulta do facto de se observar apenas essa amostra e não todas as outras possíveis (Blangiardo & Cameletti, 2015).

O parâmetro θ é modelado através de uma distribuição com função de probabilidade *a priori* $p(\theta)$.

Essa função é definida antes de se observar y e reflete a crença anterior em θ .

Assim, a verosimilhança informa sobre um parâmetro por meio dos dados, enquanto a distribuição *a priori* informa por meio de crenças ou suposições anteriores. Quando há grandes quantidades de dados, a probabilidade contribuirá mais para a estimativa do risco relativo. Quando o conjunto de dados é pobre, as distribuições *a priori* dominarão a análise.

O produto da verosimilhança e da distribuição *a priori* é chamado de distribuição *a posteriori*. Essa distribuição descreve o comportamento dos parâmetros depois de se observar os dados e as suposições *a priori* serem feitas.

Dada a distribuição de verosimilhança $p(y|\theta)$ e a distribuição *a priori* com função $p(\theta)$, numa perspetiva Bayesiana o problema inferencial é resolvido usando o teorema de Bayes:

$$p(\theta|y) = \frac{p(y|\theta) \times p(\theta)}{p(y)} \quad (2.48)$$

onde $p(\theta|y)$ designa a distribuição *a posteriori* que representa a incerteza sobre o parâmetro de interesse θ , depois de se ter observado os dados, portanto, condicionado por y . O

denominador $p(y)$ é a distribuição marginal de y , também conhecida como distribuição preditiva anterior.

Quando θ é uma variável contínua $p(y)$ é dada por:

$$p(y) = \int p(y|\theta) p(\theta) d\theta \quad (2.49)$$

Se θ é uma variável discreta, então:

$$p(y) = \sum p(y|\theta) p(\theta) \quad (2.50)$$

Como $p(y)$ não depende de θ e, portanto, é considerada uma constante de normalização, a distribuição *a posteriori* $p(\theta|y)$, dada em (2.46), pode ser definida alternativamente como:

$$p(\theta|y) \propto p(y|\theta) \times p(\theta) \quad (2.51)$$

que significa que $p(\theta|y)$ é proporcional (\propto) a $p(y|\theta) \times p(\theta)$.

Uma das vantagens de se usar uma abordagem Bayesiana é o acesso à distribuição *a posteriori* para obter o ou os parâmetros de interesse, sendo a base da inferência Bayesiana. Resumir a informação dada pela distribuição *a posteriori* sobre os parâmetros do modelo através de alguns indicadores é possível e útil. Se o modelo depende apenas de um parâmetro, uma estimativa pontual natural de θ será alguma medida de centralidade. As escolhas mais comuns são:

a média posterior

$$\hat{\theta} = E(\theta|y) = \int_{\theta \in \Theta} \theta p(\theta|y) d\theta \quad (2.52)$$

a mediana posterior

$$\hat{\theta}: \int_{-\infty}^{\hat{\theta}} p(\theta|y) d\theta = 0.5 \quad (2.53)$$

e *a moda posterior*

$$\hat{\theta}: p(\hat{\theta}|y) = \sup_{\theta \in \Theta} p(\theta|y) \quad (2.54)$$

Θ é o conjunto de todos os valores possíveis para θ .

A distribuição *a posteriori* permite fazer qualquer declaração de probabilidade desejada sobre θ recorrendo a intervalos de confiança Bayesianos (intervalos de credibilidade). Um intervalo de credibilidade pode ser facilmente obtido por:

$$p(a < \theta < b | y) = \int_a^b p(\theta|y) d\theta = 1 - \alpha \quad (2.55)$$

onde $1 - \alpha$ é o nível de credibilidade desejado. A generalização para mais de um parâmetro não é trivial, embora seja direta.

Na abordagem Bayesiana, um intervalo de credibilidade indicará explicitamente a probabilidade posterior de θ estar dentro de seus limites, $p(\theta \in CI|y)$; isto é possível pelo facto de o parâmetro de interesse estar associado a uma distribuição de probabilidade, para que se possa fazer afirmações probabilísticas e ter em consideração a incerteza subjacente. Em estudos epidemiológicos é comum calcular a probabilidade posterior de mudança do risco relativo de uma doença, $p(\theta > 1|y)$ (Blangiardo & Cameletti, 2015).

Seleção da distribuição *a priori*

A questão fundamental da inferência Bayesiana é a escolha da distribuição *a priori* dos parâmetros de interesse, pois representa a informação disponível sobre os mesmos. No entanto, a arbitrariedade dos procedimentos de seleção prévia e a falta de análise de sensibilidade realista são um argumento sério contra a prática Bayesiana atual (Simpson et al., 2017).

Blangiardo & Cameletti (2015) destacam dois aspetos importantes que precisam ser considerados na escolha da distribuição *a priori*: o tipo de distribuição que deve ser representativa da natureza dos parâmetros de interesse, e os hiperparâmetros que são responsáveis por a distribuição ser mais ou menos informativa. No respeitante ao tipo de distribuição *a priori*, por exemplo, se o parâmetro de interesse for uma contagem, a escolha típica é uma distribuição de Poisson ou uma Binomial Negativa.

Quando a distribuição *a posteriori* pertence à mesma família da distribuição *a priori*, essa *a priori* é descrita como sendo distribuição *a priori* conjugada da função de verosimilhança (Bernardo & Smith, 2004). Dada a função de verosimilhança $p(y|\theta)$, a escolha da *priori* conjugada é muito conveniente do ponto de vista algébrico, uma vez que permite obter a distribuição *a posteriori* com a integral (2.49) que é tratável. Escolhida *a priori* conjugada da função de verosimilhança, conhece-se a distribuição funcional *a posteriori* e os seus hiperparâmetros. Assim, é fácil derivar estatísticas resumidas ou outras quantidades de interesse, nomeadamente (2.57).

Na família de distribuições exponenciais, a verosimilhança é uma simples função padronizada dos parâmetros e pode-se definir a distribuição *a priori* imitando a forma da verosimilhança. A multiplicação da verosimilhança por uma distribuição *a priori* da mesma família retém a forma comum destas duas funções. Ou seja, multiplicando uma

verosimilhança por uma distribuição *a priori* da família exponencial, obtém-se uma distribuição *a posteriori* da família exponencial. E esta família de funções é facilmente integrável.

Quando se lida com uma variável que representa dados de contagem (Y) (em particular, com a contagem de casos de uma doença), portanto, uma variável discreta que pode assumir valores entre 0 e $+\infty$, esta pode ser modelada usando uma distribuição de *Poisson*(λ).

$$p(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \text{ com } \lambda > 0 \quad (2.56)$$

A distribuição *a priori* mais utilizada para a distribuição *Poisson*(λ) é

$$p(\lambda|\alpha) \propto \lambda^{\alpha_1-1} e^{-\alpha_2\lambda} \quad (2.57)$$

que corresponde a uma distribuição *Gama*($\alpha = (\alpha_1, \alpha_2)$):

$$p(\lambda|\alpha) = K(\alpha)\lambda^{\alpha_1-1} e^{-\alpha_2\lambda} \quad (2.58)$$

sendo

$$K(\alpha) = \frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} \quad (2.59)$$

Dadas N observações de Poisson, $y = (y_1, y_2, \dots, y_N)^T$, a atualização da *priori* para a *posteriori* envolve a adição de uma estatística suficiente $\sum_{n=1}^N y_n$ para λ a α_1 e o número de observações N a α_2 .

De facto, multiplicando a função de verosimilhança (2.56) pela *priori* (2.57) é imediato que a distribuição *a posteriori* tem a forma funcional de uma variável aleatória Gama

$$\lambda|\alpha \sim \text{Gama}\left(\alpha_1 + \sum_{n=1}^N y_n, \alpha_2 + N\right) \quad (2.60)$$

Esta também é a distribuição *a priori* conjugada da distribuição de Poisson, neste processo de atualização.

Para provar que a soma $T(y) = \sum_{n=1}^N y_n$ é uma estatística suficiente para o parâmetro λ de uma distribuição de Poisson, considere-se que $y = (y_1, y_2, \dots, y_N)$ é um vetor de variáveis aleatórias independentes e identicamente distribuídas, seguindo uma distribuição de Poisson com parâmetro $\lambda > 0$.

O Teorema da Fatorização de Neyman-Fisher afirma que uma estatística $T(y)$ é suficiente para λ se, e só se, a função de verosimilhança conjunta puder ser fatorizada na forma:

$$L(y_1, y_2, \dots, y_n; \lambda) = g(T(y); \lambda) \cdot h(y_1, y_2, \dots, y_n), \quad (2.61)$$

onde g é uma função que depende de λ e dos dados apenas através de $T(y)$, e h não depende de λ .

Como as variáveis y_1, y_2, \dots, y_N são independentes:

$$\begin{aligned} L(y_1, y_2, \dots, y_N; \lambda) & \quad (2.62) \\ &= \prod_{i=1}^N P(Y_i = y_i; \lambda) = \prod_{i=1}^N \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \\ &= e^{-N\lambda} \lambda^{\sum_{i=1}^N y_i} \prod_{i=1}^N \frac{1}{y_i!} = e^{-N\lambda} \lambda^T \prod_{i=1}^N \frac{1}{y_i!} \end{aligned}$$

Observa-se que:

- $g(T; \lambda) = e^{-N\lambda} \lambda^T$ depende de λ e da soma das observações $T = \sum_{i=1}^N y_i$.
- $h(y_1, y_2, \dots, y_n) = \prod_{i=1}^n \frac{1}{y_i!}$ não depende de λ .

Portanto, a função de verosimilhança pode ser escrita como (2.61), ou seja, pelo Teorema da Fatorização de Neyman-Fisher, T é uma estatística suficiente para λ .

A média e a variância da distribuição Gama são determinadas da seguinte forma:

$$E(\lambda|\alpha) = \frac{\alpha_1}{\alpha_2} \quad (2.63)$$

e

$$Var(\lambda|\alpha) = \frac{\alpha_1}{\alpha_2^2} \quad (2.64)$$

α_1 e α_2 são os hiperparâmetros que definem a distribuição Gama e representam a forma e a escala da distribuição, respetivamente.

No caso de ρ representar a taxa de uma doença (risco relativo) e E o número esperado de casos dessa doença numa determinada unidade espacial ou temporal, então $\lambda = \rho E$ e neste caso θ é ρ .

A forma da verosimilhança é, neste caso, dada por:

$$p(y|\rho) = \frac{(\rho E)^y e^{-\rho E}}{y!}, \text{ com } \rho > 0 \quad (2.65)$$

A distribuição *a priori* para ρ é, regra geral, a distribuição Gama (2.56).

Após simplificar o produto de (2.57) por (2.65) e deixando não considerando $\frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} \times \frac{1}{y!}$ que não dependem de ρ , a distribuição *a posteriori* para ρ é definida por:

$$p(\rho|y) \propto \rho^{\alpha_1+y-1} e^{-(\alpha_2+E)\rho} \quad (2.66)$$

Esta distribuição tem também a forma funcional de uma variável aleatória Gama

$$\rho|y \sim Gama(\alpha_1 + y, \alpha_2 + E) \quad (2.67)$$

No entanto, na maioria das aplicações práticas, são necessários métodos numéricos ou de simulação para realizar inferência Bayesiana, pois modelos conjugados têm uma flexibilidade muito limitada: nem todas as verosimilhanças têm uma distribuição *a priori* conjugada associada ou então a conjugação é quebrada quando são especificados modelos de regressão linear generalizados. Alguns desses métodos serão descritos na secção 2.4.2.

Após a especificação da forma da distribuição *a priori*, os seus parâmetros (hiperparâmetros) devem ser definidos de acordo com o conhecimento prévio informativo/não informativo que se tenha do problema em análise.

Se houver informação prévia sobre os parâmetros, obtida por experiências anteriores sobre o mesmo problema ou a partir da opinião de especialistas, esta deve ser incorporada no modelo por meio da distribuição *a priori*. O'Hagan et al. (2006) descrevem como é feita metodologicamente a eliciação por especialistas e como é que esta pode ser incorporada nos modelos. J. Chen et al.,(2022) recorrem à eliciação por especialistas para a distribuição *a priori* a usar para definir um Modelo Hierárquico Bayesiano, com o objetivo de estudar a subnotificação de COVID-19 ao nível dos estados nos Estados Unidos, no fim de abril de 2020.

Quando há falta de informação sobre os parâmetros, geralmente, é recomendado usar distribuições *a priori* não informativas, para permitir que os dados falem por si. Existem várias maneiras de construir distribuições *a priori* não informativas, sendo uma das distribuições *a priori* não informativa mais conhecidas aquela que é baseada na informação de Fisher (Jeffreys, 1946). Para modelos com um parâmetro, esta distribuição é função da verosimilhança e é invariante por transformações. É definida por:

$$p(\theta) \propto I(\theta)^{\frac{1}{2}} \quad (2.68)$$

onde

$$I(\theta) = E \left[\left(\frac{\partial \ln p(y|\theta)}{\partial \theta} \right)^2 \middle| \theta \right] \quad (2.69)$$

Quando o número de parâmetros do modelo é superior a um, a distribuição *a priori* é dada por:

$$p(\theta) \propto |I(\theta)|^{\frac{1}{2}} \quad (2.70)$$

onde $I(\theta)$ é a matriz de Informação de Fisher, em que cada elemento é dado por:

$$I_{ij}(\theta) = (\theta) = E \left[\left(\frac{\partial \ln p(y|\theta)}{\partial \theta_i} \frac{\partial \ln p(y|\theta)}{\partial \theta_j} \right) \middle| \theta \right] \quad (2.71)$$

Na maioria dos casos, estas distribuições *a priori* são impróprias, isto é, se as integrarmos, a integral não é 1, o que é um problema quando a distribuição *a posteriori* também é imprópria, o que em problemas complexos pode ocorrer, sendo de difícil verificação.

Em alternativa, pode-se adotar distribuições *a priori* não informativas em que a ignorância sobre os parâmetros é assegurada apenas num subconjunto dos parâmetros com verosimilhança longe de 0. Estas distribuições *a priori* dizem-se planas ou difusas no sentido em que a distribuição *a priori* é basicamente plana na região da verosimilhança onde o parâmetro tem mais peso. Regra geral, distribuições *a priori* com elevada variabilidade têm esta característica (A. B. Lawson, 2018).

Uma $Normal(0, 10^6)$ pode ser usada como a distribuição *a priori* para uma média ou um parâmetro de regressão, como sugerem (Blangiardo & Cameletti, 2015). Uma discussão sobre como formular distribuições *a priori* pode ser encontrada em Paulino et al. (2018).

A escolha mais comum de distribuições *a priori* não é subjetiva, pois, em primeiro lugar, pode não haver especialistas e, em segundo lugar, porque, para modelos hierárquicos complexos, o processo de elicição não é direto. Assim, Simpson et al. (2017) sugerem uma estrutura chamada *Complexidade Penalizada* ou *priors 'PC'* em que se definem as componentes de modelo como uma extensão flexível de um modelo básico e se definem distribuições *a priori* apropriadas para penalizar a complexidade induzida no modelo pelo desvio em relação ao modelo base mais simples. Esta estrutura ajuda o utilizador a especificar distribuições *a priori* informativas que usam apenas informações fracas, são diretas o suficiente para serem usadas por utilizadores em geral e têm significado e interpretação claros. Esta estrutura é especialmente útil na formulação de Modelos Hierárquicos Bayesianos.

Inferência Preditiva Bayesiana

Além de permitir inferências sobre um parâmetro desconhecido de interesse, a abordagem Bayesiana também permite fazer previsões sobre variáveis não observadas.

A distribuição preditiva de uma variável aleatória é a distribuição marginal (dos valores não observados) após contabilizar a incerteza nos parâmetros. Uma distribuição preditiva *a*

priori é calculada usando a distribuição *a priori* dos parâmetros. Uma distribuição preditiva *a posteriori* é calculada usando a distribuição *a posteriori* dos parâmetros, condicionada aos dados observados.

Quando Y_{n+1} é uma observação futura do mesmo modelo que gerou $y = (y_1, \dots, y_n)$, a distribuição preditiva *a posteriori* é, por definição, dada por:

$$p(y_{n+1}|y) = E_{(\theta|Y_{n+1})} = \int p(y_{n+1}|\theta, y)p(\theta|y)d\theta \quad (2.72)$$

Se Y_{n+1} for independente de y , então (2.72) simplifica para:

$$p(y_{n+1}|y) = \int p(y_{n+1}|\theta)p(\theta|y)d\theta \quad (2.73)$$

As probabilidades preditivas podem ser calculadas recorrendo à lei da probabilidade total, como médias ponderadas sobre valores possíveis de θ . No entanto, mesmo quando as distribuições condicionais de dados, dados os parâmetros, são bem conhecidas, as distribuições marginais dos dados geralmente não são. A simulação é uma ferramenta eficaz na aproximação de distribuições preditivas.

Algoritmo geral para simulação de distribuição preditiva *a posteriori* (*a priori*)

1. Para $i \in \{1, \dots, n\}$
 - Simular um valor θ_i , a partir de distribuição *a posteriori* (ou distribuição *a priori*).
 - Dado este valor de θ_i , simular um valor de y_i a partir de $f(y, \theta)$, o modelo de dados condicional em θ .
 2. Resumir os valores de y_i para aproximar a distribuição preditiva *a posteriori* (ou distribuição preditiva *a priori*).
-

2.4.2. Métodos computacionais de apoio à Inferência Bayesiana

A inferência e a modelação Bayesiana assentam na integração, e nem sempre se lida com integrais tratáveis analiticamente, sendo necessário recorrer a simulações ou integração numérica para aproximar esses integrais (Evans & Swartz, 1995).

As simulações estatísticas geram valores aleatórios a partir de uma determinada função de densidade e pode-se tornar um procedimento computacionalmente intensivo quando as distribuições envolvidas são complexas ou a quantidade de dados é grande. Como referido,

os métodos Bayesianos são baseados em simulação ou aproximações para explorar a distribuição *a posteriori* quando não é possível manipulá-la analiticamente.

De seguida, descrevem-se alguns métodos computacionais para inferência Bayesiana, nomeadamente, o *Método de Monte Carlo via Cadeias de Markov (MCMC) – algoritmo de amostragem de Gibbs* (Geman & Geman, 1984) e Aproximações de Laplace Aninhadas e Integradas -INLA (*Integrated Nested Laplace Approximations*) (Rue et al., 2009).

i) Métodos de Monte Carlo via Cadeias de Markov (MCMC)

Considerem-se alguns exemplos da inferência Bayesiana resolvidos com aproximação da distribuição *a posterior* com recurso a métodos MCMC.

1. Aproximar a média posterior de uma função real de θ , definida como:

$$E(g(\theta|y)) = \int g(\theta)p(\theta|y)d\theta \quad (2.74)$$

Dada uma amostra aleatória $\theta_1, \dots, \theta_n$ da distribuição *a posteriori* $p(\theta|y)$, o método de Monte Carlo aproxima a integral em (2.74) pela média empírica:

$$\hat{E}(g(\theta|y)) = \frac{1}{n} \sum_{i=1}^n g(\theta_i) \quad (2.75)$$

2. A função de distribuição preditiva de uma observação futura, $p(y_{n+1}|y)$, pode ser escrita como $E_{\theta|y}[p(y_{n+1}|\theta, y)]$ (Paulino et al., 2018).

Uma aproximação de Monte Carlo para esta quantidade é dada por:

$$\hat{p}(y_{n+1}|y) = \frac{1}{n} \sum_{i=1}^n p(y_{n+1}|\theta_i, y) \quad (2.76)$$

Para obter a aproximação da média posterior ou outras estatísticas posteriores da distribuição preditiva ou de intervalos de credibilidade é necessário amostrar diretamente da distribuição *a posteriori*, o que nem sempre é possível, dado que algumas distribuições *a posteriori* ou têm uma forma desconhecida ou são definidas por uma função que não é facilmente manipulável analiticamente. Uma solução possível é simular uma amostra através de uma Cadeia de Markov cuja distribuição estacionária é a distribuição *a posteriori*. Os métodos MCMC surgem dessa ideia em combinação com os métodos de Monte Carlo para calcular as estatísticas posteriores de interesse.

Os dois algoritmos MCMC mais populares são o da amostragem de Gibbs e o algoritmo Metropolis Hastings.

Algoritmo de amostragem de Gibbs

Supondo que o modelo apresenta k parâmetros $\theta = (\theta_1, \dots, \theta_k)^T$ e assumindo que as amostras podem ser geradas a partir de cada uma das distribuições condicionais completas definidas por:

$$P = \{p(\theta_i | \theta_{j \neq i}, y), i = 1, \dots, k\}, \text{ onde } \theta = (\theta_1, \dots, \theta_k)^T, \quad (2.77)$$

sempre conhecidas até à constante de proporcionalidade, uma vez que assumem a forma *Verossimilhança* \times *Priori* com todos os valores fixos exceto θ_i , o método de Gibbs permite gerar uma sequência de amostras da distribuição conjunta *a posteriori* de θ . A ideia deste método é que, sob condições suaves, a coleção de distribuições condicionais completas determina exclusivamente a distribuição conjunta posterior $p(\theta|y)$.

Para amostrar a partir da distribuição *a posteriori* $p(\theta|y)$, o amostrador de Gibbs extrai valores iterativamente de todas as distribuições condicionais de P (2.77).

Dado um conjunto arbitrário de valores iniciais $\{\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}\}$, o algoritmo de amostragem de Gibbs opera da seguinte forma:

Algoritmo de amostragem de Gibbs

Para ($t \in 1:T$), repita:

- Simular $\theta_1^{(t)}$ a partir de $p(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, y)$
 - Simular $\theta_2^{(t)}$ a partir de $p(\theta_2 | \theta_1^{(t-1)}, \theta_3^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, y)$
 -
 - Simular $\theta_k^{(t)}$ a partir de $p(\theta_k | \theta_1^{(t-1)}, \theta_2^{(t-1)}, \dots, \theta_{k-1}^{(t-1)}, y)$
-

Sob condições de regularidade moderada que geralmente são satisfeitas pela maioria dos modelos estatísticos, a realização $(\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t)})$ obtida na iteração t converge em distribuição para um k – *tuplo* extraído da verdadeira distribuição conjunta *a posteriori* $p(\theta|y)$. Portanto, para t suficientemente grande, maior que t_0 , $\{\theta^{(t)}, t = t_0 + 1, \dots, T\}$ é uma amostra da verdadeira distribuição *a posteriori*, a partir da qual quaisquer quantidades posteriores podem ser estimadas (Clark & Gelfand, 2006).

Como referido, as distribuições condicionais completas podem não ter formas familiares e, conseqüentemente, pode não ser possível obter amostras diretamente de tais distribuições.

Nesses casos, são geralmente recomendados métodos alternativos como o algoritmo Metropolis-Hastings.

Algoritmo Metropolis-Hastings

Quando a distribuição *a priori* $p(\theta)$ e a verosimilhança $f(y|\theta)$ não são um par conjugado, uma ou mais das distribuições condicionais completas de P (2.77), necessárias para implementar o algoritmo Gibbs, podem não estar disponíveis na forma fechada, o que inviabiliza a implicação deste algoritmo. No entanto, conforme observado acima, mesmo nesta configuração, $p(\theta_i|\theta_{j \neq i}, y)$ estará disponível até uma constante de proporcionalidade, pois é proporcional à porção de $f(y|\theta) \times p(\theta)$ que envolve θ_i . O algoritmo de Metropolis-Hastings requer apenas uma função proporcional à distribuição a ser amostrada, exigindo uma etapa de rejeição de uma densidade candidata e $q(\theta^*|\theta^{(t-1)})$ que é uma função de densidade válida para cada valor possível da variável condicional $\theta^{(t-1)}$.

Dado um valor inicial $\theta^{(0)}$ na iteração $t = 0$ e $h(\theta)$ a função de densidade da distribuição da qual queremos amostrar, conhecida a menos de uma constante de proporcionalidade, o algoritmo procede da seguinte forma:

Algoritmo Metropolis-Hastings

Para $(t \in 1:T)$, repita:

- Simular θ^* a partir de $q(\theta^*|\theta^{(t-1)})$
 - Calcular a razão de aceitação $r = \frac{h(\theta^*) q(\theta^{(t-1)}|\theta^*)}{h(\theta^{(t-1)}) q(\theta^*|\theta^{(t-1)})}$.
 -
 - Se $r \geq 1$, fazer $\theta^{(t)} = \theta^*$;
 - Se $r < 1$, fazer $\theta^{(t)} = \begin{cases} \theta^* & \text{com probabilidade } r \\ \theta^{(t-1)} & \text{com probabilidade } 1 - r \end{cases}$
-

Sahu (2022) desenvolveu o pacote `bmstdr` do R que facilita a modelação Bayesiana de dados espaciais e espaço-temporais referenciados em pontos e unidades de área espacial, implementando a inferência com recurso a métodos MCMC. Recorrendo a vários pacotes do R, permite que o utilizador escolha e valide o modelo sem que tenha de escrever qualquer código específico.

2.4.3. INLA - Integrated Nested Laplace Approximations

(AmaralTurkman et al., 2019; Blangiardo & Cameletti, 2015)

A análise de dados espaciais e temporais permite fazer inferências confiáveis e previsões robustas, o que é fundamental para muitas questões de interesse, nomeadamente na epidemiologia, área onde a análise de diversos tipos de dados complexos tem sido bem sucedida (Blangiardo et al., 2020; J. Chen et al., 2022; Zou et al., 2014). Em particular, a aplicabilidade geral de Modelos Hierárquicos Bayesianos tem sido aprimorada pelos avanços dos algoritmos computacionais, nomeadamente aqueles pertencentes à família dos algoritmos estocásticos baseados nas técnicas de Monte Carlo via Cadeia de Markov.

Estes modelos oferecem propriedades convenientes, como a flexibilidade em relação à distribuição de probabilidade da variável resposta, juntamente com a sua capacidade de avaliar simultaneamente as contribuições de preditores fixos e efeitos aleatórios espaciais e/ou temporais para a verosimilhança, o que garante que a variabilidade seja devidamente associada às previsões (Clark & Gelfand, 2006) e, se necessário, permitem fazer previsões em locais não amostrados, acomodando assim lacunas na amostragem e intensidade de amostragem irregular. Nesta tese utiliza-se a metodologia estatística INLA – *Integrate Nested Laplace Approximation* (Rue et al., 2009) para estimar modelos espaciais ou espaço-temporais hierárquicos, em alternativa às simulações MCMC.

Contrariamente aos métodos de MCMC que permitem aproximar a distribuição *a posteriori* através da simulação, esta alternativa analítica baseada em aproximações de Laplace Integradas e Aninhadas (INLA) proposta por Rue et al. (2009), permite a aproximação determinística de distribuições marginais *a posteriori* dos parâmetros de interesse, embora não se esgote nela. Segundo AmaralTurkman et al. (2019) este método fornece uma implementação particularmente eficiente da inferência Bayesiana com os Modelos Gaussianos Latentes (MGL), os quais, dada a sua flexibilidade, acomodam os Modelos Lineares Generalizados e os modelos espaciais e espaço-temporais. A abordagem INLA tem duas vantagens principais sobre as técnicas MCMC. A primeira é o tempo de computação para a obtenção de resultados e a segunda é a forma unificada como o INLA trata os MGL, permitindo uma substancial automatização da inferência, independente do modelo específico considerado.

A classe de Modelos Gaussianos Latentes pode ser especificada por uma estrutura hierárquica em três níveis:

1. Dados | Parâmetros (modelo amostral para y)

$$Y|\theta, \psi \sim f(y|\theta, \psi) = \prod_{i=1}^n f(y_i | \theta, \psi) \quad (2.78)$$

Neste nível a verosimilhança univariada capta a distribuição marginal dos dados e usualmente é escolhida na família exponencial, em particular, os modelos de Poisson ou Binomial Negativo.

Neste nível da especificação do modelo é assumido o pressuposto de que os valores de “amostra” de Y , dados os parâmetros, são independentes e, portanto, é possível obter o produto das contribuições individuais expressas em (2.78). Portanto, os dados são considerados condicionalmente independentes. Observe-se que, em muitas aplicações espaciais, os dados não seriam incondicionalmente independentes e estariam de facto correlacionados. Pressupõe-se a independência condicional da variável resposta Y , dado o conhecimento dos parâmetros do modelo a um nível mais elevado da hierarquia, garantindo assim que um valor observado y_i da área i pode ser considerado independente de outros resultados. No entanto, este pressuposto de independência condicional só pode ser válido se a correlação induzida por efeitos não observados for considerada em algum nível do modelo. A ideia de inclusão de correlação espacial num nível hierárquico acima da verosimilhança é uma suposição fundamental, frequentemente feita na modelação Bayesiana com dados de pequenas áreas. Isto significa que a correlação aparece em distribuições *a priori* e não na própria distribuição de probabilidade da variável resposta (A. B. Lawson, 2018).

2. Parâmetros|Hiperparâmetros (Campo aleatório Gaussiano de Markov para θ),

$$\theta|\psi \sim N(0, Q^{-1}(\psi)) \quad (2.79)$$

Neste nível assume-se que o vetor de parâmetros θ da função de verosimilhança é determinado por um campo gaussiano com estrutura markoviana (*Gaussian Markov Random Field* - GMRF), ou seja, a sua função densidade é dada por:

$$p(\theta|\psi) = (2\pi)^{-\frac{n}{2}} |Q(\psi)|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \theta^T Q(\psi) \theta\right) \quad (2.80)$$

Q é a matriz de precisão (isto é, a inversa da matriz de covariâncias) que, na prática, é indispensável que seja esparsa, do ponto de vista de tempo computacional e em termos de exigência de memória.

3. Hiperparâmetros (*Hiperprior* para ψ)

$$\psi \sim p(\psi) \tag{2.81}$$

Os hiperparâmetros podem surgir na verosimilhança como parâmetros de dispersão, em particular, como a variância da distribuição normal, o parâmetro de superdispersão do modelo Binomial Negativo ou parâmetro de forma da distribuição Gama. Podem ainda caracterizar a estrutura do modelo MGL através de variâncias, parâmetros de correlação espacial ou temporal ou coeficientes de autorregressão (Opitz, 2017).

Uma forma alternativa de especificar os MGL, central para a abordagem INLA, consiste em reconhecer o modelo como um caso especial do modelo de regressão com estrutura aditiva. Nesta formulação, a variável dependente Y pertence à família exponencial, onde a média μ_i está ligada a um preditor η_i com estrutura aditiva, através da função de ligação dada por $g(\mu_i) = \eta_i$.

A forma geral do preditor é

$$\eta_i = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} w_{ki} f^{(k)}(u_{ki}) + \epsilon_i \tag{2.82}$$

onde β_0 é o intercepto, $\beta = (\beta_1, \dots, \beta_{n_\beta})$ é o vetor com os efeitos fixos lineares das covariáveis z . As funções $(f^{(1)}, \dots, f^{(n_f)})$ de covariáveis u podem representar efeitos aleatórios estruturados, efeitos não lineares de covariáveis contínuas que incluem modelos *spline*, efeitos espaciais, efeitos temporais, efeitos espaço-temporais não separáveis, fragilidades, interseções e inclinações específicas de um sujeito ou grupo, etc. $(\{w_{ki}\})$ podem ser eventuais pesos conhecidos para cada observação. O termo ϵ_i pode acomodar efeitos aleatórios não estruturados.

Se se assumir um GMRF como distribuição *a priori* para $\theta = (\beta_0, \beta_j, f^{(k)}(u_{ki}), \eta_i)$, então obtém-se um modelo MGL.

Com o INLA obtêm-se aproximações analíticas para as distribuições marginais *a posteriori* dos parâmetros e dos hiperparâmetros do modelo.

Essas distribuições marginais são dadas por:

$$p(\theta_i | y) = \int p(\theta_i, \psi | y) d\psi = \int p(\psi | y) p(\theta_i | \psi, y) d\psi \tag{2.83}$$

e

$$p(\psi_k|y) = \int p(\psi|y)d\psi_{-k} \quad (2.84)$$

onde ψ_{-k} representa o vetor ψ sem a componente ψ_k .

Ambos as distribuições marginais dependem de $p(\psi|y)$, o que é dado por:

$$p(\psi|y) = \frac{p(\theta, \psi|y)}{p(\theta|\psi, y)} \quad (2.85)$$

onde $p(\theta, \psi|y)$ pode ser definido como:

$$\begin{aligned} p(\theta, \psi|y) &\propto p(\theta|\psi)p(\psi) \prod_i f(y_i|\theta_i, \psi) \\ &\propto p(\psi)|Q(\psi)|^{\frac{n}{2}} \exp\left(-\frac{1}{2}\theta^T(Q(\psi))\theta + \sum_i \ln(f(y_i|\theta_i, \psi))\right) \end{aligned} \quad (2.86)$$

Assim, resulta que:

$$p(\psi|y) \propto \frac{p(\theta|\psi)p(\psi)p(y|\theta, \psi)}{p(\theta|\psi, y)} \quad (2.87)$$

$$\tilde{p}(\psi|y) \propto \frac{p(\theta|\psi)p(\psi)(p(y|\theta, \psi))}{\tilde{p}(\theta|\psi, y)} \Bigg|_{\theta=\hat{\theta}(\psi)} \quad (2.88)$$

onde $\tilde{p}(\theta|\psi, y)$ é a aproximação gaussiana de $p(\theta|\psi, y)$, dada pelo método de Laplace, e $\hat{\theta}(\psi)$ é a moda para um dado ψ .

A aproximação para $p(\theta_i|\psi, y)$ pode ser feita usando uma das três abordagens seguintes:

1. Aproximar $p(\theta_i|\psi, y)$ diretamente pelas distribuições marginais de $\tilde{p}(\theta|\psi, y)$, usando uma distribuição Normal onde a decomposição de Cholesky é usada para a matriz de precisão.

No entanto, geralmente, essa aproximação não é muito boa.

2. Usar a aproximação de Laplace para obter:

$$p(\theta_i|\psi, y) = \frac{p(\theta_i, \theta_{-i}|\psi, y)}{p(\theta_{-i}|\theta_i, \psi, y)} \quad (2.89)$$

$$\propto \frac{p(\psi)p(\theta|\psi)f(y|\theta, \psi)}{p(\theta_{-i}|\theta_i, \psi, y)} \quad (2.90)$$

$$\approx \frac{p(\psi)p(\theta|\psi)f(y|\theta, \psi)}{\tilde{p}(\theta_{-i}|\theta_i, \psi, y)} \Bigg|_{\theta_{-i}=\hat{\theta}_{-i}(\theta_i, \psi)} \quad (2.91)$$

onde $\tilde{p}(\theta_{-i}|\theta_i, \psi, y)$ é a aproximação Gaussiana de Laplace para $p(\theta_{-i}|\theta_i, \psi, y)$ e $\hat{\theta}_{-i}(\theta_i, \psi)$ é a moda respectiva. Esta abordagem proporciona uma melhor aproximação do

que a anterior, mas como precisa ser recalculada para cada θ e ψ , já que a matriz de precisão depende de θ_i e ψ , é computacional mais morosa.

3. Usar a aproximação simplificada de Laplace que é baseada na expansão em série de Taylor de terceira ordem do logaritmo do numerador e do denominador em (2.91). θ_{-i} é substituído pela média condicional $E(\theta_{-i}|\theta_i)$ baseada numa aproximação normal $\tilde{p}(\theta | \psi, x)$, para reduzir o tempo computacional.

Normalmente, nesta abordagem o tempo computacional é curto e a aproximação é razoável. Detalhes sobre os resultados do INLA para essas aproximações são fornecidos em Blangiardo & Cameletti (2015) e Rue et al. (2009).

O modelo (2.82) pode ser formulado considerando uma das dimensões espaço (η_i) ou tempo (η_t), ou ainda considerando as duas dimensões num só modelo (η_{it}), de forma a poder modelar a dinâmica de processos observados no espaço, no tempo ou no espaço e tempo, em casos como é o da epidemia da COVID 19.

2.5. Modelos Hierárquicos Bayesianos para dados contagem

Os modelos de dados espaciais ou espaço-temporais devem incorporar a dependência espacial e temporal. Em particular, a dependência espacial nos modelos de dados de área pode ser considerada através de efeitos aleatórios baseados na estrutura espacial da rede de vizinhanças. No entanto, há que considerar que a inclusão de efeitos aleatórios num modelo com verosimilhança não normal pode, por vezes, induzir alguma complexidade no modelo e que a modelação hierárquica Bayesiana pode ser uma forma eficaz de lidar com essa complexidade. O Modelo Hierárquico Bayesiano vincula os submodelos, propagando corretamente as incertezas em cada submodelo de um nível para o outro.

A estrutura dos Modelos Hierárquicos Bayesianos é uma estratégia genérica de construção de modelos em que quantidades não observadas, como, por exemplo, parâmetros estatísticos, dados ausentes, efeitos aleatórios, são organizadas num pequeno número de níveis discretos com funções logicamente distintas e cientificamente interpretáveis, e relações probabilísticas entre elas que capturam características inerentes dos dados (Richardson & Best, 2003). Estes modelos são adequados para analisar muitos tipos de dados epidemiológicos e biomédicos complexos e a sua aplicabilidade tem melhorado com o avanço dos algoritmos computacionais, quer os pertencentes à família dos algoritmos estocásticos baseados nas técnicas de Monte Carlo via Cadeias de Markov, quer outros como o INLA.

Uma das formas mais gerais e versáteis de estimar o efeito de covariáveis sobre dados de contagem, tendo em consideração correlações temporais e/ou espaciais é utilizar um modelo linear generalizado hierárquico, no qual os dados são ligados a covariáveis e efeitos aleatórios espaciais e/ou temporais por meio de uma função de verosimilhança, apropriadamente escolhida, e uma função de ligação que é linear nas covariáveis e nos efeitos aleatórios.

Considerando esta abordagem de modelação, a estrutura de um modelo linear generalizado pode ser especificada em três níveis:

1. Dados – Função de ligação
2. Função de ligação – Parâmetros, termos latentes
3. Parâmetros – Hiperparâmetros

Uma das abordagens mais desenvolvidas neste âmbito é o mapeamento de doenças, também designado por epidemiologia espacial, epidemiologia ambiental ou estudos de saúde em pequenas áreas (A. B. Lawson, 2018).

O mapeamento de doenças é uma área de forte interesse epidemiológico que se concentra nas principais áreas da modelação hierárquica Bayesiana e a sua aplicação à análise de doenças. A. B. Lawson (2018) apresenta um panorama muito completo sobre o uso de modelos Bayesianos e métodos computacionais em aplicações a estudos em bioestatística espacial, com dados de saúde. A. Lawson & Choi (2016) apresentam diversas aplicações noutras áreas.

Seja y_{jt} o número de casos observados na unidade espacial (também designada por área) j , $j = 1, 2, \dots, m$ e na unidade de tempo t , $t = 1, 2, \dots, T$. Na contagem de doenças é usual assumir que o valor médio ($E(Y_{jt}) = \mu_{jt}$) tem duas componentes: 1) uma componente que representa o efeito de fundo da população numa determinada área – número de casos esperados (E_{jt}); e 2) uma componente que representa o excesso de risco nessa área – risco relativo (ρ_{jt}). Assume-se que y_{jt} segue uma distribuição de Poisson de média $\mu_{jt} = E_{jt}\rho_{jt}$ e é condicionalmente independente, conhecidos os parâmetros.

Sendo P_j a população na área j , então o risco relativo na área j é definido como $\rho_{jt} = \frac{y_{jt}}{P_j}$. O número esperado de casos de doença em cada área, geralmente, é considerado fixo para o período de tempo considerado (A. Lawson & Choi, 2016).

O modelo espaço-tempo padrão (não paramétrico) do mapeamento de doenças é dado por:

$$y_{jt} \sim \text{Poisson}(E_{jt}\rho_{jt}) \quad (2.92)$$

com

$$\log(\rho_{jt}) = \alpha_0 + u_j + v_j + \gamma_t + \varphi_t \quad (2.93)$$

onde α_0 é o intercepto que quantifica o logaritmo da taxa geral de risco, v_j é um efeito específico da área, não estruturado, modelado como permutável (a distribuição de probabilidade conjunta das observações é invariante sob permutação), u_j é outro efeito específico da área, modelado como espacialmente estruturado, φ_t é um efeito específico do tempo, não estruturado, modelado como permutável e γ_t é outro efeito específico do tempo, modelado como temporalmente estruturado.

Em relação aos efeitos aleatórios espaciais, pode-se considerar o modelo Besag–York–Mollié (BYM) (Besag et al. 1991), que estabelece a distribuição condicional do efeito espacialmente estruturado na área i .

Os pressupostos usuais do modelo são:

1. u pode ser modelado como um processo autorregressivo condicional intrínseco (*iCAR*), proposto por Besag et al. (1991), com a seguinte especificação

$$u_i | u_{-i} \sim \text{Normal} \left(\frac{1}{N_i} \sum_{j=1}^n w_{ij} u_j, s_i^2 \right) \quad (2.94)$$

onde u_{-i} indica todos os elementos em u exceto o i – ésimo, N_i é o conjunto de vizinhos da unidade espacial i , w_{ij} é o elemento (i, j) da matriz de vizinhança W que é usualmente a matriz de contiguidades cujos elementos valem 1 se as unidades espaciais i e j são vizinhas e 0 caso contrário e que é normalizada por linha, $s_i^2 = \sigma_u^2$ é a variância para a unidade espacial i .

Note-se que a matriz de vizinhanças W pode ter outras definições, nomeadamente, baseadas em distâncias entre unidades espaciais ou em covariáveis (Briz-Redón et al., 2022).

Este modelo assume apenas um parâmetro espacial estruturado e, portanto, não capta as particularidades que as áreas possam possuir. Foi proposto outro modelo para os efeitos espaciais aleatórios, que também é bastante utilizado e que é definido como modelo de convolução - modelo Besag–York–Mollié (BYM). Besag, York & Mollié (1991)

combinaram o modelo intrínseco que possui u_i espacialmente estruturado, com um efeito aleatório adicional v_i , espacialmente não estruturado.

2. O efeito espacialmente não estruturado sobre áreas, v_j , segue uma distribuição Gaussiana,

$$v_j \sim \text{Normal}(0, \sigma_{v_j}^2) \quad (2.95)$$

onde $\sigma_{v_j}^2$ é a variância do efeito.

3. γ é modelado como um passeio aleatório de primeira ordem (*AR (1)*) definido como

$$\gamma_t | \gamma_{t-1} \sim \text{Normal}(\gamma_{t-1}, \sigma_\gamma^2) \quad (2.96)$$

Note-se que γ também pode ser especificado como um passeio aleatório de segunda ordem (*AR (2)*)

$$\gamma_t | \gamma_{t-1}, \gamma_{t-2} \sim \text{Normal}(2\gamma_{t-1} + \gamma_{t-2}, \sigma_\gamma^2) \quad (2.97)$$

ou com uma regressão polinomial de tendência no tempo (A. Lawson et al., 2016).

4. O efeito temporal não estruturado, φ_t , segue uma distribuição Gaussiana:

$$\varphi_t \sim \text{Normal}(0, \sigma_{\varphi_t}^2) \quad (2.98)$$

onde $\sigma_{\varphi_t}^2$ é a variância do efeito.

Knorr-Held (2000) expande este modelo para permitir a interação entre tempo e espaço, usando o seguinte preditor linear:

$$\ln(\mu_{jt}) = \alpha_0 + u_j + v_j + \gamma_t + \varphi_t + \delta_{jt} \quad (2.99)$$

onde δ_{jt} é o efeito de interação *espaço* \times *tempo*. A matriz dos termos de interação, ϕ , assume-se com distribuição Gaussiana, $\phi \sim \text{Normal}(\mathbf{0}, \sigma_\phi^2 \Sigma_\phi)$, sendo σ_ϕ^2 uma variância global e Σ_ϕ uma matriz de estrutura expressa como o produto de Kronecker das matrizes de estrutura dos efeitos principais espaço e tempo. O autor define quatro tipos de estruturas de interação, em função dos efeitos que se assumem para interagir.

Assim, um modelo espaço-temporal puro consiste em efeitos aleatórios espaciais e temporais estruturados e não estruturados e a sua interação. Os efeitos aleatórios espaciais e temporais estruturados captam a correlação espacial e temporal, respetivamente; os efeitos aleatórios espaciais e temporais não estruturados captam a heterogeneidade espacial e temporal, respetivamente. Os quatro tipos de interação entre os efeitos aleatórios espaciais e temporais (efeito espacial estruturado \times efeito temporal estruturado ou não estruturado;

efeito espacial não estruturado \times efeito temporal estruturado ou não estruturado) são captados pelo termo de interação (Blangiardo & Cameletti, 2015; Knorr-Held, 2000).

Seguindo Blangiardo & Cameletti (2015), os quatro tipos diferentes de especificação para a matriz de precisão R_δ , que implica diferentes graus de dependência prévia para os parâmetros de interação podem resumir-se como se segue:

- O Tipo I assume que os dois efeitos não estruturados v_j e φ_t dão origem à matriz de estrutura $R_\delta = R_v \otimes R_\varphi = I \otimes I = I$, porque tanto v quanto φ não têm uma estrutura espacial ou temporal. Consequentemente, assume-se também que não há estrutura espacial ou temporal na interação e, portanto,

$$\delta_{it} \sim N\left(0, \frac{1}{\tau_\delta}\right) \quad (2.100)$$

- O Tipo II combina o efeito principal temporal estruturado γ_t com o efeito espacial não estruturado v_j . Escreve-se a matriz de estrutura como $R_\delta = R_v \otimes R_\gamma$, onde $R_v = I$ e R_γ é a estrutura de vizinhança especificada, por exemplo, através de um passeio aleatório de primeira ou segunda ordem. Isto leva à suposição de que, para a área j , o vetor de parâmetros $\{\delta_{j1}, \dots, \delta_{jT}\}$ tem uma estrutura autorregressiva na componente temporal, que é independente das demais áreas.

- O Tipo III combina o efeito temporal não estruturado φ_t com o efeito principal espacialmente estruturado u_j . Escreve-se a matriz de estrutura como $R_\delta = R_\varphi \otimes R_u$, onde $R_\varphi = I$ e R_u é uma estrutura de vizinhança definida pela especificação CAR (*Conditional Autoregressive*). Isto leva à suposição de que os parâmetros do ponto de tempo t $\{\delta_{1t}, \dots, \delta_{nt}\}$, têm uma estrutura espacial independente dos outros pontos de tempo.

- O Tipo IV é o tipo de interação mais complexo, assumindo que os efeitos espacialmente e temporalmente estruturados u_i e γ_t interagem. A matriz de estrutura pode ser escrita como $R_\delta = R_u \otimes R_\gamma$, o que assume que a estrutura de dependência temporal para cada área não é independente de todas as outras áreas, mas depende também do padrão temporal das áreas vizinhas.

Normalmente, é usada uma especificação binária: 1 ou 0, para a matriz W que conduz a uma especificação esparsa, o que torna o ajustamento destes modelos muito mais eficiente do que se W fosse uma matriz densa.

2.6. Índices de comparação de modelos não aninhados baseados na Teoria da Informação

Os índices de informação AIC (*Akaike Information Criterion*) (Akaike, 1998) e BIC (*Bayes Information Criterion*) (Schwarz, 1978) são índices relativos de ajustamento que permitem comparar modelos alternativos para um conjunto de dados.

O melhor modelo é aquele que apresenta os valores menores num, ou em mais, destes índices.

O AIC reflete a extensão em que as matrizes de covariâncias observadas e previstas diferem umas das outras contemplando um termo que penaliza o modelo em função da sua complexidade. Modelos complexos e mal ajustados resultam em índices com valores elevados.

Os índices definem-se do seguinte modo:

AIC (Akaike Information Criterion)

$$AIC = -2 \ln(\hat{L}) + 2t \quad (2.101)$$

onde \hat{L} é o valor máximo da função de Máxima Verosimilhança do modelo e t o número de parâmetros livres a estimar no modelo.

Este índice favorece modelos muito complexos em pequenas amostras pelo facto de não ter em conta o efeito da dimensão da amostra na seleção do modelo. Para resolver este problema podem ser usadas estratégias que reduzam o peso do tamanho da amostra na seleção do modelo, para melhorar o desempenho do índice, como variantes *bootstrap* do critério AIC.

BIC (Bayes Information Criterion)

$$BIC = -2 \ln(\hat{L}) + t \ln(n) \quad (2.102)$$

onde \hat{L} é o valor máximo da função de Máxima Verosimilhança do modelo, t o número de parâmetros livres a estimar no modelo e n é o número de elementos da amostra. Este índice penaliza mais do que o AIC os modelos complexos e com amostras maiores.

Com este critério seleciona-se o modelo que mais provavelmente gerou os dados no "*sentido bayesiano*".

AIC é o critério mais popular para a modelação com Modelos Lineares Generalizados. No entanto, quando o tamanho da amostra é pequeno em relação ao conjunto de modelos candidatos, o AIC é menos protetor contra *overfitting* do que muitos outros critérios

baseados em probabilidade (Cavanaugh e Shumway, 1997; McQuarrie e Tsai, 1998; Rao e Wu, 2001).

Os critérios de seleção de modelos hierárquicos Bayesianos, mais usados são métodos baseados na *Deviance Bayesiana*: os critérios DIC (*Deviance Information Criterion*) e WAIC (*Widely Applicable Information Criterion*).

A *Deviance* é definida por:

$$D(\theta) = -2 \ln(p(y|\theta)) + 2 \ln(p(y)). \quad (2.103)$$

onde $p(y|\theta)$ é a função de verosimilhança e $2 \ln(p(y))$ é uma constante que depende apenas dos dados e tem efeito meramente padronizador e que cancela em todos os cálculos que comparam modelos diferentes e que, portanto, não precisa ser conhecida.

No caso Gaussiano, a complexidade do modelo é medida pelo número efetivo de parâmetros:

$$p_D = E_{\theta|y}(D(\theta)) - D(E_{\theta|y}) = \overline{D(\theta)} - D(\bar{\theta}) \quad (2.104)$$

O critério DIC (Spiegelhalter et al., 2002) é uma medida de ajustamento do modelo global, que é calculado com base em estimativas Bayesianas posteriores dos parâmetros do modelo.

O DIC é dado por:

$$DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D = 2\overline{D(\theta)} - D(\bar{\theta}) \quad (2.105)$$

Este critério geralmente é visto como a generalização Bayesiana do AIC e é facilmente obtido como um subproduto do algoritmo de amostragem MCMC. Além disso, também faz suposições mais fracas e penaliza automaticamente a complexidade do modelo (Zhang et al., 2019). Apesar das vantagens, o critério é sujeito a críticas, sendo a principal o facto depender do nível de especificação do parâmetro ao qual a verosimilhança do modelo está condicionada e, portanto, não ser invariante por reparametrização.

O modelo com menor valor de DIC é aquele com melhor equilíbrio entre o ajustamento do modelo e a sua complexidade (Clark & Gelfand, 2006).

Uma alternativa típica ao DIC é o critério WAIC, proposto por Watanabe (2010), que é totalmente Bayesiano na medida em que utiliza toda a distribuição *a posteriori*. O WAIC é dado por

$$WAIC = -2 \sum_{l=1}^N \ln(E_{\theta|y}(p(y_l | \theta))) + 2pW \quad (2.106)$$

com

$$p_W = \sum_{i=1}^n Var_{\theta|y}(\ln(p(y_i|\theta))) \quad (2.107)$$

Como o DIC, o WAIC também equilibra entre a qualidade do ajuste e a complexidade do modelo e menores valores indicam um modelo melhor.

2.7. Avaliação da adequação de modelos

Após qualquer processo de modelação, é importante verificar a adequação do modelo. Na modelação Bayesiana, os métodos mais usados baseiam-se na distribuição preditiva. Na validação cruzada, que se fundamenta na divisão da amostra em dois grupos, um grupo de treino é utilizado para construir uma distribuição *a posteriori*, enquanto o outro é usado como amostra de teste para obter uma distribuição preditiva condicionada à amostra anterior. A abordagem Jackknife (*leave-one-out*) é uma forma de validação cruzada que consiste em repetir a validação cruzada n vezes, deixando em cada etapa uma observação fora da subamostra de treino. Essa observação desempenha então o papel de validação da amostra de teste.

Outras medidas podem ser construídas comparando características da distribuição preditiva sob o modelo condicionado a dados observados com outros dados também observados.

Nestas abordagens usam-se duas medidas:

1. a ordenada preditiva condicional (CPO – *Conditional Predictive Ordinate*) dada por

$$CPO_i = p(y_i|y_{-i}) \quad (2.108)$$

2. a probabilidade de transformação uniformizante (PIT–*Probability Integral Transforms*) dada por

$$PIT_i = p(Y_i^* \leq y_i|y_{-i}) \quad (2.109)$$

Valores extraordinariamente grandes ou pequenos de PIT indicam possíveis valores discrepantes. Além disso, um histograma do valor do PIT, muito diferente da distribuição uniforme, indica que o modelo é questionável. Valores pequenos de CPO indicam também, informalmente, observações mal ajustadas. Nesse sentido, quanto maior a soma dos valores logarítmicos do CPO, também conhecido como *log pseudo-marginal* da verosimilhança, mais adequado é um modelo (AmaralTurkman et al., 2019).

$$\sum_{i=1}^n \ln(CPO_i) = \ln\left(\prod_{i=1}^n p(x_i | x_{-i})\right) \quad (2.110)$$

A adequação do modelo pode ainda ser testada através da validação preditiva posterior, em cujos métodos se usam todas as observações para ajustar e verificar o modelo. A validação preditiva posterior baseia-se em duas quantidades:

1. a distribuição preditiva posterior:

$$p(y_i^* | y) = \int p(y_i^* | \theta_i) p(\theta_i | y) d\theta_i \quad (2.111)$$

2. o *valor-p* preditivo posterior:

$$p(Y_i^* \leq y_i | y) \quad (2.112)$$

Valores demasiado pequenos de (2.111) indiciam observações que podem ser classificadas como *outliers*. Se isso acontecer para muitas observações, o modelo não é adequado para os dados.

Se os valores de (2.112) estiverem próximos de 0 (menores que 1%) ou 1 (maiores que 99%), há evidência de que o modelo não é adequado para ajustar os dados.

2.8. Testes à homogeneidade dos coeficientes dos modelos

No capítulo 3 são implementados testes à homogeneidade dos coeficientes dos modelos de regressão de Poisson, pelo que, são apresentados seguidamente os fundamentos teóricos de dois testes de hipóteses usados, o teste ANCOVA ONE-WAY e o teste de razão de verosimilhança generalizada (GLRT) (Gillibert, André, 2020; Kay & Kay, 2011).

Um dos pressupostos de um modelo ANCOVA é a igualdade de declives dos modelos lineares ajustados em dois grupos. Se, ao ajustar um modelo ANCOVA com interação *covariável* \times *grupo*, o coeficiente de interação for estatisticamente significativo, então a hipótese de homogeneidade dos declives deve ser rejeitada. Portanto, um modelo ANCOVA ONE-WAY permite rejeitar a homogeneidade dos declives dos modelos lineares ajustados a dois grupos. Análises gráficas e comparações dos declives são úteis para visualizar a natureza e a magnitude das diferenças nos declives entre os grupos e para compreender melhor como os grupos diferem em relação à covariável (Field et al., 2012).

Para testar a igualdade dos coeficientes de dois modelos lineares generalizados na escala natural, pode-se usar o teste de razão de verosimilhança generalizada que permite testar a igualdade dos coeficientes de dois modelos lineares generalizados.

Para testar a igualdade dos coeficientes de dois modelos lineares generalizados, definimos:

$$\begin{aligned} H_0: \beta^{(1)} &= \beta^{(2)} \\ H_1: \beta^{(1)} &\neq \beta^{(2)} \end{aligned} \quad (2.113)$$

onde $\beta^{(1)}$ e $\beta^{(2)}$ são os vetores dos coeficientes de dois modelos.

O GLRT é baseado na razão de verosimilhança, mas os parâmetros são desconhecidos, quer na hipótese nula quer na hipótese alternativa. A ideia chave no GLRT é substituir os parâmetros desconhecidos pelas suas estimativas de máxima verosimilhança (MLE) (Kay & Kay, 2011).

Em geral, a estatística do teste é definida como o GLRT decide a favor de H_1 se:

$$\lambda = -2 \ln L(X) = -2 [\ln p_{H_1}(X; \hat{\Theta}_1) - \ln p_{H_0}(X; \hat{\Theta}_0)] \quad (2.114)$$

onde $\hat{\Theta}_1$ é a estimativa MLE de Θ_1 assumindo H_1 é verdadeira e $\hat{\Theta}_0$ é a estimativa MLE de Θ_0 assumindo que H_0 é verdadeira.

Rejeita-se H_0 em favor de H_1 se λ for maior que um valor crítico determinado pela distribuição de λ sob H_0 .

O GLRT não possui propriedade de otimalidade em geral, mas aproxima-se assintoticamente do teste uniformemente mais poderoso entre todos os testes invariantes de parâmetros. Na prática tem um bom desempenho (Kay & Kay, 2011).

Apesar de muitas vezes a designação do teste GLRT ser abreviada para teste de razão de verosimilhança (LRT), na verdade estes dois testes não são equivalentes. A estatística GLRT é a razão da verosimilhança dos dados observados sob a hipótese nula e a estimativa de máxima verosimilhança sobre o espaço completo de parâmetros testados enquanto o LRT definido no lema de Neyman-Pearson é baseado na razão de verosimilhança para dois parâmetros totalmente especificados θ_0 e θ_1 . O GLRT pode ser usado para rejeitar uma hipótese quando todos os valores num espaço de parâmetros são possíveis, enquanto o LRT depende de um espaço de parâmetros restrito a dois valores.

Sintetizando, o teste LRT aplica-se a hipóteses simples, onde tanto H_0 quanto H_1 são completamente especificadas (parâmetros conhecidos) e o teste GLRT estende o LRT para hipóteses compostas, onde os parâmetros são desconhecidos e substituídos pelas suas estimativas de Máxima Verosimilhança.

O pacote `glmglrt` (Gillibert, André, 2020) disponibiliza a função `p-value()` que permite implementar o teste GLRT, bilateral e unilateral.

2.9. Extração de dados por medição em gráfico de barras com distorção visual

A informação disponibilizada pelas entidades públicas de saúde nem sempre é de fácil tratamento, sendo muitas vezes disponibilizada de forma visual, a partir de gráficos. Neste caso, há necessidade extrair dados quantitativos que permitam a implementação de análises estatísticas. Esta situação ocorre com dados de COVID-19 de Portugal Continental que foram usados nesta tese.

Para resolver este problema propõe-se uma metodologia de extração visual de dados, a partir da medição de pixels na figura com os dados pretendidos, correspondentes à medida a ser extraída e sua relação com o número de pixels correspondente a uma escala conhecida.

A metodologia é aplicada no capítulo 3 à extração da medida da altura de uma barra de um gráfico de barras, que corresponde à frequência absoluta da categoria em análise, numa determinada unidade. Pretende-se extrair dados de COVID-19 de Portugal Continental, disponibilizados pelas autoridades de saúde na forma de gráficos empilhados.

Descrição da metodologia

Seja C o valor que mede a altura de uma barra num gráfico de barras como na Figura 2.1a).

O valor C da Figura 2.1a), medido no eixo vertical do gráfico, é obtido pela relação entre o número de pixels na altura da barra, aqui representado por $|P2 - P3|$, e o número de pixels entre marcas de escala com valores conhecidos. Se a imagem gráfica de onde se pretende extrair C apresenta distorção visual (Figura 2.1b)), estas medidas de pixels terão um erro. O erro absoluto de desfocagem não varia com a distância na imagem; por isso, basta escolher a marca de maior escala (k neste caso) para minimizar o erro relativo. Consequentemente, o número de pixels para as marcas da escala será o número de pixels representado por $|P1 - P3|$.

Assim, o valor de C que estabelece a relação entre o valor da escala numa determinada unidade e a altura das barras, é dada por:

$$C = \frac{k \times |P2 - P3|}{|P1 - P3|} \quad (2.115)$$

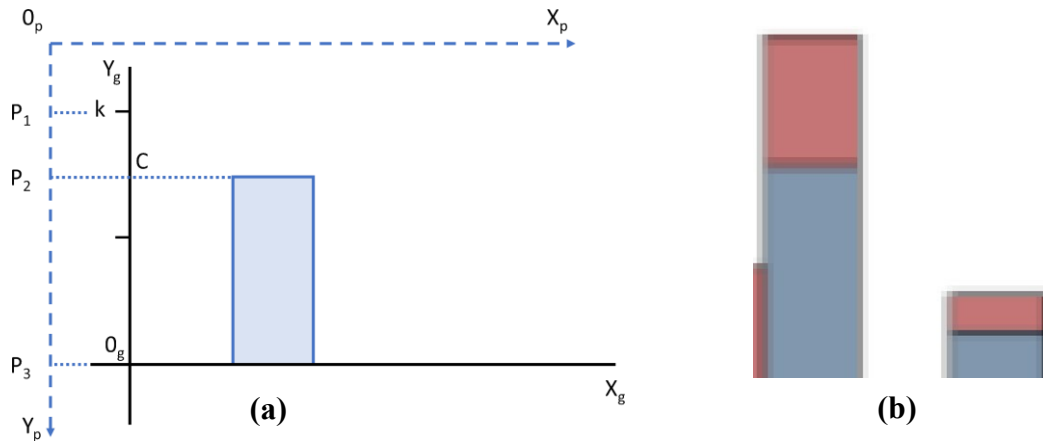


Figura 2.1: (a) Exemplo de medição de amostras; (b) Detalhe de gráfico de barras empilhadas com borrão visível.

Estimativa do erro de medida

Seja C uma quantidade final pretendida, que é a combinação de duas quantidades primárias A e B . Assuma-se que cada uma destas quantidades primárias foi medida várias vezes e o que o erro padrão da respetiva média foi determinado, σ_A e σ_B ; assumase ainda que as medições de ambas as grandezas primárias são independentes e, conseqüentemente, os erros cometidos nas mesmas são também independentes.

O erro padrão em C , σ_C , é dado pela seguinte fórmula de propagação do erro:

$$(\sigma_C)^2 = \left(\frac{\Delta C}{\Delta A} \sigma_A\right)^2 + \left(\frac{\Delta C}{\Delta B} \sigma_B\right)^2 \quad (2.116)$$

No caso de C ser um quociente, e neste caso $C = k \frac{B}{A}$, a equação (2.116) é equivalente a:

$$\left(\frac{\sigma_C}{C}\right)^2 = \left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 \quad (2.117)$$

Quando é feita apenas uma medição em cada quantidade A e B , o erro máximo para σ_A e σ_B deve ser utilizado para estimar o resultado da medição.

Quando se pretende obter um intervalo para o erro em C , é necessário determinar os valores máximo e mínimo de σ_C .

Resolvendo a equação (2.117) para σ_C , obtém-se o valor máximo:

$$\sigma_C = C \sqrt{\frac{A^2 \sigma_B^2 + B^2 \sigma_A^2}{(AB)^2}} \quad (2.118)$$

Substituindo C em (2.118), obtém-se uma expressão equivalente

$$\sigma_C = \frac{k}{A^2} \sqrt{A^2 \sigma_B^2 + B^2 \sigma_A^2} \quad (2.119)$$

2.10. Desagregação de Séries Temporais

A desagregação temporal e/ou espacial de séries de dados é fundamental para o estudo de dados históricos de epidemiologia, de que esta tese se ocupa. No capítulo 4 aplicam-se dois algoritmos de desagregação temporal às séries dos dados de COVID-19 dos concelhos do distrito do Porto e comparam-se os resultados obtidos com os dados observados.

Os dados sobre eventos de saúde variam muito em função dos objetivos, desde ensaios clínicos específicos e estudos observacionais localizados, até aos esforços nacionais (e internacionais) de vigilância de doenças. Muitas análises espaço-temporais de dados de saúde pública utilizam dados recolhidos e resumidos por agências governamentais, cuja recolha é, frequentemente, da responsabilidade de diversas entidades locais, e que são depois disponibilizados para uso público em diferentes formatos, nem sempre da melhor forma, mesmo para investigadores. Dados agregados com diferentes formatos, métricas ou granularidades exigem transformações nos mesmos com vista ao seu tratamento e análise, em função de novos objetivos. A desagregação temporal e/ou espacial inclui-se nessas transformações, com o objetivo de homogeneizar a granularidade de séries temporais para as fundir com outras ou para implementar análises com dados de granularidades temporais ou espaciais mais finas.

A desagregação de dados refere-se à separação de informações compiladas em unidades menores, para elucidar tendências e padrões subjacentes. Como referido atrás, os dados compilados podem ser provenientes de várias fontes (de setores públicos/privados e organizações nacionais/internacionais) e ter múltiplas variáveis ou “dimensões”: idade, sexo, área geográfica, educação, etnia, outras variáveis socioeconómicas ou ainda tempo.

Os dados históricos geralmente incluem relatórios sobre séries temporais agregadas em diferentes níveis, com possíveis sobreposições e dados em falta, pelo que, desagregar uma sequência temporal de eventos de vários relatórios agregados é um grande desafio na fusão de dados históricos. De facto, a soma dos relatórios diários pode não ser a mesma que a contagem de um relatório semanal, pelo que é possível a existência de relatórios conflitantes; as fontes de dados históricos podem ter diferentes níveis de confiabilidade por diferentes motivos (problemas com as fontes primárias de informação, metodologia de recolha de dados diferentes de acordo com diferentes entidades, etc.), o que pode resultar em relatórios com erros. A reconstrução adequada de eventos históricos geralmente requer desagregação razoavelmente precisa, mesmo que a partir de dados de relatórios imprecisos.

A desagregação pode ocorrer em diversas dimensões dependendo do objetivo e das características dos dados disponíveis. A desagregação por sexo, idade, por área geográfica ou tempo pode ser necessária para um estudo mais preciso de populações específicas ou para a fusão de dados de diferentes séries ou ainda porque ao longo do tempo as métricas de agregação e/ou as granularidades de séries da mesma variável foram mudando.

2.10.1. Agregação de séries espaço-temporais

A tomada de decisão em problemas do mundo real é um processo de gestão, caracterizado por tarefas complexas e difíceis. A tomada de decisão recorrendo a múltiplos critérios – objetivos ou atributos, tem sido uma das áreas de conhecimento de mais rápido crescimento nas ciências da decisão e tem sido amplamente utilizada em muitas disciplinas (Mohd & Abdullah, 2017). As variáveis de decisão podem tomar valores num domínio contínuo ou discreto. Em muitos domínios surge o problema de agregar uma coleção de leituras numéricas da(s) variável(eis) de decisão para obter um valor típico que pode ser uma média, ponderada ou não, pode ser a soma ou um operador de um conjunto de outros operadores/funções de agregação. Grabisch et al. (2009) definem agregação como o processo de combinar e mesclar vários valores (na maioria das vezes numéricos) num único valor obtido pela função de agregação e que deve representar ou sintetizar “*em algum sentido*” todas as entradas individuais, sentido esse que depende altamente do contexto.

Assim, um operador de agregação, matematicamente, é uma função que a cada n -upla (x_1, x_2, \dots, x_n) de valores a agregar associa um número real.

Formalmente, um operador de agregação é uma função

$$f: \bigcup_{n \in \mathbb{N}} (A^n) \rightarrow B \quad (2.120)$$

sendo A o conjunto suporte de valores a agregar, finito ou infinito, e B o conjunto de resultados possíveis de agregação. n é o número de argumentos de f .

A função f verifica as propriedades seguintes:

- i) É não decrescente em cada variável;
- ii) Verifica as condições fronteira:

$$\inf_{x \in A^n} f(x) = \inf A \quad e \quad \sup_{x \in A^n} f(x) = \sup A \quad (2.121)$$

Definir ou escolher a classe certa de funções de agregação para um problema específico é uma tarefa difícil, considerando a enorme variedade de potenciais funções de agregação. Detyniecki et al. (2000) e Grabisch et al. (2009) proporcionam uma análise matemática

sólida sobre diversos tipos de funções de agregação adequadas para diversos contextos e respetivas propriedades, capaz de responder a questões de ordem matemática e de ordem prática.

A agregação pode ser implementada em diversos domínios: num domínio temporal, espacial, num domínio de codificação de um atributo, num intervalo de números reais.

Na agregação temporal, a série agregada de baixa frequência é obtida através da soma de quaisquer m períodos da série de alta frequência, sendo m o nível de agregação. A agregação temporal pode ser sobreposta ou não, consoante há, ou não, períodos sobrepostos.

A agregação de dados tem vantagens e desvantagens. Por um lado, com a agregação espera-se que a variação seja menor, que haja menor ruído e que se identifiquem as características das séries subjacentes, como tendência e sazonalidade (Masselot et al., 2018). A desvantagem mais óbvia da agregação é a perda de informação. No entanto, há outro tipo de problemas que podem surgir com a agregação de dados, como a crença incorreta de que amostras pequenas provavelmente serão altamente representativas das populações das quais são extraídas, como acontece com grandes amostras – Lei dos Pequenos Números; as análises podem tornar-se vulneráveis à “falácia ecológica”: fazer inferências falsas sobre o comportamento individual com base em dados de nível populacional (“ecológicos”); ou ainda, pode ocorrer o paradoxo de Simpson, onde as relações vistas no nível individual podem ser completamente revertidas quando analisadas no nível agregado. Pollet et al. (2015) não só explicam conceptualmente como e porque podem ocorrer estas duas últimas situações, como as ilustram. Nos dados agregados em unidades geográficas deve-se considerar o problema de unidades de área modificáveis (MAUP) – a possibilidade de modificar as unidades de área alterando os seus limites ou agregando/desagregando unidades existentes produz resultados de análise inconsistentes. Este problema está intimamente relacionado com a falácia ecológica (Arbia, 1989). Um outro problema que pode surgir nas análises espaciais é o dos dados desalinhados (MIDP - *Misaligned Data Problem*). Este problema ocorre quando os dados espaciais são analisados numa escala diferente daquela em que foram registados (Banerjee et al., 2014), ou porque se pretende analisar apenas a distribuição espacial de uma variável num novo nível de agregação ou porque se pretende relacionar variáveis que estão em níveis de agregação espacial diferentes. Nos dados agregados em unidades temporais pode surgir o problema MPUP (*Modifiable Temporal Unit Problem*) – a escolha de diferentes unidades temporais (unidade de agregação temporal,

segmentação do domínio temporal) ou dos limites do domínio temporal pode originar resultados inconsistentes (Cheng & Adepeju, 2014; Deckard & Schnell, 2023).

O nível de agregação espacial e/ou temporal, portanto, é determinante na análise da distribuição geográfica de qualquer tipo de fenómeno, em particular em epidemiologia, na medida em que a agregação dos dados de ocorrência de uma doença afeta os padrões da sua distribuição geográfica, bem como a análise de potenciais fatores que poderão afetar o seu desenvolvimento. Na área da epidemiologia humana é considerada a agregação espacial e/ou temporal para preservar a confidencialidade do indivíduo. Quando diminui a granularidade da agregação, aumenta a robustez estatística dos resultados, mas diminui a variação geográfica dos padrões da doença, o que, por exemplo, diminui a possibilidade de identificação de *clusters* (Alarcon Falconi et al., 2020; Cheng & Adepeju, 2014; Jeffery et al., 2014; Ozonoff et al., 2007).

2.10.2. Desagregação (temporal) de séries temporais

A maioria das fontes governamentais detém dados de saúde que apenas torna acessíveis ao público se agregados, devido a requisitos de confidencialidade. Estes dados, geralmente disponíveis a nível municipal ou código postal/região de recenseamento, podem dar importantes perspetivas sobre a variação espacial geral de uma doença em termos de incidência ou prevalência.

A necessidade de fundir diferentes séries de dados que por vezes foram agregadas com diferentes granularidades, torna, muitas vezes, necessário proceder à desagregação das séries em granularidades mais finas (Benedetti et al., 2021a; Raim et al., 2019). Por outro lado, em diferentes áreas e também na área da saúde, as séries de dados disponibilizados por organismos públicos ao longo do tempo nem sempre mantêm as métricas de agregação. O Instituto Nacional de Estatística (INE) Português, ao longo dos anos foi aumentando o nível de granularidade dos dados de georreferenciação das unidades amostrais (Pereira, 2018); a Direção Geral de Saúde (DGS) Portuguesa desde o início da pandemia COVID-19 foi alterando a métrica temporal dos dados disponibilizados sobre infeções com o vírus SARS-COV-2.

A desagregação pode ocorrer por atributo (sexo, nível de rendimento,...), por dimensão temporal (dia, semana, mês,...) ou por dimensão espacial e pode existir ou não sobreposição de unidades agregadas, o que influencia os algoritmos/métodos de desagregação.

Na literatura, a aplicação de métodos de desagregação acontece essencialmente em:

a) séries de dados económicos - Pavía-Miralle (2010) apresenta uma revisão onde descreve uma grande quantidade de métodos, procedimentos e algoritmos propostos na literatura, essencialmente para a desagregação temporal;

b) em séries de dados meteorológicos - Breinl & Di Baldassarre (2019) propõem um algoritmo que implementa um método não paramétrico robusto para a desagregação espaço-temporal de pluviosidade e temperatura, além de apresentarem uma extensa lista de referências com modelos de desagregação de dados desta natureza.

Em epidemiologia a desagregação temporal é pouco considerada na literatura, predominando as referências à desagregação espacial (Wayant, 2019). Lee et al. (2022) propõem um modelo para desagregação espaço-temporal de dados de contagem de febre Q, agrupados no espaço e no tempo, baseando a abordagem no modelo de *link* composto penalizado (Eilers, 2007), combinado com a metodologia de *P-splines*, para obter estimativas suavizadas numa resolução mais fina a partir de dados agregados no espaço e/ou no tempo. Esta abordagem permite incluir informações populacionais em escala fina e efeitos aleatórios específicos ou estrutura de correlação adicional, se necessário.

2.10.3. Modelos de desagregação temporal

O problema básico da desagregação temporal em séries univariadas pode ser enunciado como se segue (Fonzo, 2003; Quilis, 2018).

Seja Y uma série temporal de baixa frequência observada.

$$Y = (Y_1, Y_2, \dots, Y_t, \dots, Y_N)' \quad (2.122)$$

onde Y é um vetor de dimensão $N \times 1$.

Suponha-se que se pretende estimar uma série temporal de alta frequência não observada y em que:

$$y = (y_{1,1}, y_{1,2}, \dots, y_{1,s}, y_{2,1}, \dots, y_{2,s}, \dots, y_{N,1}, \dots, y_{N,s})' \quad (2.123)$$

é o vetor de dimensão $nN \times 1$ de dados de alta frequência sendo $n = sN$. Neste caso tem-se um problema dito de distribuição temporal ou interpolação.

Assume-se que existe uma restrição (invariante no tempo) que liga Y e y que pode ser definida na forma matricial por:

$$Y = Cy \quad (2.124)$$

em que C é a matriz de agregação-extrapolação temporal que liga os valores observados de baixa frequência com os valores desconhecidos de alta frequência, definida por:

$$C = I_N \otimes c | O_{N, n-sN} \quad (2.125)$$

- N é o número de observações de baixa frequência,
- \otimes representa o produto de Kronecker,
- c é um vetor linha, de tamanho s , que define o tipo de agregação temporal (pesos),
- s é o número de pontos de dados de alta frequência para cada ponto de dados de baixa frequência (taxa de conversão de frequência),
- $O_{N, n-sN}$ é a matriz bloco de zeros de dimensão $N \times (n - sN)$ que corresponde às $n - sN$ últimas linhas de C com dados identificadores dos subperíodos de alta frequência não sujeitos a restrições de agregação temporal, por não haver informação correspondente nos dados de baixa frequência, no caso da extrapolação.

A relação (2.124) pode ser escrita na forma não matricial:

$$y_t = \sum_{u=1}^s c_u y_{t,u}, \quad t = 1, 2, \dots, N \quad (2.126)$$

Os valores do vetor c , $\{c_u\}_{u=1}^s$, são pesos conhecidos *a priori* e definem o tipo de restrição, invariante no tempo, que liga os dados de baixa frequência com os dados de alta frequência:

(i) Se $c = [1, 1, \dots, 1]$, então Y corresponde a dados agregados de alta frequência – desagregação por distribuição. Este caso consiste em encontrar os valores dos subperíodos de cada período da série de baixa frequência quando o total de cada período é conhecido. Neste caso, há restrição nas estimativas em relação ao valor do período: o total deve ser igual ao valor agregado.

(ii) Se $c = [\frac{1}{s}, \frac{1}{s}, \dots, \frac{1}{s}]$, então Y corresponde a médias dos dados de alta frequência – desagregação por distribuição. Este caso tem o mesmo objetivo do caso anterior, mas na restrição das estimativas a média deve ser igual ao valor agregado.

(iii) Se $c = [0, 0, \dots, 1]$, então Y corresponde a dados obtidos por amostragem sistemática dos dados de alta frequência e então os pesos são todos nulos com exceção do último período observado (ou o primeiro e então $c = [1, 0, \dots, 0]$). Este caso corresponde à desagregação por interpolação e consiste em estimar valores para os subperíodos componentes de cada período da série de baixa frequência. Não é imposta nenhuma restrição às estimativas.

No caso em que $n > sT$ o problema consiste na extrapolação de dados.

As matrizes C para os três casos são respetivamente definidas por:

- (i) No problema de distribuição temporal:

$$C_{(N \times sN)} = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \end{bmatrix} \quad (2.127)$$

(ii) No problema de interpolação, com valores do final do período de agregação:

$$C_{(N \times sN)} = \begin{bmatrix} 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 1 \end{bmatrix} \quad (2.128)$$

(iii) Quando nos dois casos anteriores se tem ainda um problema de extrapolação, no final de cada uma das matrizes há um bloco de dimensão $N \times (N - sN)$.

A distribuição temporal corresponde à desagregação temporal (Fonzo, 2003).

As diferentes abordagens distinguem-se ainda pela utilização ou não de séries "relacionadas", conhecidas para todos os subperíodos e que, de alguma forma, refletem os movimentos da série a desagregar nos subperíodos desconhecidos (Alba, 1988).

As técnicas de desagregação temporal clássicas e mais comumente usadas pelos profissionais que usam os dados para produzir estatísticas oficiais focam-se, em linhas gerais, em duas vertentes:

- I. métodos que dependem de critérios puramente matemáticos e/ou modelos de séries temporais para derivar trajetórias suaves para as séries não observadas considerando a representação ARIMA da série a ser desagregada.
- II. métodos que utilizam as informações obtidas com variáveis auxiliares relacionadas (indicadores) e que são observadas na (maior) frequência desejada.

2.10.4. Métodos matemáticos de desagregação temporal

Seguem-se alguns exemplos ilustrativos do primeiro caso (caso I).

a) Método de Boot, Feibes e Lisman (BFL)

O método proposto por Boot et al. (1967) e generalizado por Cohen et al. (1971) é um método puramente matemático que é baseado nas soluções de um problema de minimização de diferenças de primeira e segunda ordem, com uma restrição.

As expressões de Lagrange para as primeiras e segundas diferenças são, respetivamente,

$$L^{(1)} = \sum_{t=2}^{Ns} (y_t - y_{t-1})^2 - \sum_{T=1}^N \lambda_T \left(\sum_{t=s(T-1)+1}^{sT} y_t - Y_T \right) \quad (2.129)$$

$$L^{(2)} = \sum_{t=3}^{Ns} (\Delta y_t - \Delta y_{t-1})^2 - \sum_{T=1}^N \lambda_T \left(\sum_{t=s(T-1)+1}^{sT} y_t - Y_T \right) \quad (2.130)$$

em que $\Delta y_t = y_t - y_{t-1}$ e $Y_T = \sum_{t=s(T-1)+1}^{sT} y_t$, com $T = 1, 2, \dots, N$

b) Método de Wei e Stram

A abordagem proposta por Wei & Stram (1990) que se descreve de seguida, combina métodos matemáticos com a representação ARIMA da série agregada.

Começa-se por ajustar um modelo de média móvel integrada autorregressivo $ARIMA(p, d, r)$ que é adequado a dados agregados. Neste modelo p é a ordem do polinómio autoregressivo (AR), d a ordem de integração e r a ordem do polinómio de média móvel (MA). O modelo assume a forma seguinte (Hodgess, Erin M. & Mhoon, Kendra, 2019):

$$\alpha_p(B)(1 - B)^d Y_T = \beta_r(B) C_T \quad (2.131)$$

em que:

- $\alpha_p(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p$,
- $\beta_r(B) = 1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_r B^r$,
- B é o operador de retrocesso $B^j Y^T = Y_{T-j}$ e
- C_T é um ruído branco gaussiano $C_t \sim N(0, \sigma_C^2)$.

Os polinómios AR e AM não têm raízes comuns, e satisfazem os critérios de estacionaridade e invertibilidade.

Para remover a diferença calcula-se:

$$U_T = (1 - B)^d Y_T \quad (2.132)$$

A matriz de autocovariância de U , V_U , é construída com os valores ajustados de $\alpha_p(B)$, $\beta_r(B)$ e C_T .

A ordem do polinómio AR e a ordem de integração mantêm-se, isto é, p e d , respetivamente. A ordem máxima do polinómio AM é:

$$q = p + d + 1 \quad (2.133)$$

O modelo de desagregação da série fica:

$$\phi_p(B)(1 - B)^d Y_t = \theta_q a_t \quad (2.134)$$

sendo $\phi_i = \alpha^{\frac{1}{m}}$ e m a ordem de integração.

Seja:

$$W_t = (1 - B)Y_t \quad (2.135)$$

Os elementos da matriz de autocovariância de U , V_U , são usados para calcular os elementos da matriz de autocovariância de W , γ_W , que permite calcular os coeficientes do polinómio MA desagregado por métodos numéricos.

A série diferenciada desagregada será

$$W = (\hat{V}_W(C^d)'V_W^d)U \quad (2.136)$$

sendo $U = C^dW$

A matriz C^d de ordem $(n - d) \times (ns - d)$ é construída como

$$C_d = \begin{bmatrix} c & 0 & \mathbf{0} \\ 0 & c & 0 \\ \mathbf{0} & 0 & c \end{bmatrix} \quad (2.137)$$

sendo c os coeficientes do polinómio $(1 + B + B^2 + \dots + B^{s-1})^{d+1}$.

A última etapa consiste em produzir a estimativa da série desagregada, sem diferenciar:

$$\hat{y} = \begin{bmatrix} & \Delta_{ns}^d & \begin{bmatrix} \square \\ \square \\ \square \end{bmatrix} \\ 0 & & I_d \otimes J_s \end{bmatrix}^{-1} \begin{bmatrix} \hat{V}_W(C^d)'V_U^{-1}\Delta_n^d & \begin{bmatrix} \square \\ \square \\ \square \end{bmatrix} \\ 0 & I_d \end{bmatrix} \quad (2.138)$$

Δ_n^d é uma matriz diagonal $(k - d) \times k$ cujos elementos da diagonal principal são os coeficientes do polinómio $(B - 1)^d$.

Os métodos de Boot são casos particulares deste método, com ordens de diferenciação 1 e 2 e com matriz identidade para matriz de autocovariância desagregada (matriz de U).

Seguem-se exemplos de métodos que se enquadram no caso II, isto é, que utilizam as informações obtidas com variáveis auxiliares (relacionadas ou não) e que são observadas na (maior) frequência desejada.

c) Método de Denton

A abordagem proposta por Denton (1971) não se baseia num modelo, mas num problema de otimização quadrática com restrição. No entanto, considera a estimação preliminar de uma série dita de valores originais que não depende necessariamente de uma estrutura estocástica ou de séries observadas relacionadas. O método é implementado em duas fases: uma primeira fase em que se obtém uma série preliminar desagregada que não cumpre a restrição temporal; uma segunda etapa em que se ajusta a série de alta frequência obtida aos totais da série de baixa frequência.

Seja x a série de um indicador medido em alta frequência, não necessariamente relacionada com a série agregada Y que se pretende desagregar. Seja y a série final de alta frequência estimada por distribuição (desagregação). Seja s o número de pontos de dados de alta frequência para cada ponto de dados de baixa frequência. A série de baixa frequência tem N observações e a série de alta frequência tem $n = sN$ elementos.

O método de ajustamento original de Denton baseia-se no princípio da preservação do movimento, segundo o qual a série de alta frequência deve preservar o movimento da série do indicador, pois o movimento na série do indicador é a única informação disponível.

Para tal, especifica-se uma função de penalização, $P(x, y)$, onde $y = (y_1 \dots, y_s)'$ e $x = (x_1, \dots, x_s)'$ são vetores coluna $s \times 1$ de estimativas de alta frequência para cada período de baixa frequência e valores originais (indicador), respetivamente. Pretende-se obter estimativas y de modo a minimizar a função de penalização $P(x, y)$, sujeita às restrições de agregação temporal,

$$Y_T = \sum_{t=s(T-1)+1}^{sT} y_t, T = 1, 2, \dots, N \quad (2.139)$$

Considerando a classe de funções de penalização representada por:

$$(y - x)'A(y - x) \quad (2.140)$$

que é uma forma quadrática nas diferenças entre os valores originais e os valores da série temporal ajustados, sendo A uma matriz simétrica $n \times n$, não singular, a ser especificada posteriormente.

As expressões de Lagrange são dadas por,

$$L = (y - x)'A(y - x) - 2\lambda'(Y - Cy) \quad (2.141)$$

em que $\lambda = [\lambda_1 \lambda_2 \dots \lambda_N]'$ e C é a matriz (2.127).

A solução de minimização da penalização $p(x, y)$ é obtida pelas soluções do sistema formado pelas equações definidas pelos zeros das derivadas parciais de L em relação aos elementos de y e λ .

A solução do sistema é

$$y = x + Br \quad (2.142)$$

sendo

$$B = A^{-1}C'(CA^{-1}C')^{-1} \quad (2.143)$$

e

$$r = Y - Cx \quad (2.144)$$

o vetor das discrepâncias entre as duas séries de baixa frequência.

B representa a forma como as discrepâncias em cada período de baixa frequência são distribuídas por cada período de alta frequência. Se $A = I_{n \times n}$ então as discrepâncias de cada período de baixa frequência são distribuídas uniformemente por cada período de alta frequência, embora nem sempre esta forma de distribuição se verifique nas aplicações concretas. Denton propôs outras variantes para a função de penalização, baseadas com diferenças de primeira ordem ou maior, destacando-se as variantes com diferenças de primeira ordem e de segunda ordem aditiva e proporcional, para contornar este problema, a saber:

$$p(x, y) = \sum_{t=1}^s (\Delta(y_t - x_t))^2 \quad (2.145)$$

$$p(x, y) = \sum_{t=1}^s \left(\frac{y_t}{x_t} - \frac{y_{t-1}}{x_{t-1}} \right)^2 \quad (2.146)$$

$$p(x, y) = \sum_{t=1}^s (\Delta^2(y_t - x_t))^2 \quad (2.147)$$

$$p(x, y) = \sum_{t=1}^s \left(\Delta \left(\frac{y_t}{x_t} - \frac{y_{t-1}}{x_{t-1}} \right) \right)^2 \quad (2.148)$$

2.10.5. Métodos de regressão na desagregação de séries

No método proposto por Chow e Lin (1971) e desenvolvido posteriormente por Bournay e Laroque (1979), Fernández (1981) e Litterman (1983), considera-se uma relação linear da série a desagregar com uma ou mais séries relacionadas, observadas na alta frequência pretendida para a desagregação da série inicial.

Nos métodos de regressão, x representa uma matriz $n \times m$, onde m indica o número de indicadores (incluindo uma possível constante). O pressuposto destes métodos é que a relação linear entre a série de baixa frequência Y e Cx também se verifica entre as séries de alta frequência x e y .

A série preliminar é ajustada por regressão dos Mínimos Quadrados Generalizados (*Generalized Least Squares* – GLS):

$$p = \beta x \quad (2.149)$$

Para uma dada matriz de variância-covariância, Σ , o estimador de regressão GLS, $\hat{\beta}$, é calculado da seguinte forma padrão:

$$\hat{\beta}(\Sigma) = [x' C' (C \Sigma C')^{-1} C x]^{-1} x' C' (C \Sigma C')^{-1} Y \quad (2.150)$$

O algoritmo Chow-Lin assume que os resíduos da série de alta frequência seguem um processo autorregressivo da ordem 1 (AR1), isto é:

$$r_t = \rho r_{t-1} + \epsilon_t \quad (2.151)$$

onde ϵ é ruído branco com valor médio 0 e desvio padrão σ_ϵ e $|\rho| < 1$.

Nestas condições a matriz de covariâncias é definida por:

$$\Sigma = \frac{\sigma_\epsilon}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \ddots & \ddots & \dots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix} \quad (2.152)$$

No cálculo de Σ , σ_ϵ simplifica e ρ pode ser estimado recorrendo ao método iterativo proposto por Chow & Lin (1971). Bournay & Laroque (1979) propõem a maximização da verosimilhança da regressão GLS:

$$L(\rho, \sigma_\epsilon^2, \beta) = \frac{\exp\left(-\frac{1}{2} r' (C \Sigma C')^{-1} r\right)}{(2\pi)^{\frac{n}{2}} (\det(C \Sigma C'))^{\frac{1}{2}}} \quad (2.153)$$

onde r está definido em (2.144) e β é substituído por (2.150). O estimador de máxima verosimilhança do parâmetro autorregressivo $\hat{\rho}$, assim obtido, é um estimador para o verdadeiro valor consistente em probabilidade (consistência simples) (Fuller, 2009). Esta propriedade para a maioria das aplicações deste método de desagregação é suficiente para garantir que o estimador fornecerá resultados confiáveis com amostras grandes. Este estimador, por vezes, pode tomar um valor negativo e por isso se impõe a restrição de que seja positivo.

