Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2023/01, 36 pàg.
Working Paper 2023/01, 36 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2023/01, 36 pàg.
Working Paper 2023/01, 36 pag.

"Geographical distribution of the COVID-19 pandemic across waves in Spain"

Rosina Moreno and Esther Vayá

UBIREA
Research Institute of Applied Economics

AQR
Regional Quantitative Analysis Research Group

UBIREA Institut de Recerca en Economia Aplicada Regional i Pública
UNIVERSITAT DE BARCELONA
WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

AQR
Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group
WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

IREA Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. For that reason, IREA Working Papers may not be reproduced or distributed without the written consent of the author. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of IREA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

Abstract

This paper pursues a deep insight in the evolution of the spatial distribution of the pandemic in the Spanish provinces along the six waves. Through the use of spatial exploratory techniques, we observe that the geographical spread of the COVID-19 has been changing considerably so that the conclusions obtained for specific points in time are not transferable to other moments of the pandemic. We also take into consideration the changes in the determinants of the spread of the pandemic across waves while considering the possibility of external effects across provinces through the estimation of spatial regressions.

JEL Classification: H75, R58.

Keywords: COVID-19, Pandemic, Spatial analysis, Temperature, Non-climate factors, Spanish provinces.

Rosina Moreno: AQR-IREA, University of Barcelona. Email: rmoreno@ub.edu

Esther Vayá: AQR-IREA, University of Barcelona. Email: evaya@ub.edu

Acknowledgements

We acknowledge financial support from Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Project 2020PANDE00060.

1. Introduction

Since the end of 2019, when the SARS-CoV2 virus was detected and the COVID-19 outbreak appeared in Wuhan (China), till mid-June 2022, more than five billion people have been infected around the world and more than six million people have died due to this pandemic worldwide. According to the International Monetary Fund (IMF), the average global Gross Domestic Product (GDP) dropped by 3.9% from 2019 to 2020, making it the worst economic downturn since the Great Depression. The pandemic outbreak caught the population and governments off guard, resulting in the collapse or the near-collapse of the health system worldwide and a slew of containment and mitigation measures, such as the obligation on wearing masks in inner spaces, increased testing, contact tracing, lengthy lockdowns, quarantines, and mobility restrictions, that had an uneven impact on the spread of the virus.

The plague, however, did not impact the world uniformly. Even within Europe, some countries were struck far more severely than others. Excess mortality in the United Kingdom exceeded 59,000 in the first six months of 2020. In such period of time, COVID-19 killed 41,000 people in Spain, while it almost reached 36,000 and 21,000 deaths in Italy and France, respectively. Belgium, Italy, Spain, and the United Kingdom had the greatest incidence relative to the population. On the other hand, many Central and Eastern European nations were spared. Deaths in the Baltic States, Bulgaria, Czechia, Hungary, and Slovakia were lower than the previous five-year average. In the first half of 2020, Denmark and Germany similarly had no total excess mortality.

The economic literature analysing the pandemic has been profuse in the last two years, dealing with the study of its determinants as well as its impact, to a higher extent. Additionally, several works have tried to analyse the spatial distribution of the pandemic as well as the determinants behind the heterogeneity in such patterns. Our paper focus on the latter and pursues a deep insight in the spatial distribution of the pandemic in the Spanish provinces along the six waves, comparing its evolution, while taking into consideration the changes in the determinants of the spread of the pandemic across waves.

In an international sphere, several papers have tried to analyse the spatial distribution of the pandemic (Murugesan et al. 2020 in India, Mota et al. 2021 in Brazil, Ehlert 2021 for Germany and Guliyev 2020 for China, just to mention some). In particular for the Spanish case, Rodríguez et al. (2022) focus on the causes of the incidence of COVID-19 based on socioeconomic variables as well as spatial interaction through the consideration of possible spatial dependencies between Spanish regions, whereas Páez et al. (2021) analyze the driving factors from a geospatial point of view in Spanish provinces. In all these papers, the authors focus on specific moments in time and do not consider the evolution in time of the pandemic. In our paper, we try to go beyond and analyse the spatial distribution of the disease across the different waves, with the idea that it has been changing considerably so that the conclusions obtained for specific points in time are not transferable to other moments of the pandemic. Our results are obtained through the use of spatial exploratory techniques and spatial regressions for the case of the Spanish provinces along the six waves of the pandemic.

The outline of the paper is as follows. Section 2 overviews the extant literature. In section 3 we present the data and some basic descriptives, whereas section 4 offers the main patterns of the spatial distribution of the pandemic across the Spanish provinces along six waves through a spatial exploratory analysis. Section 5 presents the main findings on the determinants of such spatial distribution through regression analyses whereas section 6 concludes.

