Spatio-temporal analysis of environmental variables and their influence in neurodegenerative diseases

Mariana Ramos Oliveira

PHD IN HEALTH DATA SCIENCE

2024

Advisor

Professor Doutor Hernâni Manuel da Silva Lobo Maia Gonçalves

Co-advisors

Professor Doutor José Alberto da Silva Freitas

Professora Doutora Ana Cláudia Moreira Teodoro

The work of this PhD thesis was developed at the Center for Health Technology and Services Research (CINTESIS), in Faculty of Medicine of University of Porto, and supported by a PhD grant (reference SFRH/BD/147324/2019), financed by FCT (Foundation for Science and Technology) for the Doctoral Programme in Health Data Science (HEADS).

PhD jury

President:

- Doctor Cristina Maria Nogueira da Costa Santos, Full Professor at Faculty of Medicine, University of Porto

Board:

- Doctor Steve MacFeely, Adjunct Professor at the University College Cork, Ireland
- Doctor Enrico Gianluca Caiani, Associate Professor at Politecnico di Milano, Italy
- Doctor Lia Bárbara Cunha Barata Duarte, Assistant Professor at Faculty of Sciences, University of Porto
- Doctor Armando Cardoso, Assistant Professor at Faculty of Medicine, University of Porto
- Doctor Hernâni Manuel da Silva Lobo Maia Gonçalves, Assistant Professor at Faculty of Medicine, University of Porto

Acknowledgements

For one can't live on it's on, let alone accomplish great deeds, this section is dedicated to all those who stood with me throughout this journey.

First of all, to my great tutor, Professor Hernâni Gonçalves, who trusted me throughout these years and backed me up in each step of the way.

Alongside, Professor Ana Cláudia Teodoro, without whom I would never have gotten the chance to step into this wonderful world of health data. You opened the door for me.

Together with them, Professor Alberto Freitas, who's been accompanying me from day one, forgiving my forgetfulness, even when I write it down.

Also, to all my teachers, from first grade onto the PhD, you are all an inspiration and have taught me to think critically and analyse my problems from different points of view.

Second comes my family, my supporting rocks. To my mother, warrior by nature, who taught me to be the same. To my sister, who's like a mother to me, and who taught me dreams do come true. To my husband, who gets home late from work greeting me with a hug that can hold the whole world together, never refusing to help. To my twin niece and nephew, who came to cheer me up whenever I was tired, filling me with energy. To my niece, who's like my sister, and her husband, who brought to this world the tiniest bundle of cuteness I ever laid eyes upon, hypnotising me just a couple of months short of this delivery.

And because friends are the family we chose, to my chosen sister, who always answers my silly questions and helps me google things up.

And last, but not least, the true symbol behind this thesis, my father and guardian angel, the man who taught me so many great things no words can encompass, and the reason behind this work.

Say my name, so I will know you're back you're here again, for a while

Oh let us share the memories that only we can share, together

Tell me about the days before I was born, how we were as children

These colors come alive, in your heart and in your mind

I cross the borders of time, leaving today behind to be with you again

Within Temptation

Table of contents

Abstract I
Outline III
1. Introduction 1
1.1. Neurodegenerative diseases 1
1.2. Environmental factors 2
1.3. Statistical concepts 3
1.4. Study objectives 4
2. Materials and methods 7
2.1. Neurodegenerative diseases 7
2.2. Environmental factors 8
2.2.1. *Land Surface Temperature* 9
3. Geospatial analysis of environmental risk factors in neurodegenerative diseases: A systematic review 13
3.1. Introduction 14
3.2. Materials and Methods 16
3.3. Results 18
3.3.1. *Identification, screening and assessment* 18
3.3.2. *Qualitative synthesis* 26
3.4. Discussion 27
3.5. Conclusion 29
Appendix A 30
4. Geospatial analysis of environmental atmospheric risk factors in neurodegenerative diseases: A systematic review update 35
4.1. Introduction 36
4.2. Materials and methods 37
4.3. Results 38
4.3.1. *Identification, screening and assessment* 38
4.3.2. *Qualitative synthesis* 45
4.4. Discussion 46
4.5. Conclusion 47
Appendix B 48
5. Land surface temperature algorithm calibration through meteorological stations 59
5.1. Introduction 60
5.2. Materials 61

5.2.1. *Meteorological stations* ... 61
5.2.2. *Remote sensing data and auxiliary data* ... 62
5.2.3. *Study area and analysis period* ... 62
5.3. Pre-processing ... 62
5.3.1. *Meteorological stations* ... 62
5.4. Methodology ... 63
5.4.1. *Google Earth Engine* ... 63
5.4.2. *Python API* ... 67
5.4.3. *Data analysis* ... 71
5.5. Results ... 73
5.6. Conclusion ... 76
6. Neurodegenerative diseases hospitalizations in Portugal: a spatio-temporal analysis ... 77
6.1. Introduction ... 78
6.2. Materials and methods ... 79
6.2.1. *Study area* ... 79
6.2.2. *Data acquisition* ... 81
6.2.3. *Data processing* ... 81
6.3. Results ... 83
6.3.1. *Descriptive statistics* ... 83
6.3.2. *Spatio-temporal analysis* ... 86
6.4. Discussion ... 98
6.4.1. *Descriptive statistics* ... 98
6.4.2. *Spatio-temporal analysis* ... 98
6.4.3. *Limitations* ... 100
6.4.4. *Implications and future work* ... 100
6.5. Conclusion ... 100
Appendix C ... 102
7. Spatio-temporal analysis of the environmental factors in the Iberic Peninsula between 2000 and 2022 ... 112
7.1. Introduction ... 113
7.2. Materials and methods ... 114
7.2.1. *Geographic limits data* ... 114
7.2.2. *Air pollution* ... 114
7.2.3. *Environmental factors* ... 116
7.2.4. *Temporal analysis* ... 117
7.2.5. *Spatial analysis* ... 118

7.3. Results 118
7.3.1. *Validation of environmental factors measurements: ERA5 and Landsat surface temperature* 118
7.3.2. *Temporal analysis* 120
7.3.3. *Spatial analysis* 125
7.3.4. *Spatio-temporal analysis* 129
7.4. Discussion 129
7.5. Conclusion 130
Appendix D 131
8. Environmental risk factors for neurodegenerative diseases hospitalizations 141
8.1. Environmental risk factors for neurodegenerative diseases hospitalizations 142
8.1.1. Introduction 142
8.1.2. Materials and methods 143
8.1.3. Results 146
8.1.4. Discussion 149
8.1.5. Conclusion 149
8.2. Neurodegenerative risk assessment tool 150
8.2.1. Homepage 150
8.2.2. Tools page 150
8.2.3. Sample report page 152
8.2.4. Patient's form page 153
8.2.5. Results page 154
8.2.6. Conclusion 156
9. Discussion 157
10. Conclusion 161
References 163

Figures Index

Figure 1 - Moran's I example of autocorrelation coefficients. 3
Figure 2 - Scatterplot of values vs their spatial lag in LISA analysis. Different quadrants represent different types of clusters (High-high: red; High-low: orange, Low-high: light blue, Low-low: dark blue). 4
Figure 3 - Definition of the representativeness area of a particular meteorological station. First image on the left represents a land use map, overlaid with a meteorological station and a 3 Km buffer around it. In the middle, the area of the land use map in which the meteorological station falls is selected in red. Lastly, the clipped area in which the red polygon in the centre image overlays the buffer. 10
Figure 4 - Systematic Review PRISMA Flowchart 19
Figure 5 - Number of published articles on this review's scope throughout the years. 38
Figure 6 - Systematic review PRISMA flowchart, including original and update reviews and their junction. Neuro: neurodegenerative disease; env: environmental factor or pollutant, geo: geographic approach. 40
Figure 7 - Studies in each category by domain: neurodegenerative diseases, geographic approach and environmental factors and pollutants. 41
Figure 8 - Country of origin of the studies and the diseases studied in each country. 44
Figure 9 - Spatial distribution of meteorological stations in mainland Portugal. 62
Figure 10 - GEE code (part 1): example of importing Landsat 7 SR monthly data from the January 2000 and December 2018. 64
Figure 11 - GEE code (part 2): Example of importing study area boarders and emissivity dataset. 64
Figure 12 - GEE code (part 3): Definition of functions to scale bands and mask cloud pixels. 65
Figure 13 - GEE code (part 4): Definition of the function to compute LST. 65
Figure 14 - GEE code (part 5): Example of applying the functions to the Landsat ImageCollection. . 66
Figure 15 - GEE code (part 6): Example of visualization parameters for LST product. 66
Figure 16 - Computed LST for Portugal mainland during the month of January 2000 66
Figure 17 - Python's code (part 1): Importing initial libraries, downloading initial datasets (boarders, Landsat imagery, emissivity dataset and meteorological stations) and setting initial datasets. 67
Figure 18 - Python's code (part 2): Defining functions to scale bands, mask clouds, compute LST and latter convert the resulting FeatureCollection into a DataFrame. 68
Figure 19 - Python's code (part 3): Definition of a cycle to iterate on each month of the time period, application of the functions to scale bands, mask clouds and compute LST, and computation

of monthly mean and kernel mean, creating a single two banded image with the resulting means. 69
Figure 20 - Python's code (part 4): Code example of computing zonal statistics from monthly images and meteorological stations. 70
Figure 21 - Statistical analysis code (part 1): Example of a function to compute descriptive statistics on the output values of remote sensing processed data. 72
Figure 22 - Statistical analysis code (part 2): Implementing the previous descriptive statistics functions to the results dataset, iterating over each column. 73
Figure 23 - Scatter plot of satellite-computed LST and the corresponding meteorological station value measured. Different colours represent the different approaches in retrieving the LST value, as described in chapter 5.3.1. for both simple (“M”) and Kernel (“K”) mean. 74
Figure 24 - Boxplot of differences found between the computed LST value and that measured by the meteorological station. 76
Figure 25 - Population distribution per age group in Portugal – mean value from 2001, 2011 and 2021 [258]..... 79
Figure 26 - Portuguese population distribution per district and age group. Age groups as in Figure 1, population mean axis = [0 - 350 000], and population % accumulated axis = [0 - 100].... 80
Figure 27 - Moran's I example of autocorrelation coefficients. 82
Figure 28 - Scatterplot of values vs their spatial lag. Different quadrants represent different types of clusters (High-high: red; High-low: orange, Low-high: light blue, Low-low: dark blue).. 83
Figure 29 - Number of hospitalizations and respective percentage per neurodegenerative disease, followed by the percentage of those hospitalizations in which the neurodegenerative disease was the primary diagnosis. The total number of hospitalizations is 502245. 84
Figure 30 - Age standardized hospitalization rates considering all diseases, per districts. 86
Figure 31 - Age standardized principal diagnosis rates considering all diseases, per districts and per year. 87
Figure 32 - Timeline of ASHR per neurodegenerative disease. 88
Figure 33 - Yearly Moran's I for both districts and municipalities for each administrative division type. 89
Figure 34 - Neurodegenerative diseases local clusters when considering districts as the administrative area (all diseases). 90
Figure 35 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Alzheimer). 91
Figure 36 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Basal). 92
Figure 37 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Creutzfeldt). 92

Figure 38 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Dementia). 93
Figure 39 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Huntington). 93
Figure 40 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Lewy). 94
Figure 41 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Motor Neuron Disease). 94
Figure 42 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Multiple Sclerosis). 95
Figure 43 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Parkinson). 95
Figure 44 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Pick). 96
Figure 45 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Other). 96
Figure 46 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (all diseases). 97
Figure 47 - Location of the air quality stations. In orange are Portuguese stations, and in blue Spanish stations. Stations are unevenly distributed, being located mainly in the littoral areas and concentrated in the main cities 115
Figure 48 - Comparison between both ERA5 and Landsat surface temperatures and meteorological stations. 117
Figure 49 - Scatter plots of computed surface temperature from ERA5 and Landsat and air temperature from meteorological stations. 119
Figure 50 - Bland-Altman plots for each remote sensing product and station values. 119
Figure 51 - Normalized temporal data for each environmental pollutant and factor. 121
Figure 52 - Temporal data after trend was removed through differencing for all environmental variables and factors. 122
Figure 53 - Temporal data after seasonality was removed through differencing for all environmental variables and factors. 123
Figure 54 - Temporal data after all processing to remove trend, seasonality, outliers and missing values for all environmental variables and factors. 124
Figure 55 - Local spatial autocorrelation clusters for each environmental pollutant/factor (Red: high-high; Orange: high-low; Light blue: low-high, Dark blue: low-low). 128
Figure 56 - Location of the air quality stations. In orange are Portuguese stations, and in blue Spanish stations. 144

Figure 57 - Website map 150
Figure 58 - Neurodegenerative risk assessment website - Homepage 151
Figure 59 - Neurodegenerative risk assessment website - Tool page 152
Figure 60 - Neurodegenerative risk assessment website - Sample report page 153
Figure 61 - Neurodegenerative risk assessment website – Patient's form page 154
Figure 62 - Neurodegenerative risk assessment website - Results page 155

Tables Index

Table 1 - ICD codes used to retrieve hospitalization data. 7
Table 2 - Number of studies addressing each subtopic of the domains, retrieved from the included studies: neurodegenerative diseases, environmental pollutants/factors and geographic factors 20
Table 3 - Study characteristics summary: study design, study limitations, statistical methods, and effect measures 21
Table 4 - Study characteristics of the 34 included papers 22
Table 5 - Studies published by each geographic region, grouped by neurodegenerative disease 26
Table 6 - Number and percentage of papers included in the update for each domain categories. 42
Table 7 - Study methodological characteristics summary: limitations, study design, statistical methods and effect measures 43
Table 8 - Country of origin with the number of studies published (N) and the respectively studied diseases. Some studies have investigators or focus on data from more than one country. ... 44
Table 9 - Pearson's correlation coefficient computed between each monthly image, both mean and kernel mean bands, and the buffered areas of the meteorological stations, using original dataset. 75
Table 10 - Pearson's correlation coefficient computed between each monthly image, both mean and kernel mean bands, and the buffered areas of the meteorological stations, using outlier cleaned dataset. 75
Table 11 - ICD codes used to retrieve hospitalization data. 81
Table 12 - Descriptive statistics of each neurodegenerative disease included in this study, as well as their total. Categorical data (sex, age group and residence) is described by its absolute frequency and percentage, while continuous data (age and length of stay, which are both asymmetrically distributed), is described by its median and interquartile range. The total row describes the total frequency of each disease and their respective percentage. yo: years old. 85
Table 13 - Spearman correlation coefficient for each interpolation method and the stations' value. . 115
Table 14 - Comparison of the performance of Landsat and ERA5 on the measurements between Landsat and ERA5of temperature, with the meteorological stations as reference. Further description of the methodology in 2.3 (Std is the standard-deviation of the predicted values when compared to the measured values). 120
Table 15 – Spearman’s R for each environmental factor/pollutant in time series (** - significant at 0.01; * - significant at 0.05). Colours represent the strength of the correlation: greens are positive

correlations and oranges negative correlations, with darker colours meaning higher correlations (RH: Relative Humidity; SP: Surface Pressure; ST: Surface Temperature) ... 125
Table 16 - Spearman's' correlation coefficients between variables (** - significant at 0.001; * - significant at 0.05). In parenthesis are the values obtained for the primary diagnosis database. 147
Table 17 - Regression algorithms scores for all diagnosis and primary diagnosis neuro..... 149

Equations Index

Equation 1 8
Equation 2 8
Equation 3 10
Equation 4 10
Equation 5 65
Equation 6 65
Equation 7 75
Equation 8 75
Equation 9 82
Equation 10 116
Equation 11 145
Equation 12 145

Appendixes Index

Appendix A

Table A1 - List of the 56 excluded papers 30

Appendix B

Table B1 - Detailed characteristics of each included study. ALS: amyotrophic lateral sclerosis; GIS: geographic information system; MS: multiple sclerosis. 48

Table B2 - Study results summary. +: positive significative association, -: negative significative association; ?: no significative association found. 56

Appendix C

Table C1 - Hospitalizations with more than one neurodegenerative disease - general descriptive (Hunt. is Huntington, Park. is Parkinson, N is the number of occurrences and IQI is the inter quartile interval). 102

Table C2 - Descriptive statistics of hospitalizations with neurodegenerative disease as primary diagnosis included in this study. Categorical data (sex, age group and residence) is described by its absolute frequency and percentage, while continuous data (age and length of stay, which are both asymmetrically distributed), is described by its median and interquartile range. The total row describes the total frequency of each disease and their respective percentage. yo: years old. 105

Figure C1 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Alzheimer). 106

Figure C2 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Basal). 106

Figure C3 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Creutzfeldt). 107

Figure C4 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Dementia). 107

Figure C5 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Huntington). 108

Figure C6 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Lewy). 108

Figure C7 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Motor Neuron Disease). 109

Figure C8 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Multiple Sclerosis). 109

Figure C9 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Parkinson). 110
Figure C10 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Pick). 110
Figure C11 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Other) 111

Appendix D

Figure D1 - Benzene spatio-temporal distribution in the Iberian Peninsula. 131
Figure D2 - Carbon monoxide spatio-temporal distribution in the Iberian Peninsula. 132
Figure D3 - Nitrogen dioxide spatio-temporal distribution in the Iberian Peninsula. 133
Figure D4 - Ozone spatio-temporal distribution in the Iberian Peninsula. 134
Figure D5 - Particulate matter < 10 µm spatio-temporal distribution in the Iberian Peninsula. 135
Figure D6 - Particulate matter < 2.5 µm spatio-temporal distribution in the Iberian Peninsula. 136
Figure D7 - Sulphur dioxide spatio-temporal distribution in the Iberian Peninsula. 137
Figure D8 – Relative humidity spatio-temporal distribution in the Iberian Peninsula. 138
Figure D9 - Surface pressure spatio-temporal distribution in the Iberian Peninsula. 139
Figure D10 - Temperature spatio-temporal distribution in the Iberian Peninsula. 140

List of abbreviations and acronyms

Health related:

- AD: Alzheimer Disease
- ALS: Amyotrophic Lateral Sclerosis
- ASHR: Age-standardized Hospitalization Rates
- CNS: Central Nervous System
- FTD: Frontotemporal Degeneration
- ICD-*-CM: International Classification of Diseases - Clinical Modification
- LB: Lewy Bodies
- MND: Motor Neuron Disease
- MS: Multiple Sclerosis
- PD: Parkinson Disease
- PMS: Paediatric Multiple Sclerosis
- SARS-CoV-2: Severe Acute Respiratory Syndrome Corona Virus-2

Environmentally related:

- AQI: Air Quality Index
- LST: Land Surface Temperature
- PM: Particulate Matter
- PM1: Particulate Matter < 1 µm
- PM10: Particulate Matter < 10 µm
- PM25: Particulate Matter < 2.5 µm
- RH: Relative Humidity
- SP: Surface Pressure
- SR: Surface Reflectance
- ST: Surface Temperature

Geographically related:

- ASTER: Advanced Spaceborne Thermal Emission and Reflection Radiometer
- COS: Land Use and Cover Chart
- EPSG: European Petroleum Survey Group
- ERA5: Fifth generation ECMWF Atmospheric Reanalysis
- ETM+: Enhanced Thematic Mapper Plus
- ETRS: European Terrestrial Reference System
- GEE: Google Earth Engine

- GIS: Geographic Information Systems
- NIR: Near Infrared
- RGB: Red Green Blue
- SWIR: Short-wave Infrared
- TOA: Top Of the Atmosphere
- TOMS: Total Ozone Mapping Spectrometer
- UTM: Universal Transverse Mercator
- WGS: World Geodetic System

Chemical elements and formulas:

- BC: Black Carbon
- C_6H_6: Benzene
- CO: Carbon Monoxide
- NO_2: Nitrogen Dioxide
- NO_X: Nitrogen Oxides
- O_3: Ozone
- Pb: Lead
- SO_2: Sulphur Dioxide

Statistically related:

- ANOVA: Analysis of Variance
- DBSCAN: Density-Based Spatial Clustering of Applications with Noise
- IDW: Inverse Distance Weighting
- IQI: Inter Quartile Interval
- LISA: Local Indicators of Spatial Association
- OR: Odds Ratio
- STD: Standard Deviation
- SVM: Support Vector Model
- SVR: Support Vector Regression

Entities:

- ACSS: Central Administration of the Health System
- APA: Portuguese Environment Association
- CINTESIS: Centre for Health Technology and Services Research
- DGT: General Territory Direction
- FCT: Foundation for Science and Technology
- IGN: Geographic Information National Centre

- INE: National Institute of Statistics
- IPMA: Portuguese Institute of Sea and Atmosphere
- MCTES: Ministry of Science, Technology and Higher Education
- MITECO: Ministry of the Ecological Transition and the Demographic Challenge
- NASA: National Aeronautics and Space Administration
- PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- QualAr: Quality of the Air
- SNIG: National Infrastructure for Geographical Information
- SNIRH: National System of Hydrological Resources Information
- WHO: World Health Organisation
- WMO: World Meteorological Organization

Miscellaneous:

- API: Application Programming Interface
- DOI: Digital Object Identifier
- EU: European Union
- UK: United Kingdom
- URL: Uniform Resource Locator
- USA: United States of America
- ZIP: Zone Improvement Plan

List of outputs

Published articles:

- Geospatial Analysis of Environmental Atmospheric Risk Factors in Neurodegenerative Diseases: A Systematic Review
  - Mariana Oliveira, André Padrão, André Ramalho, Mariana Lobo, Ana Cláudia Teodoro, Hernâni Gonçalves, Alberto Freitas (2020)
  - International Journal of Environmental Research and Public Health 17, no. 22: 8414
  - Impact factor: 3.5 (Q1: Medicine – Public Health, Environmental and Occupational Health)
  - *https://doi.org/10.3390/ijerph17228414*

Proceedings:

- Land surface temperature algorithm calibration through meteorological stations
  - Mariana Oliveira, Ana Cláudia Teodoro, Alberto Freitas, Hernâni Gonçalves (September 2021)
  - Proc. SPIE 11863, Earth Resources and Environmental Remote Sensing/GIS Applications XII, 1186310
  - *https://doi.org/10.1117/12.2599909*
- Spatial Patterns in Neurodegenerative Disease's Hospitalizations in Portugal (2000-2016)
  - Mariana Oliveira, Alberto Freitas, Ana Cláudia Teodoro, Hernâni Gonçalves (2023)
  - Proceedings of the 9th International Conference on Geographical Information Systems Theory, Applications and Management - GISTAM
  - *http://doi.org/10.5220/0011990100003473*

Conferences:

- Land surface temperature algorithm calibration through meteorological stations
  - Oral presentation at Earth Resources and Environmental Remote Sensing/GIS Applications XII
- Spatial Patterns in Neurodegenerative Disease's Hospitalizations in Portugal (2000-2016)
  - Oral presentation at the 9th International Conference on Geographical Information Systems Theory, Applications and Management - GISTAM

Articles submitted for publication:

- Geospatial Analysis of Environmental Atmospheric Risk Factors in Neurodegenerative Diseases: A Systematic Review Update
  - Mariana Oliveira, André Padrão, Ana Cláudia Teodoro, Alberto Freitas, Hernâni Gonçalves
  - Submitted to Systematic Reviews
  - Impact factor: 3.7 (Q1: Medicine – Miscellaneous)
- Neurodegenerative diseases hospitalizations in Portugal: a spatio-temporal analysis
  - Mariana Oliveira, Alberto Freitas, Ana Cláudia Teodoro, Hernâni Gonçalves

    - Submitted to International Journal of Health Geographics
    - Impact factor: 4.9 (Q1: Public Health, Environmental and Occupational Health)

Articles at the review stage by the authors before submission:

- Spatio-temporal analysis of the environmental factors in the Iberian Peninsula between 2000 and 2022
    - Mariana Oliveira, Alberto Freitas, Ana Cláudia Teodoro, Hernâni Gonçalves
    - Draft to be submitted to the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
    - Impact factor: 5.5 (Q1: Computers in Earth Sciences)
- Environmental risk factors for neurodegenerative diseases hospitalizations: regression models
    - Mariana Oliveira, Alberto Freitas, Ana Cláudia Teodoro, Hernâni Gonçalves
    - Draft

Abstract

Our world is in a never-seen-before state, in which life expectancy keeps increasing, and with it the prevalence of many diseases primarily targeting the elders, as is the case of neurodegenerative diseases. Simultaneously, our exposure to environmental toxins is now a constant concern of public health, which brings new studies to fulfil the insights on the impacts of those toxins in our body. That is exactly the purpose of this thesis in particular respect to neurodegenerative diseases and environmental factors.

To fulfil this purpose, a couple of systematic reviews of literature were conducted to establish the state of the art and pinpoint important air pollutants and environmental factors, as well as expected results, and potential geographic approaches and methods. With them, it was confirmed the gap that no similar studies have been performed in Portugal. The first systematic review included 34 studies, 60% of which focusing on multiple sclerosis and in the second 35 studies were included, with Dementia now being the main focus. Environmental pollutants are gaining prominence in epidemiological studies.

Subsequently, between the year of 2000 to 2016, all neurodegenerative diseases' related acute care hospitalizations were considered and statistically and spatio-temporally described. An increase in age-standardized hospitalization rates was found throughout the years, as well as global and local spatial autocorrelation.

A powerful tool to assess environmental data through large temporal periods and spatial areas is remote sensing and along with it, Google Earth Engine as a tool to access and process all imagery at lower computation times. Landsat-7 and ERA5 (Fifth generation ECMWF Atmospheric Reanalysis) products were used to compute relative humidity, surface pressure and surface temperature and meteorological stations provided the remaining environmental factors (benzene, carbon monoxide, nitrogen dioxide, ozone, particulate matter with aerodynamic diameter of under 10 μm and 2.5 μm and sulphur dioxide). Environmental data presents high seasonality and seems intrinsically connected to the topography of the region.

Environmental factors show high correlation with neurodegenerative diseases' hospitalizations, in particular NO_2 (Nitrogen Dioxide) and surface pressure, with correlation coefficients of -0.34 and -0.32, respectively.

Although this study is limited by the reuse of hospitalization data and exposure assessment bias, it shows that exposure to environmental factors is related to the hospitalization rates by neurodegenerative diseases.

Keywords: neurodegenerative diseases, air pollution, environmental factors, spatio-temporal analysis

Outline

This work is a compilation of 5 papers published or to be published in scientific journals and 2 conference proceedings publications that are interconnected to explore the needed concepts, materials and methods that integrate it. Consequently, the thesis is organized like so:

- Introduction (chapter 1): gives an overview of the work done, the main key concepts involved and the research question in hand;
- Materials and methods (chapter 2): describe the data sources used, along with the respective variables and processing needed, including respective methods;
- Systematic reviews (chapters 3 and 4): show all literature directly interconnected to this work, exploring possible variables and methodological approaches to be used in this context. This section is divided in two articles, an original review and an update to that same review;
- Land Surface Temperature algorithm (chapter 5): displays the struggle in extracting and manipulating remote sensing data and confronting it with ground data;
- Spatial patterns in hospitalizations (chapter 6): maps all neurodegenerative diseases acute hospitalizations from 2000 to 2016 in Portugal mainland and explores spatial and temporal trends;
- Spatial patterns in environmental variables (chapter 7): likewise, hospitalization patterns, seeks to establish spatio-temporal trends on environmental variables from 2000 to 2022 in the Iberian Peninsula;
- Environmental risk factors (chapter 8.1): brings everything together by analysing the hospitalizations alongside the environmental variables, seeking correlations and exploring regression models;
- Risk assessment tool (chapter 8.2): describes a potentially useful tool to compute a patient's risk based on the variables of this study;
- Discussion (chapter 9): briefly presents the main results of the above studies;
- Conclusion (chapter 10): bullet points the main deductions of this thesis.

Each chapter has an introductory paragraph not included in the originally published version. This paragraph is between black lines, right after the article's title for simple identification. For better flow, all references are listed at the end of the thesis. Also, minor changes merely aesthetic or related to table, figure and equation numbers were made to the papers to increase consistency and improve the reader's experience.

1. Introduction

1.1. Neurodegenerative diseases

Due to demographic changes worldwide and the fact that neurodegenerative diseases have increased incidence in elderly population, this group of diseases is currently one of the main challenges for medicine and public health [1]. This is also a particularly challenging group due to subtle distinctions between the diseases and the three different domains these diseases usually affect (cognition, behaviour and motor) [2]. Bellow, a brief definition of each disease addressed in this thesis is presented:

- Alzheimer's disease: the leading cause of Dementia is Alzheimer's disease (AD), which is also the most common neurodegenerative disease worldwide. It is characterized by a progressive loss of cognitive function, typically associated with memory deterioration, due to the loss of neurons, white matter and synapses [1].
- Basal Ganglia degeneration: basal ganglia are brain structures anatomically symmetrical constituted of grey matter that may be affected by degenerative conditions, distressing both motor and non-motor functions and resulting primarily in disturbances of movement and muscle tone [3].
- Creutzfeldt-Jakob disease: this is a fatal neurodegenerative disease characterized by a rapid and progressive cognitive decline. It can either occur sporadically, acquired through infection or familial and it is the most common prion disease worldwide [4, 5].
- Dementia: Dementia may be defined by the loss of cognitive abilities that have a direct impact in quotidian chores and lead to independence loss. This is a broad term that encompasses many other diseases. On one hand, frontotemporal degeneration (FTD) affects both the frontal and temporal regions of the brain and is the second most common form of early-onset Dementia [6]. On the other hand, vascular Dementia is characterized by the impairment of executive functions and is usually undistinguishable from AD [7].
- Huntington's disease: described by excessive and uncontrolled motor movements and deficits in cognitive and emotional functions, Huntington is the most common hereditary neurodegenerative disease [8].
- Lewy body disease: this disease is considered to be the second most common cause of Dementia following AD [9] in people after the age of 75 and patients with this disease usually have combined Dementia and parkinsonism [10].
- Motor neuron disease: Motor neuron disease (MND) is a progressive motor syndrome that causes upper motor disfunction and denervation, initially asymmetrical. Although it is predominantly a motor disorder, it can also compromise the cognitive and occasionally the sensory systems [11]. Amyotrophic Lateral Sclerosis (ALS) is the most common form of motor

neuron disease and its degeneration primarily affects the cortical, brainstem and spinal cord areas [1].

- Multiple Sclerosis: Multiple sclerosis (MS) is an acquired inflammatory disease of the central nervous system, leading to a variety of symptoms scattered throughout the autonomic, motor, sensory and visual systems and it usually affects young adults [12].
- Parkinson's disease: Parkinson's disease (PD) is the second most prevalent neurodegenerative disease worldwide [13] and its symptoms include bradykinesia, tremor, rigidity and postural instability, alongside other non-motor symptoms such as sleep disorders and respiratory dysfunction [14].
- Pick's disease: this disease is a form of Dementia with the presence of Pick bodies – well delineated, round and cytoplasmatic inclusions of neurons [15] - and presents many similarities with frontotemporal Dementia [16].

1.2. Environmental factors

One of the main contributors to diseases worldwide is air pollution, mostly leading to a significant number of deaths from stroke, lung cancer and heart disease [17]. In Europe, this is the largest environmental health risk, posing serious long-term effects [18, 19]. However, and even though 3.8% of deaths from all causes in 2017 in Portugal were attributed to air pollution [20, 21], studies concerning the relation between air pollution and health in Portugal are scarce [22]. Current research has been linking outdoor air pollution to compromised brain health, mainly in children and elders [23]. Air pollutants such as particulate matter (PM), nitrogen dioxide (NO_2), ozone (O_3), sulphur dioxide (SO_2) and carbon monoxide (CO) have been proven to negatively affect human health [19] and, particularly, have been linked to neurodegeneration [24]. These pollutants may be retrieved through air quality and meteorological stations, which are unfortunately scarce and unevenly distributed [25], needing interpolation to fill in values in between stations. Remote sensing is another extensive source of environmental data relevant to human health [26].

In particular regard to neurodegenerative related studies, remote sensing data has been used to estimate temperature, precipitation and some pollutants [24]. Remote sensing satellite imagery data can be retrieved from numerous sensors with different characteristics. For instance, Landsat 8 is a high spatial resolution product that has a repeat cycle of 16 days and contains, apart from the typical Red, Green and Blue (RGB), the near-infrared (NIR) and short-wave infrared (SWIR) bands, which have a 30 meter resolution, and a thermal band with 60 meter resolution [27] that is essential to compute land surface temperature (LST), together with the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Emissivity Database which provides emissivity with a spatial resolution of 100 meters [28]. Products with computed land variables are also available nowadays, such as ERA5-Land (Fifth generation ECMWF Atmospheric Reanalysis) from Copernicus project which has several

climate variables available at both daily and monthly aggregates [29]. The perk on using this data is its spatial resolution, of over 10 kilometres. However, over large areas such as countries, this low spatial resolution may not have a too negative of an impact and might actually lower memory requirements and processing times [30].

1.3. Statistical concepts

This work could not have been accomplished without the use of statistics, and a brief description of the concepts associated with some of the employed methods is presented here:

- Pearson's correlation coefficient (r): this coefficient is used to measure the linear association between two variables, with at least one of them assuming normal distribution, and can assume values between -1 to +1. Positive values imply positive association, thus indicating that the increase in one variable is followed by the increase in the other variable. Also, values closer to 1 indicate better linear association while values close to 0 indicate no linear association. Additionally to this measure, a significance test is usually performed to compute the p-value of the coefficient [31];
- Spearman's rank correlation coefficient: this is the equivalent of Pearson's r if neither variables have a normal distribution or if one of the variables is discrete or ordinal and its interpretation is similar [31];
- DBSCAN: standing for Density-Based Spatial Clustering of Applications with Noise, this is a density-based clustering algorithm that groups points closely distributed and marks points further away as outliers and is one of the most common spatial clustering algorithms [32];
- Linear regression support vector model (SVM) with k-folds: SVMs are supervised models that can be associated with a number of algorithms to perform classification or regression tasks. In this particular case, a linear regression model is fit into a dataframe split between training and testing set and provide robustness to the analysis when compared to simple linear regression models [33];
- Inverse distance weighting (IDW): this spatial interpolation method estimates unsampled points using a linear combination of sampled points, weighted inversely by their distance. This method's accuracy is affected by the power used in its computation, being the most popular choice the power of 2, with smoothness of the interpolation increasing with the decrease of power. If the power is 0, then this method is a moving average and if 1, it's a linear interpolation [34];
- Moran's Index (Moran's I): this is the most common measurement of the degree of relation between geographic observations and can be used in a variety of contexts. It is a number between -1

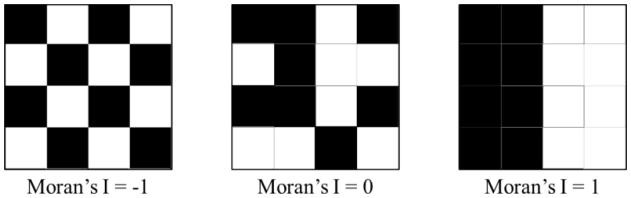


Figure 1 - Moran's I example of autocorrelation coefficients.

and 1 that quickly summarizes the existence of spatial autocorrelation and its degree. It can also identify clustering since positive values are associated to high-high or low-low values and negative values are associated to high-low or low-high values [35, 36] (Figure 1);

- Local Indicators of Spatial Association (LISA): Moran's I algorithm can be applied to each individual observation instead of all global observations, thus becoming LISA. Their use and interpretations is much like Moran's I, except locally instead of globally [35] (Figure 2);

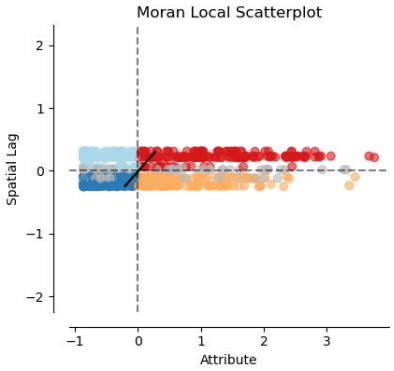


Figure 2 - Scatterplot of values vs their spatial lag in LISA analysis. Different quadrants represent different types of clusters (High-high: red; High-low: orange, Low-high: light blue, Low-low: dark blue).

- Logistic regression: this is a common approach in statistical analysis used to describe the relationship between a categorical outcome and a set of independent variables [37];
- Linear regression: this is an extremely common statistical tool used to search for a linear regression between a continuous dependent variable which is the outcome and the independent variables [38];
- Decision tree regression: this algorithm defines a tree of decision rules that model the values of the input variables, aiming to find the best tree to determine the output by measuring, at each step, which variable combination rely the lowest error to the decision [39];
- SVM regression: this method is very similar to its classification algorithm, except it allows for a continuous outcome to be modelled through the application of function estimates [40];
- Random forest regression: similarly to the random forest classifier, this method uses several random decision trees to model the outcome [41].

1.4. Study objectives

The primary aim of this study is to investigate whether environmental variables are spatially related to neurodegenerative diseases using hospitalization data with episodes coded primarily or secondarily as neurodegenerative diseases, with discharges between 2000 and 2016. To that purpose, five specific objectives were defined:

- Perform a systematic literature review to identify environmental variables being used, neurodegenerative diseases being studied, statistical and geographical approaches used and expected results;
- Statistically describe from the temporal and spatial perspectives the health data available of hospitalizations related with neurodegenerative disorders, in order to identify patterns and prepare data processing for posterior analysis;

- Explore and validate GEE as a tool to obtain and process remote sensing environmental data as a more efficient framework than that traditionally used;
- Explore and describe environmental data, starting from collection of data from 2000 to 2022, as recurring to spatial interpolation to obtain a continuous dataset of environmental factors for the whole Iberian Peninsula with monthly measurements;
- Using the conclusions from relating neurodegenerative age-standardized hospitalization rates with environmental factors, develop a tool to publish results, allowing healthcare professionals to estimate a patients' risk of developing a neurogenerative disease based on environmental exposures.

2. Materials and methods

2.1. Neurodegenerative diseases

Hospitalizations of patients with at least one neurodegenerative pathology coded as a diagnosis, both principal or secondary, were obtained from hospitalization data reuse from mainland public hospitals, with discharges between 2000 to 2016, provided by the Portuguese Central Administration of the Health System (ACSS). No further years were included since the provided data frame included only hospitalizations from 2000 to 2018, but both 2017 and 2018 were incomplete and would thus result in non-comparable results. Each record contains the year, patient's sex, age and residence (coded up to the parish), entry and leave dates, length of stay and coded diagnosis (ICD-9-CM or ICD-10-CM codes).

International Classification of Diseases (ICD) codes for neurodegenerative pathologies were retrieved from literature review [42-47], along with counselling with coding experts. The category "Other" represents neurodegenerative diseases not described elsewhere by the alternative codes and is a category included in the ICD. All resulting ICD codes used are listed in Table 1.Using *Python*, diagnosis codes were extracted and recoded as the disease they referred to. For example, ICD-9 code 331.0 and ICD-10 code G30 were both recoded as 'Alzheimer'. Age was also grouped into three categories: under 20 years old, 20 to 70 years old, and over 70 years old. These categories were chosen to adapt to the available population categories, as well as to accommodate the distinction between paediatric, early onset and late onset neurodegenerative diseases [48].

Population data for each Portuguese mainland parish was also retrieved from National Institute of Statistics (INE) publicly available Population Census for the years of 2001, 2011 and 2021 [49]. To obtain an approximate value of population by year, data was then linearly interpolated, and further used to normalize health data. All values were rounded to the closest integer, due to the nature of the population variable. District and municipality population was computed by summing the contained parishes.

Table 1 - ICD codes used to retrieve hospitalization data.

Diagnosis		ICD 9	ICD 10
Alzheimer		331.0	G30
Pick's		331.1	G31
Lewy		331.8	
Creutzfeldt		046	A81.0
Huntington		333.4	G10
Parkinson		332	G20
Motor neuron disease		335	G12.2
Multiple sclerosis		340	G35
Basal ganglia		333.0	G23
Dementia	Vascular	290.4	F01 F02 F03
	Unspecified	294	
	Senile	331.2	
	Cerebral degeneration	331.7	
	Presenile	290.1	
	Senile with delusion	290.2	
	Senile with delirium	290.3	
	Alcohol induced	291.2	
	Drug induced	292.82	
Other		331.9	G32

Data was then grouped by either district, municipality or parish to proceed to mapping, and population in each administrative area, as well as the percentage of each age group, were used to compute age-standardized hospitalization rates (ASHR), using Equation 1 [50]. This rate allows data to be normalized and therefore enables comparison of the rates, both across space and time, with a lower degree of error.

$$ASHR = \frac{N_{AgeGroup}}{Population_{AgeGroup}} \times \%_{AgeGroup} \quad \text{Equation 1}$$

In Equation 1, $N_{AgeGroup}$ represents the number of hospitalizations in each age group, $Population_{AgeGroup}$ represents the total population in the area corresponding to each age group (for example, each district) and $\%_{AgeGroup}$ is the percentage of the age group in the same population area. The resulting data was then mapped using QGIS software (version 3.28.3 - Firenze) to allow for an adequate spatio-temporal representation and consequent patterns identification.

2.2. Environmental factors

Air pollution data from 2000 to 2023 was obtained from QualAr (Quality of the Air) platform from the Portuguese Environment Association (APA) [51], IPMA (Portuguese Institute of Sea and Atmosphere) [52] and the Spanish Ministry of the Ecological Transition and the Demographic Challenge (MITECO) website [53], on January 2022, except for 2022 data, downloaded on December 2023. This dataset corresponds to meteorological stations scattered throughout the countries of Portugal and Spain and represents point measurements of the following pollutants: Benzene (C_6H_6), Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Particulate Matter with diameter under 10 µm (PM_{10}) and 2.5 µm ($PM_{2.5}$) and Sulphur Dioxide (SO_2). Spanish data was initially intended to be used to improve spatial interpolation towards the frontiers, but since meteorological stations were numerous throughout Spain, the whole Iberian Peninsula was used in the spatio-temporal analysis of environmental factors.

Retrieved values for the pollutants were then treated using Python. First, both Portuguese and Spanish data were pre-treated to match style and formatting, thus allowing for the merging of both countries' datasets. Furthermore, the data was converted from the original Excel format (.xlsx) to shapefile (.shp), to allow for spatial interpolation. IDW [34] with a weighting power of 2 was used to interpolate values outside of the stations.

Relative Humidity (RH), Surface Pressure (SP) and temperature were computed from remote sensing data using ERA5 Land monthly aggregates available in GEE data catalogue, for the years of 2000 to 2023. Both SP and temperature were directly retrieved by the product, whilst RH was computed using the dewpoint temperature and the temperature and Equation 2 [54, 55].

$$RH = \frac{6.11 \times 10^{\frac{7.5 \times Dewpoint\ temperature}{237.7 + Dewpoint\ temperature}}}{6.11 \times 10^{\frac{7.5 \times Surface\ temperature}{237.7 + Surface\ temperature}}} \times 100 \quad \text{Equation 2}$$

Equation 2 uses a frequent humidity measurement in satellites, which is the dew-point temperature (in ºC) - the temperature at which the air needs to be in order to reach saturation - and the surface temperature (also in ºC). The vapor pressure is then computed for both the dew-point temperature and

the surface temperature and, by dividing them, the relative humidity is obtained as a decimal between 0 and 1. By multiplying by 100, this result is converted into percentage.

2.2.1. *Land Surface Temperature*

Since at the beginning of this work, ERA5 was not available on GEE data catalogue, Landsat 7 was used to compute LST, since it offered images for a continuous study period between 2000 and 2022. Landsat 8 was thus disregarded for only having imagery from 2013 onward. Meteorological stations were used to validate and calibrate the computed LST and were available online through SNIRH (National System of Hydrological Resources Information) [56] and IPMA [52] for free download. A total of 117 stations were retrieved from SNIRH with mean monthly temperature and 11 stations from IPMA with minimum and maximum monthly temperature. As stations are independent from each other, their data was simply merged into a single excel file in which each line corresponded to the station and the columns contained the values measured. However, 38 of the total stations had no measured values from 2000 to 2018, and as so were disregarded. To analyse the representativeness of meteorological stations, a Land Use and Cover Chart (COS) from 2018 was used, downloaded from SNIG (National Infrastructure for Geographical Information) [57]. As meteorological stations were considered ground truth data, a representativeness analysis was performed to explore which might be the best fitting method to extract temperature from the satellite images. According to the world meteorological organization (WMO), meteorological stations may be representative of the element they are measuring in a range of 100 meters to 3 Km when used in applications such as air pollution [58]. Following these guidelines, a buffer of 3 Km was set around each station, through which the COS was clipped. The resulting polygon is solely the continuous extent of COS corresponding to that where the station is located, up to the limit of 3 Km. Through this approach, all stations have a different area of influence, depending on the land conditions surrounding it. Figure 3 shows this process for one particular station. From left to right, it represents the area of 3 Km around the meteorological station and surrounding soil occupation categories, the selection of the soil occupation category in which the station is inserted and finally, the obtained area of representativeness of that station which is the area within 3 Km with the same occupation category as that of the station. Simple buffers of 1 Km, 2 Km and 3 Km, along with the simple pixel corresponding to the location of the station were also drawn to compare methodologies.



Figure 3 - Definition of the representativeness area of a particular meteorological station. First image on the left represents a land use map, overlaid with a meteorological station and a 3 Km buffer around it. In the middle, the area of the land use map in which the meteorological station falls is selected in red. Lastly, the clipped area in which the red polygon in the centre image overlays the buffer.

The atmospherically corrected surface reflectance from Landsat 7 ETM+ (Enhanced Thematic Mapper Plus) sensor was used, taking advantage of the availability at GEE data catalogue, thus sparing further pre-processing [59], along with ASTER emissivity. The whole process was done recurring to *Python* with GEE API (Application Programming Interface) installed. Simple band scaling and cloud removing were performed and Equation 3 was used to compute LST [60].

$$LST = \frac{SR}{1 + \left(\lambda \times \left({}^{SR}/_{\rho}\right)\right) \times ln(\varepsilon)} - 273.15 \qquad \text{Equation 3}$$

In equation 3, SR (Surface Reflectance) is the Landsat 7 SR band 6 (in Kelvin), λ is the mean wavelength of emitted radiance for the same band (in meters), ρ is a constant given by Equation 4 (in meters Kelvin), and ε is the emissivity (dimensionless). In Equation 4, h is Planck's constant ($6.626*10^{34}$ *Js*), c is the velocity of light ($2.998*10^{8}$ *m/s*) and σ corresponds to the Boltzmann's constant ($1.38*10^{-23}$ *J/K*).

$$\rho = h\frac{c}{\sigma} = 1.438 \times 10^{-2} \qquad \text{Equation 4}$$

Given the computation of LST, monitoring stations were used to further calibrate the values computed through *shapefiles* containing its representativeness areas as described before. Empty values were dropped as they make the further use of machine learning models impossible. Through a DBSCAN model, outliers were separated from the original dataset, which was then used to compute Pearson's r. Later, a linear regression model was used and the corresponding coefficients saved.

Meanwhile, since ERA5 became available on GEE data catalogue, LST computed from Landsat 7 and temperature retrieved from ERA5 were compared to decide which would be best to use in this work. This was done using cross validation and both products were compared to ground meteorological

stations, considered to be ground truth. Pearson's correlations were initially computed and a linear support vector regression model was posteriorly used to predict 20% of the database, using 5 k-folds to improve model robustness. Model performance was then retrieved comparing the obtained predicted values and the real values. All the processing described in this subchapter was performed through Python.

3. Geospatial analysis of environmental risk factors in neurodegenerative diseases: A systematic review

This paper arises from the need for a solid state of the art to identify neurodegenerative diseases, what environmental factors were being studied in relation to them, and what geographical approaches were being used. A systematic review of literature seemed the best option to fulfil this task. The insight this work provided was key to the identification of neurodegenerative diseases, the selection of environmental air pollutants and factors and the comprehension of geographical approaches available to perform the desired task.

Mariana Oliveira, André Padrão, André Ramalho, Mariana Lobo, Ana Cláudia Teodoro, Hernâni Gonçalves, Alberto Freitas

International Journal of Environmental Research and Public Health 17, no. 22: 8414

November 2020 https://doi.org/10.3390/ijerph17228414

3.5 Q1 (Medicine - Public Health, Environmental and Occupational Health)

Abstract: Despite the vast evidence on the environmental influence in neurodegenerative diseases, those considering a geospatial approach are scarce. We conducted a systematic review to identify studies concerning environmental atmospheric risk factors for neurodegenerative diseases that have used geospatial analysis/tools. PubMed, Web of Science, and Scopus were searched for all scientific studies that included a neurodegenerative disease, an environmental atmospheric factor, and a geographical analysis. Of the 34 included papers, approximately 60% were related to multiple sclerosis (MS), hence being the most studied neurodegenerative disease in the context of this study. Sun exposure (n = 13) followed by the most common exhaustion gases (n = 10 for nitrogen dioxide (NO_2) and n = 5 for carbon monoxide (CO)) were the most studied atmospheric factors. Only one study used a geospatial interpolation model, although 13 studies used remote sensing data to compute atmospheric factors. In 20% of papers, we found an inverse correlation between sun exposure and multiple sclerosis. No consensus was reached in the analysis of nitrogen dioxide and Parkinson's disease, but it was related to Dementia and amyotrophic lateral sclerosis. This systematic review (number CRD42020196188 in PROSPERO's database) provides an insight into the available evidence regarding the geospatial influence of environmental factors on neurodegenerative diseases.

Keywords: neurodegenerative; environment; geospatial; epidemiology; systematic review

3.1. Introduction

The human brain suffers pathological changes in the process of ageing, leading to a range of neurodegenerative disorders [61]. Until 2050, the world's population over 60 years old is expected to nearly double compared to 2015, reaching 22% [62]. The ageing of the population raises concerns about the future of neurodegenerative diseases [63] and urges the development of better health systems with long-term care tailored to meet the needs of the elderly population [62]. These diseases are characterised by progressive damage or loss of neuronal cells, leading to compromised functioning of the brain, with either motor or cognitive roles being endangered [64]. Alzheimer's disease (AD) is the most prevalent neurodegenerative disease worldwide [65], and the most common cause of Dementia and it is a progressive disease that can be characterised by extensive cognitive damage, commonly affecting the functional competency to perform quotidian actions [66]. Following AD in prevalence, Parkinson's disease (PD) is defined by the dopaminergic neuronal loss, which presents as a Lewy Bodies (LB) disease that during the course of the disease spreads to different regions of the brain, leading to motor and non-motor symptoms [67]. The third most prevalent neurodegenerative disease is multiple sclerosis (MS), affecting around 2.5 million people worldwide, which is an inflammatory demyelination disorder of the brain, frequently of autoimmune origin [65].

Apart from ageing, the most common risk factors in the development of neurodegenerative diseases include gender, clinical history of hypertension, diabetes, cranial injury, tumours, and infections, as well as smoking and drinking habits [68]. However, the knowledge on the development of most of these diseases remains incomplete, including the influence that environmental factors might have in this context [68]. This review focuses on atmospheric pollution due to its significant impact worldwide, with 91% of the world's population living under air pollution levels exceeding the established limits by the World Health Organisation (WHO) [69]. Atmospheric pollutants are all "substances, laid by the human activities with enough concentration to cause detrimental influences to health (…)" [70], but some authors also consider natural sources of pollutants [71].

Recent studies point out that atmospheric pollution can trigger mechanisms responsible for neurodegenerative diseases, especially particulate matter with a diameter under 2.5 µm ($PM_{2.5}$) due to its capacity to reach the brain [72]. Four main pathways have been identified to air pollution reaching the brain [73]: systemic inflammation (in which the pollutant penetrates peripheral organs such as the lungs, provoking a systemic response that transfers the inflammation to the brain) [73-77], nasal olfactory (inhaled ultrafine and fine particles that penetrate the lungs reaching the systemic circulation, the trigeminal nerve, the brainstem, the hippocampus, and the frontal lobe) [78-83], adsorbed compounds (particulate matter from air pollution encloses various toxic compounds that may be adsorbed independently of the particle) [84-86] and inhaled oxidation (reactive oxygen agents such as ozone, which interact with proteins and lipids in the lungs altering them into toxic compounds that provoke oxidative stress, leading to various brain damages) [87-91]. $PM_{2.5}$ is associated with AD, PD, Dementia, autism, and stroke [92]. Other pollutants such as ozone (O_3), particulate matter with a diameter under 10 µm (PM_{10}), nitrogen dioxide (NO_2), carbon monoxide (CO), and sulphur dioxide (SO_2) are also toxic to the central nervous system (CNS) [17]. Despite the abundance of studies on neurodegenerative diseases and the environmental influence on the CNS, studies considering a geospatial approach are scarce, and many of them only consider the geographic area as a clustering factor [93-95], not proceeding to further analysis.

One of the tools for environmental data extraction and analysis is remote sensing. It has been broadly used in epidemiological studies, increasing the number of publications featuring remote sensing applied to health from 5.6% in 2007 to 13.3% in 2016 [26]. Apart from remote sensing, GIS (geographic information systems), clustering, and spatial models are other tools that can be used to analyse environmental data. Geospatial analysis is a broad concept that combines both exposure assessment methods and posterior statistical analysis, allowing us to integrate information on health and disease of specific regions with environmental data as well as other spatially distributed relevant data such as socio-economic information. Understanding the strengths and limitations of previous studies that used geospatial approach could inform us on how to better delineate future research, thereby maximising the gains regarding the understanding of the geospatial distribution of environmental risk factors associated

with neurodegenerative diseases. This type of knowledge is important to establish to aid in planning future policies that mitigate environmental factors, which may be critical in the prevention and progression of neurodegenerative diseases.

It is important to frame the issue of neurodegenerative diseases in the new coronavirus pandemic caused by SARS-CoV-2 (Severe Acute Respiratory Syndrome Corona Virus-2). This virus can penetrate the CNS mainly through two pathways: by retrograde axonal transport via the cribriform plate, and by the systemic circulatory system [96]. This last pathway is similar to one of the atmospheric pollutants reaching the CNS, raising the hypothesis that this virus may pose a higher potential for neuroinflammation and neurodegeneration [97]. The concept of Neuro-COVID-19 is, thus, being increasingly used in current research [98].

In this systematic review, we aim to identify studies concerning neurodegenerative diseases and their environmental atmospheric risk factors, which have used geographical analysis/tools as part of their work. The methods used to conduct this systematic review are presented in section 3.2. (Materials and Methods), including the query used, the inclusion and exclusion criteria, and the study characteristics gathered by the reviewers. In section 3.3. (Results), there are two subsections, the first one being identification, screening, and assessment (which explains the main numbers retrieved by the systematic review, from the total number of studies screened to the included papers and their characteristics) and the second being qualitative synthesis (in which the main relations retrieved between the environmental and neurodegenerative diseases are shown). The results from the systematic review are discussed and compared to those in the literature in section 3.4. (Discussion), and study limitations are explored. To finalise, a small conclusion is presented in section 3.5.

3.2. Materials and Methods

We searched PubMed, Web of Science, and Scopus databases from inception to the 4th February 2020, date in which the last search was run. The oldest study retrieved was from 1951 [99]. All studies with at least one neurodegenerative disease, one atmospheric pollutant or factor, and some geographic approach were included for screening, with a total of 4655 abstracts screened. No papers were excluded due to language restrictions. For the literature review, we screened the databases using the following keywords: [((alzheimer*) OR (ataxia*) OR (Chorea Minor) OR (creutzfeldt*) OR (Dementia*) OR (Frontotemporal) OR (Guillian-barre syndrome) OR (Huntington*) OR (kennedy* disease) OR (Lewy*) OR (motor neuron*) OR (Myotonic dystrophy) OR (neurodegen*) OR (parkinson*) OR (pick's) OR (Prion) OR (progressive AND palsy) OR (progressive muscular atrophy) OR (sclerosis*) OR (*senile) OR (spinal atrophy)) AND ((atmospher*) OR (carbon) OR (environment*) OR (humidity) OR (meteorologic*) OR (nitrogen) OR (ozone) OR (particulate*) OR (PM2*) OR (pollut*) OR

(sulphur*) OR (surface pressure) OR (temperature)) AND ((drone) OR (geograph*) OR (imagery) OR (landsat) OR (map) OR (mapping) OR (modis) OR (remote sens*) OR (satellite) OR (sentinel) OR (spatial) OR (topologic*))]. This query was a result of the sensitivity analysis of terms that could be relevant for this review, keywords that added no papers to the results were excluded from the query. These included the terms: geospatial, spatio*, and cartograph*. This review was submitted to PROSPERO under the registration number CRD42020196188.

All studies with at least one neurodegenerative disease, one atmospheric pollutant or factor, and some geographic approach were included for screening. The inclusion criteria were the following: (1) Studying a neurodegenerative disease; (2) Accounting for atmospheric environmental factors or pollutants; (3) Using geospatial analysis or tools; (4) All the previous criteria in the same study. Studies were excluded if: no evidence of geographic analysis was found (stating a geographic area without further analysing it or comparing it to other areas was an exclusion motive), no atmospheric pollutants or factors referred (soil and water pollutants were excluded from this revision) and no focus on neurodegenerative diseases (studies about the mechanisms and biospecimen behind the disease were discarded). Full papers were independently read by the same researchers to apply the inclusion and exclusion criteria further. Unavailable full texts were requested to the authors individually. Those studies that remained unavailable were excluded.

For the included full papers data extraction and analysis, structured forms were created and used by the reviewers. The forms included free writing inputs such as the title, year, country, authors, DOI (digital object identifier), participants, aims, and key findings. Additionally, multiple-choice inputs were available for study type, statistical methods, outcome measurements, study limitations, neurodegenerative disease, environmental risk factors, geographic approach, and type of geographical approach. A level of simplification was applied to categorise the study limitations, to better fit most studies on the main biases and issues encountered. Both reviewers filled these forms independently and discussed the extracted information at the end of the process. Nonetheless, the reviewers had the option to input free text if they considered necessary to complement any of the inputs mentioned. Risk of bias from individual studies was assessed by fitting the study limitations into one or more of the pre-defined form options: none given by the authors; conflict of interests (any conflict of interests stated by the authors); confounding factors (unassessed confounding factors); ecological bias (extrapolation of a conclusion from a population to a patient); exposure assessment (issues with assessing the patient's exposure to the environmental factor); interpolation (issues with spatially interpolating the environmental factor); recall bias (data collection relying on patient's memory); migration of patients (patients moving from one residence to another); referral bias (studies relying on the doctor's referring similar cases); sampling issues (under sampling; over sampling or non-representative sampling); statistical issues (lacking of more relevant statistical methods); study design issues (studies acknowledging inappropriate study design); survival bias (relying on a patient being alive over a period

of time); time related issues (inability to assess the amount of time a patient was exposed to the environmental factor); unassessed patients (patients outside the databases not being considered).

3.3. Results

3.3.1. *Identification, screening and assessment*

Of the 4655 articles initially screened, only 34 were included in the final study. The selection process is summarised using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [100] presented by a flow diagram (Figure 4). In the diagram of Figure 4, it is possible to identify that 4655 records were screened in this systematic review, most of which did not relate simultaneously to neurodegenerative diseases, atmospheric factors, and geographical factors (n = 3239). After screening, 90 full papers were assessed for eligibility, from which nearly half (n = 44) did not tackle any atmospheric pollutants or factors. By the end of the selection process, 34 articles were included. The discrepancy in the total number of excluded studies and the sum of the exclusion reasons categories is due to the overlapping of exclusion reasons.

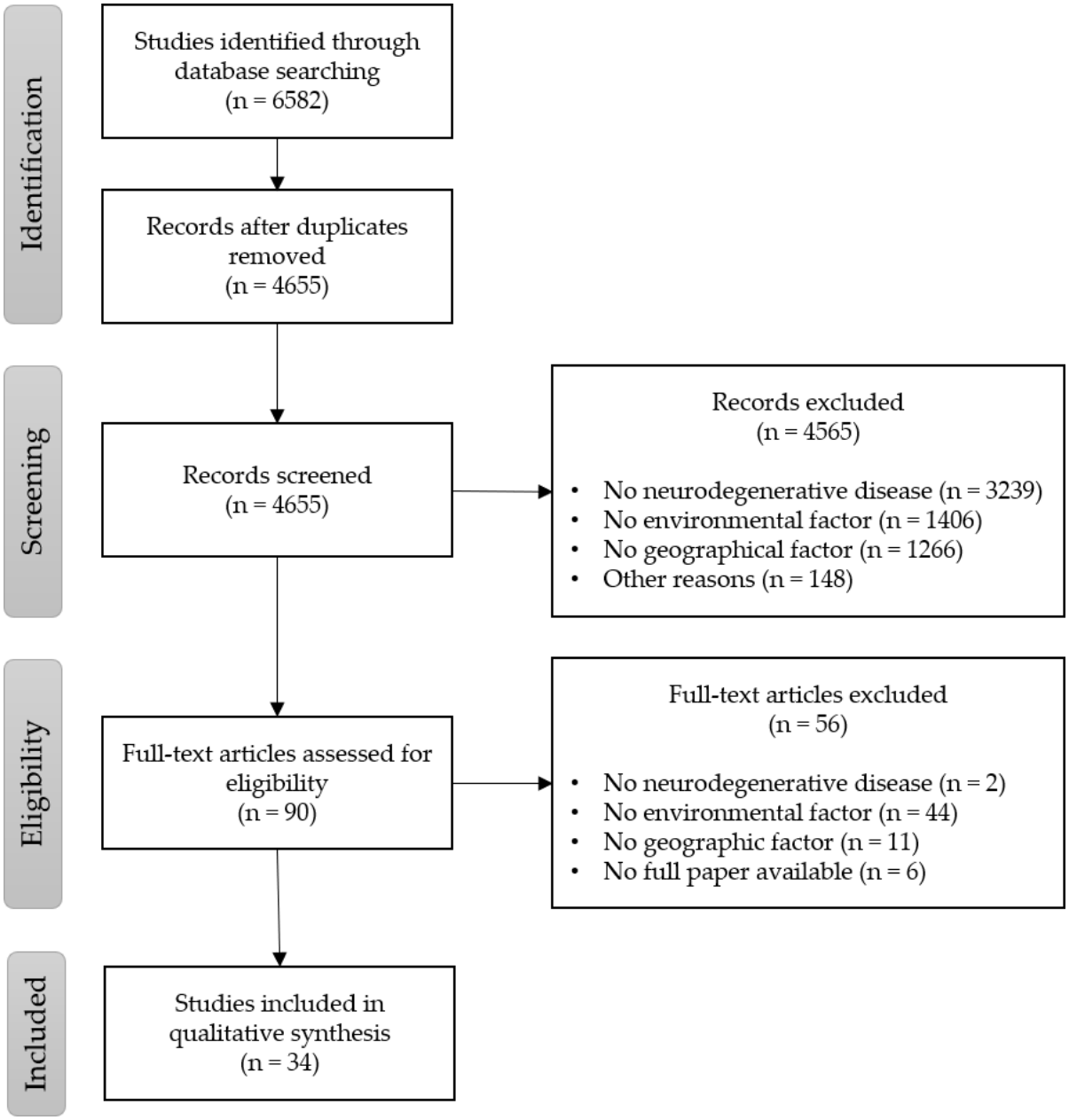


Figure 4 - Systematic Review PRISMA Flowchart

The number of times each disease, environmental factor, and geographical approach was retrieved are detailed in Table 2. The sum of the categories' absolute frequencies for environmental and geographic factors exceeds the number of studies because they are not mutually exclusive. Most papers focused on (MS) (56%), of which 63% studied sun exposure as the main environmental factor, mostly using remote sensing data (58%), always combined with another geographic factor. Administrative divisions were the most used geographical approach when considering all diseases (47%).