Além disso, Al-Osh (1989), Wei e Stram (1990), Guerrero (1990) e Guerrero e Martinez (1995) combinam uma abordagem baseada no modelo ARIMA com a utilização de séries relacionadas de alta frequência num modelo de regressão para superar algumas arbitrariedades na escolha da estrutura estocástica dos erros de alta frequência.

Fernandez (1981) propõe um modelo de passeio aleatório para os desvios que evita a estimação de parâmetros no nível de alta frequência. Litterman (1983) refina a solução de Fernandez introduzindo um processo de Markov para ter em conta a correlação serial nos

resíduos. Wei e Stram (1990) englobam as três soluções, generalizando a restrição na classe de processos ARIMA (*AutoRegressive Integrated Moving Average*).

Algoritmo 1: CHOW-LIN (REGRESSÃO)

INPUT:

- $Y_{N \times 1}$, vetor de dados de baixa frequência;
- $x_{n \times m}$, matriz de indicadores de alta frequência;
- ta , tipo de desagregação ($ta = 1, 2, 3, 4$ (1. *Soma()*, 2. *Média()*, 3. *Primeiro elemento* (interpolação), 4. *Último elemento* (interpolação));
- s , número de pontos de dados de alta frequência para cada ponto de dados de baixa frequência;
- **Tipo** = 1, 2, tipo de estimação dos coeficientes (1. *Mínimos Quadrados Ponderados ou Generalizados*; 2. *Máxima Verossimilhança*).

OUTPUT:

- $\hat{\beta}$, coeficientes de regressão estimados obtidos com o método tipo 1.
 - $\hat{\rho}$, estimativa do parâmetro autorregressivo ρ obtido com o método tipo 2.
 - \hat{y} série desagregada de alta frequência
-

Algoritmo 1: CHOW-LIN (REGRESSÃO) - IMPLEMENTAÇÃO

Iniciar: Definir um valor inicial $\hat{\rho}^{(0)}$

Construir C de acordo com ta e s .

Iteração: Para $k = 1$ até **Convergência**

- **Construir** Σ com $\hat{\rho}^{(k-1)}$
- **Estimar** $\hat{\beta}^{(k)} = [x' C' (C \Sigma C')^{-1} C x]^{-1} x' C' (C \Sigma C')^{-1} Y$
- **Calcular** $r = Y - C X \hat{\beta}^{(k)}$
- **Atualizar** $\hat{\rho}^{(k)}$
 - Maximizar a função de verossimilhança em relação a ρ usando os resíduos r .
 - Garantir que $|\rho^{(k)}| < 1$
- **Verificar convergência**
Se $|\hat{\rho}^{(k)} - \hat{\rho}^{(k-1)}| < \epsilon$ e $\|\hat{\beta}^{(k)} - \hat{\beta}^{(k-1)}\| < \epsilon$ parar

Obter $\hat{y} = x \hat{\beta} + \hat{u}$ com $\hat{u} = \Sigma C' (C \Sigma C')^{-1} (Y - C x \hat{\beta})$

Nestes dois últimos casos a matriz de variância-covariância toma a forma seguinte:

$$\Sigma(\rho) = \sigma_{\epsilon}^2(\Delta'H(\rho)'H(\rho)\Delta)^{-1} \quad (2.154)$$

Alguns autores propuseram técnicas baseadas em modelos de regressão dinâmica na identificação da relação entre as séries a serem estimadas e os indicadores relacionados. Refiram-se os trabalhos de Bisio & Moauro (2018), Poissonnier (2017) e Santos Silva & Cardoso (2001).

2.10.6. Outros métodos

São propostas outras abordagens na literatura, nomeadamente:

i) métodos que se baseiam em formulações em termos de modelos de componentes não observadas/séries temporais estruturais (em níveis ou transformadas) e o filtro de Kalman, para obter estimativas ótimas de observações ausentes por um algoritmo de suavização (Gómez et al., 2018; Gudmundsson, 1999; Hotta & Vasconcellos, 1999; Proietti, 2013). Moauro & Savio (2005) propõem uma metodologia multivariada para desagregação temporal de séries temporais que usa o modelo de equações de séries temporais aparentemente não relacionadas e é estimado pelo filtro de Kalman.

ii) técnicas que usam modelos de regressão dinâmica (e possivelmente dados transformados) na identificação da relação entre a série a ser estimada e o conjunto de séries temporais relacionadas (Bisio & Moauro, 2018; Fonzo, 2003; Proietti, 2006).

iii) técnicas usando outro tipo de abordagem: Bayesianas (Alba, 1988; Benedetti et al., 2021b); Redes Neurais Artificiais ou Convolucionais (Monteiro et al., 2019; Zaier & Abed, 2014); Análise de Wavelets (Perricone, 2018).

Bisio & Moauro (2018) descrevem e discutem os mais recentes desenvolvimentos na desagregação de séries temporais no contexto económico, nomeadamente os modelos de atraso distribuído autorregressivo (Proietti, 2006), parametrizados na forma de espaço de estados e com recurso ao filtro de Kalman para o tratamento estatístico. Comparam o desempenho dos modelos com o desempenho dos métodos tradicionais e concluem que estes mantêm superioridade em relação ao ajustamento dos resultados desagregados ao indicador da amostra.

iv) O método do *Spline* cúbico para a desagregação temporal é método matemático de desagregação temporal que Quenneville et al. (2013) mostraram ser um caso particular do método de Denton. O método tem como objetivo ajustar uma série de funções polinomiais de terceiro grau entre observações anuais e justapô-las de tal forma que a função resultante

tomada como um todo, a função *spline*, tenha primeira e segunda derivadas. Esta função é usada depois para obter estimativas de valores numa escala mais fina.

As *splines* cúbicas podem ser usadas quando não há indicadores de subperíodo para guiar os movimentos de curto prazo dos dados. No entanto, também podem ser adaptadas para os casos em que tais indicadores estão disponíveis e neste caso, a *spline* é adaptada ao indicador.

Quando nenhuma série de indicadores relacionada está disponível, a solução para a interpolação *spline* é o limite contínuo da variante aditiva da primeira diferença; quando uma série de indicadores relacionada é usada, a solução para a interpolação *spline* é o limite contínuo de uma forma ponderada da variante PFD (*Preservation of Frequency Domain*) com os pesos iguais ao inverso da série indicadora (Quenneville et al., 2013). Na desagregação de séries temporais, por algoritmos que usam séries temporais de indicadores, preservar o domínio de frequência é fundamental para garantir que a estrutura de frequência da série original seja mantida na série desagregada. Isto é importante porque diferentes componentes de frequência podem conter informações significativas sobre padrões sazonais, tendências e ciclos presentes nos dados.

2.10.7. Desagregação de séries temporais com possibilidade de períodos sobrepostos

Os dados agregados podem oferecer diversos desafios que dificultam a desagregação, nomeadamente, a sobreposição de períodos, períodos sem cobertura ou períodos com dados conflitantes (Figura 2.2).

A desagregação com períodos sobrepostos tem sido objeto de investigação. Também neste caso se verifica que é nas séries com dados económicos que se encontram mais referências na literatura. Algumas abordagens são aplicadas a problemas na área da saúde (Silva et al., 2011).

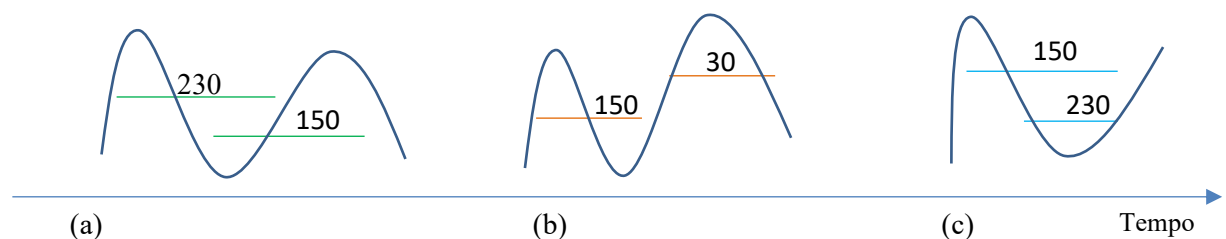


Figura 2.2: (a) Períodos temporais consecutivos com sobreposição; (b) Períodos temporais consecutivos com lacuna; (c) Períodos temporais consecutivos com conflito (o período menor tem um número maior de dados que o período maior).

Labonne et al. (2018) e Labonne & Weale (2020) desenvolvem uma abordagem de espaço de estados para filtrar e desagregar temporalmente valores do IVA com ruído e que exibem componentes dinâmicas não observadas. Desenvolveram métodos multivariados e não lineares para usar séries de indicadores e dados em logaritmos, respetivamente. Discutiram o efeito do ruído em valores agregados na estimativa de componentes desagregadas do modelo e ilustraram uma nova estratégia de agregação temporal adequada para dados sobrepostos.

Hotta et al. (1992) estudaram o efeito da agregação com sobreposição, em particular, da utilização de filtros de médias-móveis, em modelos de séries temporais para identificação, estimação e previsão, e nas componentes de tendência e sazonalidade, bem como, na identificação de pontos de reversão e em relações dinâmicas entre variáveis e concluíram que o comportamento da série é alterado, devendo ser utilizado com os cuidados necessários.

Monteiro et al. (2019) apresentam uma técnica de desagregação espacial, combinando procedimentos de análise de regressão com métodos clássicos de análise espacial, como o mapeamento dasimétrico ou interpolação piconfilética e apresentam uma discussão sobre a metodologia de desagregação espacial usada, incluindo desagregação com sobreposição espacial. Apresentam ainda uma discussão sobre a qualidade dos resultados obtidos em diferentes condições experimentais.

Na literatura encontram-se algoritmos com diferentes abordagens de desagregação que contemplam o caso em que há sobreposição.

(i) ECOTRIM

O ECOTRIM é um sistema de programas que fornece um conjunto de técnicas matemáticas e estatísticas para realizar a desagregação temporal e que foi criado por R. Barcellan e T. Di Fonzo (Barcellan, 2003), em nome da Comissão Europeia – Eurostat e escrito em Visual Basic e C++. Oferece uma abordagem amigável, sendo o utilizador que especifica as técnicas que deseja executar, além das informações óbvias sobre a série temporal a ser tratada. Implementa métodos de desagregação usando ou não séries relacionadas com a(s) série(s) que se pretende desagregar, para o caso univariado e multivariado.

No caso univariado implementa os métodos não baseados em séries relacionadas: o de Boot, Feibes e Lisman (Boot et al., 1967) e o de Denton (Denton, 1971) e os métodos

baseados no modelo ARIMA de Wei e Stram (Wei & Stram, 1990) e de Al-Osh (Al-Osh, 1989). Oferece vários métodos de desagregação por séries de alta frequência relacionadas, nomeadamente os métodos de ajustamento de Denton (Denton, 1971) e de Ginsburgh (Ginsburgh, 1973), métodos AR(1) (Fonzo, 1990), métodos ótimos no sentido dos mínimos quadrados de Fernandez (Fernandez, 1981) e de Litterman (Litterman, 1983) e o método de Guerrero (Guerrero, 1990) baseado no modelo ARIMA.

Oferece também métodos multivariados que permitem estimar séries desagregadas e cumprir restrições de agregação temporal e contemporânea, nomeadamente os procedimentos de ruído branco (Di Fonzo, 2002) e de passeio aleatório que são métodos ótimos no sentido dos mínimos quadrados e os métodos de ajustamento de Rossi (Rossi, 1982), de Denton (Denton, 1971) e de ajustamento ponderado, que exigem a estimação de séries preliminares.

(ii) H-Fuse

Liu et al. (2017) apresentam o método H-Fuse que se mostra eficaz na desagregação de dados, com possibilidade de sobreposição e dados ausentes. Além disso, tem a capacidade de proporcionar uma avaliação da confiabilidade da recuperação dos dados e em termos computacionais é aproximadamente linear no tamanho de dados de entrada.

O problema de desagregação com possível sobreposição no tempo e com lacunas pode ser representado por um sistema linear característico que comumente é subdeterminado. O objetivo é encontrar uma solução aproximada, razoavelmente precisa, que corresponda à informação desagregada.

Com o método H-Fuse usa-se o método dos mínimos quadrados e de regularização de Tikhonov (Tikhonov et al., 1995) (regularização linear) para encontrar uma solução aproximada de um sistema linear subdeterminado, introduzindo restrições adicionais como suavidade no espaço e/ou tempo para que a reconstrução reflita alguns aspetos dos dados de destino.

O problema de otimização a resolver consiste em estimar a série temporal $\vec{x} = (x_1, x_2, \dots, x_T)$, desconhecida, que pode ser agregada na série temporal $\vec{v} = (v_1, v_2, \dots, v_N)$ de dados agregados observada e que minimiza as funções de penalidade dependentes do domínio.

A função $F(\vec{x})$ que representa a norma dos desvios da série de dados observados e da série de dados desagregados que lhe deu origem e que se pretende minimizar é definida por:

$$F(\vec{x}) = \sum_{n=1}^N \left(v_n - \sum_{t=1}^T C_{nt} x_t \right)^2 \quad (2.155)$$

que na forma matricial toma a forma:

$$F(\vec{x}) = \|\vec{v} - O\vec{x}\|_2^2 \quad (2.156)$$

onde T é o número de períodos da série de alta frequência, N é o número de períodos da série de baixa frequência e $O_{N \times T}$ é a matriz de agregação ($\vec{v} = O\vec{x}$).

Uma vez que $F(\vec{x}) = 0$, solução ideal do problema de desagregação, é normalmente um problema subdeterminado, o método dos mínimos quadrados é usado para procurar a solução \vec{x} de norma mínima que satisfaz F ($\min \|\vec{x}\|_2^2$) que é obtida usando a matriz pseudo-inversa de Moore-Penrose (Penrose, 1955).

Como não há um motivo para crer que a solução é a de norma mínima, a metodologia H-Fuse propõe outra abordagem, cuja ideia base é a infundir no problema conhecimentos acerca do domínio. Por exemplo, o conhecimento da dinâmica populacional, a mudança climática, a migração de espécies biológicas, influenciam a análise de dados epidemiológicos.

Assim, a metodologia propõe minimizar a penalização total $\mathcal{L}(\vec{x})$, que tem duas componentes, $F(\vec{x})$ e $C(\vec{x})$, sendo esta última que infunde conhecimento do domínio na forma de restrições, como, por exemplo, suavização ou periodicidade.

Assim, o problema de otimização define-se da seguinte forma:

$$\min_{\vec{x}} \mathcal{L}(\vec{x}) = \min_{\vec{x}} (F(\vec{x}) + C(\vec{x})) \quad (2.157)$$

Restrições de suavidade penalizam grandes diferenças entre *time-ticks* adjacentes:

$$C_s(\vec{x}) = \sum_{t=1}^T (x_t - x_{t+1})^2 \quad (2.158)$$

Restrições a desvios de um período P que se assuma para a série:

$$C_s(\vec{x}) = \sum_{t=1}^T (x_t - x_{t+P})^2 \quad (2.159)$$

Se for introduzida mais que uma restrição, a proposta da metodologia é que se atribua o mesmo peso a todas as restrições num total de 1.

(iii) HOMERUN

Almutairi et al. (2018) introduzem uma metodologia que designam por HOMERUN para o problema da desagregação temporal. A metodologia é baseada na representação da série

através da Transformada Discreta do Cosseno (*Discrete Cosine Transform - DCT*) (Ahmed et al., 1974) e em restrições de não negatividade e suavidade. Recorre à compactação de energia do DCT encontrando a representação espectral mais esparsa da sequência alvo que contém os maiores (mais importantes) coeficientes. A DCT tem a capacidade de representar os dados de forma esparsa pois a maior parte da energia é compactada nos coeficientes de baixas frequências.

A definição mais comum da DCT é:

$$s_k = \sum_{n=0}^{N-1} x_n \alpha(k) \cos\left(\frac{\pi k(2n+1)}{2N}\right) \quad (2.160)$$

para $0 \leq k \leq N-1$ e $\alpha(k)$ definido por

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{N}} & \text{se } k = 0 \\ \sqrt{\frac{2}{N}} & \text{se } 1 \leq k \leq N-1 \end{cases} \quad (2.161)$$

A sequência original $x = \{x_n\}_{n=1}^N$ que origina a série de dados agregados $y = \{y_m\}_{m=1}^M$ pode ser obtida de forma única pela DCT inversa, definida por

$$x_n = \sum_{k=0}^{N-1} s_k \phi(k, n) \quad (2.162)$$

sendo $0 \leq n \leq N-1$ e

$$\phi(k, n) = \alpha(k) \cos\left(\frac{\pi k(2n+1)}{2N}\right) \quad (2.163)$$

Sendo $D \in \mathbb{R}^{N \times N}$ cujas entradas são as funções de base cosseno definidas em (2.163).

$$D = \begin{bmatrix} \phi(0,0) & \cdots & \phi(0,N) \\ \vdots & \ddots & \vdots \\ \phi(N-1,0) & \cdots & \phi(N-1,N-1) \end{bmatrix} \quad (2.164)$$

então as condições (2.160) e (2.162) podem ser escritas da seguinte forma:

$$s = Dx \quad (2.165)$$

E como D é ortogonal ($D^{-1} = D$)

$$x = D^T s \quad (2.166)$$

Sendo s o vetor de \mathbb{R}^N de $N-1$ CDT definidas em (2.160) e \mathbf{O} a matriz de $\mathbb{R}^{M \times N}$ em que cada linha é um vetor binário que tem 1 para os elementos de x que contribuem para y ,

dados $Ox = y$ pretende-se encontrar o vetor esparso que contém a DCT da sequência alvo x . Dado que minimizar a norma L1 promove esparsidade na solução, procura-se o mínimo $\|s\|_1 = \sum_{k=1}^{N-1} |s_k|$ que satisfaz $Ox = y$. Usando (2.163), o problema de otimização a resolver pela metodologia proposta em Almutairi et al., (2018) na sua versão básica (HOMERUN-0) toma a forma:

$$\begin{aligned} & \min_s \|s\|_1 \\ & \text{sujeito a } OD^T s = y \end{aligned} \quad (2.167)$$

Esta formulação coincide com a formulação *Basis Pursuit* de um problema de otimização.

Para resolver o problema de otimização resultante usam o Método de Direção Alternada de Multiplicadores (*Alternating Direction Method of Multipliers* - ADMM) (Boyd, 2010) com atualizações iterativas.

O algoritmo ADMM resolve problemas de otimização convexa que podem ser escritos na forma

$$\begin{aligned} & \min_{x,z} f(x) + g(z) \\ & \text{sujeito a } Ax + Bz = c \end{aligned} \quad (2.168)$$

iterando as atualizações seguintes:

$$x \leftarrow \arg \min_x f(x) + \left(\frac{\rho}{2}\right) \|Ax + Bz - c + u\|_2^2 \quad (2.169)$$

$$z \leftarrow \arg \min_z g(z) + \left(\frac{\rho}{2}\right) \|Ax + Bz - c + u\|_2^2 \quad (2.170)$$

$$u \leftarrow u + (Ax + Bz - c) \quad (2.171)$$

onde u é uma versão dimensionada das variáveis duais correspondentes à restrição de igualdade $Ax + Bz = c$, e ρ é especificado pelo utilizador (Ghadimi et al., 2015).

Almutairi et al. (2018) incrementam o HOMERUN-0 introduzindo restrições de positividade e de suavidade e deduzem os passos do algoritmo ADMM que permite resolver o problema de otimização após adicionar restrições de não negatividade e suavidade. As extensões assim introduzidas — HOMERUN-N e HOMERUN — melhoram a solução em situações específicas, reduzem o erro e aceleram a execução do algoritmo.

HOMERUN-N

Como nos problemas de desagregação se lida com sequências alvo não negativas, introduz-se esta restrição no problema de otimização e implementa-se o algoritmo HOMERUN-N para resolver o problema de otimização com a seguinte formulação:

$$\begin{aligned} & \min_s \|s\|_1 \\ & s. t. \quad OD^T s = y, \quad D^T s \geq 0 \end{aligned} \quad (2.172)$$

O algoritmo ADMM pode ser adaptado para resolver a formulação de otimização deste problema.

O problema (2.172) é convexo e pode ser reformulado introduzindo duas variáveis auxiliares r e z de \mathbb{R}^N como se segue:

$$\begin{aligned} & \min_s \mathfrak{I}_{\{r \geq 0\}} + \|s\|_1 \\ & s. t. \quad OD^T z = y, \quad D^T z = r, z = s \end{aligned} \quad (2.173)$$

sendo \mathfrak{I} uma função indicadora definida por:

$$\mathfrak{I}_{\{r \geq 0\}} = \begin{cases} 0 & \text{se } r \geq 0 \\ \infty & \text{noutros casos} \end{cases} \quad (2.174)$$

Designe-se por B , b e u as seguintes matrizes:

$$B := \begin{bmatrix} OD^T \\ D^T \\ I \end{bmatrix}; \quad b := \begin{bmatrix} y \\ r \\ s \end{bmatrix} \quad e \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \quad (2.175)$$

I é a matriz Identidade de $\mathbb{R}^{N \times N}$ e $u_1 \in \mathbb{R}^M$ e $u_2, u_3 \in \mathbb{R}^N$ são versões dimensionadas de variáveis duais e ρ é o parâmetro Lagrangeano aumentado definido por:

$$\begin{aligned} \mathcal{L}_\rho(r, s, z, u) &= \mathfrak{I}_{\{r \geq 0\}} + \|s\|_1 \\ &+ \frac{\rho}{2} (\|OD^T z - y + u_1\|_2^2 + \|D^T z - r + u_2\|_2^2 \\ &+ \|z - s + u_3\|_2^2) \end{aligned} \quad (2.176)$$

z, s, r e u obtêm-se minimizando \mathcal{L}_ρ em termos de uma variável enquanto se fixam as restantes de uma forma de otimização alternada. z é o primeiro bloco a ser atualizado, (r, s) é o segundo e u é o terceiro, através da resolução de sub-problemas de otimização:

$$z \leftarrow \arg \min_z \mathcal{L}_\rho(r, s, z, u) \quad (2.177)$$

$$\begin{cases} r \leftarrow \arg \min_r \mathcal{L}_\rho(r, s, z, u) \\ s \leftarrow \arg \min_s \mathcal{L}_\rho(r, s, z, u) \end{cases} \quad (2.178)$$

$$u \leftarrow u + (Bz - c) \quad (2.179)$$

onde a solução para z é a solução de forma fechada de mínimos quadrados via pseudo-inversa de Moore-Penrose. As soluções para s e r são casos do chamado operador de proximidade (K. Huang et al., 2015), onde s é resolvido usando um operador *soft-thresholding* (Selesnick, 2012) e a solução para r é a projeção não negativa $(\cdot)_+$ que transforma em zero os valores negativos .

Note-se que o operador *soft-thresholding* é a função $S: \mathbb{R} \rightarrow \mathbb{R}$ definida por:

$$S_k(x) = \begin{cases} x - k & \text{se } x < k \\ 0 & \text{se } |x| \leq k \\ x + k & \text{se } x < -k \end{cases} \quad (2.180)$$

Uma definição equivalente usada na biblioteca Sparse-plex do MATLAB⁴ é

$$S_k(x) = (x - \kappa)_+ - (-x - \kappa)_+. \quad (2.181)$$

O algoritmo 1 seguinte declara todas estas atualizações do HOMERUN usando notações algébricas lineares.

Uma vez que a pseudo-inversa de Moore-Penrose (B^\dagger) é usada em todas as iterações no Algoritmo 1, calcula-se uma vez na inicialização e armazena-se na variável R , para reduzir o esforço computacional. Depois de obter s , a solução aproximada da sequência alvo é:

$$x_{\text{HOMERUN-N}} = D^T s \quad (2.182)$$

O algoritmo foi avaliado com dados de epidemiologia, disponíveis no site do projeto Tycho (hepatite A, sarampo, papeira, tosse convulsa, poliomielite, rubéola e varíola) (Tycho Projetc, 2013) e com estes dados superou o desempenho do método dos Mínimos Quadrados e o H-FUSE.

Algoritmo 1: HOMERUN-N

Inicializar: Fazer $k = 1$ e $\mathbf{z}^k, \mathbf{s}^k, \mathbf{r}^k$, e \mathbf{u}^k vetores nulos; $\mathbf{b}^k = \begin{bmatrix} \mathbf{y} \\ \mathbf{r} \\ \mathbf{s} \end{bmatrix}$;

Calcular a matriz $\mathbf{R} = \mathbf{B}^\dagger$, pseudo-inversa de $\mathbf{B} = \begin{bmatrix} \mathbf{O} \mathbf{D}^T \\ \mathbf{D}^T \\ \mathbf{I} \end{bmatrix}$, $\mathbf{I} \in \mathbb{R}^{N \times N}$

e armazená-la.

Repetir

- $\mathbf{z}^{k+1} = \mathbf{R}(\mathbf{b}^k - \mathbf{u}^k)$
- $\mathbf{s}^{k+1} = \left(\mathbf{z}^{k+1} + \mathbf{u}_3^k - \frac{1}{\rho} \right)_+ - \left(-\mathbf{z}^{k+1} - \mathbf{u}_3^k - \frac{1}{\rho} \right)_+$
- $\mathbf{r}^{k+1} = (\mathbf{D}^T \mathbf{z}^{k+1} + \mathbf{u}_2^k)_+$
- **Atualizar \mathbf{b}^{k+1} usando \mathbf{r}^{k+1} e \mathbf{s}^{k+1}**
- $\mathbf{u}^{k+1} = \mathbf{u}^k + (\mathbf{B} \mathbf{z}^{k+1} - \mathbf{b}^{k+1})$
- **Fazer $k = k + 1$**

Parar quando o número máximo de iterações for atingido ($k = 3000$)

⁴ <https://sparse-plex.readthedocs.io/en/latest/intro.html>

HOMERUN – Versão final

A ideia principal do HOMERUN na sua versão final é explorar o conhecimento do domínio de suavidade, penalizando as grandes diferenças entre *timeticks* adjacentes. Assim, adiciona-se uma penalização de suavidade que resulta no problema de otimização seguinte:

$$\begin{aligned} \min_s \quad & \|s\|_1 + \frac{1}{2} \|HD^T s\|_2^2 \\ \text{s. t.} \quad & OD^T s = y, \quad D^T s \geq 0 \end{aligned} \quad (2.183)$$

onde $H \in \mathbb{R}^{(N-1) \times N}$ é uma matriz de suavidade que tem -1 na n -ésima linha e 1 nas n -ésima e $(n+1)$ -ésima colunas e zeros nas restantes posições. Pode-se adicionar um parâmetro λ de regularização (ponderação) com a penalização de suavidade acima, mas os autores observaram que $\lambda = 1$ fornece desempenho ótimo ou quase ótimo do algoritmo.

De forma similar ao Algoritmo 1, na versão final do HOMERUN reformula-se o problema (2.183) como se segue:

$$\begin{aligned} \min_s \quad & I_{\{r \geq 0\}} + \|s\|_1 + \frac{1}{2} \|HD^T s\|_2^2 \\ \text{s. t.} \quad & OD^T z = y, \quad D^T z = r, \quad s = z \end{aligned} \quad (2.184)$$

em que $r, z \in \mathbb{R}^N$ são variáveis auxiliares.

HOMERUN resolve para s a representação DCT esparsa da sequência, usando a norma L1.

O Lagrangeano aumentado é definido por:

$$\begin{aligned} \mathcal{L}_\rho(r, s, z, u) = & \mathfrak{S}_{\{r \geq 0\}} + \|s\|_1 + \frac{1}{2} \|HD^T s\|_2^2 \\ & + \frac{\rho}{2} \left(\|OD^T z - y + u_1\|_2^2 + \|D^T z - r + u_2\|_2^2 \right. \\ & \left. + \|z - s + u_3\|_2^2 \right) \end{aligned} \quad (2.185)$$

O parâmetro ρ é tomado igual a 1, pois fornece desempenho ideal ou quase ideal, resultando assim num modelo livre de parâmetros.

À semelhança do HOMERUN, designe-se por Q , q e v as seguintes matrizes:

$$Q := \begin{bmatrix} OD^T \\ D^T \\ I \\ HD^T \end{bmatrix}; \quad q := \begin{bmatrix} y \\ r \\ s \\ 0 \end{bmatrix} \quad e \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ 0 \end{bmatrix} \quad (2.186)$$

em que $0 \in \mathbb{R}^{N-1}$ é um vetor de zeros.

Também aqui se obtém z, s, r e u minimizando \mathcal{L}_ρ em termos de uma variável e fixando as outras.

Dado que a implementação do algoritmo anterior envolve a determinação da pseudo-inversa de Moore-Penrose numa única etapa, com complexidade $O(N^3)$, e as atualizações iterativas consistem em multiplicações de matrizes, adições de vetores e atualizações de vetores elemento a elemento com complexidade dominada pelo número de não-zeros na matriz de observação O , a implementação direta é aceitável com dados de tamanho pequeno a moderado. Os autores propõem um algoritmo que acelera a execução do método pois reduz o custo computacional da atualização de z em (2.185).

Algoritmo: HOMERUN- implementação direta

Inicializar:

- Fazer $\mathbf{k} = 1$ e $\mathbf{z}^k, \mathbf{s}^k, \mathbf{r}^k$, e \mathbf{u}^k vetores nulos; $\mathbf{b}^k = \begin{bmatrix} \mathbf{y} \\ \mathbf{r} \\ \mathbf{s} \end{bmatrix}$; \mathbf{q}^k e \mathbf{v}^k ;
- Calcular a matriz $\mathbf{W} = \mathbf{Q}^\dagger$, pseudo-inversa de \mathbf{Q} e armazená-la em \mathbf{W} .

Repetir

- $\mathbf{z}^{k+1} = \mathbf{W}(\mathbf{q}^k - \mathbf{v}^k)$
- $\mathbf{s}^{k+1} = (\mathbf{z}^{k+1} + \mathbf{u}_3^k - \mathbf{1})_+ - (-\mathbf{z}^{k+1} - \mathbf{u}_3^k - \mathbf{1})_+$
- $\mathbf{r}^{k+1} = (\mathbf{D}^T \mathbf{z}^{k+1} + \mathbf{u}_2^k)_+$
- Atualizar \mathbf{b}^{k+1} usando \mathbf{r}^{k+1} e \mathbf{s}^{k+1}
- Atualizar \mathbf{q}^{k+1} usando \mathbf{r}^{k+1} e \mathbf{s}^{k+1}
- $\mathbf{u}^{k+1} = \mathbf{u}^k + (\mathbf{B}\mathbf{z}^{k+1} - \mathbf{b}^{k+1})$
- Atualizar \mathbf{v}^{k+1} usando \mathbf{u}^{k+1}
- **Fazer $k = k + 1$**

Parar quando o número máximo de iterações for atingido ($k = 3000$)

(iv) ARES

Yang et al. (2018) propuseram um método de processamento de sinal para resolver o problema da desagregação temporal sem fazer qualquer suposição sobre a série temporal. Designaram a metodologia por ARES (*Automatic REStorage*). A metodologia é implementada em duas fases: primeiro, estima a sequência de contagens utilizando o conhecimento do domínio, como suavidade e periodicidade de eventos; de seguida, usa a sequência estimada para deduzir padrões notáveis na sequência alvo para refinar a série temporal reconstruída. Na segunda fase o ARES melhora a precisão da reconstrução

aplicando a filtragem de aniquilação de forma iterativa. A filtragem de aniquilação, base da metodologia proposta, permite descobrir automaticamente padrões em períodos vizinhos.

Um filtro aniquilador (*annihilating filter - AF*) é um operador linear de deslocamento invariante, de suporte finito, que suprime completamente um determinado sinal (Hormati & Vetterli, 2007). Ou seja, um filtro com resposta ao impulso $h(t)$, tal que:

$$(x * h)(t) := \sum_{\tau} x_{t-\tau} h_{\tau} = 0, \forall t \quad (2.187)$$

Na primeira fase o ARES usa a metodologia de H-Fuse com a restrição:

$$C_{sp}(\vec{x}) = \frac{1}{2} C_s(\vec{x}) + \frac{1}{2} C_p(\vec{x}) \quad (2.188)$$

(equações (2.158) para resolver o problema de otimização:

$$\min_{\vec{x}} \mathcal{L}(\vec{x}) = \min_{\vec{x}} \left(F(\vec{x}) + C_{sp}(\vec{x}) \right) \quad (2.189)$$

Na segunda fase, resolve o problema de otimização:

$$\min_{\vec{x}} \mathcal{L}(\vec{x}) = \min_{\vec{x}} \left(F(\vec{x}) + C_a(\vec{x}) \right) \quad (2.190)$$

sendo $C_a(\vec{x})$ a restrição de aniquilamento definida por:

$$C_a(\vec{x}) = \sum_{t=1}^{T-L+1} \left(\sum_{l=1}^L (c_l x_{l+t}) \right)^2 \quad (2.191)$$

T é o número de períodos da série de alta frequência e L é o tamanho do deslocamento do filtro AF. Os valores c_l são os coeficientes de AF e são tais que $\sum_{l=1}^L (c_l \vec{x}_l) = 0$, para quaisquer L tempos sucessivos de uma sequência alvo.

Em síntese, o ARES primeiro aplica restrições de suavidade e periodicidade aos relatórios agregados para obter uma sequência reconstruída \vec{x}_{sp} e depois cria uma restrição de aniquilação correspondente $C_a(\vec{x})$ com base nas informações em \vec{x}_{sp} .

O algoritmo foi avaliado em dados de epidemiologia e demonstrou elevado desempenho, superando outros algoritmos do estado da arte.

(v) TurboLift

Yang et al. (2020) propõem uma nova abordagem, chamada TurboLift, que refina as soluções fornecidas pelos métodos de desagregação existentes e aumenta significativamente a precisão das séries temporais de alta resolução reconstruídas. A partir de uma solução inicial produzida por um método de desagregação específico, o TurboLift encontra iterativamente uma nova solução que minimiza o erro de desagregação e está próxima da solução atual. Os autores proporcionam uma implementação eficiente para o processo

iterativo do TurboLift explorando a estrutura das matrizes de agregação. Ainda, deduzem uma solução analítica que permite obter a solução para a formulação do TurboLift, sem necessidade de realizar iterações demoradas.

Depois de obter a série desagregada \vec{z} por um método do estado da arte, o problema de otimização, é definido da seguinte forma:

$$\min_{x,z} (\|O\vec{x} - \vec{y}\|_2^2 + \|\vec{x} - \vec{z}\|_2^2) \quad (2.192)$$

em que \vec{y} é o vetor que contém os dados agregados, O é a matriz de agregação que mapeia os dados desagregados \vec{x} em \vec{y} .

Uma solução para o problema utiliza o algoritmo Mínimos Quadrados Alternados (*Alternating Least Squares* - ALS) para atualizar iterativamente as duas variáveis \vec{x} e \vec{z} de forma cíclica. A partir de uma solução inicial \vec{z}_0 produzida por um determinado método do estado da arte, fixa-se uma variável e atualiza-se a outra até a convergência.

Também deduziram a seguinte solução analítica:

$$x_{k+1} = (O^T O + I)^{-1} O^T \vec{y} + (O^T O + I)^{-1} \vec{z}_k \quad (2.193)$$

onde $O \in \mathbb{R}^{M \times N}$ é a matriz de Toeplitz (matriz em que cada diagonal descendente da esquerda para a direita é constante) (Heinig & Rost, 2013).

Aplicaram a metodologia a dados reais de epidemiologia, vendas, clima e crime e mostraram que esta não deteriora os resultados obtidos pelos outros algoritmos e que regra geral melhora o desempenho de diferentes métodos de reconstrução sob diferentes restrições de conhecimento de domínio (por exemplo, suavidade, esparsidade e periodicidade).

Algoritmo: TURBOLIFT- implementação iterativa (Yang et al., 2020)

Inicializar:

- Fazer $k = 0$
- z_0 = solução inicial, obtida com outro algoritmo
- Obter a decomposição de Cholesky da matriz O : $(O^T O + I) = LL^T$

Repetir

- $x_{k+1} = (L^T)^{-1} L^{-1}(O^T v + z_k)$ (usando Foward e Back substitution)
- $z_{k+1} = x_{k+1}$
- Fazer $k = k + 1$

Parar quando o critério de convergência for atingido

Algoritmo: TURBOLIFT – solução analítica (Yang et al., 2020)

Function *TurboLift* ($\mathbf{O}, \mathbf{v}, \mathbf{z}_0$)

Obter a dimensão da matriz \mathbf{O} (matriz de observação)

$$\mathbf{M}, \mathbf{N} \leftarrow \text{dimensão de } \mathbf{O}$$

Decompor em Valores Singulares

$$\mathbf{U}, \mathbf{\Sigma} \leftarrow \text{SVD}(\mathbf{O}^T \mathbf{O})$$

Extrair as \mathbf{M} primeiras colunas de \mathbf{U} como \mathbf{U}_1 e as últimas $(\mathbf{N} - \mathbf{M})$ como \mathbf{U}_2

$$\mathbf{U}_1 \leftarrow [1: \mathbf{M}]; \mathbf{U}_2 \leftarrow [\mathbf{M} + 1: \mathbf{N}]$$

Extrair os valores singulares não nulos

$$\mathbf{\Sigma}_1 \leftarrow \mathbf{\Sigma}[1: \mathbf{M} \ 1: \mathbf{M}]$$

Calcular a série reconstruída

$$\mathbf{x}_{TL} \leftarrow \mathbf{U}_1 \mathbf{\Sigma}_1^{-1} \mathbf{U}_1^T \mathbf{O}^T \mathbf{v} + \mathbf{U}_2^T \mathbf{z}_0$$

(vi) **INCOMPFUSE e A-CURE**

Xu et al. (2020b, 2020a) propõem uma metodologia de desagregação de dados que utiliza sistematicamente a incompatibilidade e a inconsistência de relatórios para a reconstrução de dados. Em particular, a estrutura proposta permite detetar fontes de dados imprecisas (ruidosas) e usar esse conhecimento no processo de reconstrução de dados, mesmo com sobreposição.

A abordagem proposta em Xu et al., (2020b) permite a avaliação da incompatibilidade e a inconsistência de relatórios $([t, t'], y)$ que dão a informação de dados agregados y num intervalo de tempo $[t, t']$, através da definição de uma função de probabilidade de incompatibilidade de dois relatórios i e j .

Metodologia

Considere-se um intervalo de valores de tempo, T , e um conjunto de relatórios, $R = \{r_1, \dots, r_n\}$, onde cada relatório tem a forma $r = ([t, t'], y)$ que representa, para o relatório r , o valor agregado v no intervalo de tempo t para t' , com t e $t' \in T$. Diz-se que a duração do relatório r é $t' - (t - 1)$.

Um par de relatórios pode ser inconsistente ou incompatível.

É inconsistente se o intervalo de tempo de um relatório desse conjunto está contido no intervalo de tempo de outro relatório, mas o valor agregado do primeiro relatório é maior que o do segundo.

Assim, dois relatórios $r_1 = ([t_1, t'_1], y_1)$ e $r_2 = ([t_2, t'_2], y_2)$ são consistentes se as seguintes condições são válidas:

$$\begin{aligned} a) \quad & t_1 \geq t_2 \quad \wedge \quad t'_1 \leq t'_2 \Rightarrow y_1 \leq y_2; \\ b) \quad & t_1 = t_2 \quad \wedge \quad t'_1 = t'_2 \Rightarrow y_1 = y_2 \end{aligned} \tag{2.194}$$

Caso contrário, os relatórios r_1 e r_2 são inconsistentes.

Dois relatórios podem parecer incompatíveis se fornecerem informações contraditórias sobre o mesmo evento ou entidade. Por exemplo, dois relatórios sobrepostos têm a mesma duração, mas um tem um valor muito superior ao outro. Se os dados forem ruidosos, ambos os relatórios podem estar errados. A incompatibilidade não é uma relação binária, pois podem existir pares mais incompatíveis do que outros. Por isso, interessa quantificar a noção de incompatibilidade usando uma medida de distância lógica.

Os autores definem uma função $inc: \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ com as seguintes propriedades:

1. Inconsistência: $inc(r_i, r_j) = 1$ se e só se r_i e r_j são inconsistentes,
2. Compatibilidade completa: $inc(r_i, r_j) = 0$ se e só se $t_i = t_j \wedge t'_i = t'_j \wedge v_i = v_j$
3. Comutatividade: $inc(r_i, r_j) = inc(r_j, r_i)$ para todo $r_i, r_j \in \mathbb{R}$,

Além disso, $inc(r_i, r'_i) < inc(r_j, r'_j)$ significa que o par r_i, r'_i é menos incompatível que o par r_j, r'_j .

A ideia geral da metodologia é que enquanto o número de relatórios com ruído for pequeno comparado com o número total de relatórios, os relatórios com ruído serão mais incompatíveis com outros relatórios do que os reais (sem ruído). Dessa forma, usa-se a incompatibilidade para encontrar os relatórios que provavelmente têm ruído.

Definição de probabilidade de incompatibilidade

a. Probabilidade de incompatibilidade básica (BIP)

Supondo que os dados se distribuem uniformemente por todo um conjunto de amplitude de valores de tempo T , define-se o valor médio de um registo como:

$$a = \frac{y}{t' - t + 1} \tag{2.195}$$

sendo y o valor agregado no intervalo $[t, t']$. Designando por d a duração de um registo r , $d = t' - (t - 1)$ e a fica:

$$a = \frac{y}{d} \quad (2.196)$$

A definição mais simples de probabilidade de incompatibilidade (*Basic incompatibility Probability* – BIP) de dois relatórios r_1 e r_2 , de valores médios a_1 e a_2 , respetivamente, pode considerar-se como:

$$BIP = \frac{|a_2 - a_1|}{\max(a_1, a_2)} \quad (2.197)$$

Como, na realidade, o pressuposto da distribuição uniforme dos dados não é plausível, introduz-se no BIP os conceitos de distância entre os dois relatórios (diferença entre os pontos médios dos respetivos intervalos) e o conceito de duração de relatório. A tolerância acerca da probabilidade de incompatibilidade de dois relatórios com valores médios diferentes é maior se estes estiverem mais afastados no tempo ou se a diferenças de durações for grande.

Assim, define-se *ratio de proximidade* entre dois relatórios r_i e r_j ($pr_{i,j}$) como:

$$pr_{i,j} = 1 - \frac{dist_{i,j}}{Cobertura_{i,j}} \quad (2.198)$$

sendo

$$dist_{i,j} = \left| \frac{t'_i + t_i}{2} - \frac{t'_j + t_j}{2} \right| \quad (2.199)$$

e

$$Cobertura_{i,j} = \max(t'_j, t'_i) - \min(t_i, t_j) \quad (2.200)$$

E define-se *ratio de distância* entre dois relatórios r_i e r_j ($dr_{i,j}$) como:

$$dr_{i,j} = \frac{|d_i - d_j|}{\max(d_i, d_j)} \quad (2.201)$$

A probabilidade de incompatibilidade entre dois relatórios r_i e r_j é então definida como:

$$IP_{i,j} = (BIP_{i,j} \times (1 - dr_{i,j})) + (1 - BIP_{i,j}) \times (1 - pr_{i,j}) \times pr_{i,j} \quad (2.202)$$

Os autores aplicam a metodologia a dados experimentais e concluem que esta tem uma elevada precisão a detetar relatórios com ruído.

Xu et al. (2020a) propõe a metodologia A-Cure, recorrendo ao INCOMFUSE para analisar a incompatibilidade de relatórios e obter um conjunto de relatórios “limpo” (sem relatórios com ruído) que é usado para a reconstrução da série de alta frequência. Esta

metodologia é implementada sequencialmente em três etapas: (1) ordenar os relatórios usando a função cumulativa de incompatibilidade para cada relatório ou um limite de probabilidade de incompatibilidade para filtrar os pares de relatórios sem ruído ou o grafo de incompatibilidade e o fluxo de energia ; (2) estimar uma sequência de relatórios ordenados por imprecisão, usando métricas RMSE sequenciais (S-RMSE – *Sequential Root-Mean-Square-Deviation*) para comparar reconstruções de diferentes subconjuntos dos relatórios; (3) eliminar os relatórios imprecisos que afetam o erro de construção e recuperar os dados de destino de alta precisão usando um algoritmo do estado da arte.

O RMSE permite estudar a diferença de reconstrução entre os resultados reconstruídos de dois conjuntos de relatórios agregados quase idênticos. A única diferença entre os dois conjuntos de relatórios é que um determinado relatório r_i é incluído no primeiro conjunto de dados, mas excluído do segundo conjunto. Uma diferença significativa entre as reconstruções resultará num grande RMSE, o que pode indicar que o relatório r_i é um relatório com ruído e que deve ser excluído. Nesta metodologia todos os relatórios são rastreados sequencialmente.

A ordem pela qual os relatórios entram no algoritmo S-RMSE afeta o impacto do erro de reconstrução. Quanto mais cedo entrarem mais preciso será o resultado.

Assim, usando um método híbrido que concilia o S-RMSE com um método de ordenação é possível encontrar uma ordem sequencial otimizada sem ser fundamentada empiricamente.

O INCOMPFUSE (Xu et al., 2020b) é um método que suporta a avaliação eficiente da confiabilidade dos dados usando a probabilidade de incompatibilidade (IP - *incompatibility probability*) de relatórios históricos. Usando esta a metodologia é possível usar o IP cumulativo de cada relatório para proceder a uma ordenação, uma vez que um valor de IP elevado representa uma maior incompatibilidade entre dois relatórios, o que também pode indicar a existência de um relatório ruidoso no par de relatórios em análise.

O IP cumulativo é definido por:

$$IP_i = \sum_{j=1}^N IP_{i,j} \quad (2.203)$$

onde N é o número de relatórios.

Os relatórios podem entrar no algoritmo por ordem decrescente de IP. O processo de filtragem de relatórios ruidosos pode ser simplificado se, em vez de acumular o IP para cada relatório, se usar um limite de IP para filtrar os pares de relatórios não ruidosos. Em Xu et

al. (2020b) os autores propõem uma forma automática de determinar um limite para IP: ordenar a distribuição do IP por ordem crescente,

Os autores propõem outra metodologia de ordenação que usa Grafos de Incompatibilidade e o método de Fluxo de Energia e que permite analisar a incompatibilidade global, contrariamente ao método INCOMPFUSE que determina a probabilidade de incompatibilidade apenas entre pares de relatórios. O grafo de incompatibilidade é construído baseado na IP e é otimizado com base no método de fluxo de energia (Xu et al., 2020a). Este método revelou-se o mais eficaz na reconstrução de dados de alta frequência na aplicações a dados de saúde anteriormente referidos (Tycho Project, 2013).

O algoritmo S-RMSE opera da seguinte forma, depois de estabelecida uma ordenação nos relatórios:

Algoritmo: S-RMSE

Inicialização: Definir o conjunto de N relatórios, incluindo relatórios com ruído, numa determinada ordem sequencial: **1, 2, 3, ..., N**; pré-definir o limiar de RMSE;

Para i de 1 até N , repetir:

- Reconstruir a série a partir do conjunto de relatórios, incluindo o relatório i ;
- Reconstruir a série a partir do conjunto de relatórios, sem o relatório i ;
- Calcular o RMSE entre os dois resultados de reconstrução:
- Se o RMSE for maior do que o limiar pré-definido então
 - Atualizar o conjunto de relatórios excluindo o relatório i ;

Apresentar os dois conjuntos de relatórios: o dos que ultrapassaram o limiar de RMSE e o dos restantes, que serão usados para a reconstrução final.

Os dados multidimensionais tornaram-se omnipresentes e são frequentemente encontrados em situações em que os mesmos dados são agregados numa ou mais dimensões, segundo diferentes perspetivas de agregação: ao longo do tempo, no espaço, segundo um atributo.

Almutairi et al. (2021) propõem um algoritmo para reconstruir dados em alta frequência de várias visualizações grosseiras, agregadas em diferentes dimensões. Este método utiliza ferramentas de fatorização tensorial de baixo nível (Jeevanjee, 2011) para fundir as múltiplas visualizações e fornecer garantias de recuperação sob certas condições e pode ser usado com

dados ausentes ou parcialmente observados, com agregação dupla e até desagregação sem qualquer referência sobre padrões de agregação (B-PREMA).

Os autores aplicaram a metodologia a dados reais de diferentes domínios: vendas a retalho, contagem de crimes e observações meteorológicas e concluíram que o algoritmo proposto é muito eficaz em diferentes cenários.

CAPÍTULO 3

MODELAÇÃO DE DADOS ESPAÇO TEMPORAIS DE COVID-19 DA ÁREA METROPOLITANA DE LISBOA

3. MODELAÇÃO DE DADOS ESPAÇO TEMPORAIS DE COVID-19 DA ÁREA METROLOPITANA DE LISBOA

3.1. Introdução

Em dezembro de 2019, uma nova doença causada por uma infeção por coronavírus foi relatada na província de Wuhan, na China, denominada Coronavirus 19 (COVID-19). Caracterizado por um alto nível de transmissão, o novo vírus é conhecido como Síndrome Respiratória Aguda Grave CoronaVirus 2 (SARS-CoV-2), conforme estabelecido pelo *Coronaviridae Study Group do International Committee on Taxonomy of Viruses (ICTV)* (Gorbalenya et al., 2020). A propagação dessa doença viral foi muito rápida e com danos graves em todo o mundo, diretamente na saúde humana e indiretamente na educação, turismo, emprego, mobilidade social, economia e política, entre outros, sendo uma das ameaças mais graves à Saúde Pública dos últimos tempos. As taxas de incidência, altamente variáveis em diferentes locais, mostram impactos distribuídos de forma desigual, de modo que a epidemiologia espacial pode ajudar a entender a distribuição e disseminação da doença no espaço e no tempo. A dinâmica da doença, especialmente o longo desfasamento espacial e temporal entre a exposição e a deteção, o potencial de transmissão de infeções assintomáticas e as implicações de uma resposta de saúde pública espacialmente fragmentada, justificam pesquisas com ferramentas espaciais.

Além disso, o comportamento humano afeta o contágio - o patogénico associado à COVID-19 é facilmente transmitido de pessoa para pessoa por via aérea ou contato direto.

Pessoas infetadas com COVID-19 são contagiosas antes de apresentarem sintomas: “A eliminação viral máxima de SARS-CoV-2 (a causa de COVID-19) ocorre no início do curso da doença. Os pacientes podem, portanto, ser contagiosos antes de desenvolverem os sintomas ou mesmo saberem que estão infetados. A transmissão de SARS-CoV-2 por indivíduos assintomáticos foi claramente documentada, e modelos matemáticos sugerem que 40-80% dos eventos de transmissão ocorrem em pessoas pré-sintomáticas ou assintomáticas” (Javid et al., 2020). Os eventos de contágio podem ocorrer a partir de contatos de superfície, mas são atribuídos principalmente à exalação de aerossóis, não apenas por tosse e espirro, mas também pela respiração normal, e aumentam ao falar (Stadnytskyi et al., 2020). Assim, ambientes fechados, incluindo veículos, são contextos de

alto risco para o contágio. O transporte público de massa, em particular, é de alto risco devido a:

- “1. *People are confined in limited space. Contagion risk increases with the level of passenger occupancy in vehicles and stations. (...)*
2. *There might be scarce access control to identify passengers or workers who may be sick.*
3. *The existence of multiple surfaces, such as seats, handrails, doors, and ticket machines, that easily transfer germs.”*

(Tirachini & Cats, 2020)

O contato entre passageiros nas redes de transporte público pode ser um mediador fundamental na propagação de doenças infecciosas (Mo et al., 2021). O tempo de proximidade também é um fator a considerar: a exalação de aerossóis por indivíduos infectados em ambiente fechado fará com que a concentração de partículas virais viáveis aumente com o tempo, aumentando significativamente o risco de contágio (Tang et al., 2020).

Assim, fatores como a mobilidade humana, o meio de transporte utilizado para viajar, bem como a distância percorrida ou o tempo de viagem podem ajudar a compreender a dinâmica da propagação do vírus e prevenir situações potenciais de superdisseminação. Diversos estudos sustentam a importância destes fatores na compreensão e previsão da dinâmica de transmissão de doenças e na eficácia das restrições de viagens no controlo do aumento de contágio, nomeadamente entre concelhos (Chinazzi et al., 2020; Kraemer et al., 2020; Melo et al., 2021; Yilmazkuday, 2020).

A influência de fatores demográficos na dinâmica da propagação do SARS-CoV-2 foi analisada em vários estudos, nomeadamente a densidade populacional, p.e. Hamidi et al. (2020), ou o índice de envelhecimento da população, p.e. Mueller et al. (2020). O aumento da idade é um fator de risco para um pior prognóstico da COVID-19, quer para o aumento do número de casos, como para o risco de doença grave (Zhao et al., 2020), mas não há evidências claras de que a densidade populacional em si seja um facilitador da propagação do SARS-CoV-2. No entanto, em áreas metropolitanas isto verificou-se até que as medidas de contenção da propagação fossem implementadas, o que sugere que um fator significativo é a facilidade de conexão entre concelhos e não a densidade populacional (Hamidi et al., 2020). Portanto, espera-se que o papel de um concelho na dinâmica regional geral da

COVID-19 dependa mais do seu fluxo diário e da distância efetiva a outros concelhos medida pelo fluxo de pessoas do que pela distância geográfica (Melo et al., 2021).

Segundo Bellomo et al. (2020), um número considerável de autores utilizou modelos baseados apenas em premissas matemáticas para tentar prever a longo prazo o número de casos de COVID-19, mas há muitos e diversos fatores que ao longo do tempo podem afetar substancialmente as previsões a mais longo prazo, como, por exemplo, mudanças nas políticas públicas de saúde, na dinâmica da doença e nas características biológicas e sociodemográficas da população. Esses autores apresentam um amplo olhar sobre essas questões emergentes da pandemia COVID-19, fornecendo recomendações para a tomada de decisões, exigindo uma contribuição interdisciplinar de especialistas, como epidemiologistas, imunologistas e economistas, além dos especialistas em modelação matemática.

No presente capítulo considera-se a modelação estatística, para um caso particular de um evento público de larga escala, realizado em Lisboa, Portugal, no início da pandemia: a Manifestação do 1.º de Maio e analisam-se variáveis de campo com características espaciais.

Tendo em conta que na altura vigoravam as primeiras medidas de confinamento definidas pelas autoridades no contexto de estado de emergência em Portugal, que restringiam a mobilidade entre concelhos e aglomerações, a exceção concedida a esta manifestação gerou uma discussão pública significativa e politizada nos meios de comunicação social. Esta discussão foi ainda alimentada por um aumento dos novos casos diários de COVID-19 na região de Lisboa, ao contrário das outras grandes regiões de Portugal, que apresentavam uma diminuição do contágio diário como se verá na secção 3.3.

Estas circunstâncias levantaram a questão da possibilidade de extrair conclusões estatisticamente sustentadas sobre possíveis efeitos do evento no padrão de contágio da COVID-19 na região de Lisboa, de forma que pudesse ser aplicada metodologicamente a outros eventos.

Em particular, considerou-se que o contágio com origem no evento pode ocorrer no próprio evento (i.e. durante a manifestação) ou durante as aglomerações associadas, nomeadamente durante as deslocações de e para o evento. Assim, pretendeu-se analisar se há evidência estatística sobre se os efeitos, a existirem, se restringiram a áreas geográficas específicas associadas às viagens, e não à região como um todo. Para responder a esta

questão implementou-se uma análise empírica baseada em modelos de regressão de Poisson, testes de hipóteses e visualização.

Esta mesma análise foi replicada usando dados do evento de manifestação contra o racismo, associada ao movimento "*Black Life matters*" e que juntou milhares de jovens em Lisboa no dia 6 de junho de 2020, tendo a maioria desfilado de máscara, mas sem respeitar a distância social imposta pela prevenção da Covid-19.⁵

Os dados sobre a COVID-19 disponibilizados publicamente pelas autoridades de saúde pública portuguesas não são consistentes, quer em termos de periodicidade, quer em termos de métricas de agregação, nomeadamente o período considerado nesta tese – 24/03/20 a 24/03/21. Para a análise implementada no capítulo 5 foi necessário realizar as transformações descritas nesse capítulo, que introduzem algum viés nos resultados. Nos períodos em análise no presente capítulo – 24/03/2020 a 6/06/2020 e 21/05 a 29/06 não há qualquer problema de métrica de agregação ou periodicidade, uma vez que neste período os relatórios com dados dos concelhos eram diários. Foi necessário lidar com outro tipo de problema que se prendeu com a forma como os relatórios diários disponibilizaram a informação – gráficos sem informação sobre os dados absolutos, dificultando o acesso aos mesmos. Na secção 3.3.2 descreve-se uma metodologia de extração de dados a partir de gráficos de barras.

3.2. Dados de COVID-19 dos concelhos da Área Metropolitana de Lisboa

3.2.1. De onde vêm, o que são e como são

Como será detalhado no capítulo 4, secção 4.1, a DGS começou a disponibilizar publicamente dados sobre a COVID-19 em Portugal, a partir do dia 24 de março de 2020, na forma de boletins com relatório diário, num formato que torna a extração de dados nada trivial. Verificou-se a existência de dados processados dentro dos boletins em formatos que não são facilmente inspecionáveis/manipuláveis, como gráficos, em detrimento dos dados absolutos que lhes deram origem, com eixos demasiado esparsos e legibilidade reduzida. Esta fragilidade aliada a alguns erros nos dados dificultam as tarefas de análise, modelação, visualização e monitorização.

⁵ <https://www.dn.pt/poder/manifestacao-contra-racismo-junta-centenas-nas-ruas-de-lisboa-12285732.html>

Os dados sobre o número cumulativo de casos COVID-19 usados neste capítulo são os que foram recolhidos diariamente através da comunidade DSSG PT⁶ de 15 de abril a 29 de junho de 2020 e que provêm dos relatórios de situação diários emitidos diariamente pela Direção-Geral de Saúde durante este período. Alguns dos dados que foram disponibilizados na forma de gráficos foram objeto da metodologia de extração de dados caracterizada no capítulo 2, secção 2.9.

3.2.2. Dados demográficos da Área Metropolitana de Lisboa

As regiões em estudo são Portugal Continental e a Área Metropolitana de Lisboa (AML). Em Portugal Continental não foram incluídas as regiões autónomas. A AML é composta pelos concelhos de Alcochete, Almada, Amadora, Barreiro, Cascais, Lisboa, Loures, Mafra, Moita, Montijo, Odivelas, Oeiras, Palmela, Seixal, Sesimbra, Setúbal, Sintra, e Vila Franca de Xira, indexados por $k = 1, 2, \dots, 18$. Os dados de georreferenciação foram retirados do site da DGSS PT (Figura 3.1).

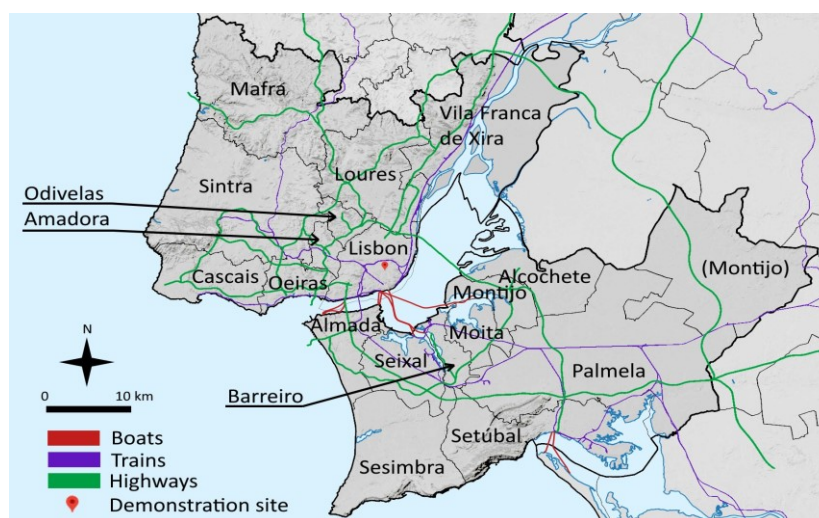


Figura 3.1: Mapa da Área Metropolitana de Lisboa. Composição de vários mapas da entrada da Wikipedia AML.

Os dados relativos a variáveis demográficas dos Concelhos da AML (*índice de envelhecimento, densidade populacional, população estimada em 2019*) foram obtidos na base de dados PORDATA⁷ e os dados relativos às distâncias espaciais e temporais, em viatura própria e em transportes públicos, na AML foram obtidos através do serviço Google

⁶ <https://github.com/dssg-pt/covid19pt-data> (acedido em 14 de julho de 2023)

⁷ <https://www.pordata.pt>

Maps, no dia 13 de julho de 2020. Como estas estimativas dadas pelo serviço Google Maps variam conforme o padrão histórico, foi escolhido o domingo seguinte ao dia de consulta (19 de julho) como dia de referência mais expectavelmente similar às deslocações de um dia feriado ou um sábado como os casos em análise (1.º de Maio de 2020 e 6 de junho de 2020). As estimativas de trajetos e tempos de viagem foram obtidas para viagens com origem nas sedes dos concelhos (no caso de Lisboa, usámos a Praça do Município) e com hora de chegada ao local da manifestação às 11:00. No caso das deslocações em viatura própria ou autocarro fretado foram usados os trajetos com o tempo de deslocação mais curto, sendo estes trajetos a fonte dos tempos de deslocação e das distâncias espaciais. No caso das deslocações em transporte público, foram selecionados os trajetos com o menor número de transbordos durante a deslocação. Os dados da mobilidade durante os meses de março e abril foram obtidos a partir do mapa disponível num artigo do jornal Expresso de 27 de junho de 2020, cuja fonte é Melo et al. (2021).

3.3. Manifestação do 1.º de Maio

Neste tópico explora-se a possibilidade de um evento público de grande escala, a Manifestação do 1.º de Maio, ter tido um efeito superdisseminador da COVID-19 de forma geograficamente heterogénea numa fase inicial da pandemia e explora-se o efeito de algumas covariáveis de cariz demográfico e de mobilidade.

No momento da realização do evento, estavam em vigor as primeiras medidas de confinamento do estado de emergência em Portugal, que limitavam a mobilidade entre concelhos e proibiam aglomerações públicas. A exceção concedida a esta manifestação gerou uma discussão pública intensa e politizada nos meios de comunicação social. Esse debate foi ainda intensificado pelo aumento dos novos casos diários de COVID-19 na região de Lisboa, em contraste com outras grandes regiões de Portugal, que apresentavam uma diminuição nos contágios diários (Figura 3.2).

Foram estas circunstâncias que motivaram esta investigação para desenvolver uma metodologia que permitisse extrair conclusões sobre os possíveis efeitos do evento no padrão de contágio na região de Lisboa, baseada em dados, de uma forma que pudesse ser aplicada a outros eventos.

A hipótese de partida é admitir a possibilidade de um evento ter um efeito superdisseminador sobre o contágio, mas não de forma homogénea na região circundante. A motivação para esta hipótese deriva da constatação de que existem diferentes contextos

de risco de contágio durante um evento potencialmente superdisseminador: a) durante o próprio evento, onde os efeitos nos concelhos de origem dos participantes não teriam uma conexão esperada; b) durante a viagem de ou para o evento, onde diferentes condições de contágio poderiam levar a comportamentos diferentes nos concelhos de origem dos participantes.

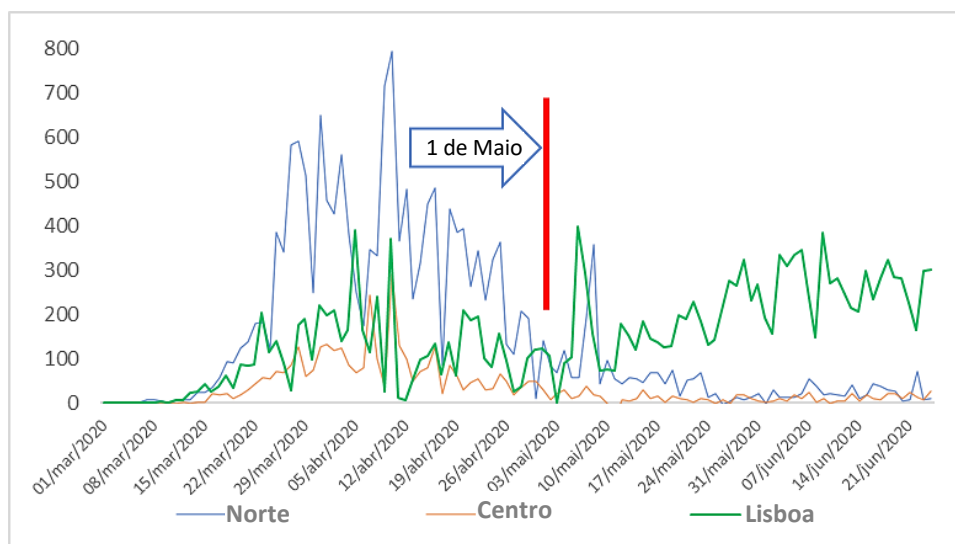


Figura 3.2: Novos casos diários em cada uma das três principais regiões de Portugal, com indicação de 1 de Maio, quando a manifestação foi realizada em Lisboa.

Assim, levantam-se duas questões, uma sobre o efeito do evento sobre a AML e a outra sobre se este efeito foi o mesmo em cada concelho da AML:

Q1: O evento não teve efeito na taxa de contágio na região de origem dos participantes, AML, em comparação com o resto do país?

Q2: O efeito do evento sobre a região de origem dos participantes, AML, não apresenta diferenças geográficas na taxa de contágio?

Tentando responder à primeira pergunta, **Q1**, procedeu-se à análise do comportamento do ritmo de contágio na AML e no resto do país, antes e depois da janela de potencial desenvolvimento dos efeitos de contágio do evento, para verificar se a relação entre estes ritmos tinha mudado ou não.

Para a segunda pergunta, **Q2**, analisou-se o comportamento dos ritmos de contágio antes e depois da referida janela do potencial desenvolvimento de efeitos, para cada concelho da AML.

A expectativa de que, entre os concelhos, o efeito sobre o ritmo de contágio da região de origem dos participantes ocorresse de uma forma geograficamente heterogénea, foi percebida por visualização gráfica. Ajustou-se um modelo duplo de Richards ao número cumulativo de casos COVID-19 relativos aos vários concelhos da AML, entre 15 de Abril e 6 de junho com recurso aos pacote `FlexParmCurve` do R (Oswald et al., 2012). Na análise de cada modelo foi possível observar, visualmente, comportamentos que indicavam uma potencial influência da manifestação do 1.º de Maio na evolução do número cumulativo de casos em alguns concelhos, enquanto noutros o evento não parece ter tido um impacto significativo (por exemplo, Figura 3.3a), Figura 3.3b).

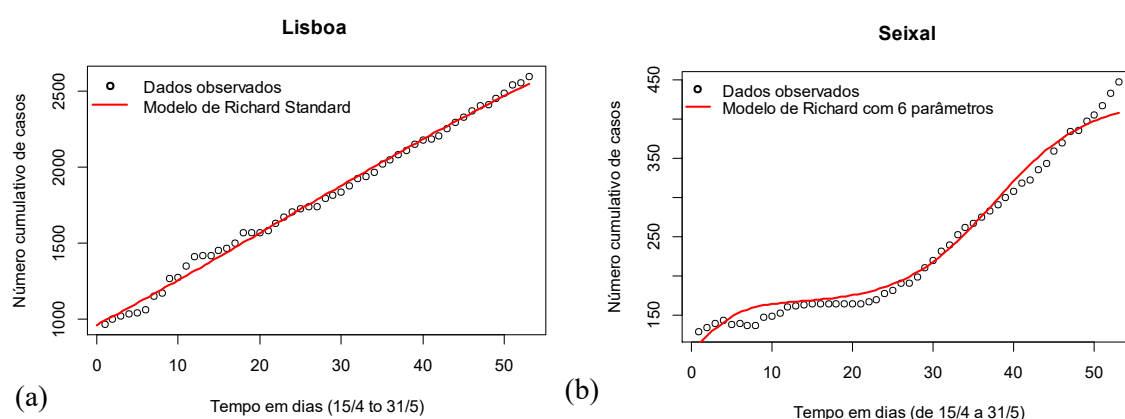


Figura 3.3: a) Modelo padrão de Richards com seis parâmetros, ajustado com número cumulativo de casos observado para o concelho de Lisboa entre 15/4 e 31/5; b) Modelo de Richards com seis parâmetros, ajustado com número cumulativo de casos observado para o concelho de Lisboa entre 15/4 e 31/5.

Para responder às questões **Q1** e **Q2**, explorou-se uma metodologia empírica baseada na modelação de regressão de Poisson, em testes de hipóteses e visualização gráfica.

A análise foi implementada com recurso a ferramentas de *software* R (R Development Core Team, 2010) e aplicada a dados correspondentes a dois períodos de tempo: um período S1 onde um caso de possível contágio no evento ainda não teria sido reportado, e um período S3 onde se esperaria que quase todos os casos de contágio desse tipo já teriam sido reportados. O período intermédio, S2, é aquele em que qualquer contágio relacionado com o evento estaria a ser reportado. Este período tem uma amplitude que corresponde ao tempo compreendido entre o contágio e o surgimento de sintomas – período de incubação acrescido de um período compreendido entre o surgimento dos sintomas e a inclusão do caso nos relatórios nacionais – *reporting delay*.

Para definir os períodos S1, S2, e S3 a partir de dados disponibilizados publicamente, foi necessário estimar o *reporting delay*. A metodologia adotada é explicitada na subsecção 3.3.2.

Definidos os períodos alvo para análise, o segundo requisito para a comparação dos padrões de contágio nos dois períodos é modelar o comportamento de contágio nesses períodos.

É expectável que o número de casos nos períodos em análise tenha um crescimento linear ou exponencial, dependendo do grau de transmissibilidade e sem pontos de inflexão nas curvas de crescimento, dado o curto período de tempo correspondente a cada segmento (menos de um mês). A observação das representações gráficas dos dados dos concelhos sustenta este raciocínio sobre as tendências de crescimento da transmissão em cada concelho e em cada um dos períodos em análise (Figura 3.5).

Durante o desenvolvimento desta tese foi feita uma abordagem preliminar com modelos de regressão linear e exponencial (publicada em Leal et al. (2023)). Uma evolução deste trabalho resultou numa abordagem com Modelos Lineares Generalizados. Uma vez que se trata de um problema com dados de contagem, esta abordagem representa a incerteza de forma mais adequada. Foram explorados modelos de regressão de Poisson e/ou Binomial Negativo, com funções de ligação *Identidade* e *Logarítmica* (modelos $Po(Id)$ e $Po(Log)$, ou $BN(Id)$ e $BN(Log)$, respetivamente), para cada período S1 e S3. A bondade de ajustamento global dos modelos foi avaliada com recurso ao teste F, ao nível de significância de 5%. A bondade do ajustamento dos modelos fez-se ainda com recurso ao Coeficiente de Determinação (R^2 ajustado) e à Raiz Quadrada do Erro Quadrático Médio (RMSE - *Root Mean Square Error*). A seleção do modelo com melhor bondade de ajustamento foi feita com base no critério AIC.

A metodologia proposta para avaliar se houve, ou não, alteração nos padrões de crescimento do número de casos de COVID-19 nos concelhos da AML, após a manifestação do 1º. de Maio, implementou-se em quatro etapas, com recurso a modelos de regressão de Poisson e/ou Binomial Negativa.

Numa primeira etapa ajustaram-se modelos $Po(Id)$ e/ou $BN(Id)$ que relacionam o *Número cumulativo de casos em cada concelho* com o *Tempo t, em dias, que decorreu desde 15 de abril de 2020* e utilizou-se o teste ANCOVA *One-Way* (*Anova Tipo III*) com interação para testar, ao nível de significância de 5%, a homogeneidade dos coeficientes do

preditor linear dos modelos ajustados aos dados dos períodos S1 e S3. Os concelhos em que a hipótese de homogeneidade das taxas de variação do número de casos diários nos períodos S1 e S3 não foi rejeitada (homogeneidade dos coeficientes de t dos modelos $Po(Id)/BN(Id)$) e, portanto, não foi rejeitada a hipótese de manutenção do padrão de crescimento após o evento, foram incluídos no grupo de concelhos para os quais o padrão de crescimento de casos não se alterou de S1 para S3.

Numa segunda etapa, analisou-se a representação gráfica dos modelos para os concelhos em que se rejeitou a homogeneidade do padrão de crescimento do período S1 para o período S3, com base nos modelos $Po(Id)/BN(Id)$, e compararam-se as inclinações das retas, bem como os valores dos declives, para identificar concelhos onde terá havido uma tendência de diminuição da taxa de crescimento ou onde o crescimento foi um artefacto, meramente devido a um aumento súbito do número de casos que ocorreu no período intermédio S2, após o qual o declive voltou, no modelo ajustado a S3, a valores muito próximos do declive do modelo ajustado a S1. Estes concelhos foram também atribuídos à categoria de locais onde provavelmente não se verificaram efeitos do evento.

Na terceira etapa desta análise, explorou-se a possibilidade de que, nos concelhos para os quais se rejeitou a hipótese de homogeneidade dos declives (em $Po(Id)/BN(Id)$) e terá havido um aumento da taxa de crescimento linear de S1 para S3, este aumento fosse apenas o que seria o crescimento esperado de um crescimento exponencial natural de contágio, já com origem anterior ao evento e não por efeito mesmo. Para tal, foram ajustados modelos de regressão de Poisson/Binomial Negativa com função de ligação logarítmica ($Po(Log)/BN(Log)$) entre a variável *Número cumulativo de casos* e *Tempo (t), em dias*, decorrido desde o dia 15/4/2020 e foi testada a hipótese de homogeneidade dos coeficientes de t dos modelos recorrendo ao teste ANCOVA *One Way (Anova Tipo III)* e ao teste *Likelihood Ratio (LRT)*. Os concelhos onde, mesmo com os modelos $Po(Log)$ ($BN(Log)$), a hipótese de homogeneidade foi rejeitada, foram atribuídos à categoria de concelhos onde pode ter havido efeito do evento no padrão de crescimento da COVID-19, devido à aceleração do crescimento do número de casos.

Na quarta e última etapa analisaram-se os concelhos para os quais se rejeitou a hipótese de não alteração do padrão de crescimento linear e não se rejeitou a hipótese de não alteração do padrão de crescimento exponencial, ao nível de significância de 5% (hipótese da homogeneidade dos coeficientes de t dos modelos $Po(Id)$ ($BN(Id)$) e

$Po(Log)$ ($BN(Log)$), respetivamente). Com esta análise pretendeu-se verificar se não rejeitar a hipótese de não alteração do padrão de crescimento exponencial do período S1 para S3 se devia à manutenção de um padrão de crescimento exponencial que já ocorria no período S1 ou se este “crescimento exponencial” não era mais do que um crescimento linear acentuado que já estava a ocorrer antes de S1 e continuou em S3, mas mais acentuado. Para isto, verificou-se qual dos modelos ajustados à série temporal completa tinha melhor bondade de ajustamento: $Po(Id)$ ($BN(Id)$) ou $Po(Log)$ ($BN(Log)$). No caso de ser o modelo de crescimento linear, antecipar-se-ia o início do período S1 para ajustar modelos $Po(Id)$ ($BN(Id)$) e $Po(Log)$ ($BN(Log)$) com os dados do concelho para este novo período S1 e selecionar-se-ia o modelo com melhor bondade de ajustamento, decidindo com base nos resultados a qual grupo atribuir este concelho.

A metodologia encontra-se sistematizada no fluxograma da Figura 3.4.

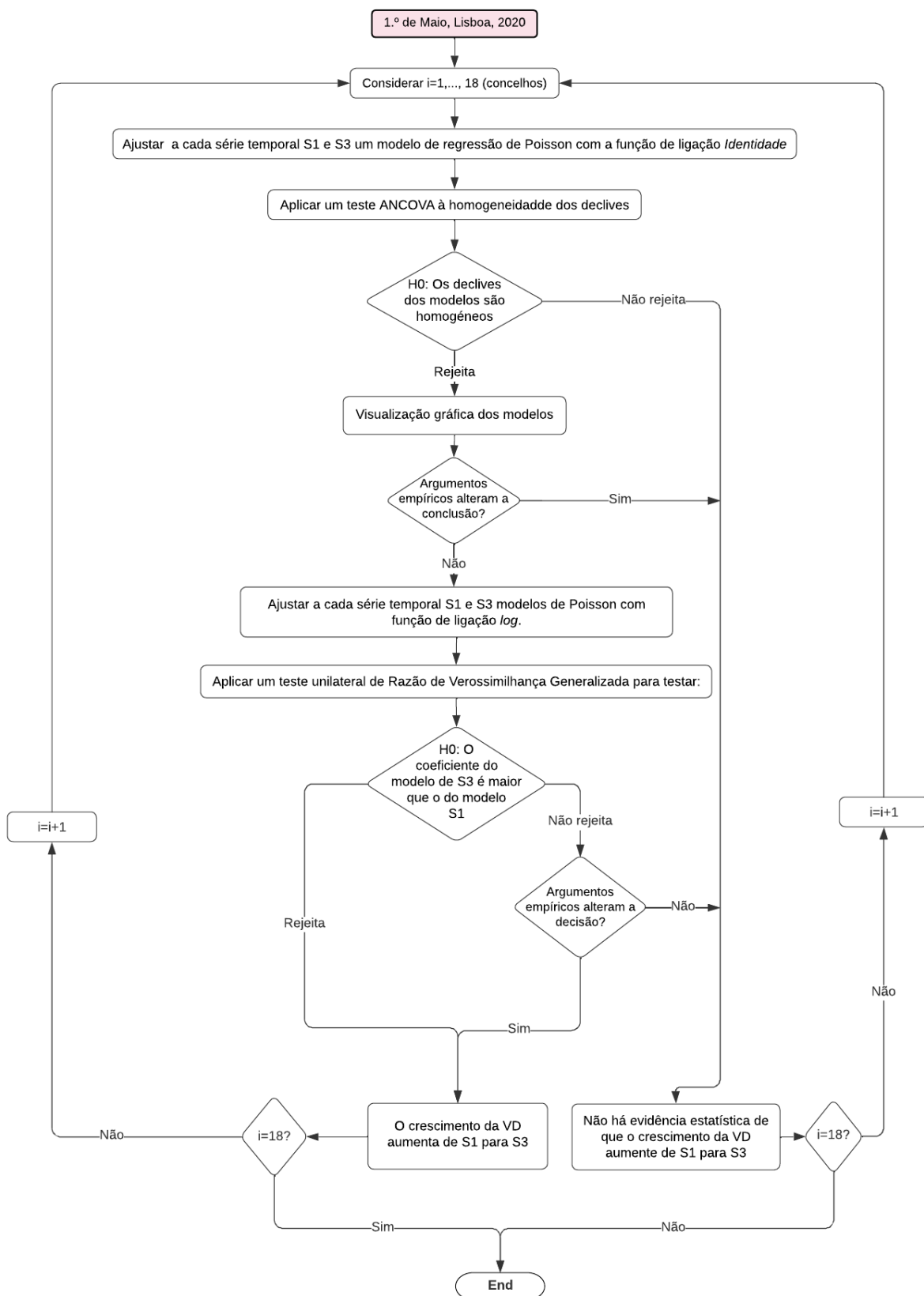


Figura 3.4: Fluxograma que ilustra a metodologia implementada no artigo para a análise de dados dos concelhos da AML.

3.3.1. Caracterização dos dados

Para a análise do potencial impacto da manifestação do 1º. de Maio em 2020, em Lisboa, foram usados os dados da COVID-19 de 24 de março a 6 de junho de 2020 nos concelhos da AML. Foram consideradas as variáveis “Número cumulativo de casos COVID-19” e “Número cumulativo de casos COVID-19 por 100 000 habitantes”, em Portugal, na AML e nos concelhos da AML e a variável t : “Número de dias desde o dia 15 de abril de 2020”.

A Figura 3.5 mostra a evolução do número de casos por 100 000 habitantes, observada no período de 15 de abril a 6 de junho de 2020, em cada um dos concelhos da AML. A linha a preto representa o número de casos por 100 000 habitantes na AML. Pode-se observar que há concelhos em que o padrão de crescimento parece manter-se estável ao longo de todo o período em análise enquanto noutros se observa uma aceleração na taxa de crescimento do número de casos por 100 000 habitantes.

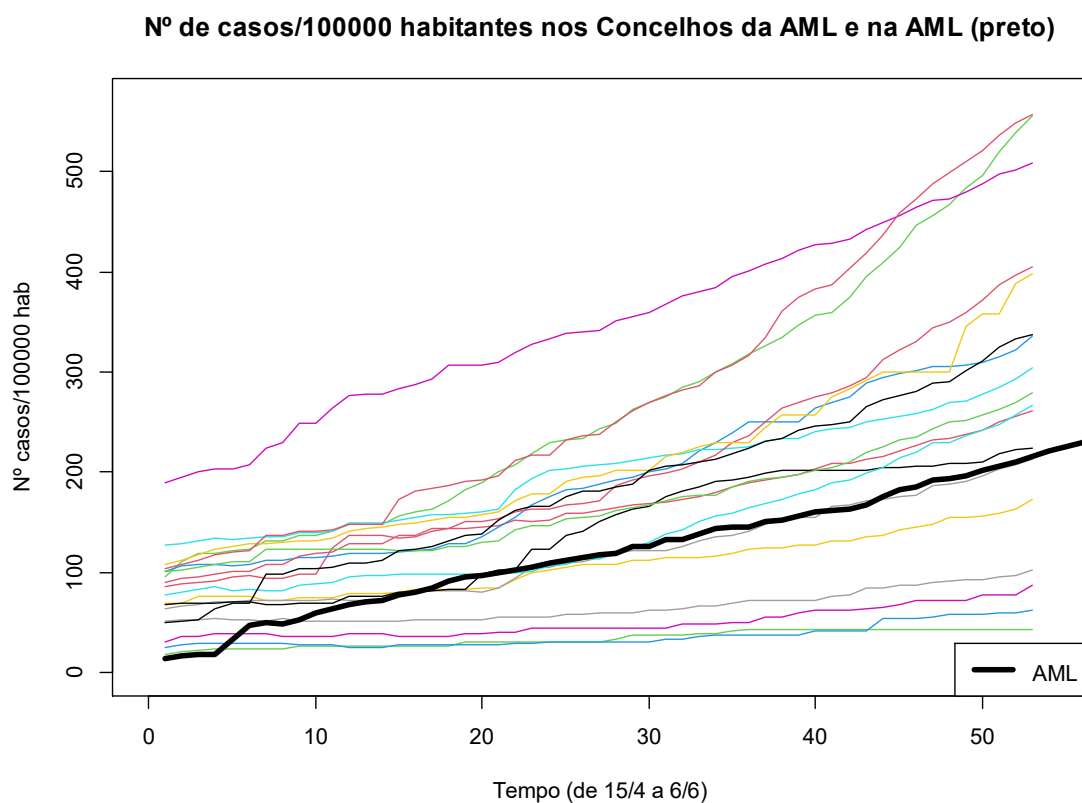


Figura 3.5: Evolução do número de casos por 100 000 habitantes observado no período de 15 de abril a 6 de junho de 2020 em cada um dos concelhos da AML. A curva a preto representa a evolução do número de casos por 100 000 habitantes na AML

A AML é uma região constituída pelos concelhos: Alcochete, Almada, Amadora, Barreiro, Cascais, Lisboa, Loures, Mafra, Moita, Montijo, Odivelas, Oeiras, Palmela, Seixal, Sesimbra, Setúbal, Sintra, e Vila Franca de Xira (*Lei n.º 75/2013*, 2013, p. 75).

Em relação a estes concelhos foram consideradas as covariáveis “Índice de Envelhecimento”, “Densidade populacional”, “Índice de mobilidade para Lisboa”, “Tempo de viagem de carro para Lisboa”, “Distância de carro para Lisboa”. As sínteses descritivas destes dados encontram-se no Anexo II.

Os gráficos da Figura 3.6 ilustram a magnitude dos valores destas covariáveis, por concelho.

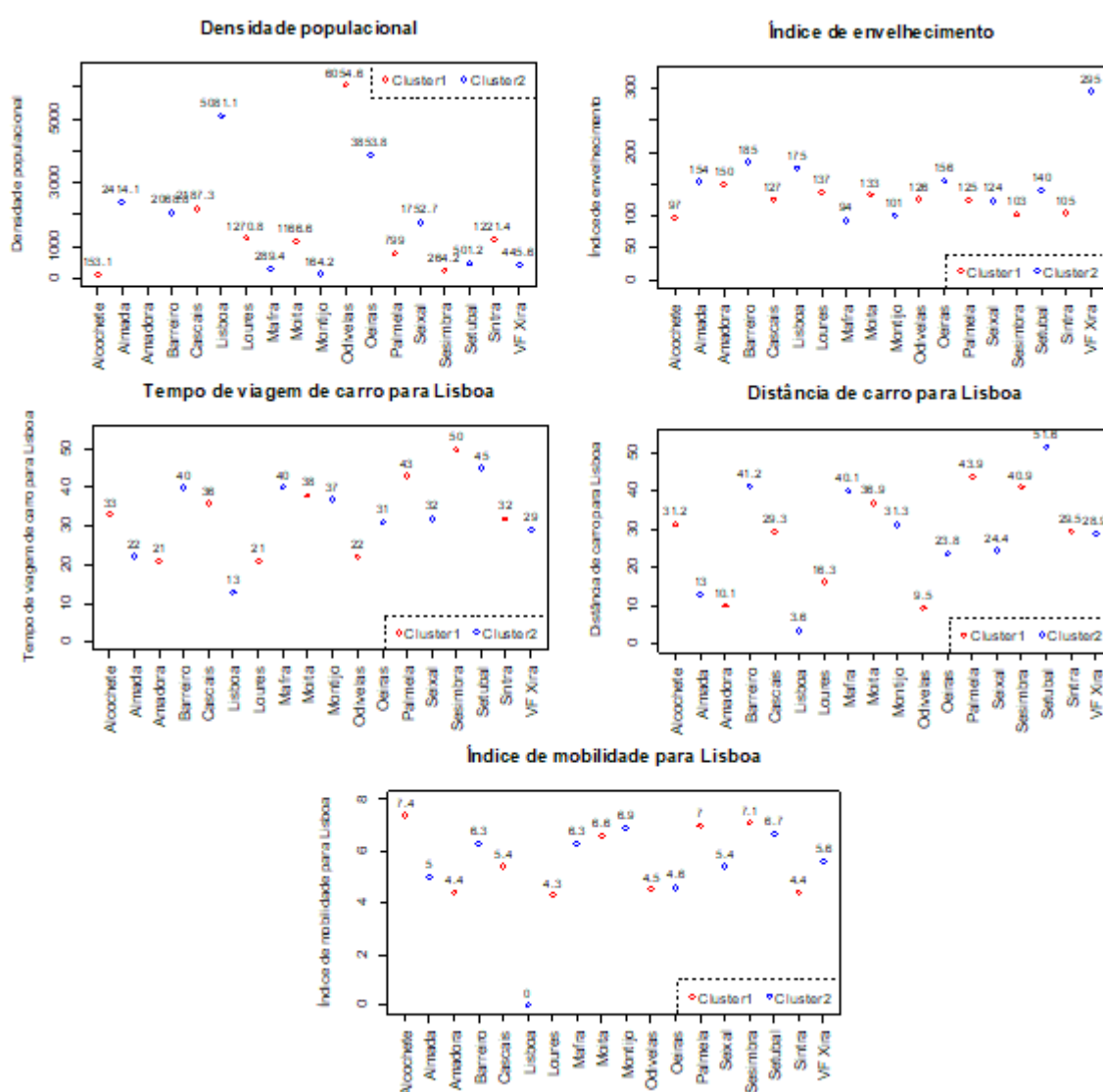


Figura 3.6: Valores das covariáveis, por concelho: *Densidade populacional*, *Índice de envelhecimento*, *Distância do carro até ao local de demonstração em Lisboa*, *Tempo de viagem até ao local de demonstração em Lisboa*, e *Índice de mobilidade até Lisboa*.

Na Figura 3.7 encontra-se a distribuição empírica do número cumulativo de casos COVID-19 nos diferentes concelhos da AML. A densidade refere-se à proporção de dias, entre 15 de abril e 6 de junho de 2020, em que se manteve até um número x de casos acumulados.