2. Literature review

Given the relevance of COVID-19, there is already an extensive body of literature studying the spreading dynamics of the virus and its effects on the economy, even if research started only two years ago. With the specific purpose of investigating the spatial distribution of the disease, and through the consideration of spatial techniques, since the beginning of the pandemic several papers have focused on different countries and cities, such as Murugesan et al. (2020) in India, Mota et al. (2021) in Brazil, Ehlert (2021) for Germany, Guliyev (2020) for China, among others. Most of these papers use spatial exploratory techniques, basically through disease mapping, clustering, and hotspot analysis, that is, the computation of spatial dependence tests such as the Moran's I test, and the identification of spatial clusters through the local versions of such tests. In addition, when analysing the determinants of the spatial pattern observed, those papers tend to estimate spatial models such as the SAR, SEM, SAC and SDM ones (Elhorst, 2014), all of them considering implicitly the presence of spatial autocorrelation in the estimation (see Fatima et al, 2021, for a review of papers using spatial techniques).

Although the findings in those papers as specific for the areas and time period under consideration, there are several issues which are worth noting. In the study of Amdaoud et al. (2021) analysing the heterogeneity of the spread of the COVID-19 pandemic across 125 regions in 12 European countries using spatial models, the authors find that spatial clusters exist, and that income and public health policies are able to explain disparities across regions. Interestingly, Sun et al (2021) find a greater spatial inequality in COVID-19 mortality in the UK than in mortality from other causes. Cao et al. (2020) examine the case-fatality rate in 209 countries and territories worldwide finding a significant correlation with population size, which may imply the healthcare strain and lower treatment efficiency in countries with large populations. Ehlert (2021) find evidence that the infections and deaths are positively and significantly related to median age, the number of people working in care of the elderly, early cases since the beginning of the pandemic and population density in 401 German counties up to June 2020. For the City of New York, Yang et al (2021) find that COVID-19 case rates are positively related to racial minority groups, household size and the elderly population, and negatively related to the number of teleworkers. Fonseca-Rodriguez et al. (2021) with a study in Sweden, found that the virus is associated with population density, the proportion of immigrants and the proportion of people aged 65+ years old.

In particular for the Spanish case and with the use of spatial econometric techniques, Pérez et al. (2021) study the spread of COVID in Spain based on spatial VAR models. Maza and Hierro (2022) focus on the distribution of COVID-19 in Madrid during the first wave of the pandemic, finding that those territories with more mobility and also those with a higher level of tourism, had a higher incidence. Romero and Arroyo (2022) point out that the pandemic was specially suffered in urban areas with higher population density and higher levels of contamination. Gullón et al. (2022) propose a study on the temporal trends in the association between area-level deprivation and COVID incidence whereas Rodríguez et al. (2022) focus on the causes of the incidence of COVID obtaining that there is spatial correlation across Spanish territories and that both socioeconomic variables, as well as those of spatial interaction, are significant. Páez et al (2021) use spatial SURE models obtaining that higher incidence is associated with higher GDP per capita and presence of mass transit systems, lower population density and percentage of older adults in March-April 2020. In the same sense, Briz-Redón and Serrano-Aroca (2020) do not obtain consistent evidence of a relationship between the accumulated number of cases in the provinces of Spain and temperature values between February and March 2020.

Reviewing the literature on the spread of the COVID-19, we observe that since the beginning of the Covid-19 pandemic, researchers from all over the world have been investigating the spatial distribution of the disease in conjunction with its main determinants, including, but not limited to, climatology and environmental conditions, demographic factors and socioeconomic factors.

First, viruses transmit more easily depending on climate conditions such as average temperatures and rainfalls (Dalziel et al., 2018). In this regard, previous studies have suggested a correlation between weather and COVID-19 pandemic in a similar way that it occurs with other viral infectious diseases such as influenza (Ficetola and Rubolini, 2021; Ma et al., 2020a, 2020b; Tosepu et al., 2020). According to Wang et al (2020) and Sajadi et al (2020), average temperatures between 5 °C and 11 °C and relative humidity levels between 50% and 70% would be the climatic characteristics of the areas in which the incidence of the COVID-19 was higher. According to this literature, an increase in temperatures and air humidity levels associated with the arrival of Spring in the Northern hemisphere could significantly reduce the transmission and spread of the coronavirus. Nonetheless, other studies have reported contradictory results showing that meteorological conditions may not be associated with COVID-19 in terms of absolute humidity (Shi et al., 2020) or temperature (Jamil et al., 2020; Xie and Zhu, 2020). According to these last authors, the previous results showing evidence for a correlation between meteorological factors with COVID-19 transmission was likely to be an artifact, reflecting the pathways of spread. In fact, several of these previous studies have been performed considering only meteorological factors, without accounting for non-meteorological variables that might be also decisive.

Second, other factors that must be taken into account when considering the spread of such a pathogen are demographic factors. Initially, higher incidences were attributed to ageing. Indeed, several studies have reported age and underlying diseases as the most important risk factors for death by COVID-19 (Morley and Vellas, 2020; Liu et al., 2020; Onder et al., 2020). Iyanda et al. (2022) carry out a broad study of 175 countries to study the health and social determinants of the spread of covid, and through the use of a spatial regressions find that age plays a central role in determining the spread all over the world. Considering that Spain has been for decades ranked in the world top-ten for having the highest life expectancy (World Data, 2022), this is an issue to consider in the Spanish case. Indeed, the ageing population in Spain could be behind the remarkably hard impact of the disease in such a country.