Table 2 - Number of studies addressing each subtopic of the domains, retrieved from the included studies: neurodegenerative diseases, environmental pollutants/factors and geographic factors

	Number of studies	Percentage of total entries
Neurodegenerative disease		
Amyotrophic lateral sclerosis	2	5.9
Dementia	3	8.8
Motor neuron disease	1	2.9
Multiple sclerosis	19	55.9
Paediatric multiple sclerosis	2	5.9
Parkinson's disease	7	20.6
Environmental pollutant/factor		
Arsenic (As)	1	1.2
Benzopyrene (C_2OH_{12})	1	1.2
Benzene (C_6H_6)	1	1.2
Cadmium (Cd)	1	1.2
Carbon monoxide (CO)	5	6.1
Cooper (Cu)	2	2.4
Hydrogen sulphide (H_2S)	1	1.2
Humidity	3	3.7
Index	2	2.4
Magnesium (Mg)	2	2.4
Manganese (Mn)	1	1.2
Nickel (Ni)	1	1.2
Nitrogen dioxide (No_x)	10	12.2
Ozone (O_3)	4	4.9
Lead (Pb)	5	6.1
PM_{10}	6	7.3
PM_{25}	8	9.8
Precipitation	6	7.3
Pressure	5	6.1
Radon (Rn)	1	1.2
Sulphur dioxide (SO_2)	6	7.3
Sun exposure	13	15.9
Temperature	8	9.8
Geographic factors		
Administrative division	16	24.2
Clustering	2	3.0
GIS	14	21.2
Latitude	5	7.6
Longitude	1	1.5
Remote sensing	13	19.7
Residence	14	21.2
Spatial interpolation	1	1.5

The main characteristics of the studies - study design, outcomes, statistical methods, and study limitations - are summarised in Table 3. Most of the studies are cross-sectional (n = 12) and correlation was the most reported effect measure (n = 19) and statistical method (n = 22) used. The most common study limitations found were related to exposure assessment (n = 18) and confounding factors n = 15). The full characteristics of the studies can be found in Table 4. This table contains all the characteristics retrieved by the reviewers for each included study: country of origin, year of publication, neurodegenerative disease, environmental factor, geographic factor, study design, study limitations, statistical methods, and outcome. A list of excluded studies [61, 67, 101-154] is also available in Table A1, Appendix A.

Table 3 - Study characteristics summary: study design, study limitations, statistical methods, and effect measures

	Number of studies	Percentage of total entries
Study design		
Case-control	9	26.5
Cohort	4	11.8
Cross-sectional	12	35.3
Ecological	6	17.6
Methodological	2	5.9
Review	1	2.9
Study limitations		
Conflict of interests	1	1.1
Confounding	15	16.1
Ecological bias	3	3.2
Exposure assessment	18	19.4
Interpolation	3	3.2
Memory bias	3	3.2
Migration	4	4.3
None given by the author	5	5.4
Referral bias	2	2.2
Sampling	7	7.5
Statistics	13	14.0
Study design	3	3.2
Survival bias	1	1.1
Time related	9	9.7
Unassessed patients	6	6.5
Statistical methods		
ANOVA	2	2.6
Chi-squared	5	6.5
Clustering	2	2.6
Correlation	22	28.6
Cox regression	4	5.2
Linear regression	11	14.3
Logistic regression	10	13.0
None	2	2.6
Poisson regression	3	3.9
Sensitivity analysis	10	13.0
Spatial autoregressive model	1	1.3
T-test	5	6.5
Effect measures		
Coefficients	14	24.1
Correlation	19	32.8
Hazard ratio	3	5.2
None	3	5.2
Odds ratio	11	19.0
Prevalence	4	6.9
Relative risk	4	6.9

Table 4 - Study characteristics of the 34 included papers

Ref	Country (year)	Neurodegenerative Disease	Environmental factor	Geographic factor	Study design	Study limitations	Statistical methods	Outcome
[155]	Canada (2012)	Multiple sclerosis	Sun exposure	GIS, Remote Sensing	Methodological	Exposure assessment	None	None
[156]	Israel (1971)	Multiple sclerosis	Sun exposure, Temperature, Precipitation, Humidity	Residence	Review	None given by the authors	None	None
[157]	Bulgaria (1987)	Multiple sclerosis	Sun exposure, Temperature, Precipitation	Administrative division, Latitude	Cross-sectional	Unassessed patients	Correlation, Chi-squared, Linear regression	Correlation, Coefficients
[158]	Australia (2001)	Multiple sclerosis	Sun exposure, Temperature, Precipitation	Administrative division, Latitude, Remote Sensing	Ecological	Confounding, Exposure assessment	Correlation, Poisson regression	Prevalence, Correlation
[159]	Canada (2011)	Multiple sclerosis	Sun exposure	Latitude, Longitude, Remote Sensing	Cross-sectional	None given by the authors	Correlation, Linear regression	Correlation
[160]	England (2011)	Multiple sclerosis	Sun exposure	GIS, Remote Sensing	Cross-sectional	Confounding, Sampling, Statistics	Correlation, Linear regression	Correlation, Coefficients
[161]	USA (2017)	Multiple sclerosis	Sun exposure, Temperature	Administrative division, GIS, Remote Sensing	Cross-sectional	Confounding, Statistics	Correlation, Linear regression	Correlation, Coefficients
[162]	USA (2018)	Multiple sclerosis	Sun exposure	Residence, Remote Sensing	Cohort	Confounding, Exposure assessment, Interpolation, Memory bias, Migration, Survival bias, Time related	Cox regression	Relative risk, Hazard ratio
[163]	USA (1983)	Multiple sclerosis	Sun exposure, Temperature, Precipitation, Humidity	Latitude	Case-control	None given by the authors	Logistic regression	Relative risk

[164]	Italy (2016)	Multiple sclerosis	Sun exposure	Administrative division, GIS	Cross-sectional	Confounding, Ecological bias, Time related	Correlation, Linear regression	Correlation, Odds ratio
[165]	Canada (2018)	Multiple sclerosis	Sun exposure	Residence, Remote Sensing	Cohort	Confounding, Exposure assessment, Time related	Linear regression	Coefficients
[166]	Norway (2010)	Multiple sclerosis	Sun exposure, Temperature, Precipitation	Administrative division	Cross-sectional	Migration, Statistics	ANOVA, Poisson regression	Prevalence
[167]	Italy (2018)	Multiple sclerosis	$PM_{2.5}$	Residence, Remote Sensing	Cross-sectional	Conflict of interests, Confounding, Study design	Correlation, Chi-squared	Correlation, Coefficients
[168]	USA (2008)	Multiple sclerosis	PM_{10}, $PM_{2.5}$, NO_X, SO_2, CO	Administrative division	Cross-sectional	None given by the authors	Correlation, T-test, Linear regression	Correlation, Coefficients
[169]	Italy (2005)	Multiple sclerosis	SO_2	Administrative division, Latitude	Cross-sectional	Exposure assessment, Interpolation	Correlation, Linear regression	Correlation, Coefficients
[170]	Iran (2014)	Multiple sclerosis	PM_{10}, NO_X, SO_2	Clustering, GIS	Cross-sectional	Confounding, Statistics, Study design	Correlation, Clustering	Correlation, Coefficients
[171]	Iran (2018)	Multiple sclerosis	Index	Administrative division, GIS, Residence	Cross-sectional	Exposure assessment, Statistics	Correlation, Logistic regression	Odds ratio, Coefficients
[172]	Norway (1997)	Multiple sclerosis	Mg	Administrative division	Methodological	Confounding	Correlation	None
[173]	England (2016)	Multiple sclerosis	Rn	Residence	Ecological	Sampling, Statistics, Unassessed patients	Correlation, Chi-squared, Linear regression	Correlation, Coefficients
[174]	USA (2017)	Paediatric Multiple sclerosis	Index	GIS, Residence	Case-control	Exposure assessment, Statistics, Time	Logistic regression	Odds ratio, Coefficients

						related, Unassessed patients		
[175]	USA (2018)	Paediatric Multiple sclerosis	PM_{10}, $PM_{2.5}$, NO_X, SO_2, CO, O_3, Pb	Administrative division, GIS, Residence	Case-control	Exposure assessment, Referral bias, Time related	T-test, Logistic regression	Odds ratio
[176]	USA (2010)	Parkinson's disease	Cu, Pb, Mg	Administrative division	Ecological	Confounding, Exposure assessment, Statistics	Logistic regression, Sensitivity analysis	Relative risk, Odds ratio
[177]	Spain (2016)	Parkinson's disease	Pb	Administrative division, GIS	Ecological	Exposure assessment, Sampling, Unassessed patients	Correlation, T-test	Correlation, Coefficients
[178]	Canada (2007)	Parkinson's disease	NO_X, Mn	Residence, Remote Sensing, Spatial interpolation	Case-control	Confounding, Exposure assessment, Interpolation, Study design, Time related	Correlation, Linear regression, Logistic regression, Cox regression, Sensitivity analysis	Prevalence, Correlation, Odds ratio
[179]	USA (2016)	Parkinson's disease	PM_{10}, $PM_{2.5}$, NO_X	GIS, Residence	Case-control	Exposure assessment, Memory bias, Statistics, Time related	Correlation, Logistic regression, Sensitivity analysis	Correlation, Odds ratio
[180]	Australia (2020)	Parkinson's disease	$PM_{2.5}$, NO_X	Residence, Remote Sensing	Cross-sectional	Memory bias, Referral bias, Sampling	Logistic regression, Sensitivity analysis	Odds ratio
[181]	Taiwan (2016)	Parkinson's disease	NO_X, CO	GIS, Residence	Case-control	Confounding, Sampling, Statistics	Correlation, Chi-squared, Logistic regression, Sensitivity analysis	Correlation, Odds ratio

[182]	France (2017)	Parkinson's disease	Sun exposure, $PM_{2.5}$	Administrative division, Remote Sensing	Ecological	Ecological bias, Exposure assessment, Migration	Correlation, Poisson regression, Sensitivity analysis	Correlation, Relative risk
[183]	USA (2019)	Dementia	Temperature	Administrative division, Residence, Remote Sensing	Cohort	Confounding, Exposure assessment, Statistics	Correlation, Cox regression, Sensitivity analysis	Correlation, Hazard ratio
[184]	Taiwan (2019)	Dementia	PM_{10}, NO_X, SO_2, CO, O_3	Clustering	Case-control	Confounding, Exposure assessment, Statistics, Unassessed patients	Correlation, Logistic regression, Sensitivity analysis	Odds ratio
[185]	Canada (2017)	Dementia	$PM_{2.5}$, NO_X, O_3	GIS, Residence, Remote Sensing	Cohort	Confounding, Exposure assessment, Time related, Unassessed patients	Cox regression, Sensitivity analysis	Hazard ratio
[186]	Spain (2018)	Amyotrophic lateral sclerosis	PM_{10}, PM_{25}, NO_X, SO_2, CO, O_3, Cu, Pb, As, Ni, Cd, C_6H_6, H_2S, C_6OH_{12}	GIS	Case-control	Ecological bias, Exposure assessment, Sampling, Statistics	T-test, Chi-squared, Linear regression, Sensitivity analysis	Prevalence, Odds ratio
[187]	Taiwan (2013)	Amyotrophic lateral sclerosis	Sun exposure, Temperature, Precipitation, Humidity, Pressure	Administrative division	Case-control	Exposure assessment, Migration, Sampling, Time related	Correlation, Spatial autoregressive model, Clustering	Correlation, Coefficients
[188]	Spain (2016)	Motor neuron disease	Pb	Administrative division, GIS	Ecological	None given by the authors	Correlation, T-test, ANOVA	Correlation, Coefficients

The country with the highest number of published papers under the criteria of this systematic review was the United States of America (USA), with nine included papers. It was followed by Canada with five papers. However, Europe, with the aggregation of articles from Bulgaria, England, France, Italy, Norway, and Spain totals twelve included papers. The Middle East and Asia had three papers, and Australia had two papers. Table 5 contains the papers studied by each geographical region, grouped by neurodegenerative disease.

Table 5 - Studies published by each geographic region, grouped by neurodegenerative disease

	Amyotrophic lateral sclerosis	***Dementia***	***Motor neuron disease***	***Multiple sclerosis***	***Paediatric multiple sclerosis***	***Parkinson***
Asia	[187]	[184]				[181]
Australia				[158]		[180]
Europe	[186]		[188]	[157, 160, 164, 166, 167, 169, 172, 173]		[177, 182]
Middle east				[156, 170, 171]		
North America		[183, 185]		[155, 159, 161-163, 165, 168]	[174, 175]	[176, 178, 179]

Of the 34 included papers, 19 were published in the last 5 years. Until 2000, only four studies were found corresponding to the criteria of this systematic review, all of which focused on multiple sclerosis. Until 2007, except for a paper focusing on Parkinson's disease, no other neurodegenerative disease had been considered.

3.3.2. *Qualitative synthesis*

Among the 19 papers focused on MS, with sun exposure being the most studied environmental factor (n = 13), one was methodological and did not measure any outcome related to MS [155]. An inverse association between MS and sun exposure was reported in seven studies [156-162]. In contrast, in the other three studies, no significant association was found [163-165], one of which attributed the results to the adjusting for latitude [163].

As for temperature, five studies found it to be negatively related to MS [156-158, 161, 166] and one found this association to be due to its high correlation with latitude [163]. Precipitation was found to be uncorrelated [156, 158, 163] or only weakly positively correlated [157, 166] to MS. Humidity was focused only on two studies, both of which found no correlation to MS [156, 163].

Concerning the environmental pollutants, $PM_{2.5}$ was found to be positively correlated with MS in one study [167] and not significantly correlated with another [168]. SO_2 had no correlation [168, 169] except in one study [170]. PM_{10} and NO_X (nitrogen oxides) were studied only twice and studies disagree in the reported correlation [168, 170]. Additionally, a study found a positive association between an air quality

index (AQI) and MS [171], and magnesium (Mg) was found to be negatively correlated [172], while radon was not found to be associated with MS [173]. In the domain of paediatric MS, PM_{10} was found to be related to it [174, 175], as well as $PM_{2.5}$, Pb (lead), SO_2, and CO [175].

Of the seven studies concerning Parkinson's Disease (PD), Pb was found to be positively spatially related [176, 177], as well as Cu (copper), Mg [176], and Mn (manganese) [178]. PM_{10} and $PM_{2.5}$ were not correlated [179], even though one study found a positive non-significant relation [180]. NO_2 is the most controversial environmental factor studied in the context of PD, with one study finding positive associations [180, 181] and another not finding any association [179]. Another study found an association only when adjusting for smoking [180] and one study found it to be both associated and not associated when comparing different cities [178]. Only one study explored sun exposure and found it to be associated when adjusting for age, positively in older subjects, and inversely in subjects below 70 years [182].

Three studies were included in the context of Dementia, which comprehended Alzheimer's, one of which found a positive relation to temperature [183]. The other two studies agreed that NO_2 and PM_{10} were positively correlated with Dementia [184, 185]. Furthermore, a study found CO to be inversely correlated with Dementia and SO_2 and O_3 not to be correlated [184].

Concerning amyotrophic lateral sclerosis (ALS), two studies were found but with no overlapping environmental factors. A study found NO (nitrogen oxide), NO_2, and Cd (cadmium) to be positively correlated with ALS and O_3 to be inversely correlated, while benzopyrene was not found to be correlated [186]. The other study found positive correlations between ALS and both precipitation and sun exposure [187].

As for motor neuron disease, in the only study included, it was found to be correlated with Pb [188].

Most studies simply used administrative divisions as geographic analysis to assess exposure to atmospheric factors, but GIS and remote sensing were also broadly used to spatially compute and model those factors. Specifically, studies that used remote sensing focused on NASA's (National Aeronautics and Space Administration) satellite TOMS (Total Ozone Mapping Spectrometer) to compute the amount of sun exposure [155, 159, 162, 165].

3.4. Discussion

Multiple sclerosis is known to be tied to latitude, thus raising the suggestion that sun exposure, as well as temperature and humidity, are probably the environmental factors behind that relationship [163]. Sun exposure is one of the strongest correlates to latitude [117] and its relation to MS may be explained by biological mechanisms that suggest that ultraviolet radiation may be immunosuppressive [189].

However, these mechanisms seem to be most relevant in early life [118]. The studies retrieved in this systematic review mostly corroborated this hypothesis [156-162], although three studies did not [163-165], with one even attributing the relationship between sun exposure and MS to the high correlation between sun exposure and latitude [163]. Amyotrophic lateral sclerosis also has an environment hypothesis linked to the genetics of the disease and its prevalence. However, studies found in this systematic review focused mostly on water and soil exposures, and are thus outside the scope of this systematic review [119, 137]. Another study identified a substantial gap in studies addressing environmental risk factors that may cause ALS [122]. The low incidence rate may be related to the low number of studies concerning the disease [122].

There is evidence that PD is mediated by environmental risk factors such as manganese or lead with the need to better model the exposure geographically model was noted [120]. This corroborates the findings of this systematic review. Dementia risk has been related to occupational exposure to air pollutants such as PM, CO_2, CO, and SO_2 [148], corroborating the findings in this systematic review. However, as Dementia encloses Alzheimer's disease, but is not limited to it, no studies were found that approached AD individually, and thus the results might have been slightly different if only AD were studied. The geographic distribution of the included studies shows a lack of studies in South America and Africa, creating the opportunity for research in the field of environmental epidemiology of neurodegenerative diseases in these regions. Analysing the source of data from each study closer, the constatation of the lack of studies from these regions remains. Although Africa has the lowest rates of neurodegenerative diseases in the world, the same cannot be said about South America, so this is a good region to develop further research.

The increasing number of papers published over the years reflects not only the general increase in scientific publications but also the availability of the technology to proceed with the geographic analysis of environmental factors. The rising interest and research investment in this thematic are also reflected in the increased number of studies over the last few years. In particular, papers referring to the new COVID-19 pandemic effects on neurodegenerative diseases have raised awareness of the synergy between this virus and the environment, which increases the risk of developing these diseases. Knowing the risk the atmospheric pollution already represents upon the central nervous system, and by adding the risk of COVID-19, which may penetrate the brain through similar pathways to those of air pollution, there is a higher potential for neuroinflammation and neurodegeneration [97]. Interesting positive relations between NO_2 levels and COVID-19 fatality rates have been retrieved from a spatial analysis study [190], indicating that this pollutant may weaken the immunity of the lungs, thus allowing the virus to access the bloodstream and eventually reach the brain [191]. Vitamin D has been described as having antiviral properties, and thus the lack of sun exposure may be problematic in people already with neurological conditions, such as MS and PD [192].

One possible limitation of our study might be not having found studies concerning other neurodegenerative diseases, such as Creutzfeldt-Jakob and Huntington's, therefore not covering the whole domain of neurodegenerative diseases defined at the start of this review. The lack of studies regarding other neurodegenerative diseases is due to the low number of studies using geospatial analysis when considering these diseases. Notwithstanding, most of the included studies focused on MS since it is mostly related to sun exposure—often derived from remote sensing—thus making it the disease most geographically studied. Grey literature was not screened, thus non-indexed papers may have escaped our research. Including this literature could have mitigated the lack of studies focussing on the remaining neurodegenerative diseases that were not found under the scope of this review. As such, further inclusion of this literature could pose an interesting opportunity for future research.

3.5. Conclusion

The most studied neurodegenerative disease in the context of this systematic review was MS, having over half the studies covering it. It is mostly unanimous that sun exposure is a preventive factor in the development of MS. The relationship to air pollutants remains unknown, and further investigation is needed to understand if such a relationship exists. As for the other diseases, air pollutants have become more broadly studied. Metals such as Pb and Mn were found to be related with Parkinson's disease and ALS, respectively. NO_X is the most common pollutant to be studied, but also the most ambiguous, being found to be both related and unrelated with Parkinson's disease, and positively correlated with Dementia and ALS. The geospatial analysis was mostly used to estimate exposure assessment, either by attributing a value to each administrative division and matching it to the residence of the patient or care unit location or by using remote sensing and GIS to compute the atmospheric factor's concentration or value geographically. The present systematic review provides a basis of the available evidence regarding the influence of environmental factors on neurodegenerative diseases using geospatial analysis. The ever-increasing amount of data and technological possibilities to advance on this analysis supports the development of further research on this topic. It is important to identify mitigation measures to reduce exposure to air pollution. Regulation on the levels of air pollution and accurate monitoring of these same levels could help ensure that the population isn't over-exposed to harmful components of the atmosphere.

Appendix A

Table A1 - List of the 56 excluded papers

REF	Country (year)	Title	Authors	Exclusion reasons
[101]	Poland (1969)	Epidemiological study of multiple sclerosis in western Poland	W. Cendrowski, M.Wender,W. Dominik, Z.	No ENV
[102]	Germany (1969)	Multiple sclerosis in Europe	R. C. Behrend	No ENV
[103]	Republic of South Africa (1975)	Comparative epidemiological studies of multiple sclerosis in South Africa and Japan	A. V. Bird, E. Satoyoshi	No ENV
[104]	Germany (1984)	Epidemiological investigations into multiple sclerosis in Southern Hesse	Klaus Lauer,Wolfgang Firnhaber, Robert Reining, Brigitte Leuchtweis	No ENV
[105]	Italy (1993)	Multiple sclerosis: does epidemiology contribute to providing etiological clues	Vittorio Govoni, Ezio Paolino, Susanna Malagù, Vincenza C. Monetti, Mirko Carreras	No ENV
[106]	Czech Republic (1994)	Geographic aspects in the epidemiology of multiple sclerosis	P. Lenský	No full paper
[107]	France (1995)	Epidemiology of Creutzfeldt-Jakob disease	N. Delasnerie-Laupretre, A. Alperovitch	No full paper
[108]	Hungria (1997)	Monthly distribution of multiple sclerosis patients' births	Padmanabhan Bharanidharan	No GEO
[109]	USA (1997)	The Epidemiology of Multiple Sclerosis	W. E. Hogancamp, M. Rodriguez, B. G. Weinshenker	No ENV
[110]	Canada (1999)	Parkinson's disease, multiple sclerosis and amyotrophic lateral sclerosis: The iodine-dopachrome-glutamate hypothesis	Harold D. Foster	No ENV
[111]	United Kingdom (2000)	Amyotrophic lateral sclerosis: toxins and environment	J. D. Mitchell	No ENV, No GEO
[112]	Spain (2002)	Epidemiologia genetica de la esclerosis multiple	D.F. Uría	No ENV
[113]	USA (2004)	Environmental risk factors in multiple sclerosis aetiology	Ruth Ann Marrie	No GEO
[114]	Italy (2004)	Genes environment and susceptibility to multiple sclerosis	Stefano Sotgiu, Maura Pugliatti, Maria Laura Fois, Giannina Arru, Alessandra Sanna, Maria Alessandra Sotgiu, Giulio Rosati	No ENV
[115]	USA (2005)	Autoimmunity: Multiple Sclerosis	Beau M. Ances, Nancy J. Newman, Laura J. Balcer	No ENV, No GEO
[116]	USA (2006)	Studies of Multiple Sclerosis in Communities Concerned about Environmental Exposures	Dhelia M.Williamson	No ENV
[117]	USA (2007)	Environmental risk factors for multiple sclerosis Part II Noninfectious factors	Alberto Ascherio, Kassandra L. Munger	No ENV, No GEO
[118]	United Kingdom (2008)	Environmental factors and multiple sclerosis	George C. Ebers	No ENV
[119]	France (2011)	Contribution of geolocalisation to neuroepidemiological studies incidence of ALS and environmental factors in Limousin France	F. Boumediène, M. Druet-Cabanac, B. Marin, Pierre-Marie Preux, P. Couratier	No ENV
[120]	USA (2012)	Environmental risk factors	Gill Nelson, Brad A. Racette	No GEO
[121]	USA (2012)	Predictors of Survival in Patients with Parkinson Disease	Allison W.Willis, Mario Schootman, Nathan Kung, Bradley A. Evano_, Joel S. Perlmutter, Brad A. Racette	No ENV

[122]	USA (2012)	Spatial clustering of amyotrophic lateral sclerosis and the potential role of BMAA	Tracie A. Caller, Nicholas C. Field, Jonathan W. Chipman, Xun Shi, Brent T. Harris, ElijahW. Stommel	No ENV
[123]	United Kingdom (2013)	Epidemiology of neurologically disabling disorders	Alan Tennant	No ENV, No GEO
[124]	Kuwait (2013)	Risk factors for multiple sclerosis in Kuwait a population-based case control study	Hanan H. Al-Afasy, Mohammed A. Al-Obaidan, Yousef A. Al-Ansari, Sarah A. Al-Yatama, Mohammed S. Al-Rukaibi, Nourah I. Makki, Anita Suresh, Saeed Akhtar	No ENV, No GEO
[125]	Spain (2014)	Geographical analysis of the sporadic Creutzfeldt Jakob disease distribution in the autonomous community of the Basque Country for the period 1995 2008	Saioa Chamosa, Ibon Tamayo, José M. Arteagoitia-Axpe, Ramón A. Juste, Ana Belém Rodríguez-Martínez, Juan J. Zarranz-Imirizaldu	No ENV
[126]	USA (2015)	Association Between Alzheimer Dementia Mortality Rate and Altitude in California Counties	Stephen Thielke, Christopher G. Slatore, William A. Banks	No ENV
[127]	United Kingdom (2015)	Geographical variation in Dementia: examining the role of environmental factors in Sweden and Scotland	Tom C. Russ, Margaret Gatz, Nancy L. Pedersen, Jean Hannah, GrantWyper, G. David Batty, Ian J. Deary, John M. Starr	No ENV
[128]	Norway (2015)	Socio economic factors and immigrant population studies of multiple sclerosis	P. Berg-Hansen, E. G. Celius	No ENV
[129]	Canada (2015)	The EnvIMS Study Design and Methodology of an International Case Control Study of Environmental Risk Factors in Multiple Sclerosis	Sandra Magalhaes, Maura Pugliatti, Ilaria Casetta, Jelena Drulovic, Enrico Granieri, Trygve Holmøy, Margitta T. Kampman, Anne-Marie Landtblom, Klaus Lauer, Kjell-Morten Myhr, Maria Parpinel, Tatjana Pekmezovic, Trond Riise, David Wolfson, Bin Zhu, ChristinaWolfson	No ENV, No GEO
[130]	Sweden (2015)	Vitamin D and multiple sclerosis from epidemiology to prevention	P. Sundström, J. Salzer	No ENV
[131]	USA (2016)	Environmental control of autoimmune inflammation in the central nervous system	Veit Rothhammer, Francisco J. Quintana	No NEURO
[132]	USA (2016)	Epidemiology of Multiple Sclerosis From Risk Factors to Prevention An Update	Alberto Ascherio, Kassandra L. Munger	No ENV
[133]	USA (2016)	Fine Particulate Matter Residential Proximity to Major Roads and Markers of Small Vessel Disease in a Memory Study Population	Elissa H. Wilker, Sergi Martinez-Ramirez, Itai Kloog, Joel Schwartz, Elizabeth Mostofsky, Petros Koutrakis, Murray A. Mittleman, Anand Viswanathan	No NEURO
[134]	United Kingdom (2016)	Geographical Variation in Dementia Mortality in Italy New Zealand and Chile The Impact of Latitude Vitamin D and Air Pollution	Tom C. Russ, Laura Murianni, Gloria Icaza, Andrea Slachevsky, John M. Starr	No ENV
[135]	Ecuador (2016)	Prevalence of multiple sclerosis in Latin America and its relationship with European migration	Edgar Correa, Víctor Paredes, Braulio Martínez	No ENV
[61]	USA (2016)	Seeking environmental causes of neurodegenerative disease and envisioning primary prevention	Peter S. Spencer, Valerie S. Palmer, Glen E. Kisby	No ENV
[136]	USA (2017)	Associations of Spatial Disparities of Alzheimer's Disease Mortality Rates and Soil Selenium Sulfur Concentrations and Risk Factors in the United States	Hongbing Sun	No ENV

[137]	Italy (2017)	Incidence of amyotrophic lateral sclerosis in the province of Novara Italy and possible role of environmental pollution	Marina Tesauro, Michela Consonni, Tommaso Filippini, Letizia Mazzini, Fabrizio Pisano, Adriano Chiò, Aniello Esposito, Marco Vinceti	No ENV
[138]	France (2017)	Small area distribution of multiple sclerosis incidence in western France in search of environmental triggers	Karima Hammas, Jacqueline Yaouanq, Morgane Lannes, Gilles Edan, Jean-François Viel	No ENV
[139]	Spain (2017)	The Geography of the Alzheimer's Disease Mortality in Spain Should We Focus on Industrial Pollutants Prevention	Èrica Martínez-Solanas, Montse Vergara-Duarte, Miquel Ortega Cerdà, Juan Carlos Martín-Sánchez, Maria Buxó, Eduard Rodríguez-Farré, Joan Benach, Glòria Pérez	No ENV
[140]	Iran (2017)	The relationship between the amount of radiation, relative humidity, and temperature with the risk of multiple sclerosis in Isfahan province, Iran, during the years 2001–2014	A. Karimi, A. Delpisheh, F. Ashtari, K. Sayehmiri, R. Meamar	No full paper
[141]	Netherlands (2018)	Assessment of residential environmental exposure to pesticides from agricultural fields in the Netherlands	Maartje Brouwer, Hans Kromhout, Roel Vermeulen, Jan Duyzer, Henk Kramer, Gerard Hazeu, Geert de Snoo, Anke Huss	No ENV
[142]	France (2018)	Environmental factors in the development of multiple sclerosis	L. Michel	No ENV, No GEO
[143]	Iran (2018)	Estimated incidence rate of multiple sclerosis and its relationship with geographical factors in Isfahan province between the years 2001 and 2014	Fereshteh Ashtari, Arezoo Karimi, Ali Delpisheh, Rokhsareh Meamar, Kourosh Sayehmiri, Salman Daliri	No ENV
[144]	Australia (2018)	Health outcomes and lifestyle in a sample of people with multiple sclerosis HOLISM Longitudinal and validation cohorts	Tracey J.Weiland, Alysha M. De Livera, Chelsea R. Brown, George A. Jelinek, Zoe Aitken, Steve L. Simpson Jr., Sandra L. Neate, Keryn L. Taylor, Emily O'Kearney,William Bevens, Claudia H. Marck	No ENV
[145]	USA (2019)	ALS and environment Clues from spatial clustering	P. S. Spencer, E. Lagrange,W. Camu	No ENV
[146]	Italy (2019)	Amyotrophic Lateral Sclerosis Descriptive Epidemiology: The Origin of Geographic Difference	Giancarlo Logroscino, Marco Piccininni	No ENV
[147]	Iran (2019)	Can environmental factors increase the risk of multiple sclerosis A narrative review	Hoda Naghshineh, Seyed Mohammad Masood Hojjati, Ali Alizadeh Khatir, Payam Saadat, Alijan Ahmadi Ahangar	No ENV
[148]	USA (2019)	Increased Dementia Mortality in West Virginia Counties with Mountaintop Removal Mining	A. K. Salm, Michael J. Benson	No ENV
[67]	China (2019)	The interplay of aging genetics and environmental factors in the pathogenesis of Parkinson's disease	Shirley Yin-Yu Pang, PhilipWing-Lok Ho, Hui-Fang Liu, Chi-Ting Leung, Lingfei Li, Eunice Eun Seo Chang, David Boyer Ramsden, Shu-Leong Ho	No ENV
[149]	Turkey (2002)	The etiology and the epidemiology of multiple sclerosis	Meral Mirza	No GEO
[150]	Russia (2009)	Risk factors of multiple sclerosis development in the population of the Rostov region	Z.A. Goncharova, V.A. Baliazin	No full paper
[151]	China (2011)	Reference value of left atrial diameter of presenile women and geographical factors based on principal component analysis	J. Jing, M. Ge, A.Z. Zhao, G.Z. Liu, S.T. Xiang, X.Wang, Y.P. Zhang	No full paper

[152]	Russia (2014)	Multiple sclerosis in the Bashkortostan Republic and the Rostov region: A comparative epidemiologic study	K. Z. Bakhtiyarova, Z. A. Goncharova	No full paper
[153]	China (2015)	Clinical features of amyotrophic lateral sclerosis in south–west China	Qianqian Wei, Xueping Chen, Zhenzhen Zheng, Rui Huang, Xiaoyan Guo, Bei Cao, Bi Zhao, Huifang Shang	No ENV
[154]	USA (2016)	Multiple Sclerosis Epidemiology	M.T.Wallin, J.F.Kurtzke	No ENV

4. Geospatial analysis of environmental atmospheric risk factors in neurodegenerative diseases: A systematic review update

Following the previous study, three years passed and the feeling that much research could have been made meanwhile in the same context led to a quick search with the same queries, on the same databases, from when the last search had been performed to the end of January 2023. This proved fructiferous since as many studies were included in this update as in the original review and new environmental factors and disease focus were obvious.

Mariana Oliveira, André Padrão, Ana Cláudia Teodoro, Alberto Freitas, Hernâni Gonçalves

Systematic Reviews

Submitted

-

3.7

Q1 (Medicine - Miscellaneous)

Abstract: Following up the previously published systematic review on the same topic, and realizing that new studies and evidence had emerged on the matter, we conducted an update on the same research terms. We searched PubMed, Web of Science and Scopus for all scientific studies considering the following three domains: neurodegenerative disease, environmental atmospheric factor, and geographical analysis; using the same keywords as in the previously published systematic review. From February 2020 to February 2023, 35 papers were included versus 34 in the previous period, with Dementia (including Alzheimer's disease) being the most focused disease (60.0%) in this update, opposed to multiple sclerosis on the last review (55.9%). Overall, multiple sclerosis remains the most studied (43.5%), followed by Dementia (34.8%). Also, environmental pollutants such as PM2.5 and NO2 have gained prominence, being represented in 65.7% and 42.9% of the new studies, opposed to 23.5% and 29.4% in the previous review, compared to environmental factors such as sun exposure (5.7% in the update vs 23.5% in the original). Globally, most studies approached PM2.5 (44.9%) and NOX (36.2%). The mostly used geographic approach remained the patient's residence (82.9% in the update vs 41.2% in the original and 62.3% in total), and remote sensing was used in 45.7% of the new studies versus 38.2% in the original review, with 42.0% of studies using it globally, being the second most common approach, usually to compute the environmental variable. This review has been registered in PROSPERO with the number CRD42020196188, and is currently set as updating.

Keywords: neurodegenerative; environment; geospatial; epidemiology; systematic review; systematic review update

4.1. Introduction

The pathological changes suffered by the human brain during the ageing process leads to a range of neurodegenerative disorders [61], which are characterized by a progressive loss or damage of neuronal cell leading to compromised brain function. While the world's population over the age of 60 is expected to reach 22% until 2050, the concern about the future of neurodegenerative diseases arises and the development of better suitable health systems for elder population is urged forward. Alzheimer's disease (AD) is the most prevalent of neurodegenerative pathologies worldwide and is responsible for an extensive cognitive damage which usually affects daily chores [67]. Parkinson's disease (PD) closely follows AD in prevalence and leads to both motor and non-motor symptoms due to dopaminergic neuronal loss. Affecting around 2.5 million people globally, multiple sclerosis (MS) is the third most prevalent neurodegenerative disease and is usually an autoimmune response causing an inflammatory demyelination of the brain [65].

Other risk factors are known in the development of neurodegenerative disorders, including gender, aggravated clinical history - hypertension, diabetes, cranial injury and tumours – smoking and drinking

[68]. Nevertheless, the knowledge on the development of these pathologies remains incomplete, and it is thought environmental factors may have a contribution [68]. With 91% of the world's population living under high pollution levels, with air quality levels above the established limits for health safety by the World Health Organization (WHO) [69], the urge for studying its implications on the population's health is rising. Recent studies have revealed that atmospheric pollution can trigger mechanisms responsible for neurodegenerative diseases [92, 193, 194].

A useful tool to both measure and analyse environmental exposure to air pollution is remote sensing data/techniques, and it has been increasingly used in epidemiological studies with the number publications featuring remote sensing applied to health increasing from 5.6% to 13.3% between 2007 and 2016 [26]. Geospatial analysis allows to integrate information on health, environmental data and socio-economic information at a local and global focus. Future research may then maximize the data used and establish future policies, which may be critical in preventing and progressing neurodegenerative diseases.

Following our previous systematic review [24], this systematic review update aims to identify new studies concerning neurodegenerative diseases and their environmental atmospheric risk factors through a geographical approach.

4.2. Materials and methods

The methods for this review are similar to the ones previously used in the original review [24]. In short, we searched PubMed, Web of Science and Scopus databases from the 1st February 2020 until the 31st January 2023, using the following keywords: [((alzheimer*) OR (ataxia*) OR (Chorea Minor) OR (creutzfeldt*) OR (Dementia*) OR (Frontotemporal) OR (Guillian-barre syndrome) OR (Huntington*) OR (kennedy* disease) OR (Lewy*) OR (motor neuron*) OR (Myotonic dystrophy) OR (neurodegen*) OR (parkinson*) OR (pick's) OR (Prion) OR (progressive AND palsy) OR (progressive muscular atrophy) OR (sclerosis*) OR (*senile) OR (Spinal Atrophy)) AND ((atmospher*) OR (carbon) OR (environment*) OR (humidity) OR (meteorologic*) OR (nitrogen) OR (ozone) OR (particulate*) OR (PM2*) OR (pollut*) OR (sulphur*) OR (surface pressure) OR (temperature)) AND ((drone) OR (geograph*) OR (imagery) OR (landsat) OR (map) OR (mapping) OR (modis) OR (remote sens*) OR (satellite) OR (sentinel) OR (spatial) OR (topologic*))]. This review has been previously registered in PROSPERO with the number CRD42020196188, and has currently been set as updated.

The search returned 3510 results and 2282 abstracts were screened after duplicates were handled. Records were included if they contained all three domains considered in this review: neurodegenerative disease, atmospheric pollutant or factor and geographical approach, and were excluded otherwise. No language restrictions were used. Exclusion criteria encompassed studying mechanisms and biospecimen

behind a neurodegenerative disease, soil and water pollutants and simply stating a geographical area without further comparing nor analysing it. Abstracts were independently read by two authors to apply the inclusion and exclusion criteria here explained and select eligible papers for full-text screening. Full-text information retrieving was completed using semi-structured forms equal to the ones used in the original review, although papers made it necessary to add fields such as Bayesian methods and PM_1 and black carbon (BC) as pollutants.

4.3. Results

4.3.1. *Identification, screening and assessment*

Of the 2282 articles initially screened, only 35 were included in the final study. The number of published papers on the scope of this review has clearly increased through the years, as visible in the graphic of Figure 5. 2021 was the year with the most published articles in this scope, with 17 papers, while until 2015 only 15 had ever been published. The selection process is summarized using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) presented by a flow diagram (Figure 6). The discrepancy in the total number of excluded studies and the sum of the exclusion reasons categories is due to the overlapping of exclusion reasons. The full list with the studies resultant from the original review and the present update is provided on Table B1 in the supplementary matterials.

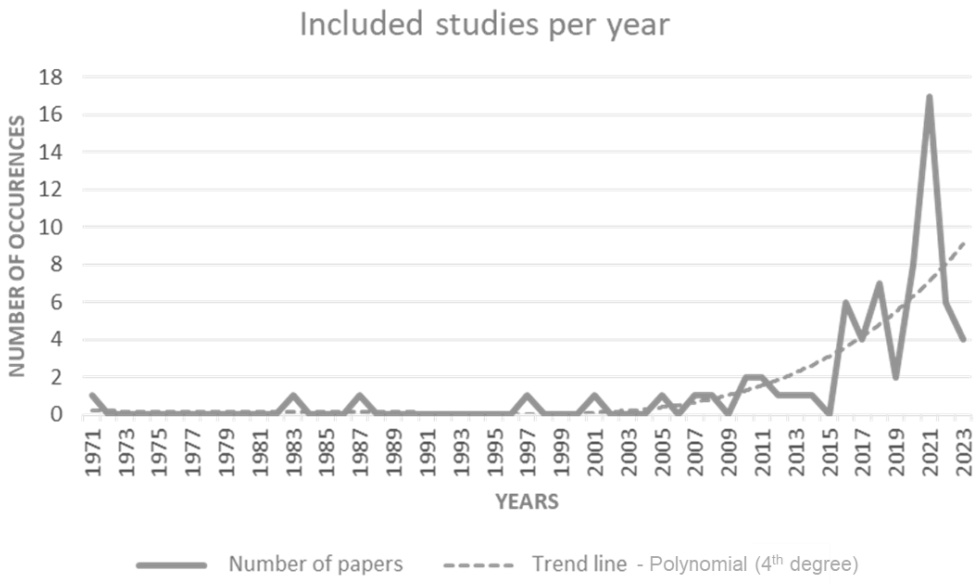


Figure 5 - Number of published articles on this review's scope throughout the years.

Overall, as in Table 6, most papers studied multiple sclerosis (43.5%) and Dementia – including AD – (34.8%) as neurodegenerative diseases and compared them with either $PM_{2.5}$, NO_X and PM_{10} (44.9%,

36.2% and 23.2%, respectively). Using the patient's residence as an estimator of exposure was used in 62.3% of the studies and remote sensing was used in 42.0%. As categories from each domain are not mutually exclusive, the sum of all categories may be higher than the total number of articles. The graphics in Figure 7 represent the number of articles in each domain with the corresponding categories.

The main methodological characteristics of the included studies are summarized on Table 7. A more detailed table with each paper's characteristics is included in supplementary materials', in Table B2. The country of origin of the studies is represented in Figure 8. The country with the most papers published was the United States of America, with 19 papers overall, 8 of which about Dementia. It was also the country with the most variety of neurodegenerative diseases studied, which included Dementia, MS, PD, amyotrophic lateral sclerosis (ALS) and paediatric multiple sclerosis (PMS). A list of all countries with their corresponding published papers is available on Table 8. As several times countries have joined their efforts to study a specific disease, the total number of studies included and the sum of the published papers by country does not match.

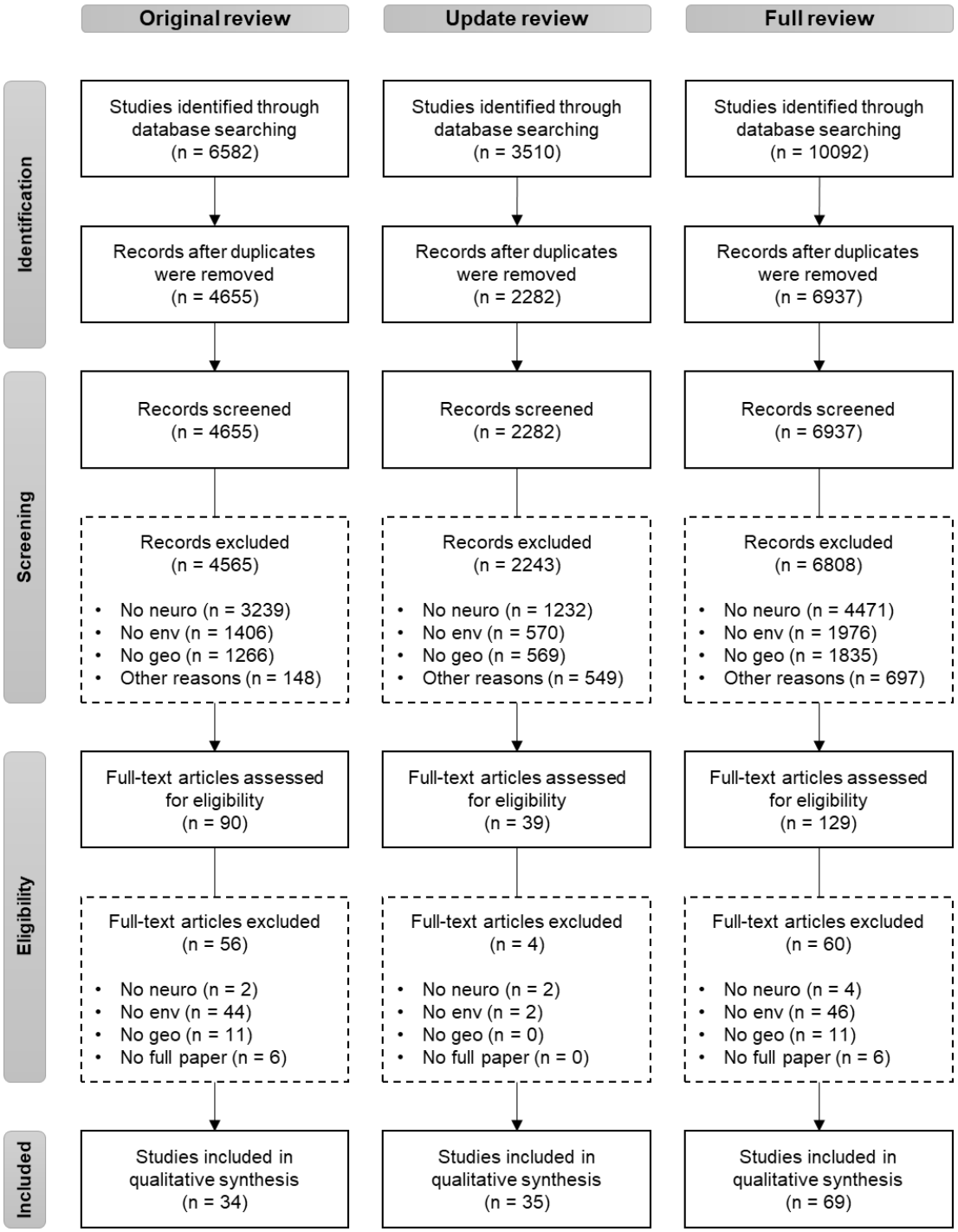


Figure 6 - Systematic review PRISMA flowchart, including original and update reviews and their junction. Neuro: neurodegenerative disease; env: environmental factor or pollutant, geo: geographic approach.

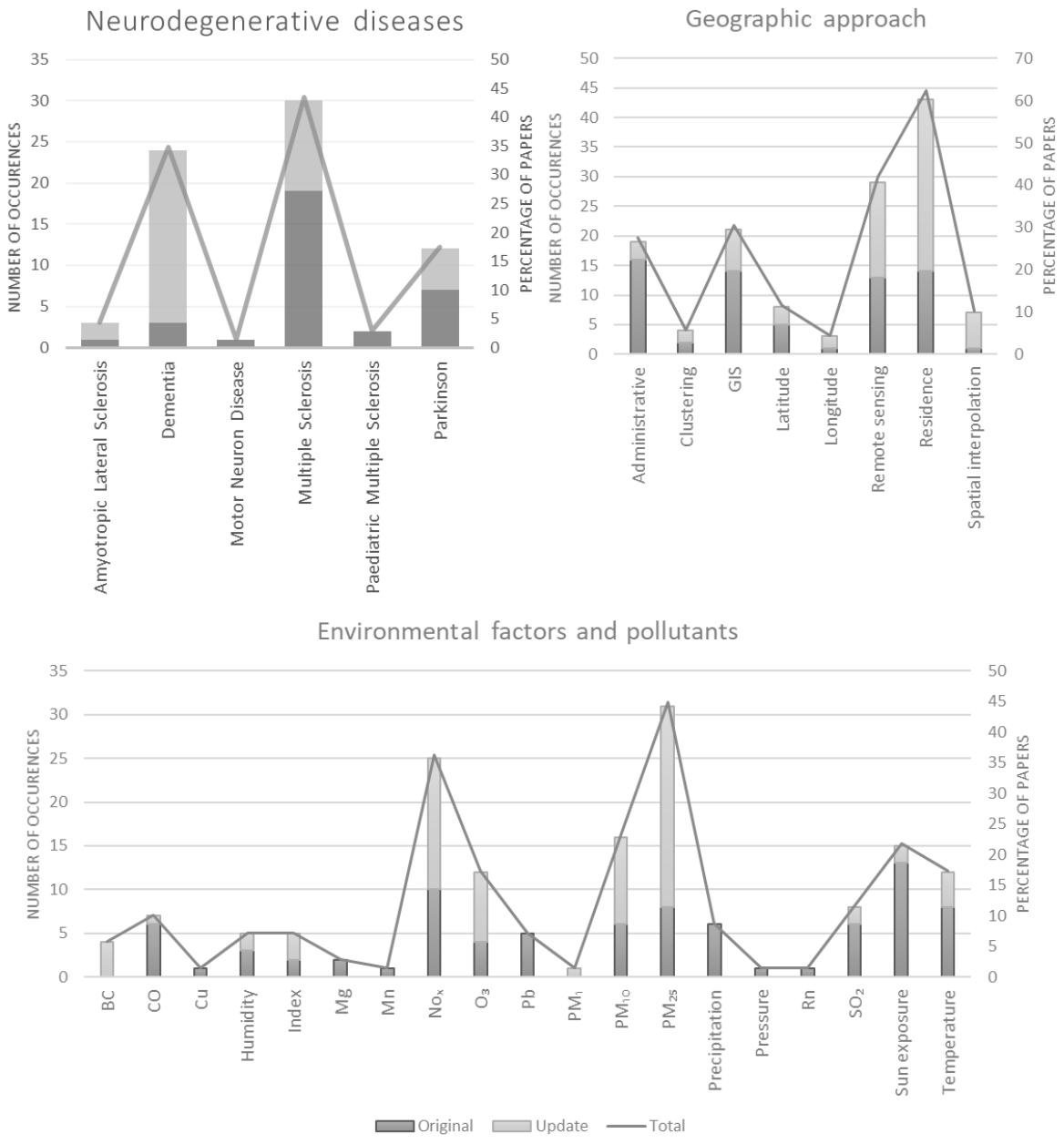


Figure 7 - Studies in each category by domain: neurodegenerative diseases, geographic approach and environmental factors and pollutants.

Table 6 - Number and percentage of papers included in the update for each domain categories.

	Update		Total	
	N	**%**	**N**	**%**
NEURO				
Amyotrophic Lateral Sclerosis	2	5.7	3	4.3
Dementia	21	60.0	24	34.8
Motor Neuron Disease	0	0.0	1	1.4
Multiple Sclerosis	11	31.4	30	43.5
Paediatric Multiple Sclerosis	0	0.0	2	2.9
Parkinson	5	14.3	12	17.4
ENV				
BC	4	11.4	4	5.8
CO	1	2.9	7	10.1
Cu	0	0.0	1	1.4
Humidity	2	5.7	5	7.2
Index	3	8.6	5	7.2
Mg	0	0.0	2	2.9
Mn	0	0.0	1	1.4
No_x	15	42.9	25	36.2
O_3	8	22.9	12	17.4
Pb	0	0.0	5	7.2
PM_1	1	2.9	1	1.4
PM_{10}	10	28.6	16	23.2
PM_{25}	23	65.7	31	44.9
Precipitation	0	0.0	6	8.7
Pressure	0	0.0	1	1.4
Rn	0	0.0	1	1.4
SO_2	2	5.7	8	11.6
Sun exposure	2	5.7	15	21.7
Temperature	4	11.4	12	17.4
GEO				
Administrative	3	8.6	19	27.5
Clustering	2	5.7	4	5.8
GIS	7	20.0	21	30.4
Latitude	3	8.6	8	11.6
Longitude	2	5.7	3	4.3
Remote sensing	16	45.7	29	42.0
Residence	29	82.9	43	62.3
Spatial interpolation	6	17.1	7	10.1

Table 7 - Study methodological characteristics summary: limitations, study design, statistical methods and effect measures

	Original		Update		Overall	
	N	**%**	**N**	**%**	**N**	**%**
Limitations						
None given by the authors	5	14.7	2	5.7	7	10.1
Conflict of interests	1	2.9	0	0.0	1	1.4
Confounding factors	15	44.1	18	51.4	33	47.8
Ecological bias	3	8.8	4	11.4	7	10.1
Exposure assessment	18	52.9	29	82.9	47	68.1
Interpolation	3	8.8	0	0.0	3	4.3
Memory bias	3	8.8	1	2.9	4	5.8
Migration of patients	4	11.8	2	5.7	6	8.7
Referral bias	2	5.9	0	0.0	2	2.9
Sampling bias	7	20.6	17	48.6	24	34.8
Statistical bias	13	38.2	6	17.1	19	27.5
Study design issues	3	8.8	2	5.7	5	7.2
Survival bias	1	2.9	1	2.9	2	2.9
Time related bias	9	26.5	7	20.0	16	23.2
Unassessed patients	6	17.6	11	31.4	17	24.6
Study design						
Case-control	9	26.5	6	17.1	15	21.7
Case-crossover	0	0.0	4	11.4	4	5.8
Cohort	4	11.8	21	60.0	25	36.2
Cross-sectional	12	35.3	1	2.9	13	18.8
Ecological	6	17.6	2	5.7	8	11.6
Methodological	2	5.9	0	0.0	2	2.9
Review	1	2.9	1	2.9	2	2.9
Statistical methods						
None	2	5.9	1	2.9	3	4.3
ANOVA	2	5.9	1	2.9	3	4.3
Bayesian	0	0.0	2	5.7	2	2.9
Chi-squared	5	14.7	1	2.9	6	8.7
Clustering	2	5.9	2	5.7	4	5.8
Correlation	22	64.7	2	5.7	24	34.8
Cox regression	4	11.8	11	31.4	15	21.7
Linear regression	12	35.3	8	22.9	20	29.0
Logistic regression	10	29.4	13	37.1	23	33.3
Poisson regression	3	8.8	1	2.9	4	5.8
Sensitivity analysis	10	29.4	8	22.9	18	26.1
Spatial autoregressive	1	2.9	0	0.0	1	1.4
T-test	5	14.7	2	5.7	7	10.1
Effect measurements						
None	2	5.9	0	0.0	2	5.9
Coefficients	14	41.2	5	14.3	14	41.2
Correlation	20	58.8	3	8.6	20	58.8
Hazard ratio	3	8.8	4	11.4	3	8.8
Odds ratio	11	32.4	11	31.4	11	32.4
Prevalence	4	11.8	13	37.1	4	11.8
Relative risk	4	11.8	5	14.3	4	11.8

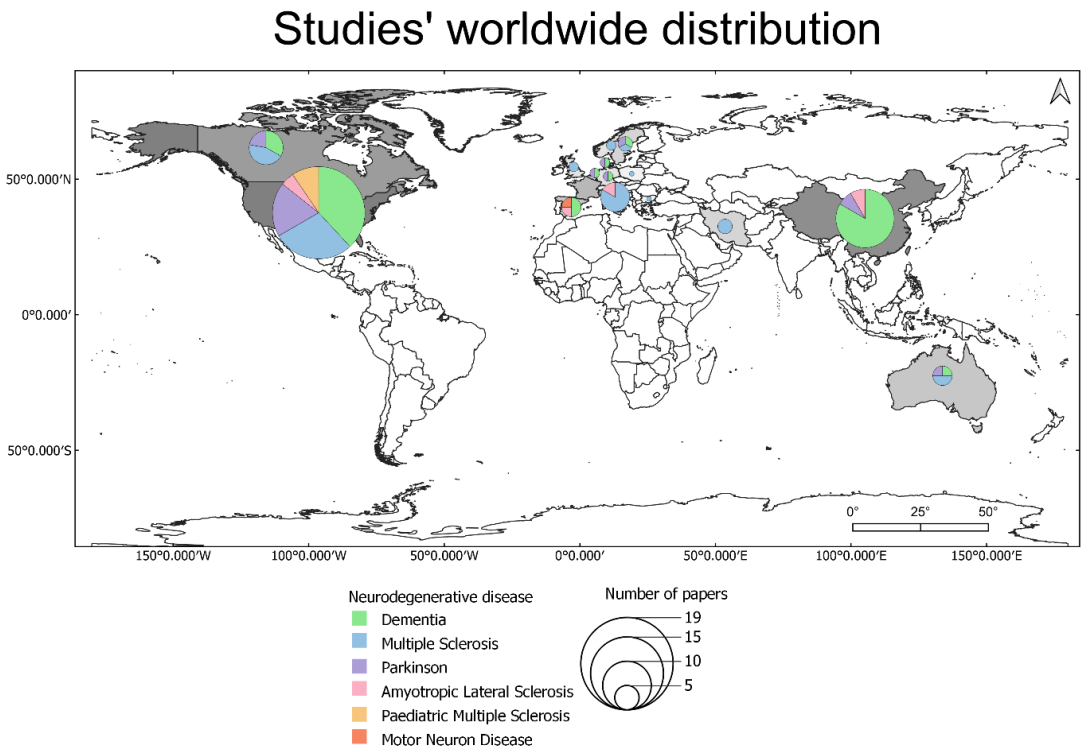


Figure 8 - Country of origin of the studies and the diseases studied in each country.

Table 8 - Country of origin with the number of studies published (N) and the respectively studied diseases. Some studies have investigators or focus on data from more than one country.

Country (N)	Amyotrophic Lateral Sclerosis	Dementia	Motor Neuron Disease	Multiple Sclerosis	Paediatric Multiple Sclerosis	Parkinson
USA (19)	1	8		6	2	4
China (12)	1	10				1
Canada (7)		3		4		2
Italy (6)	1			5		
France (5)		2		1		2
Australia (4)		1		2		1
Spain (4)	1	2	1			
Austria (3)		1		1		1
Iran (3)				3		
Sweden (3)		1		1		1
Denmark (2)		1				1
Germany (2)		1				1
Israel (2)				2		
Netherlands (2)		1				1
Norway (2)				2		
UK (2)				2		
Bulgaria (1)				1		
Poland (1)				1		
Switzerland (1)						1

4.3.2. *Qualitative synthesis*

The following analysis was performed taking into account all included articles from the original and the updated systematic review. Further detailed information regarding the results presented in each study is presented on Table B2 in the supplementary materials.

Amyotrophic lateral sclerosis: ALS was approached on 4 studies [186, 187, 195, 196]. Of these, 3 studied its interaction with PM_{10} [186, 195, 196] and 2 analysed $PM_{2.5}$ [186, 196]. Both pollutants generally showed a negative but not significant association. A study [187] identified a significant negative association between ALS and sun exposure and significant positive association with precipitation and humidity, whilst no significant associations were identified with both temperature and pressure. Another study [186] found a positive significant association with NO_x, with an odds ratio (OR) between 1.872 and 2.703. This study found no more significant associations with the remaining pollutants studied: SO_2, CO, O_3 and Pb.

Dementia: Dementia, including AD, was studied 24 times [25, 183-185, 197-216], 18 of which focused on $PM_{2.5}$, with 13 finding a significant positive association between Dementia and $PM_{2.5}$ [185, 198, 200, 201, 204-210, 214, 215], with hazard ratios ranging from 1.02 [214] to 1.67 [208] and the remaining not finding a significant association [202, 211, 212, 216]. No association was ever found between Dementia and temperature [183, 199], humidity [199] nor black carbon [202, 204, 211]. The outcomes concerning NO_x and NO_2 were controversial, with 6 papers finding a positive association [184, 185, 198, 203, 213, 216], 2 finding a negative association [199] (in particular with Alzheimer [202]) and 5 finding no significant association [201, 202, 204, 211, 212]. Furthermore, a study [213] approached several air pollutants (PM_{10}, NO_2, SO_2, CO and O_3) and found a significant positive association with all of them.

Motor neuron disease: Only one study focused on motor neuron disease (MND) [188] and it positively related the disease to lead (Pb), with a correlation coefficient of 0.824.

Multiple sclerosis: MS was studied 30 times [155-173, 202, 217-226], 13 of which in relation to sun exposure, where half the studies found a significant negative association [156-160, 163], half found no significant association [155, 161, 162, 164, 165, 225] and 1 found a positive association [217]. Other environmental factors such as temperature and precipitation brought up further conflicts: some studies found a negative relation to temperature [156, 163, 166, 221], opposing others [157, 158, 166] and precipitation was mostly inconclusive [156, 158]. Concerning the air pollutants, the results were more in line with each other, with most studies finding at least one significant positive association [167-171, 202, 222-224, 226].

Paediatric multiple sclerosis: Two studies focused on PMS [174, 175], one positively relating an air quality index to the disease [174] and another one studying several pollutants, positively associating the

disease with $PM_{2.5}$, SO_2, CO and Pb, and not finding significative association with PM_{10}, NO_X and O_3 [175].

Parkinson's disease: Finally, PD was studied 12 times [176-182, 202, 207, 210, 227, 228], mostly related to air pollutants, where significant positive relations were found between the disease and PM_{10} [227], $PM_{2.5}$ [202, 210, 228], NO_X/NO_2 [181, 202, 227] and CO [181]. No associations were found with either black carbon [202, 228], copper [176], lead [176] and manganese [178]. Negative associations were also found in relation to sun exposure [182] and O_3 [228].

4.4. Discussion

This systematic review has retrieved as many studies in this update, concerning 2020 to 2023 as in the original review, from inception to 2020, using the same methods, suggesting an increased concern in relating neurodegenerative diseases to, primarily, air pollution. Also, an increasing trend of studying Dementia opposed to MS was clearly observed, which could indicate a new focus on less studied diseases. This updated version of the systematic review has hence brought new insights on the theme.

ALS studies are scarce, as well as MND and PMS ones, probably due to the low incidence of these diseases, and gaps addressing the related environmental risk factors have been found [122]. Considering the three most prevalent diseases, Dementia (including AD) has nearly come level with MS in terms of the number of studies (24 versus 28), and PD, the second most prevalent neurodegenerative disease worldwide, has been studied not even half the times (10). Overall, all neurodegenerative diseases were positively related to air pollutants such as PM_{10}, $PM_{2.5}$, NO_2, SO_2 and CO [167-170, 175, 176, 181, 184-186, 198-202, 204-210, 213-216, 222-224, 226-228].

The lack of studies in South America and Africa rises the concern of whether this group of diseases is being overlooked in these areas. Although Africa has the lowest rates of neurodegenerative diseases in the world, the same cannot be said about South America, and so further research on this region is advised.

Our study has some limitations. For instance, despite this being an update, rarer diseases such as Creutzfeldt-Jakob and Huntington's are still unstudied in the domains of this review. It is thus identified as a potential future study case to focus on the diseases not yet found. Also, despite not having excluded any study based on its language, no grey literature was analysed, and it could potentially provide further studies of interest. However, the validity of these studies might have been limited, and its conclusions may have induced further errors.

4.5. Conclusion

The most studied neurodegenerative diseases were in line with the most prevalent ones worldwide. It is mostly unanimous that environmental pollutants significantly influence the incidence of these pathologies, increasing their rates. Particulate matter and nitric oxides were the most studied pollutants and have mostly contributed positively to the rates of neurodegenerative diseases. The present systematic review update provides an insight of the evidence being made regarding the influence of environmental factors on neurodegenerative diseases using geospatial analysis. As with the original systematic review, the ever-increasing amount of data support the development of further research on this topic. Less prevalent diseases such as ALS, MND and PMS have been targeted as less studied, as well as the regions of South America and Africa, and are an interesting starting point for future works.

Appendix B

Table B1 - Detailed characteristics of each included study. ALS: amyotrophic lateral sclerosis; GIS: geographic information system; MS: multiple sclerosis.