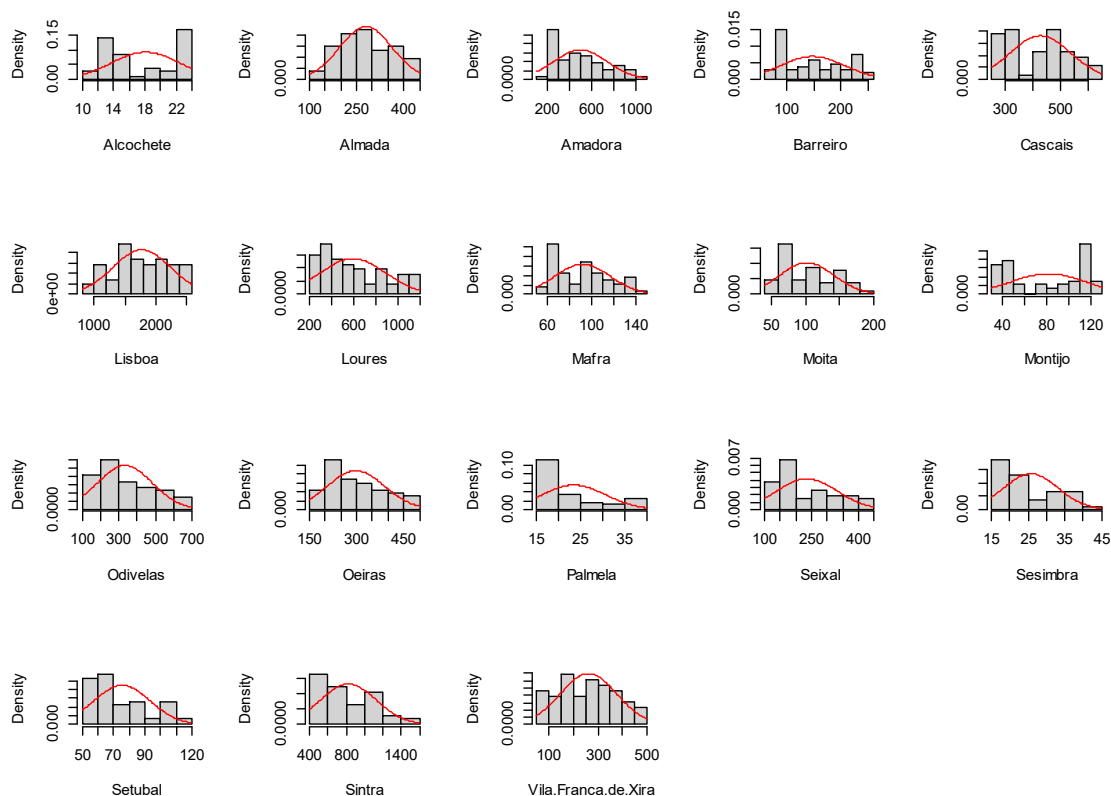


Figura 3.7: Representação gráfica da distribuição empírica do *Número cumulativo de casos*, em relação ao número de dias entre 15/4 e 6/6 de 2020.

Como mostra a Figura 3.8 estes concelhos são heterogéneos no que diz respeito às variáveis demográficas: *densidade populacional* de 151 a 5062 habitantes/ km^2 , *índice de envelhecimento* de 94 a 185 idosos por 100 jovens.

Lisboa, a capital do país e principal cidade da região, é a localização centrípeta das redes de transporte. A localização da Manifestação de 1.º de Maio é bastante central dentro da própria Lisboa, como mostra a Figura 3.1.

Dentro de Lisboa, os terminais de transporte regionais incluem estações de metro, e a rede de metro serve o local da manifestação. Os únicos concelhos fora de Lisboa servidos pela rede de metro são Amadora e Odivelas. A mobilidade é limitada pelo estuário do rio Tejo, atravessado por apenas duas pontes, de Lisboa ao Montijo (apenas rodoviário) e de

Lisboa a Almada (rodoviário e ferroviário). Existem serviços de barco de passageiros entre Lisboa e três concelhos da margem sul (Almada, Montijo e Barreiro). A limitação do estuário tem impacto no índice de mobilidade, que é mais elevado a sul do mesmo, bem como na distância e tempo de viagem de automóvel ou transportes públicos.

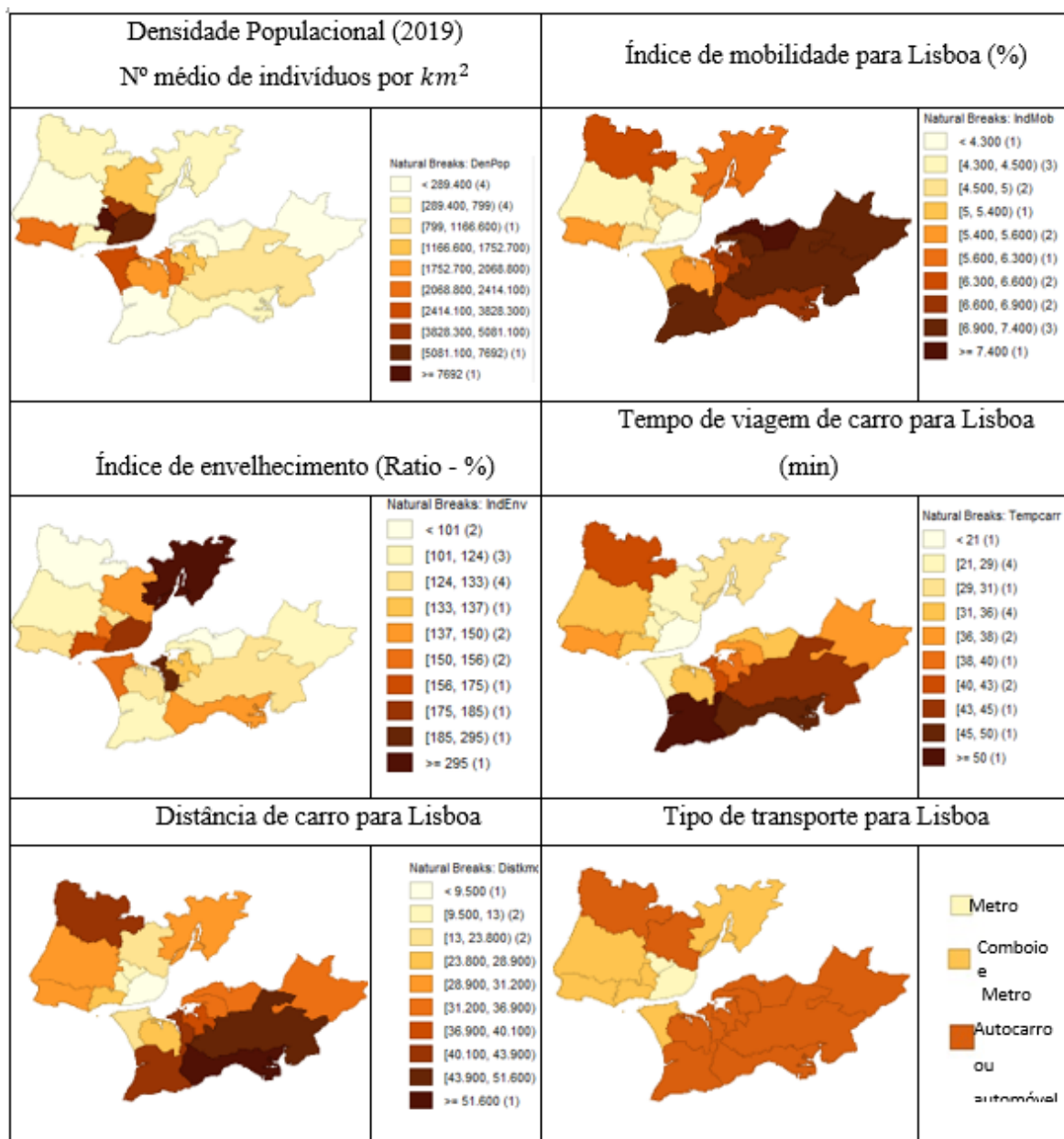


Figura 3.8: Distribuição geográfica, por concelhos, de covariáveis, maximizando as semelhanças internas (Anselin et al., 2010).

Ao considerar as opções de transporte público disponíveis entre os concelhos e o local da manifestação, o tempo estimado de viagem num feriado, a conveniência do transporte, e

as declarações públicas sobre o transporte combinado individual e de autocarro que os manifestantes usaram para a manifestação (DIF/CGTP-IN, 2020), agruparam-se os concelhos pela forma de transporte estimada.

Grupo 1: Os manifestantes com opções convenientes de transporte público combinando comboio e metro em menos de 1 hora, presume-se que tenham escolhido estes meios de transporte. Manifestantes de Lisboa, Odivelas, Amadora, Sintra, Oeiras, Almada, Cascais, Vila Franca de Xira;

Grupo 2: Pressupõe-se que os manifestantes sem essas opções tenham viajado em veículos particulares ou em autocarros alugados. Manifestantes de Loures, Seixal, Barreiro, Mafra, Moita, Setúbal, Montijo, Palmela, Sesimbra, Alcochete.

É expectável que os dados relativos ao número cumulativo de casos não se ajustem à distribuição normal. Para testar esta suposição para cada concelho representaram-se os histogramas (Figura 3.7) e aplicaram-se testes à normalidade (Shapiro-wilks e Anderson-Darling) (Korkmaz et al., 2014) e verificou-se que apenas nos concelhos de Almada, Lisboa e Vila Franca de Xira os dados se ajustam a uma distribuição normal (Anexo I). As estatísticas descritivas destes dados encontram-se em anexo (Anexo II).

A base de dados usada nesta secção contempla as variáveis que constam na Tabela 3.1.

Tabela 3.1: Variáveis em análise.

Descrição	Variável	Classificação da variável
<i>Data</i>	<i>Dia/Mês/Ano</i>	Ordinal
<i>Tempo decorrido após 15 de abril de 2020</i>	<i>Nº de dias</i>	Escala
<i>Dados sobre casos COVID-19</i>	<i>Nº cumulativo de casos</i>	Contagem
<i>Característica da População dos concelhos da AML</i>	<i>População (2019) (Nº de habitantes)</i>	Contagem
	<i>Densidade populacional (2019) (Nº médio de indivíduos por km²)</i>	Escala
	<i>Índice de envelhecimento</i>	Escala
<i>Mobilidade dos concelhos da AML para Lisboa</i>	<i>Índice de mobilidade</i>	Escala
	<i>Distância de carro para Lisboa</i>	Escala
	<i>Tempo de viagem de carro para Lisboa</i>	Escala
<i>Meio de transporte usado para a deslocação para o evento</i>	<i>Meio de transporte</i>	Nominal

3.3.2. Estimação do *Reporting Delay*

Para definir o período S2 (do último dia após 1.º de Maio garantidamente sem impacto dos casos de contágio ocorridos durante esse dia, e o primeiro dia em que esse impacto seja garantido) é necessário usar uma estimativa do período de incubação da COVID-19 e do *reporting delay* pois S2 é o período do contágio até ao diagnóstico mais o tempo para a inclusão de casos positivos nos relatórios diários da DGS.

Numa revisão de literatura sistemática e meta-análise Wassie et al. (2020) reuniram os dados de 18 estudos, representando 22 595 participantes, e estimaram um período de incubação médio combinado de SARS-CoV-2 de 5.7 dias (IC de 95%,]5.1; 6.4[). Lauer et al. (2020) estimaram o período de incubação mediano em "5.1 dias (IC a 95%: 4,5 – 5,8 dias)" e o valor médio de 5.5 dias. Este pressuposto é validado pela OMS que considera o período de incubação como um valor no intervalo de 5 a 6 dias, de acordo com dados da literatura. Nesta tese tomou-se a estimativa do período de incubação obtida por Lauer et al. (2020).

O *reporting delay* foi modelado por Kucharski et al. (2020) para casos de transmissão em Wuhan como sendo exponencialmente distribuído com uma média de 6,1 dias. A fim de estimar o *reporting delay* Abbott (2020) utilizou todos os casos de uma *linelist* (`covidregionaldata::get_linelist()` – R *software*)⁸, que estava disponível publicamente para mais de 100 países, com as datas de início dos sintomas e respetivas datas de notificação, ao longo de um período de 12 semanas. Ajustou uma distribuição log-normal para 100 subamostras de dados disponíveis (censuradas à direita e à esquerda) e obteve um *reporting delay* de 6,5 dias e um desvio padrão de 17 dias. Nesta lista, houve apenas dois casos em Portugal com informação sobre o início dos sintomas, o que torna a representatividade do modelo para Portugal quase nula. De Nicola et al. (2022) e Greene et al. (2020) estimaram a distribuição do *reporting delay* para desenvolver modelos de previsão de curto prazo (*nowcasting*) de casos COVID-19.

Nesta tese o *reporting delay* será estimado utilizando dados do número de casos em Portugal extraídos de gráficos dos relatórios diários da DGS, uma vez que não há dados quantitativos que permitam modelar esta variável, como acontece, por exemplo, na Alemanha (De Nicola et al., 2022). A extração de dados será feita usando a metodologia descrita no enquadramento teórico (2.9).

⁸ <https://www.rdocumentation.org/packages/covidregionaldata/versions/0.8.2> consultado em janeiro 2021

Como já referido, em Portugal não estão publicamente disponíveis dados numéricos sobre o número de casos por dia de início dos sintomas ou de notificação, mas apenas dados visuais constantes nos relatórios diários da DGS na forma de gráfico de barras empilhadas (Figura 3.9). As barras representam o número de casos confirmados por data de início dos sintomas (vermelho) a que acumula o número de casos por data de diagnóstico/notificação (azul). Estes últimos, apresentados nos relatórios como casos em que a data de aparecimento do sintoma é desconhecida, presumivelmente incluirão ou falhas no registo dos sintomas ou casos assintomáticos/pré-sintomáticos detetados no momento dos testes.

O número de casos entre 17 de Abril e 9 de Maio de 2020 foi extraído destes dados visuais: número total de casos, número de casos por data de início dos sintomas e número de casos por data de notificação. Esta extração ocorreu por medição gráfica, através do processo descrito na secção 2.9. A medição obteve o número de pixels da altura de cada barra (indicando o número total de casos), composto por duas barras empilhadas: vermelha (casos por data de início dos sintomas) e azul (casos por data de notificação). Foi também obtido o número de pixels na linha do eixo vertical, entre as marcas da escala 0 e 600 (valor A). Estes limites da escala foram escolhidos por serem os mais distantes e fiáveis encontrados em todos os gráficos diários: por vezes as únicas marcas na escala eram 0-200-400-600. Consequentemente, o número de casos diários em cada barra (variável C_{bar}) é:

$$C_{bar} = \frac{600 \text{ (altura da barra)}}{A} \quad (3.1)$$

Os gráficos fornecidos publicamente pela DGS estavam num formato com distorção visual (Figura 3.9 b), conferindo à medição de cada coordenada uma incerteza de $-2/+2$ pixels. Desta forma, o erro de medição da escala e da altura das barras, acumulando os erros extremos, tem uma incerteza de $-4/+4$ pixels. As alturas mínima e máxima das barras medidas em pixels entre as marcas da escala 0 e 600 foram de 210 e 296. As alturas mínima e máxima das barras, medidas em pixels, foram de 30 e 268.

Por conseguinte, pela equação (2.119):

$$\sigma_c = 600 \times 4 \sqrt{\frac{a^2 + b^2}{a^4}}, \quad (3.2)$$

para $210 < a < 296$ e $30 < b < 268$

Assim, estimando os extremos locais (Wolfram Research, 2020), conclui-se que:

$$8 < \sigma_c < 19 \quad (3.3)$$

A fim de obter a altura de cada barra empilhada, é necessário determinar após quantos dias as barras vermelhas se estabilizam (casos por data de início dos sintomas). Isto porque, à medida que os doentes apresentam sintomas e são diagnosticados, esses casos serão reportados aos correspondentes dias anteriores, levando a que a altura das barras mude de dia para dia à medida que os casos são comunicados. O número de casos reportados a dias anteriores de início de sintomas tende para zero à medida que o tempo passa.

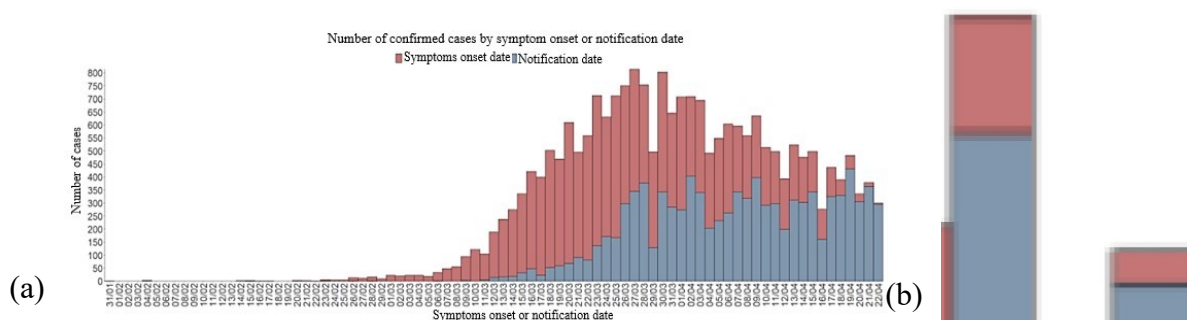


Figura 3.9: Dados visuais fornecidos publicamente através de relatórios diários da DGS. (a) Amostra de gráfico dos relatórios diários (Relatório 52)⁹ (b) Amostra de pixels amplificados, mostrando distorção.

A taxa de variação w do número de casos com início de sintomas no dia d_i , $i = 1, 2, \dots$, que entram no relatório da DGS num determinado dia j , em relação ao número de casos com início de sintomas no dia i e reportados no dia $j - 1$ é dada por

$$w = \frac{d_{ji} - d_{(j-1)i}}{d_{(j-1)i}}, i = 1, 2, \dots \quad (3.4)$$

sendo d_{ji} o número de casos notificados no dia $j = 1, 2, \dots$ com início de sintomas no dia i .

Para estimar o número de dias X até que d_i tenha estabilizado no dia t assumiu-se para w o valor alvo 0. Admitindo a margem de erro de leitura de dados anteriormente referida, assumiu-se para w o intervalo a partir do qual se considera que d_i estabilizou: $]L, U[=] - 0.036; 0.036[$, ou seja, um desvio de 0,012. Consideraram-se gráficos de Shewhart, com $k = 3$, como habitualmente nestes gráficos.

O desvio considerado origina limites de especificação nos gráficos que dão uma margem de erro de aproximadamente 4% para w , assumindo que a margem de erro de leitura é de $\pm 2\%$, a partir do momento em que a altura das barras para o número de casos reportados por dia de início de sintomas começa a estabilizar.

⁹ <https://covid19.min-saude.pt/relatorio-de-situacao/> - consultado em novembro 2020

Utilizando o pacote `qcc` de R (Scrucca, 2004) elaborou-se, para cada dia de 17 de Abril a 9 de Maio, um gráfico de Shewhart (por exemplo, Figura 3.10a,b) e c)).

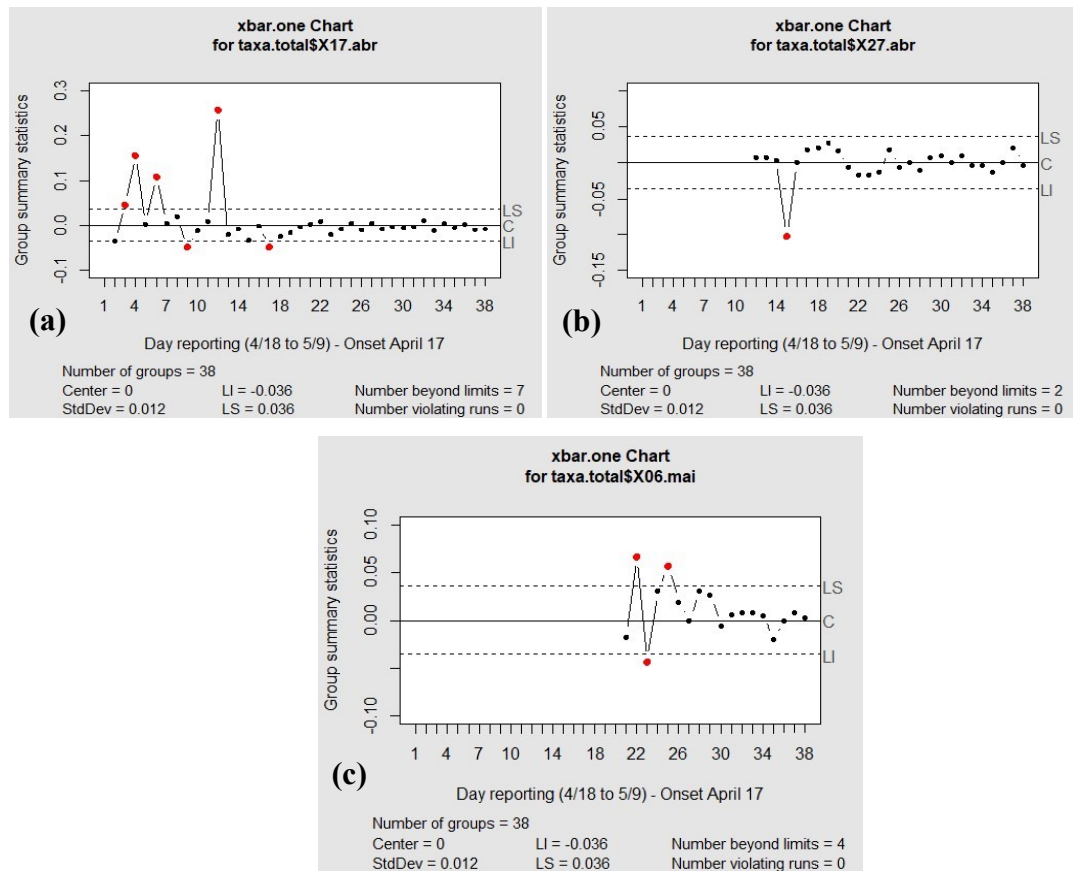


Figura 3.10: Os gráficos de Shewhart para taxas de variação diárias do número de casos notificados (w) com início de sintomas em (a) 17 de Abril, (b) 27 de Abril, e (c) 6 de Maio.

Seja X a variável que representa o número de dias até que todos os valores de w estejam dentro do intervalo $]L, U[$. Entre os dias 17 de abril e 9 de Maio de 2020 recolheram-se os valores de X e, para estimar o *reporting delay* médio com a incerteza apropriada, ajustou-se a estes dados, por Máxima Verosimilhança, a distribuição Gamma e a distribuição Lognormal, recorrendo ao pacote `fitdistribplus` do *software* R (Delignette-Muller & Dutang, 2015) (Figura 3.11). As distribuições Gama e LogNormal são usadas na literatura para modelar a distribuição do *reporting delay* (Marinović et al., 2015; Oliveira, 2018; Tariq et al., 2019). A distribuição com melhor ajustamento por Log-Verossimilhança pelos critérios AIC e BIC foi a Gama (Tabela 3.2) e, portanto, será a distribuição utilizada neste estudo, o que resulta numa distribuição do *reporting delay* com média de 4.1 dias e desvio

padrão de 1.17 dias. A estimativa *bootstrap* da mediana é de 3 dias com IC a 95%]1.7; 5.96[(Delignette-Muller & Dutang, 2015).

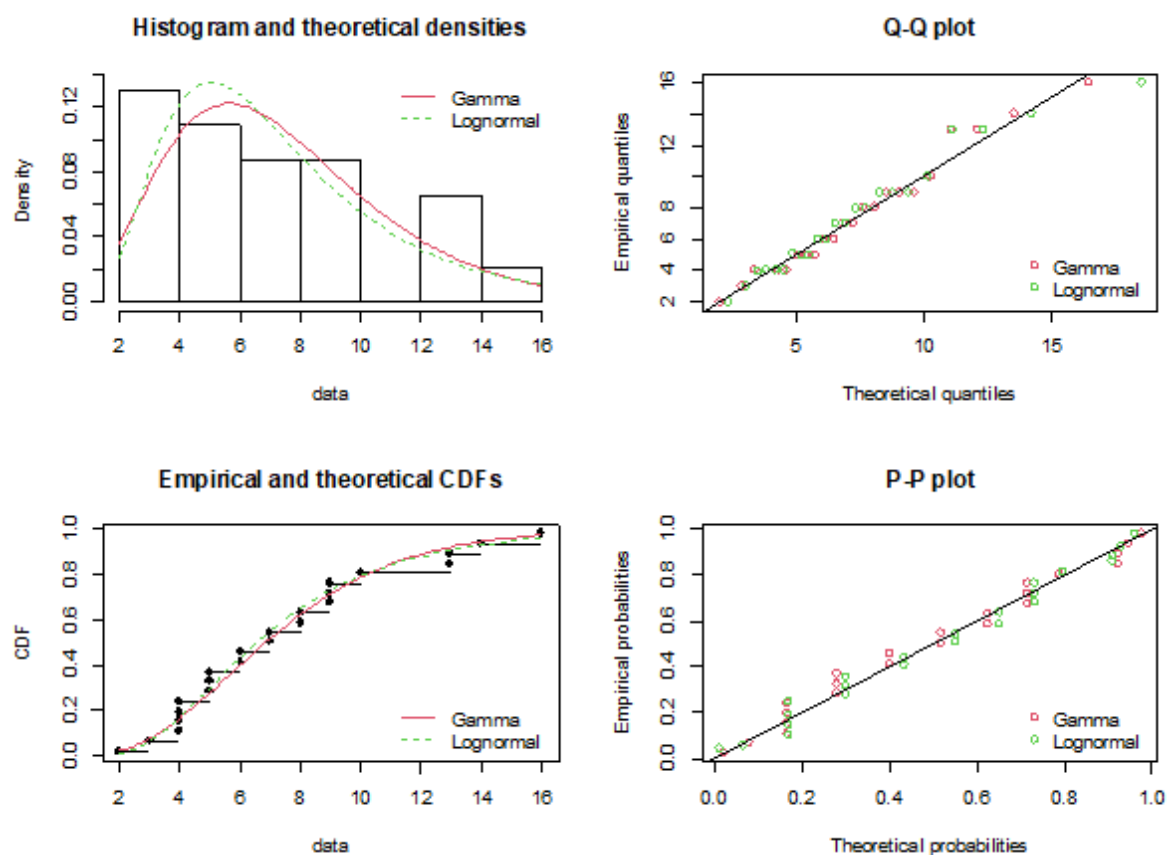


Figura 3.11: As funções PDF e CDF ajustadas aos dados de atraso de comunicação (*reporting delay*).

Tabela 3.2: Critérios de bondade de ajustamento.

Criterion	<i>Log-likelihood</i>	AIC	BIC
<i>Gamma</i>	-60.52812	125.0562	127.3272
<i>LogNormal</i>	-60.62544	125.2509	127.5219

O valor médio de 4.1 dias obtido com a distribuição Gama é sustentado, para Portugal, no raciocínio seguinte. Para uma pessoa com sintomas procurar os serviços de saúde e fazer um teste, seria necessário um dia; o resultado do teste deveria estar disponível nesse mesmo dia ou até dois dias mais tarde, uma vez que, na altura, Portugal dependia exclusivamente de testes de rRT-PCR, não tendo disponíveis testes de diagnóstico rápido (Orientação 015, 2020). E que, uma vez feito o diagnóstico, espera-se que o indivíduo seja incluído no relatório das estatísticas nacionais no dia seguinte. Por conseguinte, desde o início dos sintomas, espera-se um atraso de 4 dias no relatório. Esta fundamentação para além de ser

corroborada pela análise dos relatórios de dados reais, detalhados acima, foi também sustentada numa afirmação da Diretora-Geral da Saúde Portuguesa, que a 19 de Outubro de 2020, numa conferência de imprensa regular sobre a pandemia da COVID-19 em Portugal, mencionou um período de 3 dias entre o aparecimento dos sintomas e o diagnóstico, o que resulta num período médio de 4 dias desde o aparecimento dos sintomas e a notificação dos casos em relatórios diários da DGS.

Assumindo que o período de incubação mediano é de 5,1 dias, com intervalo de confiança de 4.5 a 5.8 dias (Lauer et al., 2020); que o *reporting delay* médio de 4,1 dias e mediano de 3 dias, com intervalo de confiança a 95%]1.7; 6[; que menos de 2,5% das pessoas infetadas apresentarão sintomas dentro de 2.2 dias após a infeção (Lauer et al., 2020) e que um número extremamente pequeno de casos de infeção ocorrida no dia 1 de Maio terá sido notificado até 6 de Maio; assumindo, ainda segundo Lauer et al. (2020), que 97.5% dos casos apresentarão sintomas no prazo de 11.5 dias após a infeção, assumiu-se que S2 decorre de 7 a 15 de Maio.

Assim, a série temporal S , do *Número cumulativo de casos* de 15 de Abril a 6 de Junho de 2020, foi dividida em três segmentos: S1 de 15 de abril a 6 de maio; S2 de 7 de maio a 15 de maio e S3 de 16 de maio a 6 de junho. Neste último período assumiu-se que já foram notificados praticamente todos os casos de possível contágio a 1 de maio.

3.3.3. Modelação e análise com modelos de regressão de Poisson

3.3.3.1. Modelação e análise da taxa de variação do número de casos diários na AML e no resto do país

Para comparar o padrão de crescimento do número cumulativo de casos diários na AML e no resto do país, foram definidas as seguintes hipóteses:

H_0 : Não houve alteração na taxa de contágio na região de origem dos participantes na manifestação, em comparação com o resto do país.

H_1 : Houve alteração na taxa de contágio na região de origem dos participantes da manifestação, em comparação com o resto do país.

A fim de estimar a taxa de variação do número cumulativo de casos para os períodos S1 e S3 ajustou-se aos dados $S1 \cup S3$, da AML e do resto do país, um modelo de regressão de

Poisson com função de ligação *Identidade*, com interação entre a variável t (*Tempo*, em dias, a partir de 15/4/2020 e até 6/6/2020) e uma variável *Dummy* que assume o valor 0 para o período S1 e o valor 1 para o período S3 (modelo $Po(Id)$). O período S2 (intermédio) não foi considerado na análise, uma vez que esta se focou na comparação entre os períodos S1 e S3, ou seja, período sem potenciais efeitos do evento e período com potenciais efeitos totais, respetivamente.

$$\begin{aligned}
 Po(Id) : g(E(NCasos)) &= \beta_0 + \beta_1 t + \beta_2 Dummy + \gamma Dummy \times t \\
 g(E(NCasos)) &= E(NCasos) \quad (\text{Função de ligação identidade}) \\
 NCasos &\sim Po(g(E(NCasos)))
 \end{aligned} \tag{3.5}$$

onde t é o tempo, em dias, entre os dias 15 de Abril e 6 de junho de 2020, $g(\cdot)$ designa a função de ligação e $E(\cdot)$ é o valor esperado.

Para testar a homogeneidade dos declives dos modelos $Po(Id)$ para S1 e $Po(Id)$ para S3, ajustados com os dados AML e do resto do país, implementou-se um teste ANCOVA *One-Way (Anova Tipo III)* recorrendo ao pacote `car` do R (Fox & Weisberg, 2019), para cada uma destas duas regiões. Neste caso, interessa testar o termo de interação para avaliar se o coeficiente de interação entre a variável t e os dois momentos temporais em análise, S1 e S3, é estatisticamente significativo (Field et al., 2012). Se for estatisticamente significativo, então a diferença entre os declives da curva ajustada aos dados S1 e da curva ajustada aos dados S3 é estatisticamente significativa.

As hipóteses a testar foram:

$$\begin{aligned}
 H_{0ANC} : \gamma &= 0 \\
 H_{1ANC} : \gamma &\neq 0
 \end{aligned} \tag{3.6}$$

Como, para as duas regiões, se obteve um $p\text{-value} < 2,2e - 16$, pode-se concluir que a diferença entre os coeficientes dos modelos de Poisson com função de ligação *Identidade*, que traduz um crescimento linear, é estatisticamente significativa, com nível de significância de 5%. Conclui-se, assim, que o ritmo de crescimento mudou nas duas regiões, do período S1 para o S3. Por outras palavras, o ritmo de contágio, tanto na AML como no resto do país, mudou do período S1 para S3, numa perspetiva de crescimento linear.

Comparando os valores dos coeficientes dos modelos $Po(Id)$ (Tabela 3.3) constata-se que os da AML aumentaram significativamente de S1 para S3 (o ritmo de crescimento do número de casos acelerou) enquanto no resto do país reduziram (abrandando a taxa de

crescimento do número de casos), tornando assim clara a diferença no padrão de crescimento de um período para o outro em ambos os casos.

Tabela 3.3: Coeficientes dos modelos de Poisson com função de ligação de *Identidade* ajustados aos *dados da AML e do resto do país*.

	Modelos de Poisson (Função de ligação <i>Identidade</i>)	
	Po(Id) (S1)	Po(Id) (S2)
Portugal Continental (sem os concelhos da AML)	290.157	74.841
AML	93.60	186.715

Ajustou-se ainda, para cada região, um modelo de regressão de Poisson com a função de ligação *log*, que se designou por $Po(Log)$ e com interação $DuMMY \times t$, aos mesmos conjuntos de dados.

$$\begin{aligned}
 Po(Log) : g(E(NCasos)) &= \beta_0 + \beta_{1t} + \beta_2 Dummy + \gamma Dummy \times t \\
 g(E(NCasos)) &= \ln(E(NCasos)) \quad (\text{Função de ligação } log) \\
 NCasos &\sim Po(g(E(NCasos)))
 \end{aligned} \tag{3.7}$$

Recorrendo à função `check_overdispersion()` do pacote `performance` (Lüdecke et al., 2021) do R, testou-se a superdispersão nos modelos de regressão de Poisson ajustados para os dois casos: AML e resto do país (Anexo III). Para o resto do país verificou-se haver superdispersão, pelo que se ajustou um modelo Binomial Negativo com funções de ligação *log* (designado $BN(Log)$) e *Identidade* (designado $BN(Id)$). Recorreu-se à função `glm.nb()` do pacote `MASS` (Venables & Ripley, 2002). O modelo que registou melhor bondade de ajustamento entre todos os modelos ajustados com este conjunto de dados foi o $BN(Id)$ (menor valor de AIC). No caso da AML o modelo não revelou superdispersão, pelo que se manteve o modelo de Poisson.

Como o coeficiente de t dos modelos de Poisson com função de ligação *log* é exponencial, com um modelo ANCOVA *OneWay* pode-se testar a homogeneidade da taxa de crescimento (dos declives) na escala logarítmica. Para comparar os coeficientes dos modelos $Po(Log)$ e $BN(Log)$, ajustados com os dados dos períodos S1 e S3 na escala natural, usou-se a função `p_value()` para implementar um teste de *Likelihood Ratio Generalizado (GLRT)*, disponível no pacote `glmglrt` (Gillibert, A., 2020) que depende do

pacote `parameters` (Lüdecke et al., 2020) e que foi especialmente concebido para os modelos MLG.

Com recurso a este teste, testou-se a hipótese nula: o coeficiente de t do modelo $Po(Log)$ ajustado aos dados do período S3 é igual ao modelo $Po(Log)$ ajustado ao período S1, contra a hipótese alternativa: o coeficiente de t no modelo ajustado a S3 é maior que o do modelo ajustado a S1. Assim, as hipóteses a testar foram:

$$\begin{aligned} H_0: \beta_1 (DuMM_y = 1) &= \beta_1 (DuMM_y = 0) \\ H_1: \beta_1 (DuMM_y = 1) &> \beta_1 (DuMM_y = 0) \end{aligned} \quad (3.8)$$

Para a AML não se rejeitou a hipótese nula ($p - value \approx 0.751$). Para o resto do país não se rejeitou a hipótese nula ($p - value \approx 1$), quer para o modelo $Po(Log)$, quer para o modelo $BN(Log)$. Quando se analisam os coeficientes de t dos modelos $Po(Log)$ para os dois períodos (Tabela 3.4) verifica-se uma diminuição no resto do país e valores muito próximos para a AML. Considerou-se a hipótese alternativa de o coeficiente do modelo ajustado a S3 ser menor que o coeficiente do modelo ajustado a S1. Neste caso, para a AML não se rejeitou a hipótese nula ($p - value = 0.2492431$) e, portanto, a diferença entre as taxas de crescimento exponencial não é estatisticamente significativa, ao nível de significância de 5%. Para o resto do país, o coeficiente do modelo $Po(Log)$ ajustado a S3 é menor que o do modelo ajustado a S1, ao nível de significância de 5% ($p - value \approx 0$ para o teste LRT).

O modelo com melhor bondade de ajustamento aos dados da AML no período S1 é o que traduz um crescimento linear ($Po(Id)$) e para o período S3 é o modelo de crescimento exponencial ($Po(Log)$), o que permite concluir que, apesar de não haver diferença estatisticamente significativa nos coeficientes do modelo $Po(Log)$ nos dois períodos, é no período S3 que o crescimento exponencial começa a acelerar. No resto do país o modelo com melhor ajustamento nos dois períodos é o que traduz um crescimento linear ($Po(Id)$ ou $BN(Id)$) - modelos com menores valores de AIC).

O comportamento dos modelos $BN(Id)$ para os dados do resto do país e $Po(Log)$ para a AML está ilustrado nos gráficos da Figura 3.12.

Em resposta à questão Q1 (secção 3.3) conclui-se que há evidências estatísticas que sustentam que o evento teve efeito na taxa de contágio na região de origem dos participantes, AML, em comparação com o resto do país.

Tabela 3.4: Coeficientes dos modelos de Poisson com função de ligação de *log*, ajustados aos dados da AML e do resto do país.

	Modelos de Poisson (Função de ligação <i>Log</i>)	
	Po(Log) (S1)	Po(Log) (S3)
	β_1	
Portugal Continental sem os concelhos da AML	0.0189 ($\exp(0.0189) = 1.02$)	0.0037 ($\exp(0.0037) = 1.00$)
AML	0.0232 ($\exp(0.0189) = 1.02$)	0.0223 ($\exp(0.0189) = 1.02$)

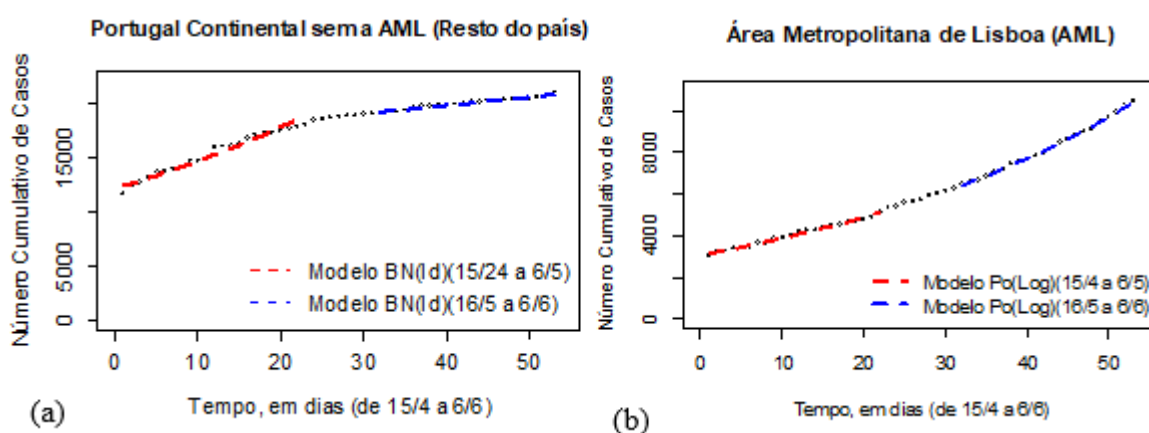


Figura 3.12: Modelos de regressão Binomial Negativo/ Poisson. ajustados às séries temporais dos períodos S1 e S3. (a) Modelo Binomial Negativo com função de ligação Identidade para Portugal Continental sem os concelhos da AML; (b) Modelo com função de ligação Log para a AML. A hipótese de homogeneidade dos declives não foi rejeitada nos dois casos, com um nível de significância de 5%.

3.3.3.2. Modelação e análise da taxa de variação do número de casos diários, por concelho

Pretende-se agora explorar a questão Q2 e avaliar se a Manifestação do 1.º de Maio teve, sobre a região de origem dos participantes, AML, efeitos homogéneos na taxa de contágio ou se apresenta diferenças geográficas, como sugerido pelas representações gráficas dos modelos de Richards ajustados com dados de diferentes concelhos da AML (exemplo na Figura 3.3).

Para cada concelho, foram formuladas as seguintes hipóteses:

H_{0i} : Não houve alteração na taxa de crescimento do número cumulativo de casos no concelho i .

H_{1i} : Houve uma alteração na taxa de crescimento do número cumulativo de casos no concelho i .

$$i=1,2,\dots, 18$$

Começou-se por aplicar a mesma metodologia usada para explorar a questão Q1.

A fim de estimar a taxa de variação do número cumulativo de casos nos períodos S1 e S3, ajustou-se um modelo de regressão de Poisson com as variáveis “Número cumulativo de casos COVID-19” e “Tempo, em dias, de 15 de Abril a 6 de junho de 2020”, com a função de ligação *Identidade* para cada concelho $i, i = 1,2,\dots,18$, ao conjunto de dados $S1 \cup S3$ ($Po(Id)_i$) (Tabela 3.5), com interação entre a variável “Tempo, em dias, de 15 de Abril a 6 de junho de 2020” (t) e a variável *Dummy*, definida anteriormente.

$$\begin{aligned} Po(Id)_i : NCasos_i &= \beta_{0_i} + \beta_{1_i} t + \beta_{2_i} Dummy_i + \gamma_i Dummy_i \times t \\ g(E(Ncasos_i)) &= E(Ncasos_i) \quad (\text{Função de ligação Identidade}) \\ NCasos_i &\sim Po(g(E(Ncasos_i))) \end{aligned} \quad (3.9)$$

Na Tabela 3.5 estão os valores dos coeficientes dos modelos $Po(Id)_i$, para os períodos S1 e S3 e para cada concelho i , com $i = 1,2,\dots,18$. As Figura 3.13 e Figura 3.14 apresentam as representações gráficas dos modelos. Na Figura 3.13. é possível observar que, para alguns concelhos, as linhas de regressão são aproximadamente paralelas e, para outros, são claramente concorrentes.

Testou-se a hipótese de haver superdispersão com recurso ao pacote `performance` (Lüdecke et al., 2021) (Anexo III). Para cada concelho não se rejeitou a hipótese de não haver superdispersão nos dados, ao nível de significância de 5%, pelo que se usou a regressão de Poisson para os dados de contagem “Número cumulativo de casos COVID-19” de cada concelho.

Aplicaram-se testes ANCOVA *OneWay* (Field et al., 2012), aos modelos $Po(Id)_i$, com $i = 1,2,\dots,18$, com recurso à função `Anova()` (Tipo III) do pacote `car` do R (Fox & Weisberg, 2019), para testar a homogeneidade dos declives dos modelos relativos aos períodos S1 e S3. Testar a homogeneidade dos declives permite testar o potencial impacto do evento 1.º de Maio no número cumulativo de casos, uma vez que é testada a diferença entre as taxas de variação do número de casos em ambos os períodos.

As hipóteses testadas, para cada $i = 1,2,\dots,18$, são as seguintes:

$$\begin{aligned} H_{0ANCI}: \gamma_i &= 0 \\ H_{0ANCI}: \gamma_i &\neq 0 \end{aligned} \tag{3.10}$$

Na Tabela 3.6 estão os resultados dos testes ANCOVA, para cada concelho. A hipótese da homogeneidade dos declives não se rejeita para os concelhos de Alcochete, Almada, Lisboa e Montijo, a um nível de significância de 5%.

Para além de avaliar o significado estatístico das diferenças entre declives dos modelos $Po(Id)_i$ (3.9) para cada concelho, para os períodos S1 e S3, é necessário avaliar se diferenças significativas resultam de uma mudança de comportamento ou simplesmente refletem um crescimento exponencial que já estava em curso e que acaba de ser confirmado.

Para isso, para cada concelho, ajustou-se um modelo de regressão de Poisson com função de ligação *log*, com interação **Dummy** × **t**, para o período S1US3 (modelo $Po(Log)_i$ - 3.11)).

$$\begin{aligned} &Po(Log)_i : \\ g(E(Ncasos_i)) &= \beta_{0_i} + \beta_{1_i}t + \beta_{2_i}Dummy_i + \gamma_i Dummy_i \times t, \\ g(E(Ncasos_i)) &= \ln(E(Ncasos_i)) \quad (\text{Função de ligação } log) \\ Ncasos_i &\sim Po(g(E(Ncasos_i))) \\ i &= 1,2,\dots,18 \end{aligned} \tag{3.11}$$

Tabela 3.5: Coeficientes dos modelos $Po(Id)(S1)$, $Po(Id)(S3)$ e $Po(Id)$ (3.9), para cada concelho.

County	<i>Modelo de regressão de Poisson (função de ligação Identidade)</i>							
	<i>Po(Id)</i>		<i>Po(Id)</i>		<i>Po(Id)</i>			
	<i>Modelo (3.9)</i>		<i>Modelo (3.9)</i>		<i>(com interação Período/Tempo)</i>			
	<i>(15/4 a 6/5)</i>		<i>(16/5 a 6/6)</i>		<i>Modelo 3.11) (15/4 a 6/6)</i>			
	β_0	β_1	β_0	β_1	β_0	β_1	β_2	γ
Alcochete	11	0.214	19	0.086	11	0.214	8	-0.127
Almada	131	6.023	73	6.781	131	6.023	-58	0.758
Amadora	180	7.945	-254	23.296	180	7.945	-434	15.351
Barreiro	74	1.365	33	4.125	74	1.365	-41	2.760
Cascais	262	4.377	206	7.762	262	4.377	-56	3.384
Lisboa	917	34.290	885	32.056	917	34.290	-32	-2.234
Loures	206	10.303	-369	29.679	206	10.303	-575	19.376
Mafra	59	0.670	24	2.178	59	0.670	-35	1.508
Moita	54	0.805	0.08	3.353	54	0.805	-54	0.085
Montijo	35	0.802	79	0.886	35	0.802	44	0.710
Odivelas	137	5.390	-168	15.374	137	5.390	-306	9.985
Oeiras	185	2.427	20	8.689	185	2.427	-165	6.263
Palmela	18	-0.015	-9	0.932	18	-0.015	-28	0.947
Seixal	130	1.923	-64	9.406	130	1.923	-194	7.483
Sesimbra	19	0.077	-4	0.864	19	0.077	-22	0.787
Setúbal	60	0.060	9	1.983	60	0.060	-50	1.923
Sintra	428	9.889	-156	30.235	428	9.889	-585	20.347
V. F. Xira	65	7.171	4	8.722	65	7.171	-61	1.551

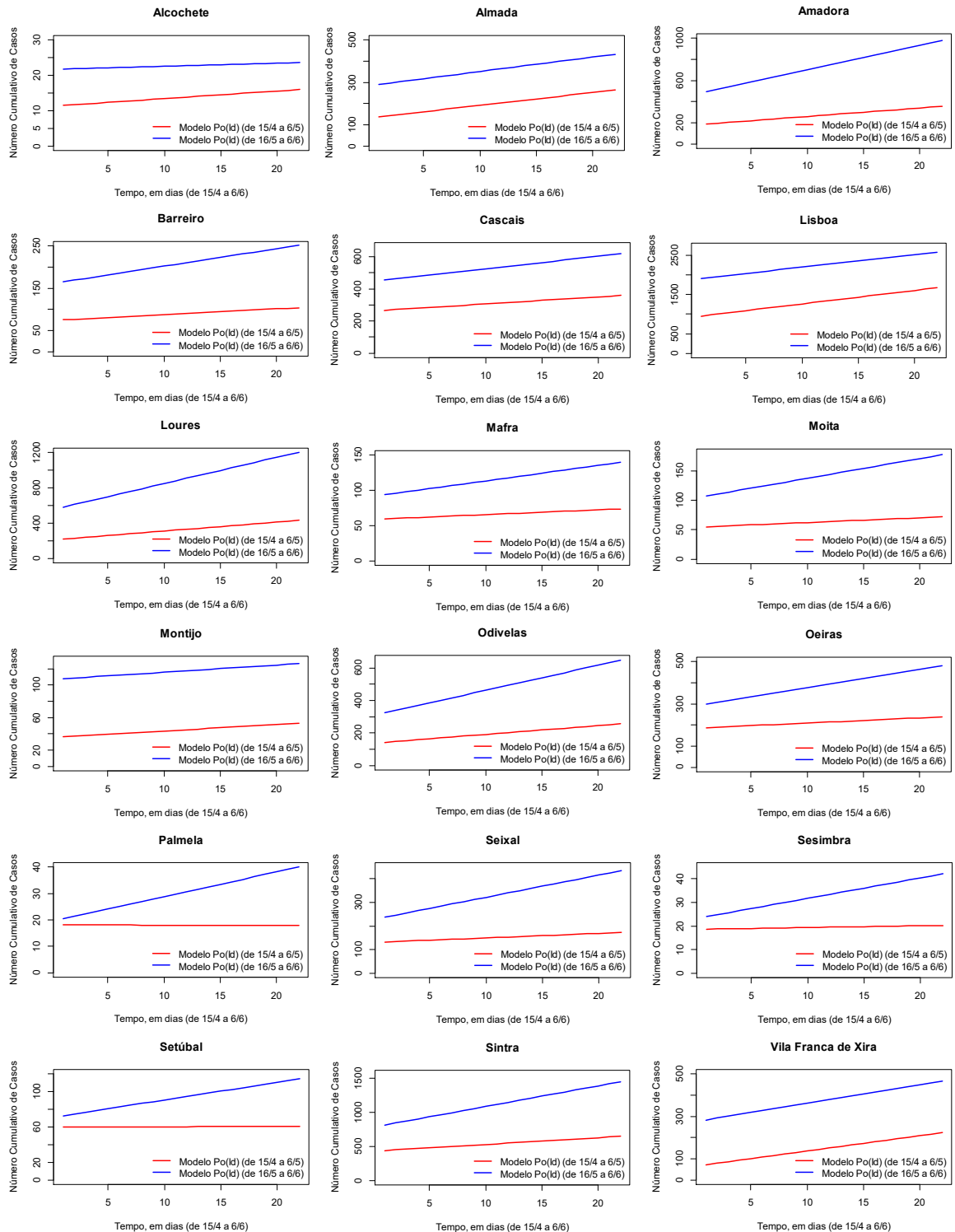


Figura 3.13: Modelos de regressão de Poisson, com função de ligação Identidade, ajustado com dados de $S_1 \cup S_3$, para cada concelho da Área Metropolitana de Lisboa ($i = 1, 2, \dots, 18$), para cada uma das subséries S1 e S3.

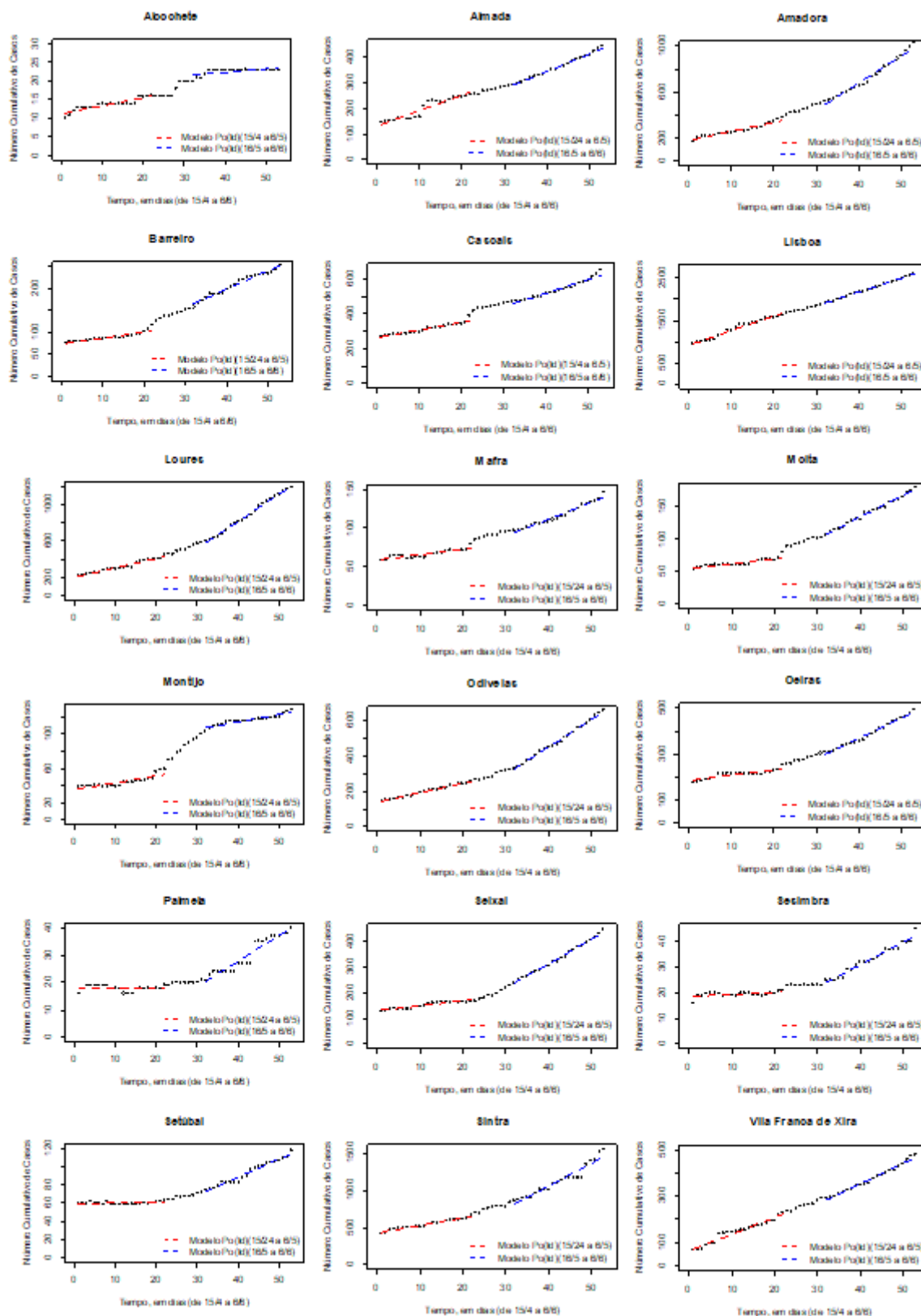


Figura 3.14: Modelos de regressão Poisson com função de ligação Identidade, ajustados com dados de S1 e S3.

Os gráficos da Figura 3.15 representam estes modelos, para os períodos S1, S3 e S1 U S3. Estão ainda representados os dados observados na série completa, incluindo os dados do período S2, período este que reflete a dinâmica da notificação de casos com possível contágio a 1.º de Maio, com atraso temporal.

Tendo testado a homogeneidade dos declives dos modelos $Po(Id)$ para os dois períodos S1 e S3 em análise, por concelho, e tendo identificado os concelhos com taxas do número de casos estatisticamente não homogéneas, ao nível de significância de 5%, testou-se o comportamento exponencial dos dados dos concelhos.

Para comparar os coeficientes de t nos modelos $Po(Log)_i$ (equação 3.11) ajustados aos dados de S1 e S3 de cada concelho, testaram-se as hipóteses:

$$\begin{aligned}
 H_0: & \text{ Há homogeneidade dos coeficientes de } t \text{ nos modelos} \\
 & \text{ajustados a S1 e a S3.} \\
 H_1: & \text{ Não há homogeneidade dos coeficientes de } t \text{ nos modelos} \\
 & \text{ajustados a S1 e a S3.}
 \end{aligned}
 \tag{3.12}$$

Estas hipóteses foram testadas recorrendo a uma ANCOVA *One-Way*, do pacote `car` (Fox & Weisberg, 2019).

Usando o teste da Razão de Verossimilhança Generalizado, disponível no pacote `glmglrt` (Gillibert, A., 2020) definiram-se as hipóteses seguintes:

$$\begin{aligned}
 H_0: & \text{ O coeficiente de } t \text{ do modelo ajustado a S3 é igual ao do} \\
 & \text{modelo ajustado a S1} \\
 H_1: & \text{ O coeficiente de } t \text{ do modelo ajustado a S3 é maior que o do} \\
 & \text{modelo ajustado a S1}
 \end{aligned}
 \tag{3.13}$$

Os resultados encontram-se na Tabela 3.6 e na Tabela 3.7. Os testes ANCOVA *OneWay* (*Anova Tipo III*), usados para testar a homogeneidade dos coeficientes de t dos modelos $Po(Id)_i$ para os períodos S1 e S3 resultaram nas mesmas conclusões no que respeita aos concelhos em que se rejeitou, ou não, a hipótese de não ter havido alteração no padrão de crescimento do número de casos COVID-19 de S1 para S3. O mesmo não se passou para os modelos $Po(Log)_i$ (3.11), como era expectável, uma vez que nem em todos os casos em que não há homogeneidade dos coeficientes dos modelos de crescimento linear, se verifica um aumento do coeficientes de S1 para S3, como definido na hipótese alternativa em (3.13).

Para os concelhos em que não se rejeitou a hipótese de homogeneidade de declives dos modelos de crescimento linear (Alcochete, Almada, Lisboa e Montijo) e, portanto, do padrão de crescimento linear do período S1 para o período S3, não se testou a hipótese de mudança no padrão de crescimento exponencial, uma vez que, não se rejeita a hipótese de não ter havido alteração na taxa de variação de crescimento de casos COVID-19 nestes concelhos. Para os restantes concelhos, verifica-se que não há evidência estatística de que a taxa de crescimento exponencial tenha aumentado de S1 para S3, ao nível de significância de 5%, nos concelhos de Cascais, Loures e Vila Franca de Xira, embora não se rejeite, ao nível de significância de 5%, a hipótese de homogeneidade das taxas de variação do número de casos em escala logarítmica dos concelhos de Cascais, Loures, Mafra e Odivelas. Rejeita-se a hipótese de homogeneidade destes coeficientes no concelho de Vila Franca de Xira. Testando a hipótese alternativa de redução do coeficiente do modelo de crescimento exponencial de S1 para S3, obtém-se um $p - value = 08.658184 \times 10^{-44}$ para o *likelihood ratio* teste e, portanto, há evidência estatística significativa de que este coeficiente tenha reduzido.

Assim, pode-se considerar três grupos de concelhos: aqueles em que não se rejeitou a hipótese de homogeneidade dos coeficientes dos modelos de crescimento linear (Alcochete, Almada, Lisboa e Montijo); aqueles em que se rejeitou a homogeneidade dos coeficientes dos modelos de crescimento linear e exponencial e aqueles em que se rejeitou a homogeneidade dos coeficientes dos modelos de crescimento linear, mas não dos modelos de crescimento exponencial.

Para avaliar se os modelos que permitiram agrupar os concelhos como descrito acima foram os que revelam melhor bondade de ajustamento e se há coerência no modelo que traduz melhor o comportamento do número cumulativo de casos de um período para o período seguinte, após o curto período S2, recorreu-se ao critério AIC. Os resultados foram os da Tabela 3.8. Observa-se que há concelhos em que o modelo que revelou melhor bondade de ajustamento (menor valor de AIC) é o mesmo para os dois períodos, outros em que o modelo com melhor bondade de ajustamento no período S1 é o linear e no período S3 é o exponencial. Outros há, ainda, em que o modelo com melhor bondade de ajustamento para S1 é o modelo de crescimento exponencial e para S3 é o linear o que, dado o curto período de transição entre S1 e S3 revela uma situação algo estranha, se se tiver em conta o que intuitivamente corresponde a um crescimento linear e um crescimento exponencial.

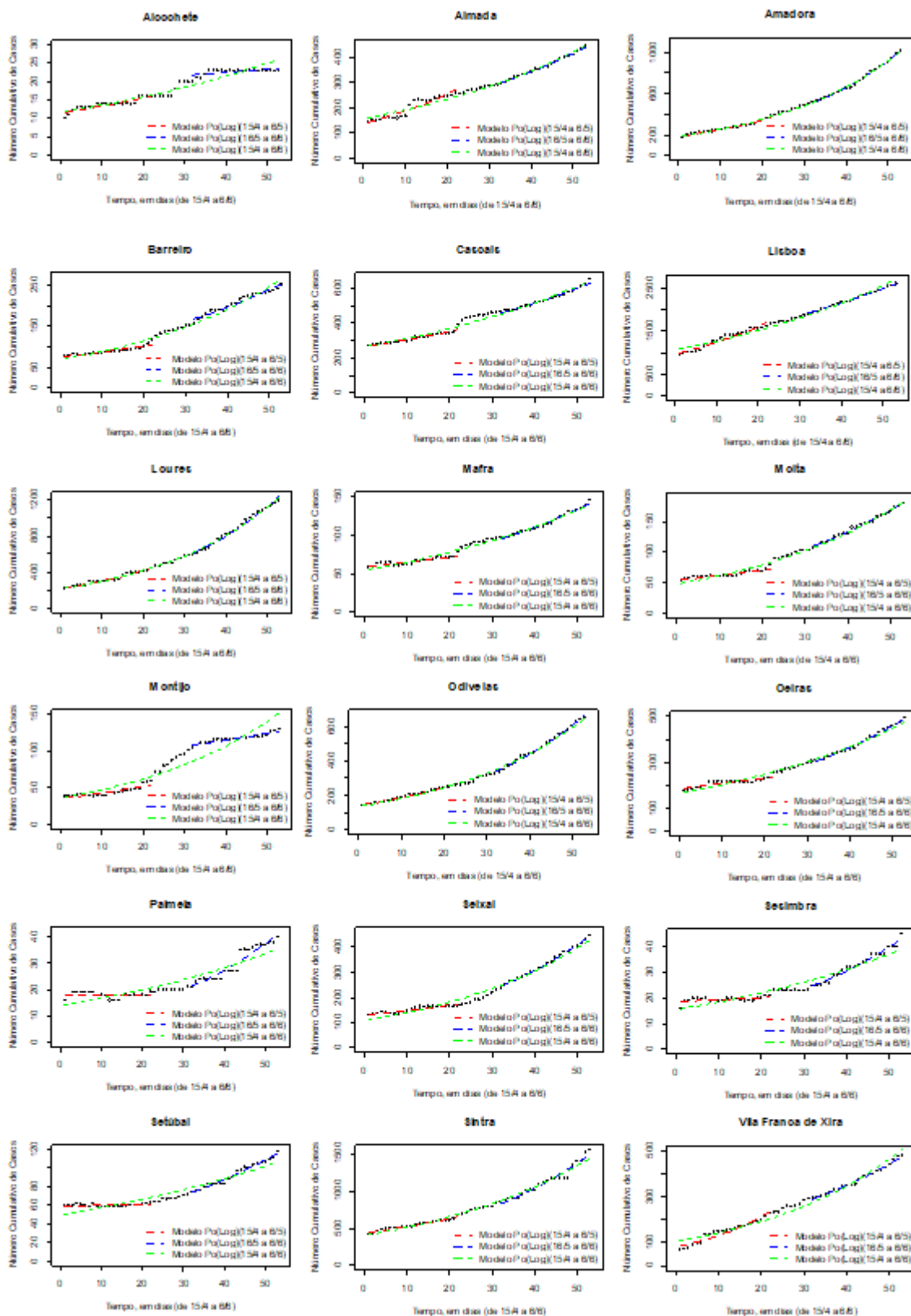


Figura 3.15: Modelos de regressão de Poisson com função de ligação \log , para a série $(S1 \cup S3)_i$ e as séries $(S1)_i$ e $(S3)_i$, sendo i o i -ésimo concelho, $i = 1, 2, \dots, 18$.

Tabela 3.6: *p* – values dos testes ANCOVA à homogeneidade dos coeficientes de *t* dos modelos *Po(Id)* e *Po(Log)* para os períodos S1 e S3 (hipóteses definidas em (3.12)).

Concelho	Com função de ligação <i>Identidade</i>	Com função de ligação <i>log</i>
	<i>p-value</i>	<i>p-value</i>
Alcochete	0.5328	
Almada	0.33562	
Amadora	< 2.2e-16 ***	0.2779
Barreiro	1.752e-06 ***	0.351
Cascais	0.0004807 ***	0.8324
Lisboa	0.2626	
Loures	< 2.2e-16 ***	0.6156
Mafra	0.0008014 ***	0.08946 .
Moita	1.01e-07 ***	0.034292 *
Montijo	0.8417077	
Odivelas	< 2.2e-16 ***	0.08847 .
Oeiras	2.125e-14 ***	7.213e-05 ***
Palmela	3.864e-05 ***	0.001230 **
Seixal	< 2.2e-16 ***	1.322e-06 ***
Sesimbra	0.001158 **	0.01972 *
Setúbal	3.347e-06 ***	0.0002397 ***
Sintra	< 2.2e-16 ***	9.567e-08 ***
Vila Franca de Xira	0.03950 *	2.577e-13 ***

Tabela 3.7: p – values dos testes de Razão de Verossimilhança para comparação dos coeficientes de t dos modelos $Po(Id)$ e $Po(Log)$ (p_{value} – pacote *parameters*) para os períodos S1 e S3 (hipóteses definidas em (3.12)).

Concelho	Com função de ligação <i>Identidade</i>	Com função de ligação <i>log</i>
	p -value	p -value
Alcochete	0.7850123	
Almada	0.1157042	
Amadora	8.026118e-67***	0.01834619*
Barreiro	6.65325e-09***	0.04452453*
Cascais	5.624479e-06***	0.3636649
Lisboa	0.920478	
Loures	1.837601e-85***	0.1655021
Mafra	1.326018e-05***	0.002430153**
Moita	8.463499e-11***	6.998145e-05***
Montijo	0.4079562	
Odivelas	6.776014e-43***	0.0008037386***
Oeiras	6.617806e-22***	1.266844e-11***
Palmela	9.663825e-08***	6.426112e-08***
Seixal	4.737208e-35***	2.695905e-18***
Sesimbra	4.737208e-35***	6.547922e-05***
Setúbal	1.28587e-09***	2.26386e-09***
Sintra	7.459431e-76***	4.169405e-21***
Vila Franca de Xira	0.007696299**	1

3.3.4. Discussão

3.3.4.1. Mudança no padrão de crescimento

No Grupo 1 (secção 3.3.1), dos concelhos onde, por hipótese, o meio de transporte utilizado para viajar para o local da manifestação do 1.º de Maio era o metro, ou uma combinação de metro e comboio e com menos de 60 minutos de tempo de viagem, destacam-se três situações: (1) para os concelhos de Almada e Lisboa, a diferença nas taxas de variação do número de casos COVID-19 em ambos os períodos não se mostrou estatisticamente significativa, a um nível de significância de 5%, uma vez que não se rejeitou hipótese de homogeneidade dos coeficientes dos modelos $Po(Id)$. Então não se pode rejeitar que não houve alteração estatisticamente significativa no padrão de crescimento linear de novos casos entre o período S1 e o período S3; (2) os concelhos de Cascais e Vila Franca de Xira mostraram uma alteração estatisticamente significativa nas taxas de crescimento linear, mas quando se ajustou um modelo de Poisson com a função log , que reflete um crescimento exponencial, verificou-se que para os coeficientes da variável t dos modelos ajustados a S1 e S3 não se rejeitou a hipótese H_0 de (3.13) ao nível de significância de 5%. Assim, não se rejeita a hipótese de o crescimento exponencial dos casos não ter aumentado de S1 para S3 para estes concelhos; (3) os concelhos de Amadora, Odivelas, Oeiras e Sintra apresentam diferenças estatisticamente significativas nos coeficientes da variável t , ao nível de significância de 5%, tanto para os modelos de crescimento linear como exponencial (em ambos os casos rejeitou-se a hipótese de homogeneidade dos coeficientes), o que leva a rejeitar a hipótese de que o padrão de crescimento se tenha mantido de um período para o outro.

Para o Grupo 2 (secção 3.3.1), o dos concelhos onde, por hipótese, os meios de transporte utilizados para viajar para o local da manifestação do 1.º de Maio foram autocarros fretados ou veículos particulares, há também três situações que se destacam: (1) os concelhos de Alcochete e Montijo não apresentam uma diferença estatisticamente significativa nas taxas de crescimento linear entre o período S1 e o período S3; (2) os concelhos de Barreiro, Mafra, Moita, Palmela, Seixal, Sesimbra e Setúbal, apresentam diferenças estatisticamente significativas nas taxas de crescimento linear do número de casos entre ambos os períodos, bem como para os coeficientes t variáveis nos modelos de crescimento exponencial (p – valor $\ll 0.005$) - levando assim a rejeitar a hipótese nula da homogeneidade de

coeficientes e portanto de ter havido aumento do crescimento de casos do período S1 para o período S3. E (3): o concelho de Loures apresenta uma diferença estatisticamente significativa nas taxas de crescimento linear entre ambos os períodos, ao nível de significância de 5%, mas os modelos de crescimento exponencial $Po(Log)$ não apresentam diferenças estatisticamente significativas para os coeficientes das variáveis t , ao nível de significância de 5% (Tabela 3.6). Poder-se-ia inferir que o crescimento exponencial já estava a ocorrer anteriormente, e que a taxa de crescimento no período S1 poderia já estar a acelerar. O critério AIC, contudo, revela que para o período S1 o modelo de crescimento exponencial ($Po(Log)$) tem o melhor ajustamento, enquanto para o período S3 o modelo de crescimento linear ($Po(Id)$) tem o melhor ajustamento (Tabela 3.8). Faltando uma razão objetiva para este comportamento (transição exponencial lenta para uma linear rápida), coloca-se a questão se o melhor ajustamento exponencial no período mais lento (S1) não pode ser simplesmente uma interpretação equivocada de um período linear. Para avaliar esta possibilidade, a janela do período S1 foi estendida, antecipando o seu início para 24 de março, o primeiro dia em que os dados dos casos COVID-19 dos concelhos foram disponibilizados. Sob esta janela estendida, o modelo linear mostrou ter o melhor ajustamento no período S1 (menor valor de AIC), o mesmo acontecendo no período S3 (Figura 3.16), esclarecendo assim a anomalia.

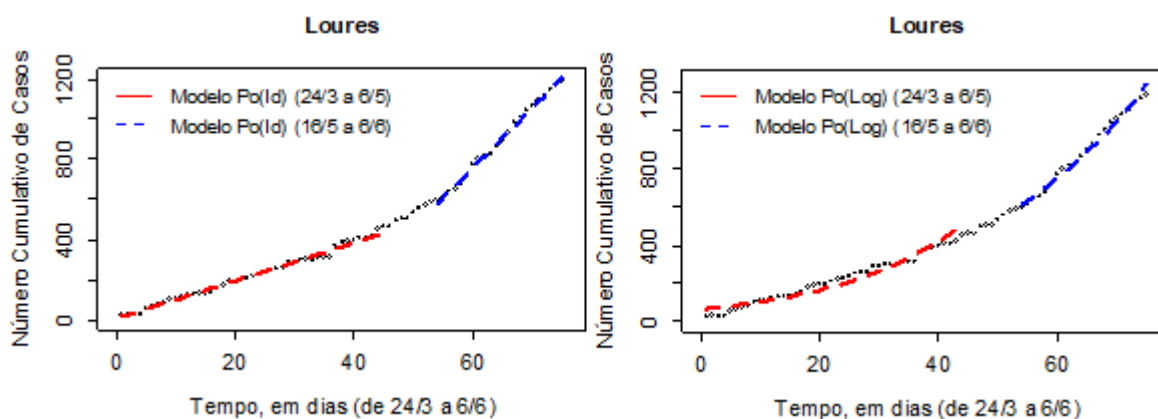


Figura 3.16: Modelos lineares (à esquerda) e exponenciais (à direita), ajustados a S1-extendido (3/24-5/6) e S3 (5/16-6/6).

Este concelho então foi associado ao grupo dos concelhos onde pode ter havido efeitos do evento, tendo em conta os modelos com a melhor bondade de ajustamento para os períodos S1 (agora estendido) e S3. A mesma análise foi conduzida para o concelho de

Cascais. Avaliando o mesmo período alargado para S1, de 24 de março a 6 de maio, constata-se, como para Loures, que neste o período o melhor ajustamento é o do modelo linear. Contudo, ao contrário de Loures, para o período S3 o melhor ajustamento é do modelo exponencial. Assim, mantém-se a decisão tomada com base nos modelos exponenciais: não rejeitar a hipótese nula.

Tabela 3.8: Valores AIC para encontrar o modelo que melhor se ajusta aos dados para ambos os períodos, S1 e S3 (refletindo o crescimento linear ou exponencial).

Concelhos	AIC					
	Período S1		Período S3		Série completa	
	Po(Id)	Po(Log)	Po(Id)	Po(Log)	Po(Id)	Po(Log)
Alcochete	102.8817	102.921	113.6088	113.6144	257.5261	258.4725
Almada	178.7218	180.3189	174.7589	174.2681	426.7524	440.0948
Amadora	177.9078	174.7062	202.6008	191.6963	726.0088	438.9635
Barreiro	146.2859	145.82	163.8726	164.8771	421.6531	386.8462
Cascais	174.8991	174.2821	186.5705	185.2613	456.1273	453.9123
Lisboa	224.0736	237.5515	215.2612	215.3478	532.8148	651.9496
Loures	181.9155	179.721	197.8376	204.3904	845.3045	454.9356
Mafra	138.3804	138.2984	150.0962	149.6482	355.2981	345.8390
Moita	137.7769	137.6297	153.8999	154.0551	389.6889	355.4585
Montijo	132.4753	131.7917	149.9141	149.9174	390.0912	404.3801
Odivelas	161.846	162.1148	182.3717	182.6918	610.0884	421.2745
Oeiras	169.1968	169.4872	178.3369	176.7638	484.7921	426.7647
Palmela	109.4255	109.4254	121.3463	121.5376	293.1822	284.5522
Seixal	157.8054	158.042	173.638	172.6875	586.4365	443.9122
Sesimbra	110.7566	110.7575	121.8019	121.8253	287.5919	280.7963
Setúbal	135.2199	135.2193	145.2664	145.1291	362.1549	349.6182
Sintra	187.1071	187.7036	248.9182	232.6616	789.4508	525.5604
V. F. de Xira	168.6007	182.7783	177.8455	175.8517	414.6792	500.7415

Nota 1: A cor rosa na coluna dos Concelhos identifica os concelhos do Grupo 1; a cor verde e a cor laranja identifica os modelos com menores valores de AIC (com função de ligação Identidade e Log, respetivamente) para cada concelho; a cor azul identifica o modelo ajustado à série completa com o menor valor de AIC.

3.3.4.2. Análise de resultados sob variáveis de mobilidade e de demografia

A dinâmica de propagação da COVID-19 pode ser influenciada por covariáveis de cariz demográfico, social, de mobilidade e meteorológico. No caso em análise foram consideradas covariáveis de cariz populacional: a densidade populacional, o índice de envelhecimento; de cariz de mobilidade: o índice de mobilidade para Lisboa, meios de transporte, particularmente o tipo de transporte público e a respetiva distância e tempo de viagem até chegar ao local de manifestação.

Os gráficos da Figura 3.8, do item 3.3.1, onde estão descritos os dados usados neste tópico, ilustram o comportamento espacial destas covariáveis. Os gráficos da Figura 3.6 do mesmo item ilustram a magnitude dos valores das covariáveis, por concelho e de acordo com os grupos definidos tendo em conta o modo como previsivelmente os participantes na manifestação se deslocaram para a mesma: metro e/ou comboio (Grupo 1) e autocarro alugado (Grupo 2).

Após a identificação dos concelhos onde não se rejeita a hipótese de que o crescimento do número de casos COVID-19 não acelerou após a manifestação de 1.º de Maio e aqueles em que este crescimento acelerou, definiu-se uma variável binária com os resultados da análise dos padrões de crescimento entre os períodos S1 e S3: *acelerou=1/não acelerou=0* com um valor para cada concelho. As covariáveis contínuas foram categorizadas agrupando os dados em classes, considerando a sua amplitude e a ocorrência de *outliers*. Dado que as tabelas de contingência incluíram células com menos de 5 dados, utilizou-se o teste Exato de Fisher para dados de contagem com *p-value* simulado (simulação Monte Carlo com base em 2.000 réplicas) (Freeman & Halton, 1951; R Development Core Team, 2010; Requena & Ciudad, 2006) para testar a hipótese da independência entre a variável binária e cada covariável considerada, e não o teste Qui-Quadrado ou teste G por se estar a trabalhar com amostras pequenas. O teste exato de Fisher para testar a independência é mais preciso que o teste Qui-Quadrado ou teste G quando os números esperados de casos são pequenos. McDonald & Delaware (2009) aconselham que se use o teste Exato de Fisher quando o tamanho total da amostra for menor que 1000 e o teste Qui-Quadrado ou o teste G para tamanhos de amostra maiores.

Assim, testou-se, para cada $j = 1,2,3,4,5$, as hipóteses:

$$\begin{aligned}
 H_{0C_j}: & \text{ Uma mudança no padrão de crescimento num concelho é} \\
 & \text{ independente da covariável } X_j. \\
 H_{1C_j}: & \text{ Uma mudança no padrão de crescimento num concelho não} \\
 & \text{ é independente da covariável } X_j.
 \end{aligned}
 \tag{3.14}$$

Os resultados dos testes para cada covariável encontram-se na Tabela 3.9.

Tabela 3.9: Resultados do teste exato de Fisher à independência entre a alteração do padrão de crescimento no número cumulativo de casos de COVID-19 e cada uma das cinco covariáveis.

Covariável X_j	Meio de transporte	Distância de carro a Lisboa	Tempo de viagem de carro para Lisboa	Índice de mobilidade	Índice de envelhecimento	Densidade populacional
<i>p-value</i>	0.3213	0.2109	0.6982	0.2994	0.1004	0.08346

Ao nível de significância de 5%, não se rejeitou hipótese H_{0C_j} definida em (3.14) para qualquer covariável, o que significa que não se pode concluir que a variável *Número de concelhos onde houve aceleração do crescimento dos casos COVID-19 após a manifestação do 1.º de Maio* dependa de alguma das covariáveis consideradas.

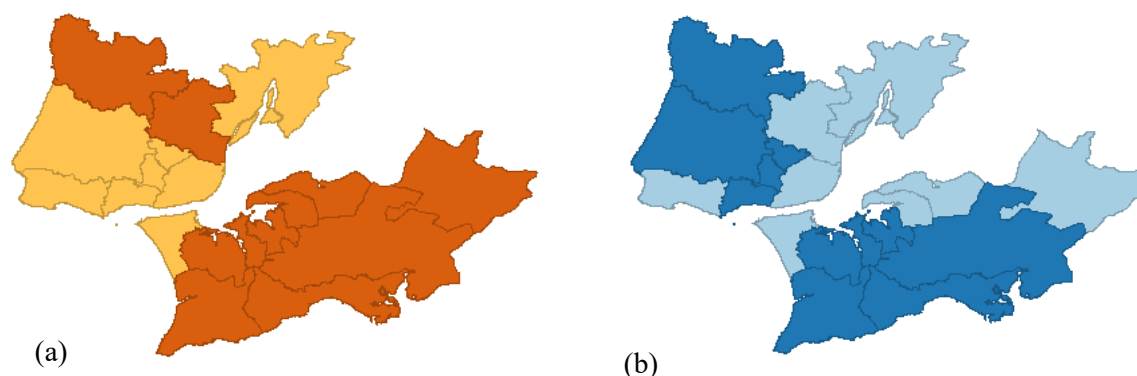


Figura 3.17: (a) Grupo 1: Laranja-claro; Grupo 2: Laranja-escuro. (b) Concelhos com mudança de padrão: Azul-escuro; Concelhos sem mudança de padrão: Azul-claro

Apesar da conclusão anterior, constata-se que em 66% dos concelhos houve aceleração do crescimento do número de casos COVID-19 após a realização da manifestação do 1.º de Maio e que, destes, 66% são concelhos do Grupo 2, concelhos a partir dos quais os participantes na manifestação, por hipótese, viajaram em autocarros fretados ou carros particulares. Não há evidência estatística de que esta situação resulte de alguma(s) das

covariáveis demográficas e de mobilidade consideradas, apesar da relação com o tipo de meio de transporte ser de alguma forma sustentada pela literatura referida na secção 3.1. que aponta a utilização de meios de transporte público durante grandes períodos como fator relevante para a disseminação desta doença.

3.4. Manifestação antirracismo

3.4.1. Caracterização do caso de estudo

No dia 6 de junho de 2020, milhares de pessoas, associando-se ao movimento iniciado nos Estados Unidos “*Black Lives Matters*”, manifestaram-se nas principais cidades portuguesas, em particular, em Lisboa, Porto e Coimbra (Lusa, 2020).

Segundo o jornal Observador¹⁰, a maior parte das pessoas que estiveram na manifestação na capital portuguesa eram jovens e a maioria estava de máscara, mas sem respeitar a distância social imposta pelas autoridades para prevenção da COVID-19.

Nesta secção replica-se a metodologia empírica usada na secção 3.3 para analisar o potencial impacto da manifestação antirracismo no padrão de contágio na AML, após a sua realização e compara-se este efeito com o potencial efeito no padrão de contágio no resto do país. Verificou-se que no período pós-impacto não houve aceleração no número de casos na AML, ao contrário do ocorrido com a manifestação de 1.º de Maio. Em seguida, analisou-se se esse padrão de não aceleração do contágio se verificou nos diferentes concelhos da AML e caso tenha havido aceleração de contágio, estatisticamente significativa, se este foi geograficamente homogéneo nos diferentes concelhos da AML.

Feita esta análise verificou-se que a proporção de concelhos onde houve potencial impacto do evento não se revelou estatisticamente significativa.

¹⁰ <https://observador.pt/2020/06/06/graca-freitas-alerta-para-riscos-das-manifestacoes-e-apela-ao-distanciamento/> Consultado em 04/03/2023

3.4.2. Modelação, análise e discussão dos resultados

3.4.2.1. Modelação e análise da taxa de variação do número de casos diários na AML e no resto do país. Discussão dos resultados.

Ajustaram-se modelos de regressão de Poisson com função de ligação *Identidade* e *Logaritmo* aos dados de contágio por COVID-19 na AML e no resto do país - modelos (3.5) e (3.7) - e testou-se a sobredispersão dos dados. Verificou-se haver sobredispersão nos dados da AML ($p - value < 0.001$), mas não nos dados do resto do país ($p - value \approx 0.995$) (Anexo IV). Ajustaram-se modelos de regressão Binomial Negativa com as funções de ligação *Identidade* e *Logaritmo* aos dados da AML de forma a acomodar a sobredispersão detetada. O modelo de regressão Binomial Negativa com função de ligação *Identidade* registou o menor valor de AIC (Figura 3.18 a)). No resto do país, o modelo de regressão de Poisson com função de ligação *Identidade* foi o que registou o menor valor de AIC (Figura 3.18 b)). Para prosseguir a análise foram seleccionados estes modelos e foram ajustados aos dados S1 U S3.

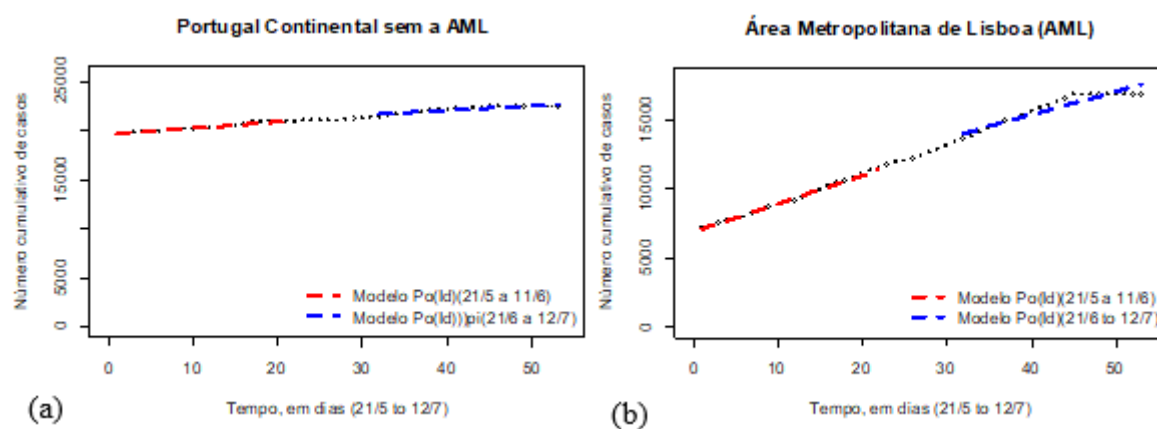


Figura 3.18: Demonstração “Black Lives Matter”: (a) Modelo de regressão de Poisson com função de ligação *Identidade* para Portugal Continental sem os concelhos da AML; (b) Modelo de regressão Binomial Negativa com função de ligação *Identidade* para a AML. O teste ANOVA rejeitou a hipótese de homogeneidade dos declives nos dois casos, com um nível de significância de 5%. O teste GLRT rejeitou a hipótese da igualdade dos declives dois casos contra a hipótese alternativa de o declive do modelo ajustado a S1 ser menor que o declive do modelo ajustado a S3.

Os valores dos coeficientes β_1 dos dois modelos encontram-se na Tabela 3.10.

Tabela 3.10: Coeficientes dos modelos de Poisson com função de ligação de Identidade ajustados aos dados da AML e do resto do país dos períodos S1 (21/05/2020 a 11/06/2020) e S3(21/06/20 a 12/07/2020).

	Período S1	Período S2
	β_1	
Portugal Continental (sem os concelhos da AML) (modelo Po(Id))	67.3	48.27
AML (modelo BN(Id))	205.645	170.58

Para testar a homogeneidade dos declives dos modelos de Poisson, para os períodos S1 e S3, implementou-se um teste ANCOVA *One-Way (Anova Tipo III)*. Assim, testou-se a significância do termo de interação entre a variável t e a variável `DuMMY` definida na secção 3.3.3. Recorreu-se ao pacote `car` do R (Fox & Weisberg, 2019).

As hipóteses testadas foram:

$$\begin{aligned} H_{0ANC} &: \gamma = 0 \\ H_{1ANC} &: \gamma \neq 0 \end{aligned} \quad (3.15)$$

Obteve-se para cada teste um $p\text{-value} < 0.01$, pelo que se pode concluir que a diferença entre os coeficientes dos modelos é estatisticamente significativa, ao nível de significância de 5%. Conclui-se, assim, que o ritmo de crescimento mudou nas duas regiões do período S1 para o S3.

Comparando os valores dos coeficientes dos modelos de Poisson verificou-se uma redução do período S1 para o período S3 (o ritmo de crescimento do número de casos desacelerou do período pré-evento para o período pós-evento) nos dois casos (Tabela 3.10).

Para testar a hipótese de os coeficientes de t dos modelos de Poisson, com funções de ligação *Identidade* e *Logaritmo*, ajustados aos dados do período S1, serem maiores que os coeficientes de t ajustados aos dados do período S3, com a variável resposta na escala natural no caso do modelo de crescimento exponencial (função de ligação *Logarítmica*) (hipóteses (3.16), usou-se o teste `p_value()` do pacote `glmglrt` (Gillibert, A., 2020), para implementar um teste de *Likelihood Ratio Generalizado (GLRT)*.

$$\begin{aligned} H_0: \beta_1(DuMMY = 1) &= \beta_1(DuMMY = 0) \\ H_1: \beta_1(DuMMY = 1) &< \beta_1(DuMMY = 0) \end{aligned} \quad (3.16)$$

A hipótese H_0 foi rejeitada para as duas regiões, AML e resto do país, confirmando, ao nível de significância de 5%, que as taxas de crescimento do número de casos COVID-19 diminuíram nos dois casos do período S1 para o período S3 ($p - value < 0.01$).

Em resposta à questão Q1 (secção 3.3) conclui-se que não há evidência estatística que sustente que o evento “*Black Lives Matter*” tenha tido efeito na taxa de contágio na região de origem dos participantes, AML, em comparação com o resto do país.

3.4.2.2. Modelação, análise da taxa de variação do número de casos diários, por concelho. Discussão dos resultados.

Para explorar a questão Q2 e avaliar estatisticamente se neste caso houve algum efeito de aceleração na taxa de contágio, geograficamente diferenciado ou não, em cada um dos concelhos da AML, ou se, à semelhança de toda a região, não há evidência estatística de efeitos, foram formuladas para cada concelho, as seguintes hipóteses ($i=1,2,\dots, 18$):

H_{0i} : Não houve alteração na taxa de crescimento do número cumulativo de casos no concelho i .

H_{1i} : Houve uma alteração na taxa de crescimento do número cumulativo de casos no concelho i .

Ajustaram-se modelos de regressão de Poisson entre a variável *Número cumulativo de casos COVID-19 e Tempo, em dias, de 21 de Maio a 12 de julho de 2020*, com as funções de ligação *Identidade e Logaritmo* para cada concelho $i, i = 1,2,\dots,18$, ao conjunto de dados $S1 \cup S3$ (modelos (3.9) e 3.11)), com interação entre a variável “*Tempo, em dias, de 21 de Maio a 12 de julho de 2020*” (t) e a variável *DuMMY*, definida na secção 3.3.3.

Na Tabela 3.11 estão os valores dos coeficientes dos modelos $Po(Id)_i$ e $Po(Log)_i$ (modelos (3.9) e (3.11)) para cada concelho i , com $i = 1,2,\dots,18$. Note-se que β_0 e β_1 são o intercepto e o coeficiente da variável t do modelo ajustado a S1 e $\beta_0 + \beta_2$ e $\beta_1 + \gamma$ são o intercepto e o coeficiente da variável t do modelo ajustado aos dados do período S3. β_2 traduz o efeito da variável *DuMMY* que identifica os períodos S1 e S3 e γ é o efeito de interação entre a variável t e os períodos S1 e S3.