Another determinant of the spread of the virus is the agglomeration of people in the same area. Indeed, population density is accounted by most of the studies wanting to analyse the factors that influenced the spread of COVID-19, observing that it is a telling factor for risk exposure (Bhadra et al., 2021; Coccia, 2020; Ahmadi et al., 2020; Pequeno et al., 2020; Wong and Li, 2020; Mollalo et al., 2020; Bayode et al., 2022). Large and dense European cities were regarded as the focus for the spread of coronavirus. However, although density may have shaped the early outbreaks, it did not have influence in its related mortality over time (Carozzi et al, 2020).

Linked to large agglomerations, we do not only consider the influence of population density but also its greater connectivity (Coelho et al., 2020). Highly connected cities were among the first ones being hit by the virus. This connectivity can be thought both with extra-regional agents as well as internally. As for extra-regional connectivity, we must consider that according to the World Tourism Organization, Spain was the country in the world with the second highest international tourist arrivals in 2019 (UNWTO, 2019). For this reason, considering the entry of tourism can be extremely valuable when trying to estimate the causes behind COVID-19 diffusion. As for the connectivity within a region, or even in cities, public transportation has been considered to be related to the spread of contagious diseases (Wang et al 2020). Paéz et al (2021) obtain that the presence of mass transit systems in the province implies a clear positive impact on the diffusion of the disease, given they are cauldrons of social contact.

A further determinant of the diffusion of the virus is economic wealth, although there is not consensus in the sign of its effect. According to previous literature, wealthier regions also tend to concentrate more activities that produce non-traded goods, which would imply that they remain more active even during lockdowns. Also, less wealthy regions have a higher proportion of workers in manual occupations who cannot telework, and have more difficulties complying with shelter-in-place orders. In this respect, Mena et al. (2021) find that mortality rates in young people in Chile are lower in high-income municipalities than in low-income municipalities.

Although not considered in our paper, other determinants that could explain the uneven geography of the COVID-19 pandemic are institutional factors, for instance, the formal institutional quality across European regions that may imply different capacity to effectively implement measures to prevent and combat the pandemic (Rodríguez-Pose and Burlina, 2021), sociological factors such as the tendency to meet with friends and family in celebrations (Rodríguez-Pose and Burlina, 2021) or the vaccination level in the last three waves, different health systems that can influence the capacity to detect and treat outbreaks (Bauer et al 2020; Liang et al 2020; Ahmed et al 2020). Although we do not include these determinants in our study, the fact that we are considering the set of provinces in Spain make that the differences in these variables are not as important as if we were considering regions across different countries.

Most of the papers surveyed above focused on the first wave of the pandemic or in a particular moment/month of it. Contrarily, we have not found research comparing the evolution of the pandemic across waves. We try to fill in this gap with two main objectives. First, to compare the uneven spatial distribution across the six waves, taking into account exploratory spatial techniques. Second, to study whether the determinants of the diffusion of the pandemic have evolved along time or if they have maintained across waves. Indeed, identifying the spatio-temporal variation of COVD-19 in Spain and the main drivers of its transmission is of utmost relevance to help public authorities to contain its spread and promote recovery.

3. Data and descriptive analysis: The geography of the six waves of the COVID-19

3.1 Data and variables

Incidence of the COVID-19

The data about the incidence of the COVID-19 are taken from the Instituto de Salud Carlos III.[1] We use two variables: number of detected cases (positive diagnosis of active infection) as well as the number of deaths. In both cases, the data are computed over 100.000 inhabitants. Although the data of hospitalized cases and emergency rooms (ER) admissions are also available, we do not use the in this analysis since the results were very similar to the ones offered by the two variables finally included. As for the number of detected cases, they consider the number of reported cases confirmed with a positive diagnostic test for active infection (PDIA) as established in the Early Detection Strategy, surveillance, and control of COVID-19 and, also the

[1] Instituto de Salud Carlos III: https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos (last Access 24th November 2022).

cases notified before May 11 that required hospitalization, admission to the ER or died with a clinical diagnosis of COVID 19, according to the current case definitions at all times.

As for the date considered, in the case of the number of detected cases, the date of diagnosis is used, and in its absence, the date of declaration to the community (see Appendix A for a more detailed explanation of the key date). As for deaths, the date of death is considered. The data are gathered for each wave, using the following information:

- 1st wave: from the beginning till 26/06/2020 (peak 26/03)
- 2nd wave: from 27/06/2020 till 10/12/2020 (peak 4/11)
- 3rd wave: from 11/12/2020 till 16/03/2021 (peak 27/01)
- 4th wave: from 17/03/2021 till 22/06/2021 (peak 26/04)
- 5th wave: from 23/06/2021 till 14/10/2021 (peak 27/07)
- 6th wave: from 15/10/2021 till 10/03/2022 (peak 21/01)

All the information is given at the provincial-level for Spain spanning from February 2020 to March 2022, as signaled above.