Ref.	Year	Country	Neurodegenerative disease	Environmental factor	Geographic approach	Study design	Limitations	Statistical methods	Outcome measurements
[187]	2013	Taiwan	ALS	Sun exposure, Temperature, Precipitation, Humidity, Pressure	Administrative	Case-control	Exposure assessment, Migration of patients, Sampling bias, Time related bias	Correlation, Spatial autoregressive, Clustering	Correlation, Coefficients
[186]	2018	Spain	ALS	PM_{10}, $PM_{2.5}$, NO_X, SO_2, CO, O_3, Pb	GIS	Case-control	Ecological bias, Exposure assessment, Sampling bias, Statistical bias	T-test, Chi-squared, Linear regression, Sensitivity analysis	Prevalence, Odds ratio
[195]	2021	Italy	ALS	PM_{10}	GIS, Residence	Case-control	Exposure assessment, Sampling bias, Statistical bias	Logistic regression	Odds ratio
[196]	2023	USA	ALS	PM_{10}, $PM_{2.5}$, NO_X, NO_2, SO_2, O_3	GIS, Residence, Spatial interpolation	Case-control	Exposure assessment, Sampling bias	Logistic regression	Relative risk
[185]	2017	Canada	Dementia	$PM_{2.5}$, NO_X, O_3	GIS, Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Time related bias, Unassessed patients	Cox regression, Sensitivity analysis	Hazard ratio
[25]	2018	USA	Dementia	$PM_{2.5}$, O_3	Residence, Spatial interpolation	Cohort	Exposure assessment, Survival bias, Time related bias	Linear regression, Sensitivity analysis	Correlation
[183]	2019	USA	Dementia	Temperature	Administrative, Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Statistical bias	Correlation, Cox regression, Sensitivity analysis	Correlation, Hazard ratio

[184]	2019	Taiwan	Dementia	PM_{10}, NO_2, SO_2, CO, O_3	Clustering	Case-control	Confounding factors, Exposure assessment, Statistical bias, Unassessed patients	Correlation, Logistic regression, Sensitivity analysis	Odds ratio
[197]	2020	Australia, Spain, China	Dementia	Temperature, $PM_{2.5}$, NO_2	GIS, Residence	Cohort	Sampling bias, Time related bias	Logistic regression	Coefficients
[198]	2020	Spain	Dementia	PM_{10}, $PM_{2.5}$, NO_2, NDVI	GIS, Residence	Cohort	Exposure assessment, Sampling bias, Statistical bias	Linear regression	Correlation
[199]	2020	China	Dementia	Temperature, Humidity, Index, PM_{10}, NO_X, O_3	Remote sensing	Cohort	Exposure assessment, Time related bias	Logistic regression	Odds ratio
[200]	2020	Taiwan	Dementia	PM_{10}, O_3	GIS, Residence, Spatial interpolation	Cohort	Confounding factors, Exposure assessment	Cox regression	Hazard ratio
[201]	2020	Canada	Dementia	PM_{10}, NO_2	Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Time related bias, Unassessed patients	Cox regression	Hazard ratio
[203]	2021	China	Dementia	PM_1, NO_2	Remote sensing	Case-control	Confounding factors, Exposure assessment, Time related bias	Logistic regression	Odds ratio
[204]	2021	France	Dementia	$PM_{2.5}$, NO_2, BC	Residence, Remote sensing	Cohort	Exposure assessment, Migration of patients, Sampling bias	Cox regression, Sensitivity analysis	Hazard ratio
[205]	2021	China	Dementia	$PM_{2.5}$	Residence, Remote sensing	Cohort	Exposure assessment, Sampling bias, Unassessed patients	Cox regression, Sensitivity analysis	Hazard ratio

[206]	2021	China	Dementia	$PM_{2.5}$	Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Sampling bias, Unassessed patients	Cox regression	Hazard ratio
[208]	2021	USA	Dementia	$PM_{2.5}$	Residence	Cohort	Confounding factors, Exposure assessment, Statistical bias	Bayesian	Hazard ratio
[209]	2021	China	Dementia	PM_{10}, $PM_{2.5}$	Remote sensing	Cross-sectional	Confounding factors, Sampling bias, Statistical bias, Study design issues	Linear regression	Odds ratio
[211]	2022	Sweden, Denmark, France, Netherlands, Germany, Austria	Dementia	$PM_{2.5}$, NO_2, O_3, BC	Residence, Remote sensing, Spatial interpolation	Cohort	Confounding factors, Ecological bias, Exposure assessment, Unassessed patients	Linear regression, Cox regression, Sensitivity analysis	Hazard ratio
[212]	2022	USA	Dementia	$PM_{2.5}$, NO_2	Residence, Spatial interpolation	Cohort	Confounding factors, Ecological bias, Exposure assessment	Cox regression	Hazard ratio
[213]	2022	China	Dementia	PM_{10}, NO_2, SO_2, CO, O_3	Administrative	Cohort	Exposure assessment, Sampling bias	Linear regression	Hazard ratio, Coefficients
[216]	2023	USA	Dementia	$PM_{2.5}$, NO_2	Residence	Case-crossover	Exposure assessment, Unassessed patients	Logistic regression	Odds ratio
[214]	2022	China	Dementia (Alzheimer)	$PM_{2.5}$	GIS, Residence, Remote sensing	Cohort	Exposure assessment, Sampling bias	T-test, Chi-squared, Cox regression, Sensitivity analysis	Hazard ratio

[215]	2022	USA	Dementia (Alzheimer)	$PM_{2.5}$	Latitude, Longitude, Residence	Cohort	Confounding factors, Ecological bias, Exposure assessment, Sampling bias	Cox regression	Hazard ratio
[202]	2020	Canada	Dementia, Alzheimer, MS, Parkinson	$PM_{2.5}$, NO_2, BC, Index	Remote sensing, Spatial interpolation, Residence	Cohort	Confounding factors, Exposure assessment, Time related bias, Unassessed patients	Logistic regression, Cox regression	Hazard ratio, Odds ratio
[207]	2021	USA	Dementia, Parkinson	$PM_{2.5}$	Residence, Remote sensing	Case-control	Confounding factors, Exposure assessment, Sampling bias	Logistic regression	Odds ratio
[210]	2021	USA	Dementia, Parkinson	$PM_{2.5}$	Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Unassessed patients	Logistic regression, Poisson regression, Sensitivity analysis	Coefficients
[188]	2016	Spain	MND	Pb	Administrative, GIS	Ecological	None given by the authors	Correlation, T-test, ANOVA	Correlation, Coefficients
[156]	1971	Israel	MS	Sun exposure, Temperature, Humidity, Precipitation	Residence	Review	None given by the authors	None	None
[163]	1983	USA	MS	Sun exposure, Temperature, Humidity, Precipitation	Latitude	Case-control	None given by the authors	Logistic regression	Relative risk
[157]	1987	Bulgaria	MS	Sun exposure, Temperature, Precipitation	Administrative, Latitude	Cross-sectional	Unassessed patients	Correlation, Chi-squared, Linear regression	Correlation, Coefficients
[172]	1997	Norway	MS	Mg	Administrative	Methodological	Confounding factors	Correlation	Correlation

[158]	2001	Australia	MS	Sun exposure, Temperature, Precipitation	Administrative, Latitude, Remote sensing	Ecological	Confounding factors, Exposure assessment	Correlation, Poisson regression	Prevalence, Correlation
[169]	2005	Italy	MS	SO_2, CO_2	Administrative, Latitude	Cross-sectional	Exposure assessment, Interpolation	Correlation, Linear regression	Correlation, Coefficients
[168]	2008	USA	MS	PM_{10}, $PM_{2.5}$, NO_X, SO_2, CO	Administrative	Cross-sectional	None given by the authors	Correlation, T-test, Linear regression	Correlation, Coefficients
[166]	2010	Norway	MS	Temperature, Precipitation	Administrative	Cross-sectional	Migration of patients, Statistical bias	ANOVA, Poisson regression	Prevalence
[159]	2011	Canada	MS	Sun exposure	Latitude, Longitude, Remote sensing	Cross-sectional	None given by the authors	Correlation, Linear regression	Correlation
[160]	2011	England	MS	Sun exposure	GIS, Remote sensing	Cross-sectional	Confounding factors, Sampling bias, Statistical bias	Correlation, Linear regression	Correlation, Coefficients
[155]	2012	Canada	MS	Sun exposure	GIS, Remote sensing	Methodological	Exposure assessment	None	None
[170]	2014	Iran	MS	PM_{10}, NO_X, SO_2	Clustering, GIS	Cross-sectional	Confounding factors, Statistical bias, Study design issues	Correlation, Clustering	Correlation, Coefficients
[173]	2016	England	MS	Rn	Residence	Ecological	Sampling bias, Statistical bias, Unassessed patients	Correlation, Chi-squared, Linear regression	Correlation, Coefficients
[164]	2016	Italy	MS	Sun exposure	Administrative, GIS	Cross-sectional	Confounding factors, Ecological bias, Time related bias	Correlation, Linear regression	Correlation, Odds ratio
[161]	2017	USA	MS	Sun exposure, Temperature	Administrative, GIS, Remote sensing	Cross-sectional	Confounding factors, Statistical bias	Correlation, Linear regression	Correlation, Coefficients
[171]	2018	Iran	MS	Index (AQ)	Administrative, GIS, Residence	Cross-sectional	Exposure assessment, Statistical bias	Correlation, Logistic regression	Odds ratio, Coefficients

[162]	2018	USA	MS	Sun exposure	Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Interpolation, Memory bias, Migration of patients, Survival bias, Time related bias	Cox regression	Relative risk, Hazard ratio
[167]	2018	Italy	MS	$PM_{2.5}$	Residence, Remote sensing	Cross-sectional	Conflict of interests, Confounding factors, Study design issues	Correlation, Chi-squared, Linear regression	Correlation, Coefficients
[165]	2018	Canada	MS	Sun exposure	Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Time related bias	Linear regression	Coefficients
[217]	2020	Poland	MS	Sun exposure	Latitude	Review	None given by the authors	None	Prevalence
[218]	2021	Italy	MS	$PM_{2.5}$	Administrative, Residence	Ecological	Exposure assessment, Statistical bias	Bayesian	Prevalence, Relative risk
[219]	2021	USA	MS	Temperature, Humidity	Clustering, Residence	Case-control	Exposure assessment, Migration of patients, Sampling bias	Logistic regression, Clustering	Odds ratio
[220]	2021	Israel	MS	PM_{10}, $PM_{2.5}$, NO_2, SO_2, O_3	Residence, Remote sensing	Case-crossover	Exposure assessment, Sampling bias, Unassessed patients	T-test, ANOVA, Logistic regression	Relative risk, Odds ratio
[221]	2021	USA	MS	Temperature	Residence	Cohort	Confounding factors, Sampling bias, Unassessed patients	Linear regression, Sensitivity analysis	Relative risk
[222]	2021	France	MS	$PM_{2.5}$	Residence	Case-crossover	Exposure assessment,	Logistic regression	Odds ratio

							Sampling bias, Time related bias		
[223]	2021	Iran	MS	$PM_{2.5}$, O_3	Administrative	Ecological	Confounding factors, Exposure assessment, Study design issues	Correlation, Linear regression	Prevalence, Coefficients
[224]	2021	Italy	MS	$PM_{2.5}$	Residence, Remote sensing	Cohort	None given by the authors	Correlation	Prevalence, Correlation
[225]	2022	Australia	MS	Sun exposure	Latitude, Longitude, Residence, Remote sensing	Cohort	Confounding factors, Ecological bias	Linear regression	Coefficients
[226]	2023	Sweden	MS	NO_X	Residence	Case-control	Confounding factors, Exposure assessment, Memory bias, Sampling bias, Unassessed patients	Logistic regression	Odds ratio
[174]	2017	USA	Paediatric MS	Index (AQ)	GIS, Residence	Case-control	Exposure assessment, Statistical bias, Time related bias, Unassessed patients	Logistic regression, Poisson regression	Odds ratio, Coefficients
[175]	2018	USA	Paediatric MS	PM_{10}, $PM_{2.5}$, NO_2, SO_2, CO, O_3, Pb	Administrative, GIS, Residence	Case-control	Exposure assessment, Referral bias, Time related bias	T-test, Logistic regression	Odds ratio
[178]	2007	Canada	Parkinson	NO_2, Mn	Residence, Remote sensing, Spatial interpolation	Case-control	Confounding factors, Exposure assessment, Interpolation, Study design issues, Time related bias	Correlation, Linear regression, Logistic regression, Cox regression, Sensitivity analysis	Prevalence, Correlation, Odds ratio
[176]	2010	USA	Parkinson	Mg, Cu, Pb	Administrative	Ecological	Confounding factors, Exposure	Logistic regression,	Relative risk, Odds ratio

							assessment, Statistical bias	Sensitivity analysis	
[179]	2016	USA	Parkinson	PM_{10}, $PM_{2.5}$, NO_2	GIS, Residence	Case-control	Exposure assessment, Memory bias, Statistical bias, Time related bias	Correlation, Logistic regression, Sensitivity analysis	Correlation, Odds ratio
[181]	2016	Taiwan	Parkinson	NO_X, CO	GIS, Residence	Case-control	Confounding factors, Sampling bias, Statistical bias	Correlation, Chi-squared, Logistic regression, Sensitivity analysis	Correlation, Odds ratio
[177]	2016	Spain	Parkinson	Pb	Administrative, GIS	Ecological	Exposure assessment, Sampling bias, Unassessed patients	Correlation, T-test	Correlation, Coefficients
[182]	2017	France	Parkinson	Sun exposure, $PM_{2.5}$	Administrative, Remote sensing	Ecological	Ecological bias, Exposure assessment, Migration of patients	Correlation, Poisson regression, Sensitivity analysis	Correlation, Relative risk
[180]	2020	Australia	Parkinson	$PM_{2.5}$, NO_2	Residence, Remote sensing	Cross-sectional	Memory bias, Referral bias, Sampling bias	Logistic regression, Sensitivity analysis	Odds ratio
[227]	2021	Switzerland	Parkinson	PM_{10}, NO_2	Clustering, GIS, Residence	Case-control	Confounding factors, Exposure assessment	Clustering	Prevalence
[228]	2023	Sweden, Denmark, France, Netherlands, Germany, Austria	Parkinson	$PM_{2.5}$, NO_2, O_3, BC	Residence, Remote sensing	Cohort	Confounding factors, Exposure assessment, Statistical bias, Unassessed patients	Cox regression, Sensitivity analysis	Hazard ratio

Table B2 - Study results summary. +: positive significative association, -: negative significative association; ?: no significative association found.

Paper	Disease	Black carbon	CO	Cu	Humidity	Index	Mg	Mn	NO_X	NO_2	O_3	Pb	$PM_{2.5}$	PM_1	PM_{10}	Precipitation	Pressure	Rn	SO_2	Sun exposure	Temperature
[187]	ALS				+											+	?			-	?
[186]	ALS		?						+		?	?	?		?				?		
[195]	ALS														?						
[196]	ALS												?		?						
[185]	Dementia									+	+		+								
[183]	Dementia																				+
[184]	Dementia		?							+	?				?				?		
[197]	Dementia																				
[198]	Dementia					-				+			+		+						
[199]	Dementia				?	-			-		+				-						?
[200]	Dementia										+		+								
[201]	Dementia									?			+								
[202]	Dementia	?				-				-			?								
[25]	Dementia										?		?								
[203]	Dementia								+					+							
[204]	Dementia	?								?			+								
[205]	Dementia												+								
[206]	Dementia												+								
[207]	Dementia												+								
[208]	Dementia												+								
[209]	Dementia												+		+						
[210]	Dementia												+								
[211]	Dementia	?								?	?		?								
[212]	Dementia									?			?								
[213]	Dementia		+							+	+				+				+		
[214]	Dementia												+								
[215]	Dementia												+								
[216]	Dementia									+			?								

Paper	Disease	Black carbon	CO	Cu	Humidity	Index	Mg	Mn	NO_X	NO_2	O_3	Pb	PM_{25}	PM_1	PM_{10}	Precipitation	Pressure	Rn	SO_2	Sun exposure	Temperature
[188]	MND											+									
[156]	MS				?											?				-	+
[163]	MS				-											-				-	+
[157]	MS															+				-	-
[172]	MS						-														
[158]	MS															?				-	-
[168]	MS		?						?				+		+				?		
[166]	MS															+					+
[159]	MS																			-	
[160]	MS																			-	
[155]	MS																			?	
[170]	MS								+						+				+		
[173]	MS																	?			
[164]	MS																			?	
[161]	MS																			?	-
[171]	MS					+															
[162]	MS																			?	
[167]	MS												+								
[165]	MS																			?	
[169]	MS		+																?		
[217]	MS																			+	
[202]	MS	?				+				?			?								
[220]	MS									?	?		?		?				?		
[221]	MS																				+
[222]	MS												+								
[223]	MS										+		?								
[224]	MS												+								
[225]	MS																			?	
[226]	MS								+												

Paper	Disease	Black carbon	CO	Cu	Humidity	Index	Mg	Mn	NO_X	NO_2	O_3	Pb	PM_{25}	PM_1	PM_{10}	Precipitation	Pressure	Rn	SO_2	Sun exposure	Temperature
[174]	PMS					+															
[175]	PMS		+						?		?	+	+		?				+		
[178]	Parkinson							?		?											
[176]	Parkinson			?			+					?									
[179]	Parkinson									?			?		?						
[181]	Parkinson		+						+												
[182]	Parkinson												?							-	
[180]	Parkinson									?			?								
[177]	Parkinson											+									
[202]	Parkinson	?				?				+			+								
[227]	Parkinson									+					+						
[207]	Parkinson												?								
[210]	Parkinson												+								
[228]	Parkinson	?								?	-		+								

5. Land surface temperature algorithm calibration through meteorological stations

One of the environmental factors identified was temperature, so this paper focusses solemnly on this factor and develops an algorithm to compute it through Landsat 7 imagery, further calibrating it using meteorological stations as ground truth. The resulting surface temperature computed was then used as one of the variables to explore the association with neurodegenerative diseases.

Mariana Oliveira, Ana Cláudia Teodoro, Alberto Freitas, Hernâni Gonçalves

Proc. SPIE 11863, Earth Resources and Environmental Remote Sensing/GIS Applications XII, 1186310

September 2021 http://doi.org/10.1117/12.2599909

Conference -

Abstract: This methodologic paper arises from the necessity to gather Land Surface Temperature (LST) data over a relatively large period and territory: 2000-2018, Portugal. The computational power required to complete this task was found to be a major barrier. However, platforms such as Google Earth Engine (GEE) offer a vast data archive freely accessible through a web interactive development environment or an application programming interface, namely, Python's API. Additionally, the computation using GEE is hosted in Google's servers, drastically reducing the processing times. However, computing LST through Landsat-7 satellite imagery resulted on a difference of -8ºC±6ºC compared to the values from meteorological ground stations. As such, this paper aims to further calibrate computed LST through meteorological stations and make the methodology and corresponding code available, thus encouraging cooperation on the development and integration of local calibration methods. A sensitivity analysis of the representativeness of each station was performed using three methods of temperature extraction: station coordinate's pixel, buffers around the station, and surrounding soil occupation (identifying the area with the same soil occupation as the station's location). Pearson's correlation coefficient was on average significant at 0.81 in the raw data and increased to 0.89 after clearing data from outliers. The best representativeness method for meteorologic stations was the one based on soil occupation, which resulted on a Pearson's r of 0.91. As a result, we advise researchers to complement their remote sensing work with ground data whenever possible through the usage of a method like the one here described.

Keywords: Google Earth Engine, Land Surface Temperature, Meteorological stations; Landsat-7

5.1. Introduction

Environmental data constitutes a key factor for human health studies, and remote sensing is a very exploited approach to retrieve such data [26, 229, 230]. In particular, temperature may be measured from remote sensing data through the computation of Land Surface Temperature (LST), which is the temperature felt at the ground level [231].

This is a theme of such scientific interest that several studies have been published on the computation of LST from various sources and using different algorithms. For instance, several researchers have compared different algorithms to attempt to reach the most accurate one [232-235], either by changing the inputs for the computation or the methodology used. After 2010, with the appearance of Google Earth Engine (GEE), different authors have shifted their studies to accompany the platform potentialities [236]. This shift allows the visualization and manipulation of a huge amount of remote sensing data, available with ease from GEE data catalogue, at minimum computational and time-consuming cost [237]. Also, Google's tutorials, inbuilt functions and communities' open-source codes allow for easy and comprehensive usage of GEE's tools.

In the case of LST retrieval, several datasets are available, for example, through the daily MOD11A1, a LST and emissivity product with a resolution of 1 Km, and the monthly OpenLandMap also at 1 Km spatial resolution which provides monthly median daytime LST. However, these datasets have relatively low spatio-temporal resolution. This has not been unnoticed, and a framework has been published for producing LST from Landsat data through GEE [238]. Also, an open-source code has been published for LST estimation using the Landsat series and GEE [239]. This study differs from ours in terms of complexity, as it uses top of the atmosphere (TOA) as well as surface reflectance (SR) to compute LST, a step skipped in our study, as the SR product from GEE is already calibrated. Also, our study focuses on the usage of Landsat 7 data, as it allows monitoring from 1999 up to the present without changing the sensor, thus maintaining any error sources that might exist.

The objective of this study was to develop a top-to-bottom and open-source methodology to gather and treat remote sensing data, validating the computed LST through ground meteorological stations. The methodology hereby presented is applicable to any time period and geographic area the user may intend.

To this purpose, this paper is structured in a comprehensive way. The materials section gives an overview of the inputs used, the pre-processing section elucidates the management of meteorological data, and the methodology section explains, step by step, the code developed. The results section explores the linear regression analysis of the retrieved LST and the meteorological stations' values in the calibration process.

5.2. Materials

5.2.1. *Meteorological stations*

Meteorological stations distributed throughout Portugal measure hourly temperature. These data are available online through SNIRH (National System of Information on Hydric Resources) [56] and IPMA (Portuguese Institute for Sea and Atmosphere) [52] for free download. A total of 117 stations were retrieved from SNIRH with mean monthly temperature and 11 stations from IPMA with minimum and maximum monthly temperature. As stations are independent from each other, their data was simply merged into a single excel file in which each line corresponded to the station and the columns contained the values measured. However, 38 of the total stations had no measured values from 2000 to 2018, and as so were disregarded. Each station had spatial coordinates associated, in the military datum from Portugal (EPSG 20790) from which they were mapped (Figure 9). All geographic information was converted to WGS84 UTM Zone 29N (EPSG 32629) for matching coordinated systems.

5.2.2. *Remote sensing data and auxiliary data*

The atmospherically corrected surface reflectance from Landsat 7 ETM+ sensor was used, taking advantage of the availability at GEE data catalogue, thus sparing further pre-processing [59].

To analyse the representativeness of meteorological stations, a Land Use and Occupation Charted (COS) from 2018 was needed. This is open-source data and was downloaded from SNIG (National Infrastructure for Geographical Information) [57]. This data is originally available in the Portuguese coordinate system ETRS89 (EPSG 3763).

5.2.3. *Study area and analysis period*

This study concerns Portugal mainland and gathers data from 2000 to 2018. This time period is related to the time span of a health dataset the researchers have available, which shall be used in the future to relate with LST. The Portuguese islands were not included as they have autonomous health datasets, which were not available to the researchers.

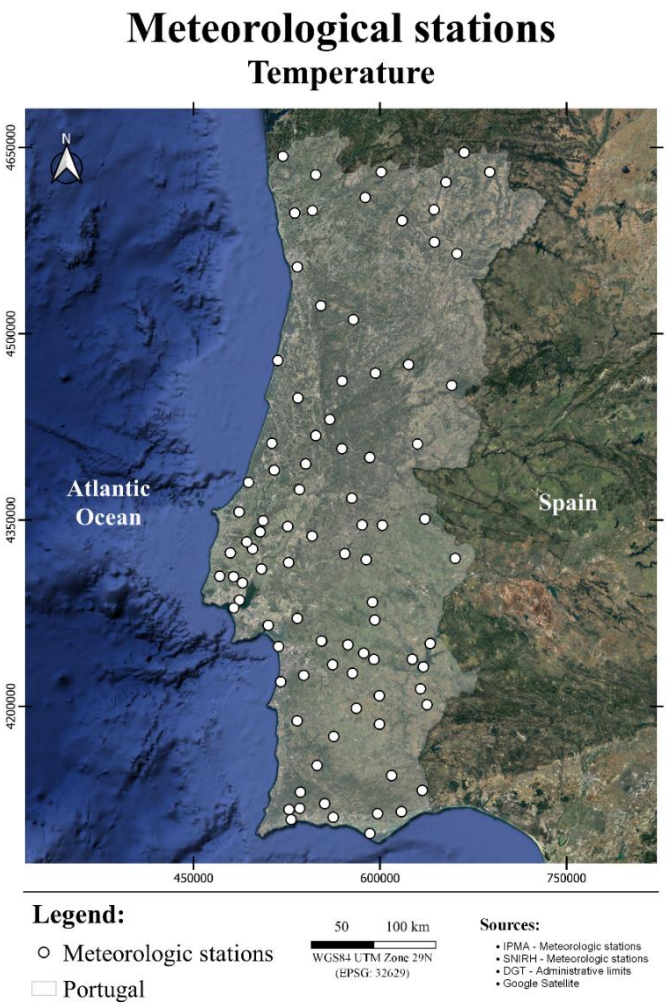


Figure 9 - Spatial distribution of meteorological stations in mainland Portugal.

5.3. Pre-processing

5.3.1. *Meteorological stations*

As meteorological stations are considered ground truth data for temperature assessment in this study, a representativeness analysis was performed to explore which may be the best fitting method to extract temperature from the satellite images.

According to the world meteorological organization (WMO), meteorological stations may be representative of the element they are measuring in a range of 100 meters to 3 Km when used in applications such as air pollution [58]. Also, an approach based on the ground characteristics surrounding the station is interesting when specifically measuring temperature [240]. For such an approach, a COS was analysed in a GIS (Geographic Information System) along with the points locating

the meteorological stations. Using the WMO guidelines referred before, a buffer of 3 Km was set around each station, through which the COS was clipped. The resulting polygon is solely the continuous extent of COS corresponding to that where the station is located, up to the limit of 3 Km. Through this approach, all stations have a different area of influence, depending on the land conditions surrounding it. As so, the temperature value measured is considered valid throughout the continuous soil use up to 3 Km away from the station.

In the representativeness analysis, temperature from the satellite images were retrieved from several methods:

i. pixels corresponding to the location of each station;
ii. pixels in a 1 Km range to the location of each station;
iii. pixels in a 2 Km range to the location of each station;
iv. pixels in a 3 Km range to the location of each station;
v. pixels that correspond to each polygon generated by the method discussed above.

5.4. Methodology

5.4.1. *Google Earth Engine*

To proceed to the steps described in the methodology, users must have an account on GEE. Also, Python's Application Programming Interface (API) needs to be installed. Developers of GEE advise users to install the API in a clean environment, as referred in the installation tutorial available on GEE website. Finally, the users need to authenticate through the API, which can be done following the same tutorial [241].

Initial remote sensing data exploration was conducted through GEE code editor. Initial code for retrieving and filtering the images, as well as computing the LST was written and tested in GEE, since the Python's API can execute that code, unaltered. Additionally, this proceeding allowed us to visualize results and ensure the algorithm was correctly implemented. As such, and since GEE is not efficient using cycles, a few ImageCollections, a native format from GEE containing a group of images, were created containing monthly Landsat images, for trial purposes. As an example, in the code in Figure 10, are the months of January 2000 and December 2018. Later, this step will not be needed, as Python is efficient using cycles.

```
//_____________________________________Monthly dataset_____________________________________
//Import Portugal borders
var PT = ee.FeatureCollection("USDOS/LSIB_SIMPLE/2017").filter(ee.Filter.eq("country_na", "Portugal"));

//Import Landsat 7 SR
var Jan2000 = ee.ImageCollection("LANDSAT/LE07/C01/T1_SR").filterDate("2000-01-01","2000-01-
31").filterBounds(PT);
var Dec2018 = ee.ImageCollection("LANDSAT/LE07/C01/T1_SR").filterDate("2018-12-01","2018-12-
31").filterBounds(PT);

//Join ImageCollection in a list
var Landsat = [Jan2000, Dec2018];

//Export
exports.LandsatMonthly = Landsat
```

Figure 10 - GEE code (part 1): example of importing Landsat 7 SR monthly data from the January 2000 and December 2018.

Temperature was computed using Landsat 7 SR product, as well as with ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) derived emissivity image, both available from GEE data catalogue. Importing these datasets was achieved through Figure 11 code.

```
//_____________________________________Import datasets_____________________________________
//Import Portugal boarders
var PT = ee.FeatureCollection("USDOS/LSIB_SIMPLE/2017").filter(ee.Filter.eq("country_na", "Portugal"));

//Import Landsat 7 SR
var Landsat = require("users/mariana_ramos_oliveira/GEE:Thesis/0_LST_Monthly").LandsatMonthly;

//Import ASTER emissivity
var Emiss = ee.Image("NASA/ASTER_GED/AG100_003").clip(PT).select("emissivity_band10").multiply(0.001);
//print(Landsat);
```

Figure 11 - GEE code (part 2): Example of importing study area boarders and emissivity dataset.

Intrinsic to most remote sensing products, some pre-processing steps were made to allow further proceedings. For pre-processing purposes, bands were scaled, and clouds were removed using the code presented in Figure 12. Scale values, as well as the cloud masking function, were retrieved from GEE website [59].

```
//_____________________________________________FUNCTIONS_____________________________________________
//Function to scale bands
var Landsat_scale = function(imageC){
  var imageColl = ee.ImageCollection(imageC);
  var Landsat_scale2 = imageColl.map(function(image){
    var B1 = image.select("B1").multiply(0.0001); //Blue band
    var B2 = image.select("B2").multiply(0.0001); //Green band
    var B3 = image.select("B3").multiply(0.0001); //Red band
    var B4 = image.select("B4").multiply(0.0001); //NIR band
    var B5 = image.select("B5").multiply(0.0001); //SWIR1 band
    var B6 = image.select("B6").multiply(0.1); // Brightness temperature band (Kelvin)
    var B7 = image.select("B7").multiply(0.0001); //SWIR2 band
    var QA = image.select("pixel_qa");
    return ee.Image([B1,B2,B3,B4,B5,B6,B7,QA]).set("SENSING_TIME",ee.Date(image.get("SENSING_TIME")));
  });
  return Landsat_scale2;
};

//Function to mask cloudy pixels
var Landsat_mask = function (imageC) {
  var imageColl = ee.ImageCollection(imageC);
  var Landsat_mask2 = imageColl.map(function(image){
    //Bits 3 and 5 are cloud shadow and cloud, respectively.
    var cloudShadowBitMask = (1 << 3);
    var cloudsBitMask = (1 << 5);
    //Get the pixel QA band.
    var qa = image.select('pixel_qa');
    //Both flags should be set to zero, indicating clear conditions.
 var mask = qa.bitwiseAnd(cloudShadowBitMask).eq(0)
                 .and(qa.bitwiseAnd(cloudsBitMask).eq(0));
    return image.updateMask(mask).set("SENSING_TIME",image.get("SENSING_TIME"));
  });
  return Landsat_mask2;
};};
```

Figure 12 - GEE code (part 3): Definition of functions to scale bands and mask cloud pixels.

The computation of LST, in Celsius degrees, follows Equation 5, and is implemented through the function on Figure 13,

$$LST = \frac{SR}{1 + \left(\lambda * \left({}^{SR}/_{\rho}\right)\right) * \ln(\varepsilon)} - 273.15 \qquad \text{Equation 5}$$

where SR is the Landsat 7 SR band 6 (Kelvin), λ is the mean wavelength of emitted radiance for the same band (meters), ρ is a constant given by Equation 6 (meters Kelvin), and ε is the emissivity (dimensionless) [60]. The constant ρ is computed as

$$\rho = h\frac{c}{\sigma} = 1.438 * 10^{-2} \qquad \text{Equation 6}$$

where h is Planck's constant ($6.626*10^{34}$ *Js*), c is the velocity of light ($2.998*10^{8}$ *m/s*) and σ corresponds to the Boltzmann's constant ($1.38*10^{-23}$ *J/K*) [60].

```
//Function to compute LST
var Landsat_LST = function(imageC){
  var imageColl = ee.ImageCollection(imageC);
  var Landsat_LST2 = imageColl.map(function(image){
    return image.expression(
    '(B6/(1 + (0.00115* (B6 / 1.4388))*log(EMISS)))-273.15', {
     'B6': image.select('B6'),
     'EMISS': Emiss.select('emissivity_band10')
    }).rename('LST').clip(PT).set("SENSING_TIME",image.get("SENSING_TIME"));
  });
  return Landsat_LST2;
};
```

Figure 13 - GEE code (part 4): Definition of the function to compute LST.

The functions need to be applied to the images to take effect, which can be done through the code on Figure 14.

```
//_____________________________________________Applying functions_____________________________________________
var Landsat_Scaled = Landsat.map(Landsat_scale);
var Landsat_Masked = Landsat_Scaled.map(Landsat_mask);
var Landsat_Temp = Landsat_Masked.map(Landsat_LST);
```

Figure 14 - GEE code (part 5): Example of applying the functions to the Landsat ImageCollection.

Results may be visualized (Figure 16) through the coding interface of GEE using the visualization parameters described in Figure 15. Missing strips of values are related to cloud covered images, whose pixels were removed through the masking function on Figure 12, to allow for the correct computation of LST.

```
//_____________________________________________Visualizing results_____________________________________________
var LST_viz = {
  min: 0,
  max: 30,
  opacity: 0.7,
  palette: ["04aeff","68ff29","fff81b","ff5210"]
};
Map.addLayer(Landsat_Temp[1], LST_viz);
```

Figure 15 - GEE code (part 6): Example of visualization parameters for LST product.

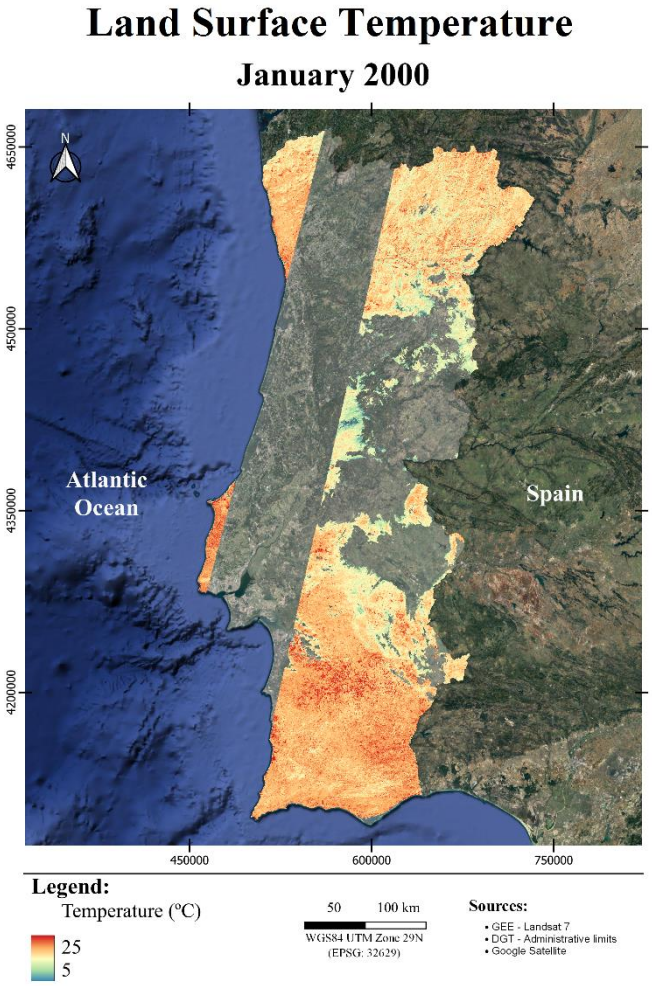


Figure 16 - Computed LST for Portugal mainland during the month of January 2000

5.4.2. *Python API*

After developing and testing the code, Python API was used to complete the code and allow for further analysis. For start, necessary libraries and initial data were imported, according to the code in Figure 17.

```
###_____________________________________________ Imports _____________________________________________
## Import libraries
from IPython.display import display, Markdown, Latex
import ee
import pandas
import datetime
import time
import numpy
from itables import show
import itables.options as opt

## Initializing GEE account
ee.Initialize()
# print("Done imports")

###_____________________________________________ Initials _____________________________________________
## Start counting time elapsed
start = time.time()

## Download datasets
# Portugal limits
PT = ee.FeatureCollection("USDOS/LSIB_SIMPLE/2017").filter(ee.Filter.eq("country_na", "Portugal"))
# Landsat images
Landsat = ee.ImageCollection("LANDSAT/LE07/C01/T1_SR").filterDate("2000-01-01", "2018-12-
31").filterBounds(PT)
# Emissivity image
Emiss = ee.Image("NASA/ASTER_GED/AG100_003").clip(PT).select("emissivity_band10").multiply(0.001)

## Get shapefiles for zonal statistics
LST_Ground_R = ee.FeatureCollection("users/MarianaRamosOliveira_GEE/GP_R_C")
LST_Ground_0 = ee.FeatureCollection("users/MarianaRamosOliveira_GEE/LST_GP_0")
LST_Ground_1000 = ee.FeatureCollection("users/MarianaRamosOliveira_GEE/LST_GP_1000m")
LST_Ground_2000 = ee.FeatureCollection("users/MarianaRamosOliveira_GEE/LST_GP_2000m")
LST_Ground_3000 = ee.FeatureCollection("users/MarianaRamosOliveira_GEE/LST_GP_3000m")

## Get stations file
stations_file = pandas.read_csv("C:/Users/pc/Desktop/GP_R/stations.csv")
# print("Done importing files")

## Filter over ImageCollection
years = range(2000, 2019)
months = range(1, 13)

## Initial list to append to
GP = pandas.DataFrame()
```

Figure 17 - Python's code (part 1): Importing initial libraries, downloading initial datasets (boarders, Landsat imagery, emissivity dataset and meteorological stations) and setting initial datasets.

Next, all necessary functions were defined (Figure 18). Adding to the scaling, masking and LST computation functions, a function to convert the resulting FeatureCollection into a Data Frame was added, to allow for posterior manipulation using Python language [242].

```
###_____________________________________________ Functions _____________________________________________
## Function to scale bands
def scale(image):
    B4 = image.select("B4").multiply(0.0001)
    B5 = image.select("B5").multiply(0.0001)
    B6 = image.select("B6").multiply(0.1)
    QA = image.select("pixel_qa")
    return ee.Image([B4, B5, B6, QA])

## Function to mask clouds
def mask(image):
    cloud_shadow = (1 << 3)
    cloud = (1 << 5)
    qa = image.select("pixel_qa")
    mask_1 = qa.bitwiseAnd(cloud_shadow).eq(0)
    mask_2 = qa.bitwiseAnd(cloud).eq(0)
    mask_final = mask_1 and mask_2
    return image.updateMask(mask_final)

## Function to compute LST
def compute_variables(image):
    Temp = image.expression("(B6/(1 + (0.00115* (B6 / 1.4388))*log(EMISS)))-273.15",
                            {"B6": image.select("B6"), "EMISS":
Emiss.select("emissivity_band10")}).rename("LST").clip(
        PT)

    return ee.Image(Temp)

## Function to convert FeatureCollection into DataFrame
def convert_FC(fc):
    features = fc.getInfo()["features"]
    dictarr = []
    for f in features:
        attr = f["properties"]
        dictarr.append(attr)
    df = pandas.DataFrame(dictarr)
    return df

# print("Done setting up functions")
```

Figure 18 - Python's code (part 2): Defining functions to scale bands, mask clouds, compute LST and latter convert the resulting FeatureCollection into a DataFrame.

In the implementation process (Figure 19), a cycle was introduced to filter images per month, allowing for the sequential and automatic mapping of each month and each year of study. Inside the cycle, the original Landsat ImageCollection was filtered by date, scaled, masked, and the LST was computed, obtaining several images per month corresponding to LST. Given those images, grouped by month, a simple mean was computed, as well as a Kernel mean with a square of 3 pixels. Both these resulting images were converted to separate bands of a single image.

```
## ____________________________________________ Implementation ____________________________________________
for i in years: # Make cycle for each year...
    for j in months: # ... and each month
        # print(j, i)
        start_date = datetime.datetime(i, j, 1) # Set start date as the first day of the month
        if j in [1, 3, 5, 7, 8, 10, 12]: # For months with 31 days...
            end_date = datetime.datetime(i, j, 31) # Set end date to the 31st
            if j in [1, 3, 5, 7, 8]: # If months are under 10
                task_name = str(i) + "0" + str(j) # Just add a 0 before the name, to have same length
            else: # If not
                task_name = str(i) + str(j) # Just name the task MonthYear
            # print(start_date, " to ", end_date)
        elif j == 2: # If month is February...
            if i in [2000, 2004, 2008, 2012, 2016]: # Then leap years ...
                end_date = datetime.datetime(i, j, 29) # Set end date to the 29th
                task_name = str(i) + "0" + str(j) # Just add a 0 before the name, to have same length
                # print(start_date, " to ", end_date)
            else: # Then non-leap years ...
                end_date = datetime.datetime(i, j, 28) # Set end date to the 28th
                task_name = str(i) + "0" + str(j) # Just add a 0 before the name, to have same length
                # print(start_date, " to ", end_date)
        elif j in [4, 6, 9, 11]: # For months with 30 days...
            end_date = datetime.datetime(i, j, 30) # Set end date to the 30th
            if j in [4, 6, 9]: # If months are under 10
                task_name = str(i) + "0" + str(j) # Just add a 0 before the name, to have same length
            else: # If not
                task_name = str(i) + str(j) # Just name the task MonthYear
            # print(start_date, " to ", end_date)
        else: # If any of the above don't work there must be an error. Just skip the month.
            print("Error in end date to filter from")
            continue
## ____________________________________________ Pre-processing ____________________________________________
        ## Filter ImageCollection by date
        Landsat_filtered = Landsat.filterDate(start_date, end_date)
        ## Apply scale to bands per image
        Landsat_scaled = Landsat_filtered.map(scale)
        ## Mask cloud and cloud shadow pixels per image
        Landsat_masked = Landsat_scaled.map(mask)
        ## Compute LST per image
        Landsat_vars = Landsat_masked.map(compute_variables)
        ## Computation of LST mean and Kernel mean per month
        Landsat_LST_M = Landsat_vars.select("LST").reduce(ee.Reducer.mean()).rename("LST_Mean")
        Landsat_LST_K_3 = Landsat_LST_M.reduceNeighborhood(ee.Reducer.mean(),
ee.Kernel.square(3)).rename("LST_Kernel_3")
        ## Join LST mean and Kernel mean in one image
        Landsat_LST = ee.Image([Landsat_LST_M, Landsat_LST_K_3])
```

Figure 19 - Python's code (part 3): Definition of a cycle to iterate on each month of the time period, application of the functions to scale bands, mask clouds and compute LST, and computation of monthly mean and kernel mean, creating a single two banded image with the resulting means.

Through this last step, the user computes LST from Landsat 7 imagery. However, monitoring stations were used to further calibrate the values computed. This process is hereafter explained. In the next step (Figure 20), *shapefiles* containing the monitoring stations representativeness areas as described in chapter 3.1 are used to compute the mean temperature in those same areas. Results were saved on a *DataFrame* to allow for posterior manipulation. The process takes around two and a half hours to be complete, using an Intel Core I7-7700 processor and a stable cable internet connection with 100/100 Mbps speed.

```
## ____________________________________________ Zonal statistics ____________________________________________
        ## Get ground points values
        Landsat_GP_R = Landsat_LST.reduceRegions(LST_Ground_R, ee.Reducer.mean(), 30).sort("CÓDIGO")
        Landsat_GP_0 = Landsat_LST.reduceRegions(LST_Ground_0, ee.Reducer.mean(), 30).sort("CÓDIGO")
        Landsat_GP_1 = Landsat_LST.reduceRegions(LST_Ground_1000, ee.Reducer.mean(), 30).sort("CÓDIGO")
        Landsat_GP_2 = Landsat_LST.reduceRegions(LST_Ground_2000, ee.Reducer.mean(), 30).sort("CÓDIGO")
        Landsat_GP_3 = Landsat_LST.reduceRegions(LST_Ground_3000, ee.Reducer.mean(), 30).sort("CÓDIGO")
        # print(task_name)

        ## Set initials for saving results
        Codes = pandas.DataFrame(stations_file["CODE"])
        Values = pandas.DataFrame(stations_file[task_name])
        #print(Values)

        ## Save results
        try:
            # Errors may have happened during the computation and GP extraction of LST.
            # This may be due to clouded areas that have NaN values.
            # As so, this cycle first tries to convert the results from FeatureCollection to DataFrame.
            Landsat_DF_R = convert_FC(Landsat_GP_R)
            Landsat_DF_0 = convert_FC(Landsat_GP_0)
            Landsat_DF_1 = convert_FC(Landsat_GP_1)
            Landsat_DF_2 = convert_FC(Landsat_GP_2)
            Landsat_DF_3 = convert_FC(Landsat_GP_3)
        except:
            # If there is an error:
            Empty_DF = pandas.DataFrame([i]*90).rename(columns={0: "Year"}) #Create Year column
            Empty_DF["Month"] = pandas.DataFrame([j]*90).rename(columns={0: "Month"}) #Create Month column
            Empty_DF["Code"] = Codes #Create Codes column
            Empty_DF["Station_Values"] = Values #Create meteorological stations values column
            Empty_DF["GP_R_LST_M"] = [numpy.nan]*90 # Fill remaining GP columns with NaN...
            Empty_DF["GP_R_LST_K"] = [numpy.nan]*90
            Empty_DF["GP_0_LST_M"] = [numpy.nan]*90
            Empty_DF["GP_0_LST_K"] = [numpy.nan]*90
            Empty_DF["GP_1_LST_M"] = [numpy.nan]*90
            Empty_DF["GP_1_LST_K"] = [numpy.nan]*90
            Empty_DF["GP_2_LST_M"] = [numpy.nan]*90
            Empty_DF["GP_2_LST_K"] = [numpy.nan]*90
            Empty_DF["GP_3_LST_M"] = [numpy.nan]*90
            Empty_DF["GP_3_LST_K"] = [numpy.nan]*90
            GP = pandas.concat([GP, Empty_DF]) # Concatenate initial dataset with the one just created
            # print("Empty")
        else:
            # If there is no error:
            Full_DF = pandas.DataFrame([i]*90).rename(columns={0: "Year"}) # Create Year column
            Full_DF["Month"] = pandas.DataFrame([j]*90).rename(columns={0: "Month"}) # Create Month column
            Full_DF["Code"] = Codes # Create Codes column
            Full_DF["Station_Values"] = Values # Create meteorological stations values column
            Full_DF["GP_R_LST_M"] = Landsat_DF_R["LST_Mean"] #Fill remaining GP columns with the values
            Full_DF["GP_R_LST_K"] = Landsat_DF_R["LST_Kernel_3"]
            Full_DF["GP_0_LST_M"] = Landsat_DF_0["LST_Mean"]
            Full_DF["GP_0_LST_K"] = Landsat_DF_0["LST_Kernel_3"]
            Full_DF["GP_1_LST_M"] = Landsat_DF_1["LST_Mean"]
            Full_DF["GP_1_LST_K"] = Landsat_DF_1["LST_Kernel_3"]
            Full_DF["GP_2_LST_M"] = Landsat_DF_2["LST_Mean"]
            Full_DF["GP_2_LST_K"] = Landsat_DF_2["LST_Kernel_3"]
            Full_DF["GP_3_LST_M"] = Landsat_DF_3["LST_Mean"]
            Full_DF["GP_3_LST_K"] = Landsat_DF_3["LST_Kernel_3"]
            GP = pandas.concat([GP, Full_DF]) # Concatenate initial dataset with the one just created
            # print("OK")

end1 = time.time() # Compute elapsed time
```

Figure 20 - Python's code (part 4): Code example of computing zonal statistics from monthly images and meteorological stations.

5.4.3. *Data analysis*

Having extracted temperature from the computed LST product through the different representativeness methods, the purpose of this study was to analyse which would be the best method and perform a linear regression to fit the LST product. For this analysis, other libraries were imported, such as *sklearn* and *matplotlib*.

Data was analysed as a continuous dataset, but also grouped monthly and yearly, to detect eventual temporal patterns. For this, three different functions were created. In the functions that take the previously created *DataFrame* as input, the name of each column is collected, to create a new DataFrame consisting only of that column, the date and the values measured at the station, dropping any empty lines, since they are not relevant for fitting the model. After, data is prepared to be searched for outliers and clusters through a DBSCAN model, and two separate *DataFrames* containing clusters and outliers were created. The cluster *DataFrame* was then used to compute the Pearson's correlation coefficient r alongside with the *DataFrame* without the identified outliers. As expected, and as presented in the results, the Pearson's coefficient is higher in the dataset without the outliers, and as so this was the dataset used to fit the linear model. This dataset was then split into a training and testing set prior to fitting the model, in a proportion of 20% test size. The results of the linear regression and the computed Pearson coefficients were saved in a *DataFrame*, which was output by the function, alongside a new *DataFrame* with all values computed from the original *DataFrame* and the linear regression coefficients.

```
## Function to compute statistics from all data
def simple_STATS(df):
    C_name = df.columns[4] # Get name of the column
    df = df[["Year", "Month", "Station_Values", C_name]].dropna() #Select relevant columns and drop NaN
    dbscan_data = df.values.astype("float32", copy=False) #Convert values to float
    dbscan_data_scaler = StandardScaler().fit(df) #Fit a scaler to values to later fit outlier/cluster
model
    dbscan_data = dbscan_data_scaler.transform(dbscan_data) # Scale values
    model = DBSCAN(eps=0.25, min_samples=12, metric="euclidean").fit(dbscan_data)# Search for
clusters/outliers
    outliers_df = df[model.labels_ == -1] # Get outliers dataframe
    clusters_df = df[model.labels_ != -1] # Get clusters dataframe
    pearson = scipy.stats.pearsonr(df["Station_Values"], df[C_name]) #Compute pearson's r with all values
    pearson_2 = scipy.stats.pearsonr(clusters_df["Station_Values"], clusters_df[C_name]) #Compute
pearson's r without outliers
    data_train, data_test = train_test_split(clusters_df, test_size=0.2, random_state=4) #Split dataset
into train and testing set
    Stations_Train = data_train["Station_Values"].values.reshape(-1, 1) #Reshape to fit model
    Stations_Test = data_test["Station_Values"].values.reshape(-1, 1) #Reshape to fit model
    LST_Train = data_train[C_name].values.reshape(-1, 1) #Reshape to fit model
    LST_Test = data_test[C_name].values.reshape(-1, 1)  Reshape to fit model
    lm = sklearn.linear_model.LinearRegression() #Initialize linear model
    lm_model = lm.fit(LST_Train, Stations_Train) #Fit linear model
    R = pearson[0].round(3) #Get pearson's r value
    R2 = pearson_2[0].round(3) #Get pearson's r value
    P = pearson[1].round(3) #Get pearson's r p-value
    P2 = pearson_2[1].round(3) #Get pearson's r p-value
    I = lm_model.intercept_[0].round(3) #Get linear model's intercept
    C = lm_model.coef_[0,0].round(3) #Get linear model's coefficient
    ## Join results in a dataframe
    Simple_results = pandas.DataFrame([[R, P, R2, P2, I, C]]).rename(columns={0: "Pearson R",
                                                                              1: "P-value",
                                                                              2: "Pearson R (Cluster)",
                                                                              3: "P-value (Cluster)",
                                                                              4: "LM Intercept",
                                                                              5: "LM
Coefficient"}).rename(index={0: C_name})
    V = (clusters_df[C_name] - C) / I #Compute new values applying linear regression
    DF_new = clusters_df.join(V, rsuffix="_new") #Add new values to dataframe
    return [Simple_results, DF_new] #Return both statistics results and dataframe with new values
```

Figure 21 - Statistical analysis code (part 1): Example of a function to compute descriptive statistics on the output values of remote sensing processed data.

The code presented in Figure 21 was then adjusted to represent months and years variation, by using a cycle to apply the code to each month or year. These functions were then applied to the dataset created in chapter 5.4.2. through the code in Figure 22, which then saves the results in three different *DataFrames*.

```
## ____________________________________________ Implementation ____________________________________________
dataset = pandas.read_csv("C:/Users/pc/Desktop/GP_R/Test.csv") # Get dataset
dataset = dataset.drop(columns=["Unnamed: 0"]) # Drop auto-generated index column
# display(dataset)

STATS_S = pandas.DataFrame() # Create empty dataframe to append to
STATS_M = pandas.DataFrame() # Create empty dataframe to append to
STATS_Y = pandas.DataFrame() # Create empty dataframe to append to
DF_S = pandas.DataFrame() # Create empty dataframe to append to
DF_M = pandas.DataFrame() # Create empty dataframe to append to
DF_Y = pandas.DataFrame() # Create empty dataframe to append to

for c in dataset.columns[4:]: # For each column...
    data = dataset[["Year", "Month","Code", "Station_Values", c]] # Gather relevant columns
    STATS_1 = simple_STATS(data)[0] # Apply function to compute stats and get first returned dataset
    STATS_2 = monthly_STATS(data)[0] # Apply function to compute stats and get first returned dataset
    STATS_3 = yearly_STATS(data)[0] # Apply function to compute stats and get first returned dataset
    D_S = simple_STATS(data)[1] # Apply function to compute stats and get second returned dataset
    D_M = monthly_STATS(data)[1] # Apply function to compute stats and get second returned dataset
    D_Y = yearly_STATS(data)[1] # Apply function to compute stats and get second returned dataset
    STATS_S = pandas.concat([STATS_S, STATS_1]) # Join results
    STATS_M = pandas.concat([STATS_M, STATS_2]) # Join results
    STATS_Y = pandas.concat([STATS_Y, STATS_3]) # Join results
    if c == dataset.columns[4]: # If it's the first column...
        DF_S = D_S # Start dataframe with own results
    else: # If it's not...
        DF_S = DF_S.join(D_S, rsuffix="_X") # Join results to previous dataframe
    DF_M = pandas.concat([DF_M, D_M]) # Join results
    DF_Y = pandas.concat([DF_Y, D_Y]) # Join results

DF_S = DF_S.drop(columns=["Year_X", "Month_X", "Station_Values_X"]) # Drop repeated columns
# print("DONE")
```

Figure 22 - Statistical analysis code (part 2): Implementing the previous descriptive statistics functions to the results dataset, iterating over each column.

5.5. Results

As illustrated in Figure 23, the relation between LST from remote sensing data and the values from the meteorological stations exhibit a linear trend, and thus the Pearson's correlation coefficient is adequate to measure the association between these variables (Figure 23).

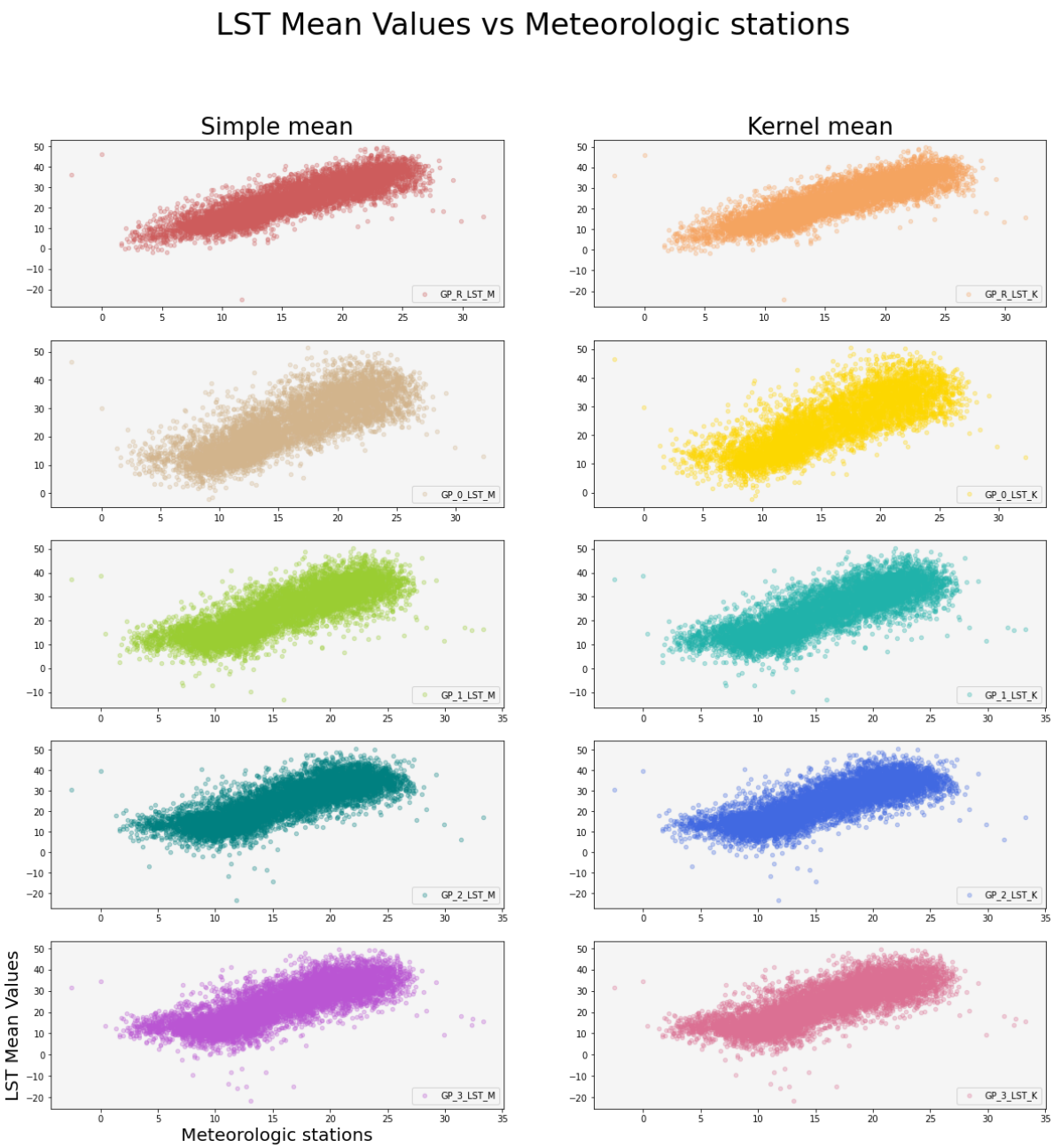


Figure 23 - Scatter plot of satellite-computed LST and the corresponding meteorological station value measured. Different colours represent the different approaches in retrieving the LST value, as described in chapter 5.3.1. for both simple ("M") and Kernel ("K") mean.

The highest correlation was observed for representativeness of meteorological stations from COS analysis with a correlation coefficient of 0.87, followed by the 1 Km buffer with an r of 0.81. However, results obtained using the simple mean and the Kernel mean when computing LST were similar with an r of 0.80 (Table 9). Additionally, there were no differences between the mean and kernel approaches.

Table 9 - Pearson's correlation coefficient computed between each monthly image, both mean and kernel mean bands, and the buffered areas of the meteorological stations, using original dataset.

Buffer type	Simple mean			Kernel mean		
	Simple	Monthly	Yearly	Simple	Monthly	Yearly
No buffer	0.80	0.80	0.80	0.80	0.80	0.80
1 Km buffer	0.81	0.81	0.81	0.81	0.81	0.81
2 Km buffer	0.80	0.80	0.80	0.80	0.80	0.80
3 Km buffer	0.80	0.80	0.80	0.80	0.80	0.80
COS	0.87	0.87	0.87	0.87	0.87	0.87

However, after removing outliers, as expected, the correlation coefficient increased, and COS analysis became more noticeably the best representativeness method (Table 10), with an r of 0.91. The mean correlation coefficient of the raw data was then 0.81, increasing to 0.89 on the clean dataset.

Table 10 - Pearson's correlation coefficient computed between each monthly image, both mean and kernel mean bands, and the buffered areas of the meteorological stations, using outlier cleaned dataset.

	Simple mean			Kernel mean		
	Simple	Monthly	Yearly	Simple	Monthly	Yearly
i. No buffer	0.89	0.21	0.79	0.89	0.22	0.80
ii. 1 Km buffer	0.89	0.19	0.87	0.89	0.19	0.87
iii. 2 Km buffer	0.88	0.06	0.85	0.88	0.06	0.84
iv. 3 Km buffer	0.86	0.08	0.85	0.88	0.07	0.85
v. COS	0.91	0.44	0.89	0.91	0.44	0.90

The linear regression model obtained for COS representativeness method versus both mean and kernel mean was given by the Equations 7 and 8, respectively.

$$LST_{Corrected,\,Mean} = 3{,}762 + 0{,}502 * LST_{Original}$$

Equation 7

$$LST_{Corrected,Kernel} = 3{,}848 + 0{,}500 * LST_{Original}$$

Equation 8

After fitting and applying the linear regression to the original dataset, the difference between the value measured by the station and the remote sensing LST value was computed.

The mean difference considering all LST retrieval methods was originally -7.96ºC with a standard deviation of 6.10 ºC. After removing outliers and applying the corresponding linear regression model, this difference increased to 9.62 ± 3.09 ºC. Furthermore, if considered only the COS representativeness method for retrieving the LST, this difference drops slightly to 9.37 ± 3.00 ºC. Although there was a slight increase on the average absolute difference, the dispersion decreased considerably, reflecting the elimination of the greater discrepancies present in the original values (Figure 24).

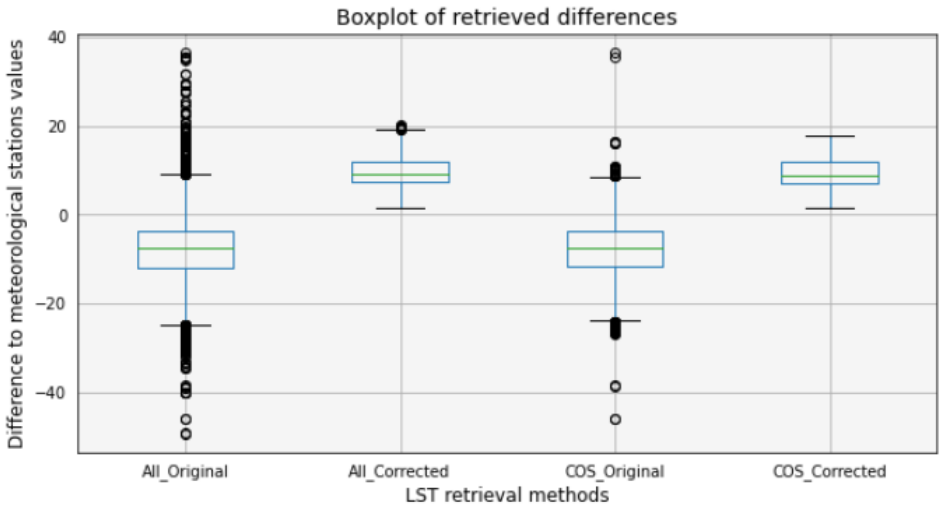


Figure 24 - Boxplot of differences found between the computed LST value and that measured by the meteorological station.

5.6. Conclusion

The computation of environmental variables from remote sensing data is of great importance to scientific studies, since remote sensing allows for easy, open access data throughout the globe, for large temporal periods. Bearing this in mind, the computation of such variables is subject to several errors, including atmospheric calibration errors. Although most of these errors are known, further calibration through ground monitoring stations is advised to improve the quality of the product. That was the purpose of this study, to develop a top-to-bottom methodology to gather and treat remote sensing data to further compute LST, which is then compared to ground meteorological stations and a linear model derived correction is applied. This methodology reduced the gap between the ground measured temperature and the LST product, and thus the purpose of this study has been fulfilled. However, the linear models obtained are fit to this particular data, and as such should not be extrapolated. To extrapolate, the user should fit the code to its own data. The same applies for extrapolating results to other temporal periods. For instance, including data from more years might change, even if slightly, the model fit. The complete code is available freely to any user by request to the authors, which contains detailed comments to help guide future users.

6. Neurodegenerative diseases hospitalizations in Portugal: a spatio-temporal analysis

Based on the neurodegenerative diseases identified in the systematic reviews, this chapter presents a spatio-temporal analysis of hospitalizations related with neurodegenerative diseases. In particular, global and local spatial autocorrelations were computed, revealing disease clusters. It also approaches cases with more than a neurodegenerative disease and differentiates neurodegenerative primary diagnosis hospitalizations from those simply having a neurodegenerative diseases as secondary diagnosis.

Mariana Oliveira, Alberto Freitas, Ana Cláudia Teodoro, Hernâni Gonçalves

International Journal of Health Geographics (submitted for publication)

Submitted

-

4.9

Q1 (Public Health, Environmental and Occupational Health)

Abstract: Background: Neurodegenerative diseases are a rising concern in the world's population increasing life expectancy. More precisely, the Portuguese population, along with that of other developed countries, is ageing at a fast rate. Understanding such diseases' temporal and spatial patterns is of utmost relevance to help manage the burden they represent in the health system. Methods: In this retrospective study, we analysed over 500 thousand hospitalizations with discharges between 2000 and 2016. In this retrospective study, we analysed over 500 thousand hospitalizations with discharges between 2000 and 2016. We used the empirical Bayes method to compute smoothed age-standardized hospitalization rates for each neurodegenerative disease and all hospitalizations per year and administrative division (districts and municipalities). We then searched for data clusters using global and local spatial autocorrelation methods based on the Moran index. Results: An increase in age-standardized hospitalization rates was found throughout the years. Statistically significant global spatial autocorrelation was found throughout the years when considering all diseases. Local analysis evidence high-high clusters in the centre and low-low clusters in the south districts, although these patterns tend to become less evident throughout time. Conclusions: The spatio-temporal analysis showed an overall increase of age-standardized hospitalization rates due to neurodegenerative diseases throughout time, with some evidence of spatial autocorrelation and clustering. The ageing of the Portuguese population, exposition to environmental variables and discrepancies in access to primary care might explain some of the observed spatial patterns.