As Figura 3.19, Figura 3.20 e Figura 3.21 apresentam os gráficos dos modelos. É de notar que para todos os concelhos as linhas de regressão são aproximadamente paralelas. Na Figura 3.21, para além dos dados estimados pelos modelos, estão representados os dados

observados, incluindo os do período S2, período este que reflete a dinâmica da notificação de casos com possível contágio a 6 de junho, com atraso temporal.

Testou-se a hipótese de haver superdispersão dos dados com recurso à função `overdispersion()` do pacote `performance` (Lüdecke et al., 2021). Para cada concelho não se rejeitou a hipótese de não haver superdispersão nos dados, ao nível de significância de 5% (Anexo IV), pelo que se usou a regressão de Poisson para os dados de contagem “Número cumulativo de casos COVID-19” de cada concelho, com exceção do concelho de Sintra. Para este concelho ajustou-se um modelo de regressão Binomial Negativa para acomodar a sobredispersão. Um teste de razão de verosimilhança implementado para os dados deste concelho, com a função `lrtest()` do pacote `lmtree` do R (Zeileis et al., 2008) resultou num $p - value = 0.001005$ e, portanto, ao nível de significância de 5%, rejeita-se a hipótese de que o modelo de regressão de Poisson é o que melhor se ajusta aos dados em relação ao modelo de regressão Binomial Negativa.

Como é possível observar na Tabela 3.11, nos modelos $Po(Id)$ o coeficiente γ na maioria dos casos é negativo, o que significa que a taxa de crescimento do número de casos COVID-19 é negativa e, portanto, houve uma tendência de diminuição da taxa de crescimento do período S1 para o período S3. Para os concelhos de Montijo, Odivelas e Setúbal a variação da taxa de crescimento é muito próxima de zero e nos restantes casos, a variação da taxa de crescimento é positiva, mas pequena. Apenas Cascais regista uma variação mais significativa. Nos modelos ajustados com a função de ligação logarítmica e, portanto, com crescimento exponencial, a variação na taxa de crescimento é negativa para todos os concelhos, o que reflete uma tendência de decrescimento na taxa de crescimento exponencial do período S1 para o período S3. Estas mesmas conclusões podem ser verificadas visualmente na inclinação das retas representadas a vermelho e azul nos gráficos da Figura 3.19 e Figura 3.20 e nos gráficos representados a azul e vermelho na Figura 3.21.

Como já referido, para o concelho de Sintra ajustou-se o modelo de regressão Binomial Negativa para acomodar a sobredispersão detetada. Os coeficientes deste modelo são muito próximos dos coeficientes do modelo de regressão $Po(Log)$, concluindo-se o mesmo com base nos dois modelos.

Tabela 3.11: Coeficientes dos modelos $Po(Id)$ (modelo de regressão de Poisson com função de ligação *Identidade* e com interação **Tempo** \times **Período**, e $Po(Log)$ (modelo de regressão de Poisson com função de ligação *Logaritmo* e com interação **Tempo** \times **Período**, para cada concelho. O coeficiente β_1 corresponde à taxa de crescimento do número de casos do período S1 e o coeficiente γ corresponde ao acréscimo de β_1 do período S1 para o período S3, ou seja, dá a variação na taxa de crescimento de um período para o outro. O coeficiente β_1 corresponde ao acréscimo do interseto do período S1 para o interseto do período S3.

County	$Po(Id)$ Modelo (3.9) 3.11				$Po(Log)$ Modelo 4.11			
	β_0	β_1	β_2	γ	β_0	β_1	β_2	γ
Alcochete	22	0.094	6	-0.094	3.107	0.004	0.224	-0.004
Almada	310	7.723	88	-3.122	5.754	0.020	0.304	-0.012
Amadora	549	26.943	350	-8.677	6.367	0.032	0.601	-0.021
Barreiro	186	3.578	33	-0.954	5.240	0.016	0.225	-0.008
Cascais	472	9.290	-66	6.165	6.168	0.016	0.026	-0.000
Lisboa	2030	33.184	291	-5.903	7.625	0.014	0.200	-0.006
Loures	700	29.434	513	-15.078	6.602	0.028	0.575	-0.021
Mafra	99	2.603	-30	1.144	4.611	0.021	0.117	-0.004
Moita	121	3.380	-18	1.091	4.819	0.021	0.231	-0.007
Montijo	111	0.975	11	0.022	4.712	0.008	0.137	-0.002
Odivelas	23	0.914	-9	0.209	3.214	0.026	0.180	-0.009
Oeiras	378	16.447	152	-2.933	5.982	0.030	0.515	-0.018
Palmela	324	9.458	-45	2.360	5.809	0.022	0.223	-0.007
Seixal	270	9.814	76	-3.757	5.637	0.026	0.342	-0.016
Sesimbra	28	0.863	11	-0.307	3.348	0.023	0.413	-0.015
Setúbal	79	2.048	-6	0.649	4.398	0.020	0.246	-0.006
Sintra	865	38.767	414	-6.395	6.806	0.031	0.568	-0.019
V. F. Xira	307	9.841	119	-1.939	5.757	0.024	0.445	-0.014

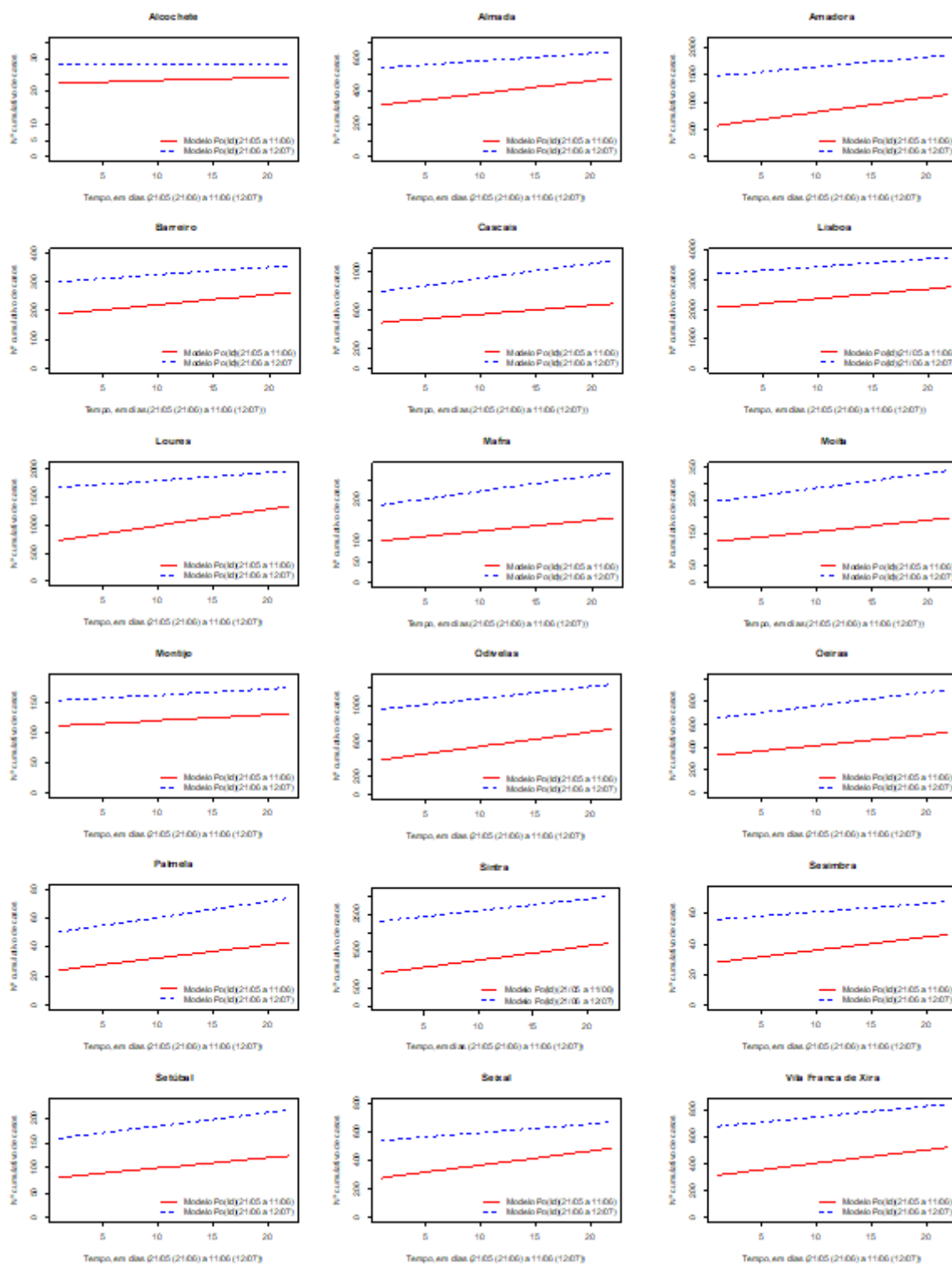


Figura 3.19: Modelos de regressão de Poisson, com função de ligação Identidade, ajustado com dados de $S_1 \cup S_3$, para cada concelho da Área Metropolitana de Lisboa ($i = 1, 2, \dots, 18$) e para cada uma das subséries S1 e S3.

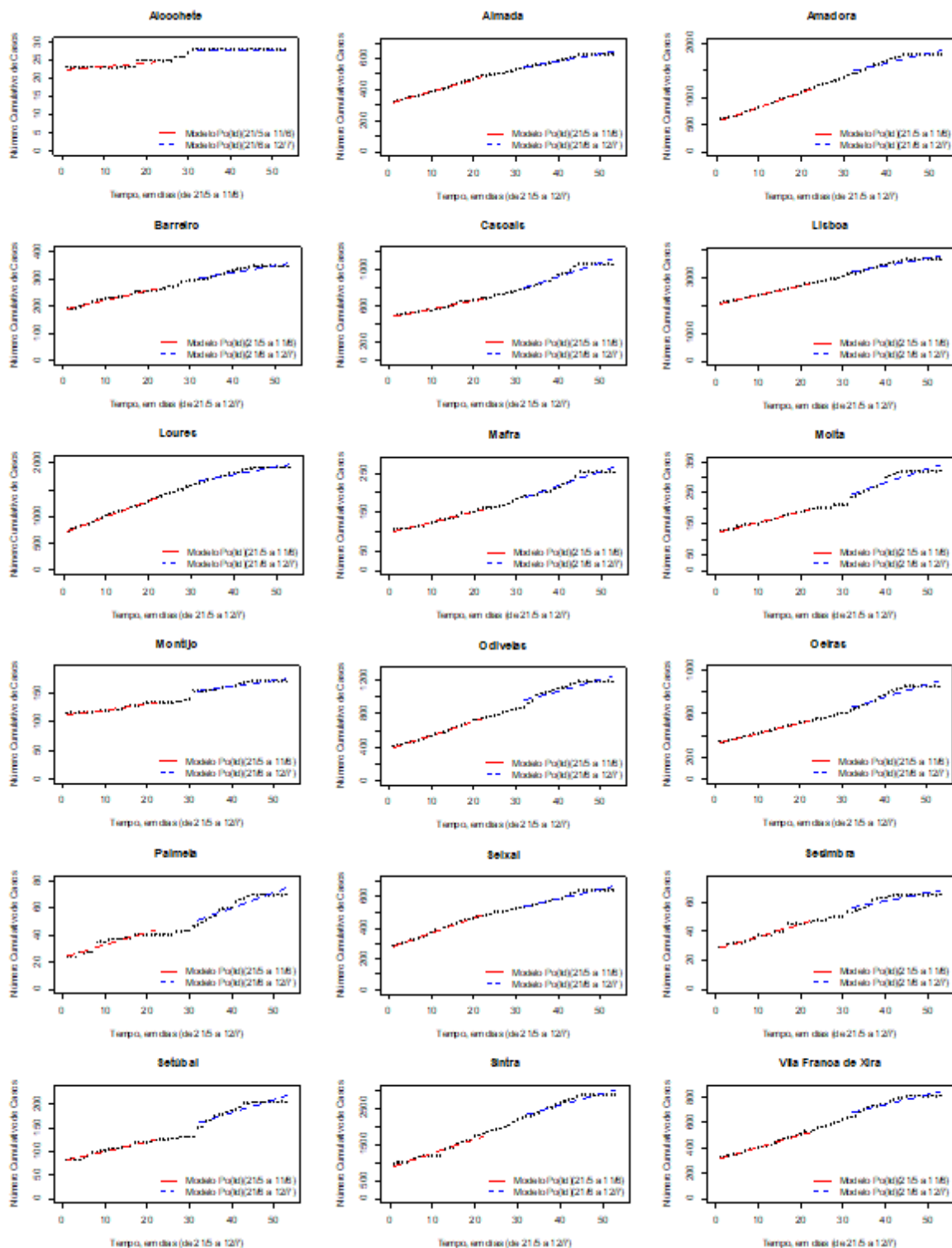


Figura 3.20: Modelos de regressão Poisson com função de ligação Identidade, ajustados com dados de S1 e S3.

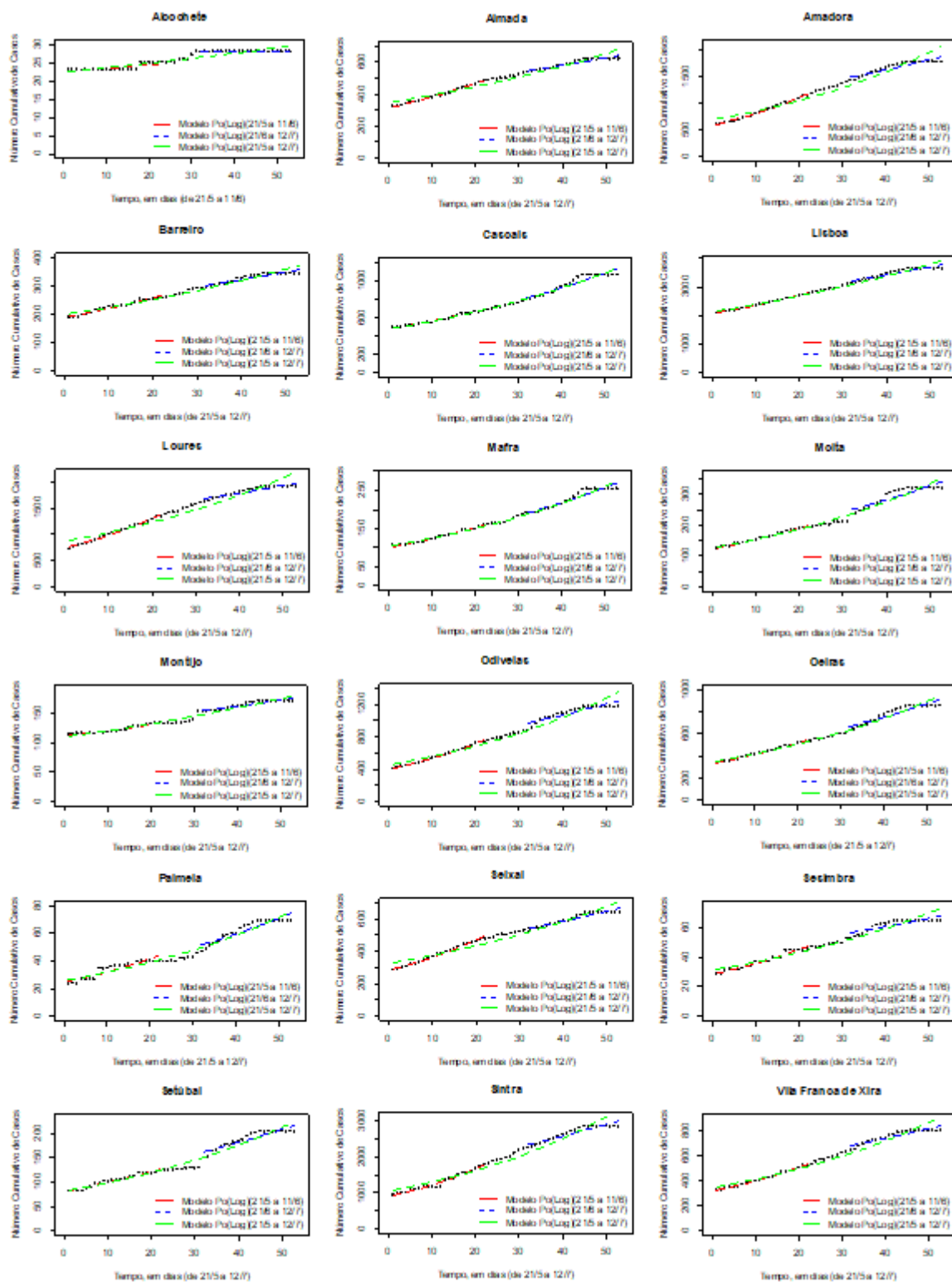


Figura 3.21: Modelos de regressão de Poisson com função de ligação *log*, para a série $(S1 \cup S3)_i$ e as séries $(S1)_i$ e $(S3)_i$, sendo i o i –ésimo concelho, $i = 1, 2, \dots, 18$.

Aplicaram-se testes ANCOVA *OneWay* (Field et al., 2012) aos modelos $Po(Id)_i$ e $Po(Log)_i$, com $i = 1, 2, \dots, l, 18$, com recurso à função `Anova()` (*Tipo III*) do pacote `car` do R (Fox & Weisberg, 2019), para testar a homogeneidade dos coeficientes de t dos modelos relativos aos períodos S1 e S3. Testar a homogeneidade dos coeficientes de t permite testar o potencial impacto do evento no número cumulativo de casos, uma vez que é testada a diferença entre as taxas de crescimento do número de casos em ambos os períodos.

As hipóteses de homogeneidade dos coeficientes de t testadas, para cada $i = 1, 2, \dots, 18$, e para os modelos $Po(Id)$ e $Po(Log)$, são as seguintes:

$$\begin{aligned} H_{0ANC_i}: \gamma_i &= 0 \\ H_{0ANC_i}: \gamma_i &\neq 0 \end{aligned} \tag{3.17}$$

Para testar se, quando houve diferença estatisticamente significativa nos coeficientes dos modelos $Po(Log)$, se registou uma diminuição estatisticamente significativa como parece pela observação dos coeficientes dos modelos ajustados aos dois períodos em análise, usou-se o teste da Razão de Verossimilhança Generalizado, disponível no pacote `glmglrt` (Gillibert, A., 2020). Testaram-se as hipóteses seguintes:

$$\begin{aligned} H_0: & \text{O coeficiente de } t \text{ do modelo ajustado a S3 é igual ao do} \\ & \text{modelo ajustado a S1} \\ H_1: & \text{O coeficiente de } t \text{ do modelo ajustado a S3 é menor que o do} \\ & \text{modelo ajustado a S1} \end{aligned} \tag{3.18}$$

Os resultados dos testes ANCOVA *One Way (Anova Tipo III)*, usados para testar a homogeneidade das taxas de crescimento nos dois modelos – coeficiente de t , encontram-se na Tabela 3.12. Os resultados do teste de Razão de Verossimilhança encontram-se na Tabela 3.12. Com exceção dos concelhos de Cascais e de Oeiras, verifica-se que, nos casos em que não se rejeita a hipótese de homogeneidade das taxas de crescimento – hipótese H_0 (3.17), também não se rejeita a hipótese H_0 (3.18), não sendo, portanto, de rejeitar que as taxas de crescimento do número cumulativo de casos COVID-19 não tenham sofrido alteração do período S1 para o período S3, ao nível de significância de 5%.

Analisando o caso do concelho de Cascais verifica-se que, apesar de se rejeitar a hipótese de homogeneidade das taxas de crescimento dos modelos $Po(Id)$, crescimento linear, não

se rejeitou a hipótese de homogeneidade da taxa de crescimento no modelo $Po(Log)$, crescimento exponencial. Este resultado é compatível com a não rejeição de H_0 (3.17), uma vez que se espera que o declive de um modelo linear ajustado a um conjunto de dados com crescimento exponencial seja mais elevado que o declive do modelo ajustado a um conjunto de dados que ocorreu num período anterior, dado que o modelo exponencial tem taxa de variação crescente se o coeficiente de t não é negativo. Os modelos com melhor qualidade de ajustamento a cada um dos períodos S1 e S3 são os modelos com função de ligação logarítmica uma vez que têm os menores valores de AIC. Também se testou a hipótese alternativa em (3.18) de os coeficientes dos modelos ajustados a S3 serem maiores que os modelos ajustados a S1, mas também neste caso não se rejeitou a hipótese H_0 ($p - valor = 0.9984585$ para o modelo $Po(Id)$ e $-valor = 0.6688771$ para o modelo $Po(Log)$).

Analisou-se o caso do concelho de Oeiras. As hipóteses de homogeneidade dos coeficientes dos modelos de crescimento linear e de crescimento exponencial foram rejeitadas ao nível de significância de 5%. Não se rejeitou a igualdade dos coeficientes pelo teste de Razão de Verosimilhança contra a hipótese alternativa de ter havido uma diminuição destes coeficientes do período S1 para S3. Podendo ter havido um aumento na taxa de crescimento de S1 para S3 testou-se a hipótese alternativa de ter havido um aumento dos coeficientes dos modelos. Rejeitou-se a hipótese nula no caso do modelo de crescimento linear, ao nível de significância de 5%, com um $p - valor = 0.006477941$. No entanto, não se rejeitou a hipótese de não ter havido aumento destes coeficientes no modelo de crescimento exponencial, $Po(Log)$. Assim, também neste concelho não se rejeita a hipótese de não ter havido aumento da taxa de crescimento do número cumulativo de casos COVID-19 em Oeiras.

Assim, para todos os concelhos da AML, do período S1 (21 de maio a 11 de junho) para o período S3 (de 21 de junho a 12 de julho) não há evidência estatística de que tenha havido aumento da taxa de crescimento do número cumulativo de casos COVID-19. Em todos os concelhos, ou não se rejeitou a hipótese nula da homogeneidade das taxas de crescimento nos dois períodos ou esta foi rejeitada contra a hipótese alternativa de ter havido diminuição desta taxa.

Conclui-se, assim, com base na metodologia implementada não haver evidência estatística que sustente que a manifestação “*Black Lives Matter*” tenha tido efeito na disseminação da COVID-19 na AML.

Tabela 3.12: *p – values* dos testes ANCOVA OneWay (*Tipo III*) à homogeneidade dos coeficientes de *t* dos modelos *Po(Id)* e *Po(Log)* para os períodos S1 e S3 (hipóteses definidas em 3.17).

Concelho	Com função de ligação <i>Identidade</i>	Com função de ligação <i>log</i>
	<i>p-value</i>	<i>p-value</i>
Alcochete	0.6941	
Almada	0.003109 **	4.274e-08 ***
Amadora	2.870e-07 ***	2.870e-07 ***
Barreiro	0.2311	
Cascais	2.983e-06 ***	0.7890
Lisboa	0.0223476 *	7.864e-12 ***
Loures	< 2.2e-16 ***	< 2.2e-16 ***
Mafra	0.07008 .	0.2532
Moita	0.1293	
Montijo	0.969658	
Odivelas	0.032937*	< 2.2e-16 ***
Oeiras	0.04469*	0.0003505 ***
Palmela	0.5314	
Seixal	0.0003621***	6.415e-13 ***
Sesimbra	0.3622	
Setúbal	0.2613	
Sintra	0.002408 **	< 2.2e-16 ***
Vila Franca de Xira	0.093717 .	3.508e-11 ***

. (Teste significativo a 0.1); * (teste significativo a 0.05); ** (teste significativo a 0.01) e *** (teste significativo a 0.001).

Tabela 3.13: *p* – *values* dos testes de Razão de Verosimilhança para comparação dos coeficientes de *t* dos modelos *Po(Id)* e *Po(Log)* (*p*_{value} – pacote *parameters*) para os períodos S1 e S3 (hipóteses definidas em (3.18)).

Concelho	Com função de ligação <i>Identidade</i>	Com função de ligação <i>log</i>
	<i>p-value</i>	<i>p-value</i>
Alcochete	0.2976549	
Almada	7.180704e-05***	2.521999e-18***
Amadora	2.078693e-10***	2.808889e-147***
Barreiro	0.0602572	
Cascais	1	
Lisboa	0.001541546***	5.290339e-27***
Loures	7.954179e-26***	6.557187e-152***
Mafra	0.9878871	
Moita	0.9692251	
Montijo	0.5199253	
Odivelas	0.004707558**	5.314894e-68***
Oeiras	0.9935221	
Palmela	0.7809214	
Seixal	2.884372e-06***	3.165046e-31***
Sesimbra	0.1264618	
Setúbal	0.9175738	
Sintra	0.0001239611***	2.66595e-181***
Vila Franca de Xira	0.01915776*	4.933369e-29***

CAPÍTULO 4

DESAGREGAÇÃO DE SÉRIES TEMPORAIS: APLICAÇÃO A DADOS DE COVID-19 EM PORTUGAL

4. DESAGREGAÇÃO DE SÉRIES TEMPORAIS: DADOS DA COVID-19 EM PORTUGAL

Nesta tese faz-se uma aplicação prática de desagregação com dados de COVID-19 dos concelhos de Portugal Continental. Uma resposta adequada e eficiente a uma pandemia exige que se identifiquem e caracterizem fatores que retardam ou aceleram a transmissão, bem como as populações mais vulneráveis ou as principais zonas de risco. Para tal, é essencial dispor de dados desagregados de alta qualidade, acessíveis, confiáveis, oportunos e abertos, para gerar informações valiosas para a tomada de decisões em tempo real. A comunidade científica pode dar um valioso contributo neste sentido, mas nem sempre lhes são disponibilizados dados desagregados com as características exigidas (Guedes, Nuno, 2020; Observador, 2020).

Como as métricas das séries do número cumulativo de casos de COVID-19 dos concelhos de Portugal Continental, disponibilizados pela Direção Geral de Saúde portuguesa, foram sendo alteradas ao longo do tempo - passaram de diários a semanais e posteriormente a incidência a 14 dias, para se poder modelar uma série mais longa ou com séries de dados de covariáveis com granularidades diferentes, há necessidade de desagregar os dados para os expressar nas mesmas métrica e granularidade ao longo do tempo para as poder modelar em conjunto. Assim, os dados são desagregados de frequência semanal para frequência diária e de incidência a 14 dias, com sobreposição de uma semana, para frequência diária. Esta aplicação permite homogeneizar a série de dados de COVID-19 de Portugal Continental, resolvendo o problema das mudanças na métrica e da granularidade de agregação temporal dos dados. Esta homogeneização permite a modelação dos dados, usando covariáveis meteorológicas de frequência diária e inferir conclusões acerca dos seus efeitos na dinâmica de propagação da COVID-19. Não se encontrou na literatura qualquer caso semelhante de desagregação.

Neste estudo são aplicados a dados de COVID-19 dos concelhos de Portugal Continental algoritmos descritos no capítulo 2, secção 2.10 e que permitem lidar com a desagregação de dados com situações descritas acima. Tanto quanto se conhece da literatura nenhum dos algoritmos de desagregação usados foi implementado na linguagem R, o que é realizado na presente secção.

Assim, dados observados em escala fina (diária) são agregados em diferentes granularidades (semanal e acumulados a 14 dias, com frequência semana) e novamente

desagregados na escala diária, de forma que se possam comparar os dados dessegregados com os dados observados. O pacote `tempdisagg` (Sax & Steiner, 2013) implementa algoritmos clássicos de desagregação temporal caracterizados na secção 2.10.4 e 2.10.5, como o método de Denton ou Chow-Lin, que são usados especialmente em dados económicos.

A desagregação espacial é importante na modelação do risco de doença pois, ao usar dados de resposta agregados em grandes regiões heterogéneas, os modelos de mapeamento de doenças permitem fazer previsões em escala fina em toda a região, usando informações de covariância em alta resolução para informar a heterogeneidade, o que é particularmente importante em muitas doenças infecciosas que têm uma distribuição geográfica heterogénea. A desagregação espacial também pode ajudar a identificar áreas de alto risco e a entender melhor as relações entre as covariáveis e a resposta numa escala espacial fina. A desagregação temporal reveste-se de igual importância, no que respeita ao estudo do comportamento dos dados no tempo e também no espaço.

Arambepola et al. (2022), entre outros, apresentam a desagregação espacial de séries de dados com recurso a modelos de regressão bayesiana, cujas respostas são combinadas com séries de covariáveis de granularidade mais fina e efeitos espaciais, de modo a poder estudar o seu efeito na dinâmica de propagação da malária. Python et al. (2020) aplica esta metodologia a dados de COVID-19 chineses. A desagregação é uma ferramenta importante na modelação do risco de doença pois permite que os modelos de mapeamento de doenças sejam mais precisos e informativos. Ao incluir informações de covariáveis de alta resolução e modelar o processo de geração de dados numa escala fina, espera-se que esses modelos possam aprender com precisão as relações entre covariáveis e a variável resposta numa escala mais fina.

O pacote `disaggregation` (Nandi et al., 2020) implementa a desagregação espacial com recurso a modelos de regressão bayesianos. Os algoritmos usados nesta tese para a desagregação temporal de dados de COVID-19, nomeadamente, o método H-Fuse, descrito na secção 2.10, subsecção 2.10.7(ii) e que são adequados para desagregar dados desta ou de outras epidemias/pandemias foram implementados no R, com as funções de desagregação definidas pela autora (códigos no Anexo X).

4.1. Caracterização dos dados de COVID-19 disponibilizados pelas entidades públicas de saúde portuguesas

A Direção Geral de Saúde (DGS) do Ministério da Saúde Português começou a disponibilizar publicamente um conjunto de dados sobre a COVID-19 em Portugal, a partir do dia 24 de março de 2020, na forma de boletins com relatório diário, num formato que torna a extração de dados nada trivial¹¹. Na tentativa de colmatar a inexistência de uma estratégia de dados abertos por parte da DGS, a *Data Science for Social Good Portugal* (DSSG PT) disponibilizou publicamente os dados que extraía diariamente dos relatórios da DGS, em formatos facilmente processáveis e familiares para a comunidade analítica. Esta equipa de investigadores, numa Carta Aberta à DGS aponta, por oposição ao caso italiano, uma série de más práticas de partilha de dados por esta Direção Geral. Entre outras, refere o nível de granularidade que é baixo – estão disponíveis dados do país, de cinco regiões (Norte, Centro, Lisboa e Vale do Tejo, Alentejo e Algarve e dos concelhos (seria mais adequado usar as NUTS - nomenclatura das unidades territoriais portuguesas usada para fins estatísticos, e engloba três níveis hierárquicos: NUTS I, NUTS II e NUTS III, consoante o nível de desagregação); dados desagregados por faixa etária ou sexo estão disponíveis apenas para o país; as constantes mudanças nos indicadores disponibilizados – não houve consistência na agregação temporal pois o número de casos acumulados diários, disponibilizados para os concelhos, passou para o número de casos acumulados semanais e posteriormente passou à incidência a 14 dias, com frequência de relatórios semanal, o que causa incertezas no planeamento de potenciais análises e projetos científicos; a existência de dados processados dentro dos boletins em formatos que não são facilmente inspecionáveis/manipuláveis: gráficos, em detrimento dos dados absolutos que lhes deram origem, com eixos demasiados esparsos e legibilidade reduzida. Estas fragilidades, aliadas a alguns erros nos dados dificultam as tarefas de análise, modelação, visualização e monitorização por parte da comunidade científica, que de outra forma melhor poderia contribuir para a tomada de decisões pelas autoridades políticas e de saúde. Apesar de, ao longo do tempo, algumas das fragilidades apontadas pela equipa da DSSD PT terem melhorado e de a DGS ter disponibilizado um formulário para que os investigadores pudessem solicitar os dados, mesmo este processo revelou-se tão demorado e recheado de

¹¹ <https://covid19.min-saude.pt/relatorio-de-situacao/>

burocracias que o acesso aos dados continuou muito limitado¹² e as fragilidades apontadas acima mantiveram-se.

Os dados relativos à COVID-19 usados nesta tese são os que foram recolhidos diariamente através da comunidade DSSG PT¹³ e que provêm dos relatórios de situação diários, emitidos de 25 de março de 2020 a 23 de março de 2021 pela Direção-Geral de Saúde. Para os concelhos, de 25 de março de 2020 a 4 de julho de 2020, o número de casos acumulados confirmados foi reportado diariamente. De 13 de julho de 2020 até 26 de outubro de 2020 foi divulgado o número acumulado de casos semanais e a partir dessa data começou a ser divulgada, semanalmente, a incidência a 14 dias, o que provoca uma sobreposição de uma semana nos dados agregados. Para as cinco regiões de Portugal Continental está disponível o número diário de casos acumulados confirmados. Para o país estão disponíveis entre outras variáveis o número diário de casos acumulados confirmados e este número desagregado por sexo e por faixa etária. A granularidade a nível dos concelhos é muito baixa e muito instável.

Os dados sobre a população total portuguesa e a população dos concelhos de Portugal Continental foram obtidos no portal Pordata¹⁴ e dizem respeito a estimativas de 2019.

A desagregação temporal é o processo que permite estimar dados de alta frequência a partir de dados de baixa frequência e, se disponíveis, obter informações de alta frequência relacionadas.

Como visto no capítulo 2, secção 2.10, na literatura é possível encontrar métodos de desagregação de séries temporais que dependem de informações básicas disponíveis, da forma como os dados foram agregados ou da existência de sobreposição (Xu et al., 2020a). Por outro lado, dependem ainda do objetivo e de possíveis problemas das séries disponíveis. Pode-se pretender desagregar dados de diferentes séries que estão disponíveis com diferentes granularidades para as poder fundir - *Change of Support Problem (COSP)* (Benedetti et al., 2021b, p. 1; Gelfand, 2001; Gotway & Young, 2002); ou ainda, uma única série pode ter, ao longo do tempo, diferentes métricas de agregação e pretende-se construir uma série temporal

¹² <https://observador.pt/2020/04/16/dgs-forneceu-dados-sobre-doentes-covid-19-a-5-grupos-de-investigadores-ha-320-pedidos-em-apreciacao/> (consultado em 19 de abril de 2020) <https://www.tsf.pt/portugal/sociedade/cientistas-contestam-protocolo-para-receberem-poucos-dados-sobre-doentes-com-covid-19-12066474.html> (consultado em 19 de abril de 2020)

¹³ <https://github.com/dssg-pt/covid19pt-data>

¹⁴ <https://www.pordata.pt>

com uma métrica homogénea. O conjunto de dados em estudo nesta tese tem estes problemas. O objetivo fundamental da desagregação é construir uma nova série temporal que seja consistente com os dados de baixa frequência disponíveis (Benedetti et al., 2021b). Dado que os dados de COVID-19 portugueses surgem agregados com métricas diferentes ao longo do tempo, vai-se proceder a transformação nos dados, desagregando parte da série relativa ao número cumulativo de casos semanais confirmados (de 7 de julho a 26 de outubro) e, como de 26 de outubro de 2020 a 23 de março de 2021, os dados são disponibilizados semanalmente através da incidência a 14 dias, em cem mil habitante, então, neste período, a desagregação deve ser implementada recorrendo a um algoritmo que permita resolver o problema da sobreposição.

A concretização de algumas metodologias faz-se na secção seguinte para o período de 25 de março de 2020 a 30 de junho de 2020 e no capítulo 5, secção 5.3. para os períodos de 7 de julho a 26 de outubro de 2020 e de 11 de novembro de 2020 a 23 de março de 2021.

4.2. Comparação dos dados reais e dos dados desagregados dos dados reais dos concelhos do distrito do Porto, com o método do Mínimos Quadrados com pseudo-inversa de Moore-Penrose e com o método H-Fuse

A agregação e desagregação de dados diários de COVID-19 é ilustrada com os dados dos 18 concelhos do distrito do Porto (Figura 2.1) e os resultados da desagregação de séries de baixa frequência para séries de alta frequência são comparados com os dados observados. Os dados das séries de alta frequência foram agregados em séries de baixa frequência: frequência diária para semanal e frequência diária para a soma acumulada dos dados diários de 14 dias consecutivos, relatados semanalmente, para os desagregar novamente para a frequência diária, utilizando algum dos métodos de desagregação temporal caracterizados no capítulo 2, secção 2.10.

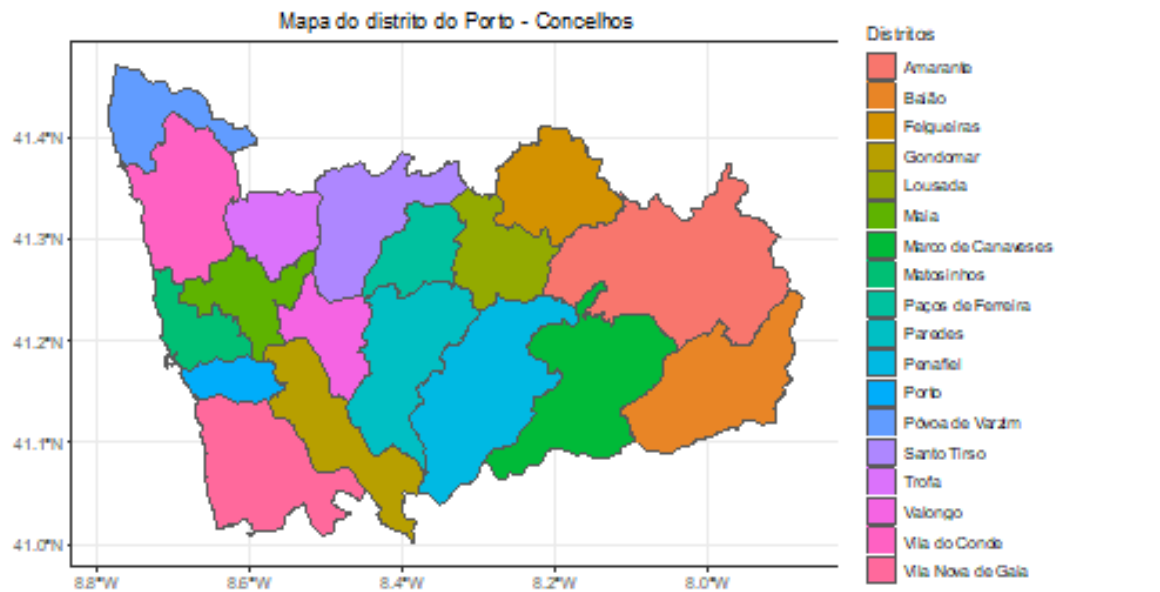


Figura 4.1: Concelhos do distrito do Porto – Portugal.

Os dados de casos portugueses de COVID-19 de 25 de março de 2020 a 30 de junho de 2020 estão disponíveis com frequência diária. Para comparar os dados reais com os dados desagregados, os dados reais foram agregados com frequência semanal, sem sobreposição (Figura 4.2), e acumulados a 14 dias consecutivos, para relatar com frequência semanal (Figura 4.3). Neste último caso há sobreposição de dados pois, a partir do segundo relatório de baixa frequência, a primeira semana de cada quinzena sobrepõe-se à segunda semana da quinzena anterior.

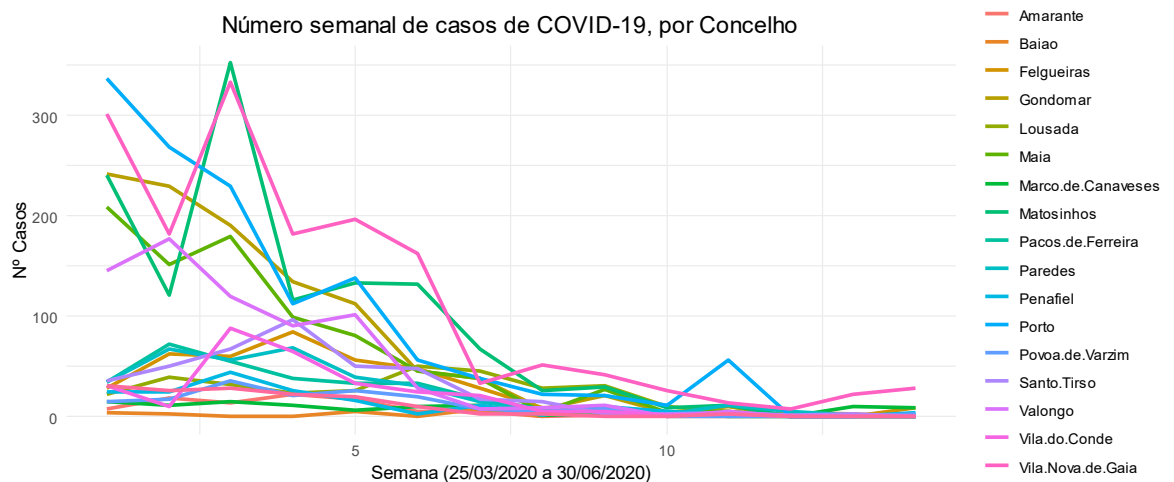


Figura 4.2: Dados de casos COVID-19 de 25/03/2020 a 30/06/2020, dos concelhos do distrito do Porto, agregados semanalmente.

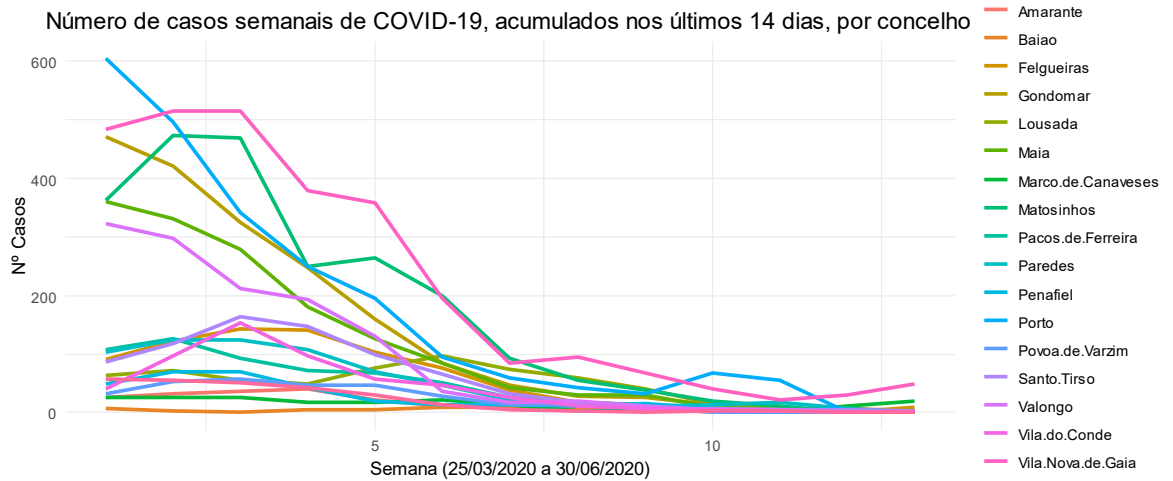


Figura 4.3: Dados de casos COVID-19 de 25/03/2020 a 30/06/2020, dos concelhos do distrito do Porto, agregados pela soma dos últimos 14 dias e relatados semanalmente – sobreposição da segunda semana da quinzena anterior com a primeira semana da quinzena atual.

As matrizes C da equação (2.124) para estes dois casos de agregação e que vão ser usadas para resolver o problema de otimização definido em (2.157) e restrição (2.158) são, respetivamente:

$$C_{(N \times 7N)} = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \end{bmatrix} \quad (4.1)$$

$$C_{(N \times 7N)} = \begin{bmatrix} 1 & 11111111111111 & 00000000000000 & \dots & 0 & 0 & \dots & 0 \\ 00000001111111 & 11111110000000 & \dots & 0 & 0 & \dots & 0 \\ 00000000000000 & 11111111111111 & \ddots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \end{bmatrix} \quad (4.2)$$

Recorrendo aos algoritmos dos Mínimos Quadrados com pseudo-inversa de Moore-Penrose (método LS) e H-Fuse (método HF) (2.10.7(ii)) (códigos R no anexo X), os dados da Figura 4.2 e Figura 4.3 foram desagregados em séries de frequência diária, com o objetivo de reconstruir os dados reais ilustrados na Figura 4.4.

A Figura 4.5 e a Figura 4.6 representam o número diário de casos de COVID-19, estimado pela distribuição dos casos semanais recorrendo aos algoritmos de desagregação dos Mínimos Quadrados com a Pseudo-Inversa de Moore-Penrose e H-Fuse, respetivamente.

A Figura 4.7 e a Figura 4.8 representam o número diário de casos de COVID-19, estimado pela distribuição dos casos acumulados a 14 dias, recorrendo aos algoritmos dos Mínimos Quadrados com a Pseudo-Inversa de Moore-Penrose e H-Fuse, respetivamente.

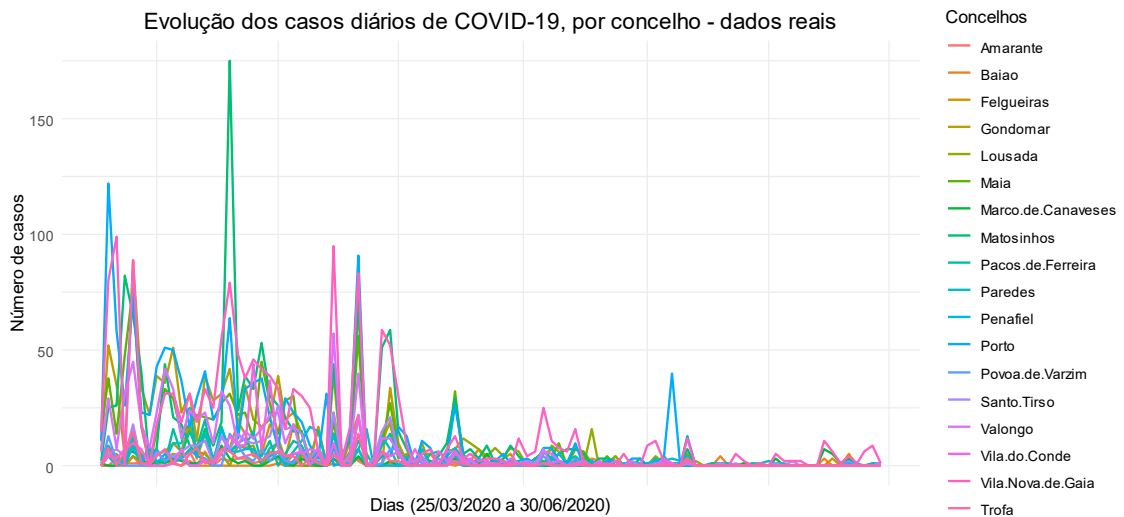


Figura 4.4: Dados diários observados nos 18 concelhos do Distrito do Porto de 25/03/2020 a 30/06/2020

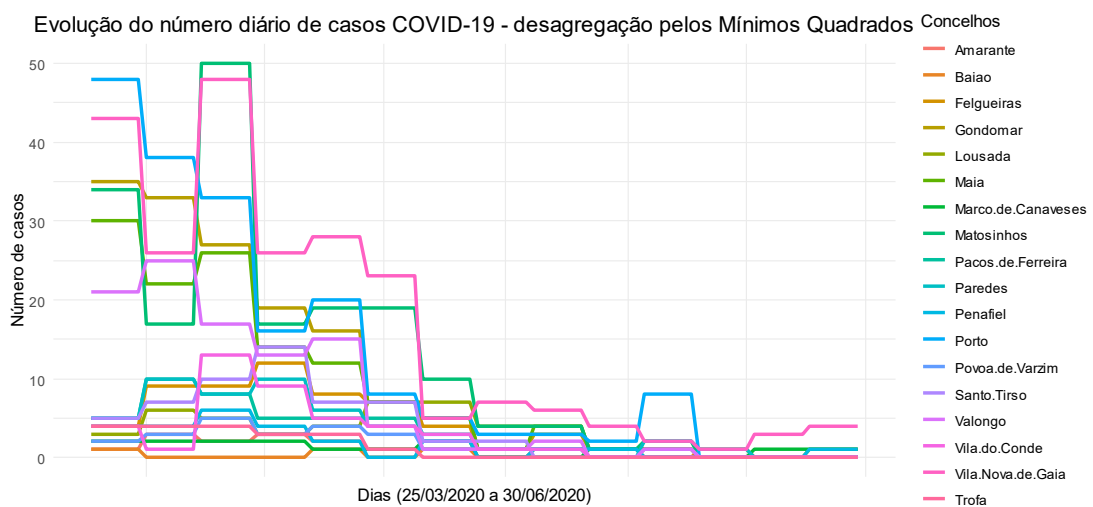


Figura 4.5: Dados de frequência semanal desagregados com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.

Pela observação das representações gráficas, verifica-se que o método HF reflete melhor a dinâmica dos casos diários observados, com suavização dos dados discrepantes, seja no caso da agregação semanal seja no caso da agregação pela soma dos dados dos 14 dias anteriores ao relatório. A suavização resulta da escolha da componente $C(x)$ em (2.157), restrição que penaliza grandes saltos entre observações sucessivas.

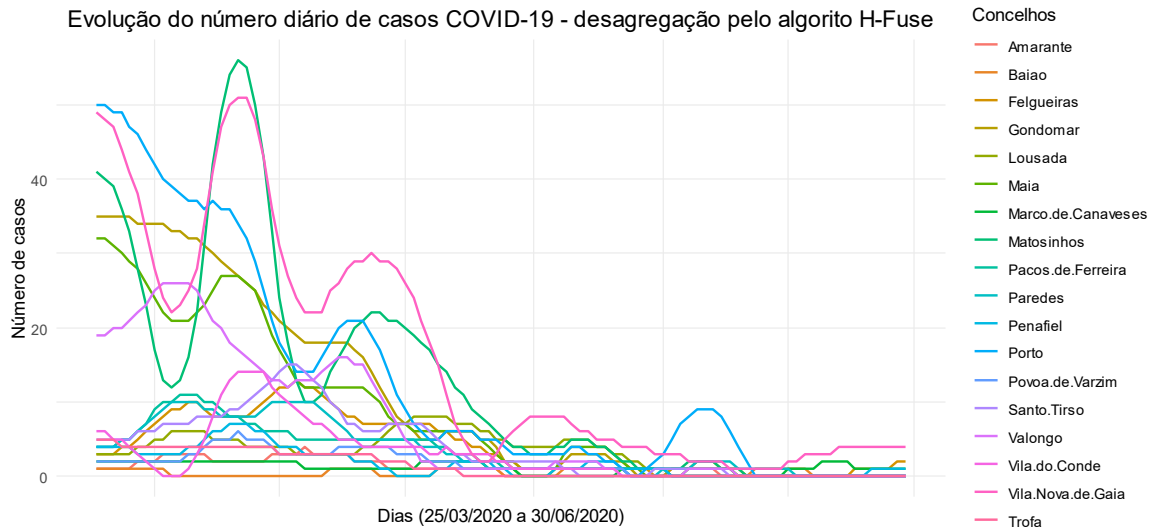


Figura 4.6: Dados de frequência semanal desagregados com recurso ao algoritmo H-Fuse.

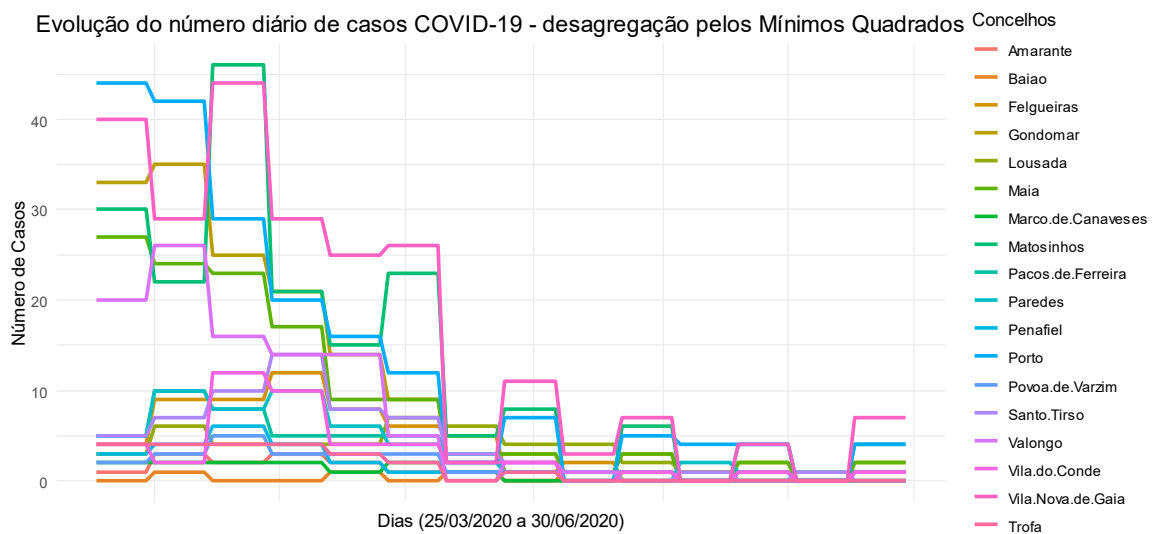


Figura 4.7: Dados acumulados a 14 dias reportados com frequência semanal, desagregados com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.

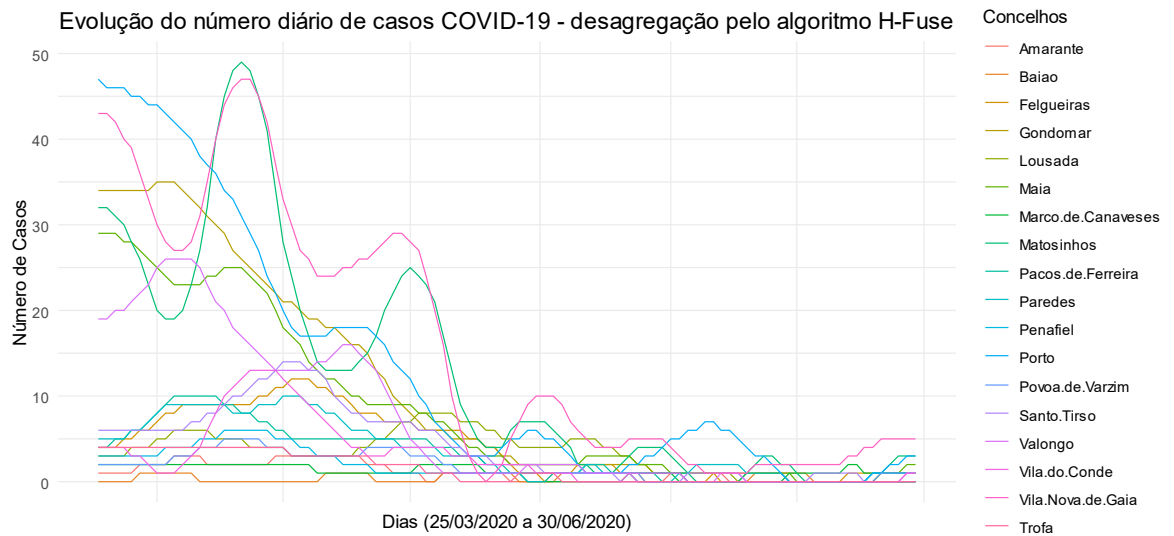


Figura 4.8: Dados acumulados a 14 dias, repostados com frequência semanal, desagregados com recurso ao algoritmo H-Fuse.

Para avaliar qual dos métodos melhor reflete os dados reais calculou-se o RMSE (*Root Mean Square Error*) e o MAE (*Mean Absolut Error*). Além disso, fez-se a normalização destas métricas pelo desvio padrão e pela média, para anular o efeito de escala. Na Tabela 4.1 e na Tabela 4.2 encontram-se os resultados da normalização pelo desvio padrão.

O RMSE normalizado pelo desvio padrão tem, para a maioria dos municípios, o valor mais baixo no método HF. No caso do erro absoluto médio (MAE), quase dois terços dos municípios têm o menor valor no método HF. Esta medida é mais robusta para valores discrepantes do que o RMSE, o que é adequado neste caso, pois é possível verificar a existência de valores discrepantes, quer valores de grande magnitude, quer de pequena magnitude, em relação aos dados contíguos. Quase todos os valores RMSE e MAE, normalizados pelo desvio padrão, são inferiores a 1, o que sugere que os valores previstos pelos diferentes modelos de desagregação estão relativamente próximos dos valores observados e que os erros não são significativamente maiores do que a variabilidade intrínseca dos dados. Também se pode assumir que há um ajuste satisfatório dos modelos aos dados.

Com este estudo conclui-se que, não dispondo dos dados diários de COVID-19 para os concelhos de Portugal Continental a partir de 4 de julho de 2020, o método H-Fuse proporciona dados desagregados que garantem maior precisão que o método dos mínimos quadrados, de forma que estes podem ser usados com covariáveis de frequência diária, que podem ser exploradas como possíveis fatores explicativos da dinâmica de disseminação da

doença. Convém, no entanto, salientar que a qualidade dos dados agregados afeta a qualidade dos resultados da desagregação, e consequentemente, os resultados da modelação.

Tabela 4.1: RMSE (*Root Mean Square Error*) dos diferentes modelos de desagregação, padronizado pelo desvio padrão.

Concelho	RMSE.S.LS	RMSE.S.HF	RMSE.14.LS	RMSE.14.HF
Amarante	0,811243	0,787724	0,811243	0,793669
Baiao	0,948766	1,009680	0,960129	1,020365
Felgueiras	0,738615	0,746643	0,742439	0,735379
Gondomar	0,686341	0,673733	0,695598	0,675436
Lousada	0,837560	0,853115	0,838840	0,852801
Maia	0,714229	0,701146	0,731932	0,708735
Marco.de.Canaveses	0,926074	0,933225	0,929656	0,907949
Matosinhos	0,795291	0,781262	0,812640	0,800129
Pacos.de.Ferreira	0,722088	0,728147	0,722088	0,712902
Paredes	0,734287	0,735632	0,734287	0,737780
Penafiel	0,766624	0,781034	0,775978	0,765768
Porto	0,693023	0,688252	0,713442	0,700920
Povoa.de.Varzim	0,881007	0,879942	0,883132	0,870840
Santo.Tirso	0,713491	0,696025	0,714680	0,692356
Valongo	0,676157	0,656308	0,679506	0,660522
Vila.do.Conde	0,821412	0,779781	0,827786	0,784211
Vila.Nova.de.Gaia	0,746863	0,714681	0,758006	0,726434
Trofa	0,805860	0,801161	0,809755	0,788494

Tabela 4.2: MAE (*Mean Absolut Error*) dos diferentes modelos de desagregação, padronizado pelo desvio padrão.

Concelho	MAE.S.LS	MAE.S.HF	MAE.14.LS	MAE.14.HF
Amarante	0,463706	0,452396	0,463706	0,446741
Baiao	0,494430	0,547029	0,473390	0,525989
Felgueiras	0,470916	0,470916	0,479636	0,477892
Gondomar	0,368554	0,355539	0,395175	0,362046
Lousada	0,568684	0,575705	0,578045	0,573365
Maia	0,405995	0,388343	0,431414	0,403171
Marco.de.Canaveses	0,628994	0,640642	0,617346	0,617346
Matosinhos	0,397338	0,400306	0,442711	0,414300
Pacos.de.Ferreira	0,459789	0,464253	0,459789	0,453093
Paredes	0,472142	0,478169	0,472142	0,474151
Penafiel	0,461000	0,471976	0,501246	0,479293
Porto	0,371147	0,385496	0,399366	0,386931
Povoa.de.Varzim	0,482476	0,488661	0,482476	0,485569
Santo.Tirso	0,416319	0,404424	0,436710	0,418018
Valongo	0,372171	0,352448	0,386749	0,358450
Vila.do.Conde	0,371615	0,363807	0,412211	0,374737
Vila.Nova.de.Gaia	0,455104	0,428025	0,466023	0,428025
Trofa	0,430034	0,422867	0,433618	0,422867

Uma medida simples de risco de doença é a taxa de incidência padronizada (SIR) (Moraga, 2020). Para cada concelho $i, i = 1, 2, \dots, 18$, o SIR é definido como a proporção entre o número de casos COVID-19 observados, Y_i , e o número de casos esperado, E_i .

$$SIR_i = \frac{Y_i}{E_i} \quad (4.3)$$

E_{jt} é calculado por padronização indireta, usando a incidência em Portugal continental no dia t , por faixas etárias, e a distribuição da população de cada concelho j , acautelando assim as diferenças que se foram revelando nas incidências por faixas etárias e as características demográficas muito diferenciadas entre concelhos.

$$E_{jt} = \sum_{k=1}^9 \left(\frac{p_{jk}}{p_k} \times n_{kt} \right) \quad (4.4)$$

onde p_{jk} é a população do concelho j , na faixa etária $k, k = 1, 2, \dots, 9, p_k$ é a população portuguesa na faixa etária k e n_{kt} é o número de casos COVID-19 na faixa etária k , no dia $t, t = 1, 2, \dots, 365$, em Portugal Continental. Note-se que foram consideradas nove classes para as faixas etárias: 0 – 9; 10 – 19; 20 – 29; 30 – 39; 40 – 49; 50 – 59; 60 – 69; 70 – 79 e mais de 80.

Na Figura 4.9, na Figura 4.10, na Figura 4.11 e na Figura 4.12 estão representados mapas de risco (SIR) calculados para os dois períodos com dados observados e dados desagregados com os algoritmos LS e HF, no início da pandemia e no período em que terminou o decréscimo acentuado que se observa nos dados originais entre 25 de março de 2020 e 30 de junho de 2020 em (Figura 4.4).

Assim, na Figura 4.9 estão representados os mapas de risco dos dias 30 e 31 de março e de 1 e 2 de abril de 2020, com os dados observados e com os dados desagregados com os algoritmos LS e HF, a partir dos dados agregados com frequência semanal. Constata-se que os dados desagregados pelo algoritmo HF refletem com maior precisão os dados observados. Na Figura 4.10 estão representados os mapas de risco para os mesmos dias, mas agora a partir dos dados desagregados a partir dos dados acumulados a 14 dias, e reportados semanalmente. Neste caso, também se verifica um melhor ajustamento aos dados observados com a desagregação pelo algoritmo HF.

Na Figura 4.11 estão representados os mapas de risco dos dias 15 de maio a 18 de maio, com os dados observados e os dados desagregados a partir dos dados agregados

semanalmente. E na Figura 4.12 estão representados os mapas de risco para os mesmos dias, mas agora a partir dos dados desagregados a partir dos dados acumulados a 14 dias, e reportados semanalmente. A maioria dos concelhos apresenta resultados muito semelhantes nos três mapas, independentemente do algoritmo usado para a desagregação e a métrica de agregação.

De acordo com os resultados obtidos nesta secção, conclui-se que o algoritmo de desagregação HF (H-Fuse) proporciona resultados mais precisos para os dados agregados semanalmente e com precisão muito semelhante para os dados acumulados nos últimos 14 dias e reportados semanalmente. Assim, no capítulo 5 serão modelados os dados de COVID-19 dos concelhos de Portugal Continental de 7 de julho de 2020 a 23 de março de 2021 desagregados recorrendo ao algoritmo H-Fuse para desagregar os dados fornecidos com agregação semanal de 7 de julho a 26 de outubro, de 2020 e acumulados a 14 dias de 11 de novembro de 2020 a 23 de março de 2021.

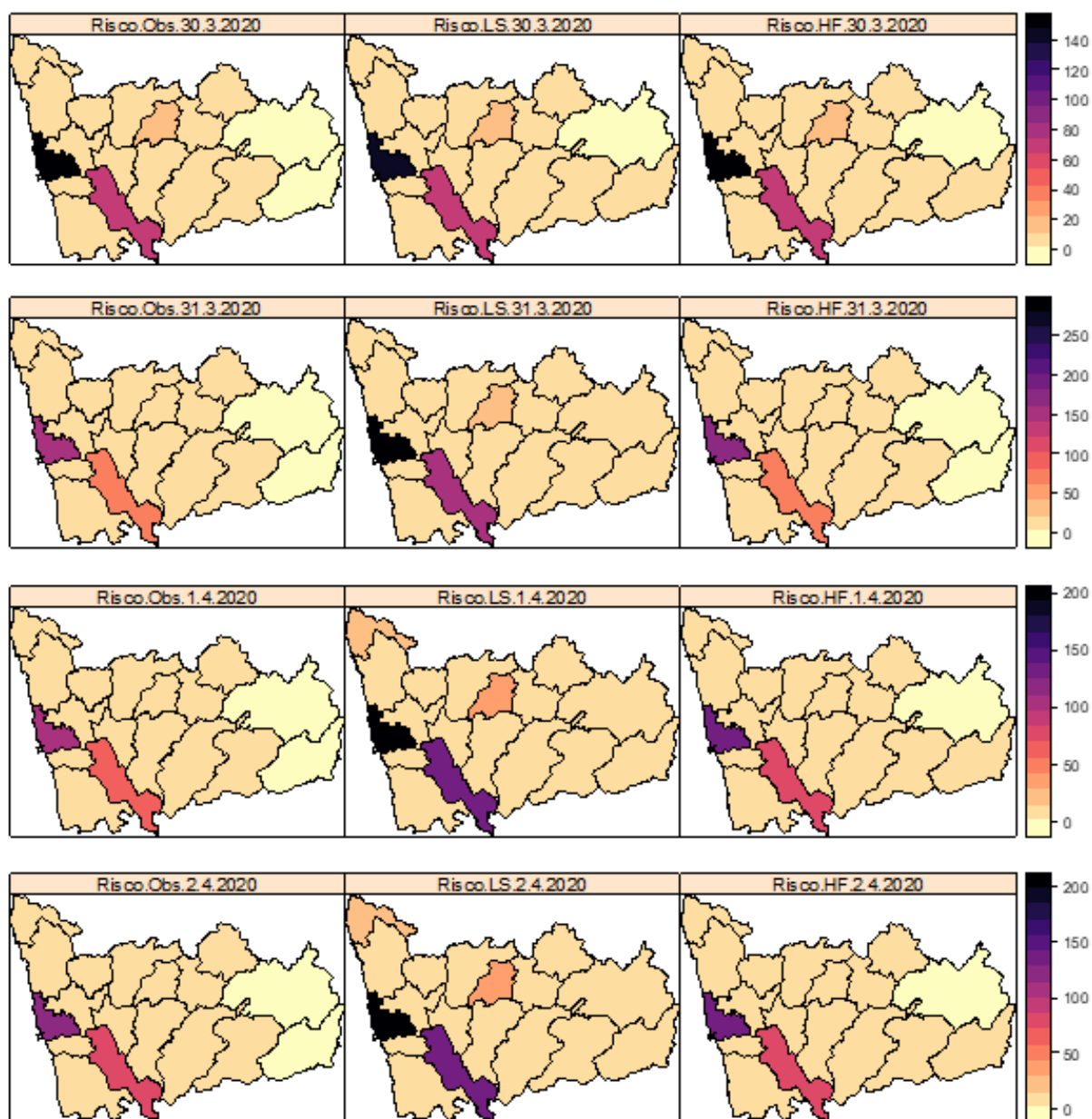


Figura 4.9: Mapas de risco dos dados observados, e os dados desagregados com os algoritmos LS e HF de dados agregados semanalmente, para os dias 30 e 31 de março, 1 e 2 de abril de 2020.

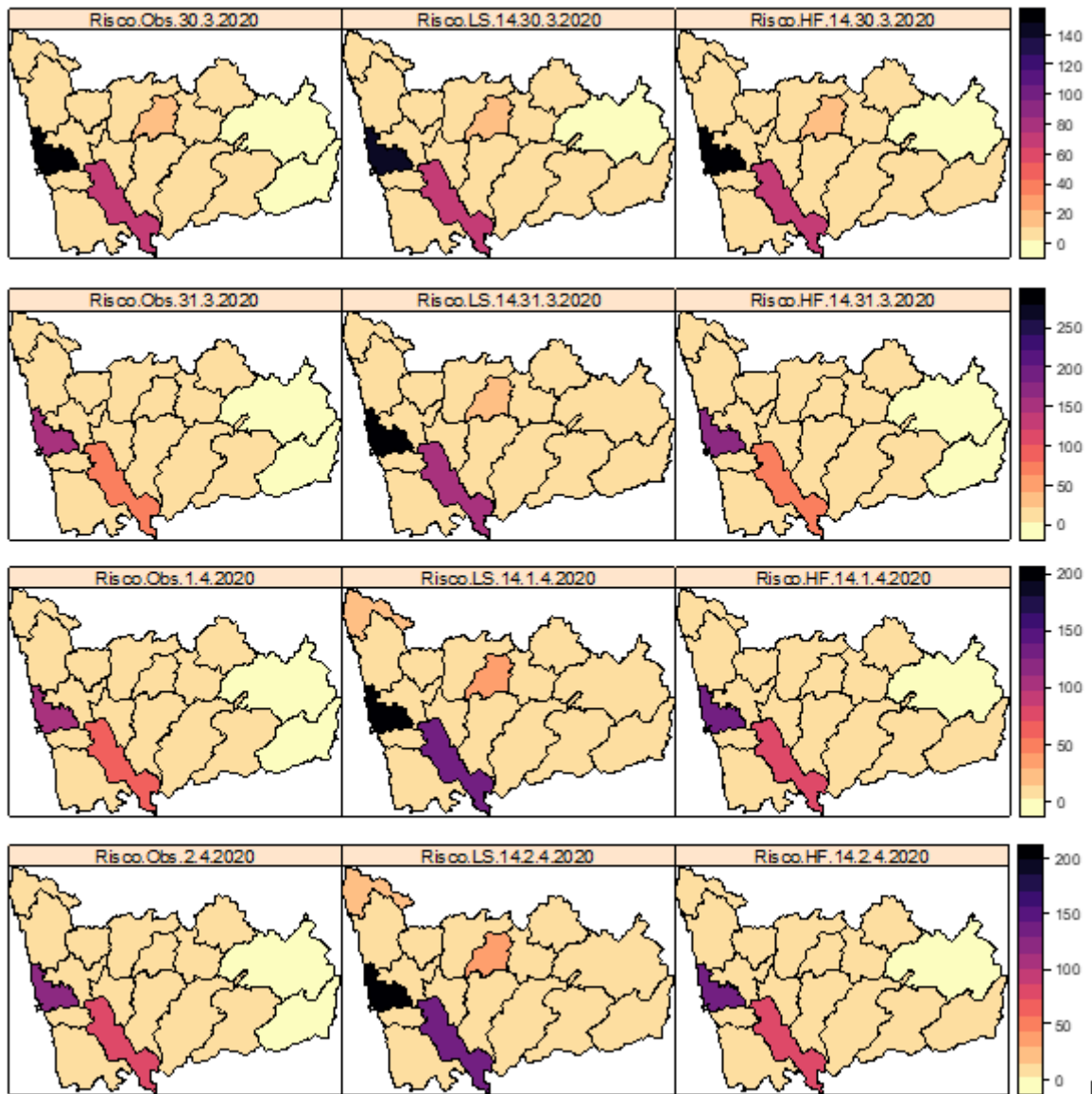


Figura 4.10: Mapas de risco dos dados observados, e os dados desagregados com os algoritmos LS e HF de dados agregados acumulados a 14 dias, para os dias 30 e 31 de março, 1 e 2 de abril de 2020

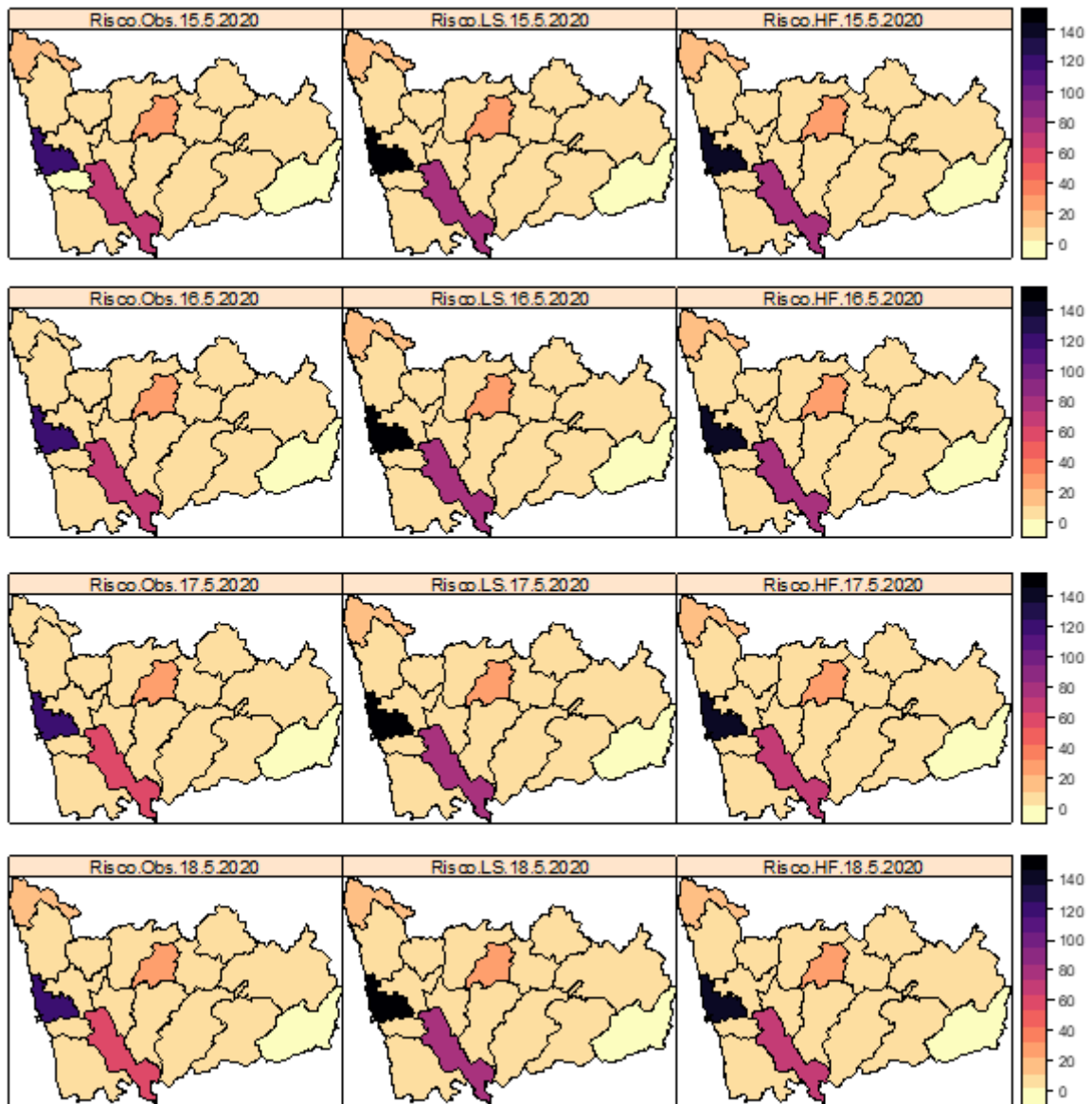


Figura 4.11: Mapas de risco dos dados observados e os dados desagregados com os algoritmos LS e HF de dados agregados semanalmente, para dias compreendidos entre 15 de maio e 18 de maio de 2020.

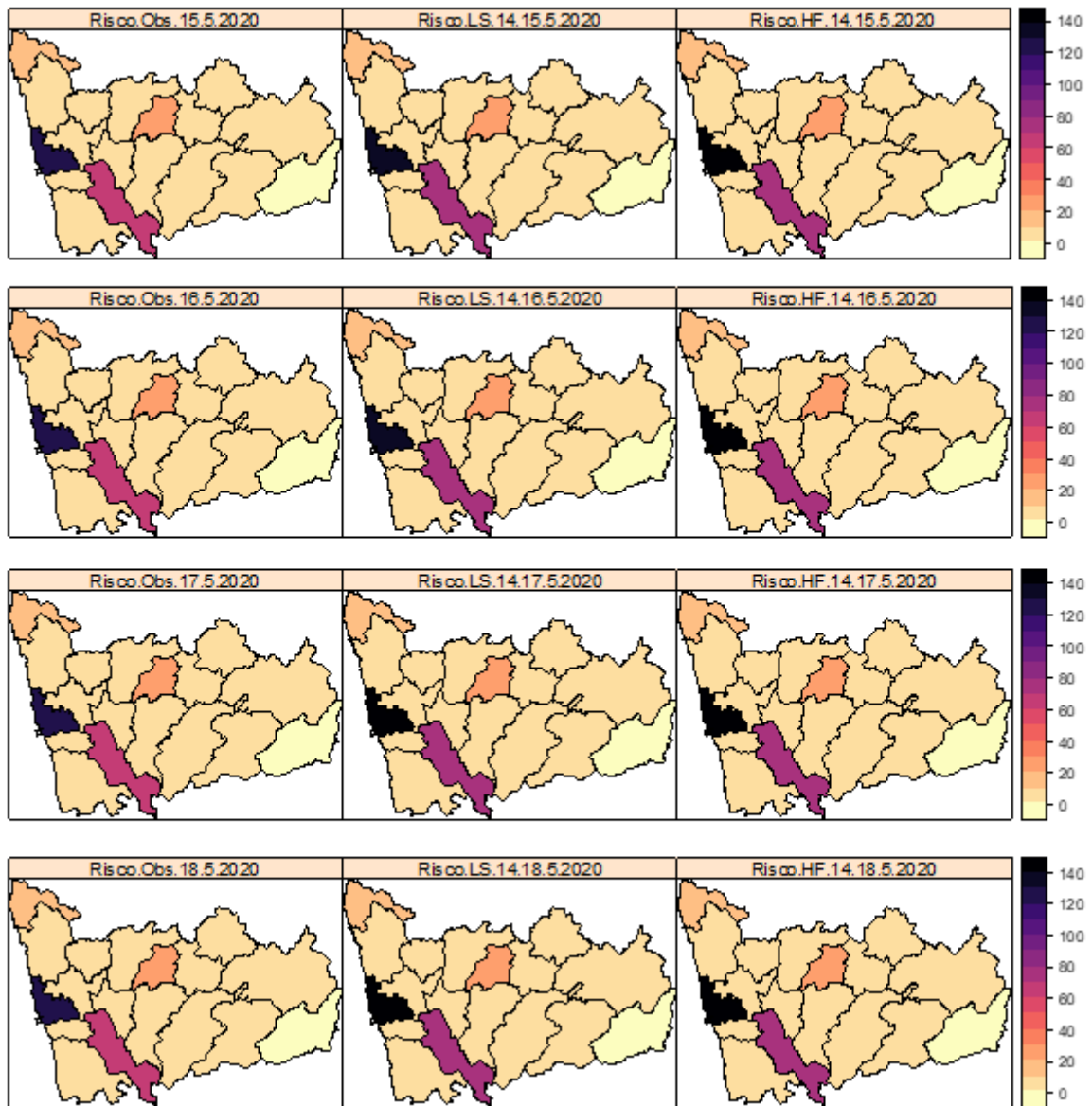


Figura 4.12: Mapas de risco dos dados observados e dos dados desagregados com os algoritmos LS e HF de dados agregados acumulados a 14 dias, para dias compreendidos entre 15 de maio e 18 de maio de 2020.

CAPÍTULO 5

MODELAÇÃO DE DADOS ESPAÇO TEMPORAIS DE COVID-19 DOS CONCELHOS DE PORTUGAL CONTINENTAL

5. MODELAÇÃO DE DADOS ESPAÇO-TEMPORAIS DE COVID-19 – CONCELHOS DE PORTUGAL CONTINENTAL

5.1. Introdução

O vírus SARS-COV-2 causa a doença COVID-19, uma das ameaças mais graves à saúde pública em tempos recentes. A epidemiologia espacial pode ajudar a compreender a distribuição e a propagação da doença no espaço e no tempo, pois as taxas de incidência altamente variáveis em diferentes locais mostram impactos distribuídos de forma desigual. A dinâmica da doença é caracterizada por um atraso espacial e temporal entre a exposição e a deteção, e é influenciada por transmissão assintomática relevante, com implicações para uma resposta de saúde pública espacialmente fragmentada, o que justifica a investigação com ferramentas espaço-temporais. Como este patógeno é facilmente propagado pela transmissão humano-para-humano pelo ar ou por contato direto, o comportamento humano afeta a dinâmica de contágio.

Como as condições atmosféricas afetam o comportamento humano, a influência na transmissão do SARS-CoV-2 de componentes climáticos como temperatura, humidade e pluviosidade tem sido investigada (Mecenas et al., 2020). Por outro lado, estudos relacionados foram motivados pelo facto de a temperatura e a humidade do ar também regularam a sobrevivência dos coronavírus de SARS e MERS (Aboubakr et al., 2021).

Como se prevê que COVID-19 possui modos de transmissão característica dos SARS, prevê-se que o SARS-CoV-2 possa ter uma relação semelhante com variáveis climáticas como temperatura, humidade e precipitação. Casanova et al. (2010), Chan et al. (2011) e Lowen et al. (2007) estudam a influência de fatores meteorológicos na sobrevivência e propagação do coronavírus SARS, concluindo que fatores como a temperatura e a humidade relativa influenciam a sobrevivência e a transmissão deste coronavírus. Em estudos publicados com dados dos primeiros meses da COVID-19 predominam os casos em que se conclui que temperatura, humidade e precipitação influenciam a transmissão do vírus. De uma forma mais ampla, McClymont e Hu (2021), Moazeni et al. (2023), Chen et al. (2023), em artigos de revisão, analisam o papel de fatores climáticos na transmissão da COVID-19, baseados na literatura disponível sobre a associação entre variáveis meteorológicas e a incidência da COVID-19. Concluem que a maioria dos estudos analisados relatou associações significativas entre temperatura e humidade e número de casos de COVID-19,

mais do que com a precipitação. Foram relatadas associações positivas ou negativas, dependendo das regiões de incidência do estudo, sugerindo que o clima desempenha um papel na dinâmica de transmissão do SARS-cov2, em concordância com os estudos relativos a outros coronavírus, mas o efeito geral dessa associação e a dinâmica da transmissão é influenciada por outros fatores, como medidas de saúde pública, padrões comportamentais humanos e fatores socioeconómicos, que podem desempenhar um papel mais significativo em futuros surtos.

Neste capítulo consideram-se as covariáveis meteorológicas: *temperatura (máxima e mínima)*, *humidade relativa (máxima e mínima)* e *pluviosidade* em Portugal Continental, covariáveis que podem ajudar a explicar padrões comportamentais que têm impacto na propagação da doença e no comportamento do vírus. Diversos estudos mostram que covariáveis de cariz sócio-demográfico como a densidade populacional ou idade também podem ser tidas em conta na dinâmica de propagação da doença (Embury et al., 2022; García, 2021).

A pandemia COVID-19 propagou-se espacialmente de forma heterogénea em todo o mundo, o mesmo acontecendo em cada país. As ferramentas de análise espacial e espaço temporal são úteis no apoio aos programas de controlo de doenças infecciosas, cuja transmissão é impulsionada por complexas variações espaciais e temporais de fatores socioambientais. Em particular, os Modelos Hierárquicos Bayesianos podem incorporar vários efeitos espaciais, temporais e espaço-temporais nos esquemas de modelação que permitem explorar o efeito destes fatores na propagação geograficamente diferenciada destas doenças e dos efeitos destes fatores na propagação.

Como descrito no capítulo 4, secção 4.1, os dados de COVID-19 disponibilizados publicamente pelas autoridades de saúde pública portuguesas não são consistentes, seja em termos de periodicidade ou em termos de métricas. Para normalizar os dados ao longo do período desta pesquisa (25/03/20 a 24/03/21) foi necessário realizar transformações que são descritas neste trabalho e que podem introduzir algum viés nos resultados. Um dos objetivos deste capítulo consiste em avaliar o possível viés introduzido pela transformação dos dados e se este pode ser observado nos mapas de distribuição geográfica de risco de COVID-19.

O número diário de casos de COVID-19 em concelhos de Portugal Continental foi modelado considerando a sua natureza espacial (dada a importância de considerar padrões espaciais para explicar e modelar a evolução da pandemia com maior precisão). Foram ainda

considerados dados meteorológicos e sociodemográficos que podem ajudar a compreender padrões que impactam a propagação da doença.

Começou-se por modelar os dados dos concelhos do Distrito do Porto uma vez que se dispõe no capítulo 4, secção 4.2.4, dos mapas de risco do período de 25 de março a 30 de junho dos dados observados e dos dados desagregados e estes podem ser comparados agora com os dados de risco estimados pelos modelos. De seguida modelou-se os dados de todos os concelhos de Portugal Continental.

5.2. Caracterização dos dados

Portugal está localizado no extremo sudoeste da Europa, na Península Ibérica. Faz fronteira com a Espanha a leste e ao norte, enquanto a oeste e sul é banhado pelo Oceano Atlântico. O país está dividido em duas regiões, Portugal Continental e duas regiões autónomas no Oceano Atlântico, os arquipélagos da Madeira e dos Açores. Portugal Continental, região sobre a qual vai incidir a análise implementada neste capítulo, possui uma área total de aproximadamente 92.000 quilómetros quadrados e está dividida em 278 concelhos.

Em Portugal Continental, o clima é geralmente classificado como mediterrânico, embora existam algumas variações regionais. A primavera, período desde o fim de março ao fim de junho é geralmente amena, com temperaturas médias entre 15°C e 20°C. Nessa estação, ocorrem frequentemente dias ensolarados e chuvosos intercalados. À medida que a primavera avança, as temperaturas tendem a aumentar gradualmente. O Verão (período desde o fim de junho até o fim de setembro) é quente e seco, com temperaturas médias que variam entre 25°C e 35°C. Nas regiões interiores, como o Alentejo, as temperaturas podem atingir valores acima de 40°C. O litoral, geralmente, tem temperaturas mais amenas devido à influência do oceano. Durante o verão, é comum haver pouca chuva, com longos períodos de sol e céu azul. O Outono (período entre fim de setembro e o fim de dezembro) é caracterizado por uma diminuição gradual das temperaturas. No início da estação, as temperaturas ainda são agradáveis, variando entre 20°C e 25°C, mas tendem a cair à medida que o outono avança. Os meses de outono podem ser instáveis em termos de precipitação, com alguns dias ensolarados e outros chuvosos. A partir de novembro, as chuvas tornam-se mais frequentes. O Inverno (período de fim de dezembro a fim de março) é suave em comparação com outros países europeus. As temperaturas médias variam entre 10°C e 15°C,

embora possam ser mais baixas nas regiões montanhosas do Norte. As áreas costeiras têm invernos mais amenos devido à influência do oceano. A precipitação é mais frequente durante o inverno, e podem ocorrer períodos de chuvas intensas.

É importante ressaltar que as características climáticas podem variar de ano para ano e existem diferenças regionais dentro de Portugal Continental. Por exemplo, as regiões do litoral norte e centro tendem a ter temperaturas mais amenas do que as regiões do interior, que são mais propensas a extremos de calor e frio.

Em resumo, o clima em Portugal Continental apresenta verões quentes e secos, invernos suaves e chuvosos, com primavera e outono intermediários em termos de temperaturas e precipitação.

Em termos socioeconómicos, Portugal possui uma economia mista, com setores público e privado desempenhando papéis significativos. É uma economia de mercado aberta, em que os principais setores económicos incluem serviços, indústria, turismo, agricultura e energia renovável. O turismo desempenha um papel importante na economia portuguesa, atraindo milhões de visitantes estrangeiros a cada ano. Tem uma infraestrutura desenvolvida, com redes de transporte bem conectadas, incluindo redes viárias, rodoviárias e aeroportos modernos. Em termos demográficos, a população portuguesa é de aproximadamente 9,8 milhões de pessoas (dados de 2019). A densidade populacional é relativamente baixa, com cerca de 106 habitantes por quilómetro quadrado. O país tem experimentado um envelhecimento da população e uma diminuição da taxa de natalidade, cenário que tem impacto nos serviços de saúde. A população está predominantemente concentrada nas áreas urbanas e no litoral. As principais cidades, como Lisboa, Porto, Braga e Coimbra, são centros populacionais significativos, enquanto as áreas rurais do interior do país têm, regra geral, uma densidade populacional baixa. Possui um sistema de saúde público, o Serviço Nacional de Saúde (SNS), que oferece cuidados médicos e serviços de saúde para os cidadãos, bem como serviços privados de saúde.

Neste capítulo vão ser usados dados meteorológicos diários dos 278 concelhos de Portugal Continental (covariáveis *temperatura máxima*, *temperatura mínima*, *humidade máxima*, *humidade mínima*, *pluviosidade*) do período compreendido entre 25 de março de 2020 e 23 de março de 2021 que foram disponibilizados pelo Instituto Português do Mar e da Atmosfera (IPMA).

São ainda considerados dados demográficos (*densidade populacional, população de cada concelho, distribuição por faixas etárias* para estimar o número esperado de casos em Portugal Continental e o *Índice de Dependência dos Idosos*) e dados sobre o *ganho médio mensal*.

Para as análises implementadas foram usados dados de COVID-19 de cada concelho de Portugal Continental, disponibilizados pela DGS na forma de relatórios periódicos e tratados pela equipa da comunidade *Data Science for Social Good Portugal – DSSG PT*¹⁵. As ilhas dos Açores e da Madeira foram descartadas para que a investigação se concentrasse inteiramente em áreas geográficas contíguas.

Assim, os dados de COVID-19 deste estudo foram descarregados do repositório *online* da comunidade DSSG PT e dizem respeito ao período de 25 de março de 2020 a 23 de março de 2021.