Determinants of the COVID-19

As for the variables considered as potential determinants of the incidence of the COVID-19, we extract them from several sources. The group of variables proxying for the climate conditions are taken from the AgriCast Resources Portal of the Joint Research Centre (JRC) of the European Commission.[2] We collect the information of the temperature (maximum, minimum and average, all of them averaged each month) as well as the rain fallen. Two variables proxy for the economic wealth of the region and it is taken from the National Institute of Statistics (INE), namely the GDPpc as well as the number of enterprises in the province (Business). As for the existence of agglomeration of population we consider population density as well as the share of population living in cities with more than 20.000 inhabitants (and more that 50.000 and more than 100.000 inhabitants), all of them extracted from the INE. Mobility of people is proxied with several variables: the number of travellers staying in hotels and the number of nights as well as the number of travellers and nights in rural hotels (all of them from the INE), the number of flight passengers (extracted from AENA), and finally, the presence of mass transit systems (subway) in any of the cities in the region. We also take into consideration the demographic structure by measuring the share of people over 70 years old. Finally, as a proxy of the dynamicity of the economy and the corresponding movement of goods and services, we use the international trade (exports + imports) over GDP as well as over population, extracted from the Institute of Foreign Commerce (ICEX). All the variables are analyzed at the provincial level for Spain and for the six waves. Table A1 in the Appendix online presents de definition, frequency and source of each variable. Although many variables vary across the different waves, the ones referring to the GDP and population structure are maintained throughout the analysis.

[2] https://agri4cast.jrc.ec.europa.eu/dataportal/ (last Access 24th November 2022).

3.2 Description of the COVID-19 incidence in Spanish provinces

The interest of this study focuses on the analysis of the incidence of the COVID-19 (number of cases per 100,000 inhabitants and number of deaths per 100,000 inhabitants). As commented above, one of the purposes of this article is to analyze the evolution of the Pandemic along the six waves in Spain (see Tables 1 and 2). Temporarily speaking, some substantial differences are worth commenting.

Table 1. Temporal evolution of COVID19 in Spain. Number of cases per 100.000 inhabitants.

Variable	Obs	Mean	Min	Max
First wave incidence	50	564.822	69.492	1722.071
Second wave incidence	50	3241.663	854.351	5539.773
Third wave incidence	50	3025.169	814.921	5549.373
Fourth wave incidence	50	1094.762	231.113	2408.844
Fifth wave incidence	50	2395.46	1490.866	4015.11
Sixth wave incidence	50	13685.67	5448.053	21489.65

Table 2. Temporal evolution of COVID19 in Spain. Number of deaths per 100.000 inhabitants

Variable	Obs	Mean	Min	Max
First wave Death	50	69.83	4.07	261.27
Second wave Death	50	55.00	8.84	147.72
Third wave Death	50	58.90	11.40	145.13
Fourth wave Death	50	11.25	1.84	27.98
Fifth wave Death	50	13.75	4.94	37.71
Sixth wave Death	50	34.41	10.84	81.15

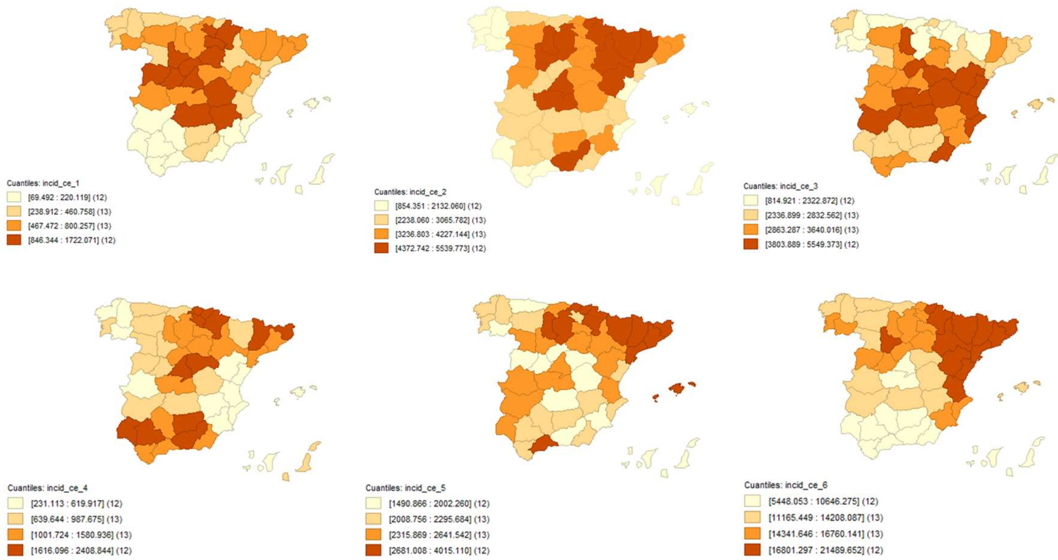
The first wave presents an interestingly low incidence, probably due to the low ability in detecting the pathogen. Afterwards, incidence has always grown between waves except in the fourth one, that presents an intriguingly low incidence corroborating the efficiency of vaccines as well as the restrictions still imposed. Finally, during the last wave the highest peak ever reached throughout the entire pandemic in Spain was observed probably due to more relaxed restrictions and Christmas celebrations.