Keywords: Neurodegenerative diseases, Spatial Epidemiology, Hospitalization Rates

6.1. Introduction

The increase in global life expectancy from 66.8 to 73.3 years between 2000 and 2019 [243] is raising the concern on neurodegenerative diseases, since age is one of the most contributing factors to developing these diseases, which occur due to pathological changes in the brain [61], resulting in the death of cells due to either necrosis or delayed apoptosis [68].

Alzheimer's disease (AD) among others are the most common form of late onset Dementia globally and contribute to 12% of the burden in disability-adjusted life years [244]. Together with Parkinson's disease (PD), these are the most common neurodegenerative diseases worldwide [13, 245]. Several complications are expected from these patients, such as longer hospital lengths of stay, higher mortality rates and added likelihood to be discharged to long-term care [246]. There are other neurodegenerative diseases that include, but are not limited to, Multiple Sclerosis (MS), Amyotrophic Lateral Sclerosis (ALS) and Huntington's.

Some risk factors to these diseases have already been discussed and include age, gender, hereditarianism and environmental exposition to pollution. For example, two thirds of Alzheimer's patients are female,

contrarily to Parkinson's which is more prevalent in men [13], while Huntington is entirely hereditary and some forms of early onset Alzheimer's are also genetically transmittable [247].

Furthermore, the importance of spatially analysing neurodegenerative diseases and its relevance to uncover new correlations with variables such as air pollution and latitude is stated in several studies [24]. Planning and mitigating for such disorders highly depend on the spatial representation and understanding the characteristics and living environment of the patients.

To the authors best knowledge, this is the most complete study developed in Portugal which spatially describes the hospitalizations by neurodegenerative diseases. With this retrospective study based on hospitalization data reuse from 2000 to 2016, we hope to provide further insight on the spatial distribution and general statistical characteristics of this group of diseases.

6.2. Materials and methods

6.2.1. *Study area*

This study was conducted with data from Portugal mainland, a country in the southwest of Europe, with approximately 89.000 km^2 and population of around 10,000,000 inhabitants with a very heterogeneous distribution per age groups (Figure 25), 21% of which being 65 years old or older [248].

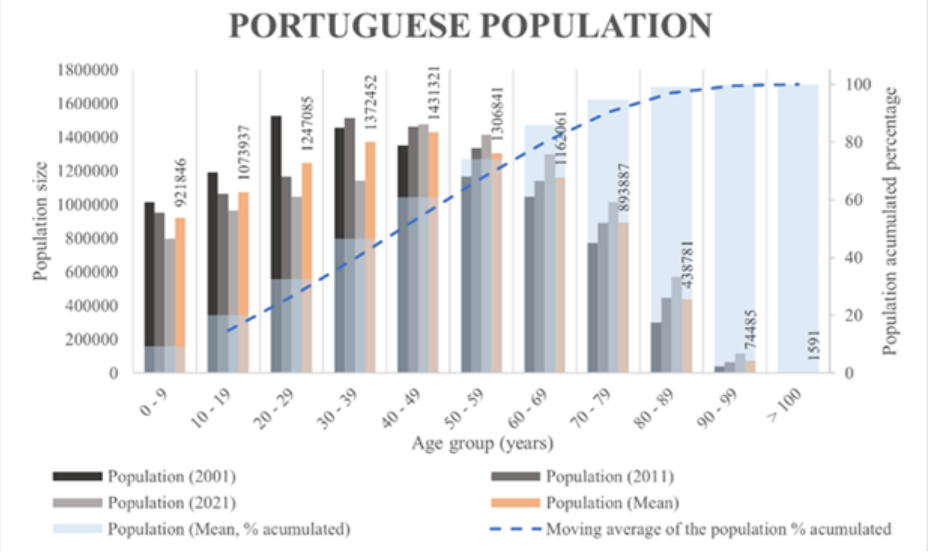


Figure 25 - Population distribution per age group in Portugal – mean value from 2001, 2011 and 2021 [258].

The most populated cities are Porto and Lisbon, as visible in the map of Figure 26, and the interior of the country is clearly very low in population.

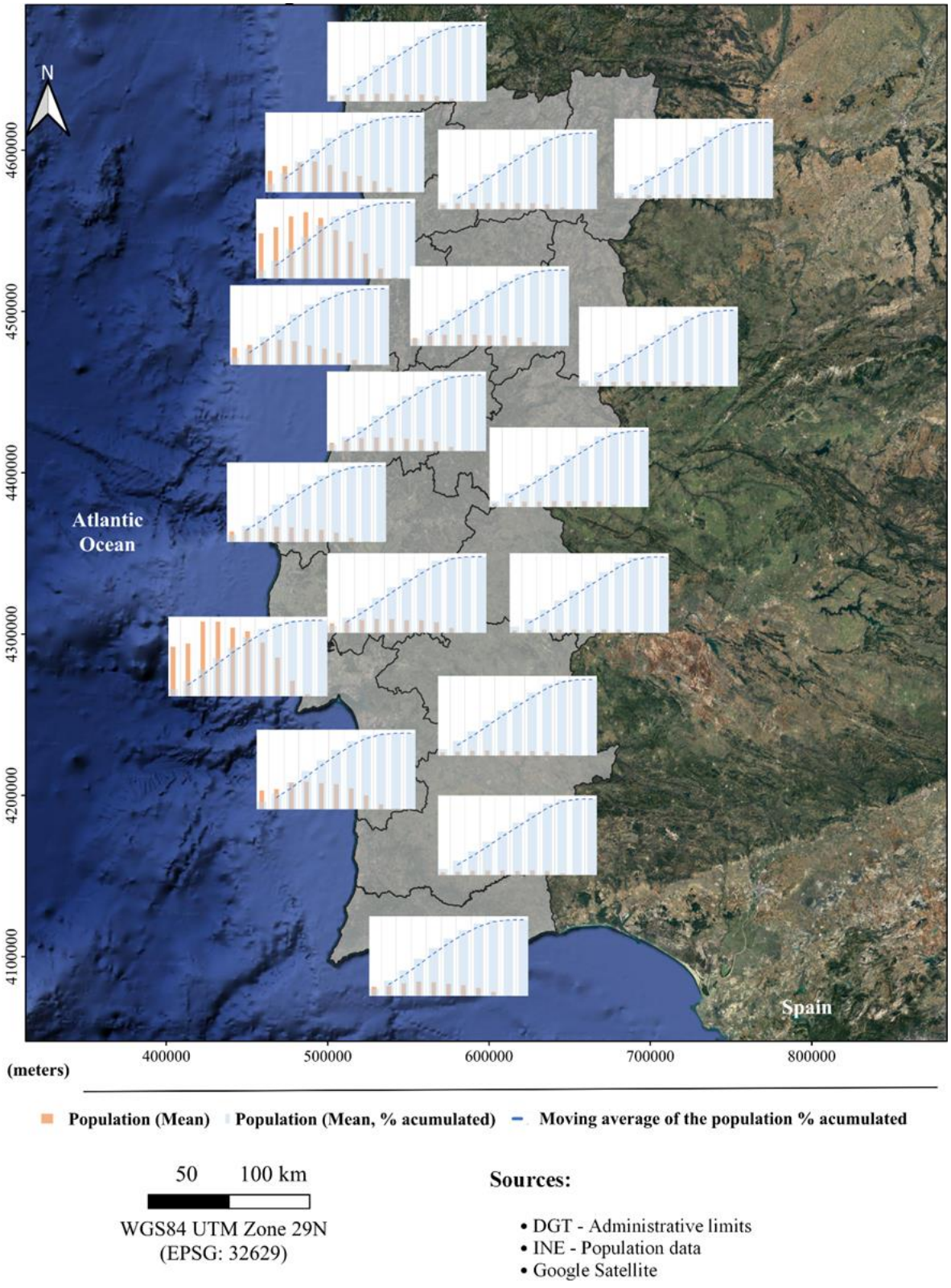


Figure 26 - Portuguese population distribution per district and age group. Age groups as in Figure 1, population mean axis = [0 - 350 000], and population % accumulated axis = [0 - 100].

6.2.2. *Data acquisition*

6.2.2.1. Health data

Hospitalizations of patients with at least one neurodegenerative disease coded as a diagnosis, both principal or secondary, were obtained from hospitalization data from public hospitals, with discharges between 2000 to 2016 in mainland Portugal. No further years were included since the provided data frame included only hospitalizations from 2000 to 2018, but both 2017 and 2018 were incomplete and would thus result in non-comparable results. Each record contains the year, patient's sex, age and residence (coded up to the parish), entry and leave dates, length of stay and coded diagnosis (ICD-9-CM or ICD-10-CM codes).

Table 11 - ICD codes used to retrieve hospitalization data.

Diagnosis		ICD 9	ICD 10
Alzheimer		331.0	G30
Pick's		331.1	G31
Lewy		331.8	
Creutzfeldt		046	A81.0
Huntington		333.4	G10
Parkinson		332	G20
Motor neuron disease		335	G12.2
Multiple sclerosis		340	G35
Basal ganglia		333.0	G23
Dementia	Vascular	290.4	F01 F02 F03
	Unspecified	294	
	Senile	331.2	
	Cerebral degeneration	331.7	
	Presenile	290.1	
	Senile with delusion	290.2	
	Senile with delirium	290.3	
	Alcohol induced	291.2	
	Drug induced	292.82	
Other		331.9	G32

International Classification of Diseases (ICD) codes for neurodegenerative diseases were retrieved from literature review [42-47], along with counselling with coding experts and are listed in Table 11. The category "Other" represents neurodegenerative diseases not described elsewhere by the alternative codes and is a category included in the ICD.

6.2.2.2. Population data

Population data for each Portuguese mainland parish was retrieved from the National Institute of Statistics (INE) publicly available Population Census for the years 2001, 2011 and 2021. To obtain an approximate value of population by year, data was then linearly interpolated, and further used to normalize health data. All values were rounded to the closest integer, due to the nature of the population variable. District and municipality population was computed by summing the contained parishes.

6.2.3. *Data processing*

Python programming language was used in all stages of data pre-processing, descriptive analysis and mapping.

Diagnosis codes were extracted from the health dataset and recoded to the disease they referred to (for example, ICD-9 code 331.0 was recoded as 'Alzheimer', as well as ICD-10 code G30). Age was grouped into three categories: under 20 years old, 20 to 70 years old, and over 70 years old, since population data is only available at 10 years periods, as well as to allow the distinction between paediatric [249], early onset and late onset neurodegenerative diseases [48]. Also, diseases were analysed as a single bulk of neurodegenerative diseases and as each individual disease, and inpatients

admitted with primary and secondary diagnosis were distinguished during the analysis. From here on, when not clarified otherwise, all diagnosis cases were used.

The data were obtained by parishes, but were also grouped by district and municipality. The population and the number of patients, in each administrative area and per age group were used to compute the age-standardized hospitalization rates (ASHR) using Equation 9 [50]. ASHR is a form of normalizing data, therefore enabling the comparison of rates across space and time with a lower degree of bias.

$$ASHR = \frac{N_{AgeGroup}}{Population_{AgeGroup}} \times \%_{AgeGroup} \qquad \text{Equation 9}$$

In Equation 9, for each administrative area, $N_{AgeGroup}$ is the number of hospitalizations per age group, $Population_{AgeGroup}$ is the total population for each age group, and $\%_{AgeGroup}$ is the percentage the age group represents in the overall population of that administrative area.

Given the very low population in some municipalities, an empirical Bayesian approach was used to smooth the ASHR [19] and reduce the statistical instability caused by the Problem of Small Numbers [20]. This approach consisted in creating a distance-based weights matrix for each administrative area which attributed a weight of 0 if further than the distance band and a non-linear weight from 0 to 1 the closest from the area. The distance band used was 150km for corresponding to half the width of the country and thus allowing for at least one district around the one being analysed to be selected as neighbour.

Afterwards, global and local spatial autocorrelation were computed to identify the presence of clustering using Moran's I [250] and LISA (Local Indicators of Spatial Association) using *Python*'s package *esda*, which has available all the formulas as simple functions [251]. Both these methods work similarly to measure how likely it is to obtain a map like the one observed due to a random pattern and if there is no evidence of randomness, then it is assumed there is a statistically significant spatial autocorrelation [252]. Moran's I evaluates global autocorrelation and ranges between -1 and 1 with positive values indicating positive spatial autocorrelation, whereas negative values imply a negative spatial autocorrelation. Values close to zero indicate that the patterns observed are random (Figure 27).[36].

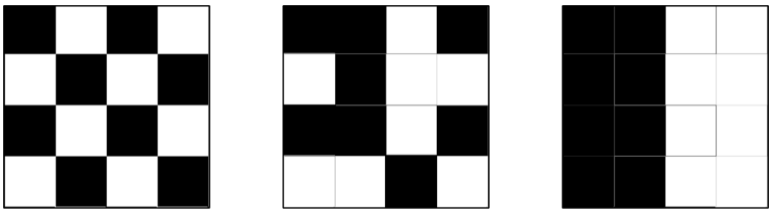


Figure 27 - Moran's I example of autocorrelation coefficients.

The LISA analysis, on the other hand, analyses local autocorrelation and produces a similar result to that of Moran's I. By working locally, it is possible to categorize a map into clusters depending on the indicator result: similar values together result in positive spatial autocorrelation, representing either high-high or low-low clusters, while different values together result in negative spatial autocorrelation, representing low-high or high-low clusters. Using the standardized spatial-lag and the standardized variable, it is then possible to further distinguish the clusters (Figure 28): areas with spatial-lag above the mean, but variable values below the mean represent low-high clusters (light blue), while spatial-lags below the mean and variable values above the mean represent high-low clusters (orange). As for positive spatial correlation, higher spatial-lags with higher variable values are high-high clusters (red) and lower than the mean spatial-lags with lower variable values represent low-low clusters (dark blue).

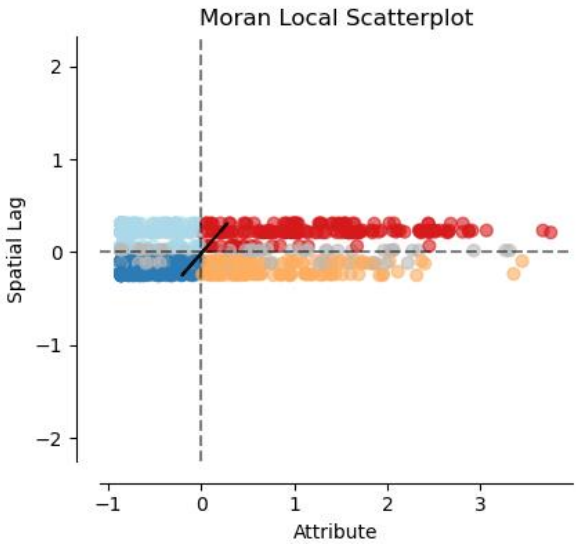


Figure 28 - Scatterplot of values vs their spatial lag. Different quadrants represent different types of clusters (High-high: red; High-low: orange, Low-high: light blue, Low-low: dark blue).

6.3. Results

6.3.1. *Descriptive statistics*

In this study, 502245 hospitalizations were included, 51220 (10.2%) of which presented two or more neurodegenerative diseases, and 34250 (6.8%) had a neurodegenerative disease as primary diagnosis. As represented in the graphic in Figure 29, the most prevalent disease was Dementia, with 43.4% of the cases, followed by Parkinson (22.7%) and Alzheimer, with 20.2% of cases. The least prevalent diseases were Basal, with 0.2% of the cases, Huntington (0.3%), Pick (0.3%) and Huntington (0.3%). Motor Neuron disease's hospitalizations were mostly primarily coded with the disease (53.5%) and Basal and Picks also reached high rates with the respective primary diagnosis (43.1% and 40.7%, respectively).

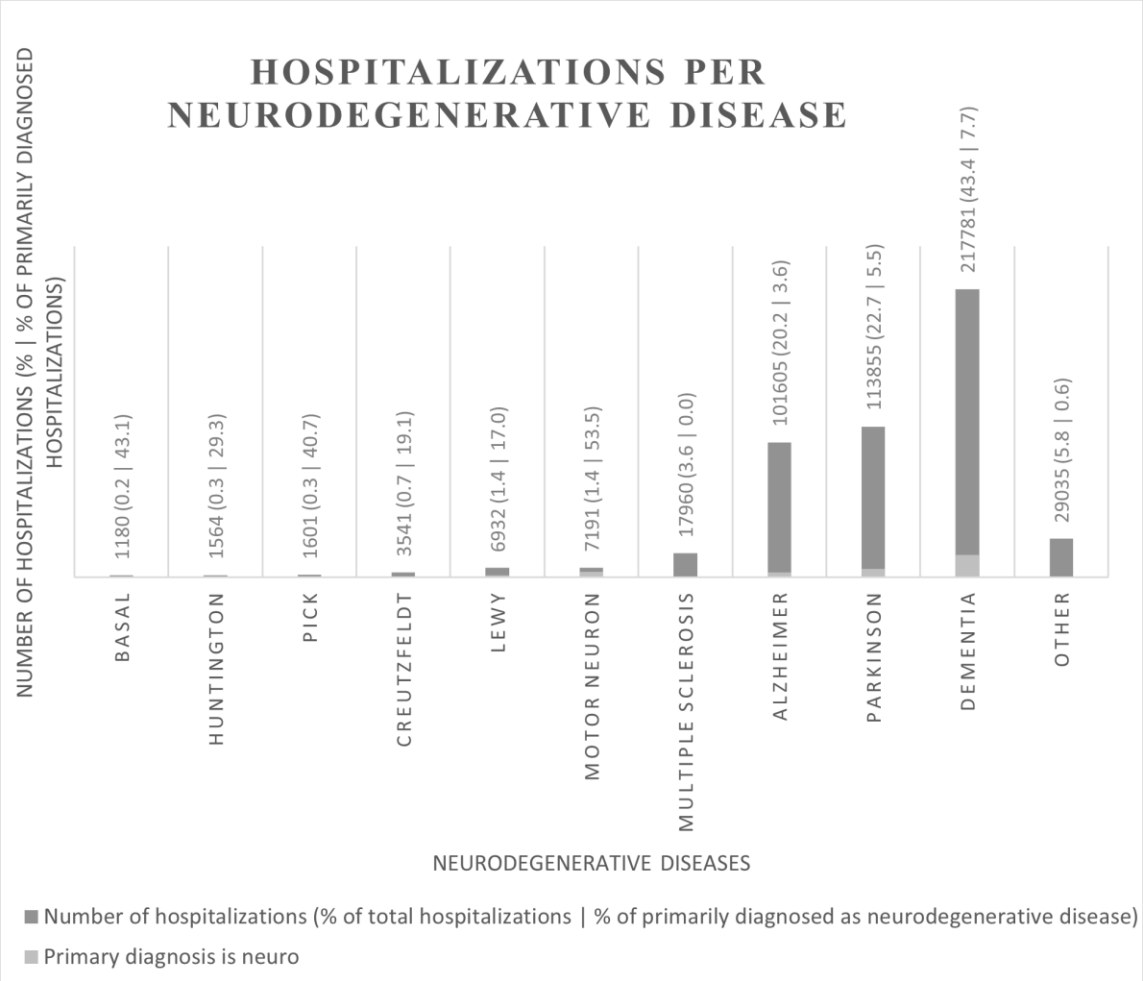


Figure 29 - Number of hospitalizations and respective percentage per neurodegenerative disease, followed by the percentage of those hospitalizations in which the neurodegenerative disease was the primary diagnosis. The total number of hospitalizations is 502245.

Most patients were female (56.3%) or over 70 years old (81.3%) and only 0.4% of the cases were of patients under 20 years old, with Motor Neuron Disease (MND) being by far the most prevalent paediatric neurodegenerative disorder (Table 12). Among the 51220 patients with multiple neurodegenerative diseases (10.2% of the total number of hospitalizations), most had 2 neurodegenerative diagnoses upon hospitalization (48692, 95.1%), with only 45 having four neurodegenerative diagnosis (0.1%). The most common combination of neurodegenerative diseases was Alzheimer and Dementia (37.3%), with a median age of 82 years old (IQI: 77, 86) followed by Parkinson and Dementia (28.2%), with a median age of 81 years old (IQI: 76, 86) (Table C1 - supplementary materials). Also, patients with neurodegenerative diseases as primary diagnosis upon hospitalization are summarized in Table C2 of supplementary materials. Dementia keeps being the most prevalent as a primary diagnosis, with 49.2% of the cases, while Multiple Sclerosis has no cases coded as a primary hospitalization diagnosis.

Table 12 - Descriptive statistics of each neurodegenerative disease included in this study, as well as their total. Categorical data (sex, age group and residence) is described by its absolute frequency and percentage, while continuous data (age and length of stay, which are both asymmetrically distributed), is described by its median and interquartile range. The total row describes the total frequency of each disease and their respective percentage. yo: years old.

		Alzheimer	Basal	Creutzfeldt	Dementia	Huntington	Lewy	Motor Neuron	Multiple Sclerosis	Parkinson	Pick	Other	Total
Sex	**Female**	62349 (61.4)	659 (55.8)	1209 (34.1)	127694 (58.6)	736 (47.1)	3458 (49.9)	3272 (45.5)	12062 (67.2)	56977 (50.0)	722 (45.1)	13775 (47.4)	282913 (56.3)
	Male	39256 (38.6)	521 (44.2)	2332 (65.9)	90087 (41.4)	828 (52.9)	3474 (50.1)	3919 (54.5)	5898 (32.8)	56877 (50.0)	879 (54.9)	15260 (52.6)	219331 (43.7)
Age (yo)		81 [76, 85]	72 [61, 79]	48 [38, 68]	82 [76, 87]	58 [47, 70]	78 [70, 84]	65 [53, 73]	43 [33, 54]	79 [73, 84]	68 [60, 77]	76 [66, 83]	80 [74, 86]
Age Group	**< 20 yo**	16 (0.0)	39 (3.3)	116 (3.3)	157 (0.1)	29 (1.9)	130 (1.9)	748 (10.4)	403 (2.2)	59 (0.1)	2 (0.1)	187 (0.6)	1886 (0.4)
	20-70 yo	9018 (8.9)	511 (43.3)	2649 (74.8)	25241 (11.6)	1169 (74.7)	1656 (23.9)	4087 (56.8)	16775 (93.4)	20052 (17.6)	956 (59.7)	9740 (33.5)	91854 (18.3)
	> 70 yo	92571 (91.1)	630 (53.4)	776 (21.9)	192383 (88.3)	366 (23.4)	5146 (74.2)	2356 (32.8)	782 (4.4)	93744 (82.3)	643 (40.2)	19108 (35.8)	408505 (81.3)
Length of stay (days)		8 [4, 13]	8 [4, 15]	15 [7, 34]	8 [5, 14]	9 [4, 16]	9 [5, 16]	7 [3, 13]	5 [2, 9]	8 [4, 14]	12 [6, 24]	9 [5, 17]	8 [4, 14]
Residence area	**North**	37116 (36.5)	513 (43.5)	973 (27.5)	79156 (36.3)	437 (27.9)	2687 (38.8)	2396 (33.3)	5411 (30.1)	35190 (30.9)	554 (34.6)	22274 (76.7)	186707 (37.2)
	Centre	49750 (49.0)	518 (43.9)	1897 (53.6)	114337 (52.5)	882 (56.4)	3642 (52.5)	3621 (50.4)	9487 (52.8)	61947 (54.4)	901 (56.3)	5225 (18.0)	252207 (50.2)
	South	14739 (14.5)	149 (12.6)	671 (18.9)	24288 (11.2)	245 (15.7)	603 (8.7)	1174 (16.3)	3062 (17.0)	16718 (14.7)	146 (9.1)	1536 (5.3)	63331 (12.6)
Total		101605 (20.2)	1180 (0.2)	3541 (0.7)	217781 (43.4)	1564 (0.3)	6932 (1.4)	7191 (1.4)	17960 (3.6)	113855 (22.7)	1601 (0.3)	29035 (5.8)	502245 (100)

6.3.2. *Spatio-temporal analysis*

After adjusting the age-standardized principal diagnosis rates using the empirical Bayes method, these rates showed to be overall lower, with districts of Beja (0.119) and Viana do Castelo (0.102) with the lowest rates, opposed to Braga (0.207) and Bragança (0.203), which had the highest rates (Figure 30). The global average ASHR was 0.168.



Figure 30 - Age standardized hospitalization rates considering all diseases, per districts.

6.3.2.1. Temporal analysis

Mapping each year's age-standardized principal diagnosis (Figure 31), showed how these rates have been steadily increasing in the past years for most districts. Only Viana do Castelo and Beja districts did not seem to have as much increase in hospitalization rates as the other districts.

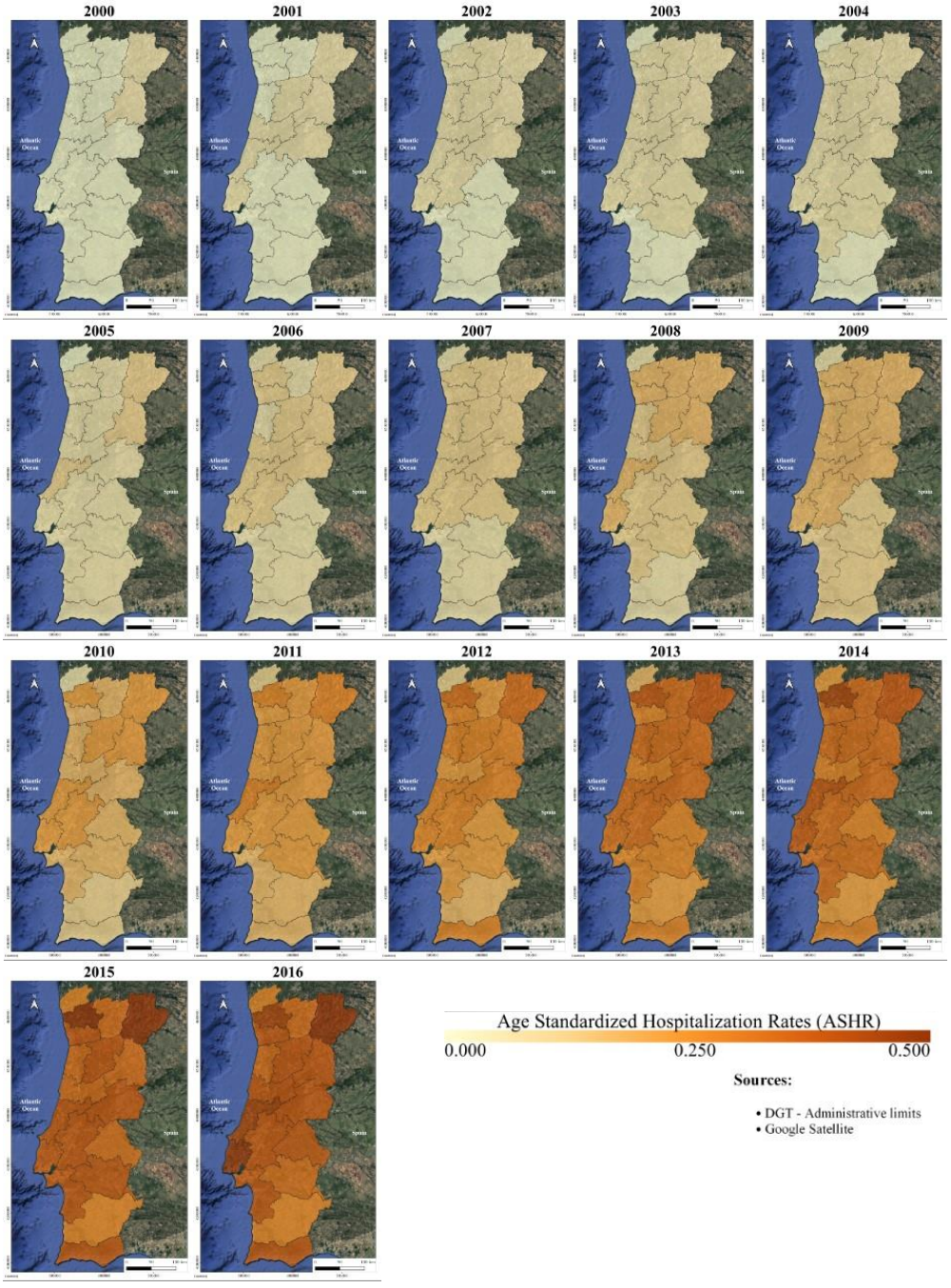


Figure 31 - Age standardized principal diagnosis rates considering all diseases, per districts and per year.

Furthermore, plotting each year's ASHR for each neurodegenerative disease brought further insight to the specific trends of each disease (Figure 32). For instance, Dementia, Parkinson and Alzheimer increased significantly when compared to the remaining diseases.

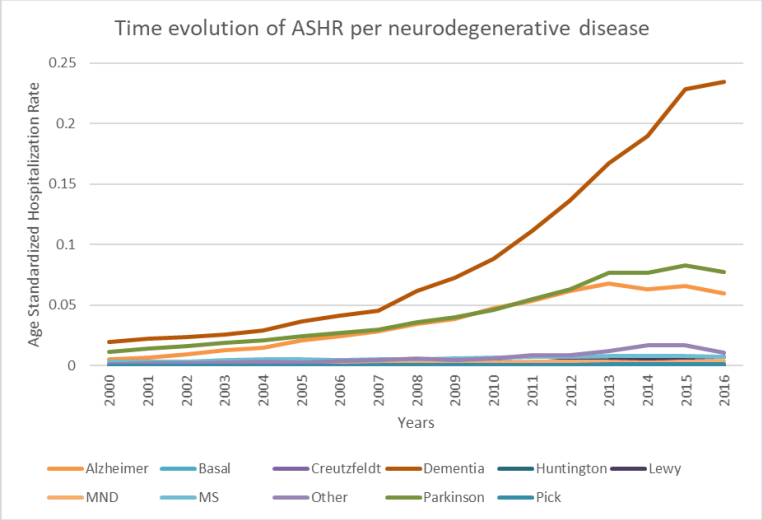


Figure 32 - Timeline of ASHR per neurodegenerative disease.

6.3.2.2. Spatial autocorrelation

6.3.2.2.1. Global analysis

Spatial autocorrelation was studied using both global and local methods. Yearly Moran's I were computed to assess global spatial trends. Figure 33 represents the obtained values for each year. For all neurodegenerative diseases, most values are statistically significant exhibiting a positive spatial autocorrelation around 0.30 until the year of 2008, decaying thereafter to values around 0.15 for municipalities division and to values around 0 (without statistical significance) from the district's perspective.

Only the latest years of 2012 onward did not show significant, and a slight peak is visible in 2004, with a decrease from then on. When analysing each disease individually, the patterns observed differed. For instance, Alzheimer and Dementia were the only showing similar patterns to those in the first graphic. Multiple sclerosis had no cases coded as primary diagnosis, but appeared to have, along with Lewy, an increasing trend toward the end of the timeline. Many values are missing since there are not enough cases spatially scattered to compute the index.

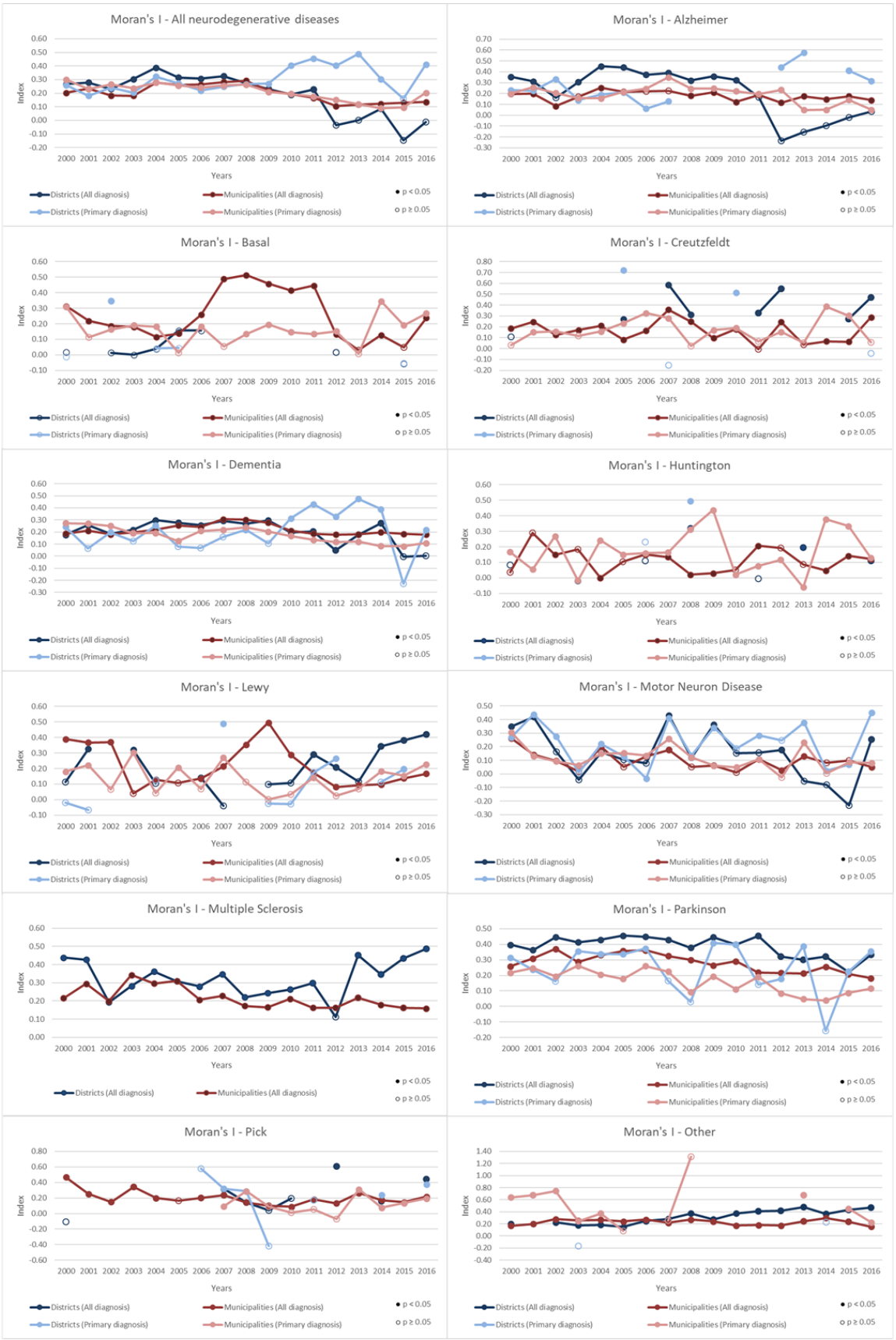


Figure 33 - Yearly Moran's I for both districts and municipalities for each administrative division type.

6.3.2.2.2. Local analysis (clusters): districts

Local spatial autocorrelation was also mapped per year for all diseases (Figure 34). The clustering patterns were very similar up to 2010, when they started varying more and even disappear completely in 2012, 2013 and 2015. The centre of Portugal had mostly high-high clusters, whilst the south had low-low.

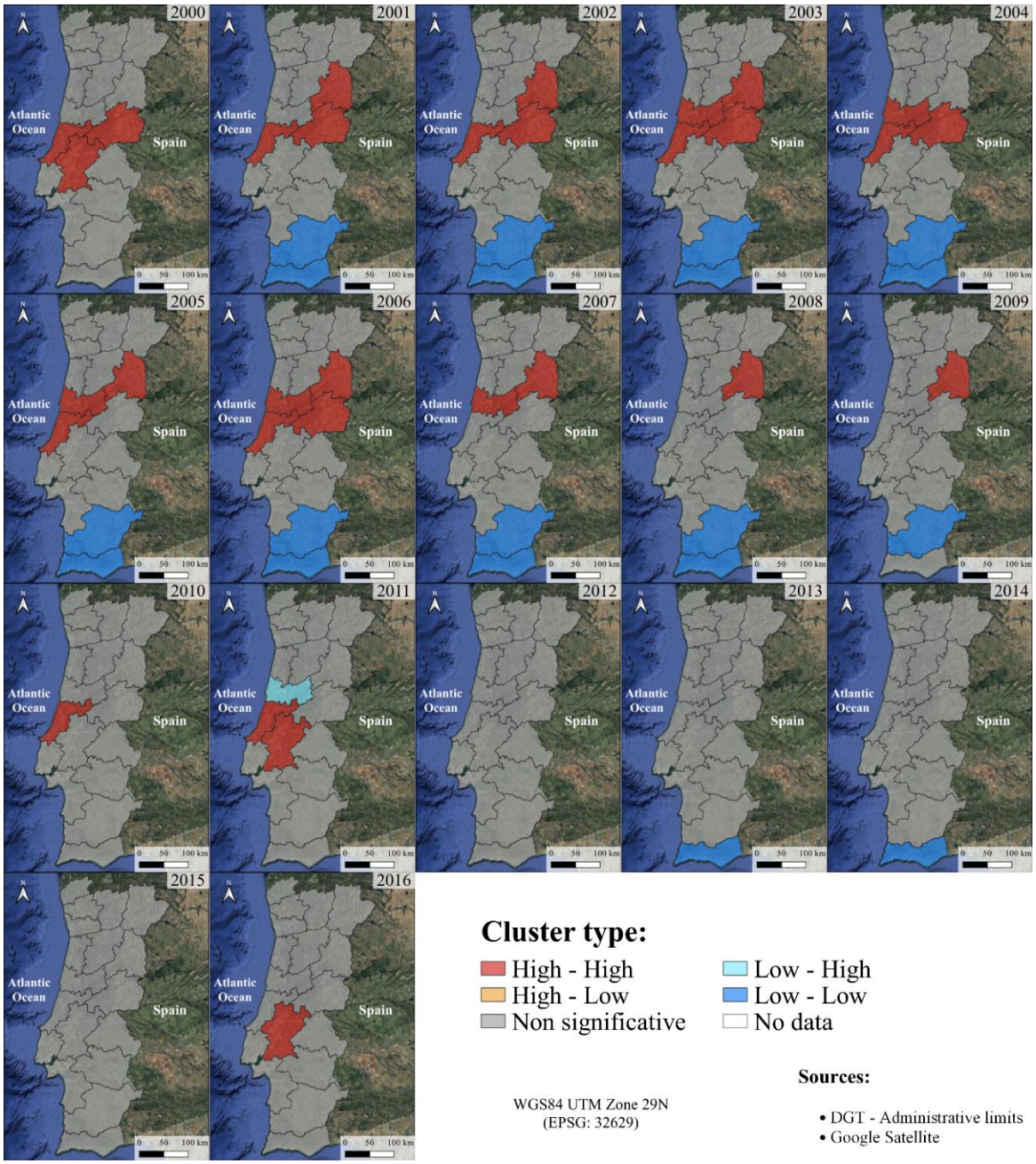


Figure 34 - Neurodegenerative diseases local clusters when considering districts as the administrative area (all diseases).

The local autocorrelation maps for each disease, per year, are presented in Figure 35 to Figure 45. The type of local clusters found differs greatly between diseases. For instance, Lewy and Other both seem to split the country in two similar areas, whereas the other diseases' cluster types appear more dispersed.

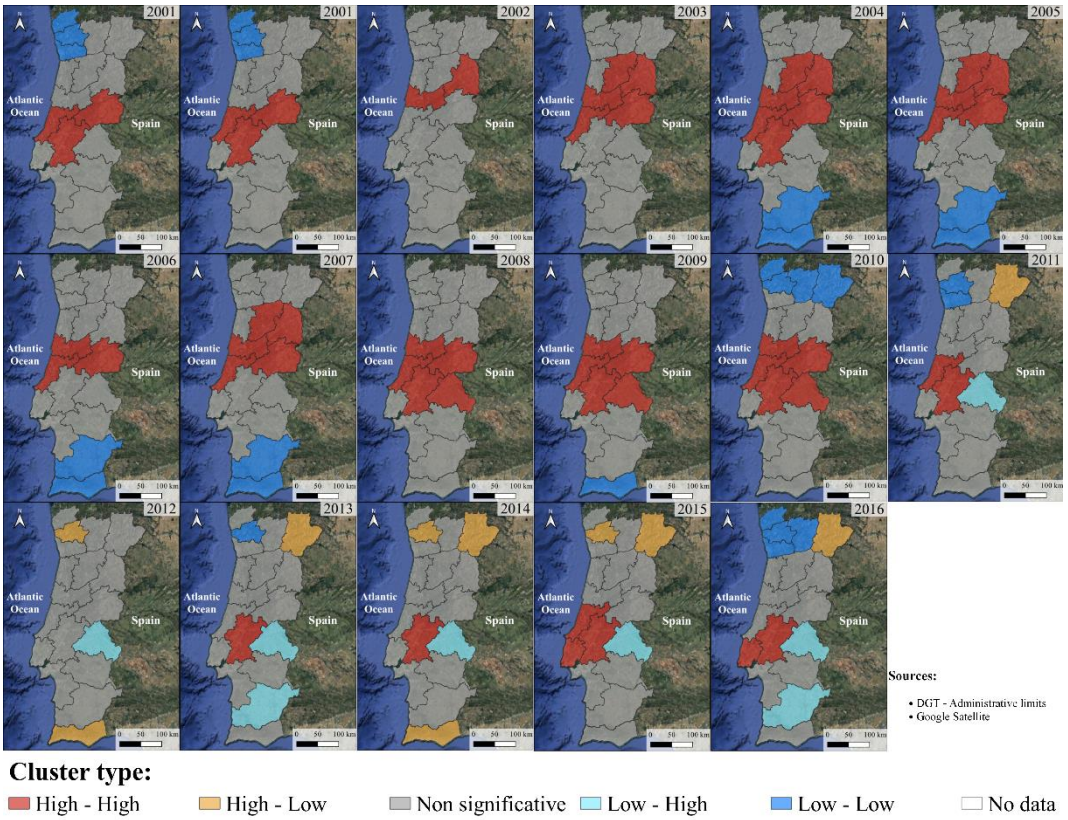


Figure 35 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Alzheimer).

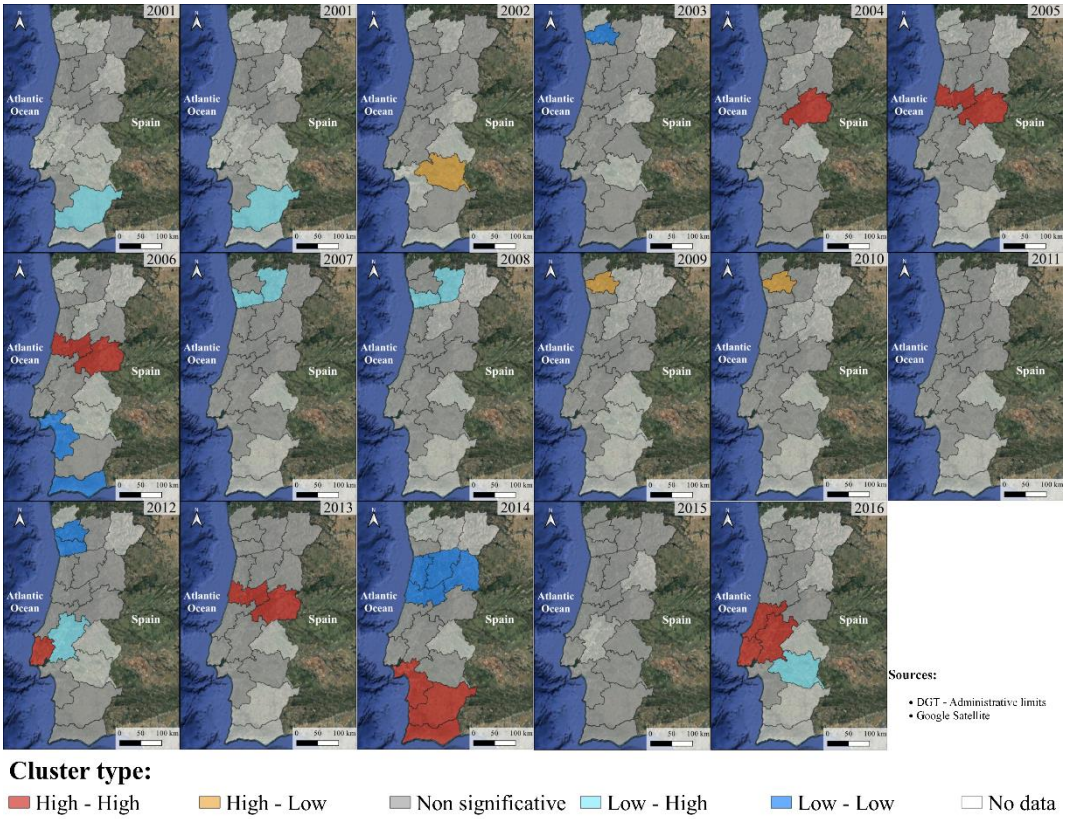


Figure 36 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Basal).

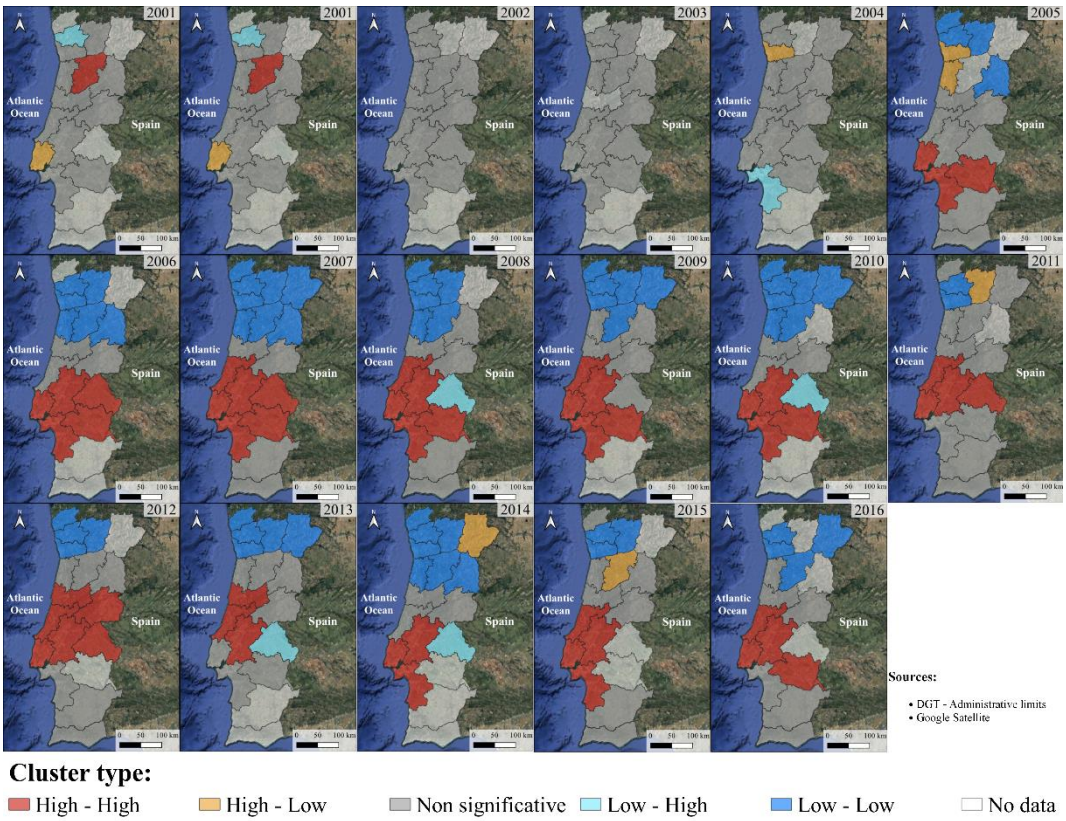


Figure 37 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Creutzfeldt).

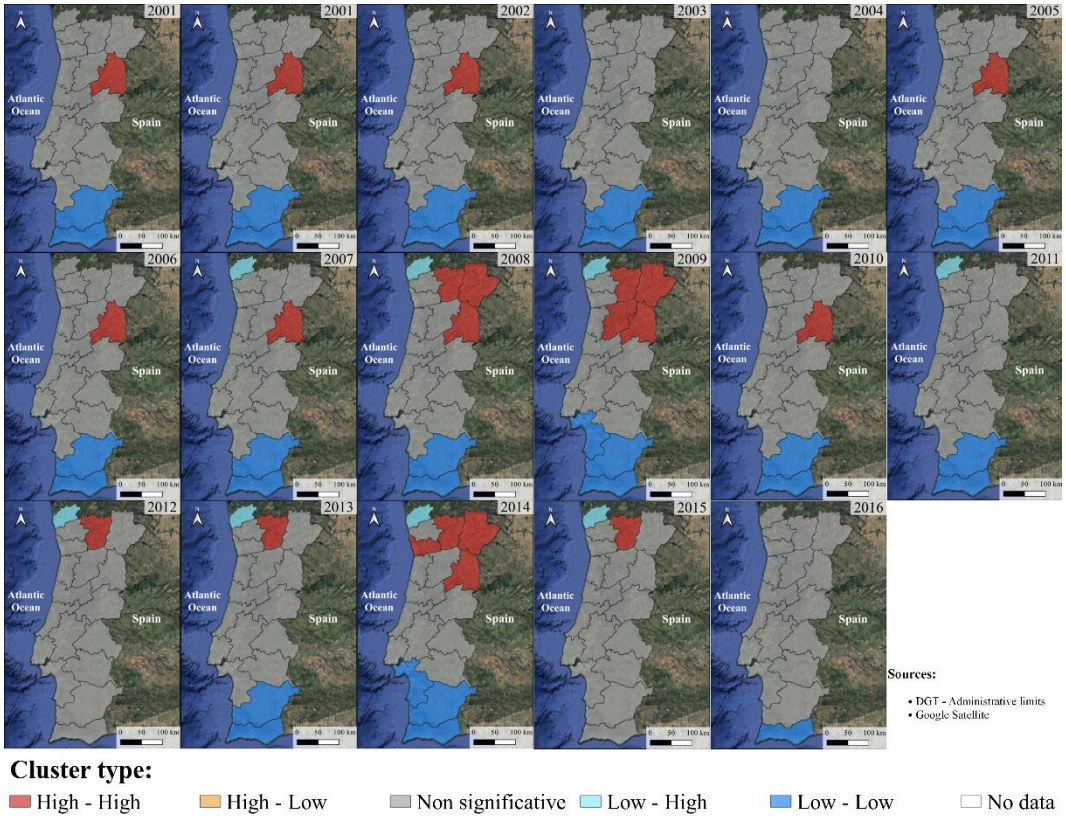


Figure 38 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Dementia).

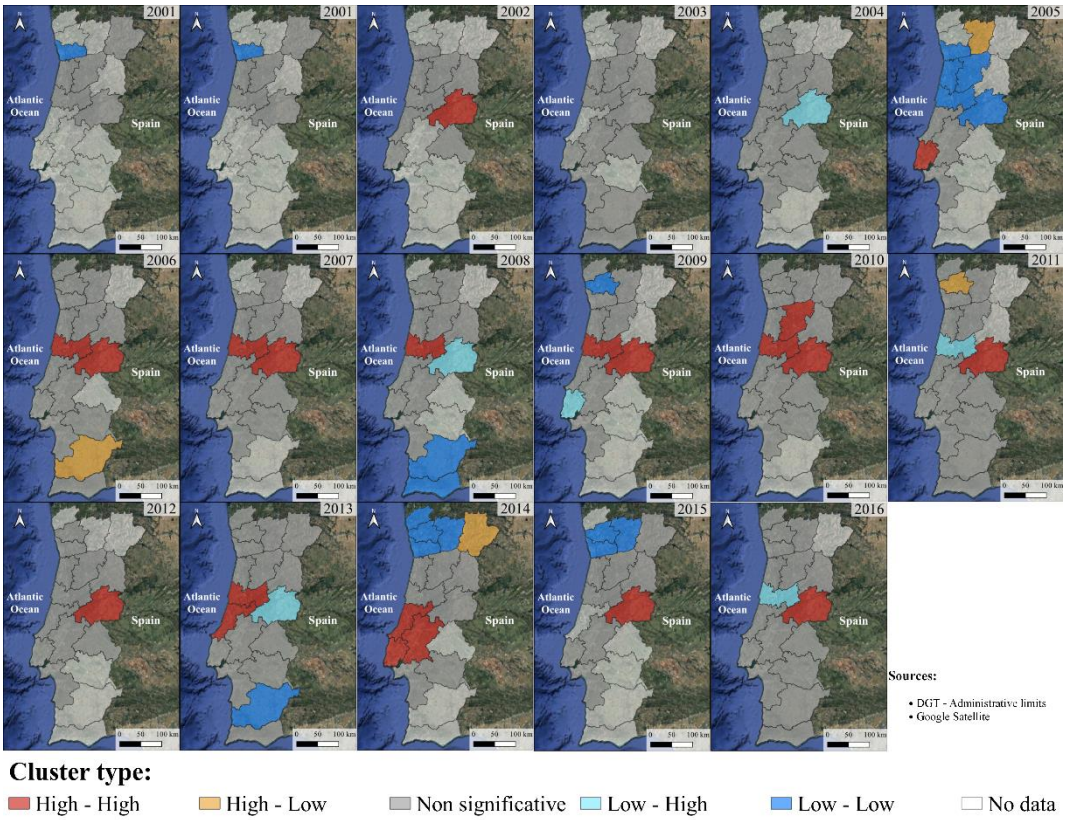


Figure 39 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Huntington).

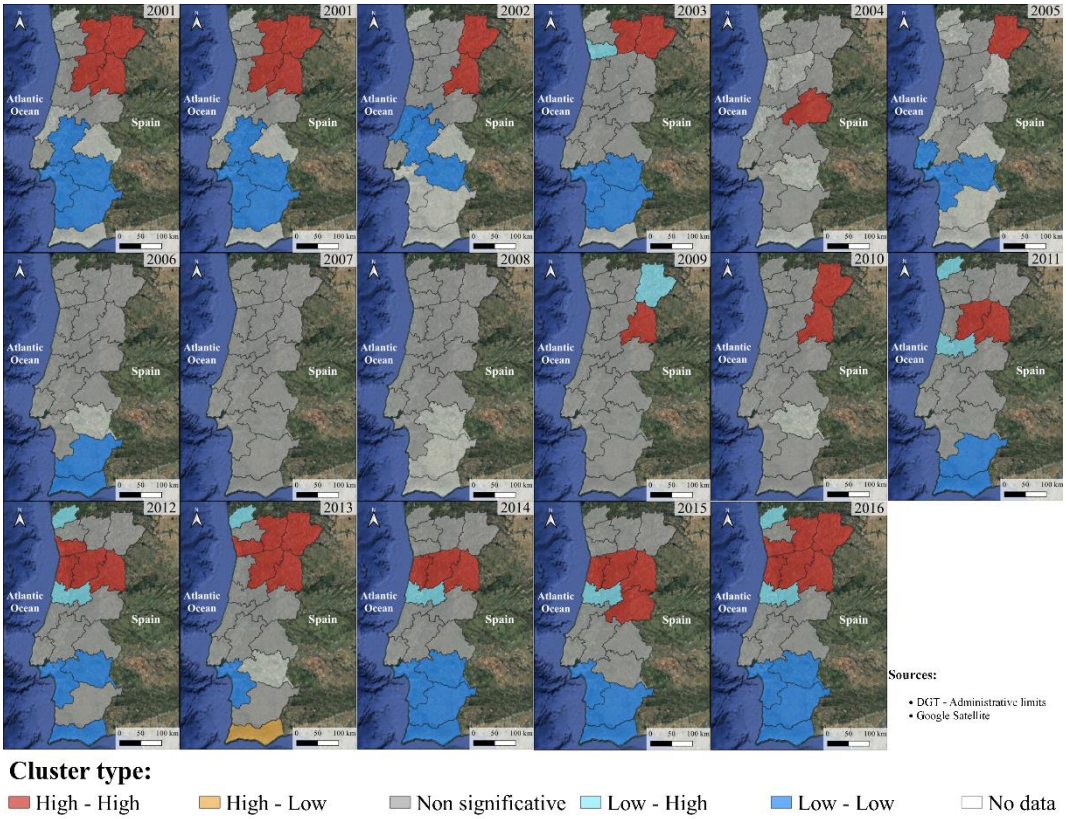


Figure 40 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Lewy).

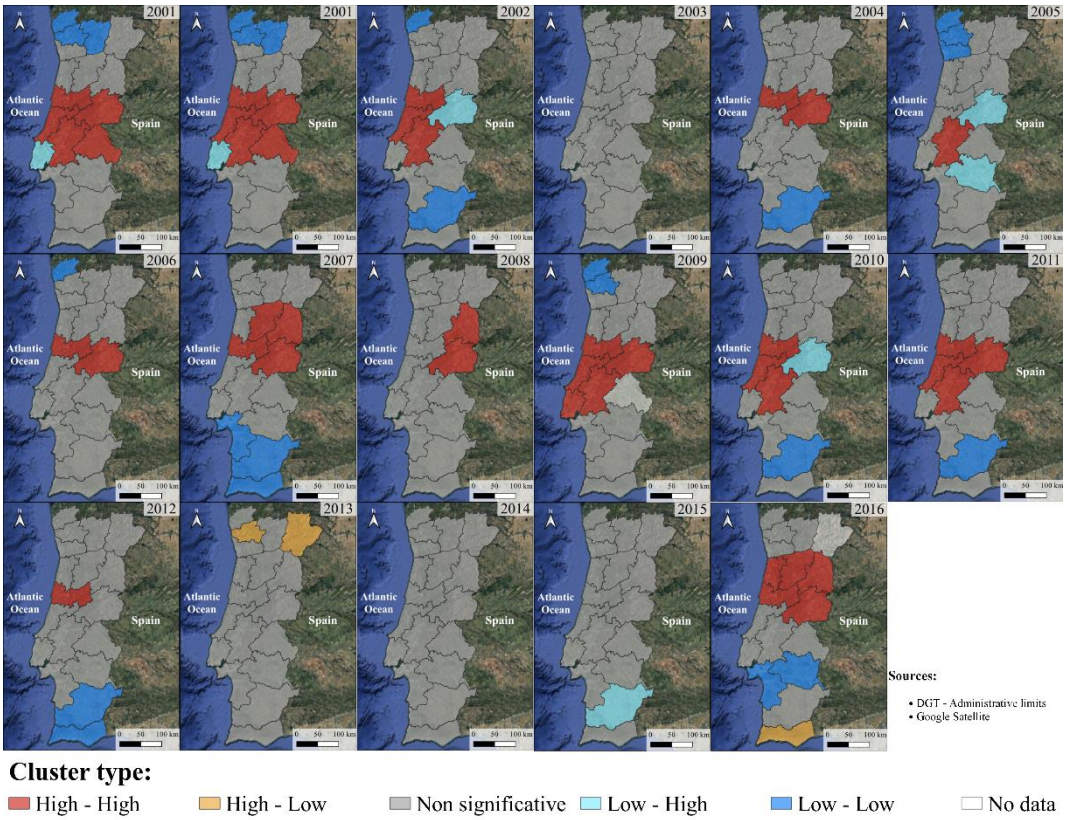


Figure 41 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Motor Neuron Disease).

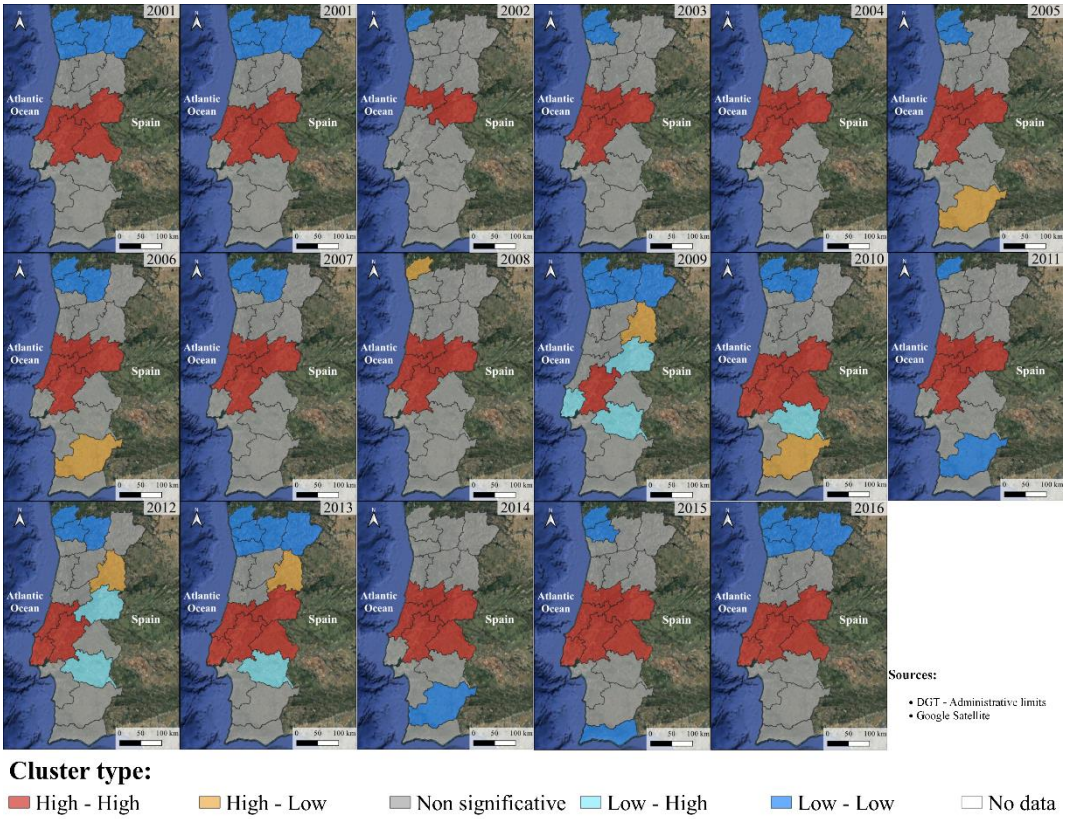


Figure 42 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Multiple Sclerosis).