Como referido anteriormente, os dados de COVID-19 dos concelhos, disponibilizados pelas autoridades de saúde neste período, enfermaram de diversos problemas. Além de se verificarem erros nos dados facilmente identificados porque de 25 de março a 26 de outubro de 2020 o número cumulativo de casos não é monótono em vários concelhos, também se verificam alterações ao longo do tempo nas metodologias de agregação, originando diferentes métricas e diferentes granularidades nos dados. Concretamente, de 25 de março a 4 de julho de 2020, o número cumulativo de casos foi relatado diariamente; de 5 de julho a 13 de julho os dados não foram atualizados. De 14 de julho a 26 de outubro de 2020 o número cumulativo de casos foi relatado semanalmente. A partir de 11 de novembro de 2020, começou a ser divulgada semanalmente a incidência cumulativa a 14 dias, o que corresponde ao quociente entre o número de novos casos confirmados nos 14 dias anteriores ao momento de análise e a população residente em cada concelho, estimada a 31 de dezembro de 2019 divulgada pelo Instituto Nacional de Estatística, expressa em número de casos por 100 000 habitantes (nota metodológica registada nos Relatórios de Situação da DGS, a 16 de novembro de 2020, <https://covid19.min-saude.pt/relatorio-de-situacao/> (DGS, 2020). O compromisso metodológico da DGS foi de que a incidência a 14 dias seria divulgada semanalmente, mas isso só começou a acontecer regularmente a partir de janeiro de 2021.

¹⁵ <https://github.com/dssg-pt/covid19pt-data>

De 11 de novembro até aí verificou-se alguma irregularidade na notificação, não sendo respeitado o período de uma semana entre notificações.

Com a incidência a 14 dias divulgada semanalmente, a primeira semana de cada quinzena é a última semana da quinzena anterior e, portanto, há sobreposição de uma semana nos dados de incidência. Assim, não é possível determinar diretamente, pelos dados reportados, o número de novos casos diários ou semanais a partir de 11/11/2020. O mesmo acontece com os casos de 13 de julho a 26 de outubro, embora não haja sobreposição, mas a agregação semanal apenas permite fazer uma estimativa do número diário por distribuição do número de casos semanais. Para poder usar os dados no período de 25 de março de 2020 a 23 de março de 2021 foi necessário proceder à desagregação de dados, usando metodologias já detalhadas no capítulo 2, secção 2.10 e aplicadas no capítulo 4.

Para o estudo efetuado neste capítulo foram consideradas as variáveis “*Número cumulativo diário de casos COVID-19*”, “*Número diário de casos COVID-19*”, “*Número cumulativo semanal de casos COVID-19*” e “*Incidência a 14 dias, por 100 000 habitantes, com sobreposição semanal*”, para cada concelho de Portugal Continental. Foi usada como variável independente o “*Número de dias desde o dia 25 de março de 2020*”. Portugal Continental está dividido em 278 concelhos (NUTS-2013), indexados por $j = 1, \dots, 278$. Não foram incluídas as regiões autónomas. Os dados de georreferenciação foram retirados do *site* da DGSS PT.

Foram ainda usadas covariáveis de natureza demográfica: *densidade populacional*, *população por faixas etárias* para estimar o *Número esperado de casos* em Portugal Continental, o *Índice de Dependência dos Idosos* (PORDATA, 2019); sociológica: *rendimento médio*; e meteorológicas: *temperatura média (máxima e mínima)*, *humidade média (máxima e mínima)* e *pluviosidade* (IPMA).

As covariáveis meteorológicas foram usadas com desfasamento temporal. Seguindo a literatura (Elias et al., 2021; Lauer et al., 2020; Wassie et al., 2020; Zaki & Mohamed, 2021) para o tempo de incubação do SAR-COV-2, considerou-se, o parâmetro *lag* de 2, 7 e de 14 dias para a defasagem temporal das variáveis meteorológicas.

Foram ainda usadas as variáveis *Concelho*, *Dico* (código do concelho), *Latitude* e *Longitude* para a localização geográfica dos concelhos de Portugal Continental.

Para os efeitos espacial e temporal, estruturados e não estruturados foram criadas as variáveis *Region_str*, *Region_unstr*, *Date_str*, *Date_unstr* e *region-date*.

Na base de dados usada neste capítulo consideram-se as variáveis da Tabela 5.1.

Tabela 5.1: Variáveis em análise.

Descrição	Variável	Classificação da variável
Data	Dia/Mês/Ano	Ordinal
Dados sobre casos COVID-19	Nº cumulativo de casos diários	Contagem
	Nº cumulativo semanal de casos	Contagem
	Incidência a 14 dias, em 100 000 habitantes	Escala
Característica da População dos concelhos	População (2019) (Nº de habitantes)	Contagem
	Densidade populacional (2019) (Nº médio de indivíduos por km^2)	Escala
	População, por faixas etárias (]0,10[, [10,20[, [20,30[, [30,40[, [40,50[, [50,60[, [60,10[, [70,80[, [80, +∞])	Escala
	População com 65 ano ou mais (OMS, 2020)	Escala
	Índice de Dependência de Idosos	Escala
Condições meteorológicas	Temperatura máxima média	Escala
	Temperatura mínima média	Escala
	Humidade máxima média	Escala
	Humidade mínima média	Escala
	Pluviosidade média	Escala
Localização geográfica	Latitude	Escala
	Longitude	Escala

A faixa etária dos 65 anos ou mais é considerada pela OMS como um fator de risco.

5.3. Desagregação da série temporal nos períodos de 7 de julho a 26 de outubro de 2020 e de 27 de outubro de 2020 a 23 de março de 2021

Os casos COVID-19 nos Concelhos tiveram frequência diária de 25 de março de 2020 a 4 de julho de 2020, sendo os relatórios disponibilizados diariamente; frequência semanal de 7 de julho de 2020 a 26 de outubro de 2020, sendo os relatórios disponibilizados semanalmente; e incidência cumulativa a 14 dias, em 100 000 habitantes de 11 de novembro de 2020 a 23 de março de 2021, reportados semanalmente, o que implica que haja dados sobrepostos, com sobreposição semanal.

A título exemplificativo, na Figura 5.1 e na Figura 5.2 estão representados o número semanal de casos COVID-19 dos 18 concelhos do distrito do Porto e o número de casos

semanal, acumulados a 14 dias, respetivamente, referentes ao período de 07/07/2020 a 26/10/2020. Na Figura 5.3 e na Figura 5.4 estão representados o número semanal de casos COVID-19 dos 18 concelhos do distrito do Porto e o número de casos semanal, acumulados a 14 dias, respetivamente, referentes ao período de 11/11/2020 a 23/03/2021.

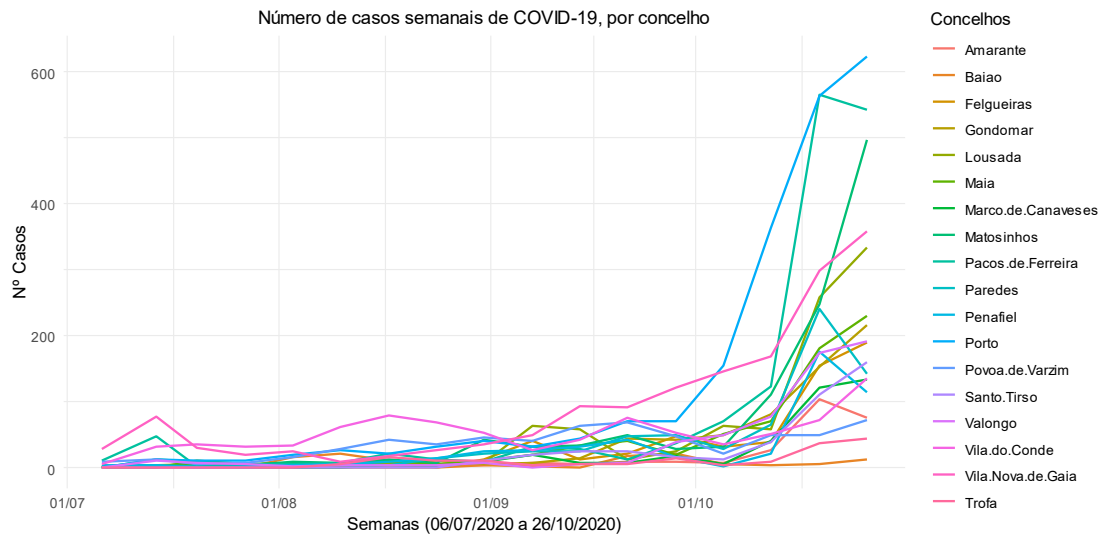


Figura 5.1: Dados semanais de COVID-19, por concelho do distrito do Porto, observados no período de 6/7/2020 a 26/10/2020.

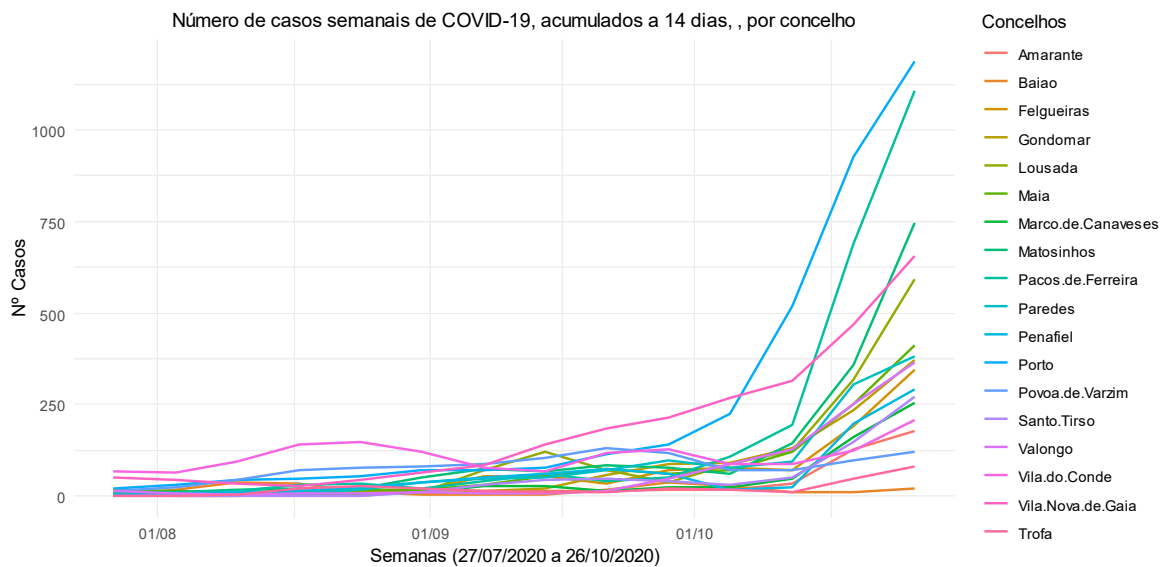


Figura 5.2: Número de casos semanais de COVID-19, acumulados a 14 dias, por concelho do distrito do Porto, observados de 14/7/2020 a 26/10/2020.

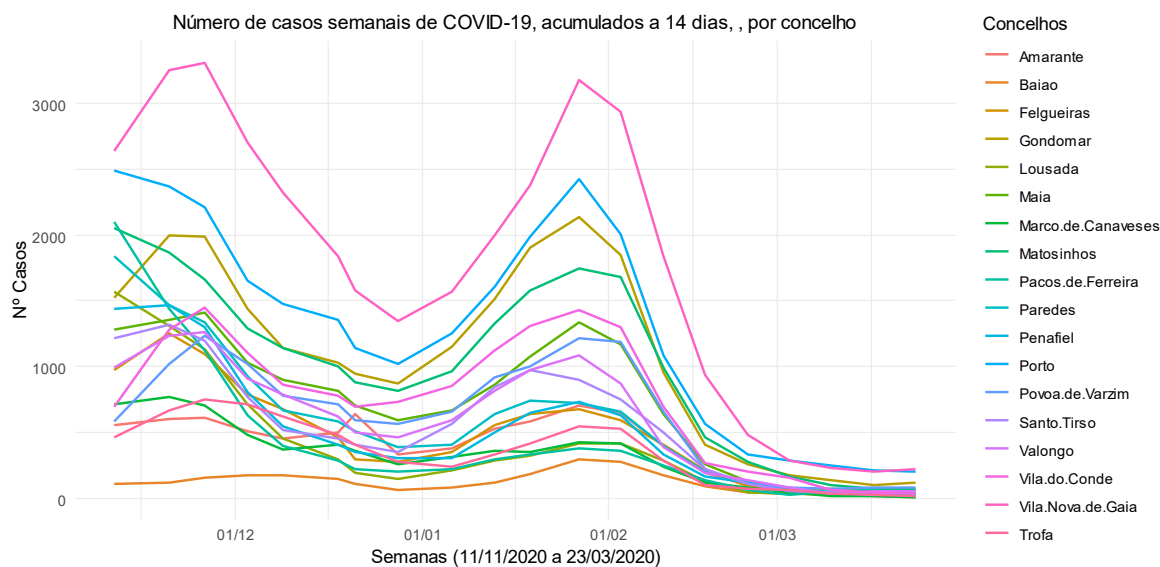


Figura 5.3: Número de casos semanais de COVID-19, acumulados a 14 dias, por concelho do distrito do Porto, observados de 11/11/2020 a 23/03/2021.

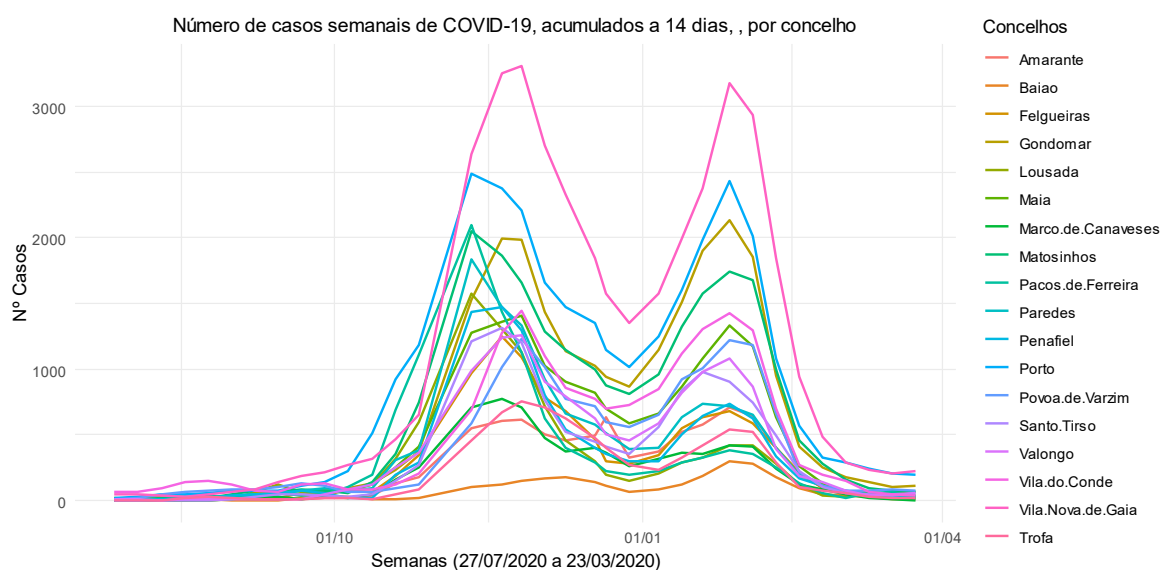


Figura 5.4: Número de casos semanais de COVID-19, acumulados a 14 dias, por concelho do distrito do Porto, observados de 27/7/2020 a 23/03/2021.

A primeira metodologia de desagregação explorada pressupõe que os casos COVID-19 se distribuem uniformemente em cada período de agregação, ou seja, que os dados semanais seguem uma distribuição uniforme $U(0,7)$ e os dados de incidência a 14 dias, depois de convertidos em número absoluto de casos em 14 dias, seguem uma distribuição uniforme $U(0,14)$. Foram depois exploradas as metodologias LS e HF descritas no capítulo 2, secção 2.10.7(ii).

5.3.1. Desagregação de dados pressupondo a distribuição uniforme

Uma das abordagens exploradas inicialmente para harmonizar os dados em todo o período de 25 de março de 2020 a 23 de março de 2021 foi usar os dados tal como disponibilizados pela DGS até 4 de julho e daqui até 26 de outubro usar a distribuição uniforme $U(0,7)$ para desagregar o número de novos casos semanais para o número de novos casos diários.

O primeiro relatório com incidência a 14 dias diz respeito aos dados de 28 de outubro a 10 de novembro de 2020 (boletim 259 da DGS de 16/11/2020 (DGS, 2020)). Os dados foram transformados desde esta data até 23 de março de 2021 (boletim 392 da DGS de 29/03/2021 (DGS, 2020)) da forma que a seguir se descreve.

A partir da incidência a 14 dias, em cem mil habitantes, Inc_{14} , determinou-se o número de novos casos de cada quinzena i , com $i = 1, 2, \dots, 20$.

$$Inc_{14} \times \frac{p_j}{p}, \quad j = 1, 2, \dots, 278 \quad (5.1)$$

onde p_j e p representam a população de cada concelho e a população portuguesa total (estimativa do INE em 2019).

Assumindo uma distribuição uniforme para o número de novos casos nos 14 dias de cada quinzena, $U(0,14)$, estimou-se o número diário de novos casos em cada intervalo de tempo sobreposto. Para tal, tomou-se a média simples do número diário de novos casos estimado pela distribuição uniforme de cada quinzena q (neste caso, metade dos casos dessa quinzena) e o número diário de novos casos estimado pela distribuição uniforme da próxima quinzena ($q + 1$) (igualmente, metade do número de casos dessa quinzena).

Assim, seja x_{di} o número de novos casos do dia d em cada semana i , com $d = 1, 2, \dots, 7$ e $i = 1, 2, \dots, 21$, a partir de 4 de novembro de 2020. Note-se que na semana de 28 de outubro a 3 de novembro, o número de novos casos diários estimado é $\frac{1}{14}$ do número cumulativo de novos casos da primeira quinzena com dados sobre incidência a 14 dias, em 100 000 habitantes.

Sendo a_{di} o número de novos casos diários em cada semana i , estimado pela distribuição uniforme $U(0, 14)$ com o número cumulativo de novos casos nos últimos quatorze dias em cada quinzena i e $i - 1$, tem-se (Figura 5.5):

$$x_{di} = \frac{a_{(d+7)(i-1)} + a_{di}}{2} \quad (5.2)$$

com $i = 2, 3, \dots, 21$ e $d = 1, 2, \dots, 7$.

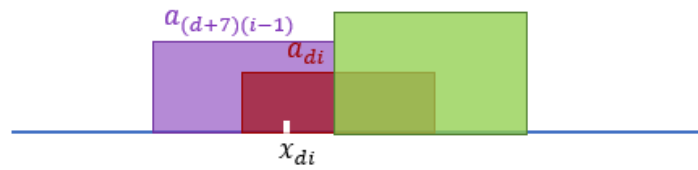


Figura 5.5: Linha do tempo que exemplifica a sobreposição de quinzenas. A caixa lilás representa a quinzena $i - 1$ com $a_{(d+7)(i-1)}$ novos casos diários estimados dessa quinzena e a caixa vermelha representa a quinzena i com a_{di} novos casos diários estimados da respetiva quinzena. x_{di} representa o número diário de novos casos estimados no dia d da semana i .

A partir das estimativas x_{di} , estimam-se valores para a variável y_t que dá o número cumulativo de casos diários, sendo t o número de dias desde 25 de março de 2020 até qualquer dia a partir de 28 de outubro, adicionando a y_{t-1} o número de casos x_t .

Esta metodologia pode ser aplicada assumindo qualquer outra distribuição que se considere adequada para o número cumulativo de casos por quinzena.

Como a 26 de outubro os dados vinham sendo reportados semanalmente, se se pretender continuar com esta agregação, usa-se o número cumulativo semanal de casos, assumindo a distribuição uniforme. Neste caso, para estimar o número cumulativo de casos semanais basta tomar para cada semana i , com $i = 1, 2, \dots, 20$, x_i tal que:

$$x_i = \frac{b_i + b_{i+1}}{4} \tag{5.3}$$

sendo b_k o número cumulativo de casos da quinzena k .

Neste caso, a partir das estimativas x_i , estimam-se valores para a variável y_t que dá o número cumulativos de casos diários, sendo t o número de dias desde 25 de março de 2020 até qualquer dia a partir de 28 de outubro, adicionando sucessivamente a cada y_{t-1} o número de casos do dia seguinte.

Na Figura 5.6 e na Figura 5.7 estão representados os dados diários dos 18 concelhos do distrito do Porto, do período de 07/07/2020 a 26/10/2020 e de 27/10/2020 a 23/03/2021, desagregados usando a distribuição uniforme, conforme descrito atrás.

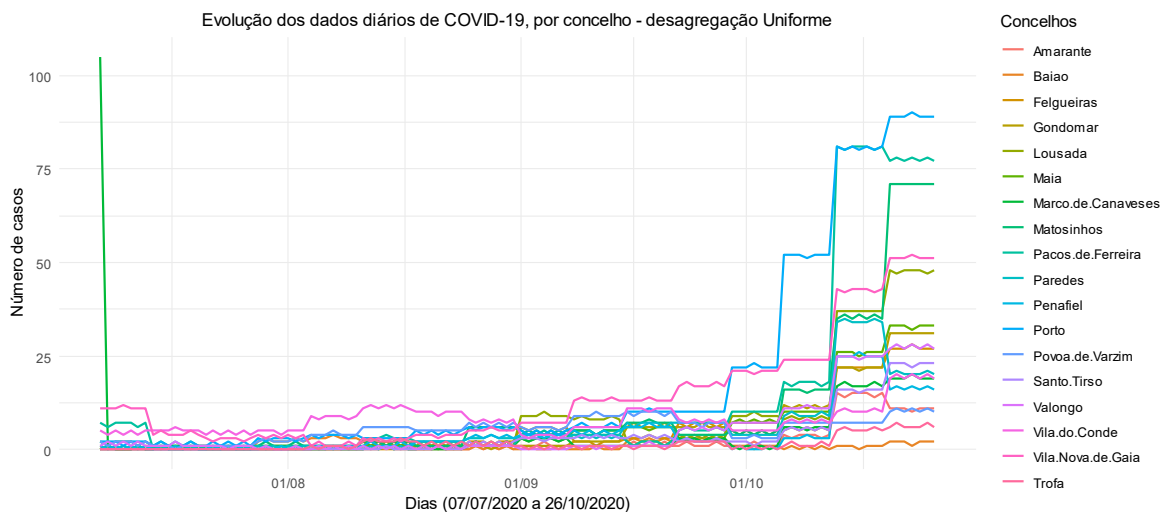


Figura 5.6: Dados diários do período de 07/07/2020 a 26/10/2020, desagregados de dados de frequência semanal com recurso ao à distribuição uniforme $U(0, 7)$.

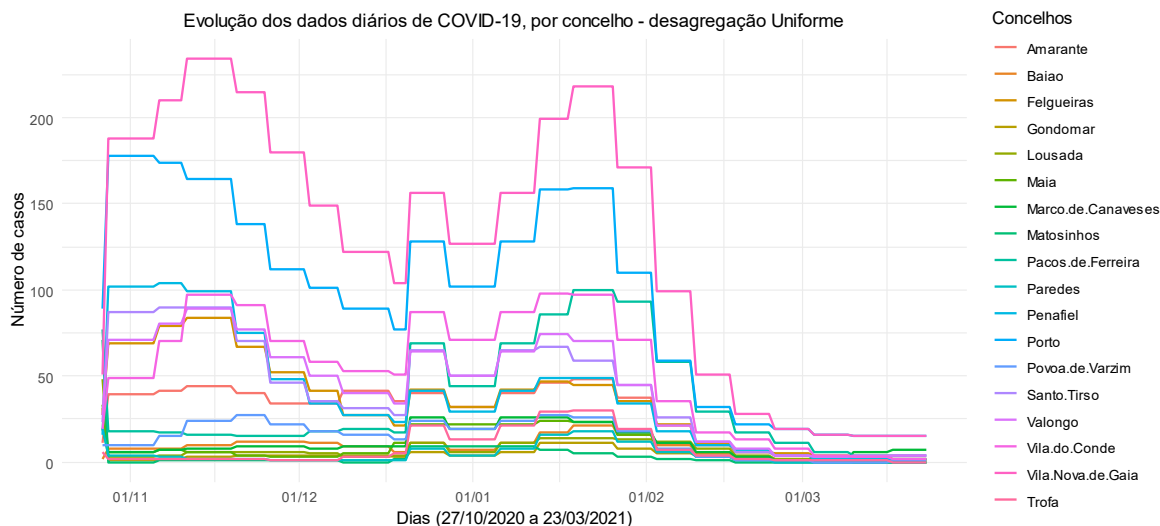


Figura 5.7: Dados diários, do período de 27/10/2020 a 23/03/2021, desagregados de dados acumulados a 14 dias, relatados semanalmente, com recurso à distribuição uniforme $U(0, 14)$.

5.3.2. Desagregação recorrendo aos algoritmos dos Mínimos Quadrados e H-FUSE descritos na secção 3.4

A. Desagregação da série temporal no período de 7 de julho a 26 de outubro – sem sobreposição

Neste período considera-se o problema definido em (2.158), para obter a solução que otimiza a função definida em (2.157).

O problema de otimização consiste em estimar a série temporal $\vec{x} = (x_1, x_2, \dots, x_{112})$, desconhecida, que pode ser agregada na série temporal $\vec{y} = (y_1, y_2, \dots, y_{16})$ de dados agregados observada e que minimiza a função $F(\vec{x})$

$$F(\vec{x}) = \|\vec{y} - C\vec{x}\|_2^2 \quad (5.4)$$

Para resolver a equação $F(\vec{x}) = 0$, é usado o método dos mínimos quadrados (LSQ – *Least Square method*) para procurar uma solução aproximada \vec{x} que satisfaz F . Na realidade, este método origina várias soluções pois este problema é subdeterminado. O método escolhe a solução que minimiza $F(\cdot)$. Quando a equação é subdeterminada, a solução apresentada é a que melhor se ajusta aos dados, uma vez que minimiza a soma dos quadrados dos resíduos. A solução que tem a norma mínima ($\min \|\vec{x}\|_2^2$) é assumida como a “melhor” solução, a que equilibra as incógnitas de forma a minimizar a influência das equações com maiores erros residuais, resultando na que se ajusta de forma mais equilibrada aos dados. No entanto, na maioria dos dados reais, a solução verdadeira não é necessariamente aquela que possui (ou está próxima da) norma mínima. Para obter esta solução de forma computacionalmente mais eficaz usa-se a matriz pseudo-inversa de Moore-Penrose (Penrose, 1955).

Os dados, no período de 07/07/2020 a 26/10/2020, estão agregados à semana e sem sobreposição, como ilustra a Figura 5.2.

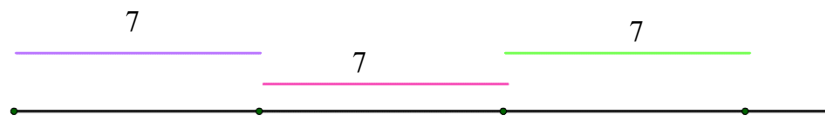


Figura 5.8: Cada segmento de reta corresponde a uma semana, ou seja, corresponde a sete períodos da escala fina diária.

Como se pretende desagregar os dados para a escala diária, a matriz C é definida por:

$$C_{(16 \times 112)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & & & & 0 & 0 & \dots & 0 & & & & \dots & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (5.5)$$

Na Figura 5.1 estão representados os casos semanais de COVID-19 dos 18 concelhos do distrito do Porto referentes ao período de 07/07/2020 a 26/10/2020.

Os dados foram desagregados com recurso ao algoritmo dos Mínimos Quadrados, com a matriz pseudo-inversa de Moore-Penrose (Penrose, 1955) e o algoritmo H-Fuse (Liu et al., 2017), com suavização temporal. Os códigos em R encontram-se no Anexo X.

Nas figuras Figura 5.9 e Figura 5.10 encontram-se as representações dos dados diários dos 18 concelhos do distrito do Porto, do período de 07/07/2020 a 26/10/2020, desagregados a partir dos dados com agregação semanal, com recurso aos algoritmos dos Mínimos Quadrados e H-Fuse.

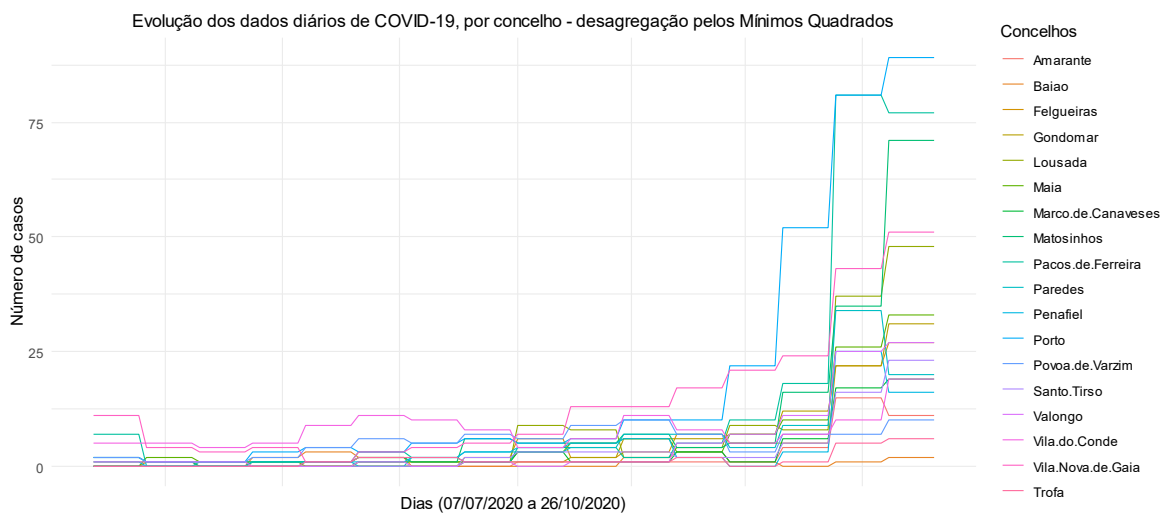


Figura 5.9: Dados diários do período de 07/07/2020 a 26/10/2020, desagregados de dados de frequência semanal com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.

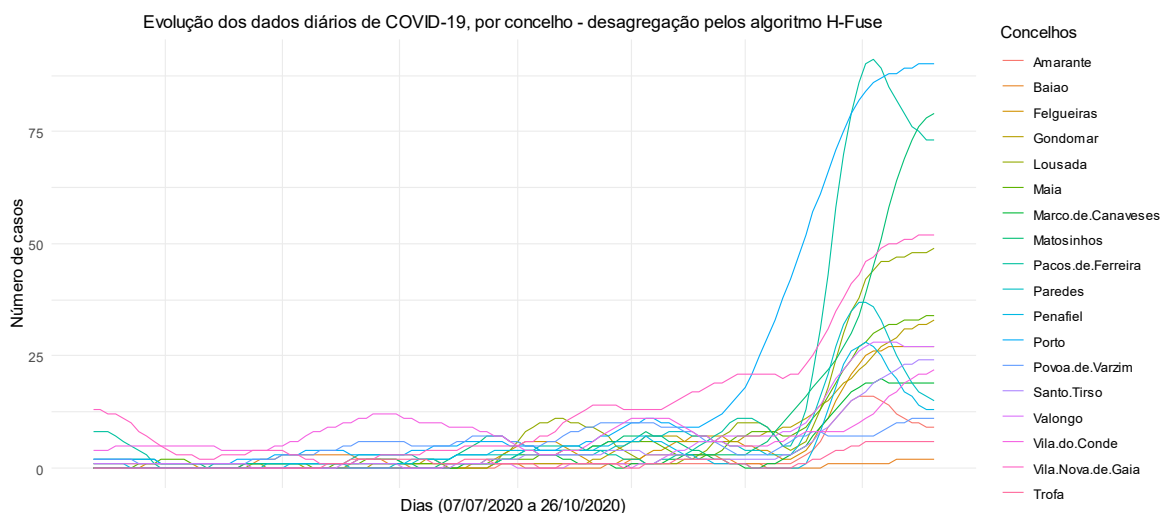


Figura 5.10: Dados diários do período de 07/07/2020 a 26/10/2020, desagregados de dados de frequência semanal com recurso ao algoritmo H-Fuse.

Quadrados e H-Fuse e os resultados encontram-se representados na Figura 5.12 e na Figura 5.13.

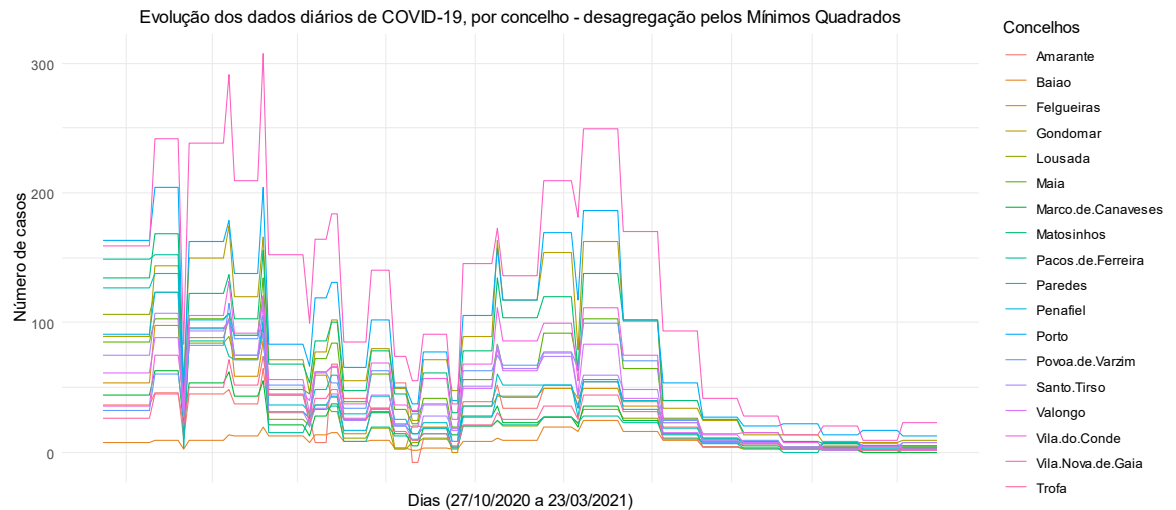


Figura 5.12: Dados diários, do período de 27/10/2020 a 23/03/2021, desagregados de dados acumulados a 14 dias, relatados semanalmente, com recurso ao método dos Mínimos Quadrados com Pseudo-Inversa de Moore-Penrose.

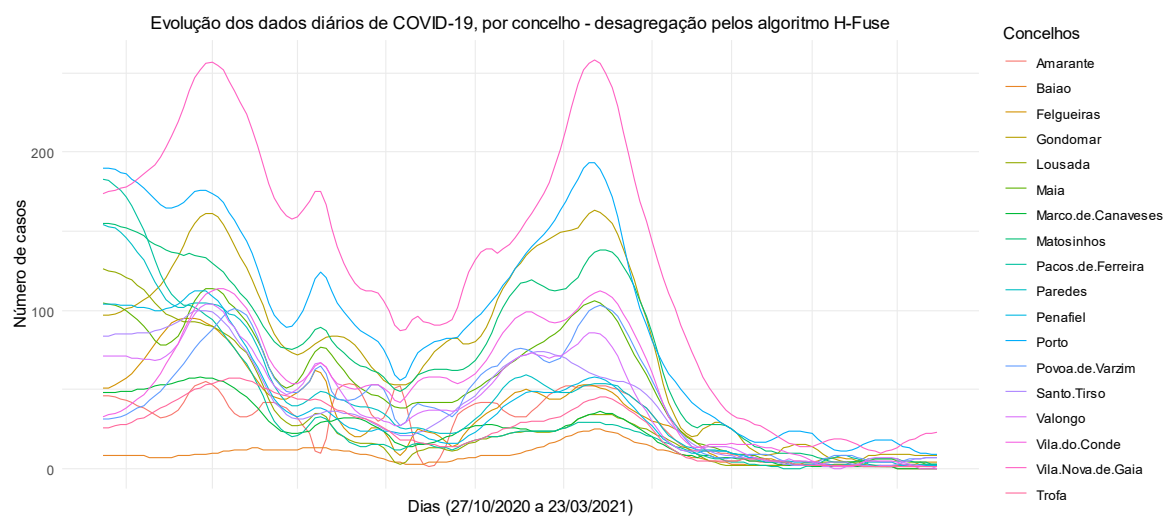


Figura 5.13: Dados diários, do período de 27/10/2020 a 23/03/2021, desagregados de dados acumulados a 14 dias, relatados semanalmente, com recurso ao algoritmo H-Fuse.

Não é possível comparar os dados assim desagregados com os dados originais, uma vez que não se dispõe dos dados diários originais. A análise comparativa dos dados desagregados com os dados reais para o período de 25 de março a 30 de junho de 2020 foi feita na secção 4.2. O que se verifica, mais uma vez, é que a desagregação apresenta dinâmicas mais suavizadas com a desagregação feita com o algoritmo H-Fuse e que, de acordo com os

resultados da secção 4.2, é o método de entre os utilizados, que oferece resultados mais precisos na desagregação.

5.4. Modelação espaço-temporal Bayesiana avançada de nível local por meio da estrutura multinível para estratégias de prevenção e controle de COVID-19.

No problema a tratar nesta secção interessa estimar o efeito de covariáveis selecionadas sobre o risco relativo da COVID-19 ao nível dos concelhos, tendo em consideração as correlações temporais e espaciais. Interessa também estudar o efeito da desagregação da série temporal na distribuição espacial e temporal do risco relativo.

Uma medida simples de risco de doença é a taxa de incidência padronizada (SIR) (Moraga, 2020), definida para cada concelho $i, i = 1, 2, \dots, 278$, como a proporção entre o número de casos COVID-19 observados, Y_i , e o número de casos esperado, E_i (equação (4.4)).

$$SIR_i = \frac{Y_i}{E_i} \tag{5.8}$$

Consoante o valor de SIR_i é maior que 1, menor que 1 ou igual a 1, o risco de contrair COVID-19 na área i é maior, menor ou igual ao da população em geral.

Em áreas pequenas o número esperado de casos pode ser muito pequeno e o SIR pode dar informação pouco confiável.

Uma das formas mais gerais e úteis de especificar este problema é usar a configuração de um modelo linear generalizado hierárquico, no qual os dados são ligados a covariáveis e efeitos aleatórios espaço-temporais através de uma verosimilhança apropriadamente escolhida e uma função de ligação que é linear nas covariáveis e nos efeitos aleatórios. Estes modelos permitem usar informação de concelhos vizinhos e de covariáveis, o que resulta na suavização dos resultados. (A. B. Lawson, 2018).

Usou-se o pacote R-INLA, do *software* R (Martins et al., 2013; Rue et al., 2009), para explorar Modelos Gaussianos Latentes, com e sem efeitos espaciais, para estimar o risco de COVID-19 nos 278 concelhos de Portugal Continental, em função das covariáveis meteorológicas: *temperatura máxima média, temperatura mínima média, humidade máxima média, humidade mínima média e pluviosidade média*, com desfasamento temporal de dois dias, sete dias e catorze dias, tendo em conta os dados da literatura sobre o período de incubação do vírus que varia de 2 a 14 dias (Lauer et al., 2020).

5.4.1. Especificação do modelo

Foi utilizado um modelo hierárquico de regressão de Poisson, especificado do seguinte modo:

$$y_{jt} | \rho_{jt} \sim \text{Poisson}(E_{jt} \rho_{jt}) \quad (5.9)$$

com função de ligação

$$\eta_{jt} = \log(\rho_{jt}) = \alpha_0 + \sum \beta_{ij} x_{ij} + u_j + v_j + \gamma_t + \varphi_t \quad (5.10)$$

onde E_{jt} é o número esperado de casos da doença para o concelho j , $j = 1, 2, \dots, 278$, durante o dia t , $t = 1, 2, \dots, 365$; ρ_{ij} é o risco relativo de contrair COVID-19 no concelho j e no dia t , e y_{jt} é o número de cumulativo/diário de casos COVID-19 no concelho j e dia t , que se assume ser condicionalmente independente; α_0 é o intercepto que quantifica o logaritmo do risco global, v_j é um efeito não estruturado, específico da área, modelado como permutável (a distribuição de probabilidade conjunta das observações é invariante sob permutação) que modela o ruído não correlacionado, u_j é outro efeito específico da área, modelado como espacialmente estruturado, que modela a dependência espacial entre os riscos relativos, φ_t é um efeito não estruturado, específico do tempo, modelado como permutável e γ_t é outro efeito específico do tempo, modelado como temporalmente estruturado tal como descrito em (2.93). $\sum \beta_{ij} x_{ij}$ representa os efeitos fixos, especificados por covariáveis x_j , $j = 1, 2, \dots, 278$, observadas nos concelhos e que quantificam possíveis fatores de risco. Pode ainda ser incluído um efeito δ_{ij} , de interação entre o tempo e o espaço, como definido em 2.97.

E_{jt} é calculado por padronização indireta como descrito na equação (4.4), usando a incidência em Portugal continental no dia t , por faixas etárias, e a distribuição da população de cada concelho j , acautelando assim as diferenças que se foram revelando nas incidências por faixas etárias e as características demográficas muito diferenciadas entre concelhos.

O risco relativo é especificado como uma função de efeitos fixos (ou seja, de covariáveis), efeitos aleatórios espaciais, temporais e efeitos de interação espaço-temporal.

Modelação de efeitos fixos.

Os efeitos fixos foram modelados, para cada concelho j e cada dia t , como combinação linear de covariáveis meteorológicas, demográficas e socioeconómicas, $X_{jt}^T \beta$. Para os parâmetros β de efeitos fixos foi considerada uma distribuição *a priori* informativa fraca

$\beta \sim N(0, \tau_\beta^{-1}I)$ com precisão pequena $\tau_\beta = 10^{-6}$ (valor assumido por defeito no R-INLA, assumindo a matriz identidade).

Uma covariável está significativamente associada ao risco de contrair COVID-19 se o intervalo de credibilidade (IC) correspondente aos seus efeitos fixos não contiver zero.

Modelação de efeitos espaciais aleatórios.

As tendências espaciais foram modeladas como a soma das heterogeneidades dos concelhos e os efeitos de agrupamento espacial. Para os efeitos aleatórios espaciais não estruturados, v_j , foi assumido um modelo latente independente e identicamente distribuído (IID) tal que $v_j \sim N\left(0, \frac{1}{\tau_{v_j}}\right)$. Esses efeitos aleatórios espaciais controlam globalmente a variabilidade adicional nos logaritmos dos riscos relativos. Sob este modelo, o efeito v_j para cada concelho j é independente de todos os outros concelhos. Para os efeitos aleatórios espaciais estruturados, u_j , assumiu-se uma especificação Besag-York Mollie (BYM) (Besag et al., 1991), em que u_j é modelado usando um modelo autorregressivo condicional intrínseco de estrutura (*iCAR*) como definido em (2.94):

$$u_j | u_{-j} \sim Normal\left(\frac{1}{N_j} \sum_{j=1}^n w_{ij} u_j, \frac{1}{\tau_{u_j}}\right) \quad (5.11)$$

onde τ_{u_j} e N_j são o parâmetro de precisão no concelho j , inverso da variância, e o número de vizinhos do concelho j , respetivamente. Sob este modelo latente, o efeito de u_j para cada concelho j é distribuído normalmente, com efeito médio igual à média dos efeitos dos vizinhos do concelho j e precisão τ_{u_j} . Com este modelo, a matriz de adjacência foi usada para caracterizar as relações espaciais entre os concelhos. Os vizinhos são definidos em termos de concelhos que compartilham pelo menos uma fronteira e o peso é definido como 1 se dois concelhos são vizinhos, e zero, caso contrário. As distribuições *a priori* para as precisões dos efeitos aleatórios espaciais não estruturados e estruturados foram assumidas como distribuições Gama inversas com parâmetros padrão do pacote R-INLA:

$$(t_{v_j}^2; \sim IG(0.5, 0.00005) \text{ e } t_{u_j}^2 \sim IG(10^{-2}, 10^{-2}) \quad (5.12)$$

Modelação de efeitos temporais e de interação espaço-temporal.

Os efeitos temporais podem ser modelados utilizando o tempo como uma variável categórica por meio da introdução de variável *dummy* – usou-se a variável *Date* que codificou cada dia de 25 de março de 2020 a 23 de março de 2021 com *Date*: 1,2,..., 364. Para modelar os efeitos temporais pode ser usada uma tendência linear paramétrica ou utilizar formulações não paramétricas, como o alisamento por médias móveis, para relaxar o pressuposto de tendências lineares. Podem ainda ser usados modelos autorregressivos como os passeios aleatórios AR(1) ou AR(2). Neste estudo, optou-se por especificar os efeitos temporais utilizando uma formulação por meio de um passeio aleatório de segunda ordem (efeito temporal estruturado γ_t), tal como definidos em 2.96 e um modelo gaussiano latente permutável (efeito temporal não estruturado φ_t).

$$\gamma_t | \gamma_{t-1}, \gamma_{t-2} \sim \text{Normal}(2\gamma_{t-1} + \gamma_{t-2}, \sigma_\gamma^2) \quad (5.13)$$

$$\varphi_t \sim \text{Normal}\left(0, \frac{1}{\tau_\varphi}\right) \quad (5.14)$$

Uma interação entre área e tempo é modelada expandindo os efeitos temporais através da adição de um termo de interação (δ_{it}). Esse termo de interação explica as diferenças na tendência temporal para diferentes concelhos, como definido em 2.97.

Na Figura 5.14 está representado o grafo que ilustra o Modelo Hierárquico Bayesiano que esteve na base da modelação dos dados de COVID-19 de concelhos de Portugal Continental, no período de 25 de março de 2020 a 23 de março de 2023. Na

Tabela 5.2 estão definidas as variáveis usadas no modelo.

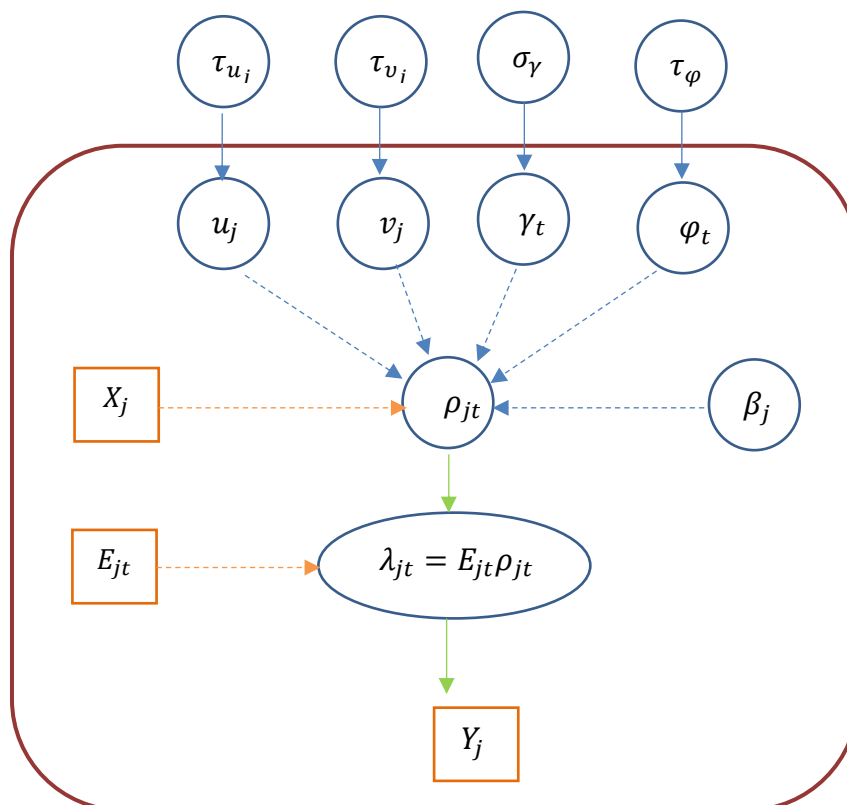


Figura 5.14: Grafo direto acíclico que ilustra a estrutura hierárquica do modelo implementado

Tabela 5.2: Definição das variáveis usadas no modelo

Variável	Definição
Y_j	Vetor de contagem observada
X_j	Vetor de covariáveis
E_j	Nº esperado de casos COVID-19
ρ_{jt}	Risco relativo de contrair COVID-19
β_j	Vetor de coeficientes dos efeitos fixos
u_j, v_j	Efeitos espaciais aleatórios
γ_t, φ_t	Efeitos aleatórios temporais
δ_{jt}	Efeitos de interação <i>tempo</i> × <i>espaço</i>
λ_{jt}	Média de Poisson
$\tau_u, \tau_v, \tau_\varphi, \sigma_\gamma$	Precisão e/ou variância

5.4.2. Resultados e discussão

5.4.2.1. Modelos ajustados com os dados dos concelhos do distrito do Porto, de 25 de março de 2020 a 30 de junho de 2020

Inicialmente investigou-se a distribuição espacial e temporal do risco de contrair COVID-19 nos diferentes concelhos do distrito do Porto de 25 de março de 2020 a 30 de junho de 2020, e a possível influência de covariáveis meteorológicas e demográficas na distribuição espacial e temporal do risco de contrair COVID-19. Investigou-se ainda a influência dos métodos de desagregação de séries temporais nos resultados dos modelos.

Modelos sem covariáveis

Começou-se por ajustar ao número de casos diários de COVID-19 um modelo sem covariáveis, apenas com efeitos temporais, espaciais e espaço-temporais.

Tabela 5.3: Modelo sem covariáveis ajustados aos dados. μ é a o intesepto, E é o número de casos esperado, u_j e v_j representam os efeitos espaciais estruturados e não estruturados, γ_t e φ_t são os efeitos temporais estruturados e não estruturados e δ_{jt} é o efeito de interação *espaço* \times *tempo*.

Modelo	$\log(\eta)$	Efeitos fixos
<i>Bym.rw.iid.M0</i>	$\mu + \log(E) + u_j + v_j$ $+ \gamma_t + \varphi_t + \delta_{jt}$	Nenhum

Tabela 5.4: Os resultados de DIC, WAIC e *Log-Likelihood* marginal dos modelos apenas com efeitos temporais e espaciais (estruturados e não estruturados), ajustados com os dados observados e com os dados agregados semanalmente e desagregados pelos métodos LS e HF.

Modelo	<i>Marginal Log-likelihood</i>			
	DIC	WAIC	Estimação por integração	Estimação Gaussiana
<i>M0</i>	2325.8934	94338.590	629.86460	5256954.9
<i>M0.HF</i>	-4939.2695	79028.594	2475.11175	4787146.0
<i>M0.LS</i>	-5606.3946	58221.929	2620.15891	482695.9

Em todas as situações os efeitos temporais e espaciais revelaram ter um efeito positivo no risco de contrair COVID-19, bem como a interação *espaço* \times *tempo* (resultados em Anexo V).

Modelos com covariáveis

Foram ajustados aos dados diários de COVID-19 modelos com as covariáveis: *temperatura mínima (valores mínimo, médio e máximo), temperatura máxima (valores mínimo, médio e máximo), humidade (valores mínimo, médio e máximo) e pluviosidade (valores mínimo, médio e máximo)*, com atrasos de 2 dias, e *temperatura mínima média, temperatura máxima média, humidade mínima média, humidade máxima média, pluviosidade média*, com atraso de 7 dias e de 14 dias, *densidade populacional, ganho médio mensal, Índice de Dependência dos Idosos (IDI)* e faixa etária dos maiores de 65 anos.

Começou-se por ajustar um modelo geral com todas as covariáveis e com todos os efeitos espaciais e temporais, estruturados e não estruturados, usando os dados observados, com as granularidades originais e os dados desagregados a partir dos dados agregados semanais. Neste modelo há variáveis correlacionadas.

Tabela 5.5: μ é o intesepto do modelo, E é o número de casos esperado, $(X_i), i = 1, 2, \dots, 29$ são as covariáveis e u_j e v_j representam os efeitos espaciais estruturados e não estruturados, γ_t e φ_t são os efeitos temporais estruturados e não estruturados e δ_{jt} é o efeito de interação *espaço \times tempo*.

Modelo	$\log(\eta) =$
<i>iid.M0 (IID)</i>	$\mu + \log(E) + \sum_{i=1}^{29} \beta_i X_i + u_j + v_j$ $+ \gamma_t + \varphi_t + \delta_{jt}$

Tabela 5.6: Os resultados dos valores de DIC, WAIC e *Log-Likelihood* marginal do modelo geral ajustado com os dados observados e com os dados agregados à semana e desagregados pelos métodos LS e HF. Os modelos foram ajustados com todas as covariáveis, e com os efeitos temporais e espaciais.

Modelo	<i>Marginal Log-likelihood</i>			
	DIC	WAIC	Estimação por integração	Estimação Gaussiana
<i>MG</i>	4689.9308	96316.972	929.93399	932.99734
MG.HF	-2098.0965	88118.362	2500.06692	2503.11525
<i>MG.LS</i>	-2586.0556	83935.367	2669.82130	2672.90110

Os resultados do modelo encontram-se no Anexo V.

Finalmente, foram ajustados modelos com as covariáveis *temperatura máxima média, temperatura mínima média, humidade máxima média, humidade mínima média, pluviosidade média*, com atrasos de 2, 7 e 14 dias, *densidade populacional, ganho médio mensal, índice de dependência do idoso* e faixa etária dos mais de 65 anos. Nas covariáveis

meteorológicas foram consideradas apenas os valores médios uma vez que os valores mínimos e máximos destas covariáveis estão correlacionados entre si e com os valores médios (exemplos na Tabela 5.7). A matriz de correlações entre todas as variáveis meteorológicas encontra-se no anexo VI.

Tabela 5.7: Matriz de correlações entre as variáveis meteorológicas usadas.

	<i>tM.mean.2</i>	<i>tm.mean.2</i>	<i>hM.mean.2</i>	<i>hm.mean.2</i>	<i>pr.mean.2</i>
<i>tM.mean.2</i>	1.000	0.587	-0.318	-0.347	-0.325
<i>tm.mean.2</i>	0.587	1.000	-0.181	0.212	0.053
<i>hM.mean.2</i>	-0.318	-0.181	1.000	0.537	0.229
<i>hm.mean.2</i>	-0.347	0.212	0.537	1.000	0.380
<i>pr.mean.2</i>	-0.325	0.053	0.229	0.380	1.000

Foram considerados diferentes modelos com efeitos temporais e espaciais (Tabela 5.8).

Tabela 5.8: μ é o intesepto do modelo, E é o número de casos esperado, \mathbf{u}_j e \mathbf{v}_j representam os efeitos espaciais estruturados e não estruturados e $\boldsymbol{\gamma}_t$ e $\boldsymbol{\varphi}_t$ são os efeitos temporais estruturados e não estruturados.

Modelo	$\log(\boldsymbol{\eta}) =$	Prior (efeitos temporais)	Prior (efeitos espaciais)
<i>iid.M (IID)</i>	$\mu + \log(E) + \sum_{i=1}^9 \beta_i X_i + \varphi_t$	Não estruturado (φ_t)	-----
<i>rw2.M (RW)</i>	$\mu + \log(E) + \sum_{i=1}^9 \beta_i X_i + \gamma_t$	Estruturado (γ_t)	-----
<i>besag.M (CAR)</i>	$\mu + \log(E) + \sum_{i=1}^9 \beta_i X_i + (u_j + v_j)t$	-----	“besag” (CAR)(u_j, v_j)
<i>bym.M (iCAR)</i>	$\mu + \log(E) + \sum_{i=1}^9 \beta_i X_i + u_j + v_j$	-----	“Besag-York-Mollie” (iCAR)
<i>bym.rw.iid.M(iCAR)</i>	$\mu + \log(E) + \sum_{i=1}^9 \beta_i X_i + u_j + v_j + \gamma_t + \varphi_t + \delta_{jt}$	Estruturado: (γ_t) Não estruturado (φ_t)	(u_j) v_j “Besag-York-Mollie” (iCAR)

Os modelos foram ajustados com as covariáveis meteorológicas, com atraso de 2 dias, com atraso de 7 dias e com atraso de 14 dias, em modelos separados. Os resultados dos DIC, WAIC e *Marginal Likelihood* encontram-se no Anexo VIII e os dados dos modelos (médias e intervalos de credibilidade) encontram-se no Anexo V. Os valores de RMSE (*Root Mean Squared Error*) e o MAE (*Mean Absolute Error*) encontram-se no Anexo VII. Na maioria

dos casos, quando se comparam os modelos ajustados com dados desagregados, os que foram ajustados com os dados obtidos com desagregação de dados semanais, pelo algoritmo H-Fuse, têm o menor valor de RMSE e MAE, e nos casos em que tal não acontece, a diferença é muito ténue. Quando se consideram os modelos ajustados com os dados observados, com as granularidades originais, apenas os que incluem os efeitos temporais têm valores de RMSE e MAE menores. Considerando os valores de RMSE e MAE normalizados pelo desvio padrão (Anexo VII) verifica-se que os valores de RMSE para os modelos que incluem efeitos espaciais, temporais e interação são muito maiores que 1 e, portanto, estes modelos têm pouca capacidade preditiva.

O modelo que combina melhores valores de DIC (menor valor) e de *Marginal Likelihood* (maior valor) são os modelos *bym.rw.iid.M(iCAR)*, com atraso de dois dias nas variáveis meteorológicas (Anexo VIII). O modelo *bym.M (iCAR)*, com as variáveis meteorológicas com atraso de 14 dias tem valores menores de WAIC, mas valores piores nas outras duas métricas. Este resultado era espectável uma vez que o WAIC penaliza os modelos mais complexos. No entanto, a *Marginal Likelihood* é a métrica que garante melhor ajustamento do modelo aos dados.

Na Tabela 5.9, na

Tabela 5.10 e na Tabela 5.11 encontram-se os resultados dos modelos *bym.rw.iid.M(iCAR)* ajustados com os dados diários do número de casos COVID-19 dos concelhos do distrito do Porto observados, com as granularidades originais, e desagregados com os algoritmos HF e LS (valor médio e intervalos de credibilidade, a 95%, dos efeitos fixos e dos efeitos espacial e temporal). Os efeitos temporal e espacial não estruturados foram ajustados supondo os modelos *iid* (2.95, 2.97)2.95. Os efeitos temporal e espacial estruturados foram modelados com um passeio aleatório de grau 2 (2.97) e com o modelo *iCAR* (2.94). O efeito de interação foi modelado com o modelo *iid* (2.100). Pode-se observar que os resultados da significância estatística das covariáveis e dos efeitos espacial e temporal são os mesmos em quase todas as variáveis para os três tipos de dados. Quanto às covariáveis socioeconómica, *Ganho Médio Mensal (GMM.T.2018)* e demográfica, população com mais de 65 anos - *Mais.65*, pode-se assumir que têm um efeito nulo, estatisticamente não significativo, uma vez que o intervalo de credibilidade a 95% contém zero. O *Índice de Dependência dos Idosos - IDI.2019* tem um efeito negativo nos três modelos, mas estatisticamente não significativo. Nas covariáveis meteorológicas, apenas a *humidade*

máxima média difere no modelo ajustado com os dados observados (com as granularidades originais) em relação aos modelos ajustados com os dados desagregados. Esta covariável tem efeito negativo sobre o risco de doença nos três casos. No primeiro modelo o efeito não é estatisticamente significativo uma vez que o intervalo de credibilidade a 95% correspondente contém zero. Nos outros dois modelos o efeito é negativo, sendo estatisticamente significativo uma vez que o intervalo de credibilidade tem extremos negativos. No entanto, em todos os casos pode-se assumir que apenas as *temperaturas médias – máxima e mínima* têm um efeito positivo e negativo, respetivamente, estatisticamente significativo, pois nas restantes covariáveis os limites dos intervalos de credibilidade a 95% estão muito próximos de zero, podendo-se, portanto, assumir que os efeitos destas covariáveis (*humidade e pluviosidade*) são muito reduzidos. Na Figura 5.15 estão representadas as distribuições posteriores dos coeficientes das covariáveis *Temperaturas médias – máxima e mínima*, *Humidades médias – máxima e mínima* e *Pluviosidade média*, do modelo ajustado com os dados desagregados pelo algoritmo HF (Wickham, 2016). É possível verificar que apenas a *temperatura mínima média* tem efeito positivo sobre o risco de COVID-19, e esse efeito é estatisticamente significativo como a seguir se explica (intervalo de credibilidade com extremos positivos).

Os efeitos espaciais e temporais são positivos nos três modelos e estatisticamente significativos, especialmente os efeitos estruturados, o que sustenta a importância de considerar estes efeitos nas análises dos efeitos de covariáveis no número de casos de COVID-19, uma vez que estes se mostraram heterogêneos.

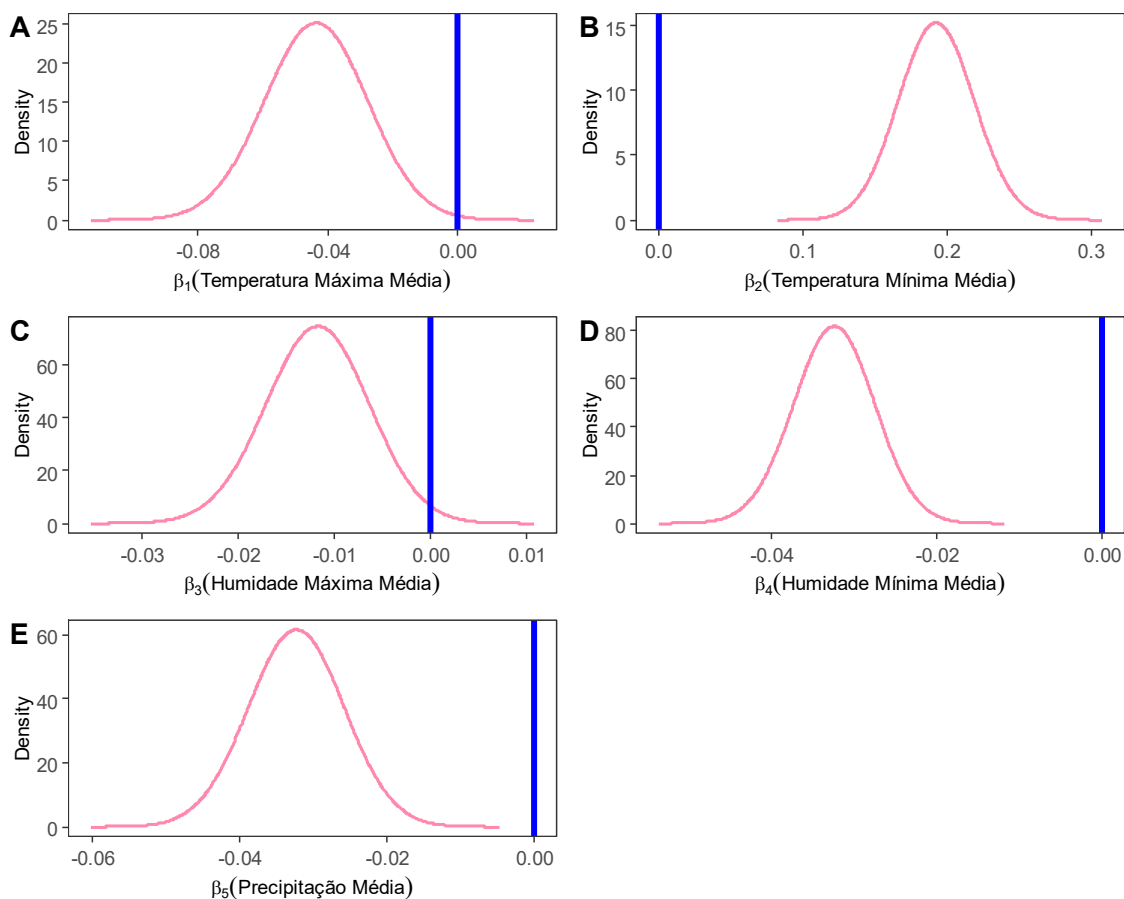


Figura 5.15: Distribuição posterior do coeficiente da covariável: **A** - Temperatura máxima média; **B** - Temperatura mínima média; **C** - Humidade máxima média; **D** - Humidade mínima média; **E** - Pluviosidade média.

Tabela 5.9: Média posterior e intervalos de credibilidade (a 95%) dos parâmetros e hiperparâmetros do modelo *bym.rw.iid.M.2*, ajustado com o número diário de casos COVID-19 observados nos concelhos do distrito do Porto, entre 25/03/2020 e 30/06/20 e com as covariáveis meteorológicas com atraso de dois dias.

Parâmetros	Média	sd	0.025Q	0.975Q
(Intercepto)	6.437	4.579	-2.649	15.525
tM.mean.2	-0.151	0.018	-0.187	-0.117
tm.mean.2	0.334	0.031	0.274	0.396
hM.mean.2	-0.004	0.006	-0.016	0.009
hm.mean.2	-0.044	0.005	-0.054	-0.034
pr.mean.2	-0.068	0.007	-0.081	-0.054
Popdensity	0.001	0.001	-0.001	0.003
GMM.T.2018	0.000	0.005	-0.010	0.009
IDI.2019	-0.172	0.098	-0.367	0.022
Mais.65	0.000	0.000	0.000	0.000

τ				
τ (Region_str (iid component))	0.3.86	0. 141	0.171	0.72
τ Region_str (spatial component)	2470	2730	207.122	9670
τ Date_str	13200	14000	1410.739	50200
τ Date_unstr	0. 511	0.088	0.360	0.706
τ region_date	23800	26100	1950.842	92900

Tabela 5.10: Média posterior e intervalos de credibilidade (a 95%) dos efeitos fixos e dos hiperparâmetros do modelo *bym.rw.iid.M.2*, ajustado com o número diário de casos COVID-19 desagregados dos dados agregados à semana com recurso ao algoritmo HF, nos concelhos do distrito do Porto, entre 25/03/2020 e 30/06/20 e com as variáveis meteorológicas com atraso de dois dias.

Modelo:				
iCAR.rw.iid.M2.2.HF : x.HF~tm.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+ Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+ f(Region_str, model="bym", graph=W.porto)+ f(Date_str, model="rw2") + f(Date_unstr, model="iid")+ f(region_date,model="iid")				
Parâmetros	Média	sd	0.025Q	0.975Q
(Intercept)	6.061	4.591	-3.053	15.176
tM.mean.2	-0.044	0.016	-0.075	-0.013
tm.mean.2	0.193	0.026	0.141	0.245
hM.mean.2	-0.012	0.005	-0.022	-0.001
hm.mean.2	-0.032	0.005	-0.042	-0.023
pr.mean.2	-0.032	0.007	-0.045	-0.020
Popdensity	0.001	0.001	-0.001	0.003
GMM.T.2018	0.000	0.005	-0.010	0.009
IDI.2019	-0.174	0.098	-0.369	0.022
Mais.65	0.000	0.000	0.000	0.000
τ				
τ (Region_str (iid component))	0.38	0.139	0.167	0.707
τ Region_str (spatial component)	2629.49	3010.00	229.093	10500.00
τ Date_str	6018.86	5170.00	1069.356	19700.00
τ Date_unstr	1.02	0.170	0.720	1.390
τ region_date	26925.46	30500.00	2322.168	107000.00

O risco de doença em pequenas áreas caracterizado pela taxa de incidência padronizada – SIR (4.3), em pequenas áreas podem apresentar valores extremos devido a pequenos tamanhos populacionais ou a amostras pequenas. Nessas situações, os SIR podem ser enganosos e insuficientemente confiáveis para relatar, pelo que estimar o risco de doença usando Modelos Hierárquicos Bayesianos, modelos que permitem “emprestar” informações de áreas vizinhas e incorporar informações de covariáveis, resulta na suavização ou encolhimento dos valores extremos (Moraga, 2018).

Tabela 5.11: Média posterior e intervalos de credibilidade (a 95%) dos efeitos fixos e dos hiperparâmetros do modelo *bym.rw.iid.M.2*, ajustado com o número diário de casos COVID-19 desagregados dos dados agregados à semana com recurso ao algoritmo LS, nos concelhos do distrito do Porto, entre 25/03/2020 e 30/06/20 e com as variáveis meteorológicas com atraso de dois dias.

Parâmetros	Média	sd	0.025Q	0.975Q
Modelo:				
<i>iCAR.rw.iid.M2.2.LS</i> : $x.LS \sim tM.mean.2 + tm.mean.2 + hM.mean.2 + hm.mean.2 + pr.mean.2 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 +$				
$f(Region_str, model="bym", graph=W.porto) +$				
$f(Date_str, model="rw2") +$				
$f(Date_unstr, model="iid") +$				
$f(region_date, model="iid")$				
(Intercept)	6.201	4.622	-2.968	15.374
tM.mean.2	-0.052	0.016	-0.084	-0.019
tm.mean.2	0.225	0.027	0.172	0.278
hM.mean.2	-0.013	0.006	-0.024	-0.003
hm.mean.2	-0.033	0.005	-0.042	-0.023
pr.mean.2	-0.037	0.006	-0.050	-0.025
Popdensity	0.001	0.001	-0.001	0.003
GMM.T.2018	0.000	0.005	-0.010	0.009
IDI.2019	-0.179	0.099	-0.376	0.018
Mais.65	0.000	0.000	0.000	0.000
τ				
$\tau(Region_str$ (iid component))	0.378	0.14	0.168	0.71
τ_{Region_str} (spatial component)	2.270	2550	185.345	9034.23
τ_{Date_str}	6.310	5440	1077.822	20745.91
τ_{Date_unstr}	94	0.155	0.671	1.28
τ_{region_date}	22300	24200	1637.708	86402.60

Na secção 4.2, do capítulo 4, foram representados mapas de risco com os dados diários acumulados observados e os dados diários acumulados desagregados. Nos Modelos Hierárquicos Bayesianos ajustados, no presente capítulo, com os dados dos concelhos do distrito do Porto de 25 de março a 30 de junho estimou-se o risco tendo em conta efeitos fixos (covariáveis) e efeitos aleatórios (espacial e temporal). Para ser possível estabelecer comparação entre o SIR e o risco estimado pelos modelos, foram ajustados modelos *iCAR.wr.iid.M2.2*. com o número diário de casos acumulados. Os mapas das figuras seguintes representam o risco estimado pelos modelos para os dias 30 e 31 de março e 1 e 2 de abril, e para os dias 15,16,17 e 18 e maio de 2020. No capítulo 4, secção 4.2, encontram-se os mapas de risco (SIR) obtidos com o número diário de casos acumulados observados e desagregados e o número diário de casos acumulados esperados (Figura 4.9, Figura 4.10, Figura 4.11, Figura 4.12). Na Figura 5.16 e na Figura 5.17 estão os mapas de risco estimados pelo modelo *iCAR.wr.iid.M2.2*. Comparando-os com os mapas dos dados observados, da Figura 4.9 e da Figura 4.11, pode-se concluir que os modelos ajustados com os dados

desagregados pelo algoritmo HF refletem de forma mais precisa a distribuição observada da incidência da doença.

5.4.2.2. Modelos ajustados com os dados dos concelhos de Portugal Continental, de 25 de março de 2020 a 23 de março de 2021

Os dados de risco (SIR) dos concelhos de Portugal Continental foram calculados de acordo com a equação (4.3), do capítulo 4, secção 4.2, com os dados reais e com os dados desagregados pela distribuição uniforme (secção 5.3.1) e pelos os algoritmos LS e HF (secção 5.3.2). Os mapas da Figura 5.18, da Figura 5.19, da Figura 5.20 e da Figura 5.21 ilustram a distribuição do risco de COVID-19 nos dias 10 de agosto e 10 de novembro de 2020, 10 de janeiro e 10 de fevereiro de 2021. A distribuição dos dados desagregados pelos algoritmos LS e Hf são muito semelhantes. A distribuição uniforme revela alguns concelhos com riscos mais extremos que nos outros dois algoritmos. Os resultados da secção 4.2 permitem concluir que os dados desagregados pelo algoritmo HF são mais precisos.

Na Figura 5.22 Figura 5.22: Mapa da distribuição do risco para (A) $SIR > 1$ (n° de casos observados $> n^\circ$ de casos esperados) e para $SIR \leq 1$ (n° de casos observados $\leq n^\circ$ de casos esperados) e (B) risco estimado pelo modelo, em 10/08/2020. e na Figura 5.23 estão representados os mapas de distribuição espacial do risco, para o caso em que este é menor ou igual a 1 e portanto o número de casos acumulados observado é menor que o número de casos acumulados esperado e para o caso em que o risco é maior que 1 e portanto o número de casos acumulados observado é maior que o número de casos acumulados esperado. Os concelhos em que se regista esta situação devem ser objeto de especial atenção pelas autoridades de saúde.

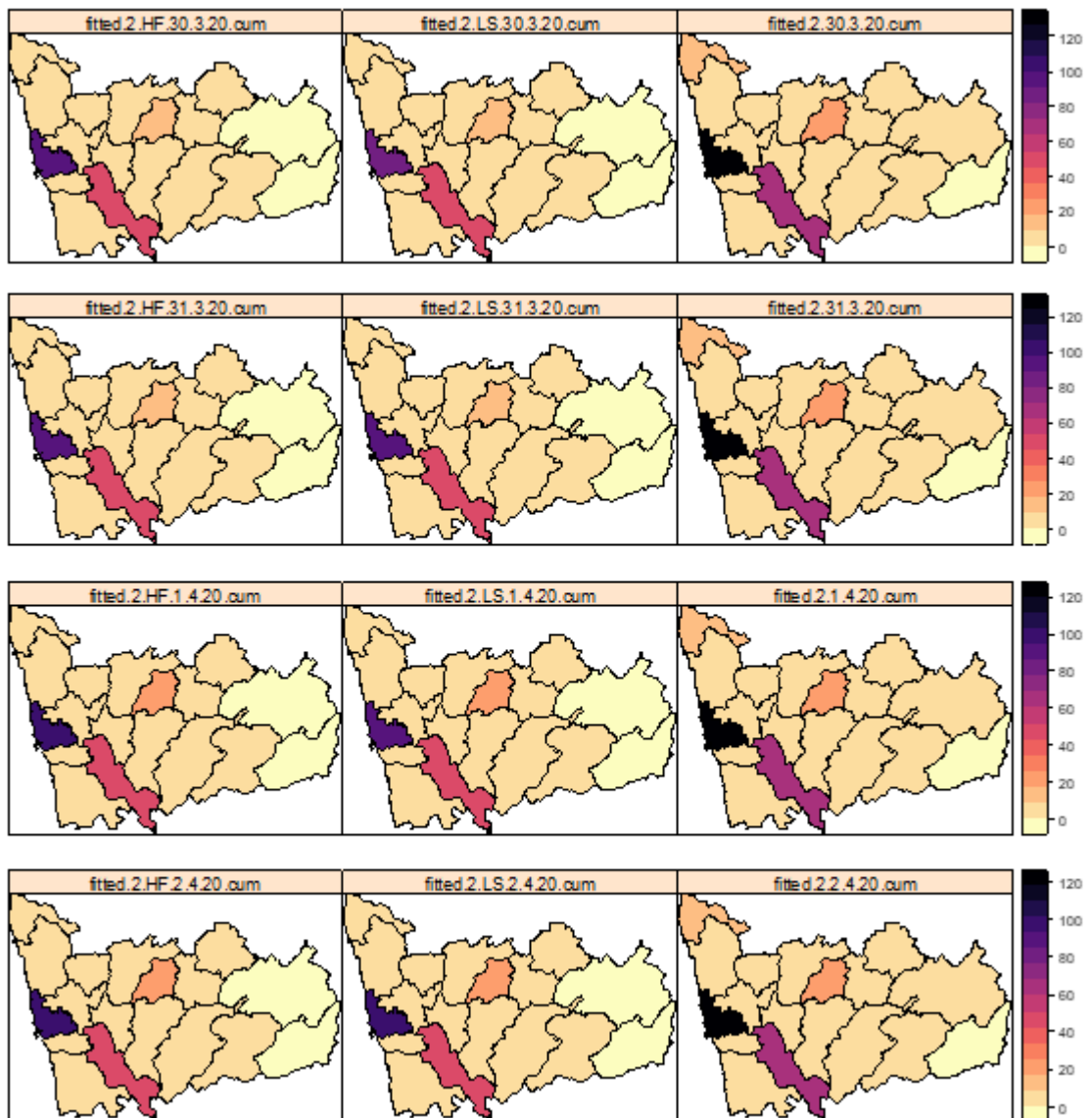


Figura 5.16: Mapas de risco nos concelhos do distrito do Porto estimados pelos modelos iCAR.wr.iid ajustados com os dados observados, e os dados desagregados com os algoritmos LS e HF de dados agregados semanalmente, para os dias 30 e 31 de março, 1 e 2 de abril de 2020, com *lag* 2 para as covariáveis meteorológicas.

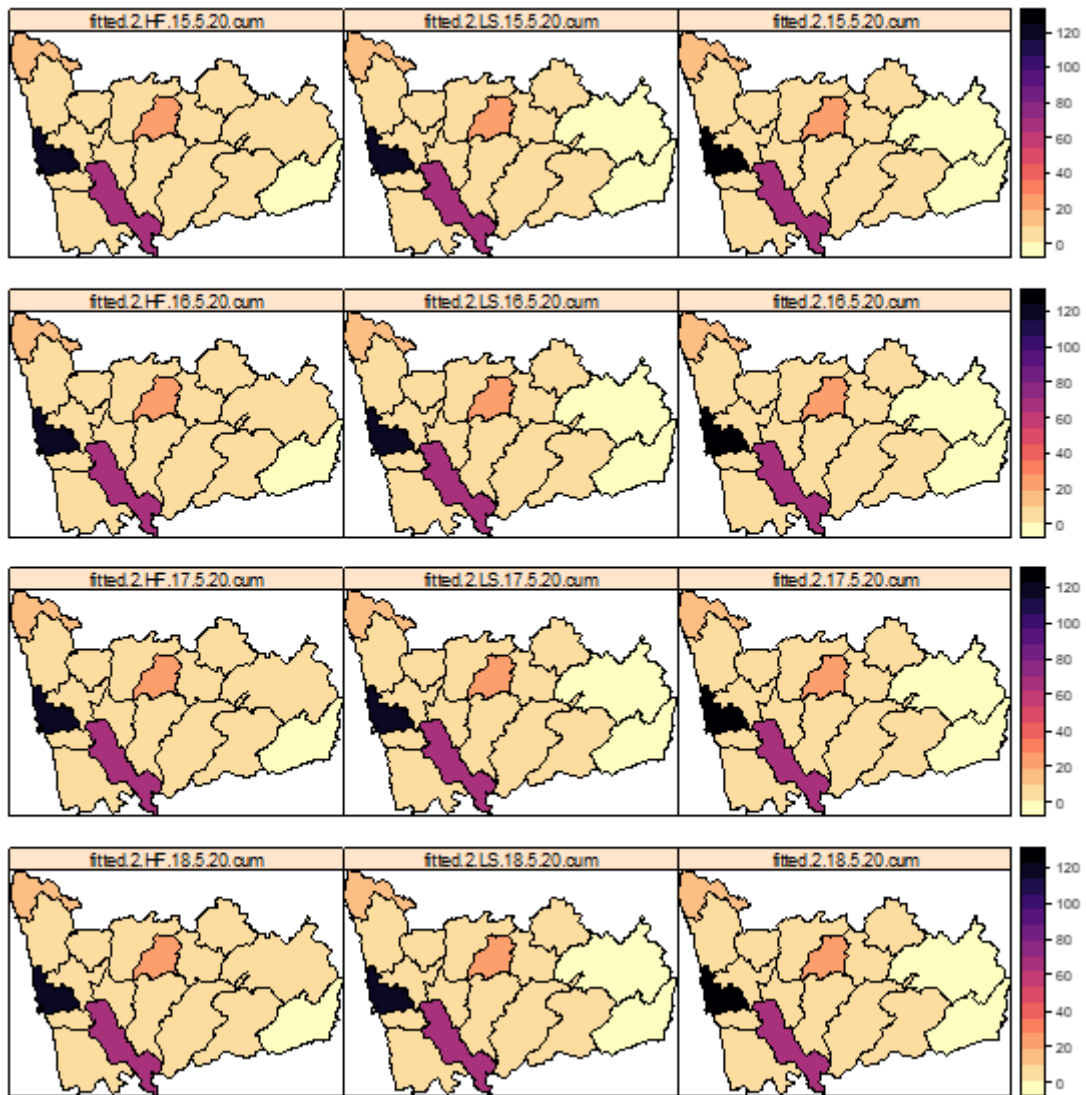


Figura 5.17: Mapas de risco nos concelhos do distrito do Porto estimados pelos modelos iCAR.wr.iid ajustados com os dados observados, e com os dados desagregados de dados agregados semanalmente, com os algoritmos LS e HF, para os dias 15,16, 17 e 18 de maio de 2020, com *lag* 2 para as covariáveis meteorológicas.

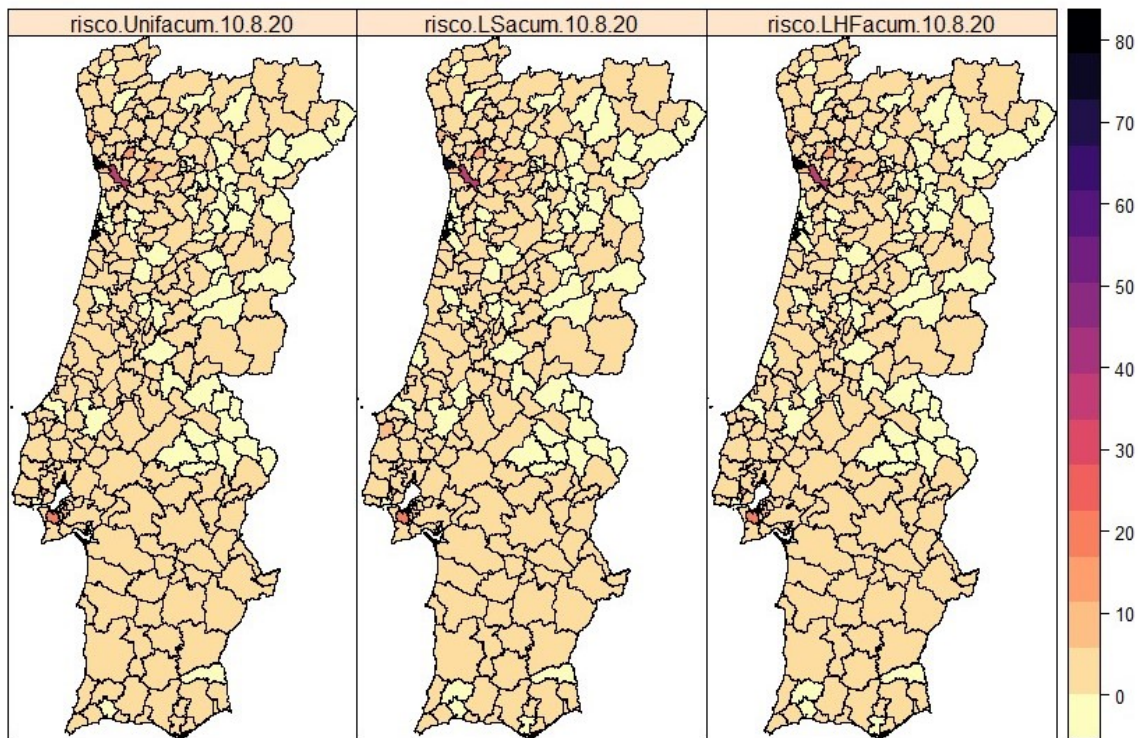


Figura 5.18: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de agosto de 2020.

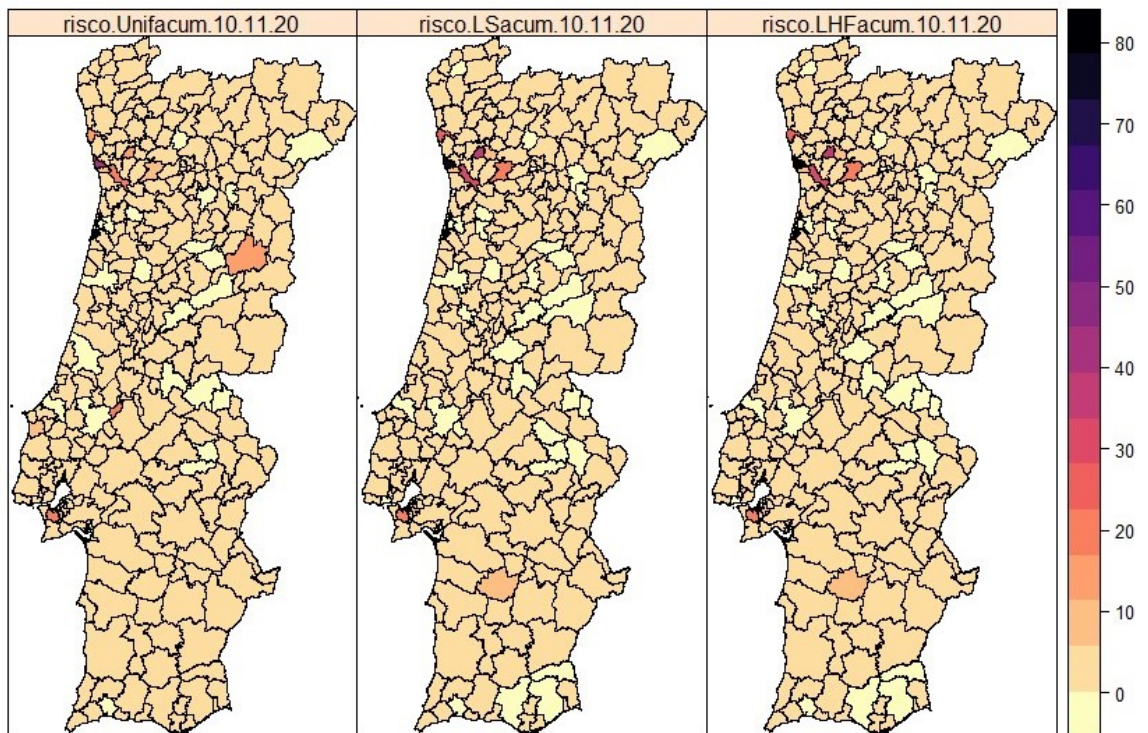


Figura 5.19: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de novembro de 2020.