In relation to the number of deaths, it is observed that the greatest number occurred during the first wave. Likewise, from the fourth wave a very pronounced change is observed, with a lower number of deaths throughout the period (going from a value of 69.83 deaths per 100,000 inhabitants in the first wave to a value of 11.25 in the fourth). Similar to the number of cases, deaths also grew in the fifth and especially in the sixth wave, although the impact of the vaccines significantly reduced the magnitude of the problem (if we take into consideration the number of deaths in relation to the number of cases).

4. Exploratory spatial analysis of the incidence of the COVID-19 in Spain: Analysis over time

We turn now to analyse the regional distribution of the incidence of the COVID-19 across the provinces in Spain, for the six consecutive waves. For the number of cases (Figure 1), in general terms, we observe that the pattern of the spatial distribution had important changes across waves.

Figure 1. Maps of the number of cases of COVID19 per 100.000 inhabitants in Spanish provinces by waves



Regarding the first wave, we can see that the regions with the highest incidence of COVID-19 are located more in the center-north of the peninsula. The two Castillas, Madrid and Barcelona, in addition to the Basque-speaking territories, with La Rioja, would specifically be the geographical areas where COVID-19 hit the hardest (special mention is required to Álava, Ávila, Salamanca, Segovia, Soria, Ciudad Real, Cuenca, La Rioja, and Madrid). In the second wave we can observe four areas with high values of the incidence: i) the province of Palencia as the epicenter of the greatest intensity and the contiguous provinces of Burgos and Valladolid and the western area of Castilla y León; ii) a second area that would present a fairly high intensity in the territories of Aragón, Navarra and the Basque Country; iii) a third one with its epicenter in Madrid; and iv) finally, a zone of strong intensity in the south, surrounding the province of Almería.

In the third wave we could observe a central belt that would cross the Iberian Peninsula from east to west and that extends towards the south-east and towards the northwest. In this third wave, the Valencian Community enters for the first time in a record of very high intensity, which it had not had in the two previous waves. In the fourth wave, the greatest intensity could be found in 4 different geographical areas. Firstly, we could find it in the region of Catalonia, practically in its entirety. The second area with the greatest intensity would be found in the Basques peaking area of the Basque Country and Navarra, which presents a positive correlation with the territories surrounding that area. A third area of strong intensity would once again be Madrid, once again presenting a positive correlation of intensity with respect to its neighbors. Finally, for the first time, the region of Andalusia would appear for the first time since the beginning of the pandemic as one of the regions with the highest intensity of cases, in this fourth wave.

In the fifth wave we observed a fairly marked spatial distribution pattern. It is the northern area, with the exception of the peninsular northwest area, the area that has a more pronounced incidence. In the sixth wave, the geospatial distribution pattern changes again and this time it is a triangular area in the northeast of Spain that clearly receives a higher incidence of cases (special attention is required to all provinces of Catalonia, their neighbors in Aragon, Valencia community and Bask Country). We can also observe that for the first time the province of Madrid is not among the provinces with the highest incidence (just the opposite), as was the case in all the previous waves.

In conclusion, regarding the incidence of COVID-19 in Spanish provinces, the territorial distribution changed along the waves. In this sense, Figure 2 collects the correlation matrix across the number of cases in the six waves. The differences between waves detected from the comparison of the maps are reinforced with this figure, in which no high correlations are detected in the number of cases across waves. Despite this, it is worth highlighting the greater similarities observed in the spatial distribution between the second and fourth waves and, to a lesser extent, between the second wave and the fifth and sixth waves or between the fifth and sixth waves themselves. In turn, it can be seen how the spatial distribution of the number of cases in the third wave is the one with the least similarity compared to the rest of the waves.