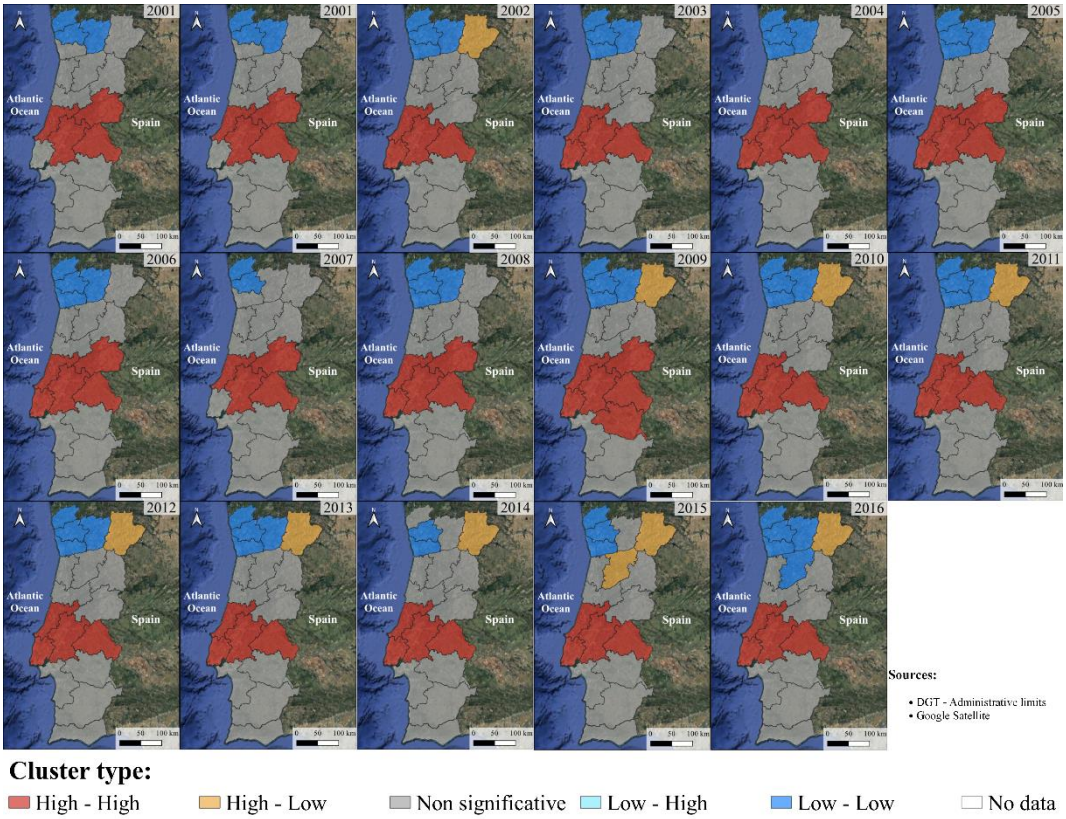


Figure 43 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Parkinson).

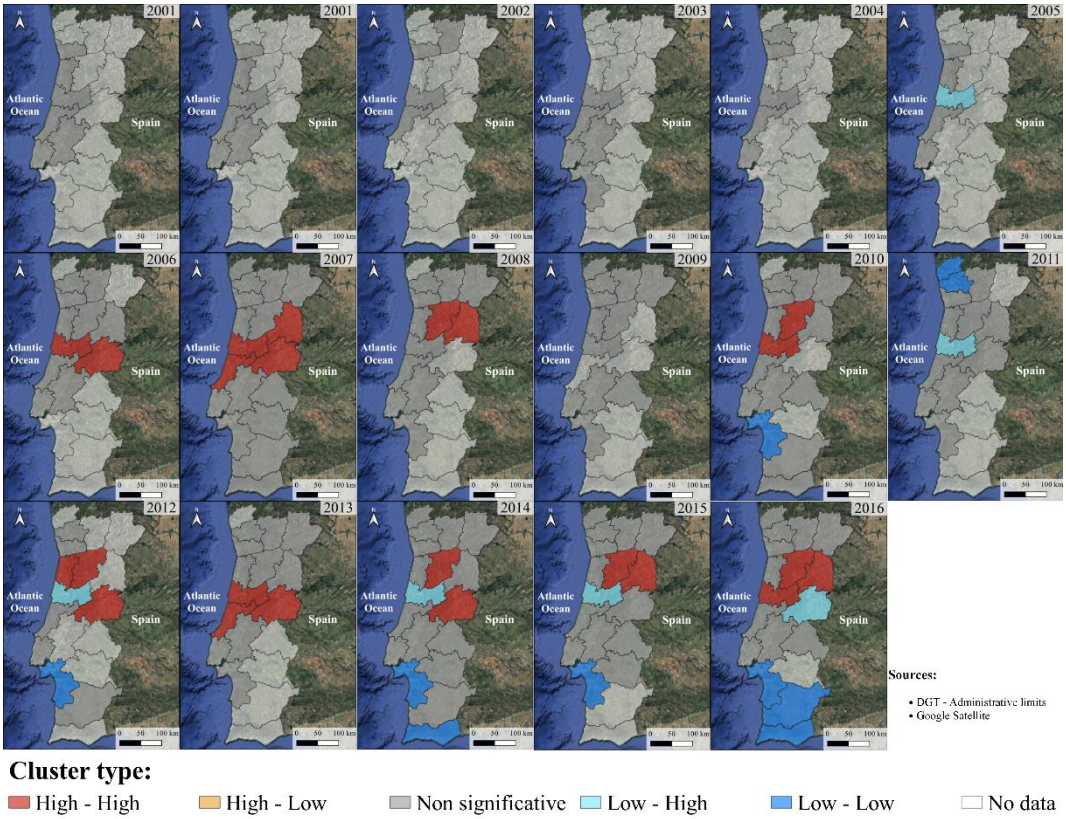


Figure 44 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Pick).

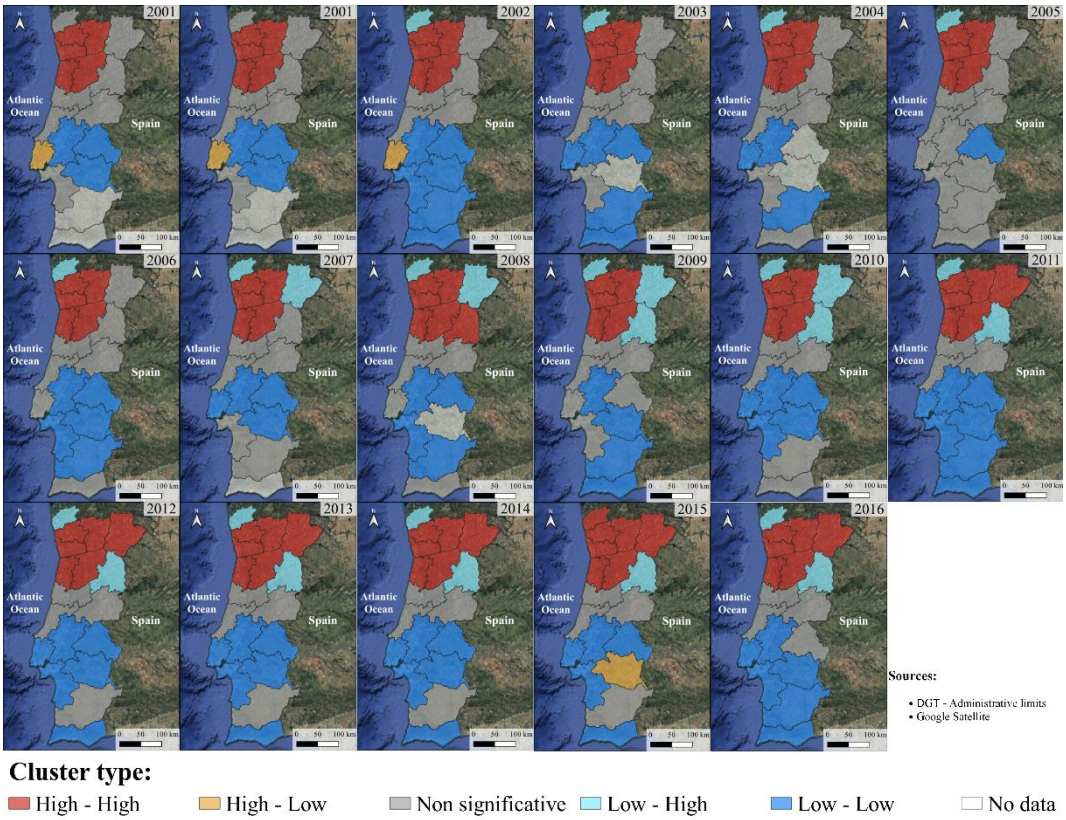


Figure 45 - Neurodegenerative diseases local clusters when considering districts as the administrative area (Other).

6.3.2.2.3. Local analysis (clusters): municipalities

Local spatial autocorrelation considering the municipalities as administrative areas was also mapped per year for all diseases (Figure 46). The clustering patterns found in the centre of Portugal were similar to those found in the districts analysis, but the south showed mostly low-high clusters, contrarily to the low-low clusters previously found in the districts analysis. Also, the northern littoral showed low-high clusters throughout almost all years, where before none had been significant.

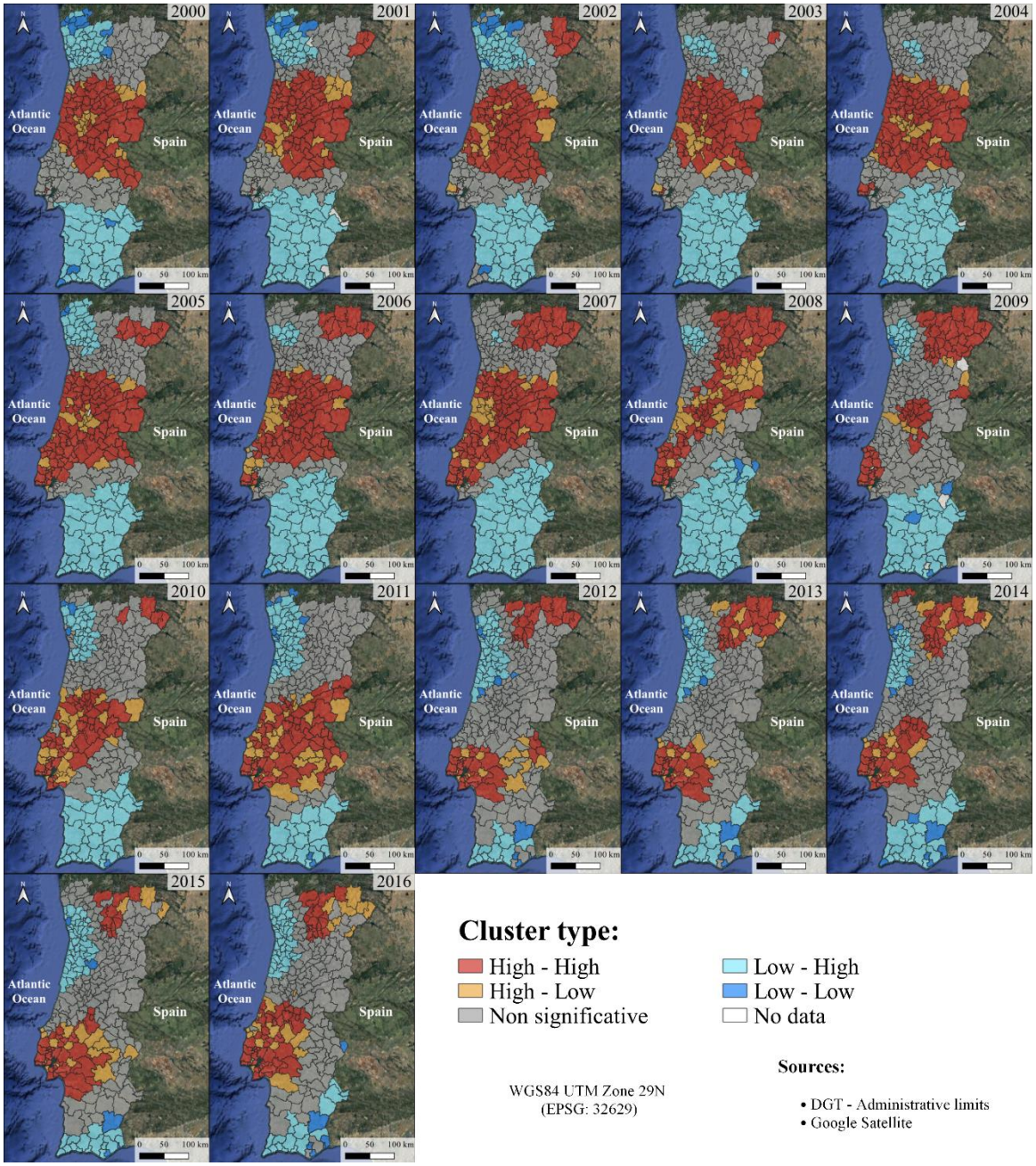


Figure 46 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (all diseases).

The local autocorrelation maps for each disease, per year, are presented in the supplementary materials Figure C1 to Figure C11. Areas with no data are now clearer and vary greatly among diseases, with Basal (Figure C2), Creutzfeldt (Figure C3). Huntington (Figure C5) and Pick (Figure C10) showing a lot of areas with no cases. These areas appear random, with the exception of Pick, in which high-high clusters seem to appear more often in the northern centre of the country. In line with the previous results for all diseases, the clusters found in municipalities did not fully correspond to those found in districts. For instance, in Alzheimer's clusters distribution (Figure C1), the south of Portugal had mostly low-high clusters from 2004 to 2007, whilst for districts, this area had been identified as a low-low cluster. This is also visible in Dementia maps (Figure C4), and Parkinson (Figure C9) follows a similar pattern, with no clusters found in the districts and low-high clusters found in the municipalities. There are also more non-significant clusters, which can be related to a lower number of cases per area unit than in the local analysis by districts.

6.4. Discussion

This paper was written on the sequence of a preliminary work recently published [253] and is, together with it, to the author's best knowledge, the first population based spatio-temporal descriptive study of neurodegenerative diseases in Portugal. The dataset from 17 years' worth hospitalizations in public Portuguese hospitals grants a significant sample of the neurodegenerative diseases in Portugal, which include patient's residence at the parish level, allowing to spatially analyse these diseases.

6.4.1. *Descriptive statistics*

As in line with literature, the most prevalent age group was that of patients over 70 years old, representing 81.3% of our sample [254]. The median age was below 70 years old for Creutzfeldt, Huntington, Motor Neuron, Multiple Sclerosis and Pick. Most of our sample were females (56.3%), which is supported by literature considering our data had mostly Alzheimer and Dementia's patients [255], even though Parkinson was the second with the most hospitalizations but, against literature [13], had the same proportion of female and male hospitalizations, and patients were hospitalized for a median of 8 days.

6.4.2. *Spatio-temporal analysis*

The temporal analysis clearly shows an increase in the ASHR per neurodegenerative diseases throughout all the districts. Braga and Bragança are the most affected districts. This increase in ASHR may arise from the ageing population of Portugal, where 21% of the population is 65 years old or older.

Neurodegenerative diseases ASHR increased from 2000 to 2016. Although a study had found rising prevalence of Alzheimer and other Dementias in all regions from 1990 to 2019, these rates appeared to stabilize when age-standardized [256]. However, multiple sclerosis prevalence, which is not as related

to population ageing, showed an increase of 30% from 2013 to 2020 and Europe had the highest prevalence of 142.81 per 100000 population, with a 32% increase from 2013 to 2020 [257]. To our knowledge, there are no other studies reporting spatial autocorrelation of neurodegenerative diseases.

This study also focuses on hospitalizations with more than one neurodegenerative disease coded, representing 10.2% of the total number of hospitalizations. Recognising this feature in our data allowed those hospitalizations to be treated as a single hospitalization when considering the overall hospitalization rates, while counting as many times as the number of neurodegenerative diseases diagnosed when accounting for each disease hospitalization rate.

Given that neurodegenerative diseases are not characterized by acute symptoms, focusing solely on primarily diagnosed hospitalizations would miss a lot of hospitalizations equally meaningful for this study. Thus, this study focused on representing all primarily and secondarily diagnosed neurodegenerative diseases, without missing the representation of primarily diagnosed hospitalizations, bringing a fuller understanding of the data.

Taking into account the high variability of the data across time, a statistical description based solely on the central tendency (mean) of the hospitalization rates using the whole 17 years at a time would be incomplete, since dispersion throughout time would not be considered. Therefore, in order to provide insight on the variation throughout the 17 years of the present study, we decided to present a more complete spatial-temporal analysis providing yearly maps, so that temporal patterns might be observed.

Global spatial autocorrelation was higher at a district level in the first years and at the municipality level in the final years of the study period, although in the final years the spatial correlation decreased reaching the absence of statistical significance. There are several factors that might explain the variation of the spatial autocorrelation throughout the years. Among them, one or multiple environmental risk factors may be behind the processes that trigger either the development of the disease or the exacerbation of symptoms caused by the disease, leading to the need for hospitalization care. The nature of the data in the present study, namely reuse of hospitalization data, might also be one such factor. Since reused data was not primarily collected with this study in mind, changes in the collection process may have occurred throughout time that are unknown to the authors. These hypotheses reflect the fact that acute episodes may represent reactiveness to environmental exposures.

Local spatial autocorrelation also showed differences between the districts and municipalities analysis. Although the centre of Portugal exhibited similar patterns in both analyses, the northern littoral and the south of the country showed different clusters. This is probably related to the spatial resolution used, in which some clusters were decomposed, showing different clusters mixed in the main cluster that had shown in the districts analysis. Also due to spatial resolution, statistical significance may have been affected, thus changing the clusters found from non-significant to significant and vice-versa. With this, areas around those may have changed its type of clustering.

6.4.3. *Limitations*

The major limitation of our work is the lack of unique patient identifier or variables to estimate it in our dataset. For this reason, a patient being hospitalized recurrently is treated as independent cases for each hospitalization, thus oversampling diseases that would require repetitive hospitalizations. To solve this problem, a unique patient identifier would be ideal and the implementation of other variables such as birthdate and ZIP code would also help provide enough information to detect duplicate entries.

Also, the increase in data availability and thus data coding raises the awareness that data is not always comparable throughout larger time periods [258] as is the case of this work. The adoption of ICD10 coding practices instead of ICD9 may also have impacted the quality and quantity of diagnosis coding in our data. With respect to the temporal range, at the time this work was started, data was unavailable from 2017 onwards, and thus it was only possible to analyse patterns from 2000 to 2016.

Hospitalization data is also not the most encompassing for neurodegenerative diseases and primary care data would be interesting to be analysed in future works. In addition, by using only public hospitals data, this study is missing part of the population of primary and private care patients.

Nevertheless, this is the best available option to study as large of a neurodegenerative diseases sample of patients, through the largest available time frame, without the implications of collecting the data prospectively. This is only possible through data reuse, which reiterates its importance in the development of new research topics.

6.4.4. *Implications and future work*

The insight it now provides in the spatial distribution and characteristics of neurodegenerative diseases in Portugal may be of relevance for any future work that relies on this topic or wants to make use of Portuguese public hospitalization data. For instance, future research work on environmental variables may allow to study the correlation between these patterns and environmental risk factors. Also, this research provides insights for public health planning and resource allocation, as well as the identification of intervention opportunities to improve patient outcomes. Public and professionals may also rely on this paper to increase their awareness on the distribution and impact of neurodegenerative diseases.

6.5. Conclusion

This work is a thorough analysis of the spatio-temporal patterns of neurodegenerative diseases' age and population-standardized hospitalization rates in Portugal mainland, from 2000 to 2016. The most prevalent disease in this setting was Dementia and the least prevalent was Basal.

The temporal analysis clearly shows the aging of the Portuguese population through the increasing incidence of neurodegenerative diseases. As for the spatial analysis, a significant positive global spatial autocorrelation was found for most diseases between administrative areas.

With this work, we hope to contribute for a better management of neurodegenerative diseases in Portugal. In particular, to enlighten future researches on this topic, exploring possible determinants of the identified spatio-temporal patterns, in order to contribute for the delineation of strategies aiming to mitigate the impact of neurodegenerative diseases, through their prevention and improvement of related health services.

Acknowledgements

This research was funded by FCT – Fundação para a Ciência e Tecnologia, grant number SFRH/BD/147324/2019 and national MCTES funds and supported also by National Funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., within CINTESIS, R&D Unit (reference UIDB/4255/2020).

Hospitalization data was provided by the Central Administration of the Health System, I.P. (ACSS).

Appendix C

Table C1 - Hospitalizations with more than one neurodegenerative disease - general descriptive (Hunt. is Huntington, Park. is Parkinson, N is the number of occurrences and IQI is the inter quartile interval).

Neurodegenerative diseases			N	%	Sex				Age		Length of stay	
					Male		Female					
					N	%	N	%	Median	IQI	Median	IQI
Alzheimer	Basal	-	-	6	0.0	2	33.3	4	66.7	74	[70, 79]	8	[6, 12]
		Dementia	-	3	0.0	2	66.7	1	33.3	73	[73, 74]	18	[12, 93]
	Creutzfeldt	-	-	13	0.0	4	30.8	9	69.2	80	[74, 86]	11	[5, 21]
		Dementia	-	4	0.0	2	50.0	2	50.0	78	[75, 81]	25	[12, 77]
		Other	-	2	0.0	1	50.0	1	50.0	82	[79, 84]	17	[11, 24]
	Dementia	-	-	19110	37.3	7243	37.9	11867	62.1	82	[77, 86]	8	[5, 14]
		Hunt.	-	2	0.0	1	50.0	1	50.0	74	[72, 77]	16	[12, 19]
			Other	1	0.0	0	0.0	1	100.0	72	[72, 72]	15	[15, 15]
		Lewy	-	131	0.3	66	50.4	65	49.6	82	[78, 85]	9	[5, 14]
			Park.	2	0.0	0	0.0	2	100.0	81	[79, 84]	12	[9, 14]
			Other	5	0.0	3	60.0	2	40.0	80	[75, 83]	6	[5, 8]
		MND	Other	1	0.0	1	100.0	0	0.0	82	[82, 82]	1	[1, 1]
		MS	-	6	0.0	4	66.7	2	33.3	80	[59, 85]	18	[13, 33]
		Parkinson	-	1143	2.2	535	46.8	608	53.2	81	[76, 85]	8	[5, 14]
			Other	27	0.1	13	48.1	14	51.9	81	[79, 86]	10	[5, 14]
		Pick	-	5	0.0	1	20.0	4	80.0	73	[71, 76]	14	[7, 16]
		Other	-	488	1.0	182	37.3	306	62.7	81	[75, 85]	10	[5, 18]
	Hunt.	-	-	8	0.0	2	25.0	6	75.0	69	[62, 77]	9	[6, 12]
	Lewy	-	-	71	0.1	36	50.7	35	49.3	80	[77, 85]	9	[5, 16]
	MND	-	-	12	0.0	7	58.3	5	41.7	72	[71, 76]	20	[10, 42]
		Parkinson	-	2	0.0	0	0.0	2	100.0	68	[68, 69]	12	[10, 13]
	MS	-	-	19	0.0	4	21.1	15	78.9	78	[75, 84]	8	[3, 14]
		Parkinson	-	1	0.0	0	0.0	1	100.0	85	[85, 85]	7	[7, 7]
	Parkinson	-	-	5391	10.5	2543	47.2	2848	52.8	81	[76, 85]	8	[5, 14]
		Lewy	-	12	0.0	10	83.3	2	16.7	84	[78, 88]	9	[6, 12]
		Pick	-	1	0.0	1	100.0	0	0.0	67	[67, 67]	8	[8, 8]
		Other	-	35	0.1	13	37.1	22	62.9	80	[77, 84]	11	[5, 19]
	Pick	-	-	13	0.0	8	61.5	5	38.5	73	[69, 80]	17	[6, 50]
	Other	-	-	372	0.7	140	37.6	232	62.4	81	[77, 85]	9	[5, 15]
Basal	Creutzfeldt	Dementia	Park.	1	0.0	1	100.0	0	0.0	42	[42, 42]	50	[50, 50]
	Dementia	-	-	94	0.2	40	42.6	54	57.4	74	[68, 80]	10	[6, 19]
		Lewy	-	4	0.0	4	100.0	0	0.0	71	[67, 77]	14	[6, 23]
		Parkinson	-	4	0.0	2	50.0	2	50.0	76	[75, 77]	9	[5, 16]
		Pick	-	3	0.0	2	66.7	1	33.3	70	[70, 72]	63	[36, 71]
		Other	-	2	0.0	0	0.0	2	100.0	63	[60, 67]	12	[11, 14]

	Hunt.	-	-	1	0.0	1	100.0	0	0.0	60	[60, 60]	6	[6, 6]
	Lewy	-	-	3	0.0	1	33.3	2	66.7	81	[77, 84]	8	[6, 19]
	MND	-	-	6	0.0	3	50.0	3	50.0	70	[52, 80]	29	[6, 89]
		Parkinson	-	2	0.0	2	100.0	0	0.0	70	[65, 76]	5	[3, 6]
	MS	-	-	3	0.0	2	66.7	1	33.3	57	[48, 58]	14	[13, 14]
	Parkinson	-	-	57	0.1	25	43.9	32	56.1	75	[66, 81]	8	[4, 16]
		Other	-	1	0.0	1	100.0	0	0.0	64	[64, 64]	9	[9, 9]
	Other	-	-	10	0.0	4	40.0	6	60.0	71	[68, 77]	12	[8, 19]
Creutzfeldt	Dementia	-	-	277	0.5	141	50.9	136	49.1	67	[58, 76]	22	[10, 50]
		Lewy	-	3	0.0	2	66.7	1	33.3	65	[64, 66]	136	[84, 378]
		MS	-	1	0.0	0	0.0	1	100.0	63	[63, 63]	29	[29, 29]
		Parkinson	-	9	0.0	5	55.6	4	44.4	72	[64, 80]	23	[12, 33]
		Other	-	10	0.0	7	70.0	3	30.0	69	[60, 74]	26	[13, 91]
	Lewy	-	-	5	0.0	4	80.0	1	20.0	43	[40, 80]	18	[10, 26]
		MS	-	1	0.0	0	0.0	1	100.0	43	[43, 43]	31	[31, 31]
		Other	-	1	0.0	1	100.0	0	0.0	34	[34, 34]	11	[11, 11]
	MND	-	-	2	0.0	1	50.0	1	50.0	68	[67, 69]	32	[17, 46]
	MS	-	-	30	0.1	5	16.7	25	83.3	37	[35, 50]	14	[5, 48]
	Parkinson	-	-	33	0.1	16	48.5	17	51.5	75	[70, 81]	11	[6, 21]
	Pick	-	-	1	0.0	1	100.0	0	0.0	35	[35, 35]	6	[6, 6]
	Other	-	-	75	0.1	51	68.0	24	32.0	45	[39, 61]	15	[10, 31]
Dementia	Hunt.	-	-	130	0.3	61	46.9	69	53.1	63	[50, 72]	10	[7, 19]
		Parkinson	-	1	0.0	0	0.0	1	100.0	75	[75, 75]	48	[48, 48]
		Pick	-	2	0.0	2	100	0	0.0	47	[47, 47]	12	[6, 17]
		Other	-	2	0.0	1	50.0	1	50.0	66	[60, 71]	21	[15, 27]
	Lewy	-	-	1965	3.8	991	50.4	974	49.6	79	[74, 84]	10	[5, 17]
		MND	-	3	0.0	1	33.3	2	66.7	74	[70, 74]	12	[11, 27]
		MS	-	1	0.0	0	0.0	1	100.0	69	[69, 69]	17	[17, 17]
		Parkinson	-	162	0.3	89	54.9	73	45.1	81	[74, 85]	8	[4, 15]
			Other	5	0.0	3	60.0	2	40.0	77	[72, 83]	29	[6, 33]
		Pick	-	5	0.0	3	60.0	2	40.0	65	[57, 76]	5	[4, 6]
			Other	1	0.0	1	100.0	0	0.0	39	[39, 39]	172	[172, 172]
		Other	-	84	0.2	40	47.6	44	52.4	78	[74, 83]	13	[9, 22]
	MND	-	-	117	0.2	60	51.3	57	48.7	75	[66, 81]	9	[4, 16]
		Parkinson	-	14	0.0	5	35.7	9	64.3	72	[67, 78]	12	[7, 18]
			Pick	1	0.0	0	0.0	1	100.0	71	[71, 71]	59	[59, 59]
		Pick	-	22	0.0	10	45.5	12	54.5	64	[60, 66]	12	[6, 23]
		Other	-	7	0.0	5	71.4	2	28.6	74	[73, 77]	15	[9, 18]
	MS	-	-	149	0.3	41	27.5	108	72.5	62	[49, 74]	9	[5, 16]
		Parkinson	-	6	0.0	4	66.7	2	33.3	79	[75, 81]	9	[4, 13]
		Other	-	6	0.0	2	33.3	4	66.7	61	[54, 64]	11	[7, 13]

		-	-	14460	28.2	7087	49.0	7373	51.0	81	[76, 86]	8	[5, 14]
	Parkinson	Pick	-	27	0.1	15	55.6	12	44.4	68	[61, 76]	9	[7, 22]
			Other	1	0.0	1	100.0	0	0.0	65	[65, 65]	8	[8, 8]
		Other	-	226	0.4	112	49.6	114	50.4	80	[74, 85]	10	[6, 18]
	Pick	-	-	614	1.2	323	52.6	291	47.4	67	[60, 76]	13	[6, 27]
		Other	-	23	0.0	13	56.5	10	43.5	64	[59, 70]	25	[10, 49]
	Other	-	-	4131	8.1	1868	45.2	2263	54.8	81	[74, 86]	10	[6, 18]
Huntington	Lewy	-	-	1	0.0	0	0.0	1	100.0	58	[58, 58]	48	[48, 48]
	MS	-	-	4	0.0	0	0.0	4	100.0	35	[34, 36]	12	[8, 14]
	Parkinson	-	-	13	0.0	5	38.5	8	61.5	72	[66, 79]	13	[6, 15]
	Other	-	-	15	0.0	11	73.3	4	26.7	60	[47, 74]	12	[8, 29]
Lewy	MND	-	-	4	0.0	0	0.0	4	100.0	72	[67, 77]	7	[6, 8]
		Parkinson	-	2	0.0	2	100.0	0	0.0	73	[72, 73]	20	[14, 25]
	MS	-	-	15	0.0	5	33.3	10	66.7	43	[34, 49]	5	[4, 11]
		Other	-	3	0.0	1	33.3	2	66.7	50	[49, 58]	11	[9, 11]
	Parkinson	-	-	391	0.8	220	56.3	171	43.7	78	[70, 83]	9	[5, 16]
		Other	-	7	0.0	4	57.1	3	42.9	65	[62, 78]	17	[11, 20]
	Pick	-	-	4	0.0	2	50.0	2	50.0	77	[68, 83]	28	[13, 56]
	Other	-	-	53	0.1	27	50.9	26	49.1	74	[66, 81]	11	[6, 22]
MND	MS	-	-	31	0.1	1	3.2	30	96.8	42	[40, 43]	4	[3, 5]
	Parkinson	-	-	83	0.2	41	49.4	42	50.6	72	[68, 79]	9	[5, 16]
		Pick	-	1	0.0	1	100.0	0	0.0	49	[49, 49]	2	[2, 2]
		Other	-	1	0.0	0	0.0	1	100.0	91	[91, 91]	26	[26, 26]
	Pick	-	-	26	0.1	15	57.7	11	42.3	63	[57, 68]	6	[4, 17]
		Other	-	1	0.0	0	0.0	1	100.0	56	[56, 56]	1	[1, 1]
	Other	-	-	41	0.1	22	53.7	19	46.3	71	[66, 78]	13	[7, 21]
MS	Parkinson	-	-	53	0.1	23	43.4	30	56.6	67	[62, 77]	10	[4, 15]
	Pick	-	-	2	0.0	0	0.0	2	100.0	83	[83, 83]	8	[8, 9]
	Other	-	-	45	0.1	21	46.7	24	53.3	47	[38, 55]	12	[6, 17]
Park.	Pick	-	-	36	0.1	23	63.9	13	36.1	70	[63, 76]	10	[6, 19]
		Other	-	1	0.0	1	100.0	0	0.0	72	[72, 72]	12	[12, 12]
	Other	-	-	674	1.3	345	51.2	329	48.8	77	[71, 83]	9	[5, 17]
Pick	Other	-	-	23	0.0	12	52.2	11	47.8	68	[62, 75]	9	[5, 22]
		Total		51200	100.0								

Table C2 - Descriptive statistics of hospitalizations with neurodegenerative disease as primary diagnosis included in this study. Categorical data (sex, age group and residence) is described by its absolute frequency and percentage, while continuous data (age and length of stay, which are both asymmetrically distributed), is described by its median and interquartile range. The total row describes the total frequency of each disease and their respective percentage. yo: years old.

		Alzheimer	Basal	Creutzfeldt	Dementia	Huntington	Lewy	Motor Neuron	Parkinson	Pick	Other	Total
Sex	**Female**	2229 (60.1)	259 (51.0)	297 (43.8)	9225 (54.8)	238 (51.9)	624 (52.8)	1852 (48.2)	2832 (45.5)	324 (49.8)	93 (56.0)	17973 (52.5)
	Male	1478 (39.9)	249 (49.0)	381 (56.2)	7609 (45.2)	221 (48.1)	557 (47.2)	1994 (51.8)	3388 (54.5)	327 (50.2)	73 (44.0)	16277 (47.5)
Age (yo)		77 [70, 82]	65 [55, 72]	58 [38, 71]	76 [69, 82]	50 [41, 60]	71 [62, 78]	65 [55, 73]	68 [60, 76]	65 [58, 72]	68 [54, 77]	73 [63, 80]
Age Group	**<20 yo**	3 (0.1)	26 (5.1)	67 (9.9)	28 (0.2)	15 (3.3)	53 (4.5)	244 (6.3)	21 (0.3)	1 (0.2)	10 (6.0)	468 (1.4)
	20-70 yo	940 (25.4)	323 (63.6)	441 (65.0)	5008 (29.7)	396 (86.3)	504 (42.7)	2359 (61.3)	3662 (58.9)	457 (70.2)	83 (50.0)	14173 (41.4)
	>70 yo	2764 (74.6)	159 (31.3)	170 (25.1)	11798 (70.1)	48 (10.5)	624 (52.8)	1243 (32.3)	2537 (40.8)	193 (29.6)	73 (44.0)	19609 (57.3)
Length of stay (days)		7 [3, 16]	9 [4, 15]	19 [7, 45]	10 [4, 20]	10 [5, 18]	9 [4, 19]	6 [3, 13]	7 [4, 13]	16 [6, 31]	8 [2, 14]	9 [4, 18]
Residence	**Norte**	1202 (32.4)	167 (32.9)	213 (31.4)	5815 (34.5)	105 (22.9)	385 (32.6)	1221 (31.7)	1928 (31.0)	207 (31.8)	57 (34.3)	11300 (33.0)
	Centro	2028 (54.7)	274 (53.9)	306 (45.1)	8710 (51.7)	279 (60.8)	679 (57.5)	1968 (51.2)	3394 (54.6)	383 (58.8)	67 (40.4)	18088 (52.8)
	Sul	477 (12.9)	67 (13.2)	159 (23.5)	2309 (13.7)	75 (16.3)	117 (9.9)	657 (17.1)	898 (14.4)	61 (9.4)	42 (25.3)	4862 (14.2)
Total		3707 (10.8)	508 (1.5)	678 (2.0)	16834 (49.2)	459 (1.3)	1181 (3.4)	3846 (11.2)	6220 (18.2)	651 (1.9)	166 (0.5)	34250 (100)

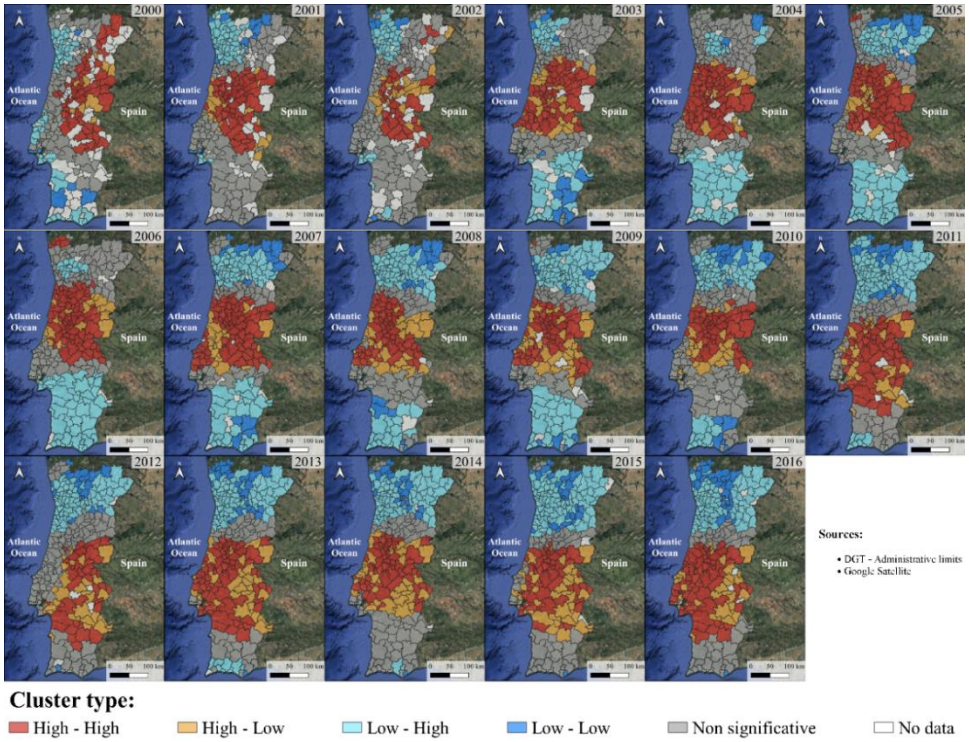


Figure C1 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Alzheimer).

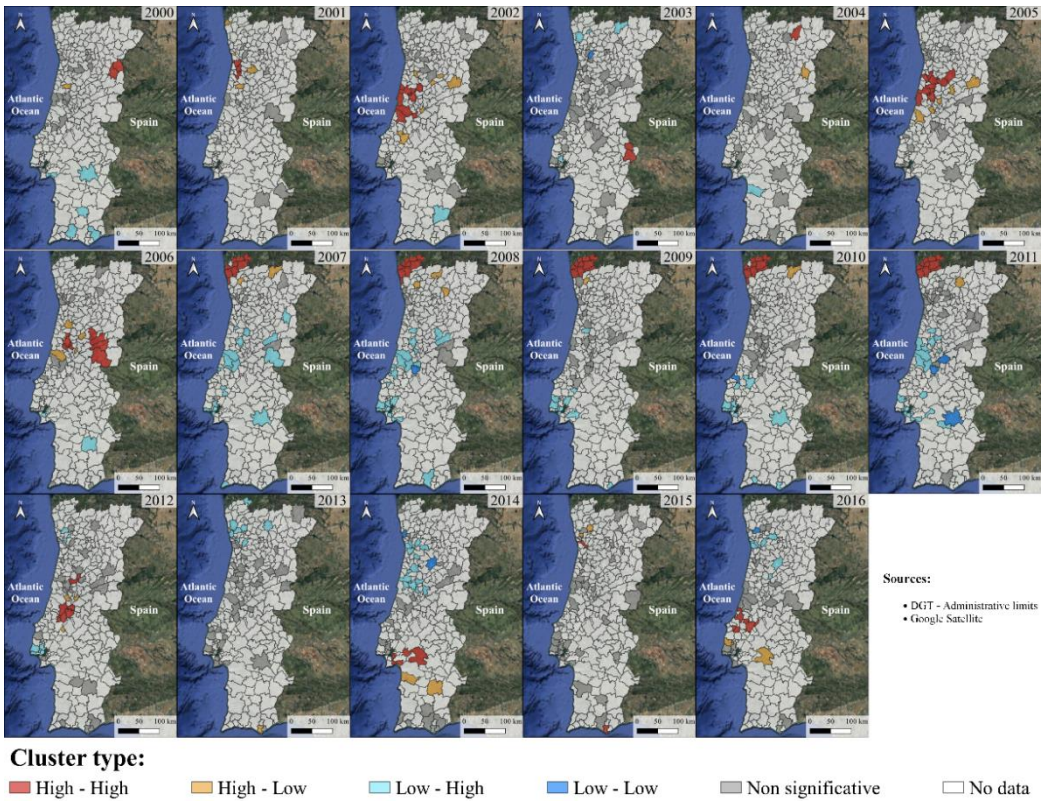


Figure C2 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Basal).

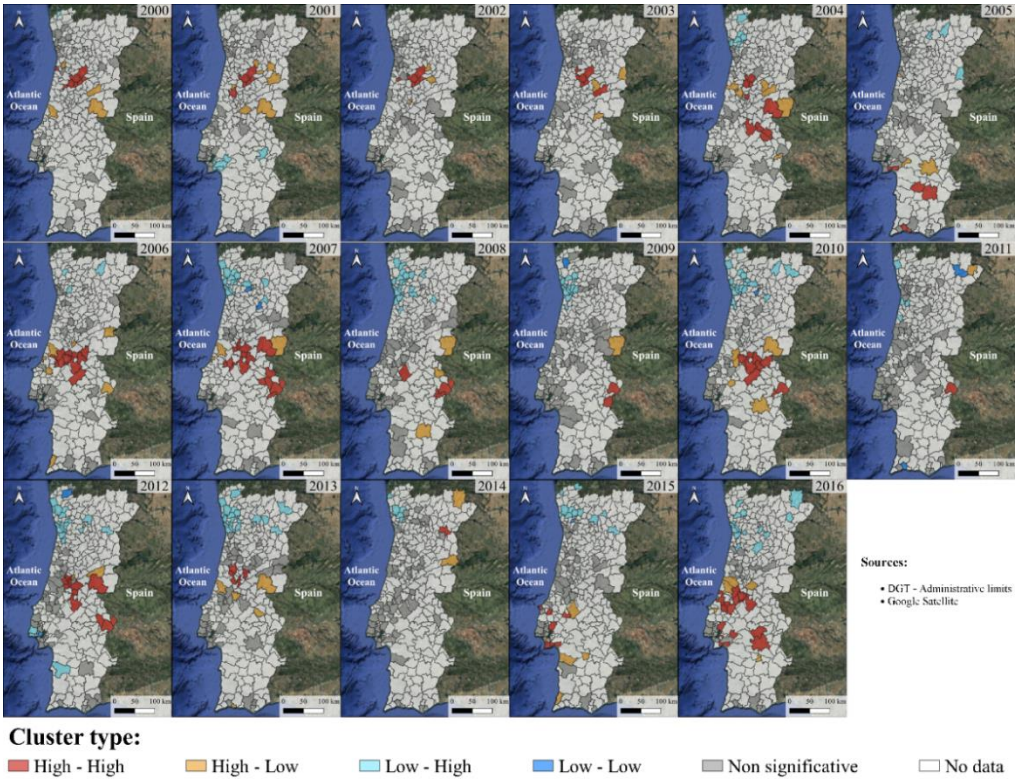


Figure C3 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Creutzfeldt).

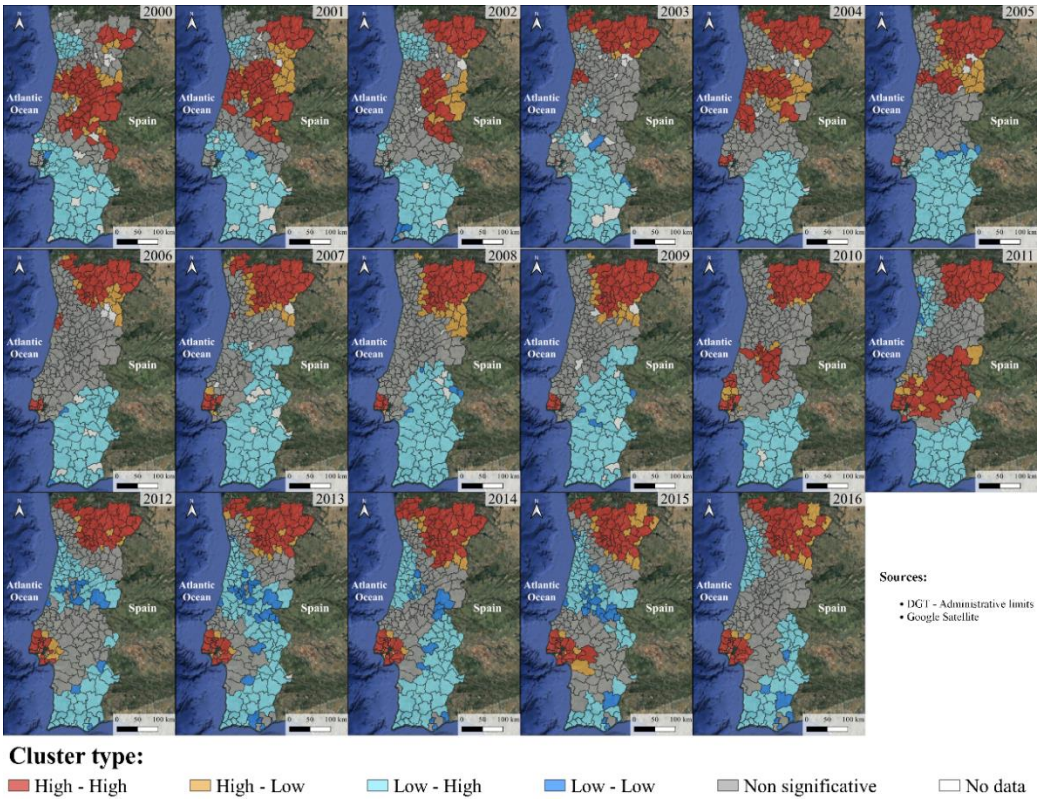


Figure C4 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Dementia).

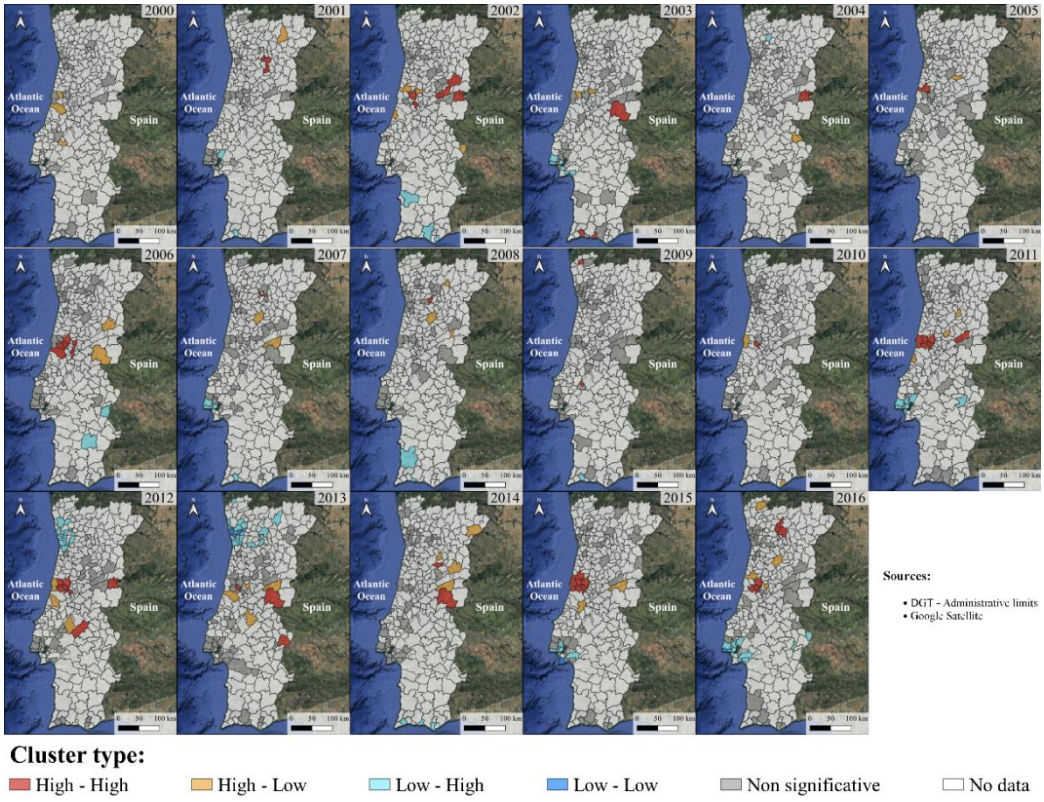


Figure C5 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Huntington).

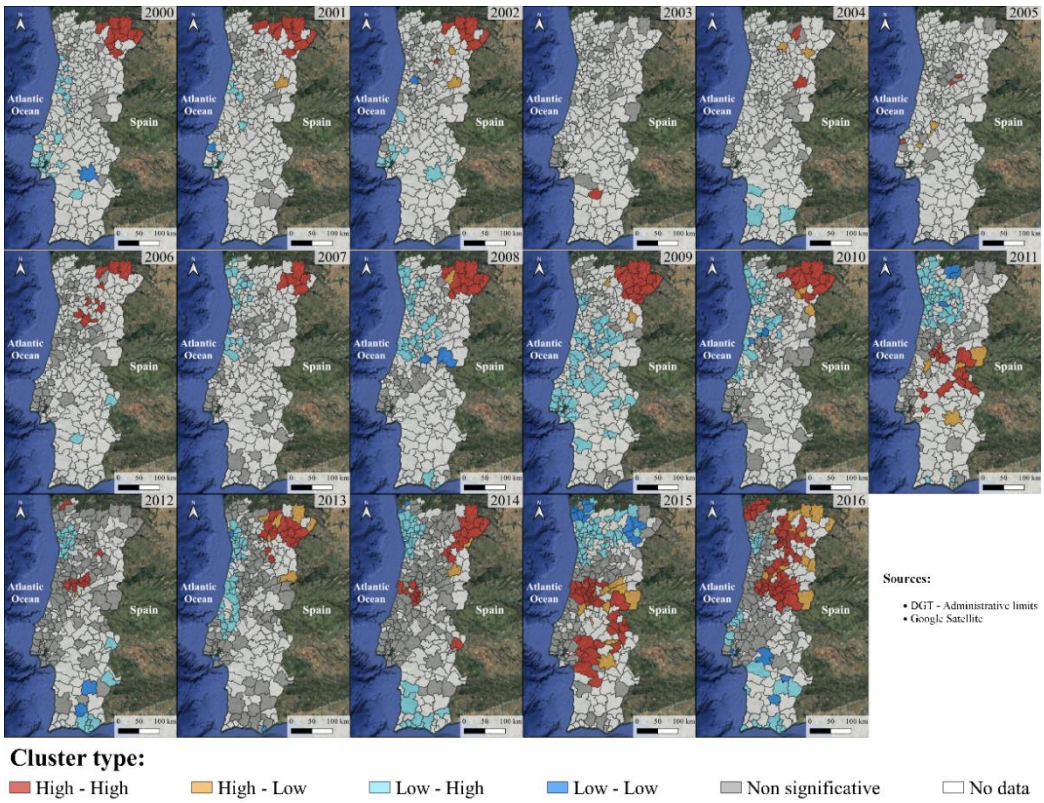


Figure C6 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Lewy).

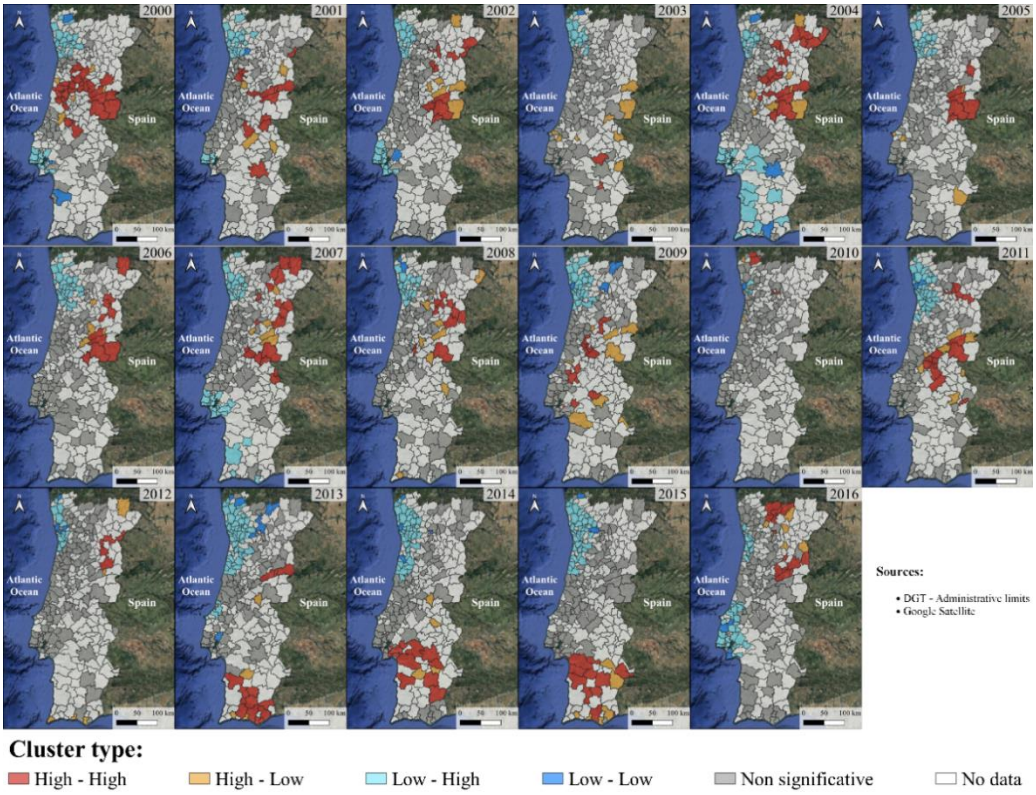


Figure C7 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Motor Neuron Disease).

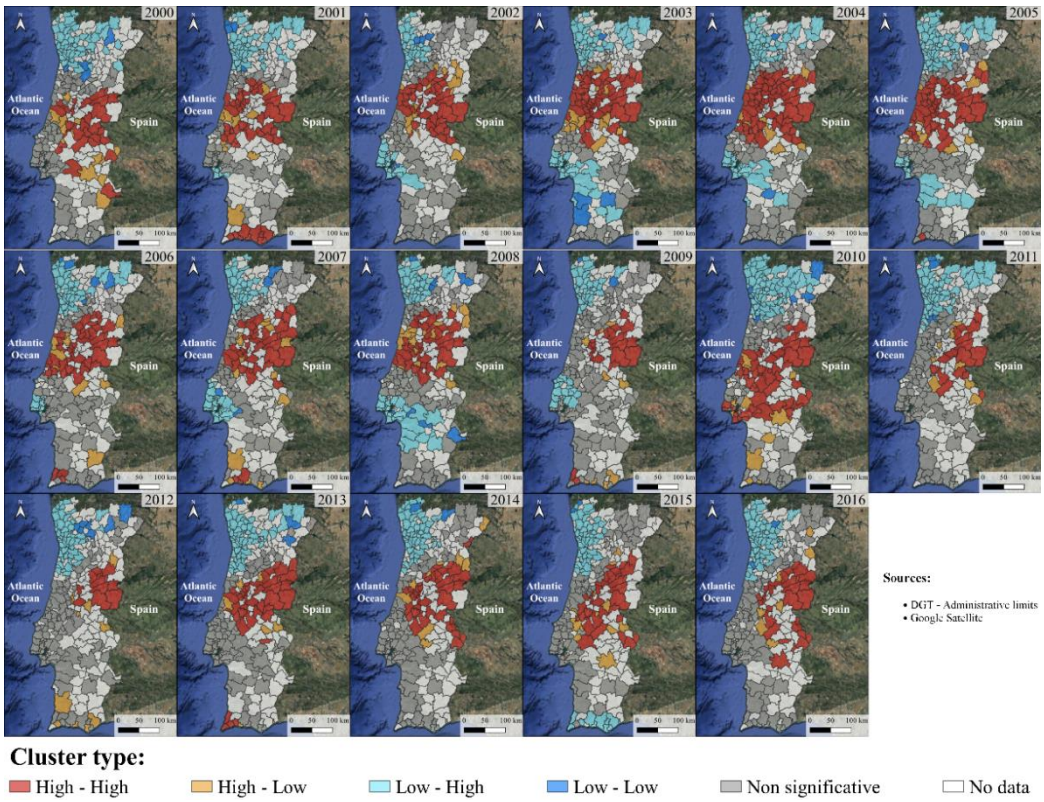


Figure C8 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Multiple Sclerosis).

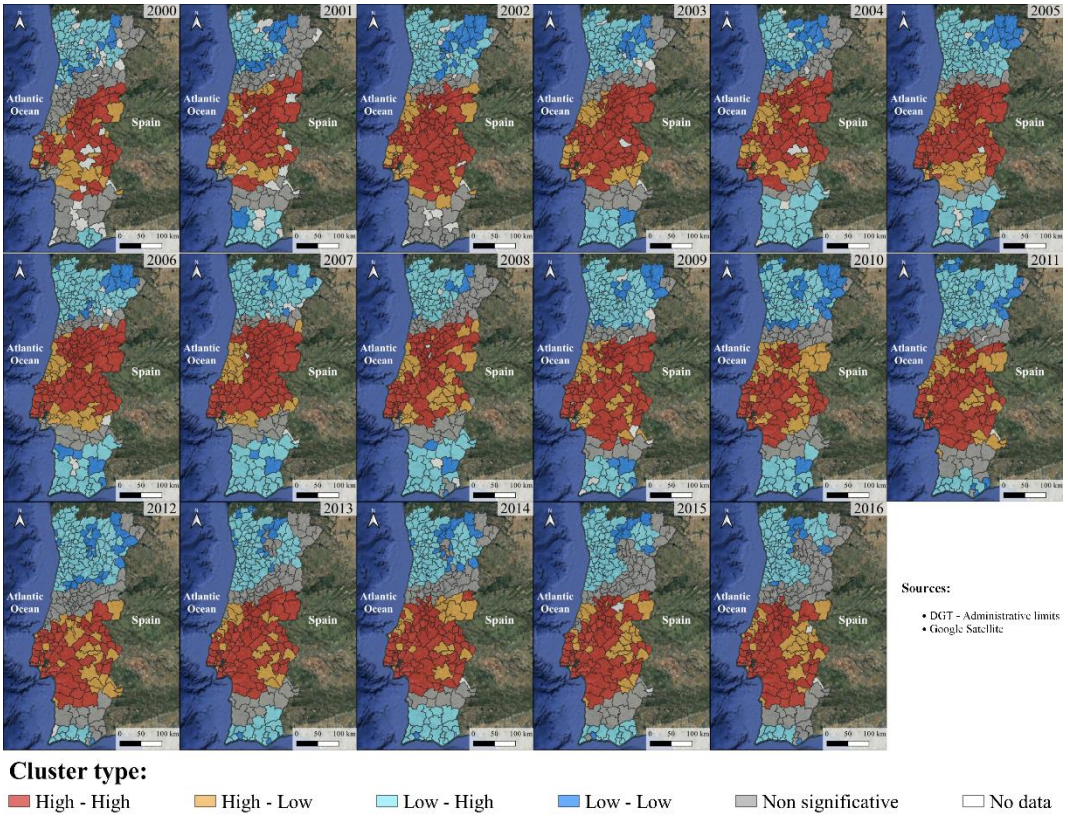


Figure C9 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Parkinson).

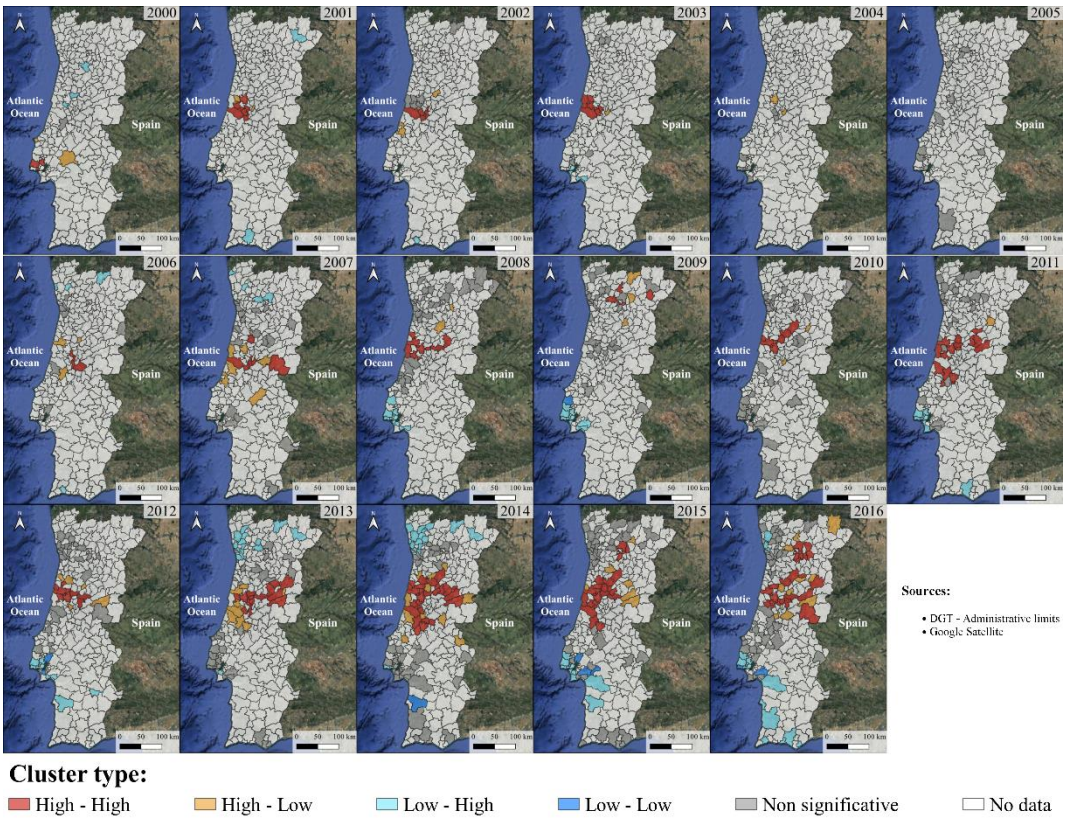


Figure C10 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Pick).

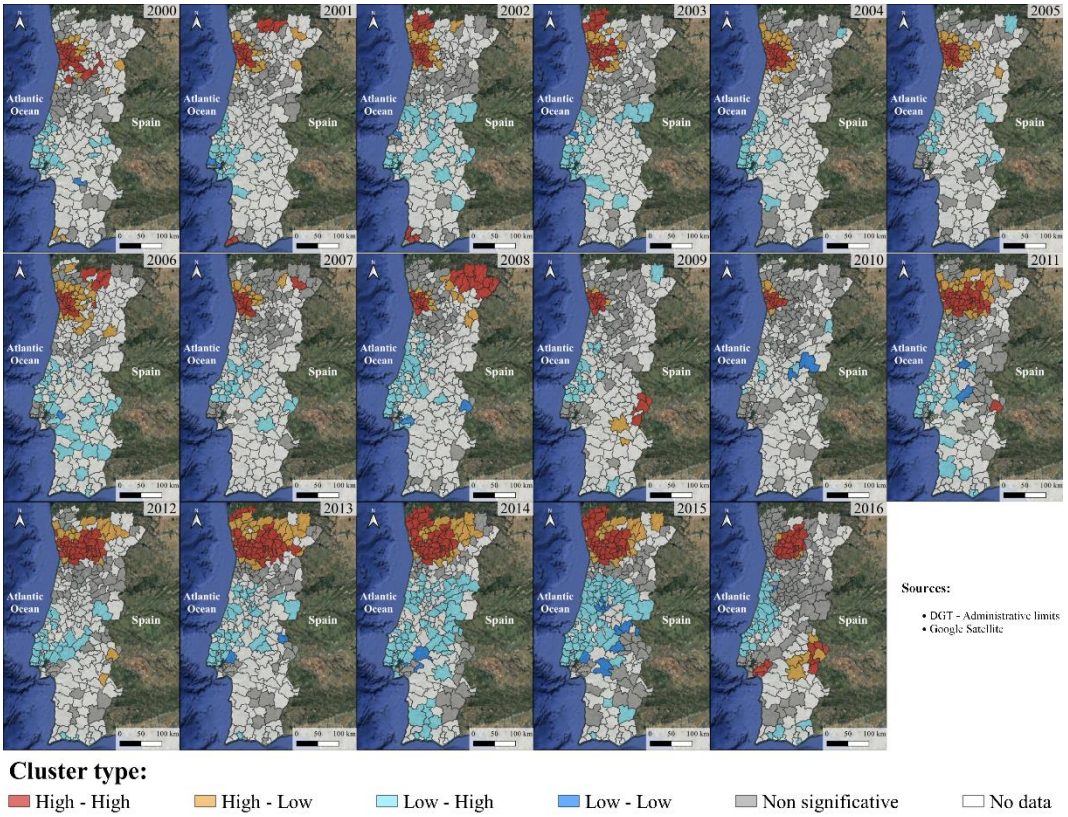


Figure C11 - Neurodegenerative diseases local clusters when considering municipalities as the administrative area (Other)

7. Spatio-temporal analysis of the environmental factors in the Iberic Peninsula between 2000 and 2022

Following the logic on the previous health descriptive papers, this paper aimed to identify spatio-temporal patterns on all environmental variables included in this thesis. Since stations were available for both Portugal and Spain, and to make the most of it with such numbers, the whole Iberian Peninsula was studied. Apart from preparing for the upcoming task of relating health and environmental variables, this paper is of particular importance as it is the only one known to have been made in Iberia regarding such an extensive collection of environmental data through such time spam.