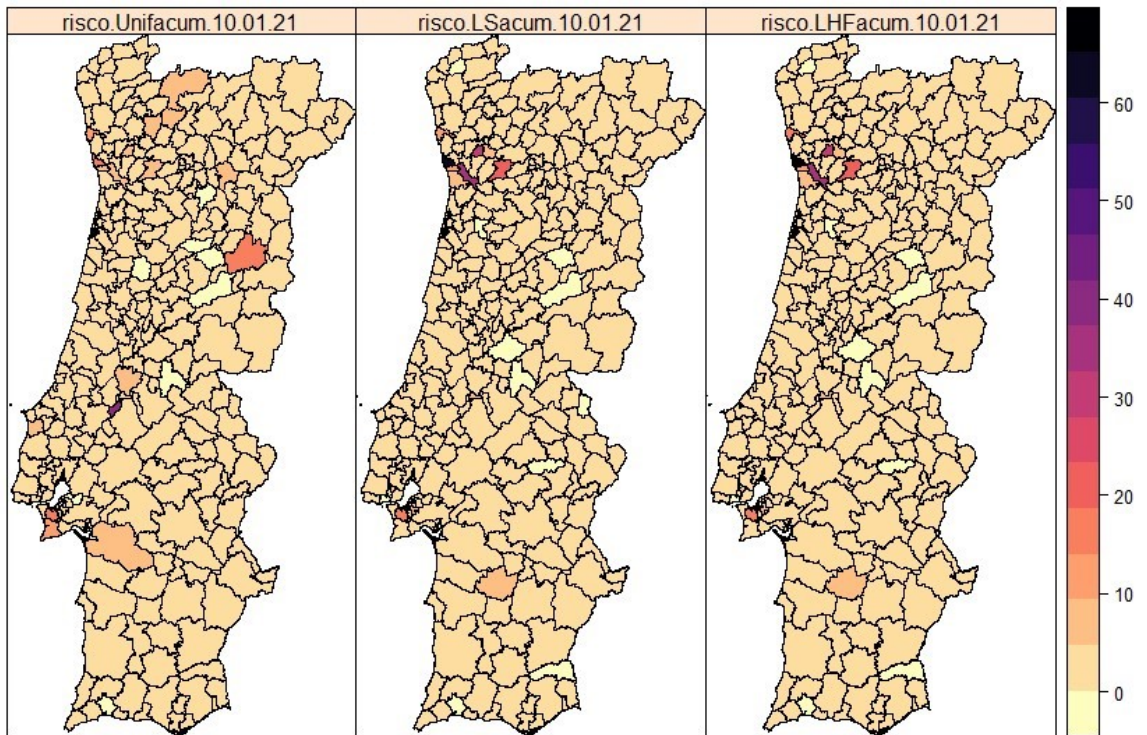


Figura 5.20: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de janeiro de 2021

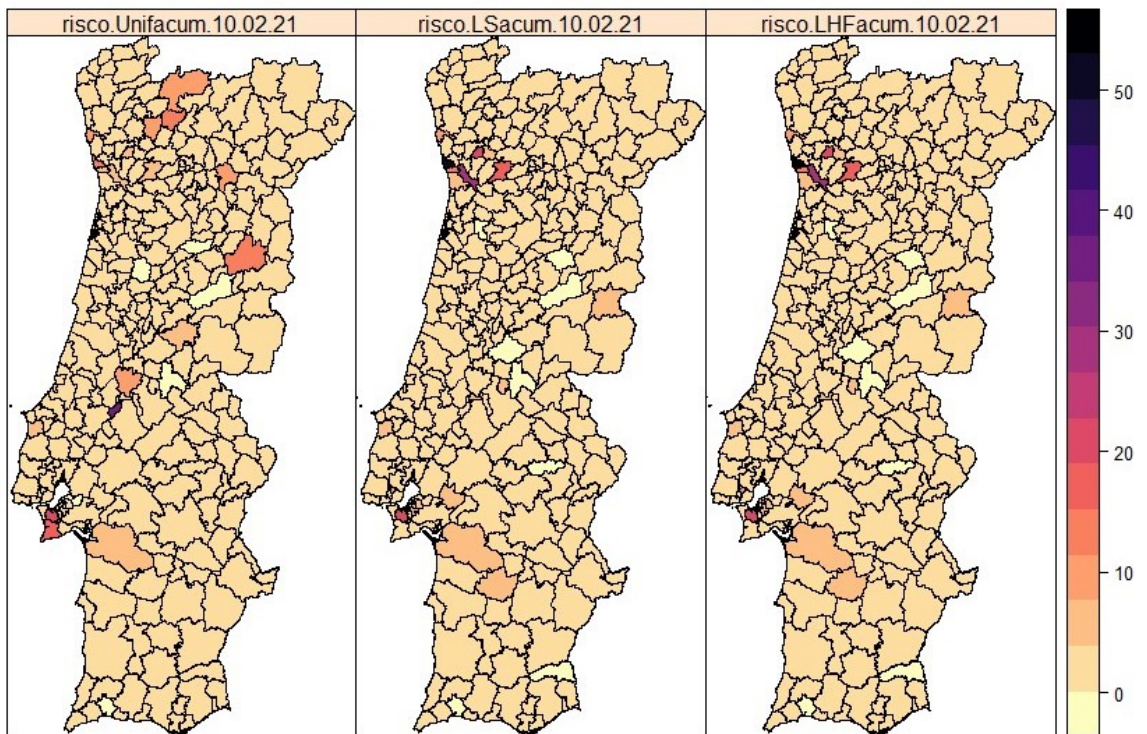


Figura 5.21: Distribuição do risco (SIR) de COVID-19 nos concelhos de Portugal Continental com base nos dados desagregados (desagregação uniforme, algoritmo LS e algoritmo HF) – dia 10 de fevereiro de 2021.

Com os dados de todos os concelhos de Portugal Continental, ajustou-se um Modelo Hierárquico Bayesiano de Poisson com as covariáveis *Temperaturas médias – máxima e mínima, Humidades médias – máxima e mínima e Pluviosidade média*, com atraso de 2 dias. Incluíram-se efeitos espaciais estruturados e não estruturados (modelo BYM) e efeitos temporais estruturados (passeio aleatório de segunda ordem) e não estruturados (modelo *iid*) e um efeito de interação **Espaço × Tempo**. O modelo foi ajustado com o número diário de casos observados, de 25 de março a 4 de julho de 2020 e com os dados desagregados para frequência diária pelo algoritmo H-Fuse, de 5 de julho de 2020 a 23 de março de 2021. Usou-se este modelo e estes dados por serem os que revelaram melhor ajustamento no estudo feito no capítulo 4, secção 4.2.

As médias posteriores das covariáveis e os respetivos intervalos de credibilidade a 95% encontram-se na Tabela 5.12.

Tabela 5.12: Média posterior e intervalos de credibilidade (a 95%) dos efeitos fixos e dos hiperparâmetros do modelo *bym.rw.iid.M.2*, ajustado com o número diário de casos COVID-19 dos concelhos de Portugal Continental, observados no período de 25/03/2020 a 4/07/2020 e desagregados pelo algoritmo H-Fuse no período de 05/07/2020 a 23/03/2021.

```

Modelo:
F2.2.HF : x.HF~tm.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popd
ensity+ GMM.T.2018 + IDI.2019+Mais.65+
f(Region_str, model="bym", graph=Regio.W)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date,model="iid")
iCAR.rw.iid.M2.2.HF<- inla(F2.2.HF,
family="poisson",
control.compute=list(dic=TRUE,cpo=TRUE,waic=TRUE,r
eturn.marginals.predictor=TRUE,config=TRUE),control.predictor = li
st(compute = TRUE),
data=BD_cap5.1,E=Exp.dia)

```

Parâmetros	Média	sd	0.025Q	0.975Q
(Intercept)	2.623	0.467	1.706	3.539
tM.mean.2	-0.029	0.001	-0.031	-0.028
tm.mean.2	-0.019	0.001	-0.021	-0.017
hM.mean.2	-0.005	0.000	-0.005	-0.004
hm.mean.2	0.001	0.000	0.001	0.002
pr.mean.2	-0.014	0.000	-0.015	-0.014
Popdensity	0.000	0.000	0.000	0.000
GMM.T.2018	-0.001	0.000	-0.002	0.000
IDI.2019	-0.029	0.005	-0.039	-0.019
Mais.65	0.000	0.000	0.000	0.000
τ				
τ (Region_str (iid component))	1.15	0.116	0.923	1.38
τ Region_str (spatial component)	1874.48	22900	4.52	10400
τ Date_str	59455.81	28700	18300	128000
τ Date_unstr	21793.69	23900	1600	85300
τ region_date	3.72	0.302	3.15	4.34

As covariáveis meteorológicas *Temperaturas médias – máxima e mínima*, *Humidade máxima média* e a *Pluviosidade média* têm efeito negativo e a *Humidade média mínima* tem efeito positivo no risco de COVID-19. No entanto, o efeito da *Humidade média – mínima e máxima* não é estatisticamente significativo, pois os extremos dos respetivos intervalos de credibilidade a 95% são muito próximos de zero se não zero. As restantes variáveis meteorológicas têm um efeito estatisticamente significativo. Os efeitos espaciais e temporais são todos positivos e estatisticamente significativos, o que sustenta a heterogeneidade espacial e temporal da COVID-19.

Nos gráficos da Figura 5.22, Figura 5.23, Figura 5.24 e Figura 5.25 de casos esperados (equação (4.3) e o risco estimado pelo modelo $iCAR.wr.iid.M2.2.HF$ ajustado com o número diário de casos acumulados, para os dias 10/08/2020, 10/11/2020, 10/01/2021 e 10/02/2021.

Os concelhos em que o risco é inferior ou igual a 1, isto é, o número de casos observados é inferior ou igual ao número de casos esperados, estão representados numa escala de cor azul, enquanto os concelhos com risco estimado superior a 1 estão representados em escala laranja.

Comparando os dois mapas para cada dia, verifica-se que, para os dias 10/8/2020 e 10/11/2020, os modelos aumentam um pouco os riscos inferiores ou iguais a um e suavizam os riscos maiores que 1. No entanto, na generalidade, os mapas que representam o risco estimado pelos modelos refletem bem a distribuição do risco estimado pelo SIR. Para os dias 10/01/2021 e 10/02/2021 os modelos suavizam os riscos menores ou iguais a 1.

Na generalidade, os mapas do risco estimado pelo modelo nestes dois dias refletem bem a distribuição espacial do risco estimado pelo SIR para a maioria dos concelhos. Em geral, em nenhuma das situações há evidências claras de desvios da distribuição dos riscos de COVID-19 estimados pelo modelo em relação ao risco estimado pelo SIR.

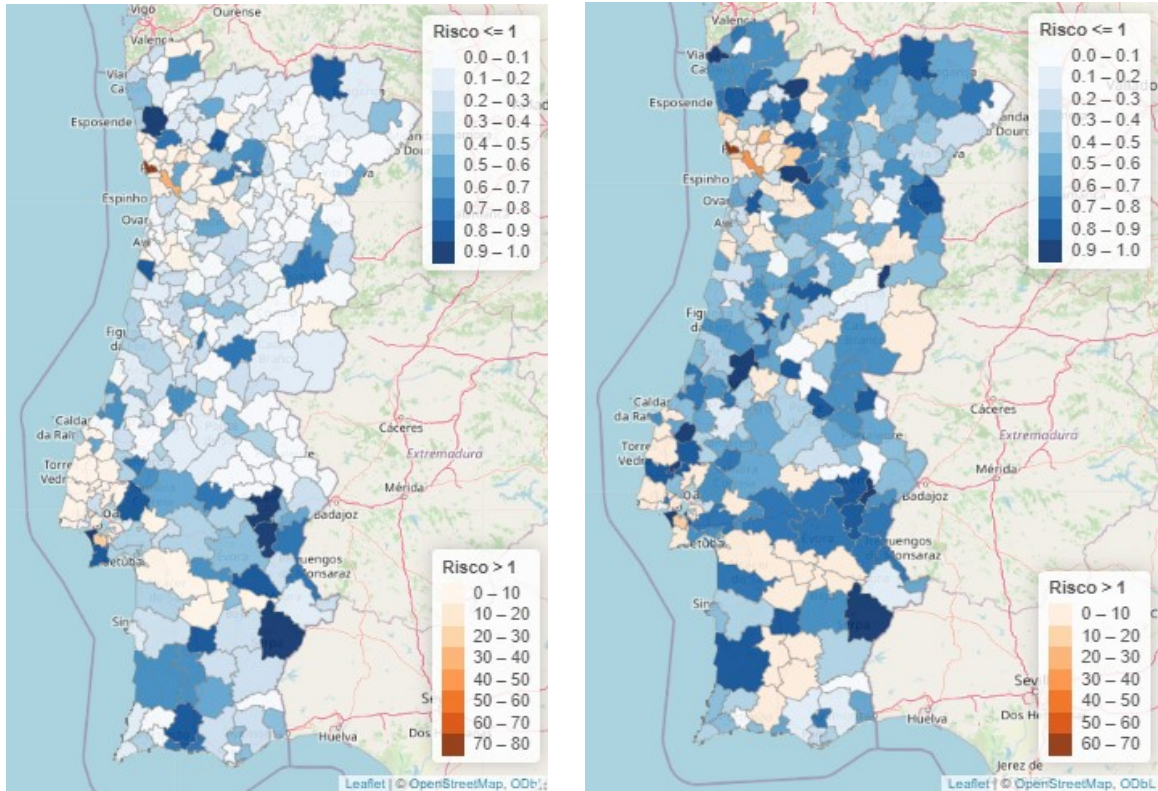


Figura 5.22: Mapa da distribuição do risco para (A) $SIR > 1$ (n° de casos observados $> n^\circ$ de casos esperados) e para $SIR \leq 1$ (n° de casos observados $\leq n^\circ$ de casos esperados) e (B) risco estimado pelo modelo, em 10/08/2020.

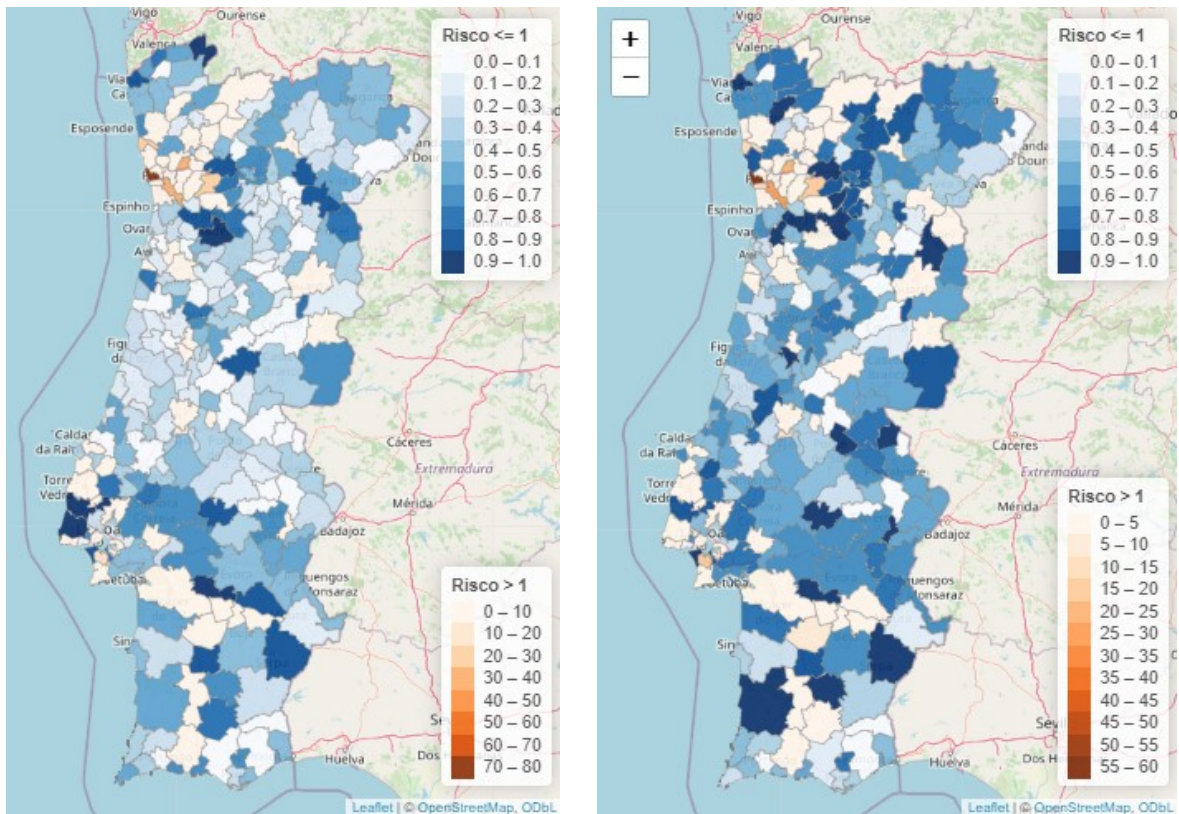


Figura 5.23: Mapa da distribuição do risco para (A) $SIR > 1$ (n° observado de casos $> n^\circ$ esperado de casos) e para $SIR \leq 1$ (n° de casos observados $\leq n^\circ$ de casos esperados) e (B) risco estimado pelo modelo, em 10/11/2020.

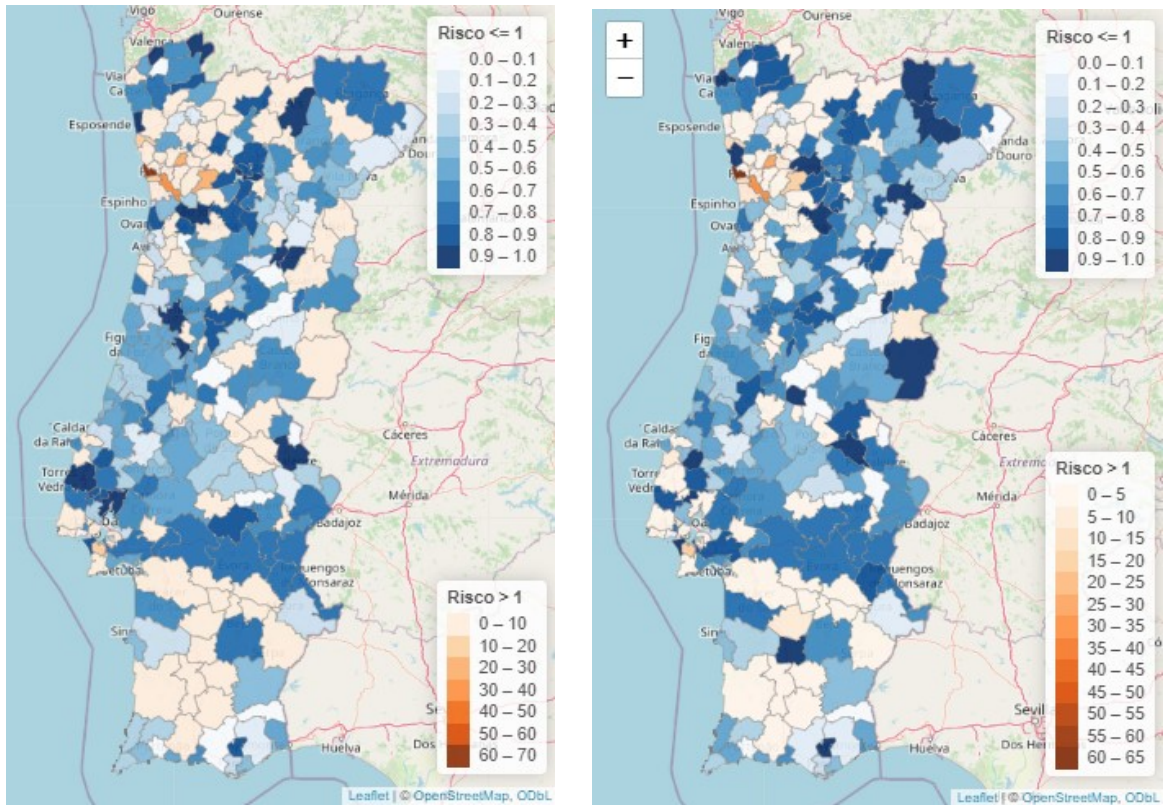


Figura 5.24: Mapa da distribuição do risco para (A) $SIR > 1$ (n° de casos observados $> n^\circ$ de casos esperados) e para $SIR \leq 1$ (n° de casos observados $\leq n^\circ$ de casos esperados) e (B) risco estimado pelo modelo, em 10/01/2021.

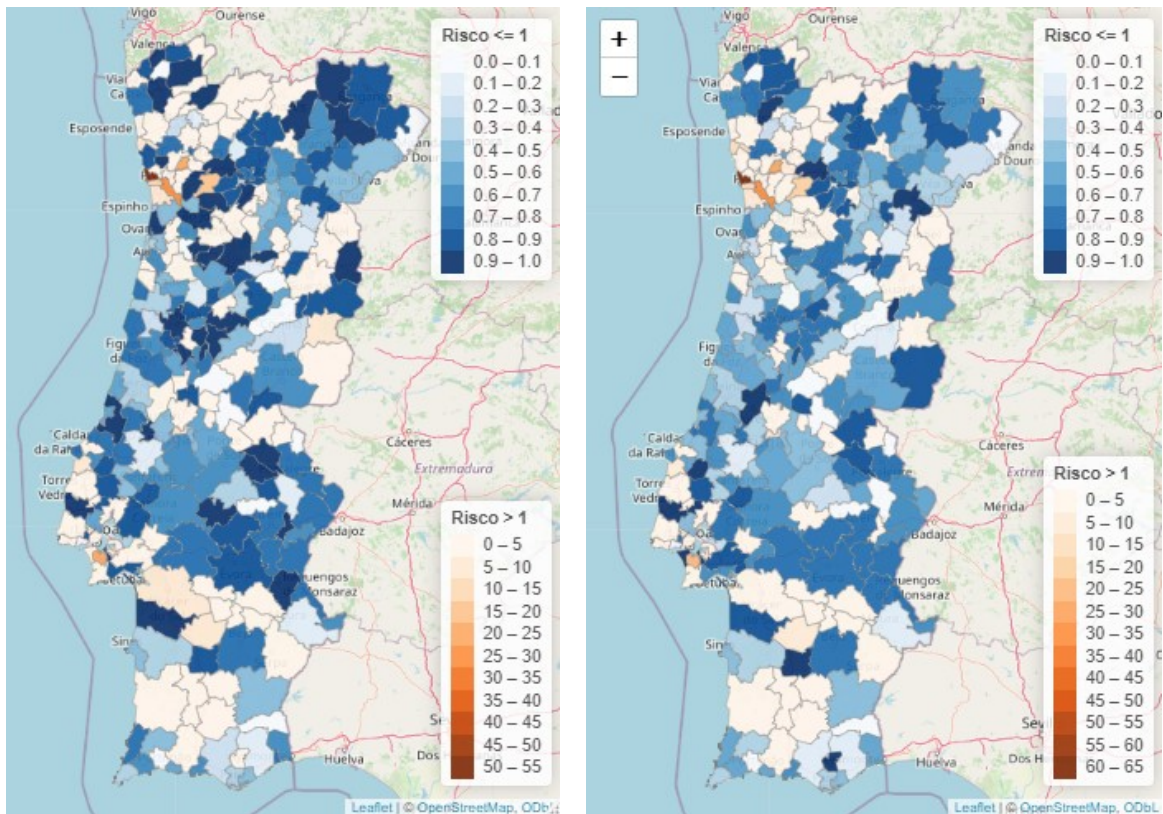


Figura 5.25: Mapa da distribuição do risco para (A) $SIR > 1$ (n° de casos observado $> n^\circ$ de casos esperado) e para $SIR \leq 1$ (n° de casos observado $\leq n^\circ$ de casos esperado) e (B) risco estimado pelo modelo, em 10/02/2021.

CAPÍTULO 6

CONCLUSÕES E CONSIDERAÇÕES FINAIS

6. CONCLUSÕES E CONSIDERAÇÕES FINAIS

Monitorizar os dados de uma pandemia é crucial para compreender a evolução da doença, implementar medidas de saúde pública eficazes e tomar decisões informadas. No contexto de uma pandemia, como a COVID-19, a aquisição de dados epidemiológicos é especialmente desafiadora por diversas razões como explanado no capítulo 2, secção 2.10 e capítulo 4.14.1: a qualidade e a confiabilidade dos dados dependem da capacidade de testagem, da notificação, da vigilância, do registo dos casos, das mortes e dos recuperados; a padronização e a integração dos dados envolvem a definição de critérios, de indicadores, de protocolos e de sistemas de informação comuns entre os diferentes níveis e setores de saúde, bem como a interoperabilidade e a segurança dos dados¹⁶; e análise e a interpretação dos dados requerem o uso de métodos estatísticos e epidemiológicos adequados, a consideração dos fatores contextuais e confundidores, e a comunicação clara e transparente dos resultados e das incertezas.

Em Portugal, durante a pandemia COVID-19, foram reconhecidos obstáculos à monitorização de dados, desde a sua recolha à publicação, verificando-se lacunas na disponibilização de dados essenciais e de dados desagregados, atrasos e atualizações irregulares, bem como falta de consistência nos relatórios, incluindo mudanças nas definições de casos, métodos de contagem ou outras variáveis, o que torna desafiador analisar e comparar dados ao longo do tempo.

Identificado no capítulo 1 o problema da dificuldade de monitorização dos dados de COVID-19 ou de outros dados epidemiológicos de natureza semelhante, foram investigadas nos capítulos 3, 4 e 5 metodologias de tratamento e monitorização de dados capazes de resolver alguns dos problemas identificados. Assim, definiu-se um conjunto de objetivos específicos que orientaram esta investigação: adequar e aplicar algoritmos de desagregação temporal de dados, para desagregar os dados diários de COVID-19 dos concelhos de Portugal Continental; estudar o impacto da desagregação dos dados sobre a precisão dos resultados; e definir uma metodologia capaz de detetar o impacto de eventos públicos na propagação da epidemia/pandemia e definir modelos para estimar o risco da COVID-19

¹⁶ [O poder dos dados na resposta a crises - dados.gov.pt - Portal de dados abertos da Administração Pública](https://dados.gov.pt)

capazes de acomodar a heterogeneidade espacial e temporal e o efeito de covariáveis no risco, em particular, covariáveis de natureza meteorológica.

Através da exploração de dois casos relevantes foram criadas formas mais sólidas de análise e monitorização dos dados.

No capítulo 3 foi apresentada uma metodologia de modelação, análise e testagem, assente em modelos lineares generalizados, com regressão de Poisson, visualização e testes de hipóteses capaz de identificar a mudança no padrão de crescimento dos casos de COVID-19 na área geográfica de realização de um evento público. Esta metodologia foi aplicada a dois eventos que se realizaram em Lisboa durante a pandemia: a Comemoração do 1.º de Maio e a Manifestação Antirracismo, ambos ocorridos em 2020.

No capítulo 4 foram aplicados algoritmos de desagregação temporal de séries temporais aos dados de COVID-19 dos concelhos do distrito do Porto e foram avaliados os resultados obtidos. No capítulo 5 estes dados e os dados observados foram modelados com recurso a Modelos Hierárquicos Bayesianos e compararam-se os resultados. Com base nestes resultados foi estimada a distribuição do risco de COVID-19 nos concelhos de Portugal Continental e o efeito da temperatura, humidade e pluviosidade. O efeito destas covariáveis foi estudado com desfasamento de 2, 7 e 14 dias, uma vez que o período de incubação do SARS-CoV-2 varia de 2 a 14 dias (Lauer et al., 2020). O modelo com melhor ajustamento foi aquele em que se usou o desfasamento de 2 dias.

A aplicação da metodologia inovadora apresentada no capítulo 3 a dois casos práticos revelou que a mesma permite identificar se a realização de um evento público tem impacto ou não no padrão de propagação de uma epidemia/pandemia na região onde este ocorreu e, aí, quais foram as áreas geográficas em que este impacto foi estatisticamente significativo. Permite ainda considerar covariáveis que podem explicar a heterogeneidade geográfica desse impacto, de forma a contribuir para a tomada de decisão por parte das autoridades competentes. De notar que não foram incluídas covariáveis nos modelos e que, portanto, a metodologia pode tornar-se mais sólida se, para além da forma que foi adotada para avaliar o efeito de covariáveis na alteração do padrão de crescimento da doença, se incluírem covariáveis de interesse nos modelos, para avaliar estatisticamente o seu efeito na variável resposta. Além disto, é de referir que os erros são autocorrelacionados e isso não foi considerado nos modelos. Para lidar com esta limitação pode ser considerada a inclusão de um termo autorregressivo (AR).

Nos capítulos 4 e 5 foram usados dois algoritmos para desagregar dados de séries temporais, com recurso ao R – o Método dos Mínimos Quadrados (LS) com Pseudo-inversa de Penrose, para resolver um problema de norma mínima dos desvios dos valores observados aos valores desagregados a estimar, e o método H-Fuse (HF – secção 2.10.7(ii)), para resolver um problema de penalização mínima, em que a função de penalização tem duas componentes, uma com os desvios e uma que infunde alguma informação do domínio. O estudo mostrou que os dados desagregados pelo método H-Fuse são os mais precisos. Pode-se considerar que a aplicação do método LS tem algumas fragilidades, uma vez que se usa o método dos Mínimos Quadrados para resolver o sistema de equações lineares minimizando os erros, mas há pressupostos deste método que podem ser críticos, nomeadamente, a independência dos erros. No entanto, a maioria das variáveis apresentou estatísticas do teste de Durbin-Watson relativamente próximas de 2 (Anexo IX), sugerindo que os resíduos são aproximadamente independentes, por a correlação ser relativamente pequena. Por outro lado, a Pseudo-inversa de Penrose lida com o problema de o sistema ser sobredeterminado. No capítulo 5 foram ajustados diferentes Modelos Hierárquicos Bayesianos aos dados observados e aos dados desagregados pelos métodos LS e HF dos concelhos do distrito do Porto, do período de 25/03/2020 a 30/06/2020, com efeitos temporais, espaciais, espaciais e temporais, com e sem covariáveis. Os resultados foram comparados e concluiu-se que o modelo com melhor ajuste foi aquele em que se consideraram os efeitos temporais estruturados e não estruturados, os efeitos espaciais estruturados e não estruturados e o efeito de interação *espaço × tempo*. Foram ainda consideradas as covariáveis meteorológicas com valores médios e *lag 2*. Este mesmo modelo foi usado para estimar a distribuição do risco de COVID-19 em todos os concelhos de Portugal Continental com os dados desagregados pelo método H-Fuse que, no estudo do capítulo 4, revelou resultados mais precisos, relativamente aos dados observados. Os mapas de distribuição de risco de COVID-19 para os concelhos de Portugal Continental, estimados pelo modelo revelaram resultados muito ajustados aos dos mapas de distribuição de risco SIR, conforme equação (4.3), que traduz a taxa de comparação entre o número de casos observados e o número de casos esperados. Considerando ainda o efeito das covariáveis, concluiu-se que, das covariáveis de natureza socioeconómica e demográfica, apenas a covariável *Índice de Dependência de Idosos* revelou um efeito negativo, estatisticamente significativo, o que pode ser resultado da eficácia das medidas de proteção adicionais aos idosos mais dependentes, nomeadamente

os residentes nos lares, decretados pelas autoridades de saúde. Esta conclusão é corroborada por Costa e Marques da Costa (2020). Das covariáveis meteorológicas, apenas a *humidade mínima média* não revelou efeito estatisticamente significativo. As restantes revelaram um efeito negativo estatisticamente significativo.

Na literatura surgem diferentes mecanismos que podem explicar a influência da temperatura, humidade e pluviosidade na disseminação do SARS-Cov-2: os que se relacionam diretamente com estabilidade e transmissão do vírus; os que se relacionam com a defesa antiviral humana; e os que se relacionam com a influência destes fatores no comportamento humano, que por sua vez, influenciam a disseminação do vírus. A maioria dos estudos aponta para um efeito negativo destes fatores na disseminação do SARS-Cov-2 (Z. Huang et al., 2020; Lowen et al., 2008; Notari, 2021, 2021; Tamerius et al., 2013). Por exemplo, Wu et al. (2020), num estudo incidindo em 166 países, concluíram que estas duas variáveis revelaram um efeito negativo na propagação do vírus. O mesmo concluiu Paez et al. (2021) sobre a propagação da COVID-19 em Espanha. No entanto, também se observam casos que apontam para a correlação positiva (Lowen et al., 2008). Por exemplo, Adhikari & Yin (2020) concluíram que a temperatura e humidade têm efeito positivo, estatisticamente significativo, na disseminação da COVID-19.

Os resultados obtidos nos estudos sobre o efeito das covariáveis meteorológicas estudadas no capítulo 5 desta tese corroboram o que se sabe da literatura, fortalecendo a confiança de que o viés introduzido pelos métodos de desagregação temporal utilizados não compromete significativamente as conclusões, podendo ser utilizados em dados de COVID-19 e em dados de outras epidemias/pandemias. Salienta-se, no entanto, que os estudos têm algumas limitações e que os resultados devem ser analisados tendo em conta estas limitações. Apesar de terem sido consideradas algumas covariáveis para ajustar os modelos, para além das covariáveis meteorológicas que podem ser encaradas como confundidores, há uma grande variedade de outras covariáveis que podem ser consideradas e que podem ter impacto significativo na dinâmica de transmissão da doença, nomeadamente o uso de máscara, as medidas de controlo impostas pelas autoridades de saúde, a restrição à mobilidade, a qualidade do ar, a heterogeneidade da testagem, a vacinação e que podem ser incluídas nos modelos tornando os resultados mais sólidos.

As metodologias propostas para a tratamento e análise de dados espaço-temporais epidemiológicos, e em particular, de dados de COVID-19, são um contributo inovador quer

para a área de investigação desta tese quer para a epidemiologia. As metodologias de desagregação temporal de dados e a metodologia proposta para a deteção de mudanças no padrão de crescimento de uma epidemia/pandemia são os contributos mais inovadores desta tese.

De facto, a desagregação temporal de dados epidemiológicos, com recurso a algoritmos que permitem infundir informação relevante do domínio é bastante promissor e contribui para a solução do problema sentido pelos investigadores da falta de dados com características adequadas à pesquisa de soluções na área da saúde ou outra, nomeadamente, a possibilidade de combinar dados de diferentes fontes que os recolhem e tratam com critérios diferentes. Acresce a possibilidade de incrementar estes algoritmos de forma a poderem ser usados na desagregação espacial e contribuir desta forma para lidar o problema MPUP (*Modifiable Temporal Unit Problem*) (Cheng & Adepeju, 2014) e com o problema MAUP (*Modifiable Areal Unit Problem*) (Openshaw, 1984). A implementação destes algoritmos de desagregação no *software* R constitui um contributo para o enriquecimento de ferramentas disponíveis numa plataforma de acesso livre, flexível e robusta, com uma vasta gama de pacotes e bibliotecas especializadas em séries temporais, o que simplifica a implementação e experimentação com diferentes métodos de desagregação. A integração destes algoritmos de desagregação no R contribui, assim, para uma abordagem mais sofisticada e eficiente no estudo de padrões temporais complexos, fortalecendo a capacidade da comunidade científica de realizar análises temporais mais precisas e avançadas.

A metodologia proposta no capítulo 3 é uma metodologia inovadora e relevante que permite analisar o impacto potencial, geograficamente heterogéneo, de um evento na disseminação do contágio numa epidemia/pandemia, seguindo um conjunto de etapas sistemáticas. A aplicação desta metodologia permite extrair conclusões bem fundamentadas e contribuir para a discussão pública informada e para a tomada de decisões, num curto período de tempo após a ocorrência do evento, como mostram a sua aplicação a eventos como a Manifestação do 1.º de maio de 2020 e da Manifestação Antirracismo, em Lisboa, no capítulo 3, seções 3.3 e 3.4 respetivamente.

Esta metodologia tem algumas fragilidades, nomeadamente a forma como se avaliou o efeito de uma covariável na mudança do padrão de crescimento da COVID-19 num concelho. Incluir as covariáveis no modelo e avaliar estatisticamente o efeito sobre o número cumulativo de casos de COVID-19 dará mais solidez às conclusões. Ajustar modelos

separados para cada concelho é uma fragilidade, uma vez que não se avalia explicitamente a variabilidade entre concelhos. Cada modelo tem os seus próprios coeficientes, perdendo-se a possibilidade de analisar e modelar a variação sistemática ou padrões comuns entre os concelhos. Para abordar esse problema e melhorar a metodologia pode-se considerar o uso de modelos de Poisson com efeitos fixos ou aleatórios, incluindo no primeiro caso uma variável *dummy* no preditor linear a identificar cada concelho na função `glm()` do R e no segundo caso incluir como efeito aleatório a variável *dummy* a identificar os concelhos na função `glmer()` do pacote `lme4` do R. No modelo de efeitos fixos cada concelho terá o seu próprio intercepto, capturando assim as diferenças nas taxas médias do número de casos entre os concelhos. Este modelo assume que a relação entre as variáveis independentes e dependente é constante entre os concelhos, mas permite que os interceptos variem. No modelo de efeitos aleatórios é tida em consideração a variabilidade entre os concelhos, permitindo que os efeitos aleatórios capturem diferenças individuais entre eles, ou seja, a variabilidade espacial é acomodada no modelo com um efeito espacial. Estes modelos são especialmente úteis quando se quer ter em consideração a variabilidade não explicada entre os concelhos.

A análise espaço-temporal de dados de COVID-19 proposta nesta tese, seja com a metodologia proposta no capítulo 3 ou na implementada no capítulo 5, secção 5.4, é inovadora e relevante, pois as duas metodologias permitem obter informações mais detalhadas e confiáveis sobre a evolução da COVID-19 em Portugal, contribuindo para o aprimoramento das políticas públicas de saúde e para o avanço do conhecimento científico na área da epidemiologia e da bioestatística. Além disso, as metodologias são flexíveis e adaptáveis, podendo ser aplicadas a outros contextos e situações que envolvam dados de saúde pública.

Nesta tese, demonstrou-se como matemática e epidemiologia se podem unir para enfrentar grandes desafios da humanidade como foi o caso da pandemia COVID-19. Através de metodologias inovadoras de desagregação temporal de dados, de deteção de mudanças no padrão de crescimento de um processo e de modelação espaço-temporal do risco, este trabalho contribui para o avanço do conhecimento científico e para a melhoria das políticas públicas de saúde, mas também para a melhoria de processos de natureza diversa. Espera-se que este trabalho inspire outros investigadores a explorar novas formas de tratar, analisar e monitorizar dados epidemiológicos ou outros, bem como a desenvolver soluções criativas e

eficazes para combater outras doenças infecciosas emergentes e para aprimorar a qualidade de dados epidemiológicos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abbott, S., Hellewell, J., Thompson, R. N., Sherratt, K., Gibbs, H. P., Bosse, N. I., Munday, J. D., Meakin, S., Doughty, E. L., Chun, J. Y., Chan, Y.-W. D., Finger, F., Campbell, P., Endo, A., Pearson, C. A. B., Gimma, A., Russell, T., CMMID COVID modelling group, Flasche, S., ... Funk, S. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*, 5, 112.
<https://doi.org/10.12688/wellcomeopenres.16006.2>
- Aboubakr, H. A., Sharafeldin, T. A., & Goyal, S. M. (2021). Stability of SARS-CoV-2 and other coronaviruses in the environment and on common touch surfaces and the influence of climatic conditions: A review. *Transboundary and Emerging Diseases*, 68(2), 296–312. <https://doi.org/10.1111/tbed.13707>
- Adhikari, A., & Yin, J. (2020). Short-Term Effects of Ambient Ozone, PM2.5, and Meteorological Factors on COVID-19 Confirmed Cases and Deaths in Queens, New York. *International Journal of Environmental Research and Public Health*, 17(11), 4047. <https://doi.org/10.3390/ijerph17114047>
- Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons Inc.
- Ahmed, N., Natarajan, T., & Rao, K. R. (1974). Discrete Cosine Transform. *IEEE Transactions on Computers*, C-23(1), 90–93. <https://doi.org/10.1109/T-C.1974.223784>
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. Em *Selected papers of hirotugu akaike* (pp. 199–213). Springer.
<https://www.gwern.net/docs/statistics/decision/1998-akaike.pdf>
- Alaimo Di Loro, P., Divino, F., Farcomeni, A., Jona Lasinio, G., Lovison, G., Maruotti, A., & Mingione, M. (2021). Nowcasting COVID-19 incidence indicators during the Italian first outbreak. *Statistics in Medicine*, 40(16), 3843–3864.
<https://doi.org/10.1002/sim.9004>
- Alarcon Falconi, T. M., Estrella, B., Sempértegui, F., & Naumova, E. N. (2020). Effects of Data Aggregation on Time Series Analysis of Seasonal Infections. *International Journal of Environmental Research and Public Health*, 17(16), 5887.
<https://doi.org/10.3390/ijerph17165887>
- Alba, E. de. (1988). Disaggregation and Forecasting: A Bayesian Analysis. *Journal of Business & Economic Statistics*, 6(2), 197–206.
<https://doi.org/10.1080/07350015.1988.10509654>
- Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., & Rousseau, J. (2012). Combining Expert Opinions in Prior Elicitation. *Bayesian Analysis*, 7(3). <https://doi.org/10.1214/12-BA717>
- Almutairi, F. M., Kanatsoulis, C. I., & Sidiropoulos, N. D. (2021). Prema: Principled Tensor Data Recovery From Multiple Aggregated Views. *IEEE Journal of Selected Topics in Signal Processing*, 15(3), 535–549.
<https://doi.org/10.1109/JSTSP.2021.3056918>
- Almutairi, F. M., Yang, F., Song, H. A., Faloutsos, C., Sidiropoulos, N., & Zadorozhny, V. (2018). Homerun: Scalable sparse-spectrum reconstruction of aggregated historical data. *Proceedings of the VLDB Endowment*, 11(11), 1496–1508.
<https://doi.org/10.14778/3236187.3236201>

- Al-Osh, M. (1989). A dynamic linear model approach for disaggregating time series data. *Journal of Forecasting*, 8(2), 85–96.
- AmaralTurkman, M. A. A., Paulino, C. D., & Müller, P. (2019). *Computational Bayesian Statistics: An Introduction* (1.^a ed.). Cambridge University Press. <https://doi.org/10.1017/9781108646185>
- Anderson, C., & Ryan, L. (2017). A Comparison of Spatio-Temporal Disease Mapping Approaches Including an Application to Ischaemic Heart Disease in New South Wales, Australia. *International Journal of Environmental Research and Public Health*, 14(2), 146. <https://doi.org/10.3390/ijerph14020146>
- Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: An introduction to spatial data analysis. Em *Handbook of applied spatial analysis* (pp. 73–89). Springer.
- Aragão, A., & dos Santos, J. G. (2019). Sistemas Sociais Complexos e Integração de Geodados no Direito e nas Políticas. *Instituto Jurídico* https://www.uc.pt/site/assets/files/433597/lc_sistemas_sociais_complexos_e_integracao_no_direito_e_nas_politicas.pdf.
- Arambepola, R., Lucas, T. C. D., Nandi, A. K., Gething, P. W., & Cameron, E. (2022). A simulation study of disaggregation regression for spatial disease mapping. *Statistics in Medicine*, 41(1), 1–16. <https://doi.org/10.1002/sim.9220>
- Arbia, G. (1989). Theoretical problems motivation. Em J. P. Ancot & A. J. H. Hallet, *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems* (Vol. 14, pp. 7–31). Springer Netherlands. https://doi.org/10.1007/978-94-009-2395-9_2
- Atluri, G., Karpatne, A., & Kumar, V. (2018). Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, 51(4), 1–41.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data* (0 ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b17115>
- Banerjee, S., & Dey, D. K. (2005). Semiparametric Proportional Odds Models for Spatially Correlated Survival Data. *Lifetime Data Analysis*, 11(2), 175–191. <https://doi.org/10.1007/s10985-004-0382-z>
- Barcellan, R. (2003). ECOTRIM: a program for temporal disaggregation of time series. *Workshop on Quarterly National Accounts, Eurostat, Theme, 2*, 79–95.
- Bellomo, N., Bingham, R., Chaplain, M. A. J., Dosi, G., Forni, G., Knopoff, D. A., Lowengrub, J., Twarock, R., & Virgillito, M. E. (2020). A multiscale model of virus pandemic: Heterogeneous interactive entities in a globally connected world. *Mathematical Models and Methods in Applied Sciences*, 30(08), 1591–1651. <https://doi.org/10.1142/S0218202520500323>
- Benedetti, M. H., Berrocal, V. J., & Little, R. J. (2021a). *Accounting for survey design in Bayesian disaggregation of survey-based areal estimates of proportions: An application to the American Community Survey*. <https://doi.org/10.48550/ARXIV.2112.06802>
- Benedetti, M. H., Berrocal, V. J., & Little, R. J. (2021b). Accounting for survey design in Bayesian disaggregation of survey-based areal estimates of proportions: An application to the American Community Survey. *arXiv preprint arXiv:2112.06802*.
- Bernardo, J. M., & Smith, A. F. M. (2004). *Bayesian theory* (Repr). Wiley.
- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20. <https://doi.org/10.1007/BF00116466>

- Bisio, L., & Moauro, F. (2018). Temporal disaggregation by dynamic regressions: Recent developments in Italian quarterly national accounts. *Statistica Neerlandica*, 72(4), 471–494.
- Blangiardo, M., Bouliari, A., Diggle, P., Piel, F. B., Shaddick, G., & Elliott, P. (2020). Advances in spatiotemporal models for non-communicable disease surveillance. *International Journal of Epidemiology*, 49(Supplement_1), i26–i37.
- Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley and Sons, Inc.
- Boot, J. C. G., Feibes, W., & Lisman, J. H. C. (1967). Further Methods of Derivation of Quarterly Figures from Annual Data. *Applied Statistics*, 16(1), 65. <https://doi.org/10.2307/2985238>
- Bournay & Laroque. (1979). Réflexions sur la méthode d'élaboration des comptes trimestriels. *Annales de l'inséé*, 36, 3. <https://doi.org/10.2307/20075332>
- Boyd, S. (2010). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, 3(1), 1–122. <https://doi.org/10.1561/22000000016>
- Breidl, K., & Di Baldassarre, G. (2019). Space-time disaggregation of precipitation and temperature across different climates and spatial scales. *Journal of Hydrology: Regional Studies*, 21, 126–146. <https://doi.org/10.1016/j.ejrh.2018.12.002>
- Britten, G. L., Mohajerani, Y., Primeau, L., Aydin, M., Garcia, C., Wang, W.-L., Pasquier, B., Cael, B. B., & Primeau, F. W. (2021). Evaluating the Benefits of Bayesian Hierarchical Methods for Analyzing Heterogeneous Environmental Datasets: A Case Study of Marine Organic Carbon Fluxes. *Frontiers in Environmental Science*, 9. <https://doi.org/10.3389/fenvs.2021.491636>
- Briz-Redón, Á., Iftimi, A., Correcher, J. F., De Andrés, J., Lozano, M., & Romero-García, C. (2022). A comparison of multiple neighborhood matrix specifications for spatio-temporal model fitting: A case study on covid-19 data. *Stochastic Environmental Research and Risk Assessment*, 36(1), 271–282.
- Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1(1), 29–53. <https://doi.org/10.1002/jae.3950010104>
- Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, 46(3), 347–364. [https://doi.org/10.1016/0304-4076\(90\)90014-K](https://doi.org/10.1016/0304-4076(90)90014-K)
- Cameron, A. C., & Trivedi, P. K. (2014). *Regression Analysis of Count Data* (2nd edition). Cambridge University Press.
- Casanova, L. M., Jeon, S., Rutala, W. A., Weber, D. J., & Sobsey, M. D. (2010). Effects of Air Temperature and Relative Humidity on Coronavirus Survival on Surfaces. *Applied and Environmental Microbiology*, 76(9), 2712–2717. <https://doi.org/10.1128/AEM.02291-09>
- Chan, K. H., Peiris, J. S. M., Lam, S. Y., Poon, L. L. M., Yuen, K. Y., & Seto, W. H. (2011). The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. *Advances in Virology*, 2011, 1–7. <https://doi.org/10.1155/2011/734690>
- Chen, J., Song, J. J., & Stamey, J. D. (2022). A Bayesian Hierarchical Spatial Model to Correct for Misreporting in Count Data: Application to State-Level COVID-19 Data in the United States. *International Journal of Environmental Research and Public Health*, 19(6), 3327. <https://doi.org/10.3390/ijerph19063327>

- Chen, S., Huang, L., Cai, D., Li, B., & Yang, J. (2023). Association between meteorological factors and COVID-19: A systematic review. *International Journal of Environmental Health Research*, 33(12), 1254–1268. <https://doi.org/10.1080/09603123.2022.2083090>
- Cheng, T., & Adepeju, M. (2014). Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. *PLoS ONE*, 9(6), e100465. <https://doi.org/10.1371/journal.pone.0100465>
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395–400. <https://doi.org/10.1126/science.aba9757>
- Chow, G. C., & Lin, A. (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics*, 53(4), 372. <https://doi.org/10.2307/1928739>
- Clark, J. S. (2004). Why environmental scientists are becoming Bayesians: Modelling with Bayes. *Ecology Letters*, 8(1), 2–14. <https://doi.org/10.1111/j.1461-0248.2004.00702.x>
- Clark, J. S., & Gelfand, A. E. (Eds.). (2006). *Hierarchical modelling for the environmental sciences: Statistical methods and applications*. Oxford University Press.
- Cohen, K. J., Müller, W., & Padberg, M. W. (1971). Autoregressive approaches to disaggregation of time series data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 20(2), 119–129.
- Coly, S., Garrido, M., Abrial, D., & Yao, A.-F. (2021). Bayesian hierarchical models for disease mapping applied to contagious pathologies. *PLOS ONE*, 16(1), e0222898. <https://doi.org/10.1371/journal.pone.0222898>
- Costa, E. M., & Marques Da Costa, N. (2020). O processo pandémico da Covid-19 em Portugal Continental. *Finisterra*, vol. 55 n.º 115 (AOP) (2020): Número especial: COVID19. <https://doi.org/10.18055/FINIS20361>
- Cressie, N. A. C. (1993). *Statistics for Spatial Data: Cressie/Statistics*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119115151>
- Cressie, N., Calder, C. A., Clark, J. S., Hoef, J. M. V., & Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, 19(3), 553–570. <https://doi.org/10.1890/07-0744.1>
- Cressie, N., & Moores, M. T. (2021). *Spatial Statistics*. <https://doi.org/10.48550/ARXIV.2105.07216>
- Dayaratna, K. D., Gonshorowski, D., & Kolesar, M. (2022). Hierarchical Bayesian spatio-temporal modeling of COVID-19 in the United States. *Journal of Applied Statistics*, 1–18. <https://doi.org/10.1080/02664763.2022.2069232>
- De Nicola, G., Schneble, M., Kauermann, G., & Berger, U. (2022). Regional now- and forecasting for data reported with delay: Toward surveillance of COVID-19 infections. *AStA Advances in Statistical Analysis*. <https://doi.org/10.1007/s10182-021-00433-5>
- de Rivera, O. R., Blangiardo, M., López-Quílez, A., & Martín-Sanz, I. (2019). Species distribution modelling through Bayesian hierarchical approach. *Theoretical Ecology*, 12(1), 49–59. <https://doi.org/10.1007/s12080-018-0387-y>

- Deckard, M., & Schnell, C. (2023). The Temporal (In)Stability of Violent Crime Hot Spots Between Months and The Modifiable Temporal Unit Problem. *Crime & Delinquency*, 69(6–7), 1312–1335. <https://doi.org/10.1177/0011287221128483>
- Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4). <https://doi.org/10.18637/jss.v064.i04>
- Denton, F. T. (1971). Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization. *Journal of the American Statistical Association*, 66(333), 99–102. <https://doi.org/10.1080/01621459.1971.10482227>
- Detyniecki, M., Bouchon-meunier, D. B., Yager, D. R., Prade, R. H., & Grabisch, E. M. (2000). *Mathematical Aggregation Operators and their Application to Video Querying*. <https://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B7FE42D48F22E9A2C68D5F8ED2404794?doi=10.1.1.21.17&rep=rep1&type=pdf>
- DGS. (2020, novembro 16). COVID-19 | RELATÓRIO DE SITUAÇÃO | 16 NOVEMBRO 2020. 2020/11/16. <https://covid19.min-saude.pt/relatorio-de-situacao/>
- Di Fonzo, T. (2002). Temporal disaggregation of a system of time series when the aggregate is known. Optimal vs. Adjustment methods. *Workshop on Quarterly National Accounts*, 63–77.
- DIF/CGTP-IN. (2020). 1º de Maio da CGTP-IN. Em *CGTP-IN*. <http://www.cgtp.pt/informacao/comunicacao-sindical/14050-1-de-maio-da-cgtp-in-2>
- Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models* (Fourth edition). CRC Press, Taylor & Francis Group.
- Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16(2), 129–138. <https://doi.org/10.1111/j.1466-8238.2006.00279.x>
- Dunson, D. B. (2001). Commentary: Practical Advantages of Bayesian Analysis of Epidemiologic Data. *American Journal of Epidemiology*, 153(12), 1222–1226. <https://doi.org/10.1093/aje/153.12.1222>
- Eilers, P. H. (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. *Statistical Modelling*, 7(3), 239–254. <https://doi.org/10.1177/1471082X0700700302>
- Elias, C., Sekri, A., Leblanc, P., Cucherat, M., & Vanhems, P. (2021). The incubation period of COVID-19: A meta-analysis. *International Journal of Infectious Diseases*, 104, 708–710. <https://doi.org/10.1016/j.ijid.2021.01.069>
- Embury, J., Tsou, M.-H., Nara, A., & Oren, E. (2022). A Spatio-Demographic Perspective on the Role of Social Determinants of Health and Chronic Disease in Determining a Population’s Vulnerability to COVID-19. *Preventing Chronic Disease*, 19, 210414. <https://doi.org/10.5888/pcd19.210414>
- Evans, M., & Swartz, T. (1995). Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems. *Statistical Science*, 10(3). <https://doi.org/10.1214/ss/1177009938>
- Fan, J., Zhang, C., & Zhang, J. (2001). Generalized Likelihood Ratio Statistics and Wilks Phenomenon. *The Annals of Statistics*, 29(1). <https://doi.org/10.1214/aos/996986505>

- Fernandez, R. B. (1981). A Methodological Note on the Estimation of Time Series. *The Review of Economics and Statistics*, 63(3), 471. <https://doi.org/10.2307/1924371>
- Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.
- Fonzo, T. D. (1990). The Estimation of M Disaggregate Time Series when Contemporaneous and Temporal Aggregates are Known. *The Review of Economics and Statistics*, 72(1), 178. <https://doi.org/10.2307/2109758>
- Fonzo, T. D. (2003). *Temporal disaggregation of economic time series: Towards a dynamic extension*. https://ec.europa.eu/eurostat/documents/3888793/5816173/KS_AN-03-035-EN.PDF/21c4417c-dbec-45ec-b440-fe8bf95661b7?version=1.0
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third edition). SAGE.
- Freeman, G. H., & Halton, J. H. (1951). Note on an Exact Treatment of Contingency, Goodness of Fit and Other Problems of Significance. *Biometrika*, 38(1/2), 141. <https://doi.org/10.2307/2332323>
- Fuller, W. A. (2009). *Introduction to statistical time series*. John Wiley & Sons.
- García, C. N. (2021). Socioeconomic, demographic and healthcare determinants of the COVID-19 pandemic: An ecological study of Spain. *BMC Public Health*, 21(1), 606. <https://doi.org/10.1186/s12889-021-10658-3>
- Gelfand, A. E. (2001). On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1), 31–45. <https://doi.org/10.1093/biostatistics/2.1.31>
- Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6), 721–741. <https://doi.org/10.1109/TPAMI.1984.4767596>
- Ghadimi, E., Teixeira, A., Shames, I., & Johansson, M. (2015). Optimal Parameter Selection for the Alternating Direction Method of Multipliers (ADMM): Quadratic Problems. *IEEE Transactions on Automatic Control*, 60(3), 644–658. <https://doi.org/10.1109/TAC.2014.2354892>
- GILLIBERT, André. (2020, agosto 7). *glmglrt: GLRT P-Values in Generalized Linear Models*. <https://cran.r-project.org/web/packages/glmglrt/index.html>
- Ginsburgh, V. A. (1973). A Further Note on the Derivation of Quarterly Figures Consistent with Annual Data. *Applied Statistics*, 22(3), 368. <https://doi.org/10.2307/2346784>
- Gómez, C., González, C., Osses, M., & Aristizábal, B. (2018). Spatial and temporal disaggregation of the on-road vehicle emission inventory in a medium-sized Andean city. Comparison of GIS-based top-down methodologies. *Atmospheric environment*, 179, 142–155.
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, -Raoul J, Drosten, C., Gulyaeva, A. A., Haagmans, B., Lauber, C., Leontovich, A., Neuman, B., & others. (2020). Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol*, 5(4), 536–544.
- Gotway, C. A., & Young, L. J. (2002). Combining Incompatible Spatial Data. *Journal of the American Statistical Association*, 97(458), 632–648. <https://doi.org/10.1198/016214502760047140>

- Grabisch, M., Marichal, J.-L., Mesiar, R., & Pap, E. (Eds.). (2009). *Aggregation functions* (1. publ). Cambridge University Press.
- Greene, S. K., McGough, S. F., Culp, G. M., Graf, L. E., Lipsitch, M., Menzies, N. A., & Kahn, R. (2020). *Evaluation of Nowcasting for Real-Time COVID-19 Tracking—New York City, March–May 2020* [Preprint]. *Epidemiology*.
<https://doi.org/10.1101/2020.10.18.20209189>
- Gudmundsson, G. (1999). Disaggregation of annual flow data with multiplicative trends. *Journal of Forecasting*, 18(1), 33–37. [https://doi.org/10.1002/\(SICI\)1099-131X\(199901\)18:1<33::AID-FOR687>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-131X(199901)18:1<33::AID-FOR687>3.0.CO;2-T)
- Guedes, Nuno. (2020, abril 14). Cientistas contestam protocolo para receberem (poucos) dados sobre doentes com Covid-19. *TSF*.
<https://www.tsf.pt/portugal/sociedade/cientistas-contestam-protocolo-para-receberem-poucos-dados-sobre-doentes-com-covid-19-12066474.html>
- Guerrero, V. M. (1990). Temporal disaggregation of time series: An ARIMA-based approach. *International Statistical Review/Revue Internationale de Statistique*, 29–46.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice* (1.^a ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511754944>
- Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does Density Aggravate the COVID-19 Pandemic?: Early Findings and Lessons for Planners. *Journal of the American Planning Association*, 86(4), 495–509.
<https://doi.org/10.1080/01944363.2020.1777891>
- Hardin, J. W., & Hilbe, J. M. (2018). *Generalized linear models and extensions* (Fourth edition). Stata Press.
- Haridy, S., Maged, A., Baker, A. W., Shamsuzzaman, M., Bashir, H., & Xie, M. (2021). Monitoring scheme for early detection of coronavirus and other respiratory virus outbreaks. *Computers & Industrial Engineering*, 156, 107235.
<https://doi.org/10.1016/j.cie.2021.107235>
- Heinig, G. & Rost. (2013). *Algebraic Methods for Toeplitz-Like Matrices and Operators*. Birkhäuser Boston.
- Hodgess, Erin M. & Mhoon, Kendra. (2019). Temporal Disaggregation of Time Series Revisited. *Management Studies*, 7(4). <https://doi.org/10.17265/2328-2185/2019.04.004>
- Hormati, A., & Vetterli, M. (2007). *Annihilating filter-based decoding in the compressed sensing framework* (D. Van De Ville, V. K. Goyal, & M. Papadakis, Eds.; p. 670121). <https://doi.org/10.1117/12.732308>
- Hotta, L. K., Morettin, P. A., & Pereira, P. L. V. (1992). The Effect of Overlapping Aggregation on Time Series Models: An Application to the Unemployment Rate in Brazil. *Brazilian Review of Econometrics*, 12(2), 223.
<https://doi.org/10.12660/bre.v12n21992.2992>
- Hotta, L. K., & Vasconcellos, K. L. (1999). Aggregation and Disaggregation of Structural Time Series Models. *Journal of Time Series Analysis*, 20(2), 155–171.
<https://doi.org/10.1111/1467-9892.00131>
- Hsieh, Y.-H. (2009). Richards Model: A Simple Procedure for Real-time Prediction of Outbreak Severity. Em Z. Ma, Y. Zhou, & J. Wu, *Series in Contemporary Applied Mathematics* (Vol. 11, pp. 216–236). CO-PUBLISHED WITH HIGHER EDUCATION PRESS. https://doi.org/10.1142/9789814261265_0009

- Hsieh, Y.-H., Fisman, D. N., & Wu, J. (2010). On epidemic modeling in real time: An application to the 2009 Novel A (H1N1) influenza outbreak in Canada. *BMC Research Notes*, 3(1), 283. <https://doi.org/10.1186/1756-0500-3-283>
- Hsieh, Y.-H., Lee, J.-Y., & Chang, H.-L. (2004). SARS Epidemiology Modeling. *Emerging Infectious Diseases*, 10(6), 1165–1167. <https://doi.org/10.3201/eid1006.031023>
- Huang, K., Sidiropoulos, N. D., & Liavas, A. P. (2015). *A Flexible and Efficient Algorithmic Framework for Constrained Matrix and Tensor Factorization*. <https://doi.org/10.48550/ARXIV.1506.04209>
- Huang, Z., Huang, J., Gu, Q., Du, P., Liang, H., & Dong, Q. (2020). Optimal temperature zone for the dispersal of COVID-19. *Science of The Total Environment*, 736, 139487. <https://doi.org/10.1016/j.scitotenv.2020.139487>
- Ibañez, M. V., Martínez-García, M., & Simó, A. (2021). A Review of Spatiotemporal Models for Count Data in R Packages. A Case Study of COVID-19 Data. *Mathematics*, 9(13), 1538.
- Javid, B., Weekes, M. P., & Matheson, N. J. (2020). Covid-19: Should the public wear face masks? *BMJ*, m1442. <https://doi.org/10.1136/bmj.m1442>
- Jeevanjee, N. (2011). *An introduction to tensors and group theory for physicists*. Birkhäuser/Springer.
- Jeffery, C., Ozonoff, A., & Pagano, M. (2014). The effect of spatial aggregation on performance when mapping a risk of disease. *International Journal of Health Geographics*, 13(1), 9. <https://doi.org/10.1186/1476-072X-13-9>
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007), 453–461. <https://doi.org/10.1098/rspa.1946.0056>
- Jewell, C. P., Kypraios, T., Neal, P., & Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(3). <https://doi.org/10.1214/09-BA417>
- Kay, S. M., & Kay, S. M. (2011). *Fundamentals of statistical signal processing. 2: Detection theory* (15. pr). Prentice Hall PTR.
- Kim, H., Sun, D., & Tsutakawa, R. K. (2001). A Bivariate Bayes Method for Improving the Estimates of Mortality Rates With a Twofold Conditional Autoregressive Model. *Journal of the American Statistical Association*, 96(456), 1506–1521. <https://doi.org/10.1198/016214501753382408>
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17–18), 2555–2567. [https://doi.org/10.1002/1097-0258\(20000915/30\)19:17/18<2555::AID-SIM587>3.0.CO;2-#](https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-#)
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R Package for Assessing Multivariate Normality. *The R Journal*, 6(2), 151–162.
- Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., Open COVID-19 Data Working Group†, du Plessis, L., Faria, N. R., Li, R., Hanage, W. P., Brownstein, J. S., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O. G., & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497. <https://doi.org/10.1126/science.abb4218>
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., Davies, N., Gimma, A., van Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, C. I., Clifford, S., Quilty, B. J., Bosse, N. I., ...

- Flasche, S. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *The Lancet Infectious Diseases*, 20(5), 553–558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
- Labonne, P., & Weale, M. (2020). Temporal disaggregation of overlapping noisy quarterly data: Estimation of monthly output from UK value-added tax data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3), 1211–1230.
- Labonne, P., Weale, M., & others. (2018). *Temporal disaggregation of overlapping noisy quarterly data using state space models: Estimation of monthly business sector output from Value Added Tax data in the UK*. Economic Statistics Centre of Excellence (ESCoE).
- Lalwani, S., Sahni, G., Mewara, B., & Kumar, R. (2020). Predicting optimal lockdown period with parametric approach using three-phase maturation SIRD model for COVID-19 pandemic. *Chaos, Solitons & Fractals*, 138, 109939. <https://doi.org/10.1016/j.chaos.2020.109939>
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9), 577–582. <https://doi.org/10.7326/M20-0504>
- Lawson, A. B. (2018). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC.
- Lawson, A. B., & Zhou, H. (2005). Spatial statistical modeling of disease outbreaks with particular reference to the UK foot and mouth disease (FMD) epidemic of 2001. *Preventive Veterinary Medicine*, 71(3–4), 141–156. <https://doi.org/10.1016/j.prevetmed.2005.07.002>
- Lawson, A., Banerjee, S., Haining, R. P., & Ugarte, M. D. (Eds.). (2016). *Handbook of spatial epidemiology*. CRC Press, Taylor & Francis Group.
- Lawson, A., & Choi, J. (2016). Spatiotemporal disease mapping. *Handbook of Spatial Epidemiology*, 335–348.
- Leal, C., Morgado, Leonel, & Oliveira, Teresa A. (2023, fevereiro). *Mathematical and Statistical Modelling for assessing COVID-19 superspreader contagion: Analysis of geographical heterogeneous impacts from public events*.
- Lee, D.-J., Durbán, M., Ayma, D., & Van de Kastele, J. (2022). Modeling latent spatio-temporal disease incidence using penalized composite link models. *Plos one*, 17(3), e0263711.
- Lee, S. Y., Lei, B., & Mallick, B. (2020). Estimation of COVID-19 spread curves integrating global data and borrowing information. *PLOS ONE*, 15(7), e0236860. <https://doi.org/10.1371/journal.pone.0236860>
- Lei n.º 75/2013 (Vol. 176/2013). (2013). <https://data.dre.pt/eli/lei/75/2013/09/12/p/dre/pt/html>
- Litterman, R. B. (1983). A Random Walk, Markov Model for the Distribution of Time Series. *Journal of Business & Economic Statistics*, 1(2), 169. <https://doi.org/10.2307/1391858>
- Liu, Z., Song, H. A., Zadorozhny, V., Faloutsos, C., & Sidiropoulos, N. (2017). H-fuse: Efficient fusion of aggregated historical data. *Proceedings of the 2017 SIAM International Conference on Data Mining*, 786–794. https://www.cs.cmu.edu/~hyunahs/papers/SDM2017_2.pdf

- Lord, D., Park, B.-J., & Model, P.-G. (2012). Negative binomial regression models and estimation methods. *Probability Density and Likelihood Functions. Texas A&M University, Korea Transport Institute*, 1–15.
- Lowen, A. C., Mubareka, S., Steel, J., & Palese, P. (2007). Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. *PLoS Pathogens*, 3(10), e151. <https://doi.org/10.1371/journal.ppat.0030151>
- Lowen, A. C., Steel, J., Mubareka, S., & Palese, P. (2008). High Temperature (30°C) Blocks Aerosol but Not Contact Transmission of Influenza Virus. *Journal of Virology*, 82(11), 5650–5652. <https://doi.org/10.1128/JVI.00325-08>
- Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). Extracting, Computing and Exploring the Parameters of Statistical Models using R. *Journal of Open Source Software*, 5(53), 2445. <https://doi.org/10.21105/joss.02445>
- Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>
- Luo, X., Duan, H., & Xu, K. (2021). A novel grey model based on traditional Richards model and its application in COVID-19. *Chaos, Solitons & Fractals*, 142, 110480. <https://doi.org/10.1016/j.chaos.2020.110480>
- Lusa. (2020, junho 6). Milhares manifestaram-se contra o racismo nas ruas de Lisboa. *SÁBADO*. sábado.pt/portugal/detalhe/milhares-manifestam-se-contraracismo-nas-ruas-de-lisboa
- Marinović, A. B., Swaan, C., van Steenberg, J., & Kretzschmar, M. (2015). Quantifying Reporting Timeliness to Improve Outbreak Control. *Emerging Infectious Diseases*, 21(2), 209–216. <https://doi.org/10.3201/eid2102.130504>
- Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3), 539–553. <https://doi.org/10.1093/biomet/ast023>
- Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, 67, 68–83. <https://doi.org/10.1016/j.csda.2013.04.014>
- Masselot, P., Chebana, F., Bélanger, D., St-Hilaire, A., Abdous, B., Gosselin, P., & Ouarda, T. B. M. J. (2018). *Aggregating the response in time series regression models, applied to weather-related cardiovascular mortality*. <https://doi.org/10.48550/ARXIV.1802.07762>
- McClymont, H., & Hu, W. (2021). Weather Variability and COVID-19 Transmission: A Review of Recent Research. *International Journal of Environmental Research and Public Health*, 18(2), 396. <https://doi.org/10.3390/ijerph18020396>
- McCullagh, P., & Nelder, J. A. (2019). *Generalized Linear Models* (2.^a ed.). Routledge. <https://doi.org/10.1201/9780203753736>
- McDonald, J. H., & Delaware, U. of. (2009). *Handbook of Biological Statistics*. Sparky House Publishing. <https://books.google.pt/books?id=AsRTywAACAAJ>
- Mecenas, P., Bastos, R. T. da R. M., Vallinoto, A. C. R., & Normando, D. (2020). Effects of temperature and humidity on the spread of COVID-19: A systematic review. *PLOS ONE*, 15(9), e0238339. <https://doi.org/10.1371/journal.pone.0238339>
- Melo, H. P. M., Henriques, J., Carvalho, R., Verma, T., da Cruz, J. P., & Araújo, N. A. M. (2021). Heterogeneous impact of a lockdown on inter-municipality mobility. *Physical Review Research*, 3(1), 013032. <https://doi.org/10.1103/PhysRevResearch.3.013032>

- Mo, B., Feng, K., Shen, Y., Tam, C., Li, D., Yin, Y., & Zhao, J. (2021). Modeling epidemic spreading through public transit using time-varying encounter network. *Transportation Research Part C: Emerging Technologies*, 122, 102893. <https://doi.org/10.1016/j.trc.2020.102893>
- Moauo, F., & Savio, G. (2005). Temporal disaggregation using multivariate structural time series models. *The Econometrics Journal*, 8(2), 214–234.
- Moazeni, M., Rahimi, M., & Ebrahimi, A. (2023). What are the effects of climate variables on COVID-19 pandemic? A systematic review and current update. *Advanced Biomedical Research*, 12(1), 33. https://doi.org/10.4103/abr.abr_145_21
- Mohd, W. R. W., & Abdullah, L. (2017). Aggregation methods in group decision making: A decade survey. *Informatica*, 41(1). <https://www.informatica.si/index.php/informatica/article/viewFile/1321/972>
- Monteiro, Martins, Murrieta-Flores, & Moura Pires. (2019). Spatial Disaggregation of Historical Census Data Leveraging Multiple Sources of Ancillary Information. *ISPRS International Journal of Geo-Information*, 8(8), 327. <https://doi.org/10.3390/ijgi8080327>
- Moraga, P. (2018). Small area disease risk estimation and visualization using r. *R J.*, 10(1), 495.
- Moraga, P. (2020). *Geospatial health data: Modeling and visualization with R-INLA and Shiny*. CRC Press/Taylor & Francis Group.
- Mueller, A. L., McNamara, M. S., & Sinclair, D. A. (2020). Why does COVID-19 disproportionately affect older people? *Aging*, 12(10), 9959–9981. <https://doi.org/10.18632/aging.103344>
- Nandi, A. K., Lucas, T. C., Arambepola, R., Gething, P., & Weiss, D. J. (2020). Disaggregation: An R package for Bayesian spatial disaggregation modelling. *arXiv preprint arXiv:2001.04847*.
- Nazia, N., Butt, Z. A., Bedard, M. L., Tang, W.-C., Sehar, H., & Law, J. (2022). Methods Used in the Spatial and Spatiotemporal Analysis of COVID-19 Epidemiology: A Systematic Review. *International journal of environmental research and public health*, 19(14), 8267.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370. <https://doi.org/10.2307/2344614>
- Notari, A. (2021). Temperature dependence of COVID-19 transmission. *Science of The Total Environment*, 763, 144390. <https://doi.org/10.1016/j.scitotenv.2020.144390>
- Observador. (2020, abril 17). DGS forneceu dados sobre doentes Covid-19 a 5 grupos de investigadores. Há 320 pedidos «em apreciação». *Observador*. <https://observador.pt/2020/04/16/dgs-forneceu-dados-sobre-doentes-covid-19-a-5-grupos-de-investigadores-ha-320-pedidos-em-apreciacao/>
- O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities* (1.^a ed.). Wiley. <https://doi.org/10.1002/0470033312>
- Oliveira, A. A. (2018). *Modelling Reporting Delays for Infectious Diseases: An Application to the Portuguese HIV-AIDS Surveillance Data*.
- OMS, W. H. O. (2020). *Clinical management of COVID-19: Interim guidance, 27 May 2020* (p. 62 p.) [Technical documents]. World Health Organization.
- Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography*.

- Opitz, T. (2017). *Latent Gaussian modeling and INLA: A review with focus on space-time applications*. <https://doi.org/10.48550/ARXIV.1708.02723>
- Orientação 015. (2020). Direção-Geral da Saúde.
- Oswald, S. A., Nisbet, I. C. T., Chiaradia, A., & Arnold, J. M. (2012). FlexParamCurve: R package for flexible fitting of nonlinear parametric curves: *Nonlinear parametric curve-fitting*. *Methods in Ecology and Evolution*, 3(6), 1073–1077. <https://doi.org/10.1111/j.2041-210X.2012.00231.x>
- Ozonoff, A., Jeffery, C., Manjourides, J., White, L. F., & Pagano, M. (2007). Effect of spatial resolution on cluster detection: A simulation study. *International Journal of Health Geographics*, 6(1), 52. <https://doi.org/10.1186/1476-072X-6-52>
- Paez, A., Lopez, F. A., Menezes, T., Cavalcanti, R., & Pitta, M. G. D. R. (2021). A Spatio-Temporal Analysis of the Environmental Correlates of COVID-19 Incidence in Spain. *Geographical Analysis*, 53(3), 397–421. <https://doi.org/10.1111/gean.12241>
- Paulino, C. D., Turkman, M. A. A., Murteira, B., & Silva, G. L. (2018). *Estatística bayesiana (2a edição revista e ampliada)*. Fundação Calouste Gulbenkian.
- Pavía-Miralles, J. M. (2010). A Survey of Methods to Interpolate, Distribute and Extrapolate Time Series. *Journal of Service Science and Management*, 03(04), 449–463. <https://doi.org/10.4236/jssm.2010.34051>
- Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3), 406–413. <https://doi.org/10.1017/S0305004100030401>
- Pereira, S. A. G. (2018). *Spatio-temporal methods and models for unemployment estimation*. https://repositorio.ul.pt/bitstream/10451/34867/1/ulsd732237_td_Soraia_Pereira.pdf
- Perricone, C. (2018). Wavelet Analysis for Temporal Disaggregation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3274721>
- Poissonnier, A. (2017). The Chow-Lin method extended to dynamic models with autocorrelated residuals. *Journal of Time Series Econometrics*, 10(1), 20160007. <https://doi.org/10.1515/jtse-2016-0007>
- Pollet, T. V., Stulp, G., Henzi, S. P., & Barrett, L. (2015). Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data: Data Aggregation and the Ecological Fallacy. *American Journal of Primatology*, 77(7), 727–740. <https://doi.org/10.1002/ajp.22405>
- PORDATA, I. (2019). *(Índice de envelhecimento, população, densidade populacional)*. PORDATA Lisboa.
- Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., Jit, M., Klepac, P., Flasche, S., Clifford, S., Pearson, C. A. B., Munday, J. D., Abbott, S., Gibbs, H., Rosello, A., Quilty, B. J., Jombart, T., Sun, F., Diamond, C., ... Hellewell, J. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet Public Health*, 5(5), e261–e270. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6)
- Proietti, T. (2006). Temporal disaggregation by state space methods: Dynamic regression methods revisited. *The Econometrics Journal*, 9(3), 357–372.
- Proietti, T. (2013). Distribution and interpolation revisited: A structural approach. *Statistica; Vol 58, No 3 (1998)*. <https://doi.org/10.6092/ISSN.1973-2201/1091>

- Python, A., Bender, A., Blangiardo, M., Illian, J. B., Lin, Y., Liu, B., Lucas, T., Tan, S., Wen, Y., Svanidze, D., & Yin, J. (2020). *A downscaling approach to compare COVID-19 count data from databases aggregated at different spatial scales* [Preprint]. *Epidemiology*. <https://doi.org/10.1101/2020.06.17.20133959>
- Qiu, P. (2019). Some recent studies in statistical process control. *Em Statistical quality technologies* (pp. 3–19). Springer.
- Quenneville, B., Picard, F., & Fortier, S. (2013). Calendarization with interpolating splines and state space models: *Calendarization*. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(3), 371–399. <https://doi.org/10.1111/rssc.12002>
- Quilis, E. M. (2018). Temporal disaggregation of economic time series: The view from the trenches: Temporal disaggregation from the trenches. *Statistica Neerlandica*, *72*(4), 447–470. <https://doi.org/10.1111/stan.12150>
- R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing: Reference Index*. R Foundation for Statistical Computing. <http://www.polsci.wvu.edu/duval/PS603/Notes/R/fullrefman.pdf>
- Raim, A. M., Holan, S. H., Bradley, J. R., & Wikle, C. K. (2019). *Spatio-Temporal Change of Support Modeling with R*. <https://doi.org/10.48550/ARXIV.1904.12092>
- Requena, F., & Ciudad, N. M. (2006). A major improvement to the Network Algorithm for Fisher's Exact Test in contingency tables. *Computational Statistics & Data Analysis*, *51*(2), 490–498. <https://doi.org/10.1016/j.csda.2005.09.004>
- Richards, F. J. (1959). A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany*, *10*(2), 290–301. <https://doi.org/10.1093/jxb/10.2.290>
- Richardson, S., & Best, N. (2003). Bayesian hierarchical models in ecological studies of health-environment effects. *Environmetrics*, *14*(2), 129–147. <https://doi.org/10.1002/env.571>
- Rossi, N. (1982). A Note on the Estimation of Disaggregate Time Series When the Aggregate Is Known. *The Review of Economics and Statistics*, *64*(4), 695–696.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- Sahu, S. K. (2022). *Bayesian modeling of spatio-temporal data with R*. Chapman and Hall/CRC.
- Sain, S. R., & Cressie, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics*, *140*(1), 226–259. <https://doi.org/10.1016/j.jeconom.2006.09.010>
- Sain, S. R., Furrer, R., & Cressie, N. (2011). A spatial analysis of multivariate output from regional climate models. *The Annals of Applied Statistics*, *5*(1). <https://doi.org/10.1214/10-AOAS369>
- Santos Silva, J. M. C., & Cardoso, F. N. (2001). The Chow-Lin method using dynamic models. *Economic Modelling*, *18*(2), 269–280. [https://doi.org/10.1016/S0264-9993\(00\)00039-0](https://doi.org/10.1016/S0264-9993(00)00039-0)
- Sax, C., & Steiner, P. (2013). Temporal Disaggregation of Time Series. *The R Journal*, *5*(2), 80. <https://doi.org/10.32614/RJ-2013-028>
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.
- Scrucca, L. (2004). qcc: An R package for quality control charting and statistical process control. *R News*, *4*(1). https://www.researchgate.net/profile/Luca-Scrucca/publication/247303593_Qcc_An_R_package_for_quality_control_charting

- _and_statistical_process_control/links/00b7d53439fa2d6033000000/Qcc-An-R-package-for-quality-control-charting-and-statistical-process-control.pdf
- Selesnick, I. (2012). *Penalty and shrinkage functions for sparse signal processing*.
- Sharkey, P., & Winter, H. C. (2019). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain: A Bayesian hierarchical model for extreme precipitation. *Environmetrics*, 30(1), e2529. <https://doi.org/10.1002/env.2529>
- Silva, E., Guerrero, V. M., & Peña, D. (2011). Temporal disaggregation and restricted forecasting of multiple population time series. *Journal of Applied Statistics*, 38(4), 799–815. <https://doi.org/10.1080/02664761003692316>
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1). <https://doi.org/10.1214/16-STS576>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4), 583–639.
- Stadnytskyi, V., Bax, C. E., Bax, A., & Anfinrud, P. (2020). The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proceedings of the National Academy of Sciences*, 117(22), 11875–11877. <https://doi.org/10.1073/pnas.2006874117>
- Tamerius, J. D., Shaman, J., Alonso, W. J., Bloom-Feshbach, K., Uejio, C. K., Comrie, A., & Viboud, C. (2013). Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLoS Pathogens*, 9(3), e1003194. <https://doi.org/10.1371/journal.ppat.1003194>
- Tang, S., Mao, Y., Jones, R. M., Tan, Q., Ji, J. S., Li, N., Shen, J., Lv, Y., Pan, L., Ding, P., Wang, X., Wang, Y., MacIntyre, C. R., & Shi, X. (2020). Aerosol transmission of SARS-CoV-2? Evidence, prevention and control. *Environment International*, 144, 106039. <https://doi.org/10.1016/j.envint.2020.106039>
- Tariq, A., Roosa, K., Mizumoto, K., & Chowell, G. (2019). Assessing reporting delays and the effective reproduction number: The Ebola epidemic in DRC, May 2018–January 2019. *Epidemics*, 26, 128–133. <https://doi.org/10.1016/j.epidem.2019.01.003>
- Tikhonov, A. N., Goncharsky, A. V., Stepanov, V. V., & Yagola, A. G. (1995). Regularization methods. In A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, & A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems* (pp. 7–63). Springer Netherlands. https://doi.org/10.1007/978-94-015-8480-7_2
- Tirachini, A., & Cats, O. (2020). COVID-19 and Public Transportation: Current Assessment, Prospects, and Research Needs. *Journal of Public Transportation*, 22(1). <https://doi.org/10.5038/2375-0901.22.1.1>
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234. <https://doi.org/10.2307/143141>
- Tycho Projetc. (2013). *Tycho. Project tycho: Data for health*. <https://www.tycho.pitt.edu/>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>
- Wang, X.-S., Wu, J., & Yang, Y. (2012). Richards model revisited: Validation by and application to infection dynamics. *Journal of Theoretical Biology*, 313, 12–19. <https://doi.org/10.1016/j.jtbi.2012.07.024>
- Wassie, G. T., Azene, A. G., Bantie, G. M., Dessie, G., & Aragaw, A. M. (2020). Incubation Period of Severe Acute Respiratory Syndrome Novel Coronavirus 2 that

- Causes Coronavirus Disease 2019: A Systematic Review and Meta-Analysis. *Current Therapeutic Research*, 93, 100607.
<https://doi.org/10.1016/j.curtheres.2020.100607>
- Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, 11(116), 3571–3594.
- Wayant, N. M. (2019). *Temporal disaggregation of annual dengue cases to monthly dengue cases*.
- Wei, W. W. S., & Stram, D. O. (1990). Disaggregation of Time Series Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3), 453–467.
<https://doi.org/10.1111/j.2517-6161.1990.tb01799.x>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wikle, C. K. (2003). Hierarchical Models in Environmental Science. *International Statistical Review*, 71(2), 181–199. <https://doi.org/10.1111/j.1751-5823.2003.tb00192.x>
- Wolfram Research, W. R. (2020). *Mathematica, Version 12.1*.
<https://www.wolfram.com/mathematica>
- Woodall, W. H., Grigg, O. A., & Burkom, H. S. (2010). Research issues and ideas on health-related surveillance. *Frontiers in Statistical Quality Control* 9, 145–155.
- Woodall, W. H., & Montgomery, D. C. (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology*, 46(1), 78–94.
- Wu, K., Darcet, D., Wang, Q., & Sornette, D. (2020). Generalized logistic growth modeling of the COVID-19 outbreak: Comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear Dynamics*, 101(3), 1561–1581.
<https://doi.org/10.1007/s11071-020-05862-6>
- Wu, Y., Jing, W., Liu, J., Ma, Q., Yuan, J., Wang, Y., Du, M., & Liu, M. (2020). Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Science of The Total Environment*, 729, 139051.
<https://doi.org/10.1016/j.scitotenv.2020.139051>
- Xu, J., Zadorozhny, V., & Grant, J. (2020a). A-Cure: An accurate information reconstruction from inaccurate data sources. *Information Systems*, 91, 101508.
<https://doi.org/10.1016/j.is.2020.101508>
- Xu, J., Zadorozhny, V., & Grant, J. (2020b). IncompFuse: A logical framework for historical information fusion with inaccurate data sources. *Journal of Intelligent Information Systems*, 54(3), 463–481. <https://doi.org/10.1007/s10844-019-00569-6>
- Yang, F., Almutairi, F. M., Song, H. A., Faloutsos, C., Sidiropoulos, N. D., & Zadorozhny, V. (2020). TurboLift: Fast accuracy lifting for historical data recovery. *The VLDB Journal*, 29(5), 1129–1148.
- Yang, F., Song, H. A., Liu, Z., Faloutsos, C., Zadorozhny, V., & Sidiropoulos, N. (2018). Ares: Automatic Disaggregation of Historical Data. *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 65–76.
<https://doi.org/10.1109/ICDE.2018.00016>
- Yilmazkuday, H. (2020). COVID-19 Deaths and Inter-County Travel: Daily Evidence from the U.S. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3568838>

- Zaier, L. H., & Abed, M. (2014). Temporal Disaggregation of Economic Time Series using Artificial Neural Networks. *Communications in Statistics - Theory and Methods*, 43(8), 1824–1833. <https://doi.org/10.1080/03610926.2012.677088>
- Zaki, N., & Mohamed, E. A. (2021). The estimations of the COVID-19 incubation period: A scoping reviews of the literature. *Journal of Infection and Public Health*, 14(5), 638–646. <https://doi.org/10.1016/j.jiph.2021.01.019>
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Journal of Statistical Software*, 27(8). <https://doi.org/10.18637/jss.v027.i08>
- Zhang, X., Tao, J., Wang, C., & Shi, N. (2019). Bayesian Model Selection Methods for Multilevel IRT Models: A Comparison of Five DIC-Based Indices. *Journal of Educational Measurement*, 56(1), 3–27. <https://doi.org/10.1111/jedm.12197>
- Zhao, M., Wang, M., Zhang, J., Gu, J., Zhang, P., Xu, Y., Ye, J., Wang, Z., Ye, D., Pan, W., & others. (2020). Comparison of clinical characteristics and outcomes of patients with coronavirus disease 2019 at different ages. *Aging (Albany NY)*, 12(11), 10070.
- Zhou, H., & Hanson, T. (2018). A Unified Framework for Fitting Bayesian Semiparametric Models to Arbitrarily Censored Survival Data, Including Spatially Referenced Data. *Journal of the American Statistical Association*, 113(522), 571–581. <https://doi.org/10.1080/01621459.2017.1356316>
- Zou, J., Karr, A. F., Datta, G., Lynch, J., & Grannis, S. (2014). A Bayesian spatio-temporal approach for real-time detection of disease outbreaks: A case study. *BMC Medical Informatics and Decision Making*, 14(1), 108. <https://doi.org/10.1186/s12911-014-0108-4>

ANEXO I

**TESTES À NORMALIDADE às variáveis demográficas e de mobilidade e ao número de casos COVID-19
nso concelhos da AML**

CAPÍTULO 3

```
> normalidade (ratio.dif$IndMob)
                                valorp
One-sample Kolmogorov-Smirnov test 0.423197
Lilliefors (Kolmogorov-Smirnov) normality test 0.059325
Cramer-von Mises normality test 0.087378
Shapiro-Wilk normality test 0.003200
Shapiro-Francia normality test 0.002759
Anderson-Darling normality test 0.023568
Pearson chi-square normality test 0.336011

> normalidade (ratio.dif$IndEnv)
                                valorp
One-sample Kolmogorov-Smirnov test 0.4013627
Lilliefors (Kolmogorov-Smirnov) normality test 0.0505586
Cramer-von Mises normality test 0.0130096
Shapiro-Wilk normality test 0.0007580
Shapiro-Francia normality test 0.0007844
Anderson-Darling normality test 0.0046172
Pearson chi-square normality test 0.1418777

> normalidade (ratio.dif$Distkmcarr2)
                                valorp
One-sample Kolmogorov-Smirnov test 0.8533
Lilliefors (Kolmogorov-Smirnov) normality test 0.5164
Cramer-von Mises normality test 0.5215
Shapiro-Wilk normality test 0.6997
Shapiro-Francia normality test 0.7532
Anderson-Darling normality test 0.5665
Pearson chi-square normality test 0.5578

> normalidade (ratio.dif$Tempcarr)
                                valorp
One-sample Kolmogorov-Smirnov test 0.9516
Lilliefors (Kolmogorov-Smirnov) normality test 0.7726
Cramer-von Mises normality test 0.6958
Shapiro-Wilk normality test 0.3584
Shapiro-Francia normality test 0.4365
Anderson-Darling normality test 0.5435
Pearson chi-square normality test 0.4369

> normalidade (ratio.dif$DensP2019)
                                valorp
One-sample Kolmogorov-Smirnov test 0.317623
Lilliefors (Kolmogorov-Smirnov) normality test 0.024858
Cramer-von Mises normality test 0.002439
Shapiro-Wilk normality test 0.002443
Shapiro-Francia normality test 0.003790
Anderson-Darling normality test 0.002023
Pearson chi-square normality test 0.005445

normalidade (Mafra$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.9146
Lilliefors (Kolmogorov-Smirnov) normality test 0.5769
Cramer-von Mises normality test 0.5570
Shapiro-Wilk normality test 0.3851
Shapiro-Francia normality test 0.3201
Anderson-Darling normality test 0.4973
Pearson chi-square normality test 0.7024
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test
```

```

> normalidade (Mafra$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.8426
Lilliefors (Kolmogorov-Smirnov) normality test 0.4161
Cramer-von Mises normality test 0.2722
Shapiro-Wilk normality test 0.2565
Shapiro-Francia normality test 0.3686
Anderson-Darling normality test 0.3007
Pearson chi-square normality test 0.5887
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Moita$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.9504
Lilliefors (Kolmogorov-Smirnov) normality test 0.6889
Cramer-von Mises normality test 0.8240
Shapiro-Wilk normality test 0.6243
Shapiro-Francia normality test 0.8480
Anderson-Darling normality test 0.8117
Pearson chi-square normality test 0.9233
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Moita$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.1074278
Lilliefors (Kolmogorov-Smirnov) normality test 0.0005307
Cramer-von Mises normality test 0.0024464
Shapiro-Wilk normality test 0.0212647
Shapiro-Francia normality test 0.0151755
Anderson-Darling normality test 0.0053389
Pearson chi-square normality test 0.0015768
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Alcochete$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 1.369e-01
Lilliefors (Kolmogorov-Smirnov) normality test 1.175e-03
Cramer-von Mises normality test 1.723e-03
Shapiro-Wilk normality test 1.238e-02
Shapiro-Francia normality test 1.286e-02
Anderson-Darling normality test 2.795e-03
Pearson chi-square normality test 3.591e-05
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Alcochete$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test NA
Lilliefors (Kolmogorov-Smirnov) normality test 1.454e-09
Cramer-von Mises normality test 1.137e-07
Shapiro-Wilk normality test 1.287e-05
Shapiro-Francia normality test 3.748e-05
Anderson-Darling normality test 1.722e-08
Pearson chi-square normality test 1.864e-10
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Almada$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.1351693
Lilliefors (Kolmogorov-Smirnov) normality test 0.0011276
Cramer-von Mises normality test 0.0003003
Shapiro-Wilk normality test 0.0013829
Shapiro-Francia normality test 0.0044173
Anderson-Darling normality test 0.0001769