Figure 2. Correlation matrix among number of cases by waves

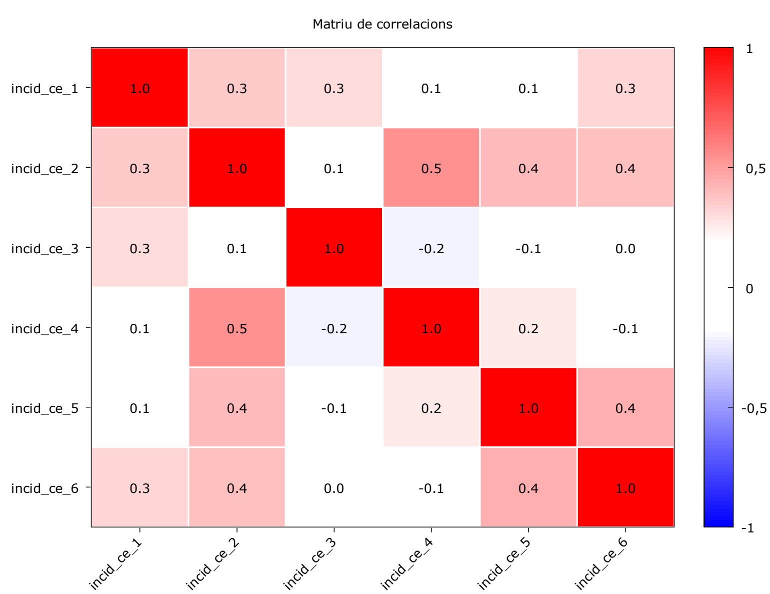


Figure 3 shows the quantile maps of the variable Death, while Figure 4 shows the correlation matrix between recorded deaths across waves. Finally, Figure 5 presents the correlation matrix between the number of cases and number of deaths across the different waves. The following conclusions can be obtained. First, there seems to be greater homogeneity between neighboring provinces in terms of the number of cases than in terms of deaths, which can be the result of the fact that the contagion of the disease is due to a process of transmission between human beings (and, therefore, their proximity is decisive). Second, the correlation across waves are higher than those observed in terms of the number of cases, indicating more similarities in the spatial distribution between waves in this variable than with the number of cases. In turn, it is worth noting the high similarities in the spatial scheme between the deaths of the second and sixth waves, as well as between the deaths of the second and fourth and second and fifth waves (similar to what happened with the number of cases). The correlation between the first and the third wave would also be remarkable.

Figure 3. Maps of the number of deaths due to COVID19 per 100.000 inhabitants in Spanish provinces by waves