Mariana Oliveira, Alberto Freitas, Ana Cláudia Teodoro, Hernâni Gonçalves

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
(to be submitted for publication)

Draft

-

5.5

Q1 (Computers in Earth Sciences)

Abstract: Despite the decrease in air pollution emissions throughout Europe between 2002 and 2011, air pollution concentrations are still above healthy limits and are leading to premature mortality. This work aims to describe environmental factors in the Iberian Peninsula to assess trends and spatial distribution. Air pollution data including benzene, carbon monoxide, nitrogen dioxide, particulate matter with diameter under 10 μm and 2.5 μm and sulphur dioxide was retrieved from air quality stations for Portugal mainland and Spain. Also, surface temperature, relative humidity and surface pressure were retrieved from remote sensing data. All these variables were used to describe the air quality in the Iberian Peninsula between 2000 and 2022. The monthly produced time-series exhibited the presence of outliers and missing values, as well as trends and seasonality. The time-series were processed to minimize these effects, so that other sources of variation can be identified. Spatial analysis showed a low to moderate global spatial autocorrelation (Moran's I between 0.164 and 0.316) and distinct clustering patterns were found for the different pollutants and factors. This study brings a new insight on the spatio-temporal patterns of environmental pollutants and factors in Iberia. Topographic characteristics of the peninsula may shape the dispersion of the pollutants and seasonality is also evident.

Keywords: air quality, atmospheric pollution, spatio-temporal patterns, environmental factors, Iberian Peninsula

7.1. Introduction

The concept of environment encompasses the conditions in which organisms live, and includes the air, water, temperature, wind and a variety of other factors [259]. Particularly, environmental air pollution may be defined as the presence of foreign substances in the atmosphere, either resultant of human activities or natural processes, which interfere on the wellbeing of live organisms [260].

Even though this is a growing issue around the world, levels of contamination vary from different parts of the globe [261], and while the main air pollutants' emissions in Europe have declined between 2002 and 2011, not all pollutants nor countries follow this trend [262]. Despite this decrease, air pollution concentrations are still harmful to human health, leading to premature mortality, more so in urban areas where most of the European population lives and air pollution levels still rise above healthy limits established by the European Union (EU) legislation and the World Health Organization (WHO) [262].

The purpose of this study is to provide a spatio-temporal analysis of environmental factors in the Iberian Peninsula – a region with around 54 million inhabitants in Europe – in order to support further studies

between 2000 and 202

where the environment has implications, namely in terms of health diseases triggered and/or exacerbated by chronic or acute harmful environmental conditions.

7.2. Materials and methods

Data used in this study may be split into two major categories: air pollution and environmental factors. As these categories have different data origins and processes, this section is thus divided. All processing was made using *Python*.

7.2.1. *Geographic limits data*

Geographic limits data was retrieved for both Portugal mainland and Spain, with respect to their municipalities. Portuguese data is available at General Territory Direction (DGT) [263] and Spanish data is available at the Geographic Information National Centre [264].

7.2.2. *Air pollution*

Air pollution data was obtained from QualAr platform from the Portuguese Environment Association (APA) [51], from the Portuguese Institute of Sea and Atmosphere (IPMA) [265] and from the Spanish Ministry for the Ecological Transition and the Demographic Challenge (MITECO) website [53], both on January 2022, except for 2022 data, downloaded on December 2023. This data corresponds to meteorological stations scattered throughout the countries of Portugal and Spain and represents point measurements in micrograms per cubic meter ($\mu g/m^3$) of the following pollutants: Benzene (C_6H_6), Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Ozone (O_3), Particulate Matter with diameter under 10 µm (PM_{10}) and 2.5 µm ($PM_{2.5}$) and Sulphur Dioxide (SO_2). All stations used in this study are represented in the map of Figure 47.



Figure 47 - Location of the air quality stations. In orange are Portuguese stations, and in blue Spanish stations. Stations are unevenly distributed, being located mainly in the littoral areas and concentrated in the main cities

Retrieved values for the pollutants were then processed using Python. First, both Portuguese and Spanish data were pre-processed to match style and formatting, thus allowing for the merging of both countries' datasets. Raw data had hourly measurements and while merging the datasets, monthly values were computed using the mean. Some outliers – values twice the interquartile range further away from the mean - were found and dealt with at this stage by inputting the median value for the corresponding variable, for that specific month and year. Furthermore, the data was converted from the original Excel format to shapefile, to allow for spatial interpolation.

Several interpolation methods were tested: nearest neighbour, linear, kriging (with spherical and Bayesian models) and inverse distance weighting with power 2 (standard) up to 20. Interpolation is a usual method for air pollution modelling [266]. A simple Spearman correlation coefficient was computed between each method's results and the station values, and the results are in Table 13. All inverse distance weighting (IDW) methods performed similarly and since the lower the power, the smother the output, IDW with a weighting power of 2 was used.

Table 13 - Spearman correlation coefficient for each interpolation method and the stations' value.

Interpolation method	Spearman correlation coefficient
Nearest neighbour	0.93 **
Linear	0.93 **
Spherical Kriging	0.71**
Gaussian Kriging	0.62 **
IDW (Power 2)	0.94 **
IDW (Power 3)	0.94 **
IDW (Power 4)	0.94 **
IDW (Power 5)	0.94 **
IDW (Power 10)	0.93 **
IDW (Power 15)	0.93 **
IDW (Power 20)	0.93 **

7.2.3. *Environmental factors*

Relative Humidity (RH), Surface Pressure (SP) and temperature were computed from remote sensing using ERA5 Land monthly aggregates available in Google Earth Engine (GEE) data catalogue, with a pixel size of around 11 km. Both SP and temperature were directly retrieved by the product, whilst RH was computed using the dewpoint temperature and the surface temperature and Equation 10 [54, 55]. This equation uses a frequent humidity measurement in satellites, which is the dew-point temperature (in ºC) - the temperature at which the air needs to be in order to reach saturation - and the surface temperature (also in ºC). The vapor pressure is then computed for both the dew-point temperature and the surface temperature and by dividing them the relative humidity is obtained as a decimal between 0 and 1. By multiplying by 100, this result is converted into percentage.

$$RH = \frac{6.11 \times 10^{\frac{7.5 \times Dewpoint\ temperature}{237.7 + Dewpoint\ temperature}}}{6.11 \times 10^{\frac{7.5 \times Surface\ temperature}{237.7 + Surface\ temperature}}} \times 100 \qquad \text{Equation 10}$$

Additionally, surface temperature retrieved from ERA5 was compared to land surface temperature (also known as skin or surface temperature) retrieved from Landsat, which has a pixel size of 30 meters, through cross validation for Portugal mainland, to analyse whether the differences were significant. Although these satellite measurements are not of air temperature, but of surface temperature, these are linearly related and are usually used as a substitute of air temperature for their continuous surface and broad availability [267]. For instance, in areas with dense vegetation, they represent the temperature of the canopy, while in areas of bare ground they represent the temperature of that surface [268]. Figure 48 shows the process of comparing both products to ground meteorological stations, considered to be ground truth. Pearson's correlations were initially computed, and a linear support vector regression (SVR) model was posteriorly used to predict 20% of the database, using 5 k-folds to improve model robustness. Model performance was then retrieved comparing the obtained predicted values and the real values. All the processing described in this subchapter was done recurring to Python and, in particular, to *sklearn* package.

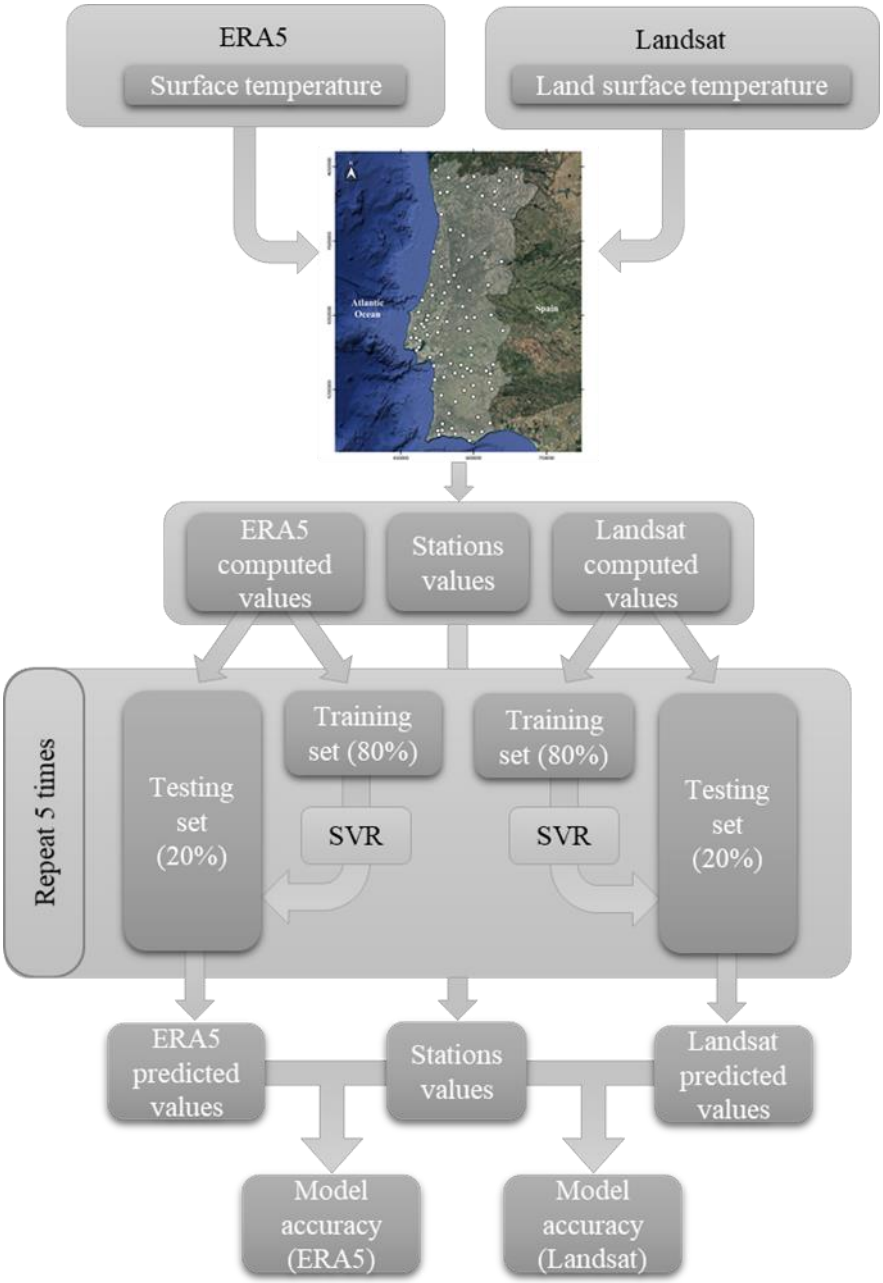


Figure 48 - Comparison between both ERA5 and Landsat surface temperatures and meteorological stations.

7.2.4. *Temporal analysis*

For temporal analysis, data was grouped by month, computing the mean of all values. Since trends and seasonality may hide variations in data, it is common practice to remove them for further analysis [269, 270]. As such, data was normalized by subtracting the mean and dividing by the standard deviation. Afterwards, trend was removed by subtracting the value of the previous observation to the current one. Seasonality was also dealt with by subtracting the month's mean to the current observation (an observation made in January would be that same observation minus the mean of all January observations). Finally, outliers (1.47% of all data) – values further away from the median than 2

interquartile ranges - were filled in with the median of the variable, obtained from a 3x3 sliding window. Outliers were considered all values below the first quartile by more than 2 interquartile ranges and above the third quartile by the same threshold. Spearman's correlation coefficient was computed between all variables.

7.2.5. *Spatial analysis*

Data was grouped by municipality, averaging values from all dates, to perform the spatial analysis, obtaining shapefiles in which each location had the corresponding average value for each environmental factor. Yearly global spatial autocorrelation was also computed using yearly grouped data, to investigate the time variation. Outliers, also corresponding to 1.47% of all data, were computed similarly to the temporal analysis and filled with the median of the respective variable, obtained from 3x3 sliding window. Afterwards, global and local spatial autocorrelation were computed to identify the presence of clustering using Moran's I [250] and LISA (Local Indicators of Spatial Association) [251]. This was done using Python package *esda*, which allows for the imputation of a geodataframe containing the geometries of each municipality, as well as the values of each environmental factor, to firstly compute a weights matrix, and then compute both the Moran's I and the LISA. The weights matrix used was distance based with a distance band of 150 kilometres. This allows for values close to each other to have higher relevance in the computation, while still allowing for whole districts to be contained within the considered distance.

7.3. Results

7.3.1. *Validation of environmental factors measurements: ERA5 and Landsat surface temperature*

All results of this section are represented in Table 14. ERA5 retrieved temperatures had a Pearson's correlation coefficient of 0.96 in relation to the meteorological station's values. On the other hand, Landsat data obtained a Pearson's correlation coefficient of 0.86 (Figure 49). Both these values were retrieved without excluding outliers, and although Landsat results slightly improved when outliers were removed (Pearson's R = 0.88), those of ERA5 did not change.

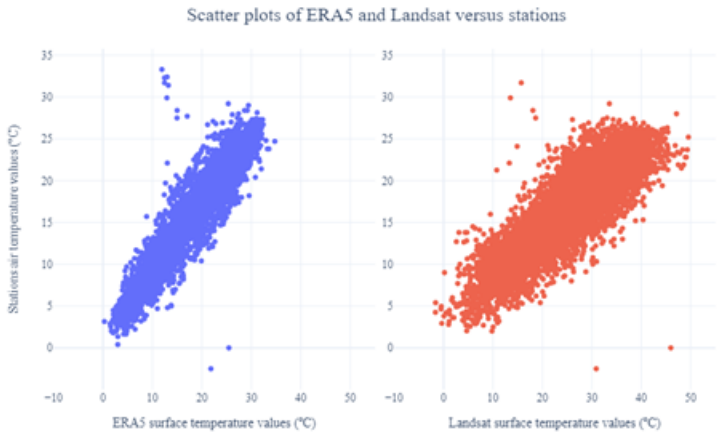


Figure 49 - Scatter plots of computed surface temperature from ERA5 and Landsat and air temperature from meteorological stations.

When analysing the Bland-Altman plots for each remote sensing product and the stations values (Figure 50), ERA5 mean difference was 2.02 with an interquartile interval of -2.19 to +6.23, while Landsat mean difference was 7.75 with an interquartile interval of -3.20 to 18.71. Landsat showed a larger bias, especially with higher values.

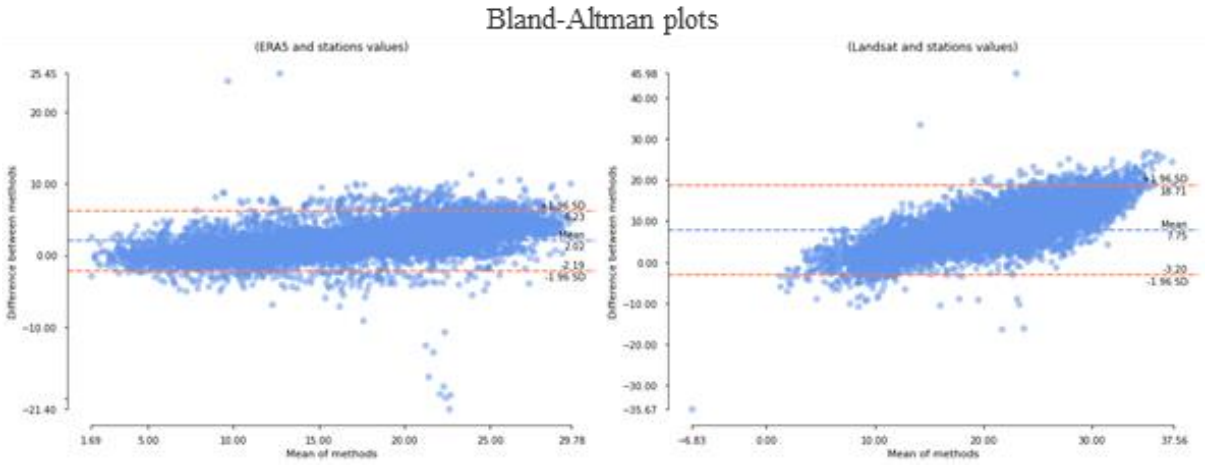


Figure 50 - Bland-Altman plots for each remote sensing product and station values.

From the linear support vector regression model, the accuracy of the Landsat was 0.75 with a standard deviation of 0.05 while ERA5 performed better, with an accuracy of 0.91 and a standard deviation of 0.05.

Table 14 - Comparison of the performance of Landsat and ERA5 on the measurements between Landsat and ERA5of temperature, with the meteorological stations as reference. Further description of the methodology in 2.3 (Std is the standard-deviation of the predicted values when compared to the measured values).

		Pearson's R		Linear SVR	
		R	**p-value**	**Accuracy**	**Std**
Landsat	**With outliers**	0.86	<0.001	0.74	0.05
	Without outliers	0.88	<0.001		
ERA5	**With outliers**	0.96	<0.001	0.91	0.05
	Without outliers	0.96	<0.001		

7.3.2. *Temporal analysis*

Firstly, temporal data was normalized so data can be represented with the mean around 0 and comparable, once scales are therefore more similar (Figure 51).

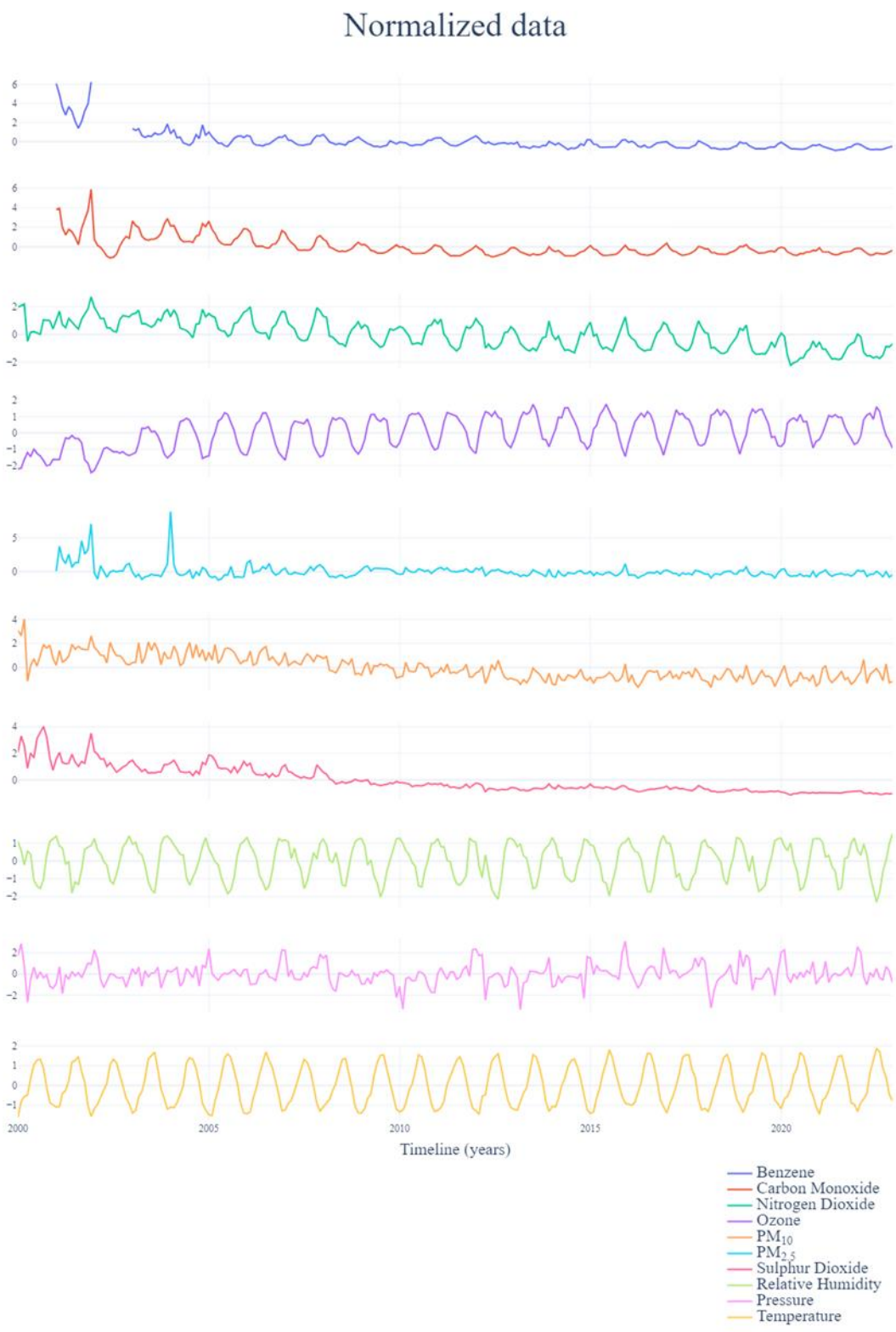


Figure 51 - Normalized temporal data for each environmental pollutant and factor.

A trend was evident in some variables, especially a decreasing trend in C_6H_6, CO, NO_2, PM_{10}, and SO_2 and an increasing trend in O_3, which led to performing differencing to remove these trends (Figure 52).

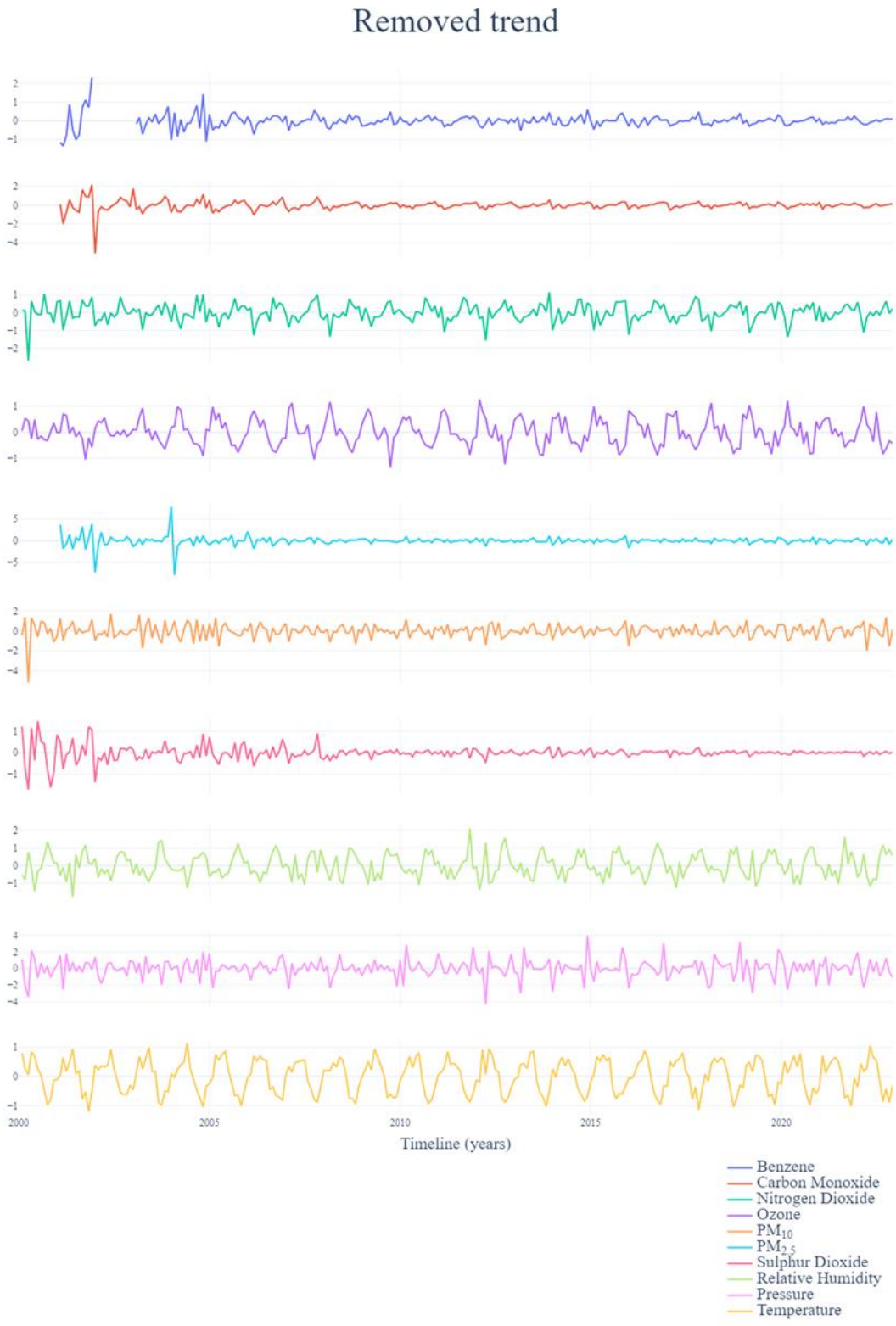


Figure 52 - Temporal data after trend was removed through differencing for all environmental variables and factors.

Seasonality was also obvious throughout most of the time series, and since it could mask other variations, it was removed through subtracting the month's mean to the current observation (Figure 53). This resulted in a more interpretable dataset, with visible peaks that are now not attributable to either seasonality nor trends, and may be due to outliers or natural variations in data.

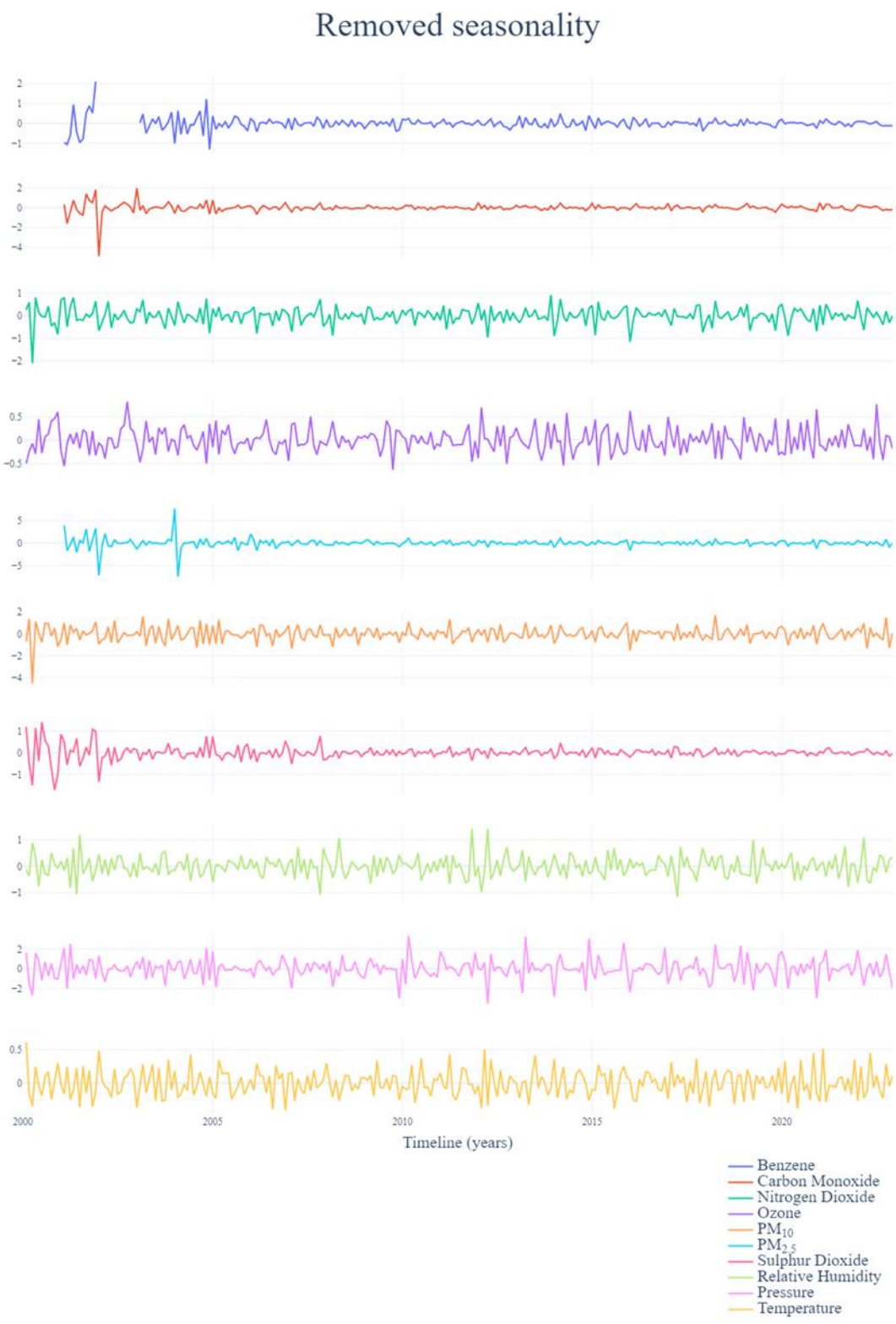


Figure 53 - Temporal data after seasonality was removed through differencing for all environmental variables and factors.

Outliers were computed to further clean the data and both these and missing values were replaced with the median of the variable in question. This process filled in most of the inconsistencies found towards the beginning of the time series (Figure 54).

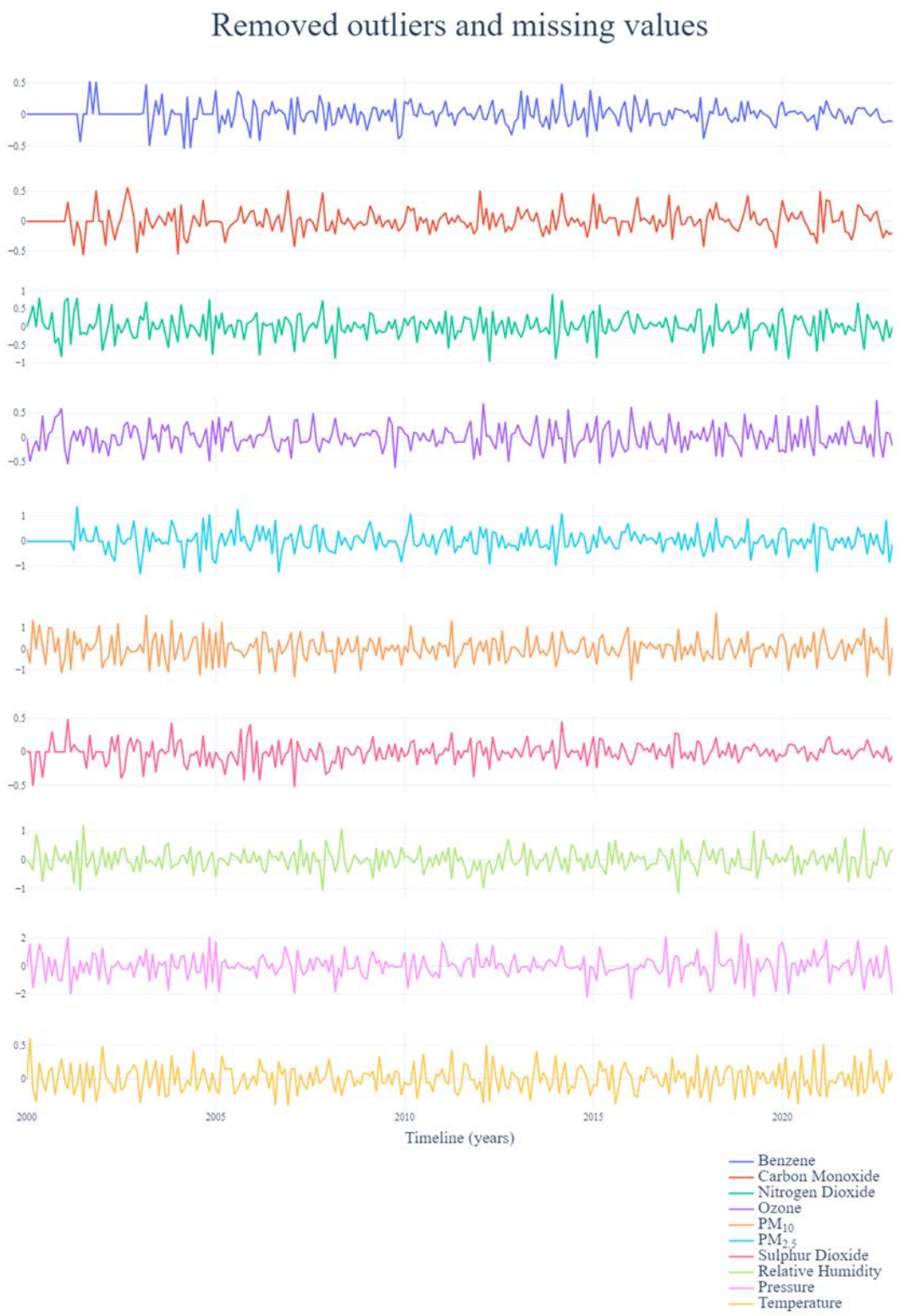


Figure 54 - Temporal data after all processing to remove trend, seasonality, outliers and missing values for all environmental variables and factors.

Furthermore, Spearman's correlation coefficient between environmental factors was computed (Table 15). As expected, most air pollutants were correlated, with the exception of O_3, which was not correlated

to PM_{10}, $PM_{2.5}$ and SO_2. Temperature was not significantly correlated to C_6H_6, CO, O_3 and SO_2. Pressure and humidity, on the other hand, were significantly correlated to all other variables.

Table 15 – Spearman's R for each environmental factor/pollutant in time series (** - significant at 0.01; * - significant at 0.05). Colours represent the strength of the correlation: greens are positive correlations and oranges negative correlations, with darker colours meaning higher correlations (RH: Relative Humidity; SP: Surface Pressure; ST: Surface Temperature)

	C_6H_6	CO	NO_2	O_3	PM_{10}	$PM_{2.5}$	SO_2	RH	SP
CO	0.54**								
NO_2	0.44**	0.48**							
O_3	-0.27**	-0.17**	-0.29**						
PM_{10}	0.32**	0.34**	0.58**	-0.10					
$PM_{2.5}$	0.33**	0.37**	0.47**	-0.07	0.60**				
SO_2	0.37**	0.50**	0.46**	-0.07	0.42**	0.35**			
RH	-0.22**	-0.17**	-0.29**	-0.15*	-0.46**	-0.18**	-0.33**		
SP	0.18**	0.25**	0.39**	-0.14*	0.45**	0.30**	0.28**	-0.34**	
ST	0.03	-0.01	0.11*	-0.04	0.40**	0.27**	0.05	-0.25**	0.30**

7.3.3. *Spatial analysis*

Spatial autocorrelation was studied using both global and local methods. Bellow, in Figure 55, are the results of the global Moran's autocorrelation index. All environmental pollutants and factors had, overall, significant spatial autocorrelation, with the highest being C_6H_6, with an index of 0.316, and the lowest being $PM_{2.5}$ with and index of 0.164. Individual years also showed similar spatial autocorrelations, increasing through the years. As in the overall Moran's I, the yearly value was highest for C_6H_6, reaching 0.362, and is lower for both $PM_{2.5}$ and PM_{10}, both with values as low as 0.021. Pressure kept the most stable Moran's I throughout the years, staying between 0.187 and 0.189 at all times.

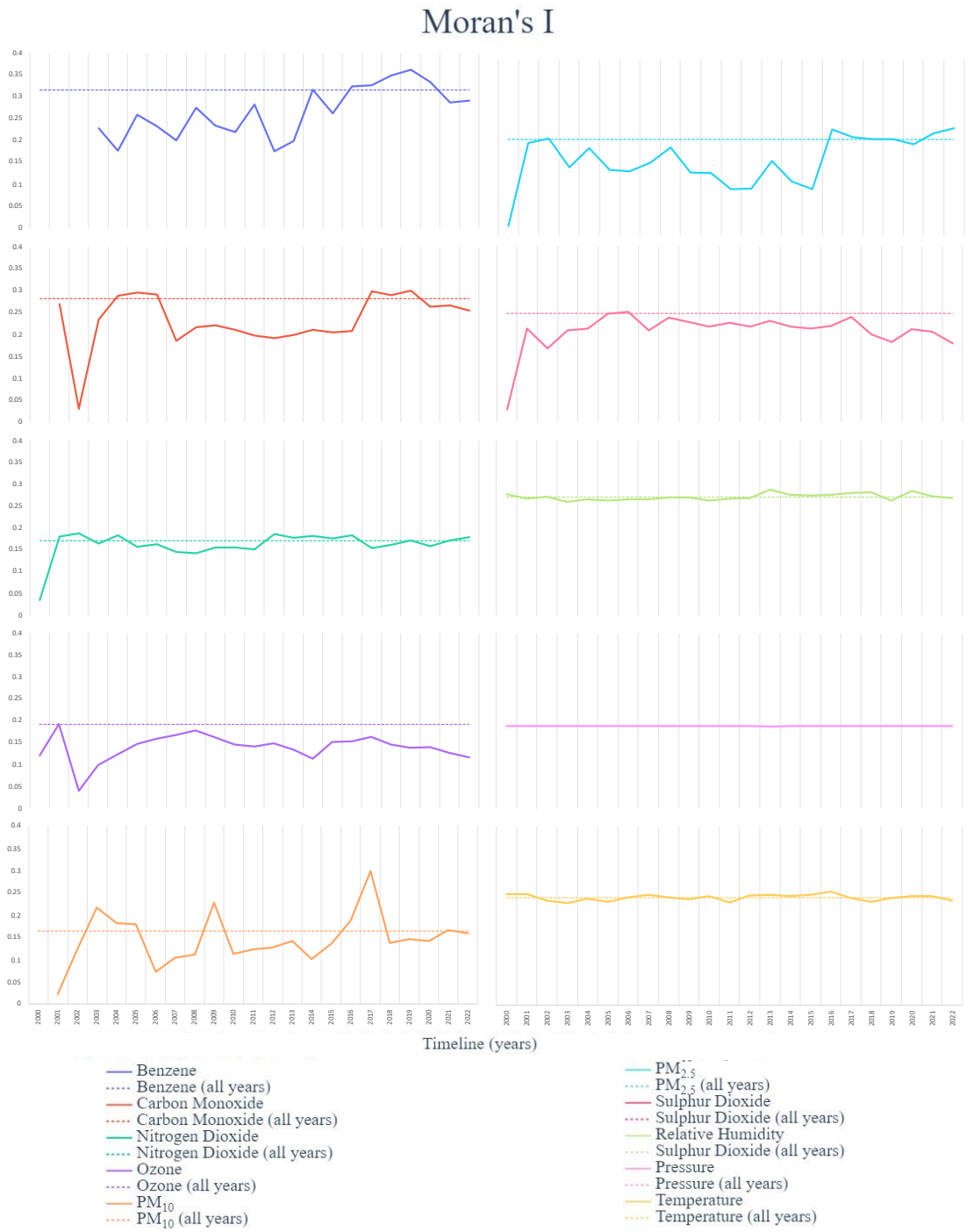


Figure 55 - Yearly Moran's I for each pollutant (with all years' value represented in dashed line).

Local spatial autocorrelation was also computed and mapped (Figure 56). Several interesting patterns were found when looking at the clusters. For instance, both surface temperature and O_3 seems to split the Iberic Peninsula in half, with the north having mostly low-low clusters and the south high-high clusters. However, O_3 had a more heterogenic pattern. CO and NO_2 had most of the east and west low-low clusters, while the centre Iberic Peninsula was mostly high-high clustered. Both PM_{10} and $PM_{2.5}$ also had high-high clusters around the centre and south of the Iberic Peninsula, and low-low clusters in the north, east and west. C_6H_6 had most of high-high clusters on the west, with most of the Iberic

Peninsula having low-low clusters, except for areas around bigger cities such as Porto, Lisbon, Madrid and Leon. Relative humidity seemed to follow a mirrored pattern to that of temperature, with high-high clusters in the north instead and the line dividing low-low clusters from high-high clusters being further north. Littoral areas were also high humidity when compared to those around, which was expected.

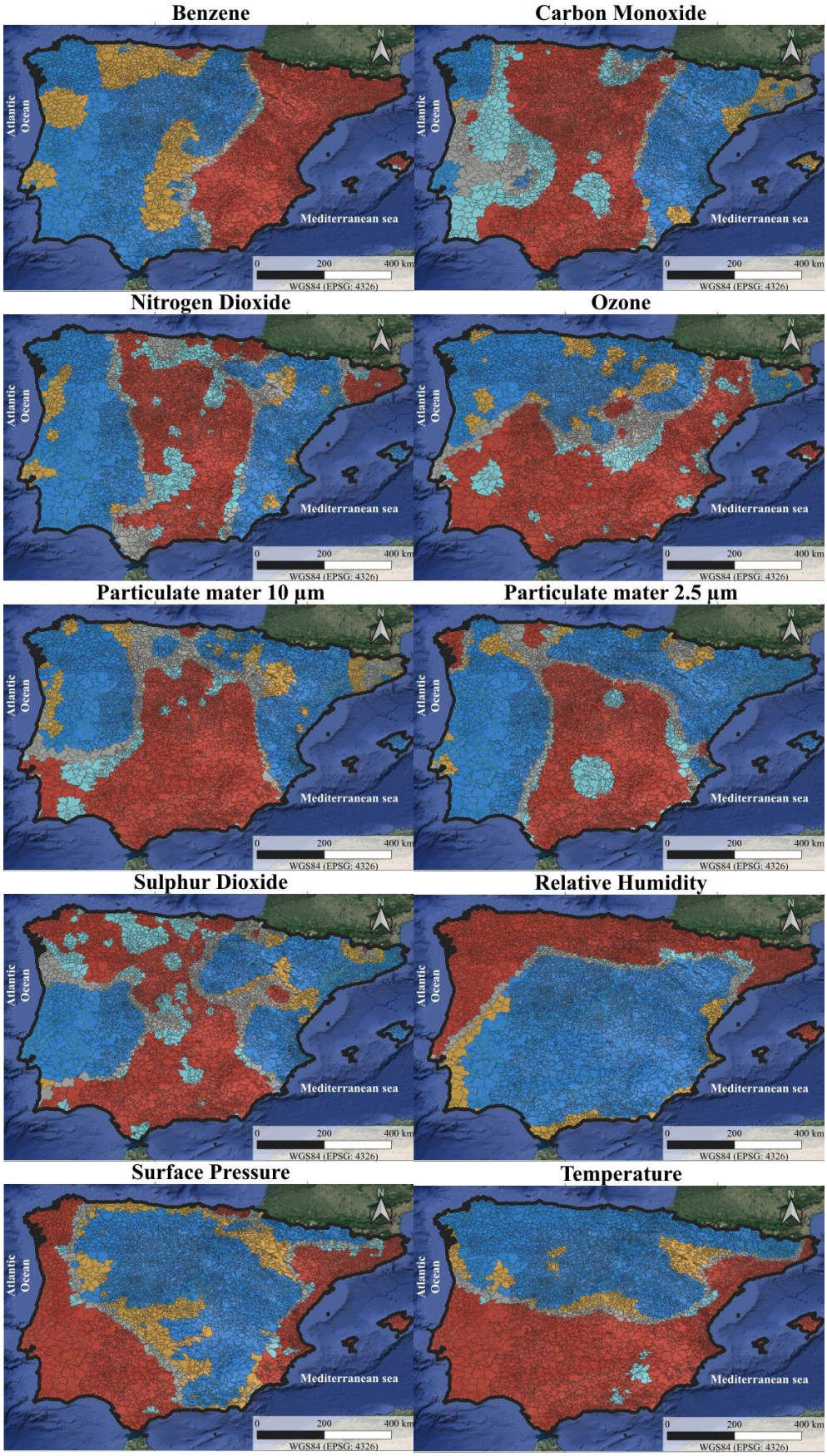


Figure 56 - Local spatial autocorrelation clusters for each environmental factor using all data and each year data (Red: high-high; Orange: high-low; Light blue: low-high, Dark blue: low-low).

7.3.4. *Spatio-temporal analysis*

Spatio-temporally, pollutants showed an overall decreasing trend (figures D1 through D7). Pollutants' distribution is very spotty, probably related to the meteorological stations positioning. Benzene (Figure D1) and nitrogen dioxide (Figure D3) and ozone (Figure D4) display seasonality in their distribution, although inversely so. Benzene thus has lower values in the summer, whilst ozone has higher values in the winter. Both surface temperature (Figure D10) and humidity (Figure D8) also present strong seasonality and are clearly inversely related, as expected. Surface pressure does not show visible changes throughout time (Figure D9). All maps representing the spatio-temporal data are available in the appendix D.

7.4. Discussion

Surface temperature data from ERA5 proved more accurate than that from Landsat. This is probably related to the extensive validation of algorithms applied in ERA5 and other satellite products, whilst Landsat's algorithm was simple. For instance, emissivity used in Landsat's algorithm was retrieved from a global emissivity dataset with no time stamp. However, Landsat's surface temperature algorithm may be more suitable when analysing smaller areas, since pixel size is substantially smaller than that from ERA5 (30 meters and 11 km, respectively). Taking into account the geographic extent of this work, ERA5 was preferred over Landsat. Also, the differences between remote sensing obtained data and meteorological stations data may be explained by the fact that they measure slightly different things: surface temperature and air temperature, respectively. Although linearly related, these are two different measurements.

All environmental data was then analysed in both temporal and spatial components separately. Temporal analysis showed clear trends, mostly decreasing, and seasonality that accompanied the seasons. Initial years also presented some instability related to the amount of missing data, which is due to less meteorological stations in early years.

Spatial analysis showed significant global spatial autocorrelation across all pollutants and factors and interesting patterns when analysing local spatial autocorrelation. High-low clusters were found in bigger cities across all Iberia and topography may be responsible for the remaining patterns observed. The Central System mountains northwest of Madrid seem to be the natural barrier to pollutants and, together with the Iberian System mountains, appear to regulate both temperature and relative humidity.

Spatio-temporally, C_6H_6 and O_3 showed seasonality. Since high-temperature and low humidity conditions are propulsive to the formation of O_3 [271], O_3 seasonality may be thus explained. Benzene also reacts to form O_3 in the atmosphere, thus explaining why it decreases while O_3 increases [272].

Overall, pollutants concentrations have decreased throughout the year, which is in line with other studies [262, 273].

This work also has some limitations. Firstly, meteorological stations are poorly distributed and don't have measurements of every pollutant for every month, leading to missing values and large spatial gaps. On the other hand, although the remote sensing products used provide continuous (both in time and space) measurement of temperature, humidity and pressure, the spatial resolution is low and do not measure pollutants. Also, the measured factors from the remote sensing products differ from the factors measured by meteorological stations, even though they are closely related.

Nevertheless, this work provides a framework for future works intending to rely on environmental factors in the Iberian Peninsula. This data can be used to assess, for example, a possible relation to increased or decreased affluence to healthcare services with respect to a particular disease or set of diseases.

7.5. Conclusion

This study brings a spatio-temporal analysis of some environmental air pollutants and factors. By comparing different methods for retrieving temperature, the care needed when picking the origin of data is clear, taking into consideration both the time and spatial granularity needed for each study case.

Environmental data presents high seasonality, which needs care when relating it to other variables. Users need to interpret the results and analyse whether to remove seasonality. For example, a user aiming to investigate trends in environmental factors may want to remove seasonality to better visualize and measure that trend, whilst an investigator aiming to relate environmental data to a particular health context may want to know if extreme values relate to extreme health reactions, thus benefiting from keeping seasonality. Both these options are interesting approaches using this data. Methodologies here explored are also of interest to either extend this work into other countries, or use it to compute other variables, such as nanoparticles and other environmental factors of interest to a particular research question.

Local topography appears to be related to the spatial dispersion of the environmental pollutants and factors.

Whilst environmental data is broadly available nowadays, no studies as this one are known to the authors, thus providing new insight on the topic.

Appendix D

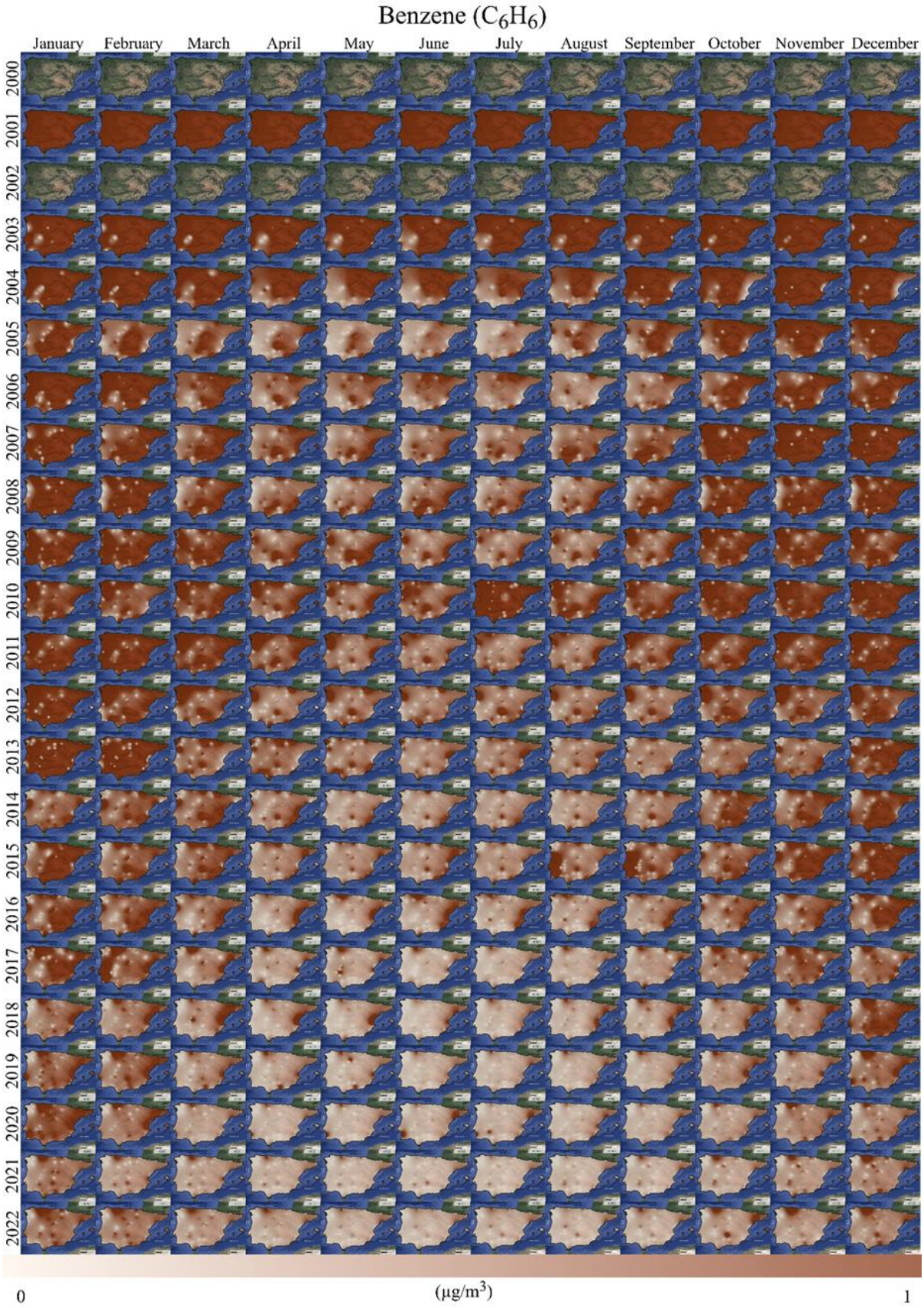


Figure D1 - Benzene spatio-temporal distribution in the Iberian Peninsula.

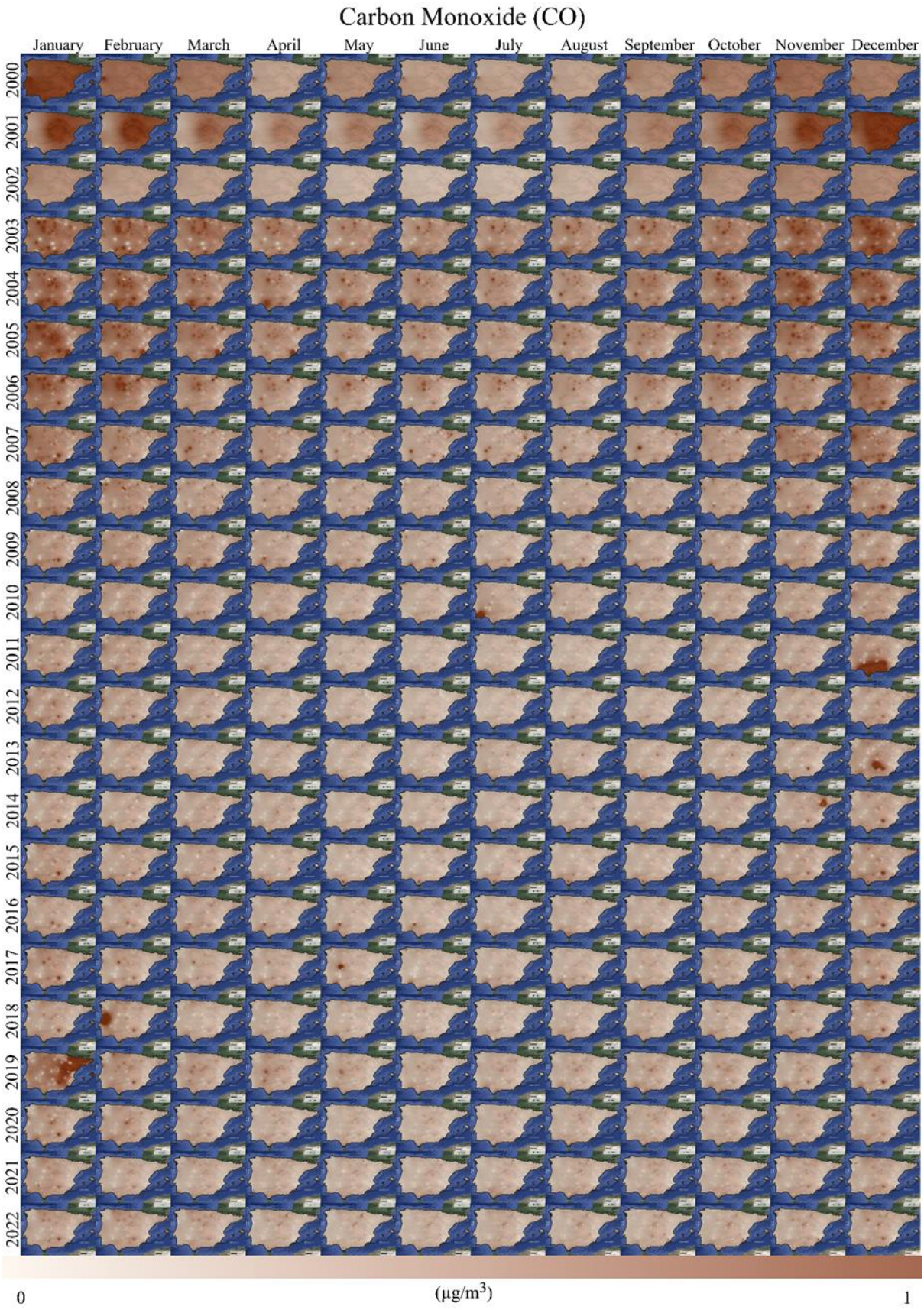


Figure D2 - Carbon monoxide spatio-temporal distribution in the Iberian Peninsula.

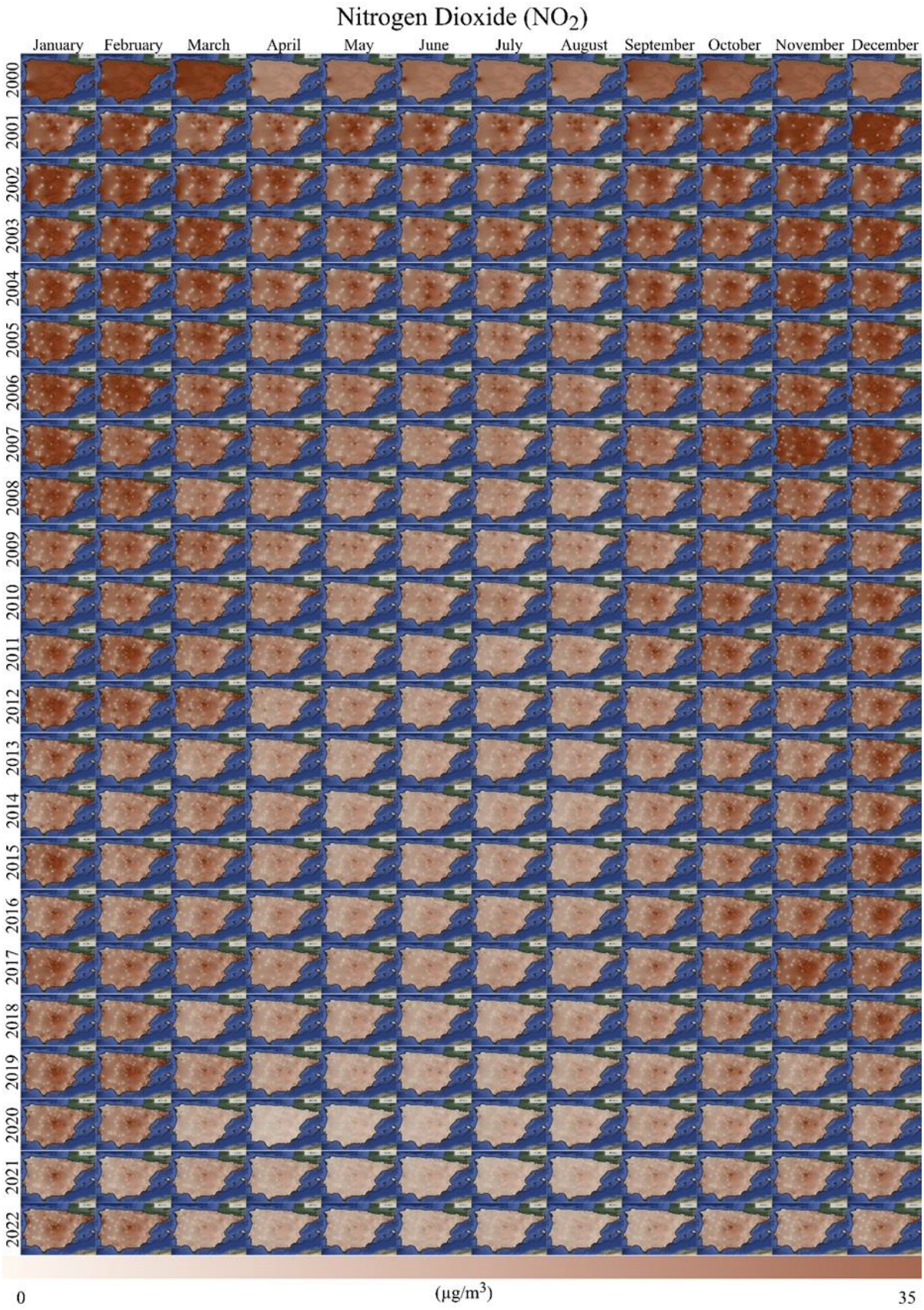


Figure D3 - Nitrogen dioxide spatio-temporal distribution in the Iberian Peninsula.

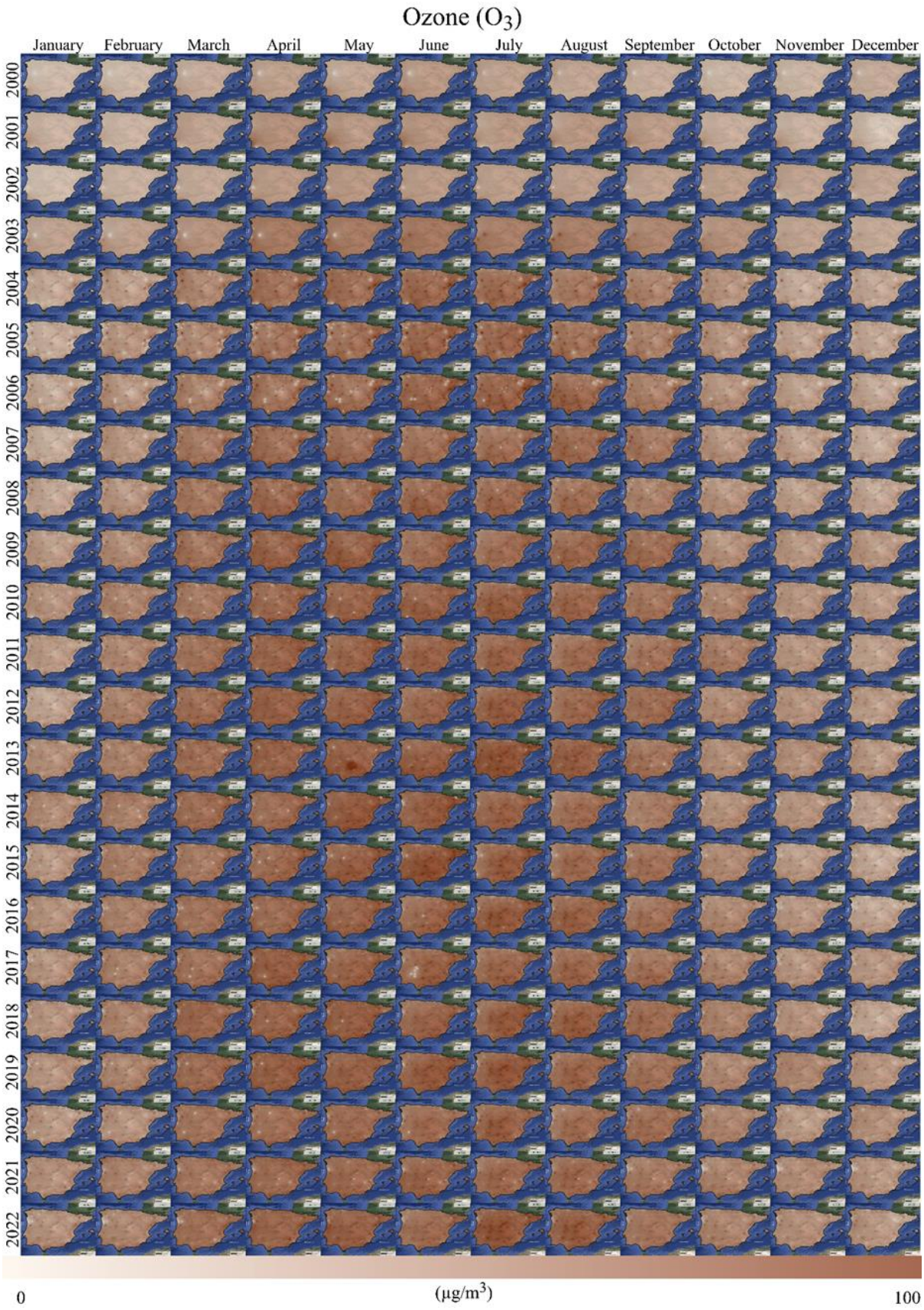


Figure D4 - Ozone spatio-temporal distribution in the Iberian Peninsula.