```

```

Pearson chi-square normality test          0.0011846
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Almada$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test    0.9906
Lilliefors (Kolmogorov-Smirnov) normality test 0.8854
Cramer-von Mises normality test      0.8255
Shapiro-Wilk normality test          0.5845
Shapiro-Francia normality test       0.8198
Anderson-Darling normality test      0.8146
Pearson chi-square normality test    0.8186
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Amadora$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test    0.8841
Lilliefors (Kolmogorov-Smirnov) normality test 0.5012
Cramer-von Mises normality test      0.5658
Shapiro-Wilk normality test          0.6852
Shapiro-Francia normality test       0.6759
Anderson-Darling normality test      0.5845
Pearson chi-square normality test    0.4848
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Amadora$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test    0.7770
Lilliefors (Kolmogorov-Smirnov) normality test 0.3847
Cramer-von Mises normality test      0.3143
Shapiro-Wilk normality test          0.1993
Shapiro-Francia normality test       0.3580
Anderson-Darling normality test      0.3067
Pearson chi-square normality test    0.8186

> normalidade (Cascais$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test    0.7188
Lilliefors (Kolmogorov-Smirnov) normality test 0.2368
Cramer-von Mises normality test      0.1700
Shapiro-Wilk normality test          0.1116
Shapiro-Francia normality test       0.1109
Anderson-Darling normality test      0.1409
Pearson chi-square normality test    0.1991
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Cascais$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test    0.8747
Lilliefors (Kolmogorov-Smirnov) normality test 0.5678
Cramer-von Mises normality test      0.5244
Shapiro-Wilk normality test          0.2671
Shapiro-Francia normality test       0.3619
Anderson-Darling normality test      0.4355
Pearson chi-square normality test    0.8186

> normalidade (Lisboa$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test    0.59581
Lilliefors (Kolmogorov-Smirnov) normality test 0.12907
Cramer-von Mises normality test      0.08153
Shapiro-Wilk normality test          0.04320
Shapiro-Francia normality test       0.10002
Anderson-Darling normality test      0.05441
Pearson chi-square normality test    0.19915

```

```

Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Lisboa$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.9919
Lilliefors (Kolmogorov-Smirnov) normality test 0.9409
Cramer-von Mises normality test    0.8126
Shapiro-Wilk normality test        0.4882
Shapiro-Francia normality test     0.7669
Anderson-Darling normality test    0.7639
Pearson chi-square normality test   0.7024

> normalidade (Loures$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.40390
Lilliefors (Kolmogorov-Smirnov) normality test 0.03739
Cramer-von Mises normality test    0.07969
Shapiro-Wilk normality test        0.15008
Shapiro-Francia normality test     0.23683
Anderson-Darling normality test    0.10582
Pearson chi-square normality test   0.25198
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Loures$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.9648
Lilliefors (Kolmogorov-Smirnov) normality test 0.8209
Cramer-von Mises normality test    0.4785
Shapiro-Wilk normality test        0.1936
Shapiro-Francia normality test     0.4030
Anderson-Darling normality test    0.3807
Pearson chi-square normality test   0.5887

> normalidade (Vila.Franca.de.Xira$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.6769
Lilliefors (Kolmogorov-Smirnov) normality test 0.1937
Cramer-von Mises normality test    0.2580
Shapiro-Wilk normality test        0.2884
Shapiro-Francia normality test     0.4096
Anderson-Darling normality test    0.2749
Pearson chi-square normality test   0.5887
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

>normalidade (Vila.Franca.de.Xira$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.8006
Lilliefors (Kolmogorov-Smirnov) normality test 0.4235
Cramer-von Mises normality test    0.5394
Shapiro-Wilk normality test        0.2614
Shapiro-Francia normality test     0.4603
Anderson-Darling normality test    0.4781
Pearson chi-square normality test   0.7024

> normalidade (Setúbal$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.1999281
Lilliefors (Kolmogorov-Smirnov) normality test 0.0040135
Cramer-von Mises normality test    0.0033184
Shapiro-Wilk normality test        0.0017205
Shapiro-Francia normality test     0.0043845
Anderson-Darling normality test    0.0012557
Pearson chi-square normality test   0.0002089
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

```

```

> normalidade (Setúbal$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.35716
Lilliefors (Kolmogorov-Smirnov) normality test 0.02540
Cramer-von Mises normality test 0.06637
Shapiro-Wilk normality test 0.13500
Shapiro-Francia normality test 0.21669
Anderson-Darling normality test 0.08449
Pearson chi-square normality test 0.05671
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Sintra$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.8932
Lilliefors (Kolmogorov-Smirnov) normality test 0.6107
Cramer-von Mises normality test 0.6417
Shapiro-Wilk normality test 0.9093
Shapiro-Francia normality test 0.9324
Anderson-Darling normality test 0.8031
Pearson chi-square normality test 0.5887

> normalidade (Sintra$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.41104
Lilliefors (Kolmogorov-Smirnov) normality test 0.03950
Cramer-von Mises normality test 0.12091
Shapiro-Wilk normality test 0.07325
Shapiro-Francia normality test 0.11560
Anderson-Darling normality test 0.09619
Pearson chi-square normality test 0.07352
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Barreiro$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.61380
Lilliefors (Kolmogorov-Smirnov) normality test 0.14154
Cramer-von Mises normality test 0.06169
Shapiro-Wilk normality test 0.03230
Shapiro-Francia normality test 0.02882
Anderson-Darling normality test 0.04714
Pearson chi-square normality test 0.31644
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Barreiro$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test 0.7979
Lilliefors (Kolmogorov-Smirnov) normality test 0.3406
Cramer-von Mises normality test 0.1785
Shapiro-Wilk normality test 0.3680
Shapiro-Francia normality test 0.4608
Anderson-Darling normality test 0.2378
Pearson chi-square normality test 0.1991
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Montijo$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test 0.2667801
Lilliefors (Kolmogorov-Smirnov) normality test 0.0100990
Cramer-von Mises normality test 0.0008527
Shapiro-Wilk normality test 0.0003180
Shapiro-Francia normality test 0.0008253
Anderson-Darling normality test 0.0001650
Pearson chi-square normality test 0.0008891
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

```

```

> normalidade (Montijo$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test      0.58708
Lilliefors (Kolmogorov-Smirnov) normality test 0.12336
Cramer-von Mises normality test        0.12226
Shapiro-Wilk normality test            0.42335
Shapiro-Francia normality test         0.31217
Anderson-Darling normality test        0.18441
Pearson chi-square normality test       0.02556
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> Odivelas<-subset(LSB.GR.3,LSB.GR.3$Municipio == ("Odivelas") )
> normalidade (Odivelas$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test      0.8725
Lilliefors (Kolmogorov-Smirnov) normality test 0.4756
Cramer-von Mises normality test        0.3857
Shapiro-Wilk normality test            0.2332
Shapiro-Francia normality test         0.4176
Anderson-Darling normality test        0.3267
Pearson chi-square normality test       0.5887
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Odivelas$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test      0.9911
Lilliefors (Kolmogorov-Smirnov) normality test 0.9362
Cramer-von Mises normality test        0.7830
Shapiro-Wilk normality test            0.4618
Shapiro-Francia normality test         0.7358
Anderson-Darling normality test        0.7308
Pearson chi-square normality test       0.7024
> normalidade (Oiras$Ncasos[32:53])
Error in ks.test(x, "pnorm", mean(x), sd(x)) : object 'Oiras' not found

> normalidade (Oeiras$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test      0.0729573
Lilliefors (Kolmogorov-Smirnov) normality test 0.0001473
Cramer-von Mises normality test        0.0007583
Shapiro-Wilk normality test            0.0235837
Shapiro-Francia normality test         0.0197089
Anderson-Darling normality test        0.0022378
Pearson chi-square normality test       0.0194989
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Oeiras$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test      0.8300
Lilliefors (Kolmogorov-Smirnov) normality test 0.3932
Cramer-von Mises normality test        0.2938
Shapiro-Wilk normality test            0.1995
Shapiro-Francia normality test         0.3641
Anderson-Darling normality test        0.2945
Pearson chi-square normality test       0.8186
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
  ties should not be present for the Kolmogorov-Smirnov test

> normalidade (Palmela$Ncasos[1:22])
                                valorp
One-sample Kolmogorov-Smirnov test      1.412e-02
Lilliefors (Kolmogorov-Smirnov) normality test 5.593e-07
Cramer-von Mises normality test        2.474e-05
Shapiro-Wilk normality test            1.142e-04
Shapiro-Francia normality test         3.847e-04
Anderson-Darling normality test        6.117e-06

```

```

Pearson chi-square normality test          1.652e-07
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
ties should not be present for the Kolmogorov-Smirnov test

```

```

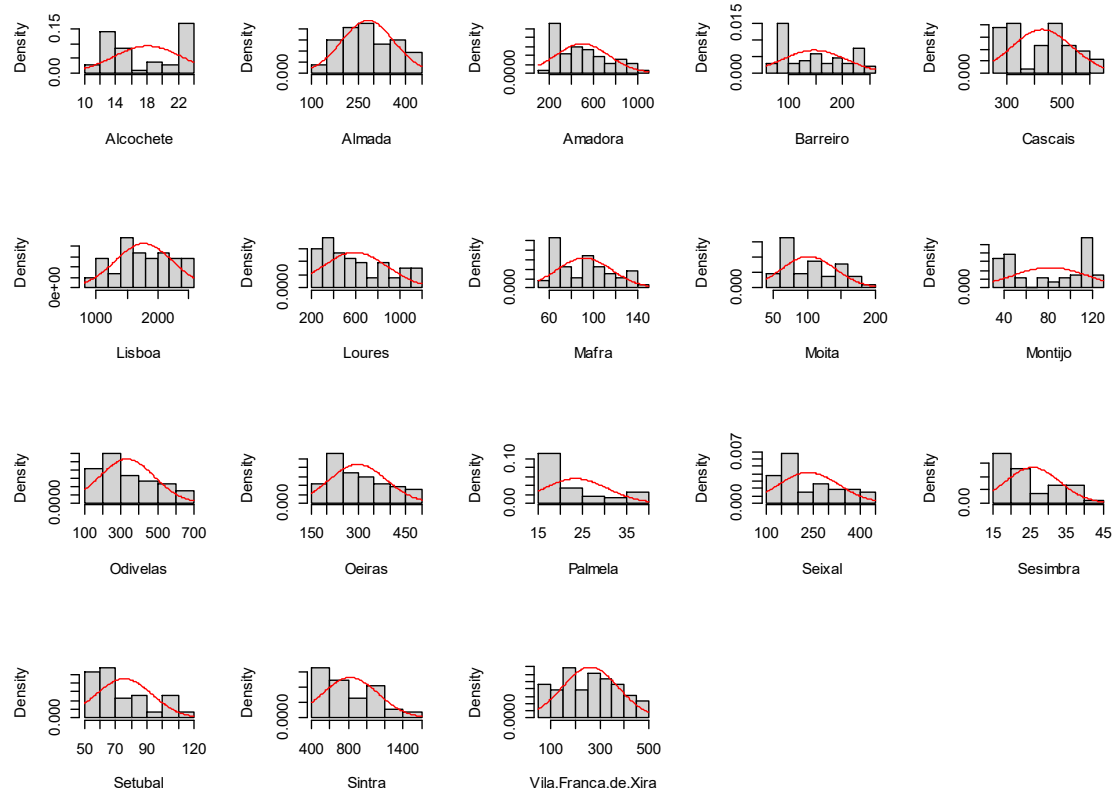
> normalidade (Palmela$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test      0.1752622
Lilliefors (Kolmogorov-Smirnov) normality test 0.0026242
Cramer-von Mises normality test         0.0005748
Shapiro-Wilk normality test             0.0030205
Shapiro-Francia normality test          0.0078607
Anderson-Darling normality test         0.0004244
Pearson chi-square normality test       0.0064736
Warning message:
In ks.test(x, "pnorm", mean(x), sd(x)) :
ties should not be present for the Kolmogorov-Smirnov test

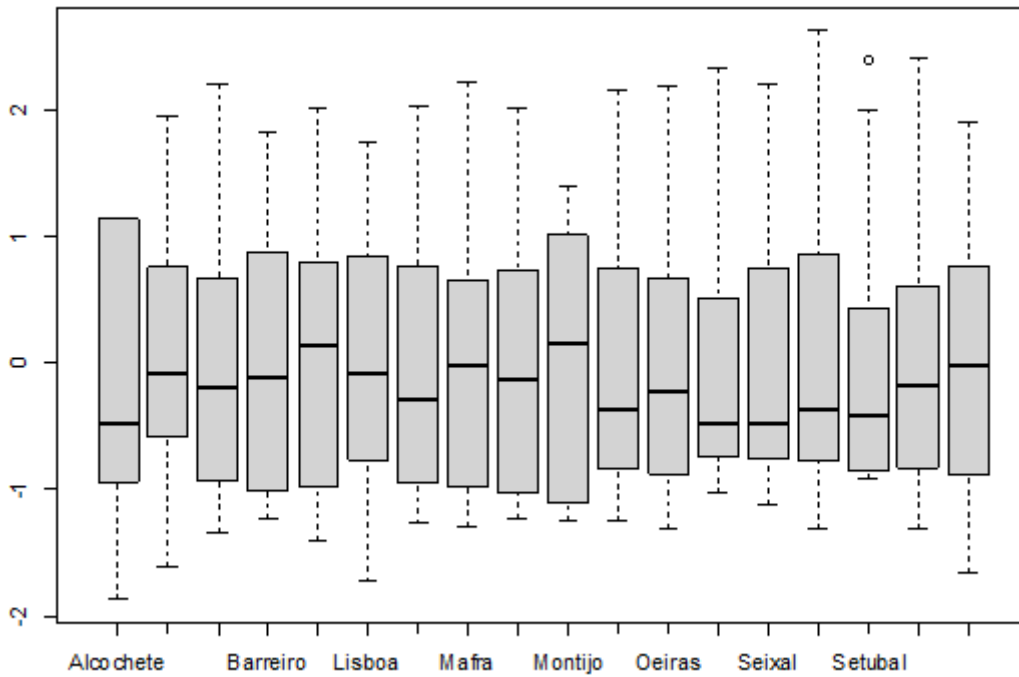
```

```

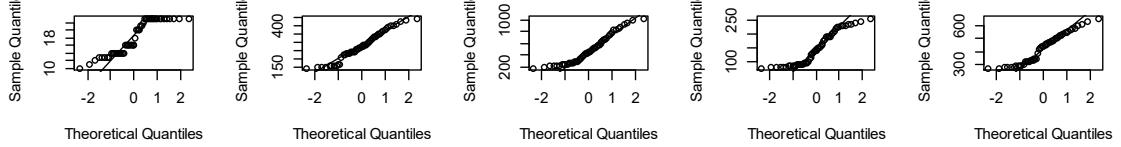
> normalidade (Seixal$Ncasos[32:53])
                                valorp
One-sample Kolmogorov-Smirnov test      0.9656
Lilliefors (Kolmogorov-Smirnov) normality test 0.8241
Cramer-von Mises normality test         0.5998
Shapiro-Wilk normality test             0.4982
Shapiro-Francia normality test          0.7210
Anderson-Darling normality test         0.6048
Pearson chi-square normality test       0.8186

```

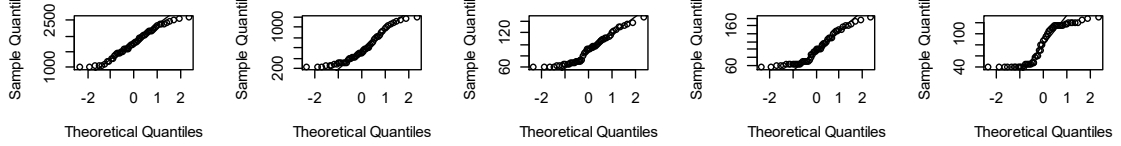




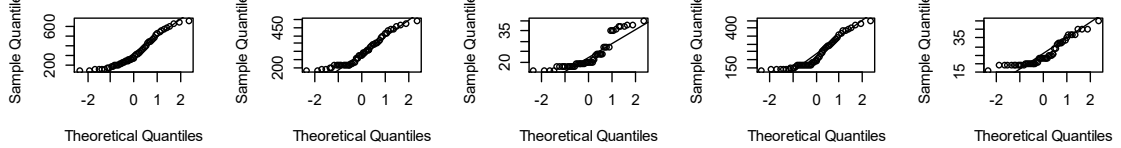
Normal Q-Q Plot (Alcoche), Normal Q-Q Plot (Almad), Normal Q-Q Plot (Amado), Normal Q-Q Plot (Barrei), Normal Q-Q Plot (Casca



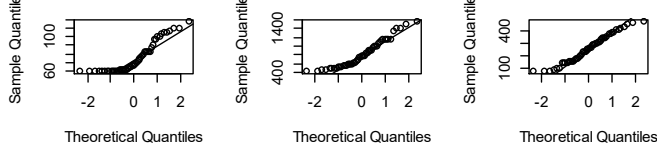
Normal Q-Q Plot (Lisboa), Normal Q-Q Plot (Loure), Normal Q-Q Plot (Mafra), Normal Q-Q Plot (Moita), Normal Q-Q Plot (Montij



Normal Q-Q Plot (Odivela), Normal Q-Q Plot (Oeira), Normal Q-Q Plot (Palme), Normal Q-Q Plot (Seixa), Normal Q-Q Plot (Sesimb



Normal Q-Q Plot (Setuba), Normal Q-Q Plot (Sintra), Normal Q-Q Plot (Vila Franca.d



ANEXO II

ESTATÍSTICAS DO NÚMERO DE CASOS COVID-19 – CONCELHOS DA AML

CAPÍTULO 3

Nº DE CASOS COVIS-19 DE 24 DE MARÇO A 6 DE JUNHO

\$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
Alcochete	53	18.07547	4.318338	16	10	23	14	23	-0.05332156	-1.6224325
Almada	53	279.64151	83.507276	273	145	442	231	343	0.09956341	-0.9750076
Amadora	53	495.03774	239.333928	448	175	1023	273	656	0.57476857	-0.8532007
Barreiro	53	147.50943	57.719688	141	76	253	89	198	0.28794109	-1.4429062
Cascais	53	427.49057	110.094125	442	273	649	320	515	0.12787777	-1.2844236
Lisboa	53	1776.66038	472.939090	1737	962	2595	1413	2177	-0.02377708	-1.1366735
Loures	53	593.69811	294.461215	507	221	1191	315	819	0.57338452	-0.9817156
Mafra	53	91.49057	25.077283	91	59	147	67	108	0.39308937	-1.0701838
Moita	53	102.01887	39.249444	97	54	181	62	131	0.42792377	-1.2140115
Montijo	53	81.60377	34.066304	87	39	129	44	116	-0.06568480	-1.7682156
Odivelas	53	333.16981	149.645959	277	146	655	208	445	0.63312162	-0.8659267
Oeiras	53	297.75472	90.467569	277	180	495	218	358	0.56154571	-0.9414373
Palmela	53	23.37736	7.152800	20	16	40	18	27	1.08577570	-0.2834024
Seixal	53	236.00000	95.585443	191	129	447	163	307	0.69441251	-0.9231944
Sesimbra	53	25.69811	7.365682	23	16	45	20	32	0.87407742	-0.4978791
Setubal	53	75.33962	17.849740	68	59	118	60	83	0.84630001	-0.6956720
Sintra	53	823.47170	305.319669	769	421	1558	568	1006	0.66686596	-0.5919811
Vila.Franca.de.Xira	53	260.86792	114.731298	258	71	479	160	349	0.12256120	-1.0800827

VARIÁVEIS DEMOGRÁFICAS E DE MOBILIDADE

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Distkmcarr2	8	23	32.42	15.13	31.30	32.44	14.68	3.60	64.60	61.00	-0.10	-0.68	3.16
IndMob	9	23	6.03	1.93	6.30	6.17	1.63	0.00	8.70	8.70	-1.07	1.76	0.40
IndEnv	10	23	140.35	42.53	126.00	134.58	34.10	94.00	295.00	201.00	1.99	4.90	8.87
DenPop	11	23	1444.11	1963.20	503.90	1027.31	672.05	50.61	7598.00	7547.39	1.83	2.63	409.35
Tempcarr	12	23	35.96	10.92	37.00	36.47	11.86	13.00	50.00	37.00	-0.35	-1.02	2.28
MeioTransp	13	23	1.78	0.67	2.00	1.74	0.00	1.00	3.00	2.00	0.24	-0.94	0.14
Densp2019	14	23	1620.94	2117.46	501.20	1226.58	529.44	86.00	7692.00	7606.00	1.52	1.24	441.52
Pop	15	23	131918.39	120964.50	84816.00	110902.26	92069.46	10651.00	509515.00	498864.00	1.57	2.25	25222.84
MedMensa1	16	23	1099.63	244.21	1080.20	1062.30	211.12	846.30	1791.50	945.20	1.21	1.02	50.92

ANEXO III

TESTE À SOBREDISPERSÃO DOS MODELOS DE POISSON – dados da AML, Resto do país e concelhos da AML

CASO “MANIFESTAÇÃO 1.º DE MAIO”

CAPÍTULO 3

```
library(performance)
> check_overdispersion(M3pi_Alcochete)
# Overdispersion test

      dispersion ratio = 0.027
Pearson's Chi-Squared = 1.073
              p-value =      1

No overdispersion detected.
> check_overdispersion(M3pi_Almada)
# Overdispersion test

      dispersion ratio = 0.473
Pearson's Chi-Squared = 18.935
              p-value = 0.998

No overdispersion detected.
> check_overdispersion(M3pi_Amadora)
# Overdispersion test

      dispersion ratio = 0.602
Pearson's Chi-Squared = 24.070
              p-value = 0.978

No overdispersion detected.
> check_overdispersion(M3pi_Barreiro)
# Overdispersion test

      dispersion ratio = 0.132
Pearson's Chi-Squared = 5.267
              p-value =      1

No overdispersion detected.
> check_overdispersion(M3pi_Cascais)
# Overdispersion test

      dispersion ratio = 0.206
Pearson's Chi-Squared = 8.224
              p-value =      1

No overdispersion detected.
> check_overdispersion(M3pi_Lisboa)
# Overdispersion test

      dispersion ratio = 0.579
Pearson's Chi-Squared = 23.155
              p-value = 0.985

No overdispersion detected.
> check_overdispersion(M3pi_Loures)
# Overdispersion test

      dispersion ratio = 0.376
Pearson's Chi-Squared = 15.051
              p-value =      1

No overdispersion detected.
> check_overdispersion(M3pi_Oeiras)
# Overdispersion test

      dispersion ratio = 0.246
Pearson's Chi-Squared = 9.824
              p-value =      1
```

```

No overdispersion detected.
> check_overdispersion(M3pi_Odivelas)
# Overdispersion test

      dispersion ratio = 0.089
Pearson's Chi-Squared = 3.576
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Mafra)
# Overdispersion test

      dispersion ratio = 0.068
Pearson's Chi-Squared = 2.727
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Moita)
# Overdispersion test

      dispersion ratio = 0.066
Pearson's Chi-Squared = 2.624
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Montijo)
# Overdispersion test

      dispersion ratio = 0.135
Pearson's Chi-Squared = 5.383
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Sesimbra)
# Overdispersion test

      dispersion ratio = 0.037
Pearson's Chi-Squared = 1.474
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Palmela)
# Overdispersion test

      dispersion ratio = 0.094
Pearson's Chi-Squared = 3.747
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Sintra)
# Overdispersion test

      dispersion ratio = 1.359
Pearson's Chi-Squared = 54.365
                p-value = 0.064

No overdispersion detected.
> check_overdispersion(M3pi_Setubal)
# Overdispersion test

      dispersion ratio = 0.043
Pearson's Chi-Squared = 1.722
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Vila.Franca.de.Xira)
# Overdispersion test

      dispersion ratio = 0.479
Pearson's Chi-Squared = 19.141
                p-value = 0.998

No overdispersion detected.

```

```

> check_overdispersion(M3pl_Vila.Franca.de.Xira)
# Overdispersion test

      dispersion ratio = 0.778
Pearson's Chi-Squared = 31.116
      p-value = 0.842

No overdispersion detected.
> check_overdispersion(M3pl_Setubal)
# Overdispersion test

      dispersion ratio = 0.040
Pearson's Chi-Squared = 1.589
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Palmela)
# Overdispersion test

      dispersion ratio = 0.090
Pearson's Chi-Squared = 3.582
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Sintra)
# Overdispersion test

      dispersion ratio = 0.961
Pearson's Chi-Squared = 38.421
      p-value = 0.541

No overdispersion detected.
> check_overdispersion(M3pl_Sesimbra)
# Overdispersion test

      dispersion ratio = 0.037
Pearson's Chi-Squared = 1.494
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Montijo)
# Overdispersion test

      dispersion ratio = 0.117
Pearson's Chi-Squared = 4.681
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Moita)
# Overdispersion test

      dispersion ratio = 0.066
Pearson's Chi-Squared = 2.630
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Mafra)
# Overdispersion test

      dispersion ratio = 0.055
Pearson's Chi-Squared = 2.194
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Odivelas)
# Overdispersion test

      dispersion ratio = 0.105
Pearson's Chi-Squared = 4.182
      p-value = 1

```

```

No overdispersion detected.
> check_overdispersion(M3pl_Oeiras)
# Overdispersion test

      dispersion ratio = 0.213
Pearson's Chi-Squared = 8.538
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Loures)
# Overdispersion test

      dispersion ratio = 0.487
Pearson's Chi-Squared = 19.495
      p-value = 0.997

No overdispersion detected.
> check_overdispersion(M3pl_Lisboa)
# Overdispersion test

      dispersion ratio = 0.920
Pearson's Chi-Squared = 36.812
      p-value = 0.615

No overdispersion detected.
> check_overdispersion(M3pl_Cascais)
# Overdispersion test

      dispersion ratio = 0.157
Pearson's Chi-Squared = 6.266
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Barreiro)
# Overdispersion test

      dispersion ratio = 0.145
Pearson's Chi-Squared = 5.781
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Amadora)
# Overdispersion test

      dispersion ratio = 0.246
Pearson's Chi-Squared = 9.836
      p-value = 1

No overdispersion detected.
> check_overdispersion(M3pl_Almada)
# Overdispersion test

      dispersion ratio = 0.507
Pearson's Chi-Squared = 20.278
      p-value = 0.996

No overdispersion detected.
> check_overdispersion(M3pl_Alcochete)
# Overdispersion test

      dispersion ratio = 0.028
Pearson's Chi-Squared = 1.115
      p-value = 1

No overdispersion detected.

No overdispersion detected.
> check_overdispersion(M3pl_AML)
# Overdispersion test

      dispersion ratio = 0.637
Pearson's Chi-Squared = 25.481
      p-value = 0.964

```

```

No overdispersion detected.
> check_overdispersion(M3pl_Resto)
# Overdispersion test

      dispersion ratio = 4.064
Pearson's Chi-Squared = 162.565
              p-value = < 0.001

Overdispersion detected.

> check_overdispersion(M3pi_AML)
# Overdispersion test

      dispersion ratio = 0.977
Pearson's Chi-Squared = 39.091
              p-value = 0.511

No overdispersion detected.
> check_overdispersion(M3pi_Resto)
# Overdispersion test

      dispersion ratio = 37.215
Pearson's Chi-Squared = 1897.942
              p-value = < 0.001

Overdispersion detected.

# LRT para comparar o ajuste dos modelos Poisson e BN
pchisq(2 * (logLik(M3bn_Resto) - logLik(M3pl_Resto)), df = 1, lower.tail = FALSE)
'log Lik.' 5.344719e-14 (df=5)
# NB é melhor

```

ANEXO IV

TESTE À SOBREDISPERSÃO DOS MODELOS DE POISSON DO CASO “MANIFESTAÇÃO ANTIRRACISMO” – dados da AML, Resto do país e concelhos da AML

CAPÍTULO 3

```
> library(performance)
> check_overdispersion(M3pi_Alcochete.r)
# Overdispersion test

      dispersion ratio = 0.008
Pearson's Chi-Squared = 0.306
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Almada.r)
# Overdispersion test

      dispersion ratio = 0.101
Pearson's Chi-Squared = 4.052
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Amadora.r)
# Overdispersion test

      dispersion ratio = 0.814
Pearson's Chi-Squared = 32.547
                p-value = 0.793

No overdispersion detected.
> check_overdispersion(M3pi_Barreiro.r)
# Overdispersion test

      dispersion ratio = 0.123
Pearson's Chi-Squared = 4.926
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Cascais.r)
# Overdispersion test

      dispersion ratio = 0.566
Pearson's Chi-Squared = 22.651
                p-value = 0.988

No overdispersion detected.
> check_overdispersion(M3pi_Lisboa.r)
# Overdispersion test

      dispersion ratio = 0.635
Pearson's Chi-Squared = 25.382
                p-value = 0.965

No overdispersion detected.
> check_overdispersion(M3pi Loures.r)
# Overdispersion test

      dispersion ratio = 0.409
Pearson's Chi-Squared = 16.372
                p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Mafra.r)
# Overdispersion test

      dispersion ratio = 0.137
Pearson's Chi-Squared = 5.494
                p-value = 1
```

```

No overdispersion detected.
> check_overdispersion(M3pi_Moita.r)
# Overdispersion test

    dispersion ratio = 0.299
Pearson's Chi-Squared = 11.966
    p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Montijo.r)
# Overdispersion test

    dispersion ratio = 0.037
Pearson's Chi-Squared = 1.460
    p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Palmela.r)
# Overdispersion test

    dispersion ratio = 0.137
Pearson's Chi-Squared = 5.474
    p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Odivelas.r)
# Overdispersion test

    dispersion ratio = 0.740
Pearson's Chi-Squared = 29.589
    p-value = 0.886

No overdispersion detected.
> check_overdispersion(M3pi_Oeiras.r)
# Overdispersion test

    dispersion ratio = 0.538
Pearson's Chi-Squared = 21.509
    p-value = 0.993

No overdispersion detected.
> check_overdispersion(M3pi_Sesimbra.r)
# Overdispersion test

    dispersion ratio = 0.057
Pearson's Chi-Squared = 2.267
    p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Seixal.r)
# Overdispersion test

    dispersion ratio = 0.137
Pearson's Chi-Squared = 5.465
    p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Setubal.r)
# Overdispersion test

    dispersion ratio = 0.160
Pearson's Chi-Squared = 6.388
    p-value = 1

No overdispersion detected.
> check_overdispersion(M3pi_Sintra.r)
# Overdispersion test

    dispersion ratio = 2.437
Pearson's Chi-Squared = 97.461
    p-value = < 0.001

Overdispersion detected.

```

```

> check_overdispersion(M3pi_Vila.Franca.de.Xira.r)
# Overdispersion test

      dispersion ratio = 0.346
Pearson's Chi-Squared = 13.822
              p-value = 1

No overdispersion detected.

> check_overdispersion(M3pi.resto.r)
# Overdispersion test

      dispersion ratio = 0.518
Pearson's Chi-Squared = 20.710
              p-value = 0.995

> check_overdispersion(M3pl.resto.r)
# Overdispersion test

      dispersion ratio = 0.527
Pearson's Chi-Squared = 21.071
              p-value = 0.994

No overdispersion detected.
> check_overdispersion(M3pi.aml.r)
# Overdispersion test

      dispersion ratio = 5.182
Pearson's Chi-Squared = 207.271
              p-value = < 0.001

Overdispersion detected.
> check_overdispersion(M3pl.aml.r)
# Overdispersion test

      dispersion ratio = 5.576
Pearson's Chi-Squared = 223.020
              p-value = < 0.001

Overdispersion detected.

> lrtest(M3pl.resto.r,M3.pl.nb.resto.r)
Likelihood ratio test

Model 1: Resto ~ DuMMY * t
Model 2: Resto ~ DuMMY * t
#Df  LogLik Df  Chisq Pr(>Chisq)
1    4 -270.17
2    5 -270.17  1 6e-04    0.9805

> lrtest(M3pl.aml.r,M3pl.nb.aml.r)
Likelihood ratio test

Model 1: (AML) ~ DuMMY * t
Model 2: (AML) ~ DuMMY * t
#Df  LogLik Df  Chisq Pr(>Chisq)
1    4 -358.74
2    5 -300.30  1 116.88 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

ANEXO V

CAPÍTULO 5

Modelos INLA ajustados aos dados nos concelhos do distrito do Porto entre 25 de março de 2020 e 30 de junho de 2020, observados, agregados a 7 dias e acumulados nos últimos 14 dias, com e sem covariáveis, com e sem efeitos espaciais estruturados e não estruturados.

```
Modelo: M0: Diario.Unif~1+
      f(Region_str, model="bym", graph=W.porto)+
      f(Date_str, model="rw2") +
      f(Date_unstr, model="iid")+
      f(region_date, model="iid")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: M0

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	0.2321331	[-0.592834, 1.056452]	Efeito Fixo
Precision for Region_str (iid component)	0.3795732	[0.1859124, 0.6691271]	Hiperparâmetro
Precision for Region_str (spatial component)	2197.128	[163.2871208, 8431.9013665]	Hiperparâmetro
Precision for Date_str	17064	[1870.4369681, 58852.0511946]	Hiperparâmetro
Precision for Date_unstr	21861.27	[1596.1189356, 83359.7521753]	Hiperparâmetro
Precision for region_date	0.8947296	[0.6207064, 1.1964225]	Hiperparâmetro

```
Modelo: M0.HF: x.HF~1+
      f(Region_str, model="bym", graph=W.porto)+
      f(Date_str, model="rw2") +
      f(Date_unstr, model="iid")+
      f(region_date, model="iid")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: M0.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	5.523886e-01	[-0.2610365, 1.3654058]	Efeito Fixo
Precision for Region_str (iid component)	3.782713e-01	[0.1856337, 0.6681177]	Hiperparâmetro
Precision for Region_str (spatial component)	2.179037e+03	[159.7533307, 8391.4120269]	Hiperparâmetro
Precision for Date_str	1.247266e+04	[1747.6595178, 43643.6397358]	Hiperparâmetro
Precision for Date_unstr	1.596900e+00	[1.1503881, 2.1346719]	Hiperparâmetro
Precision for region_date	2.160368e+04	[1588.6579548, 85571.3517672]	Hiperparâmetro

```
Modelo: M0.LS: x.LS~1+
      f(Region_str, model="bym", graph=W.porto)+
      f(Date_str, model="rw2") +
      f(Date_unstr, model="iid")+
      f(region_date, model="iid")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: M0.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	5.130324e-01	[-0.3158834, 1.3415759]	Efeito Fixo
Precision for Region_str (iid component)	3.710342e-01	[0.1834439, 0.6617601]	Hiperparâmetro
Precision for Region_str (spatial component)	2.438218e+03	[204.4294255, 9598.8958444]	Hiperparâmetro
Precision for Date_str	1.197174e+04	[1529.3970049, 41643.7592744]	Hiperparâmetro
Precision for Date_unstr	1.449943e+00	[1.0459793, 1.9395729]	Hiperparâmetro
Precision for region_date	2.356614e+04	[1805.1437756, 90921.3852998]	Hiperparâmetro

```
Modelo: MG: Diario.Unif~tm.min.2+tm.min.2+hM.min.2+hm.min.2+pr.min.2+tm.max.2+tm.max.2+hM.max.2+hm.max.2+pr.max.2+tm.mean.
2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+tm.mean.14+tm.mean.14+hM.mean.
14+hm.mean.14+pr.mean.14+Popdensity+MensalIncome+IDI.2019+ Mais.65+
```

```
      f(Region_unstr, model="iid")+
      f(Region_str, model="bym", graph=W.porto) +
      f(Date_unstr, model="iid")+
      f(Date_str, model="rw2")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: MG

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	20.0674395	[14.0657982, 26.1851519]	Efeito Fixo
tm.min.2	-0.6077627	[-0.7402757, -0.4753943]	Efeito Fixo
tm.min.2	0.0193039	[-0.1399827, 0.1786053]	Efeito Fixo
hm.min.2	-0.0528879	[-0.101511, -0.0042623]	Efeito Fixo
hm.min.2	-0.1186131	[-0.1449434, -0.092295]	Efeito Fixo
pr.min.2	-0.0550349	[-0.1089436, -0.0011287]	Efeito Fixo
tm.max.2	-0.5124169	[-0.6621373, -0.3629359]	Efeito Fixo
tm.max.2	-0.1823167	[-0.4090278, 0.0445175]	Efeito Fixo
hm.max.2	-0.0800342	[-0.1180503, -0.0420126]	Efeito Fixo
hm.max.2	-0.2528351	[-0.2848337, -0.2208537]	Efeito Fixo
pr.max.2	-0.0828984	[-0.1332795, -0.0325263]	Efeito Fixo
tm.mean.2	0.8446290	[0.5945114, 1.0949064]	Efeito Fixo
tm.mean.2	0.5908855	[0.2511113, 0.9306487]	Efeito Fixo
hm.mean.2	0.1215732	[0.04247, 0.2006719]	Efeito Fixo
hm.mean.2	0.2760621	[0.2265127, 0.325616]	Efeito Fixo
pr.mean.2	0.0646690	[-0.0279539, 0.1572964]	Efeito Fixo

tm.mean.7		0.1343863		[0.0856017 , 0.1832432]		Efeito Fixo	
tm.mean.7		-0.2747255		[-0.3546376 , -0.1956317]		Efeito Fixo	
hm.mean.7		-0.0572275		[-0.0749276 , -0.0396943]		Efeito Fixo	
hm.mean.7		0.0110756		[-0.0013754 , 0.023514]		Efeito Fixo	
pr.mean.7		0.0705737		[0.0564869 , 0.0846917]		Efeito Fixo	
tm.mean.14		0.0016613		[-0.0483832 , 0.0515431]		Efeito Fixo	
tm.mean.14		-0.4343962		[-0.5057505 , -0.3638337]		Efeito Fixo	
hm.mean.14		-0.0071147		[-0.0237936 , 0.0093394]		Efeito Fixo	
hm.mean.14		-0.0001986		[-0.0134042 , 0.0129383]		Efeito Fixo	
pr.mean.14		0.0699750		[0.0557971 , 0.0841633]		Efeito Fixo	
Popdensity		0.0013310		[-0.0002714 , 0.0029353]		Efeito Fixo	
MensalIncomeTRUE		0.0000000		[-62.0086677 , 62.0086677]		Efeito Fixo	
IDI.2019		-0.1673279		[-0.3782338 , 0.0434938]		Efeito Fixo	
Mais.65		-0.0000089		[-0.0001088 , 9.1e-05]		Efeito Fixo	
Precision for Region_unstr		26415.4631821		[2421.8729026 , 102574.2601681]		Hiperparâmetro	
Precision for Region_str (iid component)		0.3191632		[0.1451562 , 0.5866969]		Hiperparâmetro	
Precision for Region_str (spatial component)		2569.6699132		[220.6291752 , 9862.1023445]		Hiperparâmetro	
Precision for Date_unstr		0.2221117		[0.137953 , 0.2834119]		Hiperparâmetro	
Precision for Date_str		4330.4595345		[590.6880231 , 15609.3803595]		Hiperparâmetro	

Modelo: MG.HF: x.HF~1+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: MG.HF

Variável	Média	Intervalo de Credibilidade	Tipo

(Intercept)	8.1265825	[2.8840537 , 13.4054622]	Efeito Fixo
tm.min.2	-0.7082173	[-0.8429684 , -0.5735008]	Efeito Fixo
tm.min.2	0.0525315	[-0.114098 , 0.2191808]	Efeito Fixo
hm.min.2	-0.0089165	[-0.0560948 , 0.0382585]	Efeito Fixo
hm.min.2	-0.0272097	[-0.0517132 , -0.0027161]	Efeito Fixo
pr.min.2	-0.0551260	[-0.1096896 , -0.0005649]	Efeito Fixo
tm.max.2	-0.4846344	[-0.6298946 , -0.3394918]	Efeito Fixo
tm.max.2	-0.1138240	[-0.3310124 , 0.1034098]	Efeito Fixo
hm.max.2	-0.1017289	[-0.1377777 , -0.06568]	Efeito Fixo
hm.max.2	-0.0916816	[-0.1207671 , -0.0626178]	Efeito Fixo
pr.max.2	-0.0288492	[-0.0801756 , 0.0224557]	Efeito Fixo
tm.mean.2	1.1148369	[0.8624227 , 1.3672741]	Efeito Fixo
tm.mean.2	0.2731067	[-0.0694209 , 0.6156208]	Efeito Fixo
hm.mean.2	0.0978767	[0.0219201 , 0.1738329]	Efeito Fixo
hm.mean.2	0.0734509	[0.0279091 , 0.1189978]	Efeito Fixo
pr.mean.2	0.0455933	[-0.0483663 , 0.1395566]	Efeito Fixo
tm.mean.7	0.1311329	[0.0851295 , 0.1773716]	Efeito Fixo
tm.mean.7	-0.1470977	[-0.2137149 , -0.0810185]	Efeito Fixo
hm.mean.7	-0.0311394	[-0.0459555 , -0.0164086]	Efeito Fixo
hm.mean.7	0.0221682	[0.0110157 , 0.0333381]	Efeito Fixo
pr.mean.7	0.0234850	[0.0108488 , 0.0361501]	Efeito Fixo
tm.mean.14	0.0631374	[0.0209457 , 0.1053307]	Efeito Fixo
tm.mean.14	-0.1966007	[-0.2562057 , -0.1373736]	Efeito Fixo
hm.mean.14	-0.0143831	[-0.0270757 , -0.0017292]	Efeito Fixo
hm.mean.14	0.0362694	[0.0255703 , 0.0469893]	Efeito Fixo
pr.mean.14	0.0377632	[0.0238959 , 0.0516787]	Efeito Fixo
Popdensity	0.0011181	[-0.000387 , 0.0026237]	Efeito Fixo
MensalIncomeTRUE	0.0000000	[-62.0086677 , 62.0086677]	Efeito Fixo
IDI.2019	-0.1757490	[-0.3737828 , 0.022228]	Efeito Fixo
Mais.65	-0.0000048	[-9.86e-05 , 8.9e-05]	Efeito Fixo
Precision for Region_unstr	21726.3513617	[1546.3645987 , 81756.8234725]	Hiperparâmetro
Precision for Region_str (iid component)	0.3629750	[0.1657261 , 0.6686998]	Hiperparâmetro
Precision for Region_str (spatial component)	2612.9148392	[246.0471935 , 9995.8863361]	Hiperparâmetro
Precision for Date_unstr	0.5980060	[0.4204387 , 0.8313544]	Hiperparâmetro
Precision for Date_str	11985.3119856	[1323.3299389 , 48090.4040729]	Hiperparâmetro

Modelo: MG.LS: x.LS~1+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: MG.LS

Variável	Média	Intervalo de Credibilidade	Tipo

(Intercept)	8.3094618	[2.9537158 , 13.6781808]	Efeito Fixo
tm.min.2	-0.6718280	[-0.8074039 , -0.5362866]	Efeito Fixo
tm.min.2	0.0420544	[-0.1245685 , 0.2086945]	Efeito Fixo
hm.min.2	-0.0248910	[-0.0724259 , 0.02264]	Efeito Fixo
hm.min.2	-0.0121526	[-0.0371046 , 0.0127905]	Efeito Fixo
pr.min.2	-0.0474163	[-0.1026397 , 0.0078052]	Efeito Fixo
tm.max.2	-0.4450051	[-0.5921191 , -0.2980136]	Efeito Fixo
tm.max.2	-0.0910204	[-0.3084926 , 0.1265039]	Efeito Fixo
hm.max.2	-0.1175154	[-0.1539182 , -0.0811128]	Efeito Fixo
hm.max.2	-0.0973716	[-0.1272285 , -0.0675369]	Efeito Fixo
pr.max.2	-0.0380749	[-0.0895379 , 0.0133694]	Efeito Fixo
tm.mean.2	1.0294382	[0.7748475 , 1.2840551]	Efeito Fixo
tm.mean.2	0.3155756	[-0.0264531 , 0.6575948]	Efeito Fixo
hm.mean.2	0.1248155	[0.0481679 , 0.2014619]	Efeito Fixo
hm.mean.2	0.0638631	[0.0173237 , 0.1104073]	Efeito Fixo
pr.mean.2	0.0418511	[-0.0528871 , 0.1365918]	Efeito Fixo
tm.mean.7	0.1584584	[0.1119647 , 0.2051784]	Efeito Fixo
tm.mean.7	-0.1654159	[-0.232897 , -0.0984773]	Efeito Fixo
hm.mean.7	-0.0371054	[-0.0522054 , -0.0221103]	Efeito Fixo

hm.mean.7		0.0271684		[0.0160237 , 0.0383324]		Efeito Fixo	
pr.mean.7		0.0317285		[0.0190083 , 0.0444771]		Efeito Fixo	
tm.mean.14		0.0900644		[0.0470163 , 0.1331475]		Efeito Fixo	
tm.mean.14		-0.2175530		[-0.2784414 , -0.1570406]		Efeito Fixo	
hm.mean.14		-0.0186910		[-0.0315753 , -0.0058577]		Efeito Fixo	
hm.mean.14		0.0354037		[0.0245663 , 0.0462568]		Efeito Fixo	
pr.mean.14		0.0418546		[0.0278646 , 0.0558874]		Efeito Fixo	
Popdensity		0.0011300		[-0.0003968 , 0.0026568]		Efeito Fixo	
MensualIncomeTRUE		0.0000000		[-62.0086677 , 62.0086677]		Efeito Fixo	
IDI.2019		-0.1805212		[-0.3813601 , 0.0202977]		Efeito Fixo	
Mais.65		-0.0000044		[-9.95e-05 , 9.07e-05]		Efeito Fixo	
Precision for Region_unstr		27484.3519933		[2443.1949196 , 109091.9135091]		Hiperparâmetro	
Precision for Region_str (iid component)		0.3579412		[0.1647288 , 0.6656365]		Hiperparâmetro	
Precision for Region_str (spatial component)		2630.2344476		[232.0920542 , 10443.4322128]		Hiperparâmetro	
Precision for Date_unstr		0.5354725		[0.3754261 , 0.7317654]		Hiperparâmetro	
Precision for Date_str		9614.3586013		[1082.2973494 , 37696.6442449]		Hiperparâmetro	

Modelo: RW.M1.2: Diário.Unif~tm.mean.2+tm.mean.2+hM.mean.2+hM.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+ f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW.M1.2

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	1.1305861	[-0.6564885 , 2.7720427]	Efeito Fixo
tm.mean.2	-0.0023953	[-0.0328153 , 0.0276845]	Efeito Fixo
tm.mean.2	0.4474970	[0.3886528 , 0.5095723]	Efeito Fixo
hM.mean.2	-0.0187931	[-0.0331746 , -0.0030248]	Efeito Fixo
hm.mean.2	0.0482365	[0.040062 , 0.0564319]	Efeito Fixo
pr.mean.2	-0.0253384	[-0.0386192 , -0.0121058]	Efeito Fixo
Popdensity	0.0009649	[0.0009173 , 0.0010125]	Efeito Fixo
GMM.T.2018	-0.0044173	[-0.0048606 , -0.0039768]	Efeito Fixo
IDI.2019	-0.1129992	[-0.1200595 , -0.1059309]	Efeito Fixo
Mais.65	-0.0000130	[-1.46e-05 , -1.14e-05]	Efeito Fixo
Precision for Date_str	0.0811038	[0.0474985 , 0.1270958]	Hiperparâmetro

Modelo: RW.M1.2.HF: x.HF~tm.mean.2+tm.mean.2+hM.mean.2+hM.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+ f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW.M1.2.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.6768754	[3.7935222 , 5.5365461]	Efeito Fixo
tm.mean.2	0.0177495	[-0.0055448 , 0.0408378]	Efeito Fixo
tm.mean.2	0.2715656	[0.2319478 , 0.3122449]	Efeito Fixo
hM.mean.2	-0.0417610	[-0.0491909 , -0.0342064]	Efeito Fixo
hm.mean.2	0.0306344	[0.0233128 , 0.0380328]	Efeito Fixo
pr.mean.2	0.0026625	[-0.0083458 , 0.0136625]	Efeito Fixo
Popdensity	0.0008874	[0.0008403 , 0.0009345]	Efeito Fixo
GMM.T.2018	-0.0030654	[-0.0034668 , -0.0026668]	Efeito Fixo
IDI.2019	-0.1155991	[-0.1226086 , -0.1085892]	Efeito Fixo
Mais.65	-0.0000143	[-1.58e-05 , -1.27e-05]	Efeito Fixo
Precision for Date_str	0.5685050	[0.3706743 , 0.8171389]	Hiperparâmetro

Modelo: RW.M1.2.LS: x.LS~tm.mean.2+tm.mean.2+hM.mean.2+hM.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+ f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW.M1.2.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.4843023	[3.4967608 , 5.4225728]	Efeito Fixo
tm.mean.2	0.0119214	[-0.0131491 , 0.036684]	Efeito Fixo
tm.mean.2	0.3004214	[0.2585551 , 0.3438385]	Efeito Fixo
hM.mean.2	-0.0406380	[-0.0486955 , -0.0322322]	Efeito Fixo
hm.mean.2	0.0323331	[0.0246624 , 0.0400918]	Efeito Fixo
pr.mean.2	-0.0032301	[-0.0145698 , 0.0080789]	Efeito Fixo
Popdensity	0.0009084	[0.0008609 , 0.0009559]	Efeito Fixo
GMM.T.2018	-0.0032310	[-0.003642 , -0.0028233]	Efeito Fixo
IDI.2019	-0.1179908	[-0.1250597 , -0.1109214]	Efeito Fixo
Mais.65	-0.0000142	[-1.57e-05 , -1.26e-05]	Efeito Fixo
Precision for Date_str	0.4338456	[0.2744617 , 0.6342132]	Hiperparâmetro

Modelo: RW.M1.2.HF.14: x.HF.14~tm.mean.2+tm.mean.2+hM.mean.2+hM.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65 + f(Date_str, model="rw2")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: RW.M1.2.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.2396715	[3.3645822 , 5.0861651]	Efeito Fixo
tm.mean.2	0.0351004	[0.0124581 , 0.0577276]	Efeito Fixo
tm.mean.2	0.2733251	[0.2343837 , 0.3133029]	Efeito Fixo
hM.mean.2	-0.0394197	[-0.0466650 , -0.0320501]	Efeito Fixo
hm.mean.2	0.0282293	[0.0209949 , 0.0355430]	Efeito Fixo
pr.mean.2	0.0014120	[-0.0093839 , 0.0121748]	Efeito Fixo

	Popdensity		0.0008787		[0.0008318 , 0.0009256]		Efeito Fixo	
	GMM.T.2018		-0.0029835		[-0.0033809 , -0.0025890]		Efeito Fixo	
	IDI.2019		-0.1146516		[-0.1216191 , -0.1076838]		Efeito Fixo	
	Mais.65		-0.0000147		[-0.0000162 , -0.0000131]		Efeito Fixo	
	Precision for Date_str		0.603936		[0.3932824 , 0.8669537]		Hiperparâmetro	

Modelo: RW.M1.2.LS.14: x.LS.14~tm.mean.2+tm.mean.2+hm.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65
+
f(Date_str, model="rw2")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: RW.M1.2.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.9099429	[1.7207301 , 3.9746052]	Efeito Fixo
tm.mean.2	0.0333950	[0.0070769 , 0.0593743]	Efeito Fixo
tm.mean.2	0.3019979	[0.2575102 , 0.3496601]	Efeito Fixo
hm.mean.2	-0.0304410	[-0.0393960 , -0.0204384]	Efeito Fixo
hm.mean.2	0.0342992	[0.0266306 , 0.0420575]	Efeito Fixo
pr.mean.2	-0.0045177	[-0.0160639 , 0.0069697]	Efeito Fixo
Popdensity	0.0008762	[0.0008293 , 0.0009232]	Efeito Fixo
GMM.T.2018	-0.0030278	[-0.0034431 , -0.0026178]	Efeito Fixo
IDI.2019	-0.1137540	[-0.1206887 , -0.1068181]	Efeito Fixo
Mais.65	-0.0000144	[-0.0000159 , -0.0000128]	Efeito Fixo
Precision for Date_str	0.3350259	[0.1968408 , 0.5105564]	Hiperparâmetro

Modelo: RW.M1.7 : Diario.Unif~tm.mean.7+tm.mean.7+hm.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+
f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW.M1.7

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	5.1542894	[3.8869043 , 6.3460515]	Efeito Fixo
tm.mean.7	-0.0787228	[-0.1173759 , -0.0400873]	Efeito Fixo
tm.mean.7	0.3756059	[0.3207021 , 0.4311278]	Efeito Fixo
hm.mean.7	-0.0414860	[-0.0540814 , -0.028588]	Efeito Fixo
hm.mean.7	0.0509346	[0.041577 , 0.060325]	Efeito Fixo
pr.mean.7	0.0002118	[-0.0112695 , 0.0117231]	Efeito Fixo
Popdensity	0.0009442	[0.0008968 , 0.0009917]	Efeito Fixo
GMM.T.2018	-0.0042392	[-0.0046706 , -0.0038093]	Efeito Fixo
IDI.2019	-0.1114466	[-0.1185282 , -0.1043647]	Efeito Fixo
Mais.65	-0.0000128	[-1.44e-05 , -1.12e-05]	Efeito Fixo
Precision for Date_str	0.1124939	[0.0752708 , 0.158779]	Hiperparâmetro

Modelo: RW.M1.7.HF : x.HF~tm.mean.7+tm.mean.7+hm.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+
f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW.M1.7.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.2208564	[3.254184 , 5.1027541]	Efeito Fixo
tm.mean.7	-0.0123025	[-0.0474694 , 0.0227934]	Efeito Fixo
tm.mean.7	0.3119338	[0.2619347 , 0.3631211]	Efeito Fixo
hm.mean.7	-0.0321397	[-0.0427176 , -0.0212361]	Efeito Fixo
hm.mean.7	0.0283014	[0.0195635 , 0.0371158]	Efeito Fixo
pr.mean.7	-0.0208782	[-0.0314608 , -0.0102464]	Efeito Fixo
Popdensity	0.0008803	[0.0008333 , 0.0009274]	Efeito Fixo
GMM.T.2018	-0.0030800	[-0.0034927 , -0.0026711]	Efeito Fixo
IDI.2019	-0.1146791	[-0.1217765 , -0.1075809]	Efeito Fixo
Mais.65	-0.0000143	[-1.59e-05 , -1.28e-05]	Efeito Fixo
Precision for Date_str	0.3871352	[0.2402964 , 0.5707974]	Hiperparâmetro

Modelo: RW.M1.7.LS : x.LS~tm.mean.7+tm.mean.7+hm.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+
f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW.M1.7.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.9055999	[2.8688516 , 4.8576813]	Efeito Fixo
tm.mean.7	0.0009420	[-0.0344549 , 0.0362696]	Efeito Fixo
tm.mean.7	0.3198904	[0.2692959 , 0.3715644]	Efeito Fixo
hm.mean.7	-0.0328968	[-0.0438843 , -0.0215605]	Efeito Fixo
hm.mean.7	0.0311508	[0.0224506 , 0.0399232]	Efeito Fixo
pr.mean.7	-0.0160271	[-0.0267147 , -0.0052761]	Efeito Fixo
Popdensity	0.0008983	[0.0008509 , 0.0009457]	Efeito Fixo
GMM.T.2018	-0.0031837	[-0.0035977 , -0.0027726]	Efeito Fixo
IDI.2019	-0.1174430	[-0.1245881 , -0.1102975]	Efeito Fixo
Mais.65	-0.0000142	[-1.57e-05 , -1.26e-05]	Efeito Fixo
Precision for Date_str	0.3637716	[0.2294292 , 0.5323962]	Hiperparâmetro

Modelo: RW.M1.7.HF.14 : x.HF.14~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+

f(Date_str, model="rw2")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: RW.M1.7.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.0736688	[3.0866852 , 4.9593408]	Efeito Fixo
tm.mean.7	-0.0451971	[-0.0794701 , -0.0110639]	Efeito Fixo
tm.mean.7	0.3234630	[0.2731700 , 0.3751463]	Efeito Fixo
hM.mean.7	-0.0240418	[-0.0345232 , -0.0131175]	Efeito Fixo
hm.mean.7	0.0272097	[0.0185753 , 0.0359174]	Efeito Fixo
pr.mean.7	-0.0153271	[-0.0259227 , -0.0046542]	Efeito Fixo
Popdensity	0.0008696	[0.0008227 , 0.0009165]	Efeito Fixo
GMM.T.2018	-0.0030360	[-0.0034500 , -0.0026263]	Efeito Fixo
IDI.2019	-0.1124936	[-0.1195631 , -0.1054232]	Efeito Fixo
Mais.65	-0.0000144	[-0.0000159 , -0.0000128]	Efeito Fixo
Precision for Date_str	0.3517867	[0.2153241 , 0.5231991]	Hiperparâmetro

Modelo: RW.M1.7.LS.14 : x.LS.14~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+

f(Date_str, model="rw2")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: RW.M1.7.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.1591791	[1.6614905 , 4.3473128]	Efeito Fixo
tm.mean.7	-0.0563202	[-0.0905392 , -0.0221420]	Efeito Fixo
tm.mean.7	0.3426154	[0.2882238 , 0.4001366]	Efeito Fixo
hM.mean.7	-0.0159316	[-0.0281413 , -0.0021184]	Efeito Fixo
hm.mean.7	0.0308555	[0.0221106 , 0.0397143]	Efeito Fixo
pr.mean.7	-0.0142030	[-0.0252674 , -0.0030690]	Efeito Fixo
Popdensity	0.0008619	[0.0008152 , 0.0009087]	Efeito Fixo
GMM.T.2018	-0.0030125	[-0.0034418 , -0.0025903]	Efeito Fixo
IDI.2019	-0.1114774	[-0.1185101 , -0.1044439]	Efeito Fixo
Mais.65	-0.0000141	[-0.0000157 , -0.0000125]	Efeito Fixo
Precision for Date_str	0.2088377	[0.1130519 , 0.3331869]	Hiperparâmetro

Modelo: RW2.M1.14 : Diario.Unif~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+ f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW2.M1.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-0.2077422	[-1.8698798 , 1.3226283]	Efeito Fixo
tm.mean.14	0.0916440	[0.0521842 , 0.1313984]	Efeito Fixo
tm.mean.14	0.0659521	[0.0159556 , 0.1166871]	Efeito Fixo
hM.mean.14	-0.0144321	[-0.0273726 , -0.0008746]	Efeito Fixo
hm.mean.14	0.0891558	[0.0791942 , 0.0991809]	Efeito Fixo
pr.mean.14	0.0048152	[-0.0072325 , 0.0168501]	Efeito Fixo
Popdensity	0.0008312	[0.0007825 , 0.00088]	Efeito Fixo
GMM.T.2018	-0.0032129	[-0.0035904 , -0.0028362]	Efeito Fixo
IDI.2019	-0.1167746	[-0.124017 , -0.1095333]	Efeito Fixo
Precision for Date_str	0.1110339	[0.0646233 , 0.1750762]	Hiperparâmetro

Modelo: RW2.M1.14.HF : x.HF~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+ f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW2.M1.14.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	0.6033245	[-0.710623 , 1.8684054]	Efeito Fixo
tm.mean.14	0.0303234	[-0.0053452 , 0.0661863]	Efeito Fixo
tm.mean.14	0.1654461	[0.1178332 , 0.21338]	Efeito Fixo
hM.mean.14	-0.0087021	[-0.0195449 , 0.0023351]	Efeito Fixo
hm.mean.14	0.0778786	[0.0688877 , 0.0869314]	Efeito Fixo
pr.mean.14	-0.0045863	[-0.0172866 , 0.008225]	Efeito Fixo
Popdensity	0.0008310	[0.0007825 , 0.0008795]	Efeito Fixo
GMM.T.2018	-0.0035789	[-0.0039582 , -0.0032004]	Efeito Fixo
IDI.2019	-0.1141835	[-0.1213938 , -0.1069736]	Efeito Fixo
Precision for Date_str	0.2395639	[0.1478467 , 0.3554698]	Hiperparâmetro

Modelo: RW2.M1.14.LS : x.LS~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+ f(Date_str, model="rw2")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: RW2.M1.14.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	0.4538265	[-0.868152 , 1.7288853]	Efeito Fixo
tm.mean.14	0.0610295	[0.0249042 , 0.097395]	Efeito Fixo
tm.mean.14	0.1447878	[0.0971318 , 0.1927156]	Efeito Fixo
hM.mean.14	-0.0126739	[-0.0234989 , -0.001683]	Efeito Fixo
hm.mean.14	0.0791160	[0.0700356 , 0.0882523]	Efeito Fixo
pr.mean.14	-0.0026637	[-0.0154766 , 0.0102658]	Efeito Fixo
Popdensity	0.0008400	[0.0007912 , 0.0008888]	Efeito Fixo

```
|GMM.T.2018 | -0.0034544| [ -0.0038332 , -0.0030762 ] | Efeito Fixo |
|IDI.2019 | -0.1180491| [ -0.1253198 , -0.1107788 ] | Efeito Fixo |
|Precision for Date_str | 0.2310794| [ 0.1415464 , 0.3441772 ] | Hiperparâmetro |
```

Modelo: RW2.M1.14.HF.14 : x.HF.14~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019 + f(Date_str, model="rw2")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: RW2.M1.14.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.1591791	[1.6614905 , 4.3473128]	Efeito Fixo
tm.mean.7	-0.0563202	[-0.0905392 , -0.0221420]	Efeito Fixo
tm.mean.7	0.3426154	[0.2882238 , 0.4001366]	Efeito Fixo
hM.mean.7	-0.0159316	[-0.0281413 , -0.0021184]	Efeito Fixo
hm.mean.7	0.0308555	[0.0221106 , 0.0397143]	Efeito Fixo
pr.mean.7	-0.0142030	[-0.0252674 , -0.0030690]	Efeito Fixo
Popdensity	0.0008619	[0.0008152 , 0.0009087]	Efeito Fixo
GMM.T.2018	-0.0030125	[-0.0034418 , -0.0025903]	Efeito Fixo
IDI.2019	-0.1114774	[-0.1185101 , -0.1044439]	Efeito Fixo
Mais.65	-0.0000141	[-0.0000157 , -0.0000125]	Efeito Fixo
Precision for Date_str	0.2088377	[0.1130519 , 0.3331869]	Hiperparâmetro

Modelo: RW2.M1.14.LS.14 : x.LS.14~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019 + f(Date_str, model="rw2")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: RW2.M1.14.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.1591791	[1.6614905 , 4.3473128]	Efeito Fixo
tm.mean.7	-0.0563202	[-0.0905392 , -0.0221420]	Efeito Fixo
tm.mean.7	0.3426154	[0.2882238 , 0.4001366]	Efeito Fixo
hM.mean.7	-0.0159316	[-0.0281413 , -0.0021184]	Efeito Fixo
hm.mean.7	0.0308555	[0.0221106 , 0.0397143]	Efeito Fixo
pr.mean.7	-0.0142030	[-0.0252674 , -0.0030690]	Efeito Fixo
Popdensity	0.0008619	[0.0008152 , 0.0009087]	Efeito Fixo
GMM.T.2018	-0.0030125	[-0.0034418 , -0.0025903]	Efeito Fixo
IDI.2019	-0.1114774	[-0.1185101 , -0.1044439]	Efeito Fixo
Mais.65	-0.0000141	[-0.0000157 , -0.0000125]	Efeito Fixo
Precision for Date_str	0.2088377	[0.1130519 , 0.3331869]	Hiperparâmetro

Modelo: iCAR.M1.14 : Diario.Unif~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 + Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.9413105	[-3.8441617 , 11.725409]	Efeito Fixo
tm.mean.14	-0.0595322	[-0.0682363 , -0.0508281]	Efeito Fixo
tm.mean.14	-0.1297012	[-0.1418491 , -0.1175534]	Efeito Fixo
hM.mean.14	0.0194572	[0.0175155 , 0.0213988]	Efeito Fixo
hm.mean.14	0.0044149	[0.0021046 , 0.0067252]	Efeito Fixo
pr.mean.14	0.0182527	[0.015042 , 0.0214634]	Efeito Fixo
Popdensity	0.0008704	[-0.000547 , 0.0022879]	Efeito Fixo
GMM.T.2018	0.0007787	[-0.0075248 , 0.0090852]	Efeito Fixo
IDI.2019	-0.1557168	[-0.3241091 , 0.0126045]	Efeito Fixo
Mais.65	-0.0000011	[-8.23e-05 , 8e-05]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5131337	[0.2282656 , 0.9647692]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2157.9929253	[157.3725618 , 8271.138881]	Hiperparâmetro

Modelo: iCAR.M1.14.HF : x.HF~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.14.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.5693325	[-4.255064 , 11.3923543]	Efeito Fixo
tm.mean.14	-0.0329738	[-0.0415737 , -0.024374]	Efeito Fixo
tm.mean.14	-0.1452530	[-0.1569881 , -0.1335179]	Efeito Fixo
hM.mean.14	0.0201690	[0.0181812 , 0.0221568]	Efeito Fixo
hm.mean.14	0.0048455	[0.0025431 , 0.0071479]	Efeito Fixo
pr.mean.14	0.0092199	[0.0057085 , 0.0127313]	Efeito Fixo
Popdensity	0.0008764	[-0.000548 , 0.002301]	Efeito Fixo
GMM.T.2018	0.0007983	[-0.0075466 , 0.0091461]	Efeito Fixo
IDI.2019	-0.1571194	[-0.3263462 , 0.0120376]	Efeito Fixo
Mais.65	-0.0000061	[-8.22e-05 , 8.09e-05]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5082432	[0.2261443 , 0.9557888]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2158.3454464	[157.4918193 , 8277.8780642]	Hiperparâmetro

Modelo: iCAR.M1.14.LS : x.LS~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.14.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.6185404	[-4.1955247 , 11.4312168]	Efeito Fixo

tm.mean.14	-0.0393905	[-0.0480132 , -0.0307678]	Efeito Fixo
tm.mean.14	-0.1366462	[-0.1484416 , -0.1248509]	Efeito Fixo
hm.mean.14	0.0208156	[0.0188300 , 0.0228013]	Efeito Fixo
hm.mean.14	0.0041968	[0.0018965 , 0.0064972]	Efeito Fixo
pr.mean.14	0.0092423	[0.0057684 , 0.0127162]	Efeito Fixo
Popdensity	0.0008863	[-0.0005362 , 0.0023090]	Efeito Fixo
GMM.T.2018	0.0008154	[-0.0075186 , 0.0091526]	Efeito Fixo
IDI.2019	-0.1602517	[-0.3292789 , 0.0086970]	Efeito Fixo
Mais.65	-0.0000001	[-0.0000816 , 0.0000813]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5094107	[0.2266024 , 0.9577574]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2156.6595619	[157.2076206 , 8263.0603430]	Hiperparâmetro

Modelo: iCAR.M1.14.HF.14 : x.HF.14~tm.mean.14+tm.mean.14+hm.mean.14+hm.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.M1.14.HF.14

Variável	Média	Intervalo.de.Credibilidade	Tipo
(Intercept)	3.5818461	[-4.2580281 , 11.4203742]	Efeito Fixo
tm.mean.14	-0.0315817	[-0.0402258 , -0.0229375]	Efeito Fixo
tm.mean.14	-0.1313209	[-0.1430474 , -0.1195943]	Efeito Fixo
hm.mean.14	0.0195800	[0.0175785 , 0.0215815]	Efeito Fixo
hm.mean.14	0.0034530	[0.0011537 , 0.0057523]	Efeito Fixo
pr.mean.14	0.0108237	[0.0073910 , 0.0142565]	Efeito Fixo
Popdensity	0.0008804	[-0.0005469 , 0.0023079]	Efeito Fixo
GMM.T.2018	0.0007506	[-0.0076108 , 0.0091149]	Efeito Fixo
IDI.2019	-0.1562929	[-0.3258551 , 0.0131978]	Efeito Fixo
Mais.65	-0.0000012	[-0.0000830 , 0.0000805]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5062947	[0.225295 , 0.9521996]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2157.4484686	[157.423488 , 8274.3656440]	Hiperparâmetro

Modelo: iCAR.M1.14.LS.14 : x.LS.14~tm.mean.14+tm.mean.14+hm.mean.14+hm.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.M1.14.LS.14

Variável	Média	Intervalo.de.Credibilidade	Tipo
(Intercept)	3.4958815	[-4.3282760 , 11.3186705]	Efeito Fixo
tm.mean.14	-0.0347618	[-0.0434364 , -0.0260871]	Efeito Fixo
tm.mean.14	-0.1052861	[-0.1169806 , -0.0935916]	Efeito Fixo
hm.mean.14	0.0191847	[0.0171798 , 0.0211895]	Efeito Fixo
hm.mean.14	0.0008848	[-0.0014007 , 0.0031704]	Efeito Fixo
pr.mean.14	0.0114730	[0.0081377 , 0.0148082]	Efeito Fixo
Popdensity	0.0008707	[-0.0005537 , 0.0022953]	Efeito Fixo
GMM.T.2018	0.0008998	[-0.0074448 , 0.0092477]	Efeito Fixo
IDI.2019	-0.1578944	[-0.3271451 , 0.0112744]	Efeito Fixo
Mais.65	-0.0000016	[-0.0000831 , 0.0000800]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5083554	[0.2262242 , 0.95611]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2156.1406646	[157.2984302 , 8268.57526]	Hiperparâmetro

Modelo: iCAR.M1.2 : Diario.Unif~tm.mean.2+tm.mean.2+hm.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

f(Region_unstr,model="bym",graph=W.porto)

Tabele: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.2

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	1.6763911	[-6.2669769 , 9.6179778]	Efeito Fixo
tm.mean.2	-0.0024643	[-0.0082419 , 0.0033132]	Efeito Fixo
tm.mean.2	-0.1484678	[-0.1581051 , -0.1388304]	Efeito Fixo
hm.mean.2	0.0328491	[0.0299121 , 0.035786]	Efeito Fixo
hm.mean.2	0.0054934	[0.0032989 , 0.007688]	Efeito Fixo
pr.mean.2	0.0210505	[0.018107 , 0.023994]	Efeito Fixo
Popdensity	0.0008515	[-0.0005937 , 0.0022969]	Efeito Fixo
GMM.T.2018	0.0010087	[-0.0074574 , 0.0094781]	Efeito Fixo
IDI.2019	-0.1578143	[-0.3295014 , 0.0138008]	Efeito Fixo
Mais.65	0.0000003	[-8.24e-05 , 8.31e-05]	Efeito Fixo
Precision for Region_unstr (iid component)	0.4931075	[0.2192374 , 0.9265939]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2160.5536795	[157.2026377 , 8256.3189351]	Hiperparâmetro

Modelo: iCAR.M1.2.HF : x.HF~tm.mean.2+tm.mean.2+hm.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Tabele: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.2.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.4462405	[-5.3992794 , 10.2900365]	Efeito Fixo
tm.mean.2	0.0012915	[-0.0045653 , 0.0071483]	Efeito Fixo
tm.mean.2	-0.1613648	[-0.1712403 , -0.1514893]	Efeito Fixo
hm.mean.2	0.0224613	[0.0199381 , 0.0249845]	Efeito Fixo
hm.mean.2	0.0167276	[0.0145155 , 0.0189397]	Efeito Fixo
pr.mean.2	0.0209472	[0.0180582 , 0.0238362]	Efeito Fixo
Popdensity	0.0008650	[-0.0005627 , 0.0022927]	Efeito Fixo
GMM.T.2018	0.0006316	[-0.0077316 , 0.0089981]	Efeito Fixo
IDI.2019	-0.1570397	[-0.3266421 , 0.0124894]	Efeito Fixo
Mais.65	0.0000016	[-8.02e-05 , 8.33e-05]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5056532	[0.224899 , 0.9505055]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2156.2751312	[157.0614872 , 8255.3264182]	Hiperparâmetro

Modelo: iCAR.M1.2.LS : $x.LS \sim tm.mean.2 + tm.mean.2 + hm.mean.2 + hm.mean.2 + pr.mean.2 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 + f(Region_unstr, model="bym", graph=W.porto)$

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.2.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.4317579	[-5.4079709 , 10.2697441]	Efeito Fixo
tm.mean.2	0.0018363	[-0.0040348 , 0.0077074]	Efeito Fixo
tm.mean.2	-0.1654382	[-0.1753407 , -0.1555357]	Efeito Fixo
hm.mean.2	0.0224903	[0.0199489 , 0.0250318]	Efeito Fixo
hm.mean.2	0.0179832	[0.0157623 , 0.020204]	Efeito Fixo
pr.mean.2	0.0187025	[0.01576 , 0.0216451]	Efeito Fixo
Popdensity	0.0008750	[-0.0005516 , 0.0023016]	Efeito Fixo
GMM.T.2018	0.0006747	[-0.0076825 , 0.0090352]	Efeito Fixo
IDI.2019	-0.1605108	[-0.3300103 , 0.008906]	Efeito Fixo
Mais.65	0.0000023	[-7.94e-05 , 8.39e-05]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5063867	[0.2252047 , 0.9518482]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2154.4086878	[156.8196369 , 8243.3198312]	Hiperparâmetro

Modelo: iCAR.M1.2.HF.14 : $x.HF.14 \sim tm.mean.2 + tm.mean.2 + hm.mean.2 + hm.mean.2 + pr.mean.2 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 + f(Region_unstr, model="bym", graph=W.porto)$

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.M1.2.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.476270	[-5.4046424 , 10.3555051]	Efeito Fixo
tm.mean.2	0.001181	[-0.0046348 , 0.0069964]	Efeito Fixo
tm.mean.2	-0.148272	[-0.1580196 , -0.1385245]	Efeito Fixo
hm.mean.2	0.022577	[0.0200950 , 0.0250584]	Efeito Fixo
hm.mean.2	0.013201	[0.0110125 , 0.0153892]	Efeito Fixo
pr.mean.2	0.023323	[0.0204898 , 0.0261560]	Efeito Fixo
Popdensity	0.000868	[-0.0005664 , 0.0023020]	Efeito Fixo
GMM.T.2018	0.000641	[-0.0077597 , 0.0090458]	Efeito Fixo
IDI.2019	-0.156454	[-0.3268249 , 0.0138426]	Efeito Fixo
Mais.65	0.000001	[-0.0000813 , 0.0000829]	Efeito Fixo
Precision for Region_unstr (iid component)	0.501315	[0.2230318 , 0.9426355]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2155.418858	[157.0791349 , 8257.3193487]	Hiperparâmetro

Modelo: iCAR.M1.2.LS.14 : $x.LS.14 \sim tm.mean.2 + tm.mean.2 + hm.mean.2 + hm.mean.2 + pr.mean.2 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 + f(Region_unstr, model="bym", graph=W.porto)$

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.M1.2.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.345676	[-5.5266207 , 10.2162923]	Efeito Fixo
tm.mean.2	0.003092	[-0.0026707 , 0.0088548]	Efeito Fixo
tm.mean.2	-0.135587	[-0.1451564 , -0.1260169]	Efeito Fixo
hm.mean.2	0.022077	[0.0196479 , 0.0245052]	Efeito Fixo
hm.mean.2	0.011252	[0.0090943 , 0.0134100]	Efeito Fixo
pr.mean.2	0.022243	[0.0194047 , 0.0250803]	Efeito Fixo
Popdensity	0.000858	[-0.0005744 , 0.0022909]	Efeito Fixo
GMM.T.2018	0.000848	[-0.0075441 , 0.0092437]	Efeito Fixo
IDI.2019	-0.158480	[-0.3286965 , 0.0116511]	Efeito Fixo
Mais.65	0.000001	[-0.0000816 , 0.0000825]	Efeito Fixo
Precision for Region_unstr (iid component)	0.501819	[0.2231038 , 0.9428865]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2154.718179	[156.6599496 , 8232.5919755]	Hiperparâmetro

Modelo: iCAR.M1.7 : $Diario.Unif \sim tm.mean.7 + tm.mean.7 + hm.mean.7 + hm.mean.7 + pr.mean.7 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 + f(Region_unstr, model="bym", graph=W.porto)$

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.7

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.9725580	[-4.8148137 , 10.7585177]	Efeito Fixo
tm.mean.7	-0.0590135	[-0.0681588 , -0.0498683]	Efeito Fixo
tm.mean.7	-0.0848091	[-0.0981266 , -0.0714915]	Efeito Fixo
hm.mean.7	0.0272342	[0.0245629 , 0.0299056]	Efeito Fixo
hm.mean.7	0.0046312	[0.0023219 , 0.0069406]	Efeito Fixo
pr.mean.7	0.0076687	[0.0043765 , 0.0109608]	Efeito Fixo
Popdensity	0.0008806	[-0.0005367 , 0.0022981]	Efeito Fixo
GMM.T.2018	0.0007093	[-0.0075937 , 0.0090153]	Efeito Fixo
IDI.2019	-0.1567889	[-0.3251705 , 0.0115225]	Efeito Fixo
Mais.65	-0.0000012	[-8.23e-05 , 8e-05]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5114535	[0.227035 , 0.9595877]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2174.5096094	[157.6590654 , 8282.6081876]	Hiperparâmetro

Modelo: iCAR.M1.7.HF : $x.HF \sim tm.mean.2 + tm.mean.2 + hm.mean.2 + hm.mean.2 + pr.mean.2 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 + f(Region_unstr, model="bym", graph=W.porto)$

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.7.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.0257510	[-4.7747381 , 10.8248445]	Efeito Fixo
tm.mean.7	-0.0505651	[-0.0596712 , -0.041459]	Efeito Fixo
tm.mean.7	-0.0739440	[-0.0871593 , -0.0607287]	Efeito Fixo

hm.mean.7		0.0272927	[0.0246463 , 0.0299391]	Efeito Fixo
hm.mean.7		-0.0014716	[-0.0037768 , 0.0008336]	Efeito Fixo
pr.mean.7		0.0144385	[0.0112425 , 0.0176346]	Efeito Fixo
Popdensity		0.0008837	[-0.0005361 , 0.0023035]	Efeito Fixo
GMM.T.2018		0.0006783	[-0.0076388 , 0.0089984]	Efeito Fixo
IDI.2019		-0.1571967	[-0.325861 , 0.0113992]	Efeito Fixo
Mais.65		-0.0000009	[-8.22e-05 , 8.03e-05]	Efeito Fixo
Precision for Region_unstr (iid component)		0.5097146	[0.2262655 , 0.9563012]	Hiperparâmetro
Precision for Region_unstr (spatial component)		2172.9961202	[157.3809262 , 8270.4020762]	Hiperparâmetro

Modelo: iCAR.M1.7.LS : x.LS~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.M1.7.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.9295363	[-4.872762 , 10.7304077]	Efeito Fixo
tm.mean.7	-0.0458687	[-0.0549816 , -0.0367559]	Efeito Fixo
tm.mean.7	-0.0767819	[-0.089971 , -0.0635928]	Efeito Fixo
hM.mean.7	0.0278225	[0.0251289 , 0.0305161]	Efeito Fixo
hm.mean.7	-0.0020055	[-0.0043139 , 0.0003028]	Efeito Fixo
pr.mean.7	0.0153266	[0.0121299 , 0.0185233]	Efeito Fixo
Popdensity	0.0008912	[-0.0005288 , 0.0023114]	Efeito Fixo
GMM.T.2018	0.0007534	[-0.0075656 , 0.0090755]	Efeito Fixo
IDI.2019	-0.1606892	[-0.3294116 , 0.0079561]	Efeito Fixo
Mais.65	-0.0000004	[-8.17e-05 , 8.1e-05]	Efeito Fixo
Precision for Region_unstr (iid component)	0.5095528	[0.2261978 , 0.9560568]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2172.2554596	[157.1376418 , 8260.5596487]	Hiperparâmetro

Modelo: iCAR.M1.7.HF.14 : x.HF.14~tm.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

f(Region_unstr,model="bym",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.M1.7.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.974774	[-4.8550854 , 10.8032590]	Efeito Fixo
tm.mean.7	-0.054734	[-0.0637744 , -0.0456943]	Efeito Fixo
tm.mean.7	-0.072724	[-0.0858917 , -0.0595556]	Efeito Fixo
hM.mean.7	0.028404	[0.0257623 , 0.0310451]	Efeito Fixo
hm.mean.7	-0.001709	[-0.0040043 , 0.0005857]	Efeito Fixo
pr.mean.7	0.011870	[0.0086299 , 0.0151093]	Efeito Fixo
Popdensity	0.000885	[-0.0005403 , 0.0023100]	Efeito Fixo
GMM.T.2018	0.000708	[-0.0076408 , 0.0090590]	Efeito Fixo
IDI.2019	-0.156564	[-0.3258637 , 0.0126643]	Efeito Fixo
Mais.65	-0.000001	[-0.0000830 , 0.0000802]	Efeito Fixo
Precision for Region_unstr (iid component)	0.505944	[0.2246055 , 0.9493049]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2172.334118	[157.1641558 , 8261.4069100]	Hiperparâmetro

Modelo: iCAR.M1.7.LS.14 : x.LS.14~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="bym",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.M1.7.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.910454	[-4.9268778 , 10.7464066]	Efeito Fixo
tm.mean.7	-0.055010	[-0.0639705 , -0.0460499]	Efeito Fixo
tm.mean.7	-0.066196	[-0.0792162 , -0.0531757]	Efeito Fixo
hM.mean.7	0.027696	[0.0251171 , 0.0302739]	Efeito Fixo
hm.mean.7	-0.002915	[-0.0051960 , -0.0006346]	Efeito Fixo
pr.mean.7	0.009552	[0.0062715 , 0.0128330]	Efeito Fixo
Popdensity	0.000873	[-0.0005540 , 0.0022991]	Efeito Fixo
GMM.T.2018	0.000932	[-0.0074249 , 0.0092917]	Efeito Fixo
IDI.2019	-0.158536	[-0.3280289 , 0.0108759]	Efeito Fixo
Mais.65	-0.000002	[-0.0000832 , 0.0000801]	Efeito Fixo
Precision for Region_unstr (iid component)	0.505784	[0.2247852 , 0.9500271]	Hiperparâmetro
Precision for Region_unstr (spatial component)	2175.713956	[158.8882633 , 8341.7358849]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.14 : Diario.Unif~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

f(Region_str, model="bym",graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date,model="iid")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.4149248	[-5.8113914 , 12.6407805]	Efeito Fixo
tm.mean.14	0.0560082	[0.0138423 , 0.0981186]	Efeito Fixo
tm.mean.14	-0.3193900	[-0.3770655 , -0.2623104]	Efeito Fixo
hM.mean.14	-0.0016106	[-0.0142528 , 0.0108925]	Efeito Fixo
hm.mean.14	0.0133983	[0.0025193 , 0.0242565]	Efeito Fixo
pr.mean.14	0.0476702	[0.0352566 , 0.0600944]	Efeito Fixo
Popdensity	0.0009520	[-0.0007168 , 0.0026209]	Efeito Fixo
GMM.T.2018	0.0022192	[-0.0075594 , 0.0120029]	Efeito Fixo
IDI.2019	-0.1779621	[-0.3761219 , 0.0201433]	Efeito Fixo
Mais.65	-0.0000017	[-9.73e-05 , 9.38e-05]	Efeito Fixo
Precision for Region_str (iid component)	0.3699509	[0.1643198 , 0.6924617]	Hiperparâmetro
Precision for Region_str (spatial component)	2253.6616620	[170.0495293 , 8327.326013]	Hiperparâmetro

Precision for Date_str	18273.6607221	[2831.7964251 , 59340.1222022]	Hiperparâmetro
Precision for Date_unstr	0.6986726	[0.4894195 , 0.9573528]	Hiperparâmetro
Precision for region_date	21635.7882083	[1576.3671649 , 83047.7075199]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.14.HF : x.HF~tm.mean.14+tm.mean.14+hm.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.14.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.3902056	[-5.7709597 , 12.5500469]	Efeito Fixo
tm.mean.14	0.0574848	[0.0200504 , 0.0950087]	Efeito Fixo
tm.mean.14	-0.1725561	[-0.2236444 , -0.1217815]	Efeito Fixo
hm.mean.14	-0.0072326	[-0.0180193 , 0.0034674]	Efeito Fixo
hm.mean.14	0.0276586	[0.0180351 , 0.0373315]	Efeito Fixo
pr.mean.14	0.0252973	[0.012655 , 0.0380068]	Efeito Fixo
Popdensity	0.0009602	[-0.0007011 , 0.0026216]	Efeito Fixo
GMM.T.2018	0.0007394	[-0.0089948 , 0.010477]	Efeito Fixo
IDI.2019	-0.1731965	[-0.3704674 , 0.0240184]	Efeito Fixo
Mais.65	0.0000017	[-9.34e-05 , 9.69e-05]	Efeito Fixo
Precision for Region_str (iid component)	0.3816880	[0.1715846 , 0.728502]	Hiperparâmetro
Precision for Region_str (spatial component)	2285.8004577	[172.0014814 , 8849.5926389]	Hiperparâmetro
Precision for Date_str	21770.7576116	[3813.7746291 , 66432.4450095]	Hiperparâmetro
Precision for Date_unstr	1.2884376	[0.9194661 , 1.7294588]	Hiperparâmetro
Precision for region_date	23341.4324297	[1776.6690389 , 94518.6257556]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.14.LS : x.LS~tm.mean.14+tm.mean.14+hm.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.14.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.2731173	[-5.9448413 , 12.4899807]	Efeito Fixo
tm.mean.14	0.0882704	[0.0501374 , 0.1265828]	Efeito Fixo
tm.mean.14	-0.1964611	[-0.2486898 , -0.1445756]	Efeito Fixo
hm.mean.14	-0.0111781	[-0.022175 , -0.0002847]	Efeito Fixo
hm.mean.14	0.0283799	[0.018645 , 0.0381645]	Efeito Fixo
pr.mean.14	0.0279710	[0.0152398 , 0.0407596]	Efeito Fixo
Popdensity	0.0009786	[-0.0006927 , 0.0026501]	Efeito Fixo
GMM.T.2018	0.0009254	[-0.0088675 , 0.010722]	Efeito Fixo
IDI.2019	-0.1794129	[-0.3778943 , 0.0190033]	Efeito Fixo
Mais.65	0.0000019	[-9.38e-05 , 9.76e-05]	Efeito Fixo
Precision for Region_str (iid component)	0.3755544	[0.1678729 , 0.7102851]	Hiperparâmetro
Precision for Region_str (spatial component)	2796.9860249	[255.9432943 , 11069.6193553]	Hiperparâmetro
Precision for Date_str	18885.5012291	[3123.2385137 , 57976.412957]	Hiperparâmetro
Precision for Date_unstr	1.1964578	[0.8630467 , 1.6082682]	Hiperparâmetro
Precision for region_date	27161.9275857	[2524.2739178 , 107861.6808209]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.14.HF.14 : x.HF.14~tm.mean.14+tm.mean.14+hm.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.rw.iid.M2.14.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	5.5860120	[-3.6695317 , 14.8476230]	Efeito Fixo
tm.mean.2	-0.0227672	[-0.0540079 , 0.0083308]	Efeito Fixo
tm.mean.2	0.2045395	[0.1533850 , 0.2562608]	Efeito Fixo
hm.mean.2	-0.0089800	[-0.0196220 , 0.0015107]	Efeito Fixo
hm.mean.2	-0.0383806	[-0.0479602 , -0.0288996]	Efeito Fixo
pr.mean.2	-0.0313719	[-0.0441679 , -0.0186339]	Efeito Fixo
Popdensity	0.0010192	[-0.0006539 , 0.0026926]	Efeito Fixo
GMM.T.2018	-0.0000823	[-0.0098884 , 0.0097278]	Efeito Fixo
IDI.2019	-0.1745554	[-0.3732453 , 0.0240567]	Efeito Fixo
Mais.65	-0.0000043	[-0.0001001 , 0.0000915]	Efeito Fixo
Precision for Region_str (iid component)	0.3726252	[0.1662039 , 0.7018375]	Hiperparâmetro
Precision for Region_str (spatial component)	2552.6789489	[210.0619771 , 9980.5671719]	Hiperparâmetro
Precision for Date_str	4382.7767448	[752.5356990 , 13803.4696702]	Hiperparâmetro
Precision for Date_unstr	0.9316430	[0.6581982 , 1.2642232]	Hiperparâmetro
Precision for region_date	25898.9858116	[2279.3447191 , 102453.1288576]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.14.LS.14 : x.LS.14~tm.mean.14+tm.mean.14+hm.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.rw.iid.M2.14.LS.14

Variável	Média	Intervalo.de.Credibilidade	Tipo
(Intercept)	5.0226732	[-4.2373498 , 14.2860776]	Efeito Fixo
tM.mean.2	-0.0233413	[-0.0559629 , 0.0091573]	Efeito Fixo
tm.mean.2	0.2058830	[0.1541050 , 0.2579533]	Efeito Fixo
hM.mean.2	-0.0069932	[-0.0178272 , 0.0037410]	Efeito Fixo
hm.mean.2	-0.0340742	[-0.0435154 , -0.0246900]	Efeito Fixo
pr.mean.2	-0.0349998	[-0.0477944 , -0.0222437]	Efeito Fixo
Popdensity	0.0010190	[-0.0006542 , 0.0026924]	Efeito Fixo
GMM.T.2018	0.0002534	[-0.0095537 , 0.0100646]	Efeito Fixo
IDI.2019	-0.1788930	[-0.3776067 , 0.0197469]	Efeito Fixo
Mais.65	-0.0000045	[-0.0001004 , 0.0000913]	Efeito Fixo
Precision for Region_str (iid component)	0.3644492	[0.1606800 , 0.6789333]	Hiperparâmetro
Precision for Region_str (spatial component)	2196.0694554	[156.5751970 , 8308.6216177]	Hiperparâmetro
Precision for Date_str	5608.4562589	[793.1671750 , 19984.5081770]	Hiperparâmetro
Precision for Date_unstr	0.7828525	[0.5628885 , 1.0638852]	Hiperparâmetro
Precision for region_date	22397.4615802	[1719.4226845 , 87467.4260820]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.2 : Diario.Unif~tM.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019 +Mais.65+