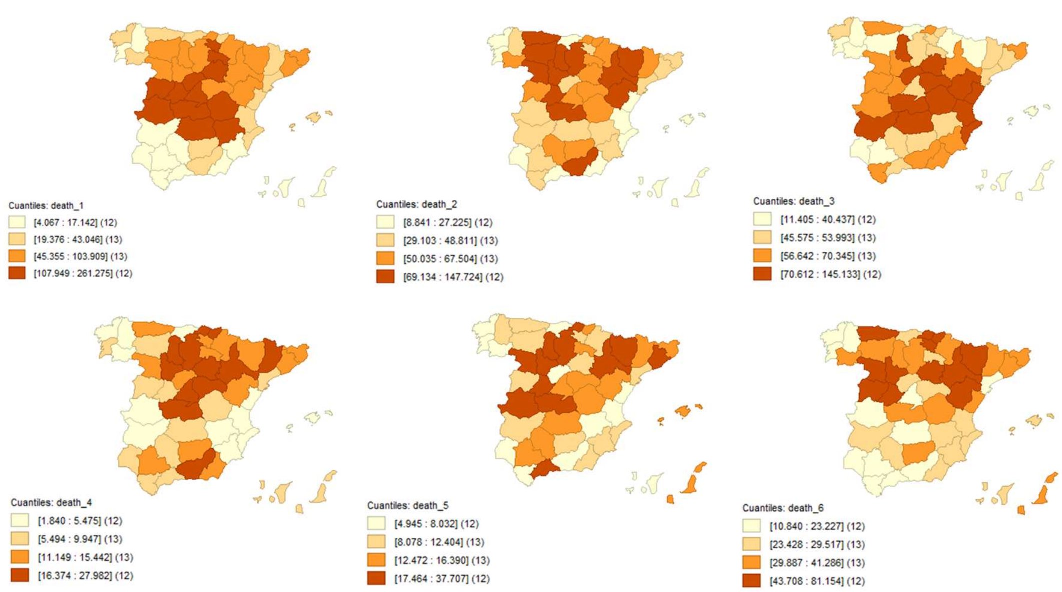


Figure 4. Correlation matrix among number of deaths by waves

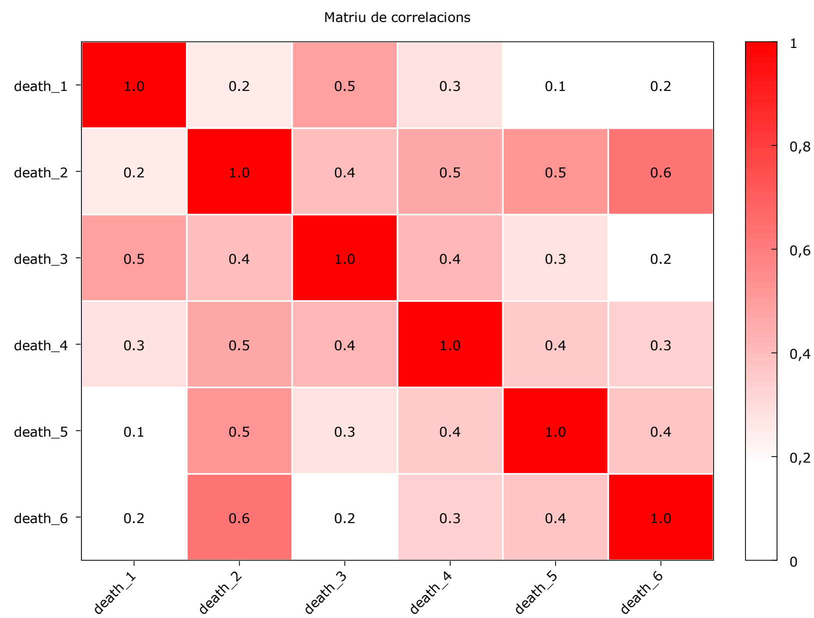
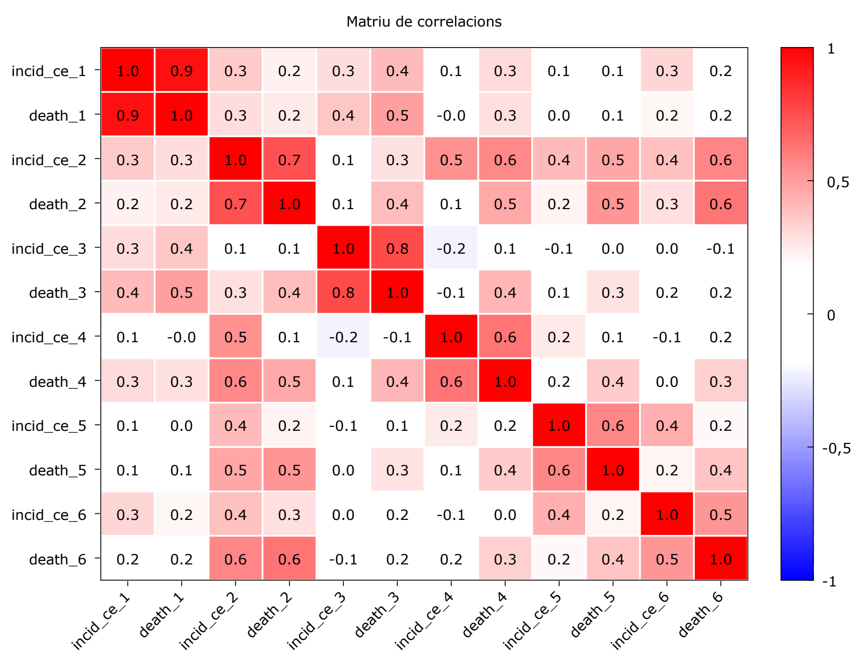


Figure 5. Correlation matrix among Cases and Deaths by waves



Third, the pattern of spatial behavior of the number of deaths is quite similar to the one of the number of cases during the first three waves (correlation of 0.9 in the first wave, 0.7 in the second and 0.8 in the third), although during the last three waves more differences were observed (correlation of 0.5 in the third wave). Thus, for example, in the case of the fifth wave, while the cases were concentrated to a greater extent in the northwestern part (especially Catalonia and part of Aragon), extending through Navarra and the Basque Country, the number of deaths took place mainly in the strip that extends from the northwest to the southwest, affecting Barcelona, Aragon, part of Castilla la Mancha and part of Andalusia (especially Málaga, Seville, Córdoba and Jaén). Finally, in the case of the sixth wave (when there is an upturn in deaths, but without ever reaching the values of the first wave), the largest number of deaths are concentrated in practically all the provinces of the Eastern strip (Catalonia, Valencian region and Aragon) extending through Castilla-León. On the other hand, the highest number of deaths is in the Northern strip (especially Aragon, the Basque Country, Asturias and Castilla León). In this sense, it is worth noting provinces such as Tarragona or Pontevedra (with high cases, but low deaths), the case of Jaén (with low cases, but high deaths) or the case of Madrid, which has come to stand out for the high values of incidences and deaths in the five waves to do so due to their low values in the two variables. A possible explanation for this lower correlation between cases and deaths in the last waves, especially in the sixth, could be in the fact that, throughout the year 2021 and especially in the last wave, the vaccination process was more advanced and, as a consequence, the high number of incidents was not necessarily accompanied by a high number of deaths (in the fourth wave the percentage of the population vaccinated with the complete regimen did not exceed 21% globally, while in the fifth and sixth waves such percentage was already 70% and 82.5%, respectively). Finally, it should be said that both in terms of the number of cases and deaths, Galicia and the Islands present very low values throughout the six waves (with some exceptions in the case of the Balearic Islands and the Palmas in the last two waves).

As observed in the maps, we detect the potential presence of a positive spatial autocorrelation process among different provinces: in all the six waves of the COVID-19, provinces with high covid incidence were surrounded by provinces with high incidence while provinces with low incidence were surrounded by provinces with low incidence. This conclusion in favor of homogeneity in the behavior of the two variables analyzed between neighboring provinces is now tested with the global Moran's I test (Tables 3 and 4). We use two weight matrices: one based on the first-order physical contiguity criterion and another based on the inverse of the distance that separates the centroids of each province.