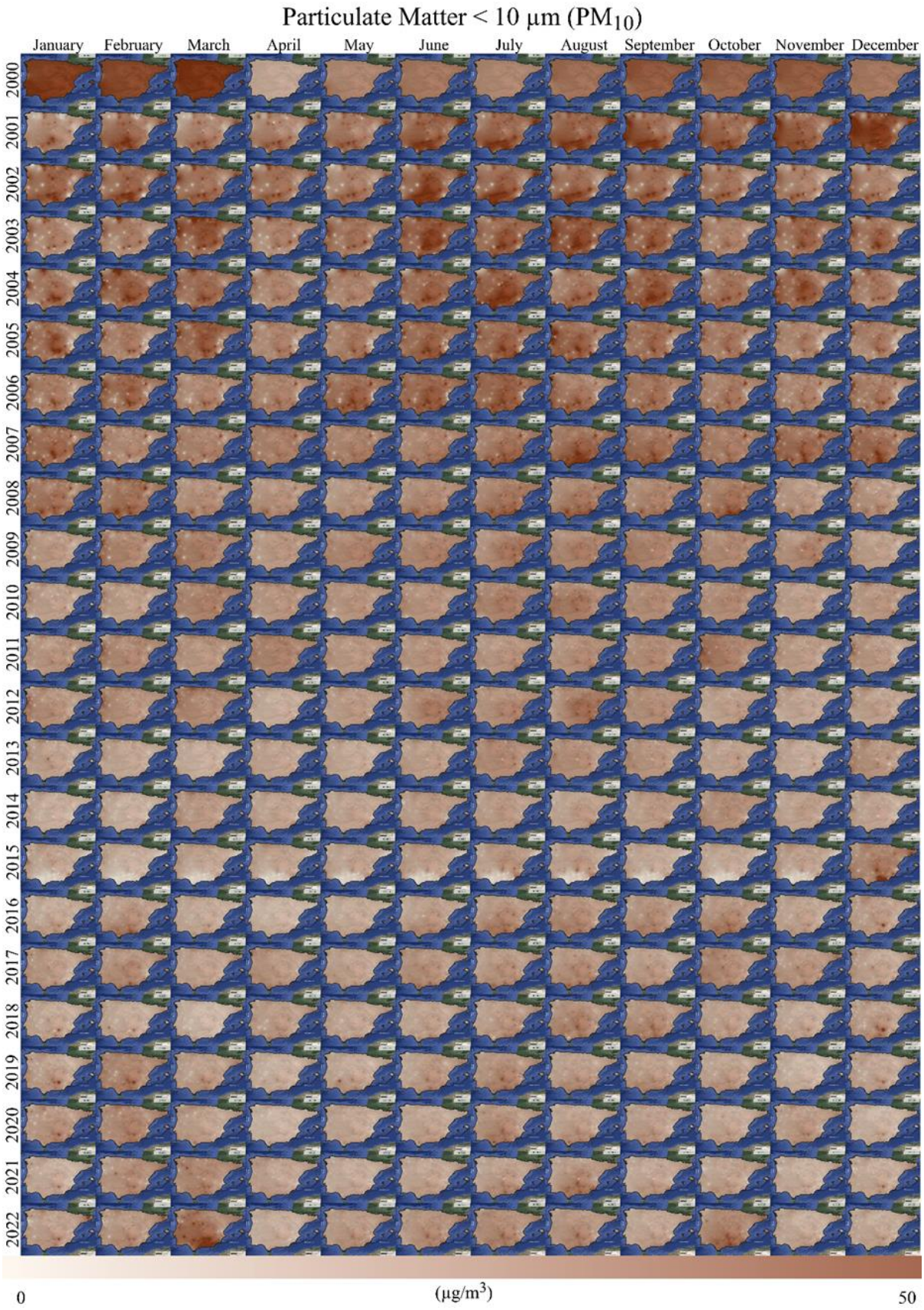


Figure D5 - Particulate matter < 10 µm spatio-temporal distribution in the Iberian Peninsula.

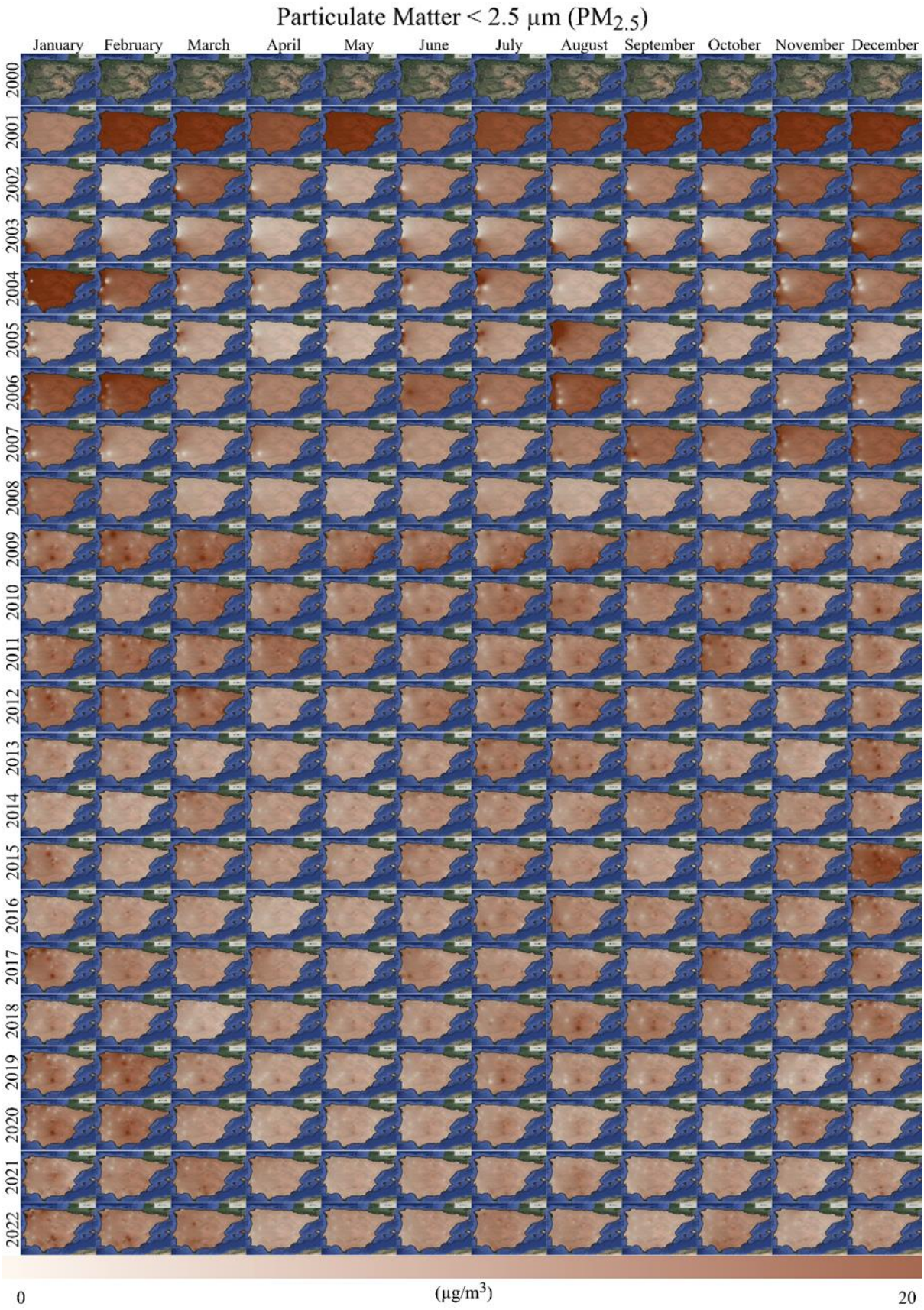


Figure D6 - Particulate matter < 2.5 µm spatio-temporal distribution in the Iberian Peninsula.

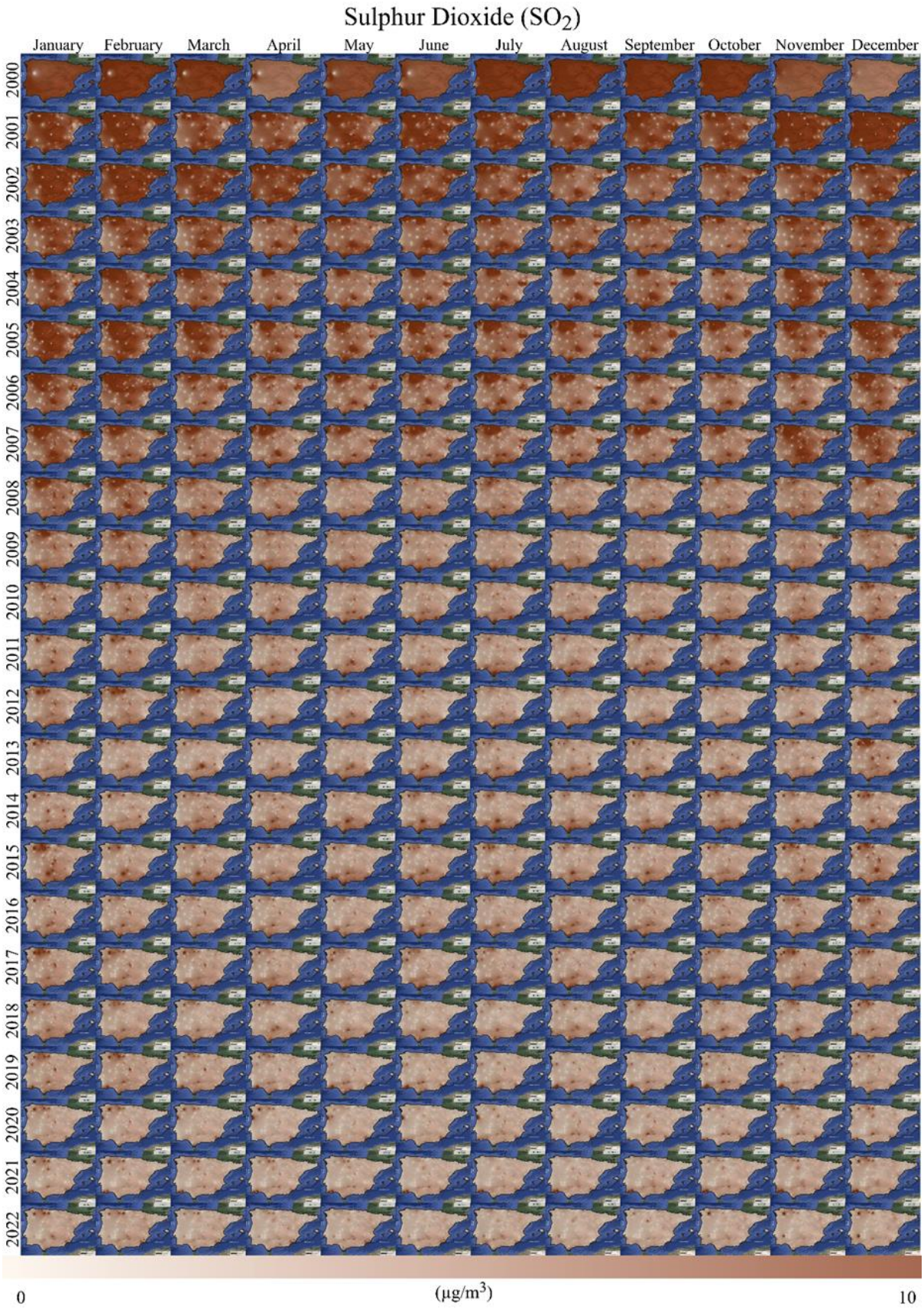


Figure D7 - Sulphur dioxide spatio-temporal distribution in the Iberian Peninsula.

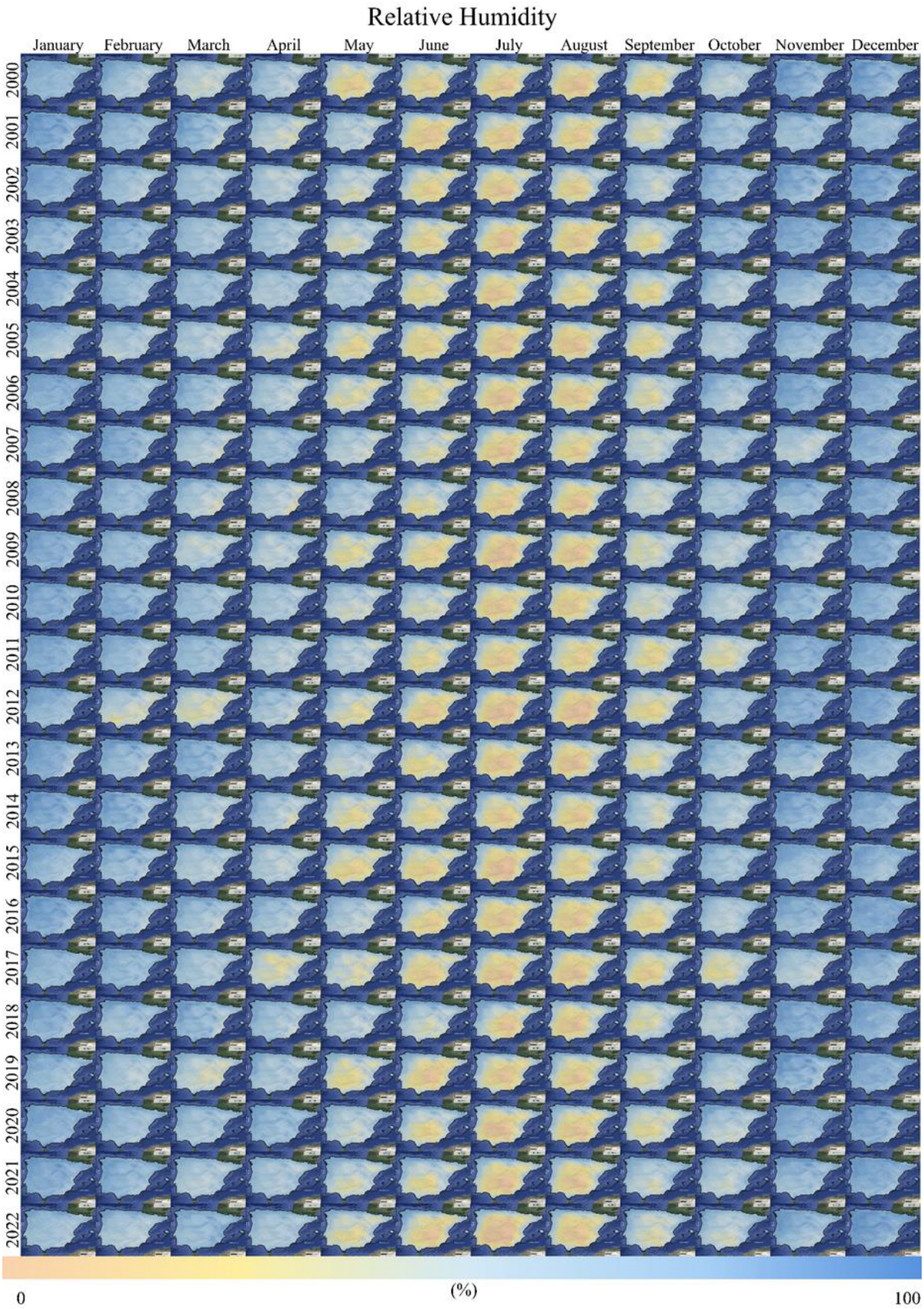


Figure D8 – Relative humidity spatio-temporal distribution in the Iberian Peninsula.

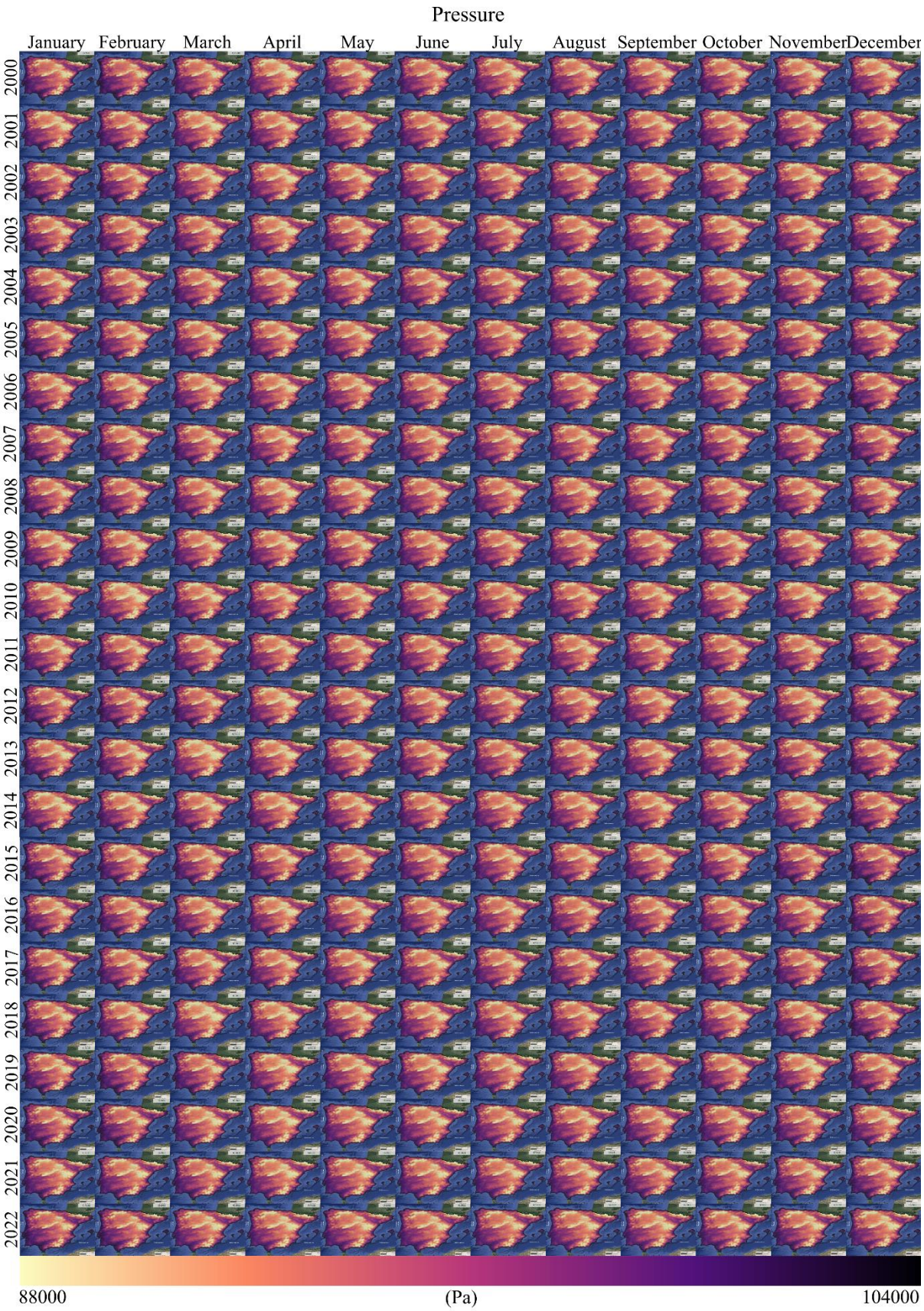


Figure D9 - Surface pressure spatio-temporal distribution in the Iberian Peninsula.

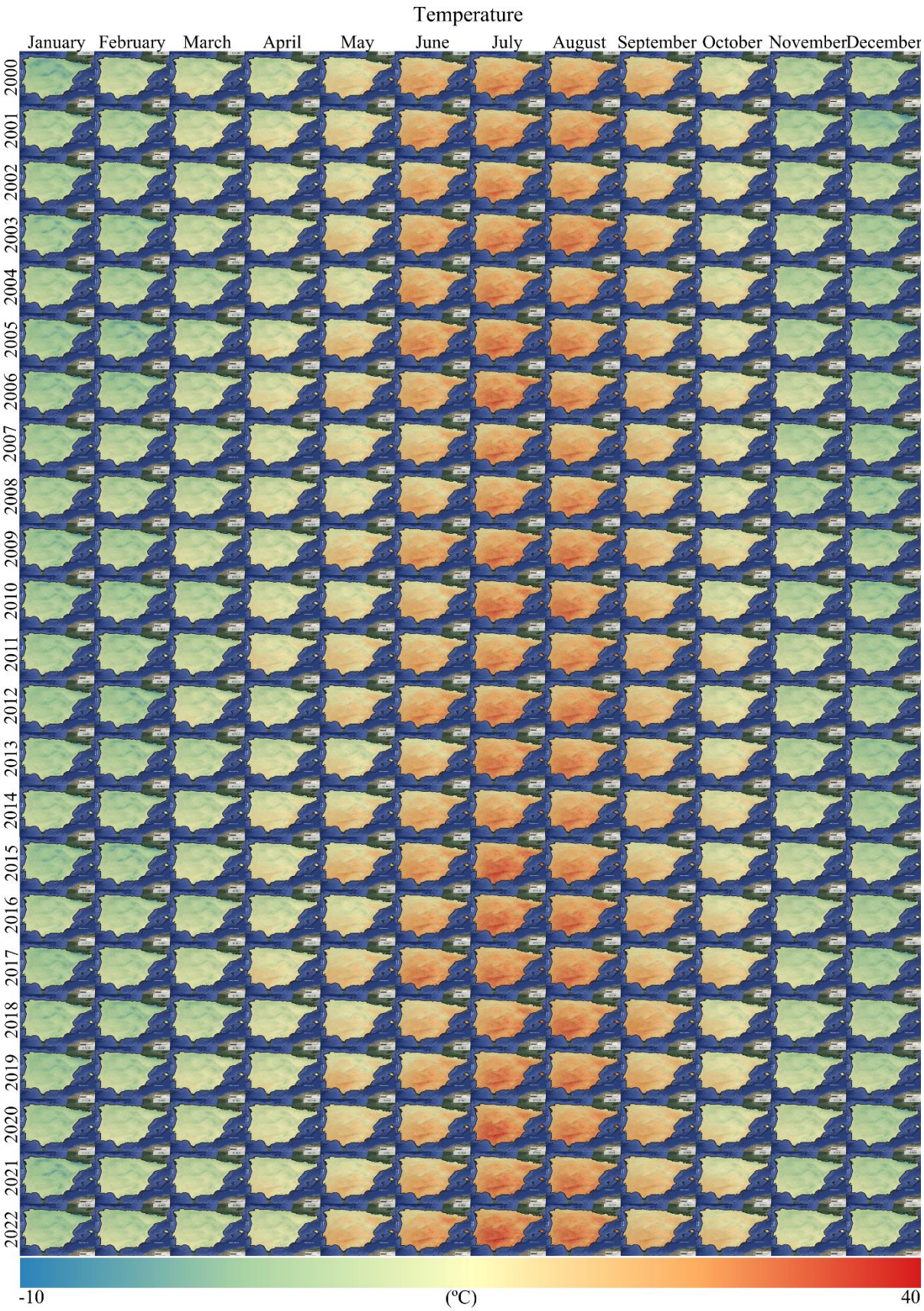


Figure D10 - Temperature spatio-temporal distribution in the Iberian Peninsula.

8. Environmental risk factors for neurodegenerative diseases hospitalizations

The culmination of all previous work is the assessment of environmental risk factors for neurodegenerative diseases. In this chapter, the health data used in chapters 6 and 7 was joined to that of the environmental factors in chapter 8 to create a dataframe in which each patient had the corresponding environmental factors levels at the time of hospitalization attributed according to their residence.

In order to make this work relevant in clinical practice, a risk tool was planned in which doctors could insert patient's data to compute a risk index of that particular patient developing a neurodegenerative disease. This tool is website based, and this is a prototype, which can be accessed through the URL https://marianaramosolivei1.wixsite.com/neurodegenerative-di. and the second part of this chapter intends to give an overview of the website designed to accommodate the tool.

Mariana Oliveira, Alberto Freitas, Ana Cláudia Teodoro, Hernâni Gonçalves

To be defined

Draft

To be defined

To be defined

To be defined

8.1. Environmental risk factors for neurodegenerative diseases hospitalizations

Abstract: Risk factors for neurodegenerative diseases have been studied for years and include some well-established risk factors such as sex, age, and some air pollutants. This study aims to attempt to relate several environmental air factors (C_6H_6, CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, SO_2, relative humidity, surface pressure and temperature) with neurodegenerative diseases' hospitalizations. Pollutants data from in situ stations, environmental factors data from remote sensing and health data from hospitalizations were merged into a dataset in which each patient had been exposed to each pollutant and environmental factor at the time of the hospitalization attributed to it according to its residence. Age-standardized hospitalization rates were considered in regression models (linear, decision tree, support vector machine and random forest), which led to low accuracies, with exception of the random forest. In conclusion, the environmental variables did not exhibit a strong relation with the studied health variables by themselves, thus suggesting the need for the exploring of alternative approaches. Primary care health data, with covariables known as risk factors for developing neurodegenerative diseases, which could be used instead of hospitalization data. Additionally, environmental exposure could also be refined using locations where the patients effectively spend more time. This work sets up a framework and identifies limitations in this type of epidemiological research.

Keywords: Neurodegenerative diseases, environmental risk factors, regression models.

8.1.1. Introduction

The aging process induces pathological changes in the human brain, leading to various neurodegenerative disorders [61]. These conditions are characterized by the gradual deterioration or loss of neuronal cells, endangering the functionality of the brain, with potential risks to either motor or cognitive functions [64]. Besides aging, risk factors contributing to the onset of neurodegenerative diseases encompass gender, a clinical history of hypertension, diabetes, cranial injury, smoking and drinking [68]. However, comprehensive understanding of the development of many of these diseases, including the potential impact of environmental factors, remains incomplete [68].

Emerging research indicates that atmospheric pollution can activate mechanisms associated with neurodegenerative diseases, particularly emphasizing the role of particulate matter with a diameter under 2.5 μm ($PM_{2.5}$) due to its ability to penetrate in the brain [72]. $PM_{2.5}$ pollutant is linked to Alzheimer's disease (AD) [185, 198, 200, 201, 204-210, 214, 215], Dementia [92], Multiple Sclerosis [167, 222, 224], Paediatric Multiple Sclerosis [175] and Parkinson's disease (PD) [202, 210, 228].

Additionally, other pollutants such as ozone (O_3), particulate matter with a diameter under 10 μm (PM_{10}), nitrogen dioxide (NO_2), carbon monoxide (CO), and sulphur dioxide (SO_2) are toxic to the

central nervous system (CNS) [17] and were related to Dementia, including AD [213], and PD [210] [202] [181, 227, 228].

Notably, no studies were found regarding basal ganglia Dementia, Creutzfeldt, Huntington's disease, Lewy body Dementia, or Pick's disease, although one study suggests that environmental factors may be related [274].

This study aims analyse the relation between several environmental air pollutants (C_6H_6, CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, SO_2) and some environmental factors (relative humidity, surface pressure and temperature) with neurodegenerative diseases' hospitalizations. From our knowledge, this is the first study encompassing such a large variety of environmental and health variables.

8.1.2. Materials and methods

8.1.2.1. *Health data*

Health data was used as a global database, with age-standardized rates computed for each location and each month of the time period, with each line representing a municipality encompassing the period the values refer to and the respective age-standardized hospitalization rate. This was used in regression models to try to relate risk factors with the development of neurodegenerative diseases.

Hospitalizations with neurodegenerative diseases may be coded as primary and/or secondary diagnosis. The algorithms were applied separately to all coded diagnosis (primary and/or secondary diagnosis) and to cases of only primary diagnosis.

Details on the origin and pre-processing of the data may be found in previous works [253]. In short, hospitalizations data with discharges between 2000 to 2016 were included. For this particular study, admission date was considered instead of discharge date, since this would be the date the patient would have been more closely exposed to the pollutants and environmental factors. Additionally, since diseases were coded according to both ICD-9-CM and ICD-10-CM, they were recoded as the disease they represented (for example, ICD-9 code 331.0 and ICD-10 code G30 were both recoded as Alzheimer).

Health data included date of admission (split into year and month for posterior convenience), age (in years), length of stay (in days) and previously computed age-standardized hospitalization rates. Although the date of admission was available at the day level, we used the month and year in the remaining of the analysis. Health data only contains patients with neurodegenerative disease, incapacitating the use of other diseases as possible differentiating outcomes.

8.1.2.2. Environmental data

Environmental data was retrieved from public sources such as QualAr [51], IPMA (Portuguese Institute of Sea and Atmosphere) [52] and SNIRH (National Information System for hydric resources) [56] for Portugal. Spanish data was also used to allow for more accurate interpolation towards the boarders, and it was available at MITECO [53]. These entities provided with air pollutants data from meteorological stations (Figure 57), which were interpolated using inverse distance weighting (IDW) algorithm with a weighting power of 2 and then summarized by municipality using the mean and standard deviation. Although data was available hourly, it was resampled to monthly values using the average.



Figure 57 - Location of the air quality stations. In orange are Portuguese stations, and in blue Spanish stations.

Further environmental factors were retrieved through remote sensing data using ERA5 product available from Google Earth Engine (GEE), with a pixel size of over 11 km and monthly aggregated rasters [29]. This product has surface temperature and pressure ready-to-use and relative humidity was computed using the dewpoint temperature and the temperature according to Equation 11 [54, 55]. This equation uses dew-point temperature (in ºC) and surface temperature (also in ºC) to compute vapor pressure and saturated vapor pressure. By dividing both vapor pressure results, a decimal between 0 and 1 is obtained representing relative humidity and multiplying it by 100, it is converted into percentage.

$$RH = \frac{6.11 \times 10^{\frac{7.5 \times Dewpoint\ temperature}{237.7 + Dewpoint\ temperature}}}{6.11 \times 10^{\frac{7.5 \times Surface\ temperature}{237.7 + Surface\ temperature}}} \times 100 \quad \text{Equation 11}$$

8.1.2.3. *Regression algorithms*

For initial assessment of the relation between variables, and since data was not normally distributed, spearman correlation coefficients were computed using "*scipy stats"* module.

Regression algorithms were used with the age-standardized hospitalizations rates database and the environmental database. These methods attempt to model the hospitalization rates regarding the environmental risk factors. The methods used were: Linear regression: this is an extremely common statistical tool used to search for a linear regression between a continuous dependent variable which is the outcome and the independent variables [38]; Decision tree regression: this algorithm defines a tree of decision rules that model the values of the input variables, aiming to find the best tree to determine the output by measuring, at each step, which variable combination rely the lowest error to the decision [39]; Support vector machine regression: this method is very similar to its classification algorithm, except it allows for a continuous outcome to be modelled through the application of function estimates [40]; Random forest regression: similarly to the random forest classifier, this method used several random decision trees to model the outcome [41].

All regression algorithms were computed using Python's package "*sklearn"* and the default parameters for each method. To evaluate the models, the dataset was split into training and testing sets with 80% and 20% of data, respectively. The models were computed using the training set and then used to predict the remaining test set. Models were evaluated using the pre-defined score method from "*sklearn"*, which computes the coefficient of determination R^2 using Equation 12.

$$R^2 = \left(1 - \frac{\sum(true_value - predicted_value)^2}{\sum(true_value - \overline{true_value})^2}\right) \quad \text{Equation 12}$$

This score uses the predicted and true values of the ASHR and computes the square differences between the two values, divided by the square differences between the true values and the average of the true values. The result can be negative, indicating an arbitrarily worse model and up to 1, indicating perfect prediction.

8.1.3. Results

8.1.3.1. Correlation

Spearman correlation coefficients were computed between all variables in the patient-oriented dataset. Hospitalization rates were significantly correlated to all other variables except for benzene. Correlation results are summarized in Table 16. Primarily diagnosed neurodegenerative diseases' hospitalizations were analysed separately and showed usually more correlation to environmental factors and less to personal variables and environmental pollutants, keeping the significance.

Table 16 - Spearman's' correlation coefficients between variables (** - significant at 0.001; * - significant at 0.05). In parenthesis are the values obtained for the primary diagnosis database.

	Hospitalization rate	Year	Month	Age	Internment Days	C_6H_6	CO	NO_2	O_3	PM_{10}	$PM_{2.5}$	SO_2	Humidity	Pressure
Year	0.40** (0.22**)													
Month	-0.01* (0.00)	-0.01** (0.00)												
Age	0.44** (0.09**)	0.17** (0.05**)	0.00 (0.01*)											
Length of stay (days)	0.01** (0.00)	-0.02** (0.06**)	0.00 (0.01)	0.02** (-0.05**)										
C_6H_6	0.00 (0.01)	-0.08** (-0.06**)	0.06** (0.06**)	0.00 (0.02**)	-0.00* (-0.01*)									
CO	-0.16** (-0.07**)	-0.45** (-0.45**)	0.01** (0.00)	-0.07** (-0.03**)	0.00* (-0.04**)	0.31** (0.31**)								
NO_2	-0.34** (-0.37**)	-0.44** (-0.36**)	0.08** (0.06**)	-0.13** (-0.02**)	0.01** (0.02**)	0.24** (0.20**)	0.52** (0.47**)							
O_3	0.24** (0.19**)	0.29** (-0.16**)	-0.19** (-0.16**)	0.08** (0.02**)	0.01* (0.01*)	-0.20** (-0.18**)	-0.38** (-0.40**)	-0.45** (-0.40**)						
PM_{10}	-0.32** (-0.23**)	-0.66** (-0.63**)	0.04** (0.04**)	-0.13** (-0.05**)	0.02** (-0.02**)	0.12** (0.14**)	0.52** (0.51**)	0.60** (0.56**)	-0.19** (-0.19**)					
$PM_{2.5}$	-0.09**	-0.23**	0.04**	-0.03**	0.01**	-0.01**	0.17**	0.27**	0.01**	0.48**				

	(-0.07**)	(-0.19**)	(0.03**)	(0.00)	(-0.01)	(0.01)	(0.11**)	(0.24**)	(0.04**)	(0.44**)				
SO_2	-0.12**	-0.49**	-0.02**	-0.05**	0.00	0.13**	0.45**	0.21**	-0.24**	0.37**	0.05**			
	(0.05**)	(-0.52**)	(-0.04**)	(-0.02**)	(-0.04**)	(0.16**)	(0.47**)	(0.16**)	(-0.27**)	(0.35**)	(-0.02**)			
Humidity	-0.09**	0.00	-0.03**	-0.02**	-0.01**	0.26**	0.30**	0.26**	-0.53**	-0.07**	-0.06**	-0.03**		
	(-0.12**)	(0.01)	(-0.05**)	(0.01)	(-0.02**)	(0.24**)	(0.28**)	(0.23**)	(-0.51**)	(-0.07**)	(-0.04**)	(-0.02**)		
Pressure	-0.32**	-0.05**	-0.03**	-0.10**	0.00	-0.12**	-0.06**	0.19**	-0.08**	0.20**	0.20**	0.05**	-0.30**	
	(-0.43**)	(0.02**)	(-0.03**)	(-0.02**)	(0.02**)	(-0.11**)	(-0.08**)	(0.23**)	(-0.07**)	(0.17**)	(0.20**)	(-0.34**)	(0.08**)	
Temperature	-0.04**	0.02**	0.24**	-0.02**	0.00	-0.30**	-0.45**	-0.30**	0.44**	0.00	0.02**	-0.74**	-0.10**	0.13**
	(-0.06**)	(0.03**)	(0.27**)	(-0.02**)	(0.02**)	(-0.27**)	(-0.45**)	(-0.27**)	(0.45**)	(-0.01*)	(0.01*)	(-0.14**)	(-0.72**)	(0.11**)

8.1.3.2. Regression algorithms

Regression algorithms performed poorly (Table 17). The best algorithm was the random forest, with an accuracy of 0.518. Thus, these models are not suitable to predict age-standardized hospitalization rates. In this case, the primary diagnosis performed worse than the whole database.

Table 17 - Regression algorithms scores for all diagnosis and primary diagnosis neuro.

	All diagnosis	**Primary diagnosis**
Linear regression	0.249	0.201
Decision tree	0.122	-0.227
Random forest	0.518	0.461
Support vector machine	-4.632	-11.408

8.1.4. Discussion

Although spearman correlation coefficients showed promising correlation between the age standardized rates and the remaining variables, regression methods are inefficient when trying to predict hospitalizations rates. Age was the most correlated variable to ASHR and NO_2 and Pressure were also highly correlated, with correlation coefficients of -0.34 and -0.32, respectively, for all diagnosis and -0.37 and -0.43 for primary diagnosis. Only C_6H_6 wasn't correlated to the ASHR.

Among the limitations of this work, the lack of more precise clinical data may lead to insufficient variables to predict such complex diseases. Furthermore, acute hospitalizations' data such as the one in this study have limited evidence of the true development of a neurodegenerative disease, since hospitalizations rarely occur due to acute episodes of the diseases. Since distinguishing primarily diagnosed neurodegenerative diseases increased the classification models' accuracy, adding clinically relevant variables and limiting the analysis to neurodegenerative diseases related episodes may drastically increase model performance.

Also, patient's exposure to environmental factors is assigned to their residence, which may be incorrect, or may not be the place they spend the most time in. Environmental data is also limited, since it relies on meteorological stations unevenly scattered which needed interpolation, and remote sensing imagery with low spatial resolution.

8.1.5. Conclusion

This is the first study, to the author's knowledge, to attempt to model neurodegenerative diseases from environmental data in Portugal. All environmental variables were correlated to the age-standardized hospitalization rates, with exception of benzene. In future works, primary care data may be used to improve health data and complementary variables such as information on health history of diabetes and

other known risk factors should be added. Limiting the analysis to neurodegenerative related episodes instead of using any episode, as long as the patient has a neurodegenerative disease coded, may also provide better outputs.

8.2. Neurodegenerative risk assessment tool

Abstract: Theoretical results may be of interest to understand the mechanisms underlying the interactions between neurodegenerative diseases and environmental factors. However, applying the results to real-life scenarios would raise its value and turn it into practical knowledge. Understanding the need for applicability in healthcare, this tool aims to be as straight-forward as possible, collecting relevant data from a semi-structured patient form quick to fill, and returning to the user easy interpretable results in an efficient way. Even though this is yet but a prototype, it acts as a foundation for applying the real-world data and the chosen algorithms.

Keywords: Risk assessment tool, website, neurodegenerative diseases

8.2.1. Homepage

This website was designed using Wix, an online platform to create websites without coding. In Figure 58 is a map of the website.

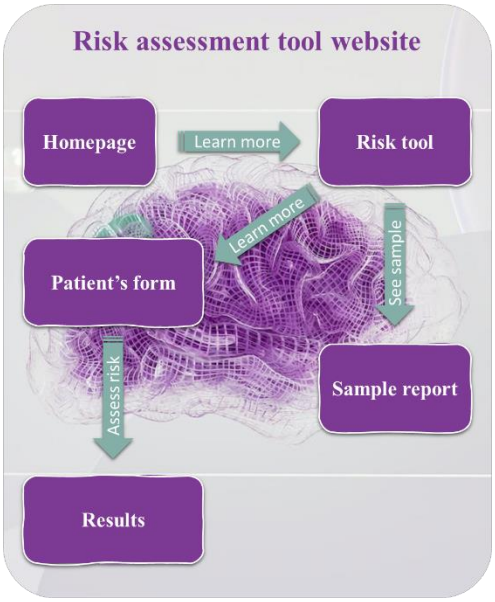


Figure 58 - Website map

Its homepage was designed to give an overview of the tool's objectives and features. It presents some key concepts inherent to the tool (Figure 59).

8.2.2. Tools page

By clicking in the "Learn more" button, users are redirected to the Tools page. In this page, a brief description of the tool is available and benefits of using the tool are explained. A comprehensive paragraph is also written on the functionality of the tool

Figure 60).

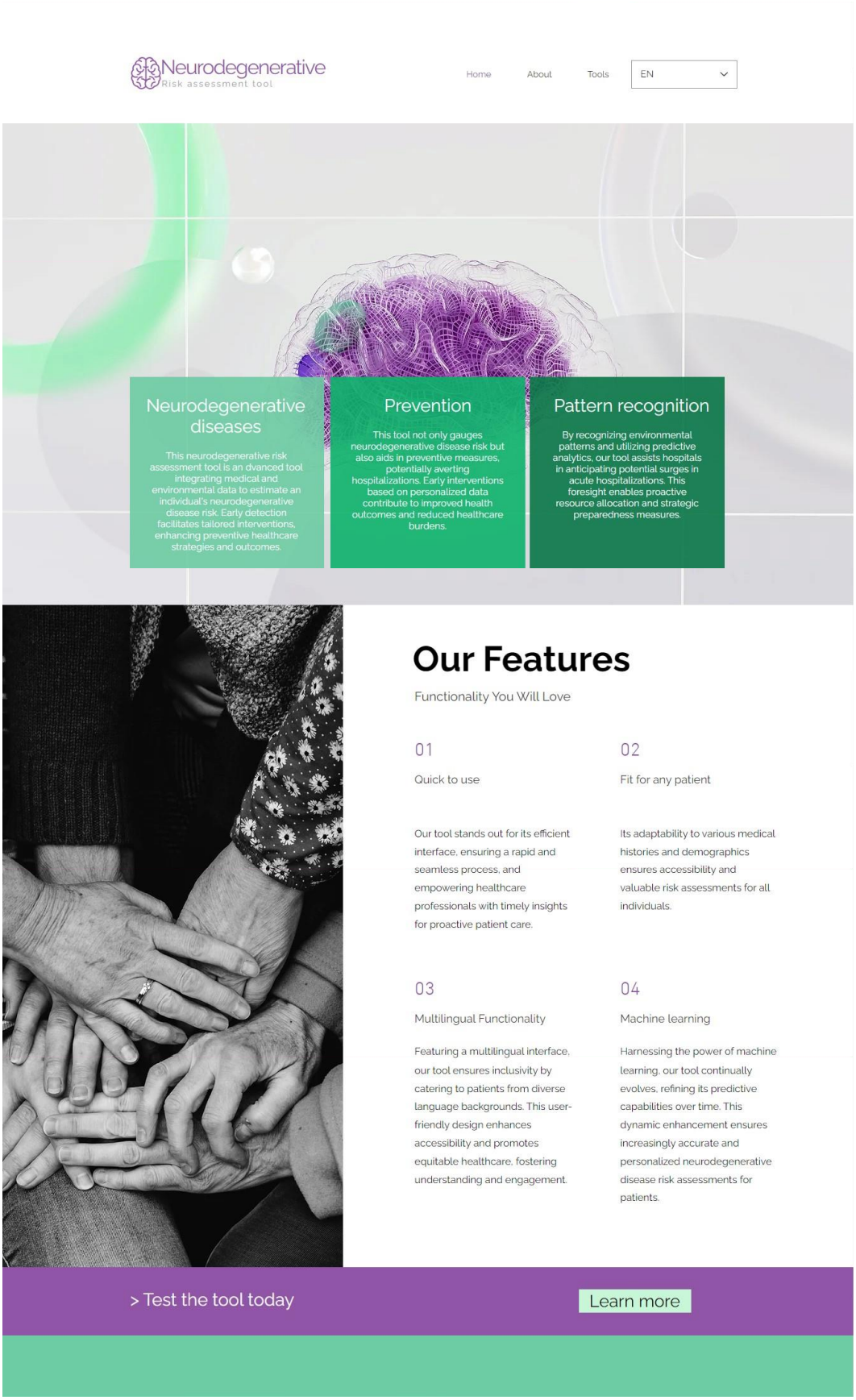


Figure 59 - Neurodegenerative risk assessment website - Homepage

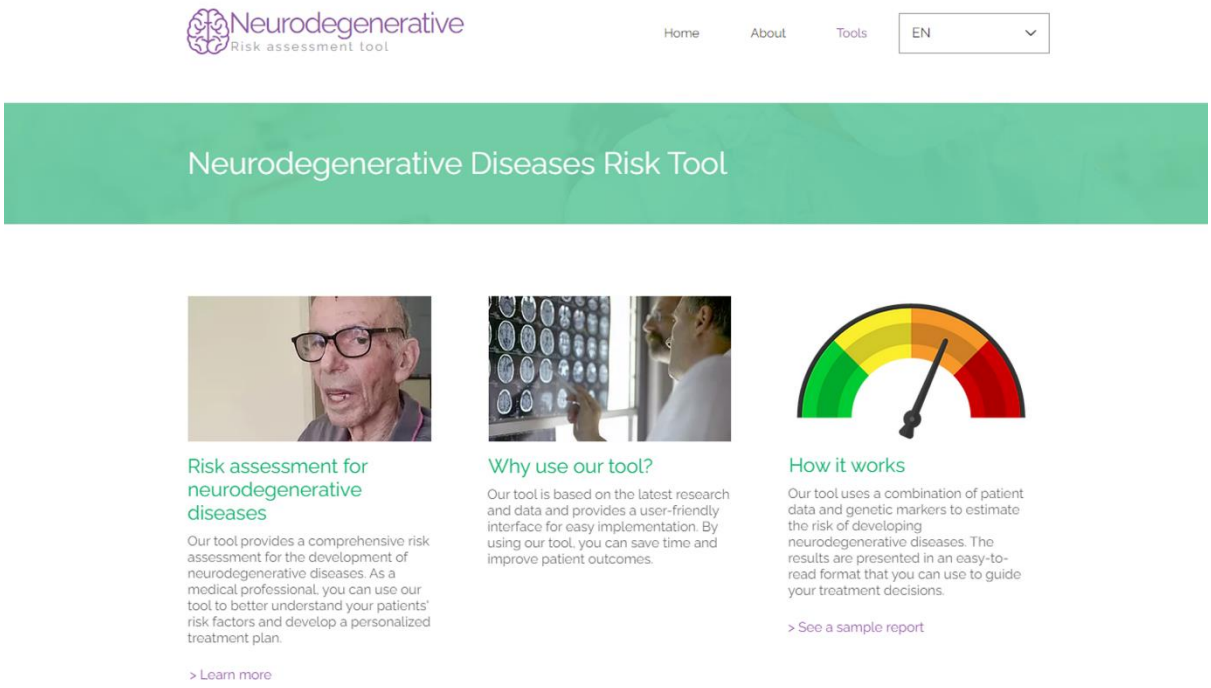


Figure 60 - Neurodegenerative risk assessment website - Tool page

8.2.3. Sample report page

A sample report of a patient is provided to help illustrate how the results of the tool should be presented to the user. A small description of the patient with the variables collected during the form filling phase of the tool is at the top, followed by a quick case report that may or may not be filled in by the user, and finally a visual of the risk result, accompanied by a small clinical recommendation Figure 61.

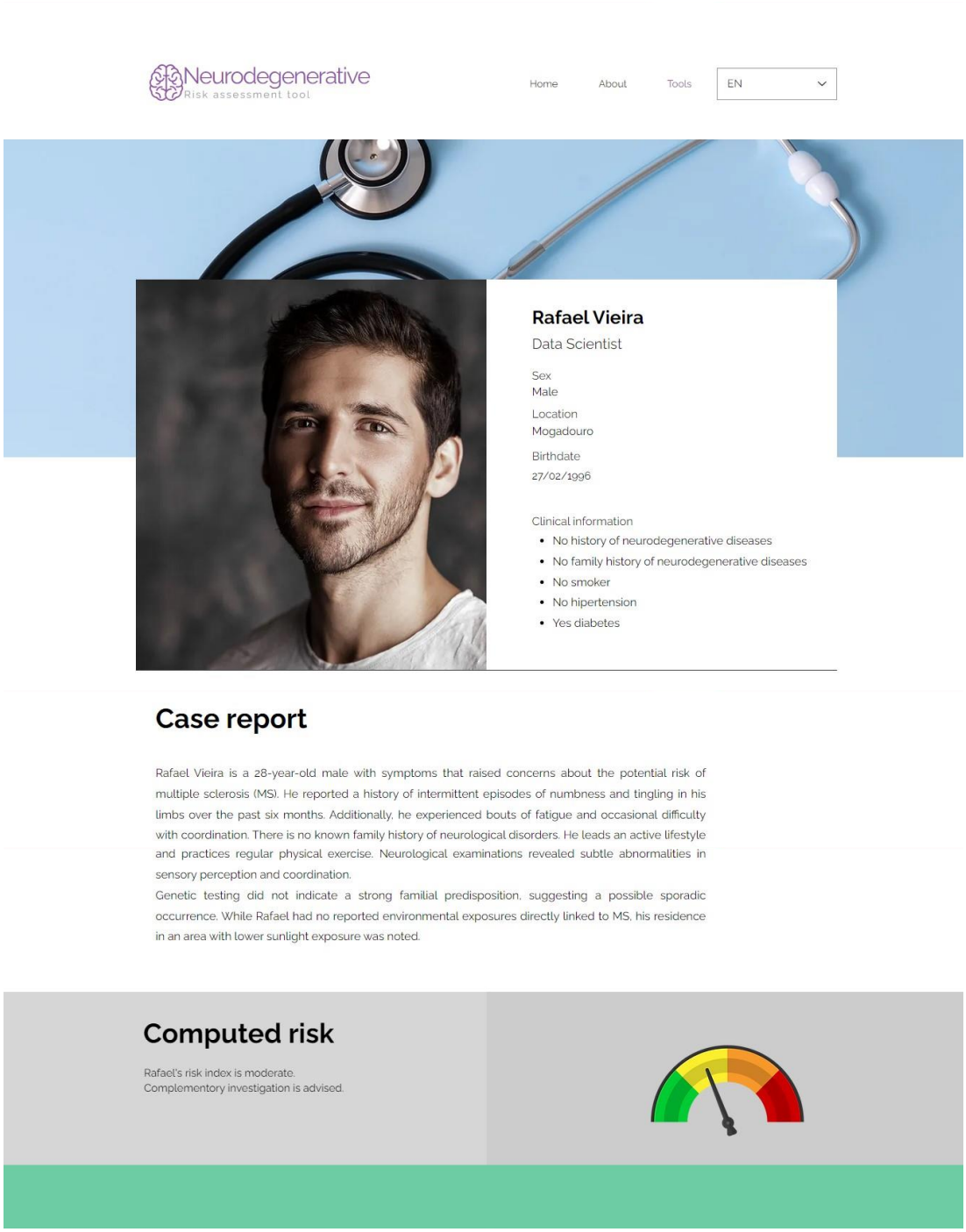


Figure 61 - Neurodegenerative risk assessment website - Sample report page

8.2.4. Patient's form page

A simulation of a patient's input form is also provided, with different types of both mandatory and optional fields. These include patient's name, address, location, sex and birthdate as socio-demographic features and some checkboxes representing relevant clinical information (Figure 62).

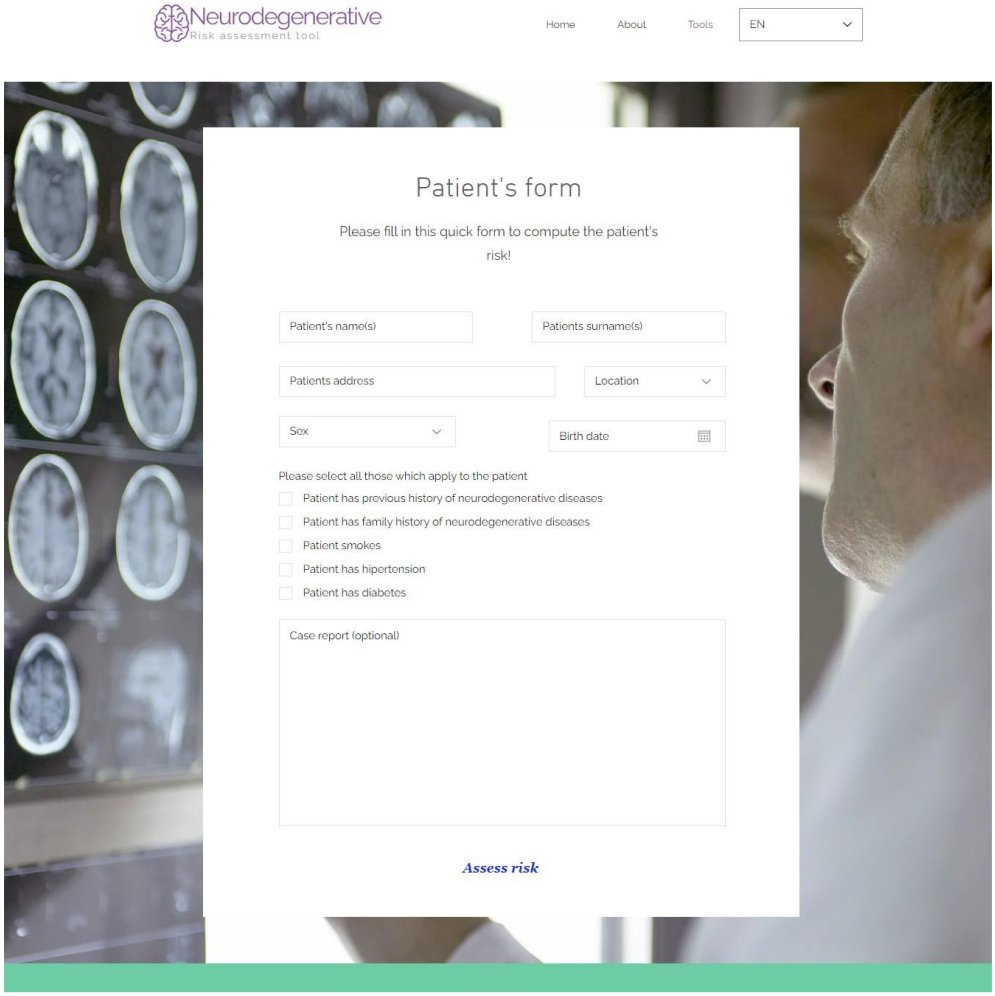


Figure 62 - Neurodegenerative risk assessment website – Patient's form page

8.2.5. Results page

After clicking "Assess risk", the user is redirected to the results page, in which a success message appears, together with a similar report to that in the sample report (Figure 63).

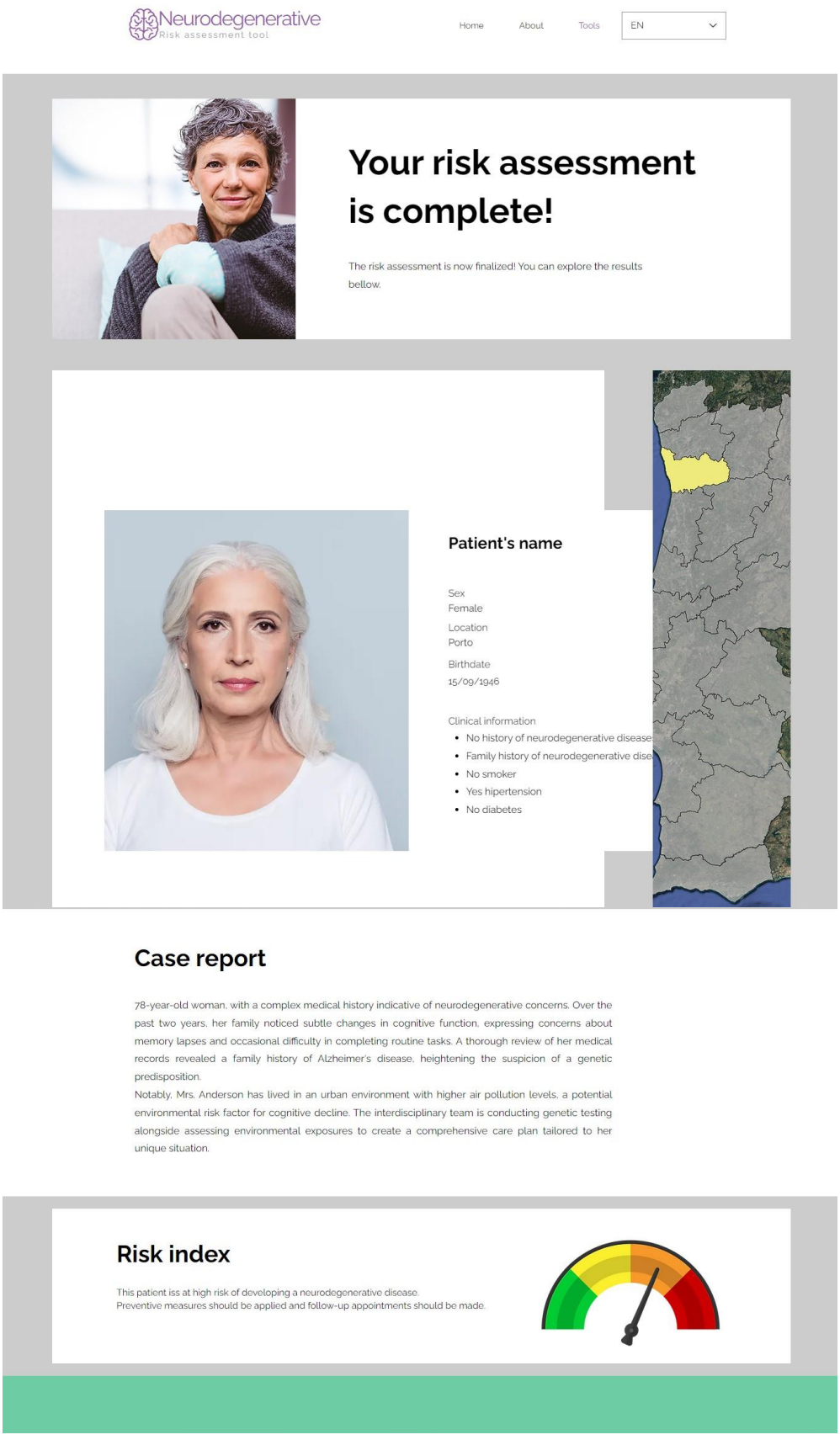


Figure 63 - Neurodegenerative risk assessment website - Results page

8.2.6. Conclusion

Assessing air pollution risk in various health problems is a rising concern, with several epidemiological studies finding correlation between environmental factors and a variety of diseases [275, 276]. Other tools such as AirQ+ [277] from WHO have been created to address this issue. In the particular case of neurodegenerative diseases, some tools can predict Dementia fairly accurately, but none are recommended for clinical use [278]. Risk assessment is a complex problem, with the need to understand the interaction between variables to accurately predict it. Even though this chapter only represents a prototype of what a website-based risk assessment tool could be, it is a good example of a framework to collect patient data and process it to compute risk.

9. Discussion

This thesis brought important insights on the relation between neurodegenerative hospitalizations and air pollution and environmental factors in Portugal.

Firstly, the systematic review showed which main pollutants were being related to neurodegenerative diseases, which neurodegenerative diseases showed correlations to these pollutants and what geographical approaches are being used. From this review, the most studied neurodegenerative disease was MS, with 43.5% of the studies, followed by Dementia with 34.8%. Diseases such as Basal, Creutzfeldt, Huntington, Lewy and Pick, which were addressed in this thesis, did not seem to have been studied with regard to environmental variables under any form of geographical approach. Environmentally speaking, $PM_{2.5}$ was the most approached, followed by NO_X. Geographically, most studies used the patient's residence to assess exposure (62.3%), as in this thesis, and used remote sensing data to retrieve the environmental variables (42.0%). In particular, ALS was related to humidity and NO_X (both significant), but no significant results arise with CO, O_3, Pb, PM_{10}, $PM_{2.5}$, SO_2, temperature and pressure. In regard to Dementia, $PM_{2.5}$ was mostly positively associated, whereas BC, humidity, NO_X and temperature did not show consistent results. MND was positively associated to Pb in one study. MS also had controversial results in relation to sun exposure, temperature and precipitation. In the particular case of PMS, air quality was linked to the disease, as well as CO, Pb, $PM_{2.5}$ and SO_2. PM_{10}, NO_X and O_3 were not associated with PMS. Finally, PD was related to CO, NO_X, NO_2, sun exposure, O_3, PM_{10} and $PM_{2.5}$, but not with BC, Cu, Pb and Mg.

Through the LST algorithm calibration, high-resolution LST was derived and, with the help of meteorological stations, calibrated. This resulted in a better LST product than that originated simply through Landsat imagery. However, when compared to ERA5's LST product, Landsat proved to be less accurate, probably because ERA5 algorithms are more developed and calibrated than the one developed for Landsat. In this case, the user must compromise on either accuracy (and choose Landsat) or spatial resolution (and choose ERA5). Since the extent needed for the computation of LST was rather big, ERA5 was used throughout this thesis. Along with LST, ERA5 provides both surface pressure and relative humidity.

The analysis of health data, although limited by the inherent issues related to its source (hospitalizations data reuse), brought insights on the characteristics and spatio-temporal distribution of neurodegenerative-related hospitalizations. In this subject, results of the descriptive analysis were in line with those in literature. Spatio-temporally, an increase in age-standardized hospitalization rates is noted, probably related to the higher awareness for better coding and the ageing of the Portuguese population. Global spatial autocorrelation was higher at a district level in the first years and at the

municipality level in the final years of the study period. From this, an associated environmental risk may be posed as a hypothesis.

Following that line of thought, environmental variables were analysed. Although Portuguese data was of main relevance to this thesis, since the Spanish data was available and useful to improve the analysis around Portugal borders, the environmental analysis was extended to the whole of Iberia. Temporally, environmental data exhibited high seasonality and overall decreasing trends, and spatially all variables showed significant global and local autocorrelation. The spatial patterns identified may be mostly explained through the conjoined analysis with the Iberian Peninsula topography.

In the exploitation of the relation between health and environmental data, correlation analysis indicated promising results. However, when attempting to model the data into classification and regression models, the algorithms performance was generally low. This may be related to the fact of relying solely on hospitalizations and to the lack of clinically relevant variables, namely the use of data from the primary care services, since neurodegenerative diseases progress is usually gradual and require primary care, leading to insufficient variables to predict such complex diseases. Since neurodegenerative diseases develop gradually throughout time and are not typically diagnosed based on acute events, as is the case of those related with the cardiovascular system, it turns out to be a much more difficult topic in terms of prediction. Two approaches were attempted to improve the algorithms' accuracy: grouping neurodegenerative diseases by their main pathological characteristics, and analysing primarily diagnosed neurodegenerative diseases separately. Both improved the models and together reached the higher accuracy score.

Finally, a tool prototype was developed, even though with no model attached to it yet. This prototype aims to establish a foundation onto which a more robust model can be applied to work as a diagnosis assistance tool by signalling patients at higher risk of developing a particular disease.

In future work, this study may be improved with primary care data, ideally integrating further environmental variables such as BC and PM_1, which were identified in the systematic review but whose estimation methods are not very broad yet. For these particular exposures, using road proximity might be an interesting approach, since most of these pollutants are produced by traffic.

Apart from retrospectively using this work, a prospective coorte study across several years would fit the methodology even better, allowing for collection of a larger number of variables, both clinical and environmental as, for example, blood and genetic tests results, smoking habits and working and living locations and type of environments. For instance, in Portugal, a prospective cohort study was in place from the Public Health Institute of the University of Porto with a sample size of nearly 2500 people aged 18 to 92 that, although focusing on no disease specific, collected a wide collection of variables, including each person's address. With data purposedly collected to explore the relation between

environmental factors and neurodegenerative diseases, the limitations connected to the reuse of data would be overcome and those related to the assessment bias of environmental exposure decreased.

10. Conclusion

Altogether, this is an innovative work relating neurodegenerative diseases to environmental air pollutants and factors. As far as is known, this is the most extensive work involving hospitalization episodes (between 2000 and 2016) which included diagnosed cases of a wide range of neurodegenerative diseases, and its relation with several of the most common environmental factors, through both temporal and spatial analysis. Further work on this complex and highly relevant topic might rely on the use of a database specifically gathered for its purpose, using health variables that are already known to be related to neurodegenerative pathologies, including lifestyle (e.g. smoking habits) and clinical (e.g. hypertension and diabetes) variables, ideally in a prospective cohort, but also elaborating simulation (in silico) studies.