```
f(region_str,model="bym",graph=W.porto)+
f(Date_str,model="rw2") +
f(Date_unstr,model="iid")+
f(region_date,model="iid")
```

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.2

Variável	Média	Intervalo.de.Credibilidade	Tipo
(Intercept)	6.4373010	[-2.6489335 , 15.5247350]	Efeito Fixo
tM.mean.2	-0.1513698	[-0.1866775 , -0.1165421]	Efeito Fixo
tm.mean.2	0.3343839	[0.2737501 , 0.3963352]	Efeito Fixo
hM.mean.2	-0.0039253	[-0.0162579 , 0.0085775]	Efeito Fixo
hm.mean.2	-0.0438013	[-0.0538992 , -0.0337642]	Efeito Fixo
pr.mean.2	-0.0676378	[-0.0812609 , -0.0540488]	Efeito Fixo
Popdensity	0.0010761	[-0.0005608 , 0.0027132]	Efeito Fixo
GMM.T.2018	-0.0004124	[-0.0100094 , 0.0091880]	Efeito Fixo
IDI.2019	-0.1723767	[-0.3667621 , 0.0219550]	Efeito Fixo
Mais.65	-0.0000070	[-0.0001008 , 0.0000867]	Efeito Fixo
Precision for Region_str (iid component)	0.3856208	[0.1711932 , 0.7201180]	Hiperparâmetro
Precision for Region_str (spatial component)	2466.9514119	[207.1222865 , 9672.7271051]	Hiperparâmetro
Precision for Date_str	13216.0078498	[1410.7392828 , 50211.8567150]	Hiperparâmetro
Precision for Date_unstr	0.5110358	[0.3600433 , 0.7062831]	Hiperparâmetro
Precision for region_date	23769.3353201	[1950.8421667 , 92939.1480189]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.2.HF : x.HF~tM.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(region_str,model="bym",graph=W.porto)+
f(Date_str,model="rw2") +
f(Date_unstr,model="iid")+
f(region_date,model="iid")
```

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.2.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	6.0610701	[-3.0529135 , 15.1758765]	Efeito Fixo
tM.mean.2	-0.0436515	[-0.0750701 , -0.0125678]	Efeito Fixo
tm.mean.2	0.1928222	[0.1414587 , 0.2449395]	Efeito Fixo
hM.mean.2	-0.0118254	[-0.0224756 , -0.0013411]	Efeito Fixo
hm.mean.2	-0.0323799	[-0.0420198 , -0.0228519]	Efeito Fixo
pr.mean.2	-0.0323841	[-0.0451837 , -0.0196427]	Efeito Fixo
Popdensity	0.0010153	[-0.0006313 , 0.0026621]	Efeito Fixo
GMM.T.2018	-0.0001999	[-0.0098507 , 0.0094547]	Efeito Fixo
IDI.2019	-0.1735699	[-0.3691058 , 0.0219002]	Efeito Fixo
Mais.65	-0.0000034	[-9.77e-05 , 9.09e-05]	Efeito Fixo
Precision for Region_str (iid component)	0.3795165	[0.1671253 , 0.7071357]	Hiperparâmetro
Precision for Region_str (spatial component)	2629.4943031	[229.09279 , 10534.1207346]	Hiperparâmetro
Precision for Date_str	6018.8610029	[1069.3558092 , 19732.3484028]	Hiperparâmetro
Precision for Date_unstr	1.0192409	[0.719672 , 1.385611]	Hiperparâmetro
Precision for region_date	26925.4621708	[2322.1683622 , 107371.4089923]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.2.LS : x.LS~tM.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(region_str,model="bym",graph=W.porto)+
f(Date_str,model="rw2") +
f(Date_unstr,model="iid")+
f(region_date,model="iid")
```

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.2.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	6.2006989	[-2.968158 , 15.3743631]	Efeito Fixo
tM.mean.2	-0.0515853	[-0.0840671 , -0.0194345]	Efeito Fixo
tm.mean.2	0.2245118	[0.1722011 , 0.2775913]	Efeito Fixo
hM.mean.2	-0.0134652	[-0.0243739 , -0.0026994]	Efeito Fixo
hm.mean.2	-0.0325522	[-0.0423695 , -0.022833]	Efeito Fixo
pr.mean.2	-0.0371652	[-0.0498632 , -0.0245132]	Efeito Fixo
Popdensity	0.0010462	[-0.0006102 , 0.0027027]	Efeito Fixo
GMM.T.2018	-0.0002881	[-0.0099968 , 0.0094244]	Efeito Fixo
IDI.2019	-0.1789166	[-0.3756338 , 0.017724]	Efeito Fixo
Mais.65	-0.0000032	[-9.81e-05 , 9.16e-05]	Efeito Fixo
Precision for Region_str (iid component)	0.3781993	[0.1684405 , 0.7101724]	Hiperparâmetro
Precision for Region_str (spatial component)	2269.0676618	[185.3452621 , 9034.2337691]	Hiperparâmetro
Precision for Date_str	6313.3568330	[1077.8218664 , 20745.9112354]	Hiperparâmetro

```
| Precision for Date_unstr | 0.9396238 | [ 0.6706706 , 1.280247 ] | Hiperparâmetro |
| Precision for region_date | 22274.5068175 | [ 1637.7080588 , 86402.5989732 ] | Hiperparâmetro |
```

Modelo: iCAR.rw.iid.M2.2.HF.14: x.HF.14~tm.mean.2+tm.mean.2+hM.mean.2+hM.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.rw.iid.M2.2.HF.14

Variável	Média	Intervalo.de.Credibilidade	Tipo
(Intercept)	5.5860120	[-3.6695317 , 14.8476230]	Efeito Fixo
tm.mean.2	-0.0227672	[-0.0540079 , 0.0083308]	Efeito Fixo
tm.mean.2	0.2045395	[0.1533850 , 0.2562608]	Efeito Fixo
hM.mean.2	-0.0089800	[-0.0196220 , 0.0015107]	Efeito Fixo
hM.mean.2	-0.0383806	[-0.0479602 , -0.0288996]	Efeito Fixo
pr.mean.2	-0.0313719	[-0.0441679 , -0.0186339]	Efeito Fixo
Popdensity	0.0010192	[-0.0006539 , 0.0026926]	Efeito Fixo
GMM.T.2018	-0.0000823	[-0.0098884 , 0.0097278]	Efeito Fixo
IDI.2019	-0.1745554	[-0.3732453 , 0.0240567]	Efeito Fixo
Mais.65	-0.0000043	[-0.0001001 , 0.0000915]	Efeito Fixo
Precision for Region_str (iid component)	0.3726252	[0.1662039 , 0.7018375]	Hiperparâmetro
Precision for Region_str (spatial component)	2552.6789489	[210.0619771 , 9980.5671719]	Hiperparâmetro
Precision for Date_str	4382.7767448	[752.5356990 , 13803.4696702]	Hiperparâmetro
Precision for Date_unstr	0.9316430	[0.6581982 , 1.2642232]	Hiperparâmetro
Precision for region_date	25898.9858116	[2279.3447191 , 102453.1288576]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.2.LS.14: x.LS.14~tm.mean.2+tm.mean.2+hM.mean.2+hM.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.rw.iid.M2.2.LS.14

Variável	Média	Intervalo.de.Credibilidade	Tipo
(Intercept)	5.0226732	[-4.2373498 , 14.2860776]	Efeito Fixo
tm.mean.2	-0.0233413	[-0.0559629 , 0.0091573]	Efeito Fixo
tm.mean.2	0.2058830	[0.1541050 , 0.2579533]	Efeito Fixo
hM.mean.2	-0.0069932	[-0.0178272 , 0.0037410]	Efeito Fixo
hM.mean.2	-0.0340742	[-0.0435154 , -0.0246900]	Efeito Fixo
pr.mean.2	-0.0349998	[-0.0477944 , -0.0222437]	Efeito Fixo
Popdensity	0.0010190	[-0.0006542 , 0.0026924]	Efeito Fixo
GMM.T.2018	0.0002534	[-0.0095537 , 0.0100646]	Efeito Fixo
IDI.2019	-0.1788930	[-0.3776067 , 0.0197469]	Efeito Fixo
Mais.65	-0.0000045	[-0.0001004 , 0.0000913]	Efeito Fixo
Precision for Region_str (iid component)	0.3644492	[0.1606800 , 0.6789333]	Hiperparâmetro
Precision for Region_str (spatial component)	2196.0694554	[156.5751970 , 8308.6216177]	Hiperparâmetro
Precision for Date_str	5608.4562589	[793.1671750 , 19984.5081770]	Hiperparâmetro
Precision for Date_unstr	0.7828525	[0.5628885 , 1.0638852]	Hiperparâmetro
Precision for region_date	22397.4615802	[1719.4226845 , 87467.4260820]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.7 : Diario.Unif~tm.mean.7+tm.mean.7+hM.mean.7+hM.mean.7+pr.mean.7+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.7

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	6.5387759	[-2.4022473 , 15.4888729]	Efeito Fixo
tm.mean.7	0.0761083	[0.0323488 , 0.1200678]	Efeito Fixo
tm.mean.7	-0.0701696	[-0.1356221 , -0.0057035]	Efeito Fixo
hM.mean.7	-0.0489120	[-0.0631545 , -0.0349193]	Efeito Fixo
hM.mean.7	0.0204130	[0.0098723 , 0.0309742]	Efeito Fixo
pr.mean.7	0.0054270	[-0.0058457 , 0.016704]	Efeito Fixo
Popdensity	0.0010531	[-0.0005629 , 0.0026692]	Efeito Fixo
GMM.T.2018	0.0000753	[-0.0093965 , 0.0095545]	Efeito Fixo
IDI.2019	-0.1752202	[-0.3671355 , 0.0166109]	Efeito Fixo
Mais.65	-0.0000029	[-9.54e-05 , 8.96e-05]	Efeito Fixo
Precision for Region_str (iid component)	0.3919700	[0.1717011 , 0.7286418]	Hiperparâmetro
Precision for Region_str (spatial component)	2484.7881263	[199.9637677 , 9643.2964736]	Hiperparâmetro
Precision for Date_str	13403.7853673	[1281.6674886 , 55705.0926058]	Hiperparâmetro
Precision for Date_unstr	0.7991030	[0.5639448 , 1.1022648]	Hiperparâmetro
Precision for region_date	28926.2405150	[2598.0807068 , 116679.6066787]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.7.HF : x.HF~tm.mean.7+tm.mean.7+hM.mean.7+hM.mean.7+pr.mean.7+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.7.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.4165858	[-4.6735999 , 13.50426]	Efeito Fixo
tm.mean.7	0.1017681	[0.0599476 , 0.1440796]	Efeito Fixo

tm.mean.7		-0.0645576		[-0.1220246 , -0.007645]		Efeito Fixo	
hM.mean.7		-0.0268093		[-0.0393812 , -0.0143629]		Efeito Fixo	
hm.mean.7		0.0177916		[0.0077542 , 0.0278787]		Efeito Fixo	
pr.mean.7		-0.0077288		[-0.0187551 , 0.003288]		Efeito Fixo	
Popdensity		0.0009915		[-0.0006551 , 0.0026382]		Efeito Fixo	
GMM.T.2018		0.0001209		[-0.0095289 , 0.0097747]		Efeito Fixo	
IDI.2019		-0.1739432		[-0.3694784 , 0.0215184]		Efeito Fixo	
Mais.65		0.0000002		[-9.41e-05 , 9.45e-05]		Efeito Fixo	
Precision for Region_str (iid component)		0.3830999		[0.1699155 , 0.7171146]		Hiperparâmetro	
Precision for Region_str (spatial component)		2509.9854236		[204.0836115 , 9943.0176061]		Hiperparâmetro	
Precision for Date_str		10747.1361587		[1643.3766138 , 35291.8160091]		Hiperparâmetro	
Precision for Date_unstr		1.3476319		[0.9618179 , 1.8224532]		Hiperparâmetro	
Precision for region_date		26291.9855663		[2282.4394376 , 105770.1903386]		Hiperparâmetro	

Modelo: iCAR.rw.iid.M2.7.LS : x.LS~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: iCAR.rw.iid.M2.7.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.4512073	[-4.6051146 , 13.5063427]	Efeito Fixo
tm.mean.7	0.1350301	[0.0926353 , 0.177836]	Efeito Fixo
hM.mean.7	-0.0821385	[-0.1403469 , -0.0244184]	Efeito Fixo
hm.mean.7	-0.0349623	[-0.0478854 , -0.0221757]	Efeito Fixo
pr.mean.7	0.0237061	[0.0136549 , 0.0338023]	Efeito Fixo
Popdensity	0.0006907	[-0.0104 , 0.0117811]	Efeito Fixo
GMM.T.2018	0.0010179	[-0.0006216 , 0.0026576]	Efeito Fixo
IDI.2019	0.0001158	[-0.0094935 , 0.0097289]	Efeito Fixo
Mais.65	-0.1796517	[-0.3743623 , 0.0149962]	Efeito Fixo
Precision for Region_str (iid component)	0.0000006	[-9.33e-05 , 9.45e-05]	Efeito Fixo
Precision for Region_str (spatial component)	0.3783672	[0.1649795 , 0.701174]	Hiperparâmetro
Precision for Date_str	2268.1183547	[178.0115058 , 8873.3255216]	Hiperparâmetro
Precision for Date_unstr	10708.3969269	[1606.5232573 , 36218.4089991]	Hiperparâmetro
Precision for region_date	1.1941827	[0.859059 , 1.6211574]	Hiperparâmetro
Precision for region_date	24510.0082914	[2043.5181474 , 94842.7324983]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.7.HF.14 : x.HF.14~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.rw.iid.M2.7.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	4.186080	[-5.0108439 , 13.3818733]	Efeito Fixo
tm.mean.7	0.059217	[0.0195601 , 0.0991600]	Efeito Fixo
hM.mean.7	-0.068102	[-0.1250159 , -0.0117313]	Efeito Fixo
hm.mean.7	-0.015583	[-0.0278218 , -0.0034254]	Efeito Fixo
pr.mean.7	0.014973	[0.0050752 , 0.0249074]	Efeito Fixo
Popdensity	-0.003613	[-0.0145611 , 0.0073366]	Efeito Fixo
GMM.T.2018	0.000980	[-0.0006863 , 0.0026467]	Efeito Fixo
IDI.2019	0.000435	[-0.0093307 , 0.0102050]	Efeito Fixo
Mais.65	-0.172932	[-0.3708089 , 0.0248872]	Efeito Fixo
Precision for Region_str (iid component)	-0.000000	[-0.0000957 , 0.0000951]	Efeito Fixo
Precision for Region_str (spatial component)	0.378286	[0.1696366 , 0.7173872]	Hiperparâmetro
Precision for Date_str	2284.367854	[172.5071519 , 8783.7619866]	Hiperparâmetro
Precision for Date_unstr	12084.470371	[1501.4546396 , 43588.9595698]	Hiperparâmetro
Precision for region_date	1.340232	[0.9640149 , 1.8098660]	Hiperparâmetro
Precision for region_date	23494.740377	[1900.7421667 , 91163.9889597]	Hiperparâmetro

Modelo: iCAR.rw.iid.M2.7.LS.14 : x.LS.14~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

```
f(Region_str, model="bym", graph=W.porto)+
f(Date_str, model="rw2") +
f(Date_unstr, model="iid")+
f(region_date, model="iid")
```

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: iCAR.rw.iid.M2.7.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	3.784427	[-5.4281037 , 12.9957796]	Efeito Fixo
tm.mean.7	0.069213	[0.0297247 , 0.1088760]	Efeito Fixo
hM.mean.7	-0.094672	[-0.1520502 , -0.0376013]	Efeito Fixo
hm.mean.7	-0.015965	[-0.0284245 , -0.0035383]	Efeito Fixo
pr.mean.7	0.019246	[0.0093957 , 0.0291230]	Efeito Fixo
Popdensity	-0.001466	[-0.0126668 , 0.0097386]	Efeito Fixo
GMM.T.2018	0.000975	[-0.0006931 , 0.0026424]	Efeito Fixo
IDI.2019	0.000916	[-0.0088575 , 0.0106939]	Efeito Fixo
Mais.65	-0.177624	[-0.3756719 , 0.0203595]	Efeito Fixo
Precision for Region_str (iid component)	-0.000001	[-0.0000961 , 0.0000949]	Efeito Fixo
Precision for Region_str (spatial component)	0.364077	[0.1572194 , 0.673003]	Hiperparâmetro
Precision for Date_str	2709.856559	[254.7797300 , 10489.168331]	Hiperparâmetro
Precision for Date_unstr	17643.163143	[1915.6475470 , 69866.555812]	Hiperparâmetro

```
| Precision for Date_unstr | 0.963608 | [ 0.6968467 , 1.294874 ] | Hiperparâmetro |
| Precision for region_date | 23469.696155 | [ 1907.5687527 , 91005.473767 ] | Hiperparâmetro |
```

Modelo: IID.M1.14 : Diario.Unif~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+ f(Date_unstr, model="iid")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.6217716	[-4.3755098 , -0.9165573]	Efeito Fixo
tm.mean.14	0.1122401	[0.0724301 , 0.1521294]	Efeito Fixo
tm.mean.14	0.0927411	[0.0402258 , 0.1454252]	Efeito Fixo
hM.mean.14	0.0027451	[-0.0108295 , 0.0164451]	Efeito Fixo
hm.mean.14	0.0943355	[0.0844278 , 0.104259]	Efeito Fixo
pr.mean.14	0.0032794	[-0.0091684 , 0.0157305]	Efeito Fixo
Popdensity	0.0008357	[0.0007871 , 0.0008843]	Efeito Fixo
GMM.T.2018	-0.0032752	[-0.0036558 , -0.0028945]	Efeito Fixo
IDI.2019	-0.1161480	[-0.1233575 , -0.1089395]	Efeito Fixo
Precision for Date_unstr	0.1307836	[0.0884248 , 0.1852877]	Hiperparâmetro

Modelo: IID.M1.14.HF : x.HF~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+ f(Date_unstr, model="iid")

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.14.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.0247971	[-3.6604182 , -0.4332724]	Efeito Fixo
tm.mean.14	0.0544522	[0.0167417 , 0.0922512]	Efeito Fixo
tm.mean.14	0.2026140	[0.1512865 , 0.254064]	Efeito Fixo
hM.mean.14	0.0083999	[-0.0041314 , 0.0210498]	Efeito Fixo
hm.mean.14	0.0872400	[0.0783428 , 0.0961505]	Efeito Fixo
pr.mean.14	0.0008314	[-0.0130171 , 0.0146895]	Efeito Fixo
Popdensity	0.0008454	[0.0007972 , 0.0008937]	Efeito Fixo
GMM.T.2018	-0.0037602	[-0.0041416 , -0.0033786]	Efeito Fixo
IDI.2019	-0.1135899	[-0.1207654 , -0.1064153]	Efeito Fixo
Precision for Date_unstr	0.1430861	[0.0981477 , 0.1999948]	Hiperparâmetro

Modelo: IID.M1.14.LS : x.LS~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019+ f(Date_unstr, model="iid")

Table: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.14.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.1905331	[-3.8192947 , -0.6048121]	Efeito Fixo
tm.mean.14	0.0889516	[0.051013 , 0.1269786]	Efeito Fixo
tm.mean.14	0.1792905	[0.1280399 , 0.2306694]	Efeito Fixo
hM.mean.14	0.0038650	[-0.0085574 , 0.0163953]	Efeito Fixo
hm.mean.14	0.0888477	[0.0799339 , 0.0977744]	Efeito Fixo
pr.mean.14	0.0033778	[-0.0105885 , 0.0173537]	Efeito Fixo
Popdensity	0.0008545	[0.0008059 , 0.000903]	Efeito Fixo
GMM.T.2018	-0.0036328	[-0.0040137 , -0.0032519]	Efeito Fixo
IDI.2019	-0.1175172	[-0.1247517 , -0.1102835]	Efeito Fixo
Precision for Date_unstr	0.1381697	[0.0950547 , 0.1926055]	Hiperparâmetro

Modelo: IID.M1.14.HF.14 : x.HF.14~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019 + f(Date_unstr, model="iid")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: IID.M1.14.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.4528327	[-4.1176149 , -0.8288761]	Efeito Fixo
tm.mean.14	0.0779928	[0.0396935 , 0.1163739]	Efeito Fixo
tm.mean.14	0.2295470	[0.1776149 , 0.2815734]	Efeito Fixo
hM.mean.14	0.0079534	[-0.0047715 , 0.0207740]	Efeito Fixo
hm.mean.14	0.0876593	[0.0785966 , 0.0967339]	Efeito Fixo
pr.mean.14	-0.0118190	[-0.0256318 , 0.0019997]	Efeito Fixo
Popdensity	0.0008427	[0.0007947 , 0.0008907]	Efeito Fixo
GMM.T.2018	-0.0039336	[-0.0043175 , -0.0035495]	Efeito Fixo
IDI.2019	-0.1117517	[-0.1188798 , -0.1046243]	Efeito Fixo
Precision for Date_unstr	0.1330538	[0.0910638 , 0.1865836]	Hiperparâmetro

Modelo: IID.M1.14.LS.14 : x.LS.14~tm.mean.14+tm.mean.14+hM.mean.14+hm.mean.14+pr.mean.14+Popdensity+ GMM.T.2018 + IDI.2019 + f(Date_unstr, model="iid")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: IID.M1.14.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.9018889	[-4.5411921 , -1.3012649]	Efeito Fixo
tm.mean.14	0.1098931	[0.0715867 , 0.1482682]	Efeito Fixo
tm.mean.14	0.2090595	[0.1575357 , 0.2606930]	Efeito Fixo
hM.mean.14	0.0043647	[-0.0082630 , 0.0170767]	Efeito Fixo

hm.mean.14	0.0922876	[0.0832065 , 0.1013795]	Efeito Fixo
pr.mean.14	-0.0053677	[-0.0191837 , 0.0084548]	Efeito Fixo
Popdensity	0.0008313	[0.0007835 , 0.0008791]	Efeito Fixo
GMM.T.2018	-0.0037702	[-0.0041524 , -0.0033880]	Efeito Fixo
IDI.2019	-0.1120192	[-0.1190975 , -0.1049414]	Efeito Fixo
Precision for Date_unstr	0.1286143	[0.0885015 , 0.1793917]	Hiperparâmetro

Modelo: IID.M1.2 : Diario.Unif~tm.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65
+
f(Date_unstr, model="iid")

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.2

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.3981217	[-4.2617222 , -0.5749727]	Efeito Fixo
tm.mean.2	-0.0241806	[-0.0555664 , 0.0071638]	Efeito Fixo
tm.mean.2	0.5412259	[0.480215 , 0.6027578]	Efeito Fixo
hM.mean.2	0.0152814	[0.000384 , 0.0303705]	Efeito Fixo
hm.mean.2	0.0493171	[0.0410111 , 0.0576267]	Efeito Fixo
pr.mean.2	-0.0287706	[-0.0424483 , -0.0151039]	Efeito Fixo
Popdensity	0.0009702	[0.0009225 , 0.0010179]	Efeito Fixo
GMM.T.2018	-0.0047017	[-0.0051419 , -0.0042621]	Efeito Fixo
IDI.2019	-0.1110409	[-0.1180595 , -0.1040223]	Efeito Fixo
Mais.65	-0.0000121	[-1.37e-05 , -1.05e-05]	Efeito Fixo
Precision for Date_unstr	0.0981664	[0.0685691 , 0.1346912]	Hiperparâmetro

Modelo: IID.M1.2.HF : x.HF~tm.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+
f(Date_unstr, model="iid")

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.2.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	0.0425994	[-1.675843 , 1.7135385]	Efeito Fixo
tm.mean.2	-0.0121032	[-0.0454409 , 0.0211524]	Efeito Fixo
tm.mean.2	0.4057353	[0.3486335 , 0.4637432]	Efeito Fixo
hM.mean.2	0.0021072	[-0.0119562 , 0.016549]	Efeito Fixo
hm.mean.2	0.0359774	[0.0276064 , 0.0443515]	Efeito Fixo
pr.mean.2	0.0002746	[-0.0130523 , 0.0135865]	Efeito Fixo
Popdensity	0.0009019	[0.0008547 , 0.000949]	Efeito Fixo
GMM.T.2018	-0.0036321	[-0.0040569 , -0.0032081]	Efeito Fixo
IDI.2019	-0.1123425	[-0.1193938 , -0.105291]	Efeito Fixo
Mais.65	-0.0000129	[-1.45e-05 , -1.13e-05]	Efeito Fixo
Precision for Date_unstr	0.1789175	[0.124311 , 0.2464152]	Hiperparâmetro

Modelo: IID.M1.2.LS : x.LS~tm.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+
f(Date_unstr, model="iid")

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.2.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-0.7827332	[-2.5768999 , 0.9646358]	Efeito Fixo
tm.mean.2	-0.0210335	[-0.0555093 , 0.0133534]	Efeito Fixo
tm.mean.2	0.4517346	[0.3929356 , 0.5113483]	Efeito Fixo
hM.mean.2	0.0092487	[-0.0054614 , 0.0243137]	Efeito Fixo
hm.mean.2	0.0378225	[0.0291958 , 0.0464528]	Efeito Fixo
pr.mean.2	-0.0071437	[-0.0204016 , 0.0060997]	Efeito Fixo
Popdensity	0.0009245	[0.000877 , 0.000972]	Efeito Fixo
GMM.T.2018	-0.0038543	[-0.0042837 , -0.0034256]	Efeito Fixo
IDI.2019	-0.1144130	[-0.1215174 , -0.1073087]	Efeito Fixo
Mais.65	-0.0000126	[-1.42e-05 , -1.1e-05]	Efeito Fixo
Precision for Date_unstr	0.1517332	[0.1054287 , 0.2090746]	Hiperparâmetro

Modelo: IID.M1.2.HF.14 : x.HF.14~tm.mean.2+tm.mean.2+hM.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+
f(Date_unstr, model="iid")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: IID.M1.2.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-0.7947212	[-2.4943297 , 0.8563968]	Efeito Fixo
tm.mean.2	0.0180394	[-0.0146261 , 0.0506558]	Efeito Fixo
tm.mean.2	0.4090527	[0.3519897 , 0.4670499]	Efeito Fixo
hM.mean.2	0.0056179	[-0.0081086 , 0.0197265]	Efeito Fixo
hm.mean.2	0.0344457	[0.0261747 , 0.0427220]	Efeito Fixo
pr.mean.2	-0.0023850	[-0.0154831 , 0.0106957]	Efeito Fixo
Popdensity	0.0008953	[0.0008483 , 0.0009423]	Efeito Fixo
GMM.T.2018	-0.0035646	[-0.0039901 , -0.0031400]	Efeito Fixo
IDI.2019	-0.1117066	[-0.1187119 , -0.1047011]	Efeito Fixo
Mais.65	-0.0000133	[-0.0000148 , -0.0000117]	Efeito Fixo
Precision for Date_unstr	0.190637	[0.1317942 , 0.2638037]	Hiperparâmetro

Modelo: IID.M1.2.LS.14 : x.LS.14~tm.mean.2+tm.mean.2+hM.mean.2+hM.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+

f(Date_unstr, model="iid")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: IID.M1.2.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.1681803	[-3.8806520 , -0.4969881]	Efeito Fixo
tm.mean.2	0.0065061	[-0.0269050 , 0.0398453]	Efeito Fixo
tm.mean.2	0.4481198	[0.3901168 , 0.5068353]	Efeito Fixo
hM.mean.2	0.0168862	[0.0030519 , 0.0310081]	Efeito Fixo
hM.mean.2	0.0390186	[0.0306915 , 0.0473522]	Efeito Fixo
pr.mean.2	-0.0088605	[-0.0218747 , 0.0041383]	Efeito Fixo
Popdensity	0.0008935	[0.0008467 , 0.0009404]	Efeito Fixo
GMM.T.2018	-0.0036277	[-0.0040546 , -0.0032015]	Efeito Fixo
IDI.2019	-0.1106398	[-0.1175919 , -0.1036878]	Efeito Fixo
Mais.65	-0.0000129	[-0.0000145 , -0.0000113]	Efeito Fixo
Precision for Date_unstr	0.1639197	[0.1132419 , 0.2272045]	Hiperparâmetro

Modelo: IID.M1.7 : Diario.Unif~tm.mean.7+tm.mean.7+hM.mean.7+hM.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+

f(Date_unstr, model="iid")

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.7

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	2.8813602	[1.2785016 , 4.439713]	Efeito Fixo
tm.mean.7	-0.0798302	[-0.1193083 , -0.0403143]	Efeito Fixo
tm.mean.7	0.4151326	[0.3571442 , 0.4735299]	Efeito Fixo
hM.mean.7	-0.0220112	[-0.0361164 , -0.0077561]	Efeito Fixo
hM.mean.7	0.0557410	[0.0462887 , 0.0652102]	Efeito Fixo
pr.mean.7	0.0033313	[-0.0084572 , 0.015128]	Efeito Fixo
Popdensity	0.0009468	[0.0008993 , 0.0009942]	Efeito Fixo
GMM.T.2018	-0.0044035	[-0.004484 , -0.0039681]	Efeito Fixo
IDI.2019	-0.1102390	[-0.1173157 , -0.1031625]	Efeito Fixo
Mais.65	-0.0000120	[-1.36e-05 , -1.04e-05]	Efeito Fixo
Precision for Date_unstr	0.1807921	[0.1272606 , 0.2457459]	Hiperparâmetro

Modelo: IID.M1.7.HF : x.HF~tm.mean.7+tm.mean.7+hM.mean.7+hM.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+

f(Date_unstr, model="iid")

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.7.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-1.6837712	[-3.4276129 , 0.0049635]	Efeito Fixo
tm.mean.7	-0.0051654	[-0.0436647 , 0.0333924]	Efeito Fixo
tm.mean.7	0.4460058	[0.3872629 , 0.5051738]	Efeito Fixo
hM.mean.7	0.0156367	[0.0007047 , 0.0308043]	Efeito Fixo
hM.mean.7	0.0394926	[0.0301702 , 0.0488305]	Efeito Fixo
pr.mean.7	-0.0177004	[-0.0295207 , -0.0058824]	Efeito Fixo
Popdensity	0.0008967	[0.0008498 , 0.0009436]	Efeito Fixo
GMM.T.2018	-0.0037326	[-0.0041566 , -0.0033093]	Efeito Fixo
IDI.2019	-0.1114200	[-0.1184929 , -0.1043479]	Efeito Fixo
Mais.65	-0.0000125	[-1.41e-05 , -1.09e-05]	Efeito Fixo
Precision for Date_unstr	0.1593880	[0.1105059 , 0.2201482]	Hiperparâmetro

Modelo: IID.M1.7.LS : x.LS~tm.mean.7+tm.mean.7+hM.mean.7+hM.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+

f(Date_unstr, model="iid")

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: IID.M1.7.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-1.6652181	[-3.4050588 , 0.0213354]	Efeito Fixo
tm.mean.7	0.0066116	[-0.0319309 , 0.0452112]	Efeito Fixo
tm.mean.7	0.4436525	[0.3852855 , 0.5024448]	Efeito Fixo
hM.mean.7	0.0124971	[-0.0024108 , 0.0276291]	Efeito Fixo
hM.mean.7	0.0418231	[0.0325928 , 0.0510682]	Efeito Fixo
pr.mean.7	-0.0115531	[-0.0234175 , 0.0003104]	Efeito Fixo
Popdensity	0.0009130	[0.0008658 , 0.0009602]	Efeito Fixo
GMM.T.2018	-0.0037783	[-0.0042022 , -0.0033552]	Efeito Fixo
IDI.2019	-0.1144640	[-0.1215851 , -0.1073436]	Efeito Fixo
Mais.65	-0.0000124	[-1.4e-05 , -1.08e-05]	Efeito Fixo
Precision for Date_unstr	0.1571680	[0.109439 , 0.2161393]	Hiperparâmetro

Modelo: IID.M1.7.HF.14 : x.HF.14~tm.mean.7+tm.mean.7+hM.mean.7+hM.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+

f(Date_unstr, model="iid")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: IID.M1.7.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-1.3468068	[-3.0892430 , 0.3336633]	Efeito Fixo
tm.mean.7	-0.0416647	[-0.0787371 , -0.0045326]	Efeito Fixo
tm.mean.7	0.4473951	[0.3888337 , 0.5064095]	Efeito Fixo
hM.mean.7	0.0204943	[0.0057634 , 0.0355106]	Efeito Fixo
hM.mean.7	0.0370907	[0.0279148 , 0.0462853]	Efeito Fixo
pr.mean.7	-0.0110667	[-0.0227525 , 0.0006190]	Efeito Fixo

Popdensity	0.0008846	[0.0008379 , 0.0009314]	Efeito Fixo
GMM.T.2018	-0.0036356	[-0.0040596 , -0.0032123]	Efeito Fixo
IDI.2019	-0.1094365	[-0.1164812 , -0.1023924]	Efeito Fixo
Mais.65	-0.0000127	[-0.0000143 , -0.0000111]	Efeito Fixo
Precision for Date_unstr	0.1824896	[0.1250191 , 0.2547809]	Hiperparâmetro

Modelo: IID.M1.7.LS.14 : x.LS.14~tm.mean.7+tm.mean.7+hM.mean.7+hM.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+

f(Date_unstr, model="iid")

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: IID.M1.7.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	-2.1957546	[-3.9580548 , -0.4852699]	Efeito Fixo
tm.mean.7	-0.0443520	[-0.0806168 , -0.0080353]	Efeito Fixo
tm.mean.7	0.4557336	[0.3981243 , 0.5136608]	Efeito Fixo
hM.mean.7	0.0271619	[0.0125321 , 0.0420047]	Efeito Fixo
hm.mean.7	0.0399314	[0.0309712 , 0.0489078]	Efeito Fixo
pr.mean.7	-0.0113279	[-0.0231547 , 0.0005003]	Efeito Fixo
Popdensity	0.0008760	[0.0008295 , 0.0009225]	Efeito Fixo
GMM.T.2018	-0.0035379	[-0.0039570 , -0.0031193]	Efeito Fixo
IDI.2019	-0.1090247	[-0.1160088 , -0.1020415]	Efeito Fixo
Mais.65	-0.0000125	[-0.0000141 , -0.0000109]	Efeito Fixo
Precision for Date_unstr	0.1565449	[0.1075009 , 0.218427]	Hiperparâmetro

Modelo: CAR.M1.14 : Diario.Unif~tm.mean.14+tm.mean.14+hM.mean.14+hM.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="besag",graph=W.porto)

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	13.7150437	[4.0535556 , 23.379]	Efeito Fixo
tm.mean.14	-0.0595348	[-0.0682389 , -0.0508308]	Efeito Fixo
tm.mean.14	-0.1297083	[-0.1418562 , -0.1175604]	Efeito Fixo
hM.mean.14	0.0194590	[0.0175173 , 0.0214006]	Efeito Fixo
hm.mean.14	0.0044140	[0.0021037 , 0.0067244]	Efeito Fixo
pr.mean.14	0.0182545	[0.0150438 , 0.0214652]	Efeito Fixo
Popdensity	0.0013961	[-0.0001176 , 0.0029102]	Efeito Fixo
GMM.T.2018	-0.0115186	[-0.0223304 , -0.0007084]	Efeito Fixo
IDI.2019	-0.1141222	[-0.2726256 , 0.0443129]	Efeito Fixo
Mais.65	0.0000403	[-6.29e-05 , 0.0001435]	Efeito Fixo
Precision for Region_unstr	0.1300697	[0.0544881 , 0.2388768]	Hiperparâmetro

Modelo: CAR.M1.14.HF : x.HF~tm.mean.14+tm.mean.14+hM.mean.14+hM.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="besag",graph=W.porto)

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.14.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	13.3604166	[3.6548805 , 23.0683407]	Efeito Fixo
tm.mean.14	-0.0329767	[-0.0415764 , -0.0243769]	Efeito Fixo
tm.mean.14	-0.1452580	[-0.1569931 , -0.133523]	Efeito Fixo
hM.mean.14	0.0201709	[0.0181831 , 0.0221587]	Efeito Fixo
hm.mean.14	0.0048443	[0.0025419 , 0.0071467]	Efeito Fixo
pr.mean.14	0.0092216	[0.0057102 , 0.0127331]	Efeito Fixo
Popdensity	0.0014000	[-0.0001206 , 0.002921]	Efeito Fixo
GMM.T.2018	-0.0115111	[-0.022372 , -0.0006517]	Efeito Fixo
IDI.2019	-0.1158231	[-0.2750469 , 0.0433323]	Efeito Fixo
Mais.65	0.0000411	[-6.26e-05 , 0.0001447]	Efeito Fixo
Precision for Region_unstr	0.1288793	[0.0539895 , 0.2366876]	Hiperparâmetro

Modelo: CAR.M1.14.LS : x.LS~tm.mean.14+tm.mean.14+hM.mean.14+hM.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="besag",graph=W.porto)

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.14.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	13.3595906	[3.6767505 , 23.0448614]	Efeito Fixo
tm.mean.14	-0.0393934	[-0.048016 , -0.0307707]	Efeito Fixo
tm.mean.14	-0.1366516	[-0.148447 , -0.1248563]	Efeito Fixo
hM.mean.14	0.0208175	[0.0188318 , 0.0228031]	Efeito Fixo
hm.mean.14	0.0041957	[0.0018954 , 0.0064961]	Efeito Fixo
pr.mean.14	0.0092440	[0.00577 , 0.0127179]	Efeito Fixo
Popdensity	0.0014011	[-0.000116 , 0.0029185]	Efeito Fixo
GMM.T.2018	-0.0114808	[-0.0223163 , -0.0006468]	Efeito Fixo
IDI.2019	-0.1171984	[-0.2760568 , 0.0415869]	Efeito Fixo
Mais.65	0.0000415	[-6.19e-05 , 0.0001449]	Efeito Fixo
Precision for Region_unstr	0.1294616	[0.0542314 , 0.2377637]	Hiperparâmetro

Modelo: CAR.M1.14.HF.14 : x.HF.14~tm.mean.14+tm.mean.14+hM.mean.14+hM.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019 +Mais.65+f(Region_unstr,model="besag",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: CAR.M1.14.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	13.3972228	[3.6772179 , 23.1197152]	Efeito Fixo

	tm.mean.14	-0.0315842	[-0.0402284 , -0.0229401]	Efeito Fixo	
	tm.mean.14	-0.1313260	[-0.1430525 , -0.1195995]	Efeito Fixo	
	hm.mean.14	0.0195818	[0.0175803 , 0.0215833]	Efeito Fixo	
	hm.mean.14	0.0034517	[0.0011524 , 0.0057510]	Efeito Fixo	
	pr.mean.14	0.0108255	[0.0073927 , 0.0142582]	Efeito Fixo	
	Popdensity	0.0014092	[-0.0001137 , 0.0029324]	Efeito Fixo	
	GMM.T.2018	-0.0115953	[-0.0224725 , -0.0007197]	Efeito Fixo	
	IDI.2019	-0.1145460	[-0.2740054 , 0.0448442]	Efeito Fixo	
	Mais.65	0.0000402	[-0.0000636 , 0.0001440]	Efeito Fixo	
	Precision for Region_unstr	0.1285072	[0.0538339 , 0.2360055]	Hiperparâmetro	

Modelo: CAR.M1.14.LS.14 : x.LS.14~tm.mean.14+tm.mean.14+hm.mean.14+hm.mean.14+pr.mean.14+Popdensity+GMM.T.2018 + IDI.2019+Mais.65+f(Region_unstr,model="besag",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: CAR.M1.14.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	13.2948445	[3.5776639 , 23.0143026]	Efeito Fixo
tm.mean.14	-0.0347637	[-0.0434384 , -0.0260891]	Efeito Fixo
tm.mean.14	-0.1052915	[-0.1169860 , -0.0935970]	Efeito Fixo
hm.mean.14	0.0191864	[0.0171816 , 0.0211913]	Efeito Fixo
hm.mean.14	0.0008836	[-0.0014020 , 0.0031691]	Efeito Fixo
pr.mean.14	0.0114747	[0.0081394 , 0.0148100]	Efeito Fixo
Popdensity	0.0013833	[-0.0001391 , 0.0029061]	Efeito Fixo
GMM.T.2018	-0.0114871	[-0.0223608 , -0.0006147]	Efeito Fixo
IDI.2019	-0.1141553	[-0.2735735 , 0.0451887]	Efeito Fixo
Mais.65	0.0000408	[-0.0000629 , 0.0001446]	Efeito Fixo
Precision for Region_unstr	0.1285186	[0.0538376 , 0.2360233]	Hiperparâmetro

Modelo: CAR.M1.2 : Diario.Unif~tm.mean.2+tm.mean.2+hm.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65 + f(Region_unstr,model="besag",graph=W.porto)

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.2

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	11.7654868	[1.9156703 , 21.6173258]	Efeito Fixo
tm.mean.2	-0.0024689	[-0.0082465 , 0.0033088]	Efeito Fixo
tm.mean.2	-0.1484756	[-0.158113 , -0.1388382]	Efeito Fixo
hm.mean.2	0.0328524	[0.0299152 , 0.0357897]	Efeito Fixo
hm.mean.2	0.0054923	[0.0032978 , 0.0076869]	Efeito Fixo
pr.mean.2	0.0210513	[0.0181078 , 0.0239949]	Efeito Fixo
Popdensity	0.0013977	[-0.0001449 , 0.0029407]	Efeito Fixo
GMM.T.2018	-0.0117302	[-0.0227484 , -0.0007134]	Efeito Fixo
IDI.2019	-0.1132618	[-0.2747884 , 0.0481994]	Efeito Fixo
Mais.65	0.0000429	[-6.22e-05 , 0.0001481]	Efeito Fixo
Precision for Region_unstr	0.1252022	[0.0524531 , 0.2299226]	Hiperparâmetro

Modelo: CAR.M1.2.HF : x.HF~tm.mean.2+tm.mean.2+hm.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+ f(Region_unstr,model="besag",graph=W.porto)

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.2.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.3672459	[2.628483 , 22.1081451]	Efeito Fixo
tm.mean.2	0.0012866	[-0.0045703 , 0.0071436]	Efeito Fixo
tm.mean.2	-0.1613723	[-0.1712479 , -0.1514968]	Efeito Fixo
hm.mean.2	0.0224639	[0.0199405 , 0.0249873]	Efeito Fixo
hm.mean.2	0.0167264	[0.0145143 , 0.0189385]	Efeito Fixo
pr.mean.2	0.0209485	[0.0180595 , 0.0238375]	Efeito Fixo
Popdensity	0.0013962	[-0.0001292 , 0.002922]	Efeito Fixo
GMM.T.2018	-0.0118842	[-0.0227793 , -0.0009906]	Efeito Fixo
IDI.2019	-0.1136222	[-0.2733457 , 0.0460352]	Efeito Fixo
Mais.65	0.0000438	[-6.02e-05 , 0.0001478]	Efeito Fixo
Precision for Region_unstr	0.1280591	[0.0536489 , 0.2351727]	Hiperparâmetro

Modelo: CAR.M1.2.LS : x.LS~tm.mean.2+tm.mean.2+hm.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+ f(Region_unstr,model="besag",graph=W.porto)

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.2.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.3072675	[2.5862754 , 22.0304435]	Efeito Fixo
tm.mean.2	0.0018313	[-0.00404 , 0.0077026]	Efeito Fixo
tm.mean.2	-0.1654457	[-0.1753482 , -0.1555431]	Efeito Fixo
hm.mean.2	0.0224930	[0.0199514 , 0.0250346]	Efeito Fixo
hm.mean.2	0.0179819	[0.015761 , 0.0202027]	Efeito Fixo
pr.mean.2	0.0187039	[0.0157613 , 0.0216464]	Efeito Fixo
Popdensity	0.0013984	[-0.0001242 , 0.0029214]	Efeito Fixo
GMM.T.2018	-0.0118314	[-0.0227065 , -0.0009578]	Efeito Fixo
IDI.2019	-0.1154177	[-0.2748557 , 0.0439495]	Efeito Fixo
Mais.65	0.0000443	[-5.95e-05 , 0.0001482]	Efeito Fixo
Precision for Region_unstr	0.1285293	[0.0538425 , 0.2360482]	Hiperparâmetro

Modelo: CAR.M1.2.HF.14 : x.HF.14~tm.mean.2+tm.mean.2+hm.mean.2+hm.mean.2+pr.mean.2+Popdensity+ GMM.T.2018 + IDI.2019+Mais.65+ f(Region_unstr,model="besag",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: CAR.M1.2.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.4436300	[2.6695240 , 22.2199586]	Efeito Fixo
tM.mean.2	0.0011760	[-0.0046397 , 0.0069918]	Efeito Fixo
tm.mean.2	-0.1482800	[-0.1580276 , -0.1385323]	Efeito Fixo
hM.mean.2	0.0225792	[0.0200973 , 0.0250611]	Efeito Fixo
hm.mean.2	0.0131998	[0.0110114 , 0.0153881]	Efeito Fixo
pr.mean.2	0.0233240	[0.0204909 , 0.0261571]	Efeito Fixo
Popdensity	0.0014059	[-0.0001251 , 0.0029372]	Efeito Fixo
GMM.T.2018	-0.0119395	[-0.0228743 , -0.0010062]	Efeito Fixo
IDI.2019	-0.1124341	[-0.2727371 , 0.0478020]	Efeito Fixo
Mais.65	0.0000427	[-0.0000616 , 0.0001471]	Efeito Fixo
Precision for Region_unstr	0.127152	[0.0532678 , 0.2335123]	Hiperparâmetro

Modelo: CAR.M1.2.LS.14 : $x.LS.14 \sim tM.mean.2 + tm.mean.2 + hM.mean.2 + hm.mean.2 + pr.mean.2 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 +$

$f(Region_unstr, model="besag", graph=W.porto)$

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: CAR.M1.2.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.3070795	[2.5240814 , 22.0921009]	Efeito Fixo
tM.mean.2	0.0030877	[-0.0026752 , 0.0088506]	Efeito Fixo
tm.mean.2	-0.1355950	[-0.1451648 , -0.1260252]	Efeito Fixo
hM.mean.2	0.0220791	[0.0196502 , 0.0245079]	Efeito Fixo
hm.mean.2	0.0112511	[0.0090933 , 0.0134089]	Efeito Fixo
pr.mean.2	0.0222435	[0.0194057 , 0.0250813]	Efeito Fixo
Popdensity	0.0013815	[-0.0001509 , 0.0029142]	Efeito Fixo
GMM.T.2018	-0.0117827	[-0.0227273 , -0.0008393]	Efeito Fixo
IDI.2019	-0.1125890	[-0.2730448 , 0.0477950]	Efeito Fixo
Mais.65	0.0000434	[-0.0000610 , 0.0001479]	Efeito Fixo
Precision for Region_unstr	0.1268703	[0.0531492 , 0.2329921]	Hiperparâmetro

Modelo: CAR.M1.7: $Diario.Unif \sim tM.mean.7 + tm.mean.7 + hM.mean.7 + hm.mean.7 + pr.mean.7 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 +$

$f(Region_unstr, model="besag", graph=W.porto)$

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.7

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.8030077	[3.1617998 , 22.4467863]	Efeito Fixo
tM.mean.7	-0.0590067	[-0.0681519 , -0.0498615]	Efeito Fixo
tm.mean.7	-0.0848371	[-0.0981548 , -0.0715193]	Efeito Fixo
hM.mean.7	0.0272365	[0.0245651 , 0.0299079]	Efeito Fixo
hm.mean.7	0.0046321	[0.0023228 , 0.0069414]	Efeito Fixo
pr.mean.7	0.0076745	[0.0043823 , 0.0109667]	Efeito Fixo
Popdensity	0.0014070	[-0.0001033 , 0.0029176]	Efeito Fixo
GMM.T.2018	-0.0117354	[-0.0225226 , -0.0009499]	Efeito Fixo
IDI.2019	-0.1121377	[-0.270278 , 0.045939]	Efeito Fixo
Mais.65	0.0000406	[-6.24e-05 , 0.0001435]	Efeito Fixo
Precision for Region_unstr	0.1306365	[0.0547263 , 0.2399095]	Hiperparâmetro

Modelo: CAR.M1.7.HF : $x.HF \sim tM.mean.7 + tm.mean.7 + hM.mean.7 + hm.mean.7 + pr.mean.7 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 + f(Region_unstr, model="besag", graph=W.porto)$

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.7.HF

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.8509601	[3.1988011 , 22.5056455]	Efeito Fixo
tM.mean.7	-0.0505574	[-0.0596635 , -0.0414514]	Efeito Fixo
tm.mean.7	-0.0739708	[-0.0871862 , -0.0607554]	Efeito Fixo
hM.mean.7	0.0272947	[0.0246482 , 0.0299412]	Efeito Fixo
hm.mean.7	-0.0014709	[-0.0037761 , 0.0008343]	Efeito Fixo
pr.mean.7	0.0144444	[0.0112483 , 0.0176405]	Efeito Fixo
Popdensity	0.0014042	[-0.0001078 , 0.0029165]	Efeito Fixo
GMM.T.2018	-0.0117590	[-0.0225585 , -0.0009612]	Efeito Fixo
IDI.2019	-0.1126016	[-0.2709223 , 0.0456565]	Efeito Fixo
Mais.65	0.0000411	[-6.2e-05 , 0.0001441]	Efeito Fixo
Precision for Region_unstr	0.1303368	[0.0546018 , 0.2393541]	Hiperparâmetro

Modelo: CAR.M1.7.LS : $x.LS \sim tM.mean.7 + tm.mean.7 + hM.mean.7 + hm.mean.7 + pr.mean.7 + Popdensity + GMM.T.2018 + IDI.2019 + Mais.65 + f(Region_unstr, model="besag", graph=W.porto)$

Tabela: Dados dos Efeitos Fixos e Hiperparâmetros - Modelo: CAR.M1.7.LS

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.7244730	[3.084182 , 22.367323]	Efeito Fixo
tM.mean.7	-0.0458606	[-0.0549735 , -0.0367478]	Efeito Fixo
tm.mean.7	-0.0768092	[-0.0899985 , -0.0636199]	Efeito Fixo
hM.mean.7	0.0278246	[0.0251309 , 0.0305183]	Efeito Fixo
hm.mean.7	-0.0020048	[-0.0043132 , 0.0003035]	Efeito Fixo
pr.mean.7	0.0153327	[0.012136 , 0.0185295]	Efeito Fixo
Popdensity	0.0014053	[-0.0001049 , 0.0029158]	Efeito Fixo
GMM.T.2018	-0.0116947	[-0.0224807 , -0.0009103]	Efeito Fixo
IDI.2019	-0.1142752	[-0.2724054 , 0.043788]	Efeito Fixo
Mais.65	0.0000415	[-6.14e-05 , 0.0001445]	Efeito Fixo

|Precision for Region_unstr | 0.1306612 | [0.0547365 , 0.2399561] | Hiperparâmetro |

Modelo: CAR.M1.7.HF.14 : x.HF.14~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais .65+f(Region_unstr,model="besag",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: CAR.M1.7.HF.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.8483362	[3.1674771 , 22.5318211]	Efeito Fixo
tm.mean.7	-0.0547266	[-0.0637667 , -0.0456865]	Efeito Fixo
tm.mean.7	-0.0727508	[-0.0859190 , -0.0595826]	Efeito Fixo
hM.mean.7	0.0284057	[0.0257643 , 0.0310472]	Efeito Fixo
hm.mean.7	-0.0017086	[-0.0040036 , 0.0005865]	Efeito Fixo
pr.mean.7	0.0118756	[0.0086358 , 0.0151154]	Efeito Fixo
Popdensity	0.0014143	[-0.0001023 , 0.0029311]	Efeito Fixo
GMM.T.2018	-0.0117972	[-0.0226289 , -0.0009673]	Efeito Fixo
IDI.2019	-0.1113400	[-0.2701296 , 0.0473857]	Efeito Fixo
Mais.65	0.0000402	[-0.0000632 , 0.0001436]	Efeito Fixo
Precision for Region_unstr	0.1295649	[0.0542782 , 0.2379391]	Hiperparâmetro

Modelo: CAR.M1.7.LS.14 : x.LS.14~tm.mean.7+tm.mean.7+hM.mean.7+hm.mean.7+pr.mean.7+Popdensity+GMM.T.2018 + IDI.2019 +Mais .65+f(Region_unstr,model="besag",graph=W.porto)

Tabela de Dados de efeitos fixos e hiperparâmetros- Modelo: CAR.M1.7.LS.14

Variável	Média	Intervalo de Credibilidade	Tipo
(Intercept)	12.7908918	[3.0921768 , 22.4920325]	Efeito Fixo
tm.mean.7	-0.0550017	[-0.0639620 , -0.0460414]	Efeito Fixo
tm.mean.7	-0.0662236	[-0.0792441 , -0.0532032]	Efeito Fixo
hM.mean.7	0.0276974	[0.0251190 , 0.0302758]	Efeito Fixo
hm.mean.7	-0.0029144	[-0.0051951 , -0.0006337]	Efeito Fixo
pr.mean.7	0.0095584	[0.0062776 , 0.0128392]	Efeito Fixo
Popdensity	0.0013889	[-0.0001305 , 0.0029085]	Efeito Fixo
GMM.T.2018	-0.0116387	[-0.0224902 , -0.0007885]	Efeito Fixo
IDI.2019	-0.1113626	[-0.2704526 , 0.0476584]	Efeito Fixo
Mais.65	0.0000410	[-0.0000625 , 0.0001446]	Efeito Fixo
Precision for Region_unstr	0.1289764	[0.054029 , 0.2368549]	Hiperparâmetro

ANEXO VI

MATRIZ_CORRELAÇÕES - CAPÍTULO 5

	tm.min.2	tm.min.2	hm.min.2	hm.min.2	pr.min.2	tm.max.2	tm.max.2	hm.max.2	hm.max.2	pr.max.2
tm.min.2	1.00	0.59	-0.34	-0.39	-0.33	0.98	0.60	-0.34	-0.36	-0.34
tm.min.2	0.59	1.00	-0.16	0.17	0.05	0.57	0.98	-0.21	0.20	0.04
hm.min.2	-0.34	-0.16	1.00	0.59	0.24	-0.33	-0.21	0.94	0.60	0.26
hm.min.2	-0.39	0.17	0.59	1.00	0.39	-0.40	0.14	0.52	0.97	0.40
pr.min.2	-0.33	0.05	0.24	0.39	1.00	-0.34	0.04	0.20	0.36	0.97
tm.max.2	0.98	0.57	-0.33	-0.40	-0.34	1.00	0.59	-0.32	-0.36	-0.35
tm.max.2	0.60	0.98	-0.21	0.14	0.04	0.59	1.00	-0.24	0.18	0.03
hm.max.2	-0.34	-0.21	0.94	0.52	0.20	-0.32	-0.24	1.00	0.54	0.22
hm.max.2	-0.36	0.20	0.60	0.97	0.36	-0.36	0.18	0.54	1.00	0.37
pr.max.2	-0.34	0.04	0.26	0.40	0.97	-0.35	0.03	0.22	0.37	1.00
tM.mean.2	0.99	0.58	-0.34	-0.39	-0.34	1.00	0.60	-0.33	-0.36	-0.35
tm.mean.2	0.59	0.99	-0.19	0.16	0.04	0.58	1.00	-0.23	0.19	0.04
hm.mean.2	-0.34	-0.18	0.99	0.56	0.23	-0.32	-0.23	0.98	0.57	0.24
hm.mean.2	-0.37	0.19	0.60	0.99	0.37	-0.38	0.16	0.53	0.99	0.38
pr.mean.2	-0.34	0.04	0.25	0.40	0.99	-0.35	0.04	0.21	0.36	0.99
tM.mean.7	0.49	0.44	-0.16	-0.15	-0.31	0.49	0.43	-0.12	-0.12	-0.30
tm.mean.7	0.40	0.50	-0.09	-0.01	-0.13	0.39	0.49	-0.09	0.02	-0.11
hm.mean.7	-0.07	-0.10	0.10	0.04	0.09	-0.06	-0.09	0.10	0.03	0.10
hm.mean.7	-0.09	0.02	0.10	0.17	0.06	-0.09	0.03	0.06	0.17	0.09
pr.mean.7	-0.13	-0.10	0.13	0.13	0.07	-0.13	-0.11	0.11	0.13	0.07
tM.mean.14	0.24	0.23	-0.21	-0.22	-0.09	0.24	0.23	-0.14	-0.23	-0.10
tm.mean.14	0.19	0.30	-0.13	-0.06	-0.10	0.19	0.30	-0.07	-0.05	-0.10
hm.mean.14	0.04	-0.06	0.22	0.04	-0.16	0.04	-0.06	0.21	0.04	-0.13
hm.mean.14	-0.05	-0.08	0.16	0.14	-0.07	-0.06	-0.06	0.15	0.14	-0.04
pr.mean.14	-0.10	-0.05	0.09	0.11	0.06	-0.11	-0.05	0.07	0.11	0.07

	tM.mean.2	tm.mean.2	hm.mean.2	hm.mean.2	pr.mean.2	tM.mean.7	tm.mean.7	hm.mean.7	hm.mean.7	
tM.min.2	0.99	0.59	-0.34	-0.37	-0.34	0.49	0.40	-0.07	-0.07	-0.09
tm.min.2	0.58	0.99	-0.18	0.19	0.04	0.44	0.50	-0.10	0.10	0.02
hm.min.2	-0.34	-0.19	0.99	0.60	0.25	-0.16	-0.09	0.10	0.10	0.10
hm.min.2	-0.39	0.16	0.56	0.99	0.40	-0.15	-0.01	0.04	0.04	0.17
pr.min.2	-0.34	0.04	0.23	0.37	0.99	-0.31	-0.13	0.09	0.09	0.06
tM.max.2	1.00	0.58	-0.32	-0.38	-0.35	0.49	0.39	-0.06	-0.09	-0.09
tm.max.2	0.60	1.00	-0.23	0.16	0.04	0.43	0.49	-0.09	0.10	0.03
hm.max.2	-0.33	-0.23	0.98	0.53	0.21	-0.12	-0.09	0.03	0.03	0.06
hm.max.2	-0.36	0.19	0.57	0.99	0.36	-0.12	0.02	0.03	0.03	0.17
pr.max.2	-0.35	0.04	0.24	0.38	0.99	-0.30	-0.11	0.10	0.10	0.09
tM.mean.2	1.00	0.59	-0.34	-0.38	-0.34	0.49	0.40	-0.07	-0.07	-0.09
tm.mean.2	0.59	1.00	-0.21	0.18	0.04	0.44	0.50	-0.10	0.03	0.03
hm.mean.2	-0.34	-0.21	1.00	0.57	0.23	-0.14	-0.09	0.11	0.11	0.08
hm.mean.2	-0.38	0.18	0.57	1.00	0.38	-0.14	0.01	0.03	0.03	0.17
pr.mean.2	-0.34	0.04	0.23	0.38	1.00	-0.31	-0.12	0.09	0.09	0.08
tM.mean.7	0.49	0.44	-0.14	-0.14	-0.31	1.00	0.55	-0.31	-0.31	-0.41
tm.mean.7	0.40	0.50	-0.09	0.01	-0.12	0.55	1.00	-0.08	-0.08	0.31
hm.mean.7	-0.07	-0.10	0.11	0.03	0.09	-0.31	-0.08	1.00	0.56	0.56
hm.mean.7	-0.09	0.03	0.08	0.17	0.08	-0.41	0.31	0.56	1.00	1.00
pr.mean.7	-0.13	-0.11	0.12	0.13	0.08	-0.37	0.12	0.28	0.28	0.44
tM.mean.14	0.24	0.23	-0.18	-0.23	-0.09	0.29	0.34	-0.14	-0.14	-0.10
tm.mean.14	0.19	0.30	-0.10	-0.05	-0.10	0.33	0.37	-0.08	-0.08	-0.06
hm.mean.14	0.04	-0.06	0.22	0.04	-0.15	-0.05	-0.17	0.10	0.10	0.02
hm.mean.14	-0.06	-0.07	0.15	0.15	-0.05	-0.05	-0.06	0.04	0.04	0.06
pr.mean.14	-0.11	-0.05	0.08	0.11	0.07	-0.03	-0.11	0.01	0.01	-0.01

	pr.mean.7	tM.mean.14	tm.mean.14	hm.mean.14	hm.mean.14	pr.mean.14	
tM.min.2	-0.13	0.24	0.19	0.04	-0.05	-0.10	
tm.min.2	-0.10	0.23	0.30	-0.06	-0.08	-0.05	
hm.min.2	0.13	-0.21	-0.13	0.22	0.16	0.09	
hm.min.2	0.13	-0.22	-0.06	0.04	0.14	0.11	
pr.min.2	0.07	-0.09	-0.10	-0.16	-0.07	0.06	
tM.max.2	-0.13	0.24	0.19	0.04	-0.06	-0.11	
tm.max.2	-0.11	0.23	0.30	-0.06	-0.07	-0.05	
hm.max.2	0.11	-0.14	-0.07	0.21	0.15	0.07	
hm.max.2	0.13	-0.23	-0.05	0.04	0.14	0.11	
pr.max.2	0.07	-0.10	-0.10	-0.13	-0.04	0.07	
tM.mean.2	-0.13	0.24	0.19	0.04	-0.06	-0.11	
tm.mean.2	-0.11	0.23	0.30	-0.06	-0.07	-0.05	
hm.mean.2	0.12	-0.18	-0.10	0.22	0.15	0.08	
hm.mean.2	0.13	-0.23	-0.05	0.04	0.15	0.11	
pr.mean.2	0.08	-0.09	-0.10	-0.15	-0.05	0.07	
tM.mean.7	-0.37	0.29	0.33	-0.05	-0.05	-0.03	
tm.mean.7	0.12	0.34	0.37	-0.17	-0.06	-0.11	
hm.mean.7	0.28	-0.14	-0.08	0.10	0.04	0.01	
hm.mean.7	0.44	-0.10	-0.06	0.02	0.06	-0.01	
pr.mean.7	1.00	-0.04	-0.09	0.04	0.04	-0.05	
tM.mean.14	-0.04	1.00	0.57	-0.28	-0.42	-0.34	
tm.mean.14	-0.09	0.57	1.00	-0.03	0.29	0.14	
hm.mean.14	0.04	-0.28	-0.03	1.00	0.55	0.26	
hm.mean.14	0.04	-0.42	0.29	0.55	1.00	0.44	
pr.mean.14	-0.05	-0.34	0.14	0.26	0.44	1.00	

ANEXO VII

RMSE E MAE DOS MODELOS AJUSTADOS NO CAPÍTULO 5

Root Mean Square Error e Mean Absolute Error

Modelo	Algoritmo HF		Algoritmo LS		Dados Observados	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
M0	258.484	32.523	299.142	35.120	136.583	26.888
MG	514.532	50.190	653.854	57.692	367.259	57.898
RW.M1.2	10.645	4.685	10.819	4.715	12.131	4.817
RW.M1.2.14	10.971	4.772	11.237	4.894		
RW.M1.7.	10.539	4.713	10.742	4.727	12.124	4.817
RW.M1.7.14	11.033	4.829	11.581	4.978		
RW2.M1.14	10.668	4.717	10.900	4.733	12.078	4.715
RW2.M1.14.14	11.187	4.834	11.670	4.958		
iCAR.M1.14	23.730	10.693	23.727	10.681	25.179	11.142
iCAR.M1.14.14	23.734	10.803	24.402	11.117		
iCAR.M1.2	22.239	10.473	21.996	10.398	24.373	11.198
iCAR.M1.2.14	22.646	10.606	23.301	10.874		
iCAR.M1.7.	23.598	10.822	23.483	10.769	24.772	11.292
iCAR.M1.7.14	23.666	10.843	24.283	11.070		
iCAR.rw.iid.M2.14	276.383	33.834	318.847	36.632	151.638	28.569
iCAR.rw.iid.M2.14.14	321.000	36.496	430.123	44.507		
iCAR.rw.iid.M2.2	329.380	35.734	388.181	38.983	157.967	30.264
iCAR.rw.iid.M2.2.14	401.745	39.374	525.337	47.127		
iCAR.rw.iid.M2.7	295.830	34.294	359.409	37.922	145.537	28.012
iCAR.rw.iid.M2.7.14	331.024	36.601	444.002	44.445		
IID.M1.14	10.865	4.758	11.125	4.780	12.082	4.726
IID.M1.14.14	11.418	4.876	11.862	4.996		
IID.M1.2	10.696	4.735	10.797	4.755	12.145	4.836
IID.M1.2.14	10.971	4.816	11.154	4.930		
IID.M1.7	11.003	4.796	11.241	4.809	12.130	4.829
IID.M1.7	11.003	4.796	11.241	4.809		
IID.M1.7.14	11.568	4.918	12.093	5.056		
CAR.M1.14	23.729	10.692	23.726	10.681	25.178	11.141
CAR.M1.14.14	23.733	10.802	24.402	11.117		
CAR.M1.2	22.238	10.472	21.995	10.398	24.372	11.197
CAR.M1.2.14	22.645	10.606	23.300	10.874		
CAR.M1.7	23.597	10.821	23.482	10.768	24.771	11.291
CAR.M1.7.14	23.665	10.843	24.282	11.069		

Root Mean Square Error e Mean Absolute Error

normalizados pelo desvio padrão

Modelo	Algoritmo HF		Algoritmo LS			
	MRME	MAE	MRME	MAE	MRME	MAE
M0	20.737	2.649	23.992	2.861	10.927	2.151
MG	41.214	4.059	52.361	4.663	29.380	4.632
RW.M1.2	1.109	0.448	1.129	0.456	0.970	0.385
RW.M1.2.14	1.131	0.454	1.156	0.467		
RW.M1.7	1.103	0.449	1.125	0.457	0.970	0.385
RW.M1.7.14	1.136	0.458	1.179	0.473		
RW2.M1.14	1.114	0.447	1.137	0.454	0.966	0.377
RW2.M1.14.14	1.148	0.455	1.188	0.468		
iCAR.M1.14	2.018	0.901	2.015	0.900	2.014	0.891
iCAR.M1.14.14	2.017	0.910	2.065	0.935		
iCAR.M1.2	1.903	0.886	1.888	0.880	1.950	0.896
iCAR.M1.2.14	1.933	0.896	1.985	0.918		
iCAR.M1.7	2.002	0.913	1.995	0.909	1.982	0.903
iCAR.M1.7.14	2.007	0.914	2.054	0.932		
iCAR.rw.iid.M2.14	22.168	2.753	25.568	2.981	12.131	2.285
iCAR.rw.iid.M2.14.14	25.734	2.966	34.464	3.609		
iCAR.rw.iid.M2.2	26.406	2.907	31.112	3.171	12.637	2.421
iCAR.rw.iid.M2.2.14	32.192	3.197	42.080	3.819		
iCAR.rw.iid.M2.7	23.722	2.790	28.810	3.084	11.643	2.241
iCAR.rw.iid.M2.7.14	26.535	2.975	35.574	3.605		
IID.M1.14	1.128	0.451	1.153	0.458	0.967	0.378
IID.M1.14.14	1.164	0.459	1.201	0.472		
IID.M1.2	1.112	0.450	1.127	0.458	0.972	0.387
IID.M1.2.14	1.131	0.456	1.151	0.468		
IID.M1.7	1.134	0.456	1.157	0.464	0.970	0.386
IID.M1.7.14	1.171	0.465	1.213	0.480		
CAR.M1.14	2.018	0.901	2.015	0.900	2.014	0.891
CAR.M1.14.14	2.017	0.910	2.065	0.935		
CAR.M1.2	1.903	0.886	1.887	0.880	1.950	0.896
CAR.M1.2.14	1.933	0.896	1.985	0.918		
CAR.M1.7	2.002	0.913	1.995	0.908	1.982	0.903
CAR.M1.7.14	2.007	0.914	2.054	0.932		

ANEXO VIII

**DIC, WAIC E MARGINAL LIKELIHOOD DOS MODELOS
CAPÍTULO 5**

Modelo		DIC	WAIC	Marginal log-Likelihood	
				Estimación por integração	Estimación Gaussiana
M0	Sem efeitos	2325.8934	94338.590	629.86460	632.88243
M0.HF		-4939.2695	79028.594	2475.11175	2478.09846
M0.LS		-5606.3946	58221.929	2620.15891	2623.21365
MG	Só com efeitos fixos (todos)	4689.9308	96316.972	929.93399	932.99734
MG.HF		-2098.0965	88118.362	2500.06692	2503.11525
MG.LS		-2586.0556	83935.367	2669.82130	2672.90110
RW.M1.2	Lag 2 - agregados/desagregados - semana/14 dias f(Date_str, model="rw2")	7677.1317	30587.711	-6540.86490	-6541.27364
RW.M1.2.HF		7976.0895	24676.580	-4853.09521	-4853.48631
RW.M1.2.LS		7683.0216	22942.931	-4702.27209	-4702.66418
RW.M1.2.HF.14		8395.6339	18368.397	-4811.13948	-4811.52906
RW.M1.2.LS.14		8291.0564	19601.623	-4830.84699	-4831.24664
RW.M1.7	Lag 7 - agregados/desagregados - semana/14 dias f(Date_str, model="rw2")	7682.5015	30376.236	-6571.24590	-6571.64834
RW.M1.7.HF		7893.1131	24566.256	-4856.20680	-4856.59868
RW.M1.7.LS		7674.3804	21107.931	-4707.70107	-4708.09184
RW.M1.7.HF.14		8164.8000	17304.065	-4815.81262	-4816.20374
RW.M1.7.LS.14		8127.3629	17948.611	-4841.43272	-4841.82757
RW2.M1.14	Lag 14 - agregados/desagregados - semana/14 dias f(Date_str, model="rw2")	7541.5007	37588.854	-6851.59444	-6852.01005
RW2.M1.14.HF		8199.5007	23777.369	-4970.88836	-4971.28194
RW2.M1.14.LS		7952.9140	21179.560	-4856.28884	-4856.68393
RW2.M1.14.HF.14		8419.6788	18721.766	-4942.02376	-4942.42565
RW2.M1.14.LS.14		8378.1183	18821.882	-4969.05708	-4969.45470
iCAR.M1.14	Lag 14 - agregados/desagregados - semana/14 dias f(Region_unstr,model="bym",graph=W.porto)	5489.1838	31560.484	-3169.46630	-3169.10190
iCAR.M1.14.HF		-465.7591	6546.582	-298.92670	-298.56234
iCAR.M1.14.LS		-492.1595	6222.151	-301.60294	-301.23862
iCAR.M1.14.HF.14		-759.2454	7455.062	-152.76841	-152.40416
iCAR.M1.14.LS.14		-706.4972	7706.504	-148.97297	-148.60882
iCAR.M1.2	Lag 2 - agregados/desagregados - semana/14 dias f(Region_unstr,model="bym",graph=W.porto)	4925.0083	33712.468	-3320.54797	-3320.18341
iCAR.M1.2.HF		-458.9139	6263.600	-333.90475	-333.54045
iCAR.M1.2.LS		-558.5282	7614.625	-306.36337	-305.99918
iCAR.M1.2.HF.14		-763.2964	7538.966	-133.03884	-132.67468
iCAR.M1.2.LS.14		-957.2300	8436.789	-62.51723	-62.15293
iCAR.M1.7	Lag 7 - agregados/desagregados - semana/14 dias f(Region_unstr,model="bym",graph=W.porto)	5522.8747	32852.201	-3482.75050	-3482.38416
iCAR.M1.7.HF		529.5575	15776.512	-610.54416	-610.17784
iCAR.M1.7.LS		398.1499	12419.688	-606.35788	-605.99152
iCAR.M1.7.HF.14		-189.9005	11448.104	-310.31568	-309.94931
iCAR.M1.7.LS.14		-560.5228	9928.744	-192.59884	-192.23323

Modelo		DIC	WAIC	Marginal log-Likelihood	
				Estimaco por integrao	Estimaco Gaussiana
iCAR.rw.iid.M2.14	Lag 14 - agregados/desagregados - semana/14 dias f(Region_str, model="bym",graph=W.porto)+ f(Date_str, model="rw2") + f(Date_unstr, model="iid")+ f(region_date,model="iid")	2514.0166	96690.420	658.98980	661.99632
iCAR.rw.iid.M2.14.HF		-4411.1354	75761.860	2436.46964	2439.46455
iCAR.rw.iid.M2.14.LS		-4884.4520	82741.627	2582.82644	2585.89903
iCAR.rw.iid.M2.14.HF.14		-5866.0743	81355.960	2563.33035	2566.31970
iCAR.rw.iid.M2.14.LS.14		-6229.4887	62864.286	2794.55277	2797.55488
iCAR.rw.iid.M2.2	Lag 2 - agregados/desagregados - semana/14 dias f(Region_str, model="bym",graph=W.porto)+ f(Date_str, model="rw2") + f(Date_unstr, model="iid")+ f(region_date,model="iid")	2236.2860	92641.500	676.13037	679.16676
iCAR.rw.iid.M2.2.HF		-4504.8794	81630.890	2449.50087	2452.59537
iCAR.rw.iid.M2.2.LS		-5070.0001	89803.550	2605.67638	2608.69932
iCAR.rw.iid.M2.2.HF.14		-6249.6136	79372.137	2610.12276	2613.19410
iCAR.rw.iid.M2.2.LS.14		-6459.3320	63559.839	2826.14489	2829.13990
iCAR.rw.iid.M2.7	Lag 7 - agregados/desagregados - semana/14 dias f(Region_str, model="bym",graph=W.porto)+ f(Date_str, model="rw2") + f(Date_unstr, model="iid")+ f(region_date,model="iid")	2540.1161	92324.080	577.01172	580.10320
iCAR.rw.iid.M2.7.HF		-4411.1969	78694.220	2411.89490	2414.96744
iCAR.rw.iid.M2.7.LS		-4805.4866	78818.653	2564.67822	2567.70549
iCAR.rw.iid.M2.7.HF.14		-5812.1029	73609.016	2546.14196	2549.17559
iCAR.rw.iid.M2.7.LS.14		-6239.7356	63500.631	2767.58547	2770.62002
IID.M1.14	Lag 14 - agregados/desagregados - semana/14 dias f(Date_unstr, model="iid") f(Date_unstr, model="iid")+ f(region_date,model="iid")	7508.1812	37655.545	-6817.63393	-6818.04218
IID.M1.14.HF		8082.2133	23636.915	-4952.68885	-4953.08230
IID.M1.14.LS		7802.7930	21238.996	-4836.96293	-4837.35668
IID.M1.14.HF.14		8320.5953	18634.219	-4910.47664	-4910.86995
IID.M1.14.LS.14		8333.2112	18588.905	-4938.53517	-4938.93035
IID.M1.2	Lag 2 - agregados/desagregados - semana/14 dias f(Date_unstr, model="iid") f(Date_unstr, model="iid")+ f(region_date,model="iid")	7535.9290	31244.966	-6500.49898	-6500.90729
IID.M1.2.HF		7860.9933	25194.460	-4853.66522	-4854.05650
IID.M1.2.LS		7374.4013	23669.489	-4690.77913	-4691.17094
IID.M1.2.HF.14		8174.7749	18793.039	-4806.08393	-4806.47491
IID.M1.2.LS.14		8109.8011	20748.261	-4801.72260	-4802.11643
IID.M1.7	Lag 7 - agregados/desagregados - semana/14 dias f(Date_unstr, model="iid") f(Date_unstr, model="iid")+ f(region_date,model="iid")	7668.8318	30174.163	-6534.74697	-6535.15368
IID.M1.7.HF		7572.6961	20152.150	-4832.40161	-4832.79329
IID.M1.7.LS		7460.3222	21021.788	-4688.65003	-4689.04191
IID.M1.7.HF.14		8016.7583	17522.526	-4784.86535	-4785.25717
IID.M1.7.LS.14		7990.1943	18104.519	-4797.22165	-4797.61697
CAR.M1.14	Lag 14 - agregados/desagregados - semana/14 dias f(Region_unstr,model="besag",graph=W.porto)	5489.1215	31561.007	-3189.58458	-3189.96742
CAR.M1.14.HF		-465.7960	6546.939	-319.04242	-319.42544
CAR.M1.14.LS		-492.1961	6222.485	-321.70291	-322.08627
CAR.M1.14.HF.14		-759.2804	7455.369	-172.87818	-173.26103
CAR.M1.14.LS.14		-706.5208	7706.668	-169.11377	-169.49745
CAR.M1.2	Lag 2 - agregados/desagregados - semana/14 dias f(Region_unstr,model="besag",graph=W.porto)	4924.8742	33714.171	-3340.64773	-3341.03100
CAR.M1.2.HF		-458.9666	6264.208	-354.03026	-354.41348
CAR.M1.2.LS		-558.5805	7615.175	-326.47405	-326.85722
CAR.M1.2.HF.14		-763.3410	7539.818	-153.15503	-153.53786
CAR.M1.2.LS.14		-957.2676	8437.544	-82.65555	-83.03899
CAR.M1.7	Lag 7 - agregados/desagregados - semana/14 dias f(Region_unstr,model="besag",graph=W.porto)	5522.7805	32856.032	-3502.78218	-3503.16536
CAR.M1.7.HF		529.5400	15778.308	-630.56872	-630.95192
CAR.M1.7.LS		398.1275	12421.031	-626.36117	-626.74441
CAR.M1.7.HF.14		-189.9261	11449.532	-330.32875	-330.71193
CAR.M1.7.LS.14		-560.5364	9929.590	-212.65156	-213.03606

ANEXO IX

Resultados do teste de Durbin-Watson para avaliar a independência das diferenças entre os valores observados e estimados para cada variável nos concelhos do distrito do Porto

Concelho	Autocorrelacao
Amarante	2.745396
Baiao	2.099469
Felgueiras	2.634051
Gondomar	2.597060
Lousada	2.280333
Maia	1.962644
Marco de Canaveses	1.846776
Matosinhos	2.185767
Pacos de Ferreira	2.597478
Paredes	2.434295
Penafiel	2.723214
Porto	2.454724
Povoa de Varzim	2.314174
Santo Tirso	2.156524
Valongo	2.437189
Vila do Conde	2.332245
Vila Nova de Gaia	2.413701
Trofa	2.425768

ANEXO X

Códigos R para a implementação da desagregação temporal

```
library(Matrix)
library(pracma)
library(gsignal)
library(MASS)
install.packages("openxlsx")
library(openxlsx)
library(writexl)
library(dplyr)
      library(ggplot2)
library(reshape2)
library(lubridate)
library(clipr)
library(flare)
library(tidyr)

# Período de 25 de março de 2020 a 30 de junho de 2020 - Capítulo 4
# Agregar a 7 dias
#####
# Criar uma nova data.frame para armazenar os resultados
resultados.1<- data.frame(ID_Region = integer(), Semana = numeric())
# Loop para somar os valores em blocos de 7 dias para cada ID_Region
for (region_id in unique(Dist_Porto_1.1$ID_Region)) {
  data_region <- filter(Dist_Porto_1.1, ID_Region == region_id)
  semana <- numeric()
  for (i in seq(1, nrow(data_region), by = 7)) {
    valores_semana <- data_region$Diario.Unif[i:(i + 6)]
    soma_semana <- sum(valores_semana)
    semana <- c(semana, soma_semana)
  }
  resultados.1 <- rbind(resultados.1, data.frame(ID_Region = region_id,
Semana = semana))
}
# Visualizar os resultados
print(resultados.1)
```

```

# Agregar a 14 dias
#####
# Carregar o pacote dplyr
library(dplyr)

# Loop para somar os valores de duas semanas consecutivas para cada ID_Region
for (region_id in unique(resultados.1$ID_Region)) {
  data_region <- filter(resultados.1, ID_Region == region_id)
  duas_semanas <- numeric()

  for (i in 2:(nrow(data_region))) {
    soma_duas_semanas <- data_region$Semana[i] + data_region$Semana[i -
1]
    duas_semanas <- c(duas_semanas, soma_duas_semanas)
  }

  data_region <- data_region[-nrow(data_region), ]
  data_region$DuasSemanas <- duas_semanas

  if (region_id == unique(resultados.1$ID_Region)[1]) {
    resultados_finais.1 <- data_region
  } else {
    resultados_finais.1 <- rbind(resultados_finais.1, data_region)
  }
}

# Desagregar os dados agregados a 14 dias, com sobreposição
clipboard_data <- read_clip()
Y<-read.table(text = clipboard_data, header = TRUE, sep="\t") #Dados da
nova_data_frame.Y
# tem todos os valores de dados de conelhos acumulados a 14 dias desde
25/3/2020

# Definir a matriz A
Overlap <-7 # dados com sobreposição de uma semana, agregados a 14 dias
fit <- 1
A<- create_obs_matrix(N, 14, Overlap, fit)

```

```

# Desagregar com o método LS
n<-98 #N° de períodos de dados desagregados
resultados.1_LS <- matrix(0, nrow =n, ncol = ncol(Y))
# Loop para substituir y por cada coluna de Y
for (i in 1:ncol(Y)) {
  y_i <- Y[, i] # Extrai a coluna i de Y
  x_LS_i <- ginv(A) %*% y_i # Calcula x_LS para a coluna i de Y
  resultados.1_LS[, i] <- round(x_LS_i,0)
}

# Desagregar pelo método HF

# Definir a função para implementar o método H-Fuse
H_Fuse <- function(A, y, H) {
# reconstruction using smoothness regularization (using ginv for final
reconstruction)
  N <- ncol(H)
  A <- rbind(A, H)
  y <- c(y, rep(0, N-2+1))
  xhat_sm <- t(ginv(A) %*% y)
  return(xhat_sm)
}

# Definir a matriz H
N<-98
h <- c(1, -1)
c <- c(h[1], rep(0, N - 2))
r <- rep(0, N)
r[1:2] <- h
H<-Toeplitz(c, r)

n<-98 #N° de períodos de dados desagregados
resultados.1_HF<- matrix(0, nrow = n, ncol = ncol(Y))
  for (i in 1:ncol(Y)) {
    y_i <- Y[, i] # Extrai a coluna i de Y
    x_HF_i <- H # Calcula x_LS para a coluna i de Y
    x_Hfuse_i <- H_Fuse(A, y_i, H) # Calcula x_Hfuse para a coluna i de Y
    resultados.1_HF[, i] <- round(x_Hfuse_i,0)
  }

```

```
#Desagregar os dados semanais sem sobreposição de 25/3 a 30/6/2020
```

```
N <- 98
```

```
# Definir a matriz A e a matriz H - a matriz de suavização
```

```
h <- c(1, -1)
```

```
c <- c(h[1], rep(0, N - 2))
```

```
r <- rep(0, N)
```

```
r[1:2] <- h
```

```
H1<-Toeplitz(c, r)
```

```
dim(H1)
```

```
# Definir a matriz A
```

```
Overlap <-0 # dados sem sobreposição agregados a 7 dias
```

```
fit <- 1
```

```
A<- create_obs_matrix(N, 7, Overlap, fit)
```

```
Overlap <-7 # dados com sobreposição de uma semana, agregados a 14 dias
```

```
fit <- 1
```

```
A<- create_obs_matrix(N, 14, Overlap, fit)
```

```
clipboard_data <- read_clip()
```

```
Y1<- read.table(text = clipboard_data, header = TRUE, sep="\t") # tem  
todos os valores de dados semanais de concelhos de 25/3 até 7/7 depois de  
agregados
```

```
#Método LS
```

```
n<-98 #N° de períodos de dados desagregados
```

```
resultados.1_LS.S <- matrix(0, nrow =n, ncol = ncol(Y1))
```

```
# Loop para substituir y por cada coluna de Y
```

```
for (i in 1:ncol(Y1)) {
```

```
  y1_i <- Y1[, i] # Extrai a coluna i de Y
```

```
  x_LS.S_i <- ginv(A1) %*% y1_i # Calcula x_LS para a coluna i de Y
```

```
  resultados.1_LS.S[, i] <- round(x_LS.S_i,0)
```

```
}
```

```
resultados.1_LS.S.df<-as.data.frame(resultados.1_LS.S)
```

```

#Método HF
n<-98 #N° de períodos de dados desagregados
resultados.1_HF.S<- matrix(0, nrow = n, ncol = ncol(Y1))
  for (i in 1:ncol(Y1)) {
y1_i <- Y1[, i] # Extrai a coluna i de Y
x_HF.S_i <- H1 # Calcula x_LS para a coluna i de Y
x_Hfuse.S_i <- H_Fuse(A1, y1_i, H1) # Calcula x_Hfuse para a coluna i de Y
resultados.1_HF.S[, i] <- round(x_Hfuse.S_i,0)
}

resultados.1_HF.S.df<-as.data.frame(resultados.1_HF.S)

# Desagregar os dados de 7 de julho a 27 de outubro de 2020 e de 28 de outubro de 2020 a 23 de março de 2021

H_Fuse <- function(A, y, H) {
  # reconstruction using smoothness regularization (using ginv for final reconstruction)
  N <- ncol(H)
  A <- rbind(A, H)
  y <- c(y, rep(0, N-2+1))
  xhat_sm <- t(ginv(A) %*% y)
  return(xhat_sm)
}

# Definir a matriz A e a matriz H

# Definir a matriz A
Overlap <-0 # dados sem sobreposição agregados a 7 dias (7 de julho a 27 de outubro de 2020)
fit <- 1
A<- create_obs_matrix(N, 7, Overlap, fit)

Overlap <-7 # dados com sobreposição de uma semana, agregados a 14 dias (28 de outubro de 2020 a 23 de março de 2021)
fit <- 1
A<- create_obs_matrix(N, 14, Overlap, fit)

```

```

# Definir a matriz H

N<-147
h <- c(1, -1)
c <- c(h[1], rep(0, N - 2))
r <- rep(0, N)
r[1:2] <- h
H<-Toeplitz(c, r)

# Carregar a matriz Y (Dados observados)

Y<-read.table("clipboard", header = TRUE, sep="\t") # tem todos os
valores de dados de concelhos acumulados a 14 dias desde
11/11/2020

#Calcular as matrizes de dados desagregados segundo LS e HF a partir de
28/10/2020 e até 23/03/2021
n<-147 #N° de períodos de dados desagregados
resultados_LS <- matrix(0, nrow =n, ncol = ncol(Y))

# Loop para substituir y (dados desagregados) por cada coluna de Y (dados
observados)
for (i in 1:ncol(Y)) {
  y_i <- Y[, i] # Extrai a coluna i de Y
  x_LS_i <- ginv(A) %*% y_i # Calcula x_LS para a coluna i de Y
  resultados_LS[, i] <- round(x_LS_i,0)
}

n<-147 #N° de períodos de dados desagregados
resultados_HF<- matrix(0, nrow = n, ncol = ncol(Y))
for (i in 1:ncol(Y)) {
  y_i <- Y[, i] # Extrai a coluna i de Y
  x_HF_i <- H # Calcula x_LS para a coluna i de Y
  x_Hfuse_i <- H_Fuse(A, y_i, H) # Calcula x_Hfuse para a coluna i de Y
  resultados_HF[, i] <- round(x_Hfuse_i,0)
}

```

```
#Desagregar os dados semanais sem sobreposição de 7/7 a 27/10
```

```
N <- 112
```

```
Y1<- read.table("clipboard", header = TRUE, sep="\t") # tem todos os  
valores de dados semanais de concelhos de 14/7 até 27/10
```

```
# Método LS
```

```
n<-112 #Nº de períodos de dados desagregados
```

```
resultados_LS.1 <- matrix(0, nrow =n, ncol = ncol(Y1))
```

```
# Loop para substituir y por cada coluna de Y
```

```
for (i in 1:ncol(Y1)) {
```

```
  y1_i <- Y1[, i] # Extrai a coluna i de Y
```

```
  x_LS.1_i <- ginv(A1) %*% y1_i # Calcula x_LS para a coluna i de Y
```

```
  resultados_LS.1[, i] <- round(x_LS.1_i,0)
```

```
}
```

```
resultados_LS.1.df<-as.data.frame(resultados_LS.1)
```

```
# Método HF
```

```
n<-112 #Nº de períodos de dados desagregados
```

```
resultados_HF.1<- matrix(0, nrow = n, ncol = ncol(Y1))
```

```
  for (i in 1:ncol(Y1)) {
```

```
    y1_i <- Y1[, i] # Extrai a coluna i de Y
```

```
    x_HF.1_i <- H1 # Calcula x_LS para a coluna i de Y
```

```
    x_Hfuse.1_i <- H_Fuse(A1, y1_i, H1) # Calcula x_Hfuse para a coluna i de
```

```
Y
```

```
    resultados_HF.1[, i] <- round(x_Hfuse.1_i,0)
```

```
  }
```

```
  resultados_HF.1.df<-as.data.frame(resultados_HF.1)
```