References

1. Logroscino, G. and R. Tortelli, *Epidemiology of neurodegenerative diseases.* Imaging in Neurodegenerative Disorders, 2014. **6**: p. 1.
2. Logroscino, G., D. Urso, and R. Savica, *Descriptive Epidemiology of Neurodegenerative Diseases: What Are the Critical Questions?* Neuroepidemiology, 2022. **56**(5): p. 309-318. DOI: https://doi.org/10.1159/000525639.
3. Paprocka, J., et al., *Neuroimaging of Basal Ganglia in Neurometabolic Diseases in Children.* Brain Sciences, 2020. **10**(11): p. 849.
4. Iwasaki, Y., *Creutzfeldt-Jakob disease.* Neuropathology, 2017. **37**(2): p. 174-188. DOI: https://doi.org/https://doi.org/10.1111/neup.12355.
5. Head, M., et al., *Greenfield's neuropathology*. 2015, CRC Press Boca Raton, FL, USA:.
6. Grossman, M., *Frontotemporal dementia: a review.* J Int Neuropsychol Soc, 2002. **8**(4): p. 566-83. DOI: https://doi.org/10.1017/s1355617702814357.
7. Korczyn, A.D., V. Vakhapova, and L.T. Grinberg, *Vascular dementia.* J Neurol Sci, 2012. **322**(1-2): p. 2-10. DOI: https://doi.org/10.1016/j.jns.2012.03.027.
8. Finkbeiner, S., *Huntington's Disease.* Cold Spring Harbor Perspectives in Biology, 2011. **3**(6). DOI: https://doi.org/10.1101/cshperspect.a007476.
9. McKeith, I., *Dementia with Lewy bodies.* Dialogues Clin Neurosci, 2004. **6**(3): p. 333-41. DOI: https://doi.org/10.31887/DCNS.2004.6.3/imckeith.
10. Gomperts, S.N., *Lewy Body Dementias: Dementia With Lewy Bodies and Parkinson Disease Dementia.* Continuum (Minneap Minn), 2016. **22**(2 Dementia): p. 435-63. DOI: https://doi.org/10.1212/con.0000000000000309.
11. Talbot, K., *Motor neuron disease.* THE BARE ESSENTIALS, 2009. **9**(5): p. 303-309. DOI: https://doi.org/10.1136/jnnp.2009.188151.
12. Doshi, A. and J. Chataway, *Multiple sclerosis, a treatable disease* Clin Med (Lond), 2017. **17**(6): p. 530-536. DOI: https://doi.org/10.7861/clinmedicine.17-6-530.
13. de Lau, L.M.L. and M.M.B. Breteler, *Epidemiology of Parkinson's disease.* The Lancet Neurology, 2006. **5**(6): p. 525-535. DOI: https://doi.org/https://doi.org/10.1016/S1474-4422(06)70471-9.
14. Jellinger, K.A., *Neuropathobiology of non-motor symptoms in Parkinson disease.* Journal of Neural Transmission, 2015. **122**(10): p. 1429-1440. DOI: https://doi.org/10.1007/s00702-015-1405-5.
15. Devi, G., *Chapter 12 - The tauopathies*, in *Handbook of Clinical Neurology*, D.S. Younger, Editor. 2023, Elsevier. p. 251-265.
16. Takeda, N., Y. Kishimoto, and O. Yokota, *Pick's Disease*, in *Neurodegenerative Diseases*, S.I. Ahmad, Editor. 2012, Springer US: New York, NY. p. 300-316.
17. Shah, A.S.V., et al., *Short term exposure to air pollution and stroke: systematic review and meta-analysis.* BMJ, 2015. **350**: p. h1295. DOI: https://doi.org/10.1136/bmj.h1295.
18. Lotufo, P.A., I.J.M. Benseñor, and J.M.F. Furtado, *Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016.* Lancet, 2017. **390**(10100): p. 1345-1422. DOI: https://doi.org/10.1016/s0140-6736(17)32366-8.
19. He, H., B. Schäfer, and C. Beck, *Spatial heterogeneity of air pollution statistics in Europe.* Scientific Reports, 2022. **12**(1): p. 12215. DOI: https://doi.org/10.1038/s41598-022-16109-2.
20. Lim, S.S., et al., *A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010.* The Lancet, 2012. **380**(9859): p. 2224-2260. DOI: https://doi.org/https://doi.org/10.1016/S0140-6736(12)61766-8.
21. Tomczak, A., et al., *Long-term exposure to fine particulate matter air pollution and the risk of lung cancer among participants of the Canadian National Breast Screening Study.* International

Journal of Cancer, 2016. **139**(9): p. 1958-1966. DOI: https://doi.org/https://doi.org/10.1002/ijc.30255.
22. Brito, J., et al., *Quantitative analysis of air pollution and mortality in Portugal: Current trends and links following proposed biological pathways.* Science of The Total Environment, 2021. **755**: p. 142473. DOI: https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.142473.
23. Guxens, M. and J. Sunyer, *A review of epidemiological studies on neuropsychological effects of air pollution.* Swiss Med Wkly, 2012. **142**: p. w13322. DOI: https://doi.org/10.57187/smw.2012.13322.
24. Oliveira, M., et al., *Geospatial Analysis of Environmental Atmospheric Risk Factors in Neurodegenerative Diseases: A Systematic Review*. International Journal of Environmental Research and Public Health, 2020. **17**(22): p. 8414.
25. Cleary, E.G., et al., *Association of Low-Level Ozone with Cognitive Decline in Older Adults.* J Alzheimers Dis, 2018. **61**(1): p. 67-78. DOI: https://doi.org/10.3233/jad-170658.
26. Viana, J., et al., *Remote Sensing in Human Health: A 10-Year Bibliometric Analysis.* Remote Sensing, 2017. **9**(12): p. 1225. DOI: https://doi.org/10.3390/rs9121225
27. USGS. *Landsat 7*. [cited 2023 December]; Available from: https://www.usgs.gov/landsat-missions/landsat-7.
28. USGS. *ASTER Global Emissivity Dataset 100 meter HDF5*. [cited 2023 December]; Available from: https://lpdaac.usgs.gov/products/ag100v003/.
29. Copernicus. *ERA5-Land monthly averaged data from 1950 to present*. [cited 2023 December]; Available from: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=overview.
30. Suwanprasit, C. and N. Srichai, *Impacts of spatial resolution on land cover classification.* Proceedings of the Asia-Pacific Advanced Network, 2012. **33**: p. 39. DOI: https://doi.org/10.7125/APAN.33.4.
31. Sedgwick, P., *Pearson's correlation coefficient.* Bmj, 2012. **345**.
32. Ester, M., et al. *A density-based algorithm for discovering clusters in large spatial databases with noise*. in *kdd*. 1996.
33. Drucker, H., et al., *Support vector regression machines.* Adv Neural Inform Process Syst, 1997. **28**: p. 779-784.
34. Li, J., *A Review of Spatial Interpolation Methods for Environmental Scientists*. 2008.
35. Arribas-Bel, D., *A course on Geographic Data Science.* Journal of Open Source Education, 2019. **2**: p. 42. DOI: https://doi.org/10.21105/jose.00042.
36. Glen, S. *Moran's I: Definition, Examples*. 2024 [cited 2024 January]; Available from: https://www.statisticshowto.com/morans-i/.
37. Das, A., *Logistic Regression*, in *Encyclopedia of Quality of Life and Well-Being Research*, F. Maggino, Editor. 2020, Springer International Publishing: Cham. p. 1-2.
38. Maulud, D. and A.M. Abdulazeez, *A review on linear regression comprehensive in machine learning*. Journal of Applied Science and Technology Trends, 2020. **1**(4): p. 140-147.
39. Pekel, E., *Estimation of soil moisture using decision tree regression.* Theoretical and Applied Climatology, 2020. **139**(3): p. 1111-1119. DOI: https://doi.org/10.1007/s00704-019-03048-8.
40. Zhang, F. and L.J. O'Donnell, *Chapter 7 - Support vector regression*, in *Machine Learning*, A. Mechelli and S. Vieira, Editors. 2020, Academic Press. p. 123-140.
41. Ao, Y., et al., *The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling*. Journal of Petroleum Science and Engineering, 2019. **174**: p. 776-789. DOI: https://doi.org/https://doi.org/10.1016/j.petrol.2018.11.067.
42. Wetmore, J.B., et al., *Increases in institutionalization, healthcare resource utilization, and mortality risk associated with Parkinson disease psychosis: Retrospective cohort study.* Parkinsonism Relat Disord, 2019. **68**: p. 95-101. DOI: https://doi.org/10.1016/j.parkreldis.2019.10.018.
43. Lee, S.Y., et al., *Increased risk of neurodegenerative dementia in women with migraines: A nested case-control study using a national sample cohort.* Medicine (Baltimore), 2019. **98**(7): p. e14467. DOI: https://doi.org/10.1097/md.0000000000014467.

44. Min, C., et al., *Rheumatoid arthritis and neurodegenerative dementia: a nested case-control study and a follow-up study using a national sample cohort.* Clin Rheumatol, 2020. **39**(1): p. 159-166. DOI: https://doi.org/10.1007/s10067-019-04769-x.
45. Rochoy, M., et al., *Factors associated with the onset of Alzheimer's disease: Data mining in the French nationwide discharge summary database between 2008 and 2014.* PLoS One, 2019. **14**(7): p. e0220174. DOI: https://doi.org/10.1371/journal.pone.0220174.
46. St Germaine-Smith, C., et al., *Recommendations for optimal ICD codes to study neurologic conditions: a systematic review.* Neurology, 2012. **79**(10): p. 1049-55. DOI: https://doi.org/10.1212/WNL.0b013e3182684707.
47. Kim, S.Y., et al., *Risk of neurodegenerative dementia in asthma patients: a nested case-control study using a national sample cohort.* BMJ Open, 2019. **9**(10): p. e030227. DOI: https://doi.org/10.1136/bmjopen-2019-030227.
48. Fatima, K., A.M. Mehendale, and H. Reddy, *Young-Onset Dementia and Neurodegenerative Disorders of the Young With an Emphasis on Clinical Manifestations.* Cureus, 2022. **14**(10): p. e30025. DOI: https://doi.org/10.7759/cureus.30025.
49. INE. *Censos 2021 - Indicadores*. [cited 2022 December]; Available from: https://tabulador.ine.pt/censos2021/.
50. Curado, M.-P., et al., *Cancer incidence in five continents, Volume IX*. 2007: IARC Press, International Agency for Research on Cancer.
51. APA. *QualAr*. [cited 2022 January]; Available from: https://qualar.apambiente.pt/downloads.
52. IPMA. *IPMA - Long data series*. [cited 2021 March]; Available from: https://www.ipma.pt/en/oclima/series.longas/list.jsp.
53. MITECO. *Datos Calidad del Aire 2001-2021*. [cited 2022 January]; Available from: https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/calidad-del-aire/evaluacion-datos/datos/datos_oficiales_2021.html.
54. NOAA. *Vapor Pressure*. [cited 2023 November]; Available from: https://www.weather.gov/media/epz/wxcalc/vaporPressure.pdf.
55. Sampaio, D.A., *Análise espácio-temporal da asma em Portugal Continental com base em dados de deteção remota*. 2012, University of Porto.
56. APA. *SNIRH - National System of Information on Hydric Resources*. [cited 2021 February]; Available from: https://snirh.apambiente.pt/.
57. DGT. *SNIG - National Infrastructure for Geographical Information*. [cited 2020 December]; Available from: https://snig.dgterritorio.gov.pt/rndg/srv/por/catalog.search#/search?anysnig=COS&fast=index .
58. WMO, *Guide to Instruments and Methods of Observation Volume 1–Measurement of Meteorological Variables.* 2018.
59. GEE. *Earth Engine Data Catalog - USGS Landsat 7 Level 2*. Available from: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LE07_C02_T1_L2.
60. Avdan, U. and G. Jovanovska, *Algorithm for Automated Mapping of Land Surface Temperature Using LANDSAT 8 Satellite Data.* Journal of Sensors, 2016. **2016**: p. 1480307. DOI: https://doi.org/10.1155/2016/1480307.
61. Spencer, P., V. Palmer, and G. Kisby, *Seeking environmental causes of neurodegenerative disease and envisioning primary prevention.* Neurotoxicology, 2016. **56**. DOI: https://doi.org/10.1016/j.neuro.2016.03.017.
62. WHO, W.H.O. *Ageing and health*. 2020 [cited 2023 June]; Available from: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health.
63. Garre-Olmo, J., *[Epidemiology of Alzheimer's disease and other dementias].* Rev Neurol, 2018. **66**(11): p. 377-386.
64. Liu, Z., et al., *Oxidative Stress in Neurodegenerative Diseases: From Molecular Mechanisms to Clinical Applications.* Oxid Med Cell Longev, 2017. **2017**: p. 2525967. DOI: https://doi.org/10.1155/2017/2525967.
65. WHO, W.H.O., *Neurological Disorders: Public Health Challenges*. 2006.
66. Campanozzi, M.D., et al., *Evaluation of the slopes of cognitive impairment and disability in Alzheimer's disease (AD) patients treated with acetylcholinesterase inhibitors (AChEI).* Arch

Gerontol Geriatr, 2007. **44 Suppl 1**: p. 91-6. DOI: https://doi.org/10.1016/j.archger.2007.01.013.

67. Pang, S.Y.-Y., et al., *The interplay of aging, genetics and environmental factors in the pathogenesis of Parkinson's disease.* Translational Neurodegeneration, 2019. **8**(1): p. 23. DOI: https://doi.org/10.1186/s40035-019-0165-9.
68. Brown, R.C., A.H. Lockwood, and B.R. Sonawane, *Neurodegenerative Diseases: An Overview of Environmental Risk Factors.* Environmental Health Perspectives, 2005. **113**(9): p. 1250-1256. DOI: https://doi.org/doi:10.1289/ehp.7567.
69. WHO, W.H.O. *Air pollution*. 2020 [cited 2023 June]; Available from: https://www.who.int/health-topics/air-pollution#tab=tab_2.
70. Almetwally, A.A., M. Bin-Jumah, and A.A. Allam, *Ambient air pollution and its influence on human health and welfare: an overview*. Environ Sci Pollut Res Int, 2020. **27**(20): p. 24815-24830. DOI: https://doi.org/10.1007/s11356-020-09042-2.
71. Putta, S.N., *Atmospheric-pollution, its history, origins and prevention.* Journal of the American Chemical Society, 1984. **106**(10): p. 3066-3066.
72. Calderón-Garcidueñas, L., et al., *Air pollution and your brain: what do you need to know right now.* Prim Health Care Res Dev, 2015. **16**(4): p. 329-45. DOI: https://doi.org/10.1017/s146342361400036x.
73. Block, M.L. and L. Calderón-Garcidueñas, *Air pollution: mechanisms of neuroinflammation and CNS disease.* Trends in Neurosciences, 2009. **32**(9): p. 506-516. DOI: https://doi.org/10.1016/j.tins.2009.05.009.
74. Calderón-Garcidueñas, L., et al., *Air pollution, cognitive deficits and brain abnormalities: A pilot study with children and dogs.* Brain and Cognition, 2008. **68**(2): p. 117-127. DOI: https://doi.org/https://doi.org/10.1016/j.bandc.2008.04.008.
75. Tamagawa, E. and S.F. van Eeden, *Impaired Lung Function and Risk for Stroke: Role of the Systemic Inflammation Response?* CHEST, 2006. **130**(6): p. 1631-1633. DOI: https://doi.org/10.1378/chest.130.6.1631.
76. Rückerl, R., et al., *Air Pollution and Inflammation (Interleukin-6, C-Reactive Protein, Fibrinogen) in Myocardial Infarction Survivors.* Environmental Health Perspectives, 2007. **115**(7): p. 1072-1080. DOI: https://doi.org/doi:10.1289/ehp.10021.
77. Calderón-Garcidueñas, L., et al., *Systemic Inflammation, Endothelial Dysfunction, and Activation in Clinically Healthy Children Exposed to Air Pollutants.* Inhalation Toxicology, 2008. **20**(5): p. 499-506. DOI: https://doi.org/10.1080/08958370701864797.
78. Valavanidis, A., K. Fiotakis, and T. Vlachogianni, *Airborne Particulate Matter and Human Health: Toxicological Assessment and Importance of Size and Composition of Particles for Oxidative Damage and Carcinogenic Mechanisms.* Journal of Environmental Science and Health, Part C, 2008. **26**(4): p. 339-362. DOI: https://doi.org/10.1080/10590500802494538.
79. Nemmar, A. and I.M. Inuwa, *Diesel exhaust particles in blood trigger systemic and pulmonary morphological alterations.* Toxicology Letters, 2008. **176**(1): p. 20-30. DOI: https://doi.org/https://doi.org/10.1016/j.toxlet.2007.09.006.
80. Oberdörster, G., et al., *Translocation of Inhaled Ultrafine Particles to the Brain.* Inhalation Toxicology, 2004. **16**(6-7): p. 437-445. DOI: https://doi.org/10.1080/08958370490439597.
81. Wang, B., et al., *Transport of Intranasally Instilled Fine Fe2O3 Particles into the Brain: Micro-distribution, Chemical States, and Histopathological Observation.* Biological Trace Element Research, 2007. **118**(3): p. 233-243. DOI: https://doi.org/10.1007/s12011-007-0028-6.
82. Wang, J., et al., *Time-dependent translocation and potential impairment on central nervous system by intranasally instilled TiO2 nanoparticles.* Toxicology, 2008. **254**(1): p. 82-90. DOI: https://doi.org/https://doi.org/10.1016/j.tox.2008.09.014.
83. Calderón-Garcidueñas, L., et al., *Long-term Air Pollution Exposure Is Associated with Neuroinflammation, an Altered Innate Immune Response, Disruption of the Blood-Brain Barrier, Ultrafine Particulate Deposition, and Accumulation of Amyloid β-42 and α-Synuclein in Children and Young Adults.* Toxicologic Pathology, 2008. **36**(2): p. 289-310. DOI: https://doi.org/10.1177/0192623307313011.

84. Mühlfeld, C., et al., *Interactions of nanoparticles with pulmonary structures and cellular responses.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2008. **294**(5): p. L817-L829. DOI: https://doi.org/10.1152/ajplung.00442.2007.
85. Simkhovich, B.Z., M.T. Kleinman, and R.A. Kloner, *Air Pollution and Cardiovascular Injury: Epidemiology, Toxicology, and Mechanisms.* Journal of the American College of Cardiology, 2008. **52**(9): p. 719-726. DOI: https://doi.org/https://doi.org/10.1016/j.jacc.2008.05.029.
86. Peters, A., et al., *Translocation and potential neurological effects of fine and ultrafine particles a critical update.* Particle and Fibre Toxicology, 2006. **3**(1): p. 13. DOI: https://doi.org/10.1186/1743-8977-3-13.
87. Pryor, W.A., G.L. Squadrito, and M. Friedman, *A new mechanism for the toxicity of ozone.* Toxicology Letters, 1995. **82-83**: p. 287-293. DOI: https://doi.org/https://doi.org/10.1016/0378-4274(95)03563-X.
88. Hollingsworth, J.W., S.R. Kleeberger, and W.M. Foster, *Ozone and Pulmonary Innate Immunity.* Proceedings of the American Thoracic Society, 2007. **4**(3): p. 240-246. DOI: https://doi.org/10.1513/pats.200701-023AW.
89. Guevara-Guzmán, R., et al., *Estradiol prevents ozone-induced increases in brain lipid peroxidation and impaired social recognition memory in female rats.* Neuroscience, 2009. **159**(3): p. 940-950. DOI: https://doi.org/10.1016/j.neuroscience.2009.01.047.
90. Pereyra-Muñoz, N., et al., *Oxidative damage in substantia nigra and striatum of rats chronically exposed to ozone.* Journal of Chemical Neuroanatomy, 2006. **31**(2): p. 114-123. DOI: https://doi.org/https://doi.org/10.1016/j.jchemneu.2005.09.006.
91. Angoa-Pérez, M., et al., *Estrogen counteracts ozone-induced oxidative stress and nigral neuronal death.* NeuroReport, 2006. **17**(6).
92. Fu, P., et al., *The association between PM(2.5) exposure and neurological disorders: A systematic review and meta-analysis.* Sci Total Environ, 2019. **655**: p. 1240-1248. DOI: https://doi.org/10.1016/j.scitotenv.2018.11.218.
93. Willis, A.W., et al., *Geographic and ethnic variation in Parkinson disease: a population-based study of US Medicare beneficiaries.* Neuroepidemiology, 2010. **34**(3): p. 143-51. DOI: https://doi.org/10.1159/000275491.
94. Scott, K.M., et al., *Geographical clustering of amyotrophic lateral sclerosis in South-East England: a population study.* Neuroepidemiology, 2009. **32**(2): p. 81-8. DOI: https://doi.org/10.1159/000177032.
95. Noonan, C.W., et al., *Temporal and geographic variation in United States motor neuron disease mortality, 1969-1998.* Neurology, 2005. **64**(7): p. 1215-21. DOI: https://doi.org/10.1212/01.Wnl.0000156518.22559.7f.
96. Achar, A. and C. Ghosh, *COVID-19-Associated Neurological Disorders: The Potential Route of CNS Invasion and Blood-Brain Barrier Relevance.* Cells, 2020. **9**(11): p. 2360.
97. Calderón-Garcidueñas, L., et al., *Environmental Nanoparticles, SARS-CoV-2 Brain Involvement, and Potential Acceleration of Alzheimer's and Parkinson's Diseases in Young Urbanites Exposed to Air Pollution.* Journal of Alzheimer's Disease, 2020. **78**: p. 479-503. DOI: https://doi.org/10.3233/JAD-200891.
98. Bougakov, D., K. Podell, and E. Goldberg, *Multiple Neuroinvasive Pathways in COVID-19.* Molecular Neurobiology, 2021. **58**(2): p. 564-575. DOI: https://doi.org/10.1007/s12035-020-02152-5.
99. Rossby, C.G., *Über die Vertikalverteilung von Windgeschwindigkeit und Schwerestabilität in Freistrahlbewegungen der oberen Troposphäre.* Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie A, 1951. **4**(1): p. 3-23. DOI: https://doi.org/10.1007/BF02246789.
100. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* Bmj, 2009. **339**: p. b2535. DOI: https://doi.org/10.1136/bmj.b2535.
101. Cendrowski, W., et al., *Epidemiological study of multiple sclerosis in western Poland.* Eur Neurol, 1969. **2**(2): p. 90-108. DOI: https://doi.org/10.1159/000113777.
102. Behrend, R.C., *Multiple sclerosis in Europe.* Eur Neurol, 1969. **2**(3): p. 129-45. DOI: https://doi.org/10.1159/000113782.

103. Bird, A.V. and E. Satoyoshi, *Comparative epidemiological studies of multiple sclerosis in South Africa and Japan.* Journal of Neurology, Neurosurgery & Psychiatry, 1975. **38**(9): p. 911-918. DOI: https://doi.org/10.1136/jnnp.38.9.911.
104. Lauer, K. and W. Firnhaber, *Epidemiological investigations into multiple sclerosis in southern Hesse. II. The distribution of cases in relation to exogenous features.* Acta Neurol Scand, 1984. **70**(4): p. 266-73.
105. Granieri, E., et al., *Multiple sclerosis: does epidemiology contribute to providing etiological clues?* J Neurol Sci, 1993. **115**: p. S16-23.
106. Lenský, P., *[Geographic aspects in the epidemiology of multiple sclerosis].* Epidemiol Mikrobiol Imunol, 1994. **43**(4): p. 174-6.
107. Delasnerie-Laupretre, N. and A. Alperovitch, *[Epidemiology of Creutzfeldt-Jakob disease].* Pathol Biol (Paris), 1995. **43**(1): p. 22-4.
108. Bharanidharan, P., *Monthly distribution of multiple sclerosis patients' births.* International Journal of Biometeorology, 1997. **40**(2): p. 117-118. DOI: https://doi.org/10.1007/s004840050030.
109. Hogancamp, W.E., M. Rodriguez, and B.G. Weinshenker, *The Epidemiology of Multiple Sclerosis.* Mayo Clinic Proceedings, 1997. **72**(9): p. 871-878. DOI: https://doi.org/10.4065/72.9.871.
110. Foster, H.D., *Parkinson's disease, multiple sclerosis and amyotrophic lateral sclerosis: the iodine-dopachrome-glutamate hypothesis.* Journal of Orthomolecular Medicine, 1999. **14**: p. 128-136.
111. Mitchell, J.D., *Amyotrophic lateral sclerosis: toxins and environment.* Amyotroph Lateral Scler Other Motor Neuron Disord, 2019. **1**(4): p. 235-50. DOI: https://doi.org/10.1080/14660820050515061.
112. Uría, D.F., *[Genetic epidemiology of multiple sclerosis].* Rev Neurol, 2002. **35**(10): p. 979-84.
113. Marrie, R.A., *Environmental risk factors in multiple sclerosis aetiology.* The Lancet Neurology, 2004. **3**(12): p. 709-718. DOI: https://doi.org/10.1016/S1474-4422(04)00933-0.
114. Sotgiu, S., et al., *Genes, environment, and susceptibility to multiple sclerosis.* Neurobiology of Disease, 2004. **17**(2): p. 131-143. DOI: https://doi.org/https://doi.org/10.1016/j.nbd.2004.07.015.
115. Ances, B.M., N.J. Newman, and L.J. Balcer, *Autoimmunity – Multiple Sclerosis.* Measuring Immunity, 2005: p. 515-24. DOI: https://doi.org/10.1016/b978-012455900-4/50307-x.
116. Williamson, D.M., *Studies of multiple sclerosis in communities concerned about environmental exposures.* J Womens Health (Larchmt), 2006. **15**(7): p. 810-4. DOI: https://doi.org/10.1089/jwh.2006.15.810.
117. Ascherio, A. and K.L. Munger, *Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors.* Annals of Neurology, 2007. **61**(6): p. 504-513. DOI: https://doi.org/https://doi.org/10.1002/ana.21141.
118. Ebers, G.C., *Environmental factors and multiple sclerosis.* The Lancet Neurology, 2008. **7**(3): p. 268-277. DOI: https://doi.org/10.1016/S1474-4422(08)70042-5.
119. Boumediene, F., et al., *Contribution of geolocalisation to neuroepidemiological studies: incidence of ALS and environmental factors in Limousin, France.* J Neurol Sci, 2011. **309**(1): p. 115-22.
120. Nelson, G.R., B.A., *Environmental risk factors*, in *Handbook of Parkinson's Disease*. 2013, CRC Press. p. 341--357.
121. Willis, A.W., et al., *Predictors of Survival in Patients With Parkinson Disease.* Archives of Neurology, 2012. **69**(5): p. 601-607. DOI: https://doi.org/10.1001/archneurol.2011.2370.
122. Caller, T.A., et al., *Spatial clustering of amyotrophic lateral sclerosis and the potential role of BMAA.* Amyotroph Lateral Scler, 2012. **13**(1): p. 25-32. DOI: https://doi.org/10.3109/17482968.2011.621436.
123. Tennant, A., *Epidemiology of neurologically disabling disorders.* Handb Clin Neurol, 2013. **110**: p. 77-92. DOI: https://doi.org/10.1016/b978-0-444-52901-5.00007-1.
124. Al-Afasy, H.H., et al., *Risk factors for multiple sclerosis in Kuwait: a population-based case-control study.* Neuroepidemiology, 2013. **40**(1): p. 30-5. DOI: https://doi.org/10.1159/000341240.

125. Chamosa, S., et al., *Geographical analysis of the sporadic Creutzfeldt-Jakob disease distribution in the autonomous community of the Basque Country for the period 1995-2008.* Eur Neurol, 2014. **72**(1-2): p. 20-5. DOI: https://doi.org/10.1159/000358298.
126. Thielke, S., C.G. Slatore, and W.A. Banks, *Association Between Alzheimer Dementia Mortality Rate and Altitude in California Counties.* JAMA Psychiatry, 2015. **72**(12): p. 1253-4. DOI: https://doi.org/10.1001/jamapsychiatry.2015.1852.
127. Russ, T.C., et al., *Geographical variation in dementia: examining the role of environmental factors in Sweden and Scotland.* Epidemiology, 2015. **26**(2): p. 263-70. DOI: https://doi.org/10.1097/ede.0000000000000230.
128. Berg-Hansen, P. and E.G. Celius, *Socio-economic factors and immigrant population studies of multiple sclerosis.* Acta Neurol Scand, 2015. **132**(199): p. 37-41. DOI: https://doi.org/10.1111/ane.12429.
129. Magalhaes, S., et al., *The EnvIMS Study: Design and Methodology of an International Case-Control Study of Environmental Risk Factors in Multiple Sclerosis.* Neuroepidemiology, 2015. **44**(3): p. 173-81. DOI: https://doi.org/10.1159/000381779.
130. Sundström, P. and J. Salzer, *Vitamin D and multiple sclerosis—from epidemiology to prevention.* Acta Neurol Scand, 2015. **132**(199): p. 56-61. DOI: https://doi.org/10.1111/ane.12432.
131. Rothhammer, V. and F.J. Quintana, *Environmental control of autoimmune inflammation in the central nervous system.* Curr Opin Immunol, 2016. **43**: p. 46-53. DOI: https://doi.org/10.1016/j.coi.2016.09.002.
132. Ascherio, A. and K.L. Munger, *Epidemiology of Multiple Sclerosis: From Risk Factors to Prevention-An Update.* Semin Neurol, 2016. **36**(2): p. 103-14. DOI: https://doi.org/10.1055/s-0036-1579693.
133. Wilker, E.H., et al., *Fine Particulate Matter, Residential Proximity to Major Roads, and Markers of Small Vessel Disease in a Memory Study Population.* J Alzheimers Dis, 2016. **53**(4): p. 1315-23. DOI: https://doi.org/10.3233/jad-151143.
134. Russ, T.C., et al., *Geographical Variation in Dementia Mortality in Italy, New Zealand, and Chile: The Impact of Latitude, Vitamin D, and Air Pollution.* Dementia and Geriatric Cognitive Disorders, 2016. **42**(1-2): p. 31-41. DOI: https://doi.org/10.1159/000447449.
135. Correa, E., V. Paredes, and B. Martínez, *Prevalence of multiple sclerosis in Latin America and its relationship with European migration.* Mult Scler J Exp Transl Clin, 2016. **2**: p. 2055217316666407. DOI: https://doi.org/10.1177/2055217316666407.
136. Sun, H., *Associations of Spatial Disparities of Alzheimer's Disease Mortality Rates with Soil Selenium and Sulfur Concentrations and Four Common Risk Factors in the United States.* J Alzheimers Dis, 2017. **58**(3): p. 897-907. DOI: https://doi.org/10.3233/jad-170059.
137. Tesauro, M., et al., *Incidence of amyotrophic lateral sclerosis in the province of Novara, Italy, and possible role of environmental pollution.* Amyotroph Lateral Scler Frontotemporal Degener, 2017. **18**(3-4): p. 284-290. DOI: https://doi.org/10.1080/21678421.2017.1281961.
138. Hammas, K., et al., *Small-area distribution of multiple sclerosis incidence in western France: in search of environmental triggers.* Int J Health Geogr, 2017. **16**(1): p. 35. DOI: https://doi.org/10.1186/s12942-017-0108-6.
139. Martínez-Solanas, È., et al., *The Geography of the Alzheimer's Disease Mortality in Spain: Should We Focus on Industrial Pollutants Prevention?* Healthcare (Basel), 2017. **5**(4). DOI: https://doi.org/10.3390/healthcare5040089.
140. Karimi, A., et al., *The Relationship between the Amount of Radiation, Relative Humidity, and Temperature with the Risk of Multiple Sclerosis in Isfahan Province, Iran, during the Years 2001-2014.* Journal of Isfahan Medical School, 2017. **35**(427): p. 434-439.
141. Brouwer, M., et al., *Assessment of residential environmental exposure to pesticides from agricultural fields in the Netherlands.* J Expo Sci Environ Epidemiol, 2018. **28**(2): p. 173-181. DOI: https://doi.org/10.1038/jes.2017.3.
142. Michel, L., *Environmental factors in the development of multiple sclerosis.* Rev Neurol (Paris), 2018. **174**(6): p. 372-377. DOI: https://doi.org/10.1016/j.neurol.2018.03.010.

143. Ashtari, F., et al., *Estimated Incidence Rate of Multiple Sclerosis and Its Relationship with Geographical Factors in Isfahan Province between the Years 2001 and 2014.* Int J Prev Med, 2018. **9**: p. 103. DOI: https://doi.org/10.4103/ijpvm.IJPVM_317_16.
144. Weiland, T.J., et al., *Health Outcomes and Lifestyle in a Sample of People With Multiple Sclerosis (HOLISM): Longitudinal and Validation Cohorts.* Front Neurol, 2018. **9**: p. 1074. DOI: https://doi.org/10.3389/fneur.2018.01074.
145. Spencer, P.S., E. Lagrange, and W. Camu, *ALS and environment: Clues from spatial clustering?* Revue Neurologique, 2019. **175**(10): p. 652-663.
146. Logroscino, G. and M. Piccininni, *Amyotrophic Lateral Sclerosis Descriptive Epidemiology: The Origin of Geographic Difference.* Neuroepidemiology, 2019. **52**(1): p. 93-103.
147. Naghshineh, H., et al., *Can environmental factors increase the risk of multiple sclerosis? A narrative review.* Biomedical Research and Therapy, 2019. **6**(12): p. 3513-3517.
148. Salm, A.K. and M.J. Benson, *Increased Dementia Mortality in West Virginia Counties with Mountaintop Removal Mining?* Int J Environ Res Public Health, 2019. **16**(21). DOI: https://doi.org/10.3390/ijerph16214278.
149. Mirza, M., *The Etiology And The Epidemiology of Multiple Sclerosis.* J Clin Pract Res, 2002. **24**(1): p. 40-47.
150. Goncharova, Z.A. and V.A. Baliazin, *[Risk factors of multiple sclerosis development in the population of the Rostov region].* Zh Nevrol Psikhiatr Im S S Korsakova, 2009. **109**(7 Suppl 2): p. 10-5.
151. Jing, J., et al., *Reference value of left atrial diameter of presenile women and geographical factors based on principal component analysis.* Journal of Jilin University Medicine Edition, 2011. **37**: p. 1144-1148.
152. Bakhtiiarova, K.Z. and Z.A. Goncharova, *[Multiple sclerosis in the Bashkortostan Republic and the Rostov region: a comparative epidemiologic study].* Zh Nevrol Psikhiatr Im S S Korsakova, 2014. **114**(2 Pt 2): p. 5-9.
153. Wei, Q., et al., *Clinical features of amyotrophic lateral sclerosis in south-west China.* Amyotroph Lateral Scler Frontotemporal Degener, 2015. **16**(7-8): p. 512-9. DOI: https://doi.org/10.3109/21678421.2015.1069849.
154. Wallin, M.T. and J. Kurtzke, *Multiple Sclerosis; Epidemiology*. 2016. p. 153-160.
155. Schuurman, N., et al., *A proposed methodology to estimate the cumulative life-time UVB exposure using geographic information systems: An application to multiple sclerosis*. Mult Scler Relat Disord, 2012. **2**(1): p. 29-35. DOI: https://doi.org/10.1016/j.msard.2012.07.003.
156. Leibowitz, U., *Multiple sclerosis: Progress in epidemiologic and experimental research: A review.* Journal of the Neurological Sciences, 1971. **12**(3): p. 307-318. DOI: https://doi.org/https://doi.org/10.1016/0022-510X(71)90065-7.
157. Kalafatova, O., *Geographic and climatic factors and multiple sclerosis in some districts of Bulgaria.* Neuroepidemiology, 1987. **6**(3): p. 116-9. DOI: https://doi.org/10.1159/000110106.
158. Van Der Mei, I.A.F., et al., *Regional variation in multiple sclerosis prevalence in Australia and its association with ambient ultraviolet radiation.* Neuroepidemiology, 2001. **20**(3): p. 168-174. DOI: https://doi.org/10.1159/000054783.
159. Sloka, S., et al., *A quantitative analysis of suspected environmental causes of MS.* Can J Neurol Sci, 2011. **38**(1): p. 98-105. DOI: https://doi.org/10.1017/s0317167100011124.
160. Ramagopalan, S.V., et al., *Relationship of UV exposure to prevalence of multiple sclerosis in England.* Neurology, 2011. **76**(16): p. 1410-1414. DOI: https://doi.org/10.1212/WNL.0b013e318216715e.
161. Sun, H., *Temperature dependence of multiple sclerosis mortality rates in the United States.* Mult Scler, 2017. **23**(14): p. 1839-1846. DOI: https://doi.org/10.1177/1352458516688954.
162. Gallagher, L.G., et al., *Lifetime exposure to ultraviolet radiation and the risk of multiple sclerosis in the US radiologic technologists cohort study.* Mult Scler, 2018. **25**(8): p. 1162-1169. DOI: https://doi.org/10.1177/1352458418783343.
163. Norman, J.E., Jr., J.F. Kurtzke, and G.W. Beebe, *Epidemiology of multiple sclerosis in U.S. veterans: 2. Latitude, climate and the risk of multiple sclerosis.* J Chronic Dis, 1983. **36**(8): p. 551-9. DOI: https://doi.org/10.1016/0021-9681(83)90142-x.

164. Monti, M.C., et al., *Is Geo-Environmental Exposure a Risk Factor for Multiple Sclerosis? A Population-Based Cross-Sectional Study in South-Western Sardinia*. PLoS One, 2016. **11**(9): p. e0163313. DOI: https://doi.org/10.1371/journal.pone.0163313.
165. Amram, O., et al., *The use of satellite data to measure ultraviolet-B penetrance and its potential association with age of multiple sclerosis onset.* Mult Scler Relat Disord, 2018. **21**: p. 30-34. DOI: https://doi.org/10.1016/j.msard.2018.02.005.
166. Risberg, G., et al., *Prevalence and incidence of multiple sclerosis in Oppland County - a cross-sectional population-based study in a landlocked county of Eastern Norway*. Acta Neurologica Scandinavica, 2010. **124**(4): p. 250-257. DOI: https://doi.org/10.1111/j.1600-0404.2010.01465.x.
167. Tateo, F., et al., *PM2.5 levels strongly associate with multiple sclerosis prevalence in the Province of Padua, Veneto Region, North-East Italy*. Mult Scler, 2018. **25**(13): p. 1719-1727. DOI: https://doi.org/10.1177/1352458518803273.
168. Gregory, A.C., 2nd, et al., *Multiple Sclerosis disease distribution and potential impact of environmental air pollutants in Georgia*. Sci Total Environ, 2008. **396**(1): p. 42-51. DOI: https://doi.org/10.1016/j.scitotenv.2008.01.065.
169. Iuliano, G., *Geography based hypotheses about Multiple Sclerosis.* Rivista Italiana di Neurobiologia, 2005. **2**: p. 213-220.
170. Heydarpour, P., et al., *Potential impact of air pollution on multiple sclerosis in Tehran, Iran.* Neuroepidemiology, 2014. **43**(3-4): p. 233-8. DOI: https://doi.org/10.1159/000368553.
171. Ashtari, F., et al., *An 8-year study of people with multiple sclerosis in Isfahan, Iran: Association between environmental air pollutants and severity of disease.* J Neuroimmunol, 2018. **319**: p. 106-111. DOI: https://doi.org/10.1016/j.jneuroim.2018.02.019.
172. Bølviken, B., R. Nilsen, and Å. Ukkelberg, *A new method for spatially moving correlation analysis in geomedicine.* Environmental Geochemistry and Health, 1997. **19**(4): p. 143-153. DOI: https://doi.org/10.1023/A:1018410807648.
173. Groves-Kirkby, C.J., et al., *Is environmental radon gas associated with the incidence of neurodegenerative conditions? A retrospective study of multiple sclerosis in radon affected areas in England and Wales.* Journal of Environmental Radioactivity, 2016. **154**: p. 1-14. DOI: https://doi.org/10.1016/j.jenvrad.2015.12.003.
174. Lavery, A.M., et al., *Examining the contributions of environmental quality to pediatric multiple sclerosis.* Mult Scler Relat Disord, 2017. **18**: p. 164-169. DOI: https://doi.org/10.1016/j.msard.2017.09.004.
175. Lavery, A.M., et al., *Urban air quality and associations with pediatric multiple sclerosis.* Annals of Clinical and Translational Neurology, 2018. **5**(10): p. 1146-1153. DOI: https://doi.org/10.1002/acn3.616.
176. Willis, A.W., et al., *Metal emissions and urban incident Parkinson disease: a community health study of Medicare beneficiaries by using geographic information systems.* Am J Epidemiol, 2010. **172**(12): p. 1357-63. DOI: https://doi.org/10.1093/aje/kwq303.
177. Santurtún, A., et al., *[Geographical distribution of mortality by Parkinson's disease and its association with air lead levels in Spain].* Med Clin (Barc), 2016. **147**(11): p. 481-487. DOI: https://doi.org/10.1016/j.medcli.2016.07.022.
178. Finkelstein, M.M. and M. Jerrett, *A study of the relationships between Parkinson's disease and markers of traffic-derived and environmental manganese air pollution in two Canadian cities.* Environ Res, 2007. **104**(3): p. 420-32. DOI: https://doi.org/10.1016/j.envres.2007.03.002.
179. Liu, R., et al., *Ambient Air Pollution Exposures and Risk of Parkinson Disease.* Environ Health Perspect, 2016. **124**(11): p. 1759-1765. DOI: https://doi.org/10.1289/ehp135.
180. Salimi, F., et al., *Associations between long-term exposure to ambient air pollution and Parkinson's disease prevalence: A cross-sectional study.* Neurochem Int, 2020. **133**: p. 104615. DOI: https://doi.org/10.1016/j.neuint.2019.104615.
181. Lee, P.C., et al., *Gene-environment interactions linking air pollution and inflammation in Parkinson's disease.* Environ Res, 2016. **151**: p. 713-720. DOI: https://doi.org/10.1016/j.envres.2016.09.006.

182. Kravietz, A., et al., *Association of UV radiation with Parkinson disease incidence: A nationwide French ecologic study.* Environ Res, 2017. **154**: p. 50-56. DOI: https://doi.org/10.1016/j.envres.2016.12.008.
183. Wei, Y., et al., *Associations between seasonal temperature and dementia-associated hospitalizations in New England.* Environ Int, 2019. **126**: p. 228-233. DOI: https://doi.org/10.1016/j.envint.2018.12.054.
184. Li, C.Y., et al., *Association between air pollution and risk of vascular dementia: A multipollutant analysis in Taiwan.* Environ Int, 2019. **133**(Pt B): p. 105233. DOI: https://doi.org/10.1016/j.envint.2019.105233.
185. Chen, H., et al., *Exposure to ambient air pollution and the incidence of dementia: A population-based cohort study.* Environ Int, 2017. **108**: p. 271-277. DOI: https://doi.org/10.1016/j.envint.2017.08.020.
186. Povedano, M., et al., *Spatial Assessment of the Association between Long-Term Exposure to Environmental Factors and the Occurrence of Amyotrophic Lateral Sclerosis in Catalonia, Spain: A Population-Based Nested Case-Control Study.* Neuroepidemiology, 2018. **51**(1): p. 33-49. DOI: https://doi.org/10.1159/000489664.
187. Tsai, C.P. and C. Tzu-Chi Lee, *Climatic factors associated with amyotrophic lateral sclerosis: a spatial analysis from Taiwan.* Geospat Health, 2013. **8**(1): p. 45-52. DOI: https://doi.org/10.4081/gh.2013.53.
188. Santurtún, A., et al., *Trends in motor neuron disease: association with latitude and air lead levels in Spain.* Neurol Sci, 2016. **37**(8): p. 1271-5. DOI: https://doi.org/10.1007/s10072-016-2581-2.
189. Deluca, H.F. and M.T. Cantorna, *Vitamin D: its role and uses in immunology.* Faseb j, 2001. **15**(14): p. 2579-85. DOI: https://doi.org/10.1096/fj.01-0433rev.
190. Ogen, Y., *Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality.* Science of The Total Environment, 2020. **726**: p. 138605. DOI: https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138605.
191. Bornstein, S.R., et al., *Is There a Role for Environmental and Metabolic Factors Predisposing to Severe COVID-19?* Horm Metab Res, 2020. **52**(07): p. 540-546. DOI: https://doi.org/10.1055/a-1182-2016.
192. Hribar, C.A., P.H. Cobbold, and F.C. Church, *Potential Role of Vitamin D in the Elderly to Resist COVID-19 and to Slow Progression of Parkinson's Disease.* Brain Sciences, 2020. **10**(5): p. 284.
193. Moulton, P.V. and W. Yang, *Air pollution, oxidative stress, and Alzheimer's disease.* J Environ Public Health, 2012. **2012**: p. 472751. DOI: https://doi.org/10.1155/2012/472751.
194. Chin-Chan, M., J. Navarro-Yepes, and B. Quintanilla-Vega, *Environmental pollutants as risk factors for neurodegenerative disorders: Alzheimer and Parkinson diseases*. Front Cell Neurosci, 2015. **9**: p. 124. DOI: https://doi.org/10.3389/fncel.2015.00124.
195. Filippini, T., et al., *Risk of Amyotrophic Lateral Sclerosis and Exposure to Particulate Matter from Vehicular Traffic: A Case-Control Study*. International Journal of Environmental Research and Public Health, 2021. **18**(3): p. 973.
196. Malek, A.M., et al., *Long-term air pollution and risk of amyotrophic lateral sclerosis mortality in the Women's Health Initiative cohort.* Environmental Research, 2023. **216**. DOI: https://doi.org/10.1016/j.envres.2022.114510.
197. Cerin, E., et al., *International Mind, Activities and Urban Places (iMAP) study: Methods of a cohort study on environmental and lifestyle influences on brain and cognitive health.* BMJ Open, 2020. **10**(3). DOI: https://doi.org/10.1136/bmjopen-2019-036607.
198. Crous-Bou, M., et al., *Impact of urban environmental exposures on cognitive performance and brain structure of healthy individuals at risk for Alzheimer's dementia.* Environ Int, 2020. **138**: p. 105546. DOI: https://doi.org/10.1016/j.envint.2020.105546.
199. Ho, H.C., et al., *The associations between social, built and geophysical environment and age-specific dementia mortality among older adults in a high-density Asian city.* Int J Health Geogr, 2020. **19**(1): p. 53. DOI: https://doi.org/10.1186/s12942-020-00252-y.

200. Jung, C.R., Y.T. Lin, and B.F. Hwang, *Ozone, particulate matter, and newly diagnosed Alzheimer's disease: a population-based cohort study in Taiwan.* J Alzheimers Dis, 2015. **44**(2): p. 573-84. DOI: https://doi.org/10.3233/jad-140855.
201. Smargiassi, A., et al., *Exposure to ambient air pollutants and the onset of dementia in Québec, Canada.* Environmental Research, 2020. **190**. DOI: https://doi.org/10.1016/j.envres.2020.109870.
202. Yuchi, W., et al., *Road proximity, air pollution, noise, green space and neurologic disease incidence: a population-based cohort study.* Environmental Health, 2020. **19**(1): p. 8. DOI: https://doi.org/10.1186/s12940-020-0565-4.
203. Chen, G.B., et al., *Long-term exposures to ambient PM1 and NO2 pollution in relation to mild cognitive impairment of male veterans in China.* Environmental Research Letters, 2021. **16**(2). DOI: https://doi.org/10.1088/1748-9326/abde5c.
204. Mortamais, M., et al., *Long-term exposure to ambient air pollution and risk of dementia: Results of the prospective Three-City Study.* Environ Int, 2021. **148**: p. 106376. DOI: https://doi.org/10.1016/j.envint.2020.106376.
205. Ran, J., et al., *Long-term exposure to fine particulate matter and dementia incidence: A cohort study in Hong Kong.* Environ Pollut, 2021. **271**: p. 116303. DOI: https://doi.org/10.1016/j.envpol.2020.116303.
206. Ran, J., et al., *The joint association of physical activity and fine particulate matter exposure with incident dementia in elderly Hong Kong residents.* Environ Int, 2021. **156**: p. 106645. DOI: https://doi.org/10.1016/j.envint.2021.106645.
207. Rhew, S.H., J. Kravchenko, and H.K. Lyerly, *Exposure to low-dose ambient fine particulate matter PM2.5 and Alzheimer's disease, non-Alzheimer's dementia, and Parkinson's disease in North Carolina.* PLoS One, 2021. **16**(7): p. e0253253. DOI: https://doi.org/10.1371/journal.pone.0253253.
208. Sullivan, K.J., et al., *Ambient fine particulate matter exposure and incident mild cognitive impairment and dementia.* Journal of the American Geriatrics Society, 2021. **69**(8): p. 2185-2194. DOI: https://doi.org/10.1111/jgs.17188.
209. Tan, J., et al., *Associations of particulate matter with dementia and mild cognitive impairment in China: A multicenter cross-sectional study.* Innovation (Camb), 2021. **2**(3): p. 100147. DOI: https://doi.org/10.1016/j.xinn.2021.100147.
210. Yitshak-Sade, M., et al., *PM2.5 and hospital admissions among Medicare enrollees with chronic debilitating brain disorders.* Science of The Total Environment, 2021. **755**: p. 142524. DOI: https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.142524.
211. Andersen, Z.J., et al., *Long-term exposure to air pollution and mortality from dementia, psychiatric disorders, and suicide in a large pooled European cohort: ELAPSE study.* Environment International, 2022. **170**. DOI: https://doi.org/10.1016/j.envint.2022.107581.
212. Wang, X., et al., *Association of improved air quality with lower dementia risk in older women.* Proceedings of the National Academy of Sciences of the United States of America, 2022. **119**(2). DOI: https://doi.org/10.1073/pnas.2107833119.
213. Xie, J. and C. Lu, *Is there a casual relation between air pollution and dementia?* Environmental Science and Pollution Research, 2022. DOI: https://doi.org/10.1007/s11356-022-23226-y.
214. Yang, L., et al., *Associations between PM2.5 exposure and Alzheimer's Disease prevalence Among elderly in eastern China.* Environmental Health, 2022. **21**(1): p. 119. DOI: https://doi.org/10.1186/s12940-022-00937-w.
215. Younan, D., et al., *Racial/Ethnic Disparities in Alzheimer's Disease Risk: Role of Exposure to Ambient Fine Particles.* J Gerontol A Biol Sci Med Sci, 2022. **77**(5): p. 977-985. DOI: https://doi.org/10.1093/gerona/glab231.
216. Zhang, H., et al., *Short-term associations between ambient air pollution and emergency department visits for Alzheimer's disease and related dementias.* Environ Epidemiol, 2023. **7**(1): p. e237. DOI: https://doi.org/10.1097/ee9.0000000000000237.
217. Adamczyk-Sowa, M., B. Gębka-Kępińska, and M. Kępiński, *MULTIPLE SCLEROSIS - RISK FACTORS.* Wiad Lek, 2020. **73**(12 cz 1): p. 2677-2682.

218. Bergamaschi, R., et al., *PM2.5 exposure as a risk factor for multiple sclerosis. An ecological study with a Bayesian mapping approach.* Environmental Science and Pollution Research, 2021. **28**(3): p. 2804-2809. DOI: https://doi.org/10.1007/s11356-020-10595-5.
219. Chacko, G., et al., *Heat Exposure and Multiple Sclerosis-A Regional and Temporal Analysis.* International Journal of Environmental Research and Public Health, 2021. **18**(11). DOI: https://doi.org/10.3390/ijerph18115962.
220. Elgabsi, M., et al., *An impact of air pollution on moderate to severe relapses among multiple sclerosis patients.* Multiple Sclerosis and Related Disorders, 2021. **53**. DOI: https://doi.org/10.1016/j.msard.2021.103043.
221. Elser, H., et al., *Anomalously warm weather and acute care visits in patients with multiple sclerosis: A retrospective study of privately insured individuals in the US.* Plos Medicine, 2021. **18**(4). DOI: https://doi.org/10.1371/journal.pmed.1003580.
222. Januel, E., et al., *Fine Particulate Matter Related to Multiple Sclerosis Relapse in Young Patients.* Front Neurol, 2021. **12**: p. 651084. DOI: https://doi.org/10.3389/fneur.2021.651084.
223. Kazemi Moghadam, V., et al., *Association of the global distribution of multiple sclerosis with ultraviolet radiation and air pollution: an ecological study based on GBD data.* Environmental Science and Pollution Research, 2021. **28**(14): p. 17802-17811. DOI: https://doi.org/10.1007/s11356-020-11761-5.
224. Scartezzini, A., et al., *Association of Multiple Sclerosis with PM 2.5 levels. Further evidence from the highly polluted area of Padua Province, Italy.* Mult Scler Relat Disord, 2021. **48**: p. 102677. DOI: https://doi.org/10.1016/j.msard.2020.102677.
225. Vitkova, M., et al., *Association of Latitude and Exposure to Ultraviolet B Radiation With Severity of Multiple Sclerosis An International Registry Study.* Neurology, 2022. **98**(24): p. E2401-E2412. DOI: https://doi.org/10.1212/wnl.0000000000200545.
226. Hedström, A.K., et al., *Association between exposure to combustion-related air pollution and multiple sclerosis risk.* Int J Epidemiol, 2023. DOI: https://doi.org/10.1093/ije/dyac234.
227. Fleury, V., et al., *Geospatial analysis of individual-based Parkinson's disease data supports a link with air pollution: A case-control study.* Parkinsonism Relat Disord, 2021. **83**: p. 41-48. DOI: https://doi.org/10.1016/j.parkreldis.2020.12.013.
228. Cole-Hunter, T., et al., *Long-term air pollution exposure and Parkinson's disease mortality in a large pooled European cohort: An ELAPSE study.* Environ Int, 2023. **171**: p. 107667. DOI: https://doi.org/10.1016/j.envint.2022.107667.
229. Alvarez-Mendoza, C.I., et al., *Assessment of Remote Sensing Data to Model PM10 Estimation in Cities with a Low Number of Air Quality Stations: A Case of Study in Quito, Ecuador.* Environments, 2019. **6**(7): p. 85.
230. Alvarez-Mendoza, C., A. Teodoro, and L. Ramírez-Cando, *Spatial estimation of surface ozone concentrations in Quito Ecuador with remote sensing data, air pollution measurements and meteorological variables.* Environmental Monitoring and Assessment, 2019. **191**. DOI: https://doi.org/10.1007/s10661-019-7286-6.
231. A, R., *Estimation of land surface temperature of dindigul district using landsat 8 data.* International Journal of Research in Engineering and Technology, 2014. **03**: p. 122-126. DOI: https://doi.org/10.15623/ijret.2014.0305025.
232. Mustafa, E.K., et al., *Study for Predicting Land Surface Temperature (LST) Using Landsat Data: A Comparison of Four Algorithms.* Advances in Civil Engineering, 2020. **2020**: p. 7363546. DOI: https://doi.org/10.1155/2020/7363546.
233. Wang, L., Y. Lu, and Y. Yao, *Comparison of Three Algorithms for the Retrieval of Land Surface Temperature from Landsat 8 Images.* Sensors, 2019. **19**(22): p. 5049.
234. Jimenez, J.-C., et al., *Land Surface Temperature Retrieval Methods From Landsat-8 Thermal Infrared Sensor Data.* Geoscience and Remote Sensing Letters, IEEE, 2014. **11**: p. 1840-1843. DOI: https://doi.org/10.1109/LGRS.2014.2312032.
235. Guo, J., et al., *Evaluation of Land Surface Temperature Retrieval from Landsat 8/TIRS Images before and after Stray Light Correction Using the SURFRAD Dataset.* Remote Sensing, 2020. **12**(6): p. 1023.
236. Amani, M., et al., *Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review.* IEEE Journal of Selected Topics in Applied

Earth Observations and Remote Sensing, 2020. **13**: p. 5326-5350. DOI: https://doi.org/10.1109/JSTARS.2020.3021052.
237. Tamiminia, H., et al., *Google Earth Engine for geo-big data applications: A meta-analysis and systematic review.* ISPRS Journal of Photogrammetry and Remote Sensing, 2020. **164**: p. 152-170. DOI: https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.04.001.
238. Wang, M., et al., *An Efficient Framework for Producing Landsat-Based Land Surface Temperature Data Using Google Earth Engine.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020. **PP**: p. 1-1. DOI: https://doi.org/10.1109/JSTARS.2020.3014586.
239. Ermida, S.L., et al., *Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series.* Remote Sensing, 2020. **12**(9): p. 1471.
240. Liu, S., et al., *An analysis of spatial representativeness of air temperature monitoring stations.* Theoretical and Applied Climatology, 2018. **132**(3): p. 857-865.
241. GEE. *Guides - Python Installation*. Available from: https://developers.google.com/earth-engine/guides/python_install-conda.
242. Marty, R. *Iterate over ImageCollection returning pandas dataframe using earth engine & python*. 2017 [cited 2021 April]; Available from: https://gis.stackexchange.com/a/257748.
243. WHO, W.H.O., *World Health Statistics 2022*. 2022. p. 21. DOI: https://doi.org/9789240051157.
244. Rabinovici, G.D., *Late-onset Alzheimer Disease.* CONTINUUM: Lifelong Learning in Neurology, 2019. **25**(1): p. 14-33. DOI: https://doi.org/10.1212/con.0000000000000700.
245. Jankovic, J., *Parkinson's disease: clinical features and diagnosis.* Journal of Neurology, Neurosurgery & Psychiatry, 2008. **79**(4): p. 368-376. DOI: https://doi.org/10.1136/jnnp.2007.131045.
246. Timmons, S., et al., *Dementia in older people admitted to hospital: a regional multi-hospital observational study of prevalence, associations and case recognition.* Age and Ageing, 2015. **44**(6): p. 993-999. DOI: https://doi.org/10.1093/ageing/afv131.
247. Price, D.L., S.S. Sisodia, and D.R. Borchelt, *Genetic Neurodegenerative Diseases: The Human Illness and Transgenic Models.* Science, 1998. **282**(5391): p. 1079-1083. DOI: https://doi.org/doi:10.1126/science.282.5391.1079.
248. INE. *Projections of resident population in Portugal*. 2020 [cited 2023 January]; Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=406534255&DESTAQUESmodo=2&xlang=en.
249. Hardin, A.P., et al., *Age Limit of Pediatrics.* Pediatrics, 2017. **140**(3). DOI: https://doi.org/10.1542/peds.2017-2151.
250. Moran, P.A.P., *Notes on Continuous Stochastic Phenomena.* Biometrika, 1950. **37**(1/2): p. 17-23. DOI: https://doi.org/10.2307/2332142.
251. Anselin, L., *Local Indicators of Spatial Association—LISA.* Geographical Analysis, 1995. **27**(2): p. 93-115. DOI: https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.
252. Getis, A. and J.K. Ord, *The Analysis of Spatial Association by Use of Distance Statistics.* Geographical Analysis, 1992. **24**(3): p. 189-206. DOI: https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.
253. Oliveira, M.F., A.; Cláudia Teodoro, A.; Gonçalves, H. *Spatial Patterns in Neurodegenerative Disease's Hospitalizations in Portugal (2000-2016)*. in *Proceedings of the 9th International Conference on Geographical Information Systems Theory, Applications and Management - GISTAM*. 2023. SciTePress. DOI: https://doi.org/10.5220/0011990100003473.
254. Hou, Y., et al., *Ageing as a risk factor for neurodegenerative disease.* Nat Rev Neurol, 2019. **15**(10): p. 565-581. DOI: https://doi.org/10.1038/s41582-019-0244-7.
255. Piscopo, P., et al., *A Sex Perspective in Neurodegenerative Diseases: microRNAs as Possible Peripheral Biomarkers.* Int J Mol Sci, 2021. **22**(9). DOI: https://doi.org/10.3390/ijms22094423.
256. Javaid, S., et al., *Epidemiology of Alzheimer's disease and other dementias: rising global burden and forecasted trends [version 1; peer review: 1 approved with reservations].* F1000Research, 2021. **10**(425). DOI: https://doi.org/10.12688/f1000research.50786.1.

257. Walton, C., et al., *Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition.* Multiple Sclerosis Journal, 2020. **26**(14): p. 1816-1821. DOI: https://doi.org/10.1177/1352458520970841.
258. Freitas, A., I. Lema, and A. da Costa-Pereira. *Comorbidity Coding Trends in Hospital Administrative Databases*. in *New Advances in Information Systems and Technologies*. 2016. Cham: Springer International Publishing.
259. Appannagari, R.R., *Environmental pollution causes and consequences: a study.* North Asian International Research Journal of Social Science & Humanities, 2017. **3**(8): p. 151-161.
260. Braga, A., et al., *Atmospheric pollution.* J Bras Pneumol, 2006. **32**(Supl 1): p. S5-S11.
261. Fayiga, A.O., M.O. Ipinmoroti, and T. Chirenje, *Environmental pollution in Africa.* Environment, Development and Sustainability, 2018. **20**(1): p. 41-73. DOI: https://doi.org/10.1007/s10668-016-9894-4.
262. Cristina, B.B.G., F. Valentin, and Frank, *Air quality status and trends in Europe.* Atmospheric Environment, 2014. **98**: p. 376-384. DOI: https://doi.org/https://doi.org/10.1016/j.atmosenv.2014.09.017.
263. DGT. *Oficial Administrative Map of Portugal*. [cited 2020 10]; Available from: https://www.dgterritorio.gov.pt/cartografia/cartografia-tematica/caop.
264. CNIG. *Reference Geographic Information*. [cited 2022 06]; Available from: https://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=CAANE.
265. IPMA. *Surface climatologic stations' list*. [cited 2021 March]; Available from: https://www.ipma.pt/pt/oclima/series.longas/list.jsp.
266. Gianquintieri, L., et al., *State-of-art in modelling particulate matter (PM) concentration: a scoping review of aims and methods.* Environment, Development and Sustainability, 2024. DOI: https://doi.org/10.1007/s10668-024-04781-5.
267. Almeida, C.R.d., et al., *Remote Sensing Image-Based Analysis of the Urban Heat Island Effect in Bragança, Portugal.* Environments, 2022. **9**(8): p. 98.
268. Almeida, C.R.d., A.C. Teodoro, and A. Gonçalves, *Study of the Urban Heat Island (UHI) Using Remote Sensing Data/Techniques: A Systematic Review.* Environments, 2021. **8**(10): p. 105.
269. Zhang, G.P. and M. Qi, *Neural network forecasting for seasonal and trend time series.* European Journal of Operational Research, 2005. **160**(2): p. 501-514. DOI: https://doi.org/https://doi.org/10.1016/j.ejor.2003.08.037.
270. Gonsamo, A. and P. D'Odorico, *Citizen science: best practices to remove observer bias in trend analysis.* International Journal of Biometeorology, 2014. **58**(10): p. 2159-2163. DOI: https://doi.org/10.1007/s00484-014-0806-8.
271. Ninneman, M. and D. Jaffe, *Observed Relationship between Ozone and Temperature for Urban Nonattainment Areas in the United States.* Atmosphere, 2021. **12**(10): p. 1235.
272. Ji, Y., et al., *A review of atmospheric benzene homologues in China: Characterization, health risk assessment, source identification and countermeasures.* Journal of Environmental Sciences, 2020. **95**: p. 225-239. DOI: https://doi.org/https://doi.org/10.1016/j.jes.2020.03.035.
273. Sicard, P., et al., *Urban population exposure to air pollution in Europe over the last decades.* Environmental Sciences Europe, 2021. **33**(1): p. 28. DOI: https://doi.org/10.1186/s12302-020-00450-2.
274. Denouel, A., et al., *The role of environmental factors on sporadic Creutzfeldt-Jakob disease mortality: evidence from an age-period-cohort analysis.* European Journal of Epidemiology, 2023. **38**(7): p. 757-764. DOI: https://doi.org/10.1007/s10654-023-01004-5.
275. Bey, I., et al., *Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation.* Journal of Geophysical Research: Atmospheres, 2001. **106**(D19): p. 23073-23095. DOI: https://doi.org/https://doi.org/10.1029/2001JD000807.
276. Billionnet, C., D. Sherrill, and I. Annesi-Maesano, *Estimating the Health Effects of Exposure to Multi-Pollutant Mixture.* Annals of Epidemiology, 2012. **22**(2): p. 126-141. DOI: https://doi.org/https://doi.org/10.1016/j.annepidem.2011.11.004.
277. WHO, W.H.O. *AirQ+*. 2016 [cited 2024 January]; Available from: https://www.who.int/tools/airq.

278. Tang, E.Y.H., L. Robinson, and B.C.M. Stephan, *Dementia risk assessment tools: an update.* Neurodegenerative Disease Management, 2017. **7**(6): p. 345-347. DOI: https://doi.org/10.2217/nmt-2017-0031.