Table 3. Moran's I test (standardized). Number of cases per 100.000 inhabitants.

Z-Moran (p-value)	Binary contiguity	Inverse distance
First Wave Incidence	5,22 (0,002)	5,20 (0,002)
Second Wave Incidence	5,04 (0,002)	6,02 (0,002)
Third Wave Incidence	4,17 (0,002)	5,55 (0,002)
Fourth Wave Incidence	5,19 (0,002)	3,57 (0,004)
Fifth Wave Incidence	4,63 (0,002)	5,43 (0,002)
Sixth Wave Incidence	8,74 (0,002)	10,11 (0,002)

Note: P-value in brackets.

Table 4. Moran's I test (standardized). Number of deaths per 100.000 inhabitants.

Z-Moran (p-value)	Binary contiguity	Inverse distance
First Wave Death	5,06 (0,002)	4,40 (0,008)
Second Wave Death	2,60 (0,016)	3,09 (0,01)
Third Wave Death	1,80 (0,052)	3,45 (0,008)
Fourth Wave Death	3,03 (0,008)	3,59 (0,006)
Fifth Wave Death	1,54 (0,064)	0,97 (0,150)
Sixth Wave Death	3,44 (0,006)	3,57(0,004)

Note: P-value in brackets.

Starting with the number of cases, it can be seen that, regardless of the matrix used, for each of the waves, we reject the null hypothesis of no spatial autocorrelation and therefore confirm that indeed, as we showed in the maps, the COVID-19 pandemic followed a clear pattern of positive spatial autocorrelation. Additionally, the autocorrelation seems to be more intense when the concept of neighborhood is extended and the inverse distance matrix is used instead of the first-order physical contiguity matrix, a fact that reinforces the amplitude of the diffusion. Likewise, it is important to highlight that the autocorrelation is more intense in the last wave (where the number of incidents is also higher), as could already be observed in the map, in which there clearly seems to be a division of the country into two parts: high number of cases in the Northeast and fewer cases in the Southwest.

When the analysis is repeated, but this time for the case of deaths, it is observed that, with the exception of the first wave, the null hypothesis of no spatial autocorrelation is rejected in practically all the scenarios but with a test value less than that obtained for the number of cases. In fact, in the case of the number of deaths in the fifth wave, the hypothesis of non-autocorrelation is not rejected, concluding that the distribution in space of such variable is random, in the sense that it does not seem to respond to any pattern of geographic proximity. Likewise, it should be noted that, while the autocorrelation was much stronger in the last wave for the number of cases, the greatest spatial association for deaths was observed in the first wave (which was also observed in the map).

Despite its significance, the Global Moran's I test itself has some limitations since it only reveals a global behavioral pattern, but a variety of local issues can arise. To check for the abovementioned limitation, we next carry out a LISA analysis wave by wave (Figures 6, 7, 8, and 9).

Figure 6. Local Moran's I. Number of cases by waves. Physical contiguity matrix.

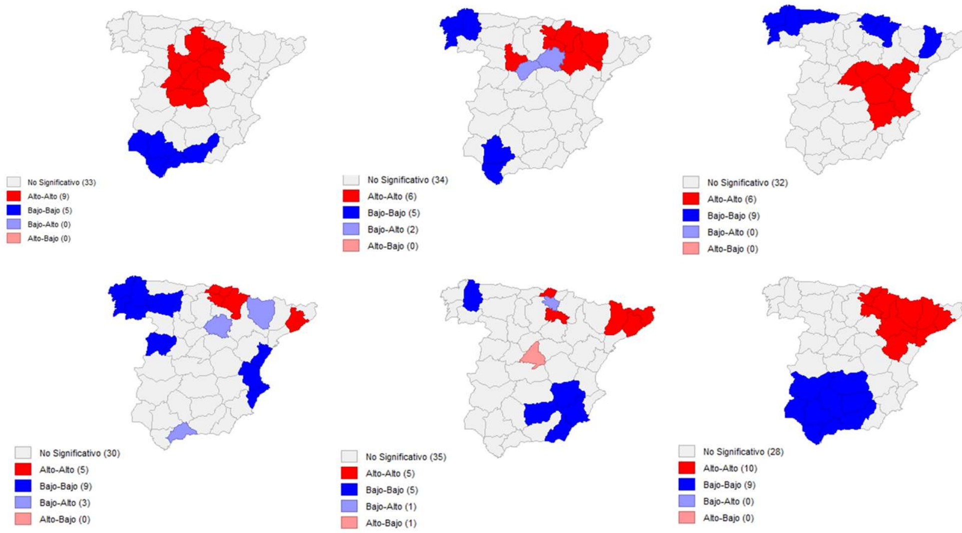


Figure 7. Local Moran's I. Number of cases by waves. Inverse Distance matrix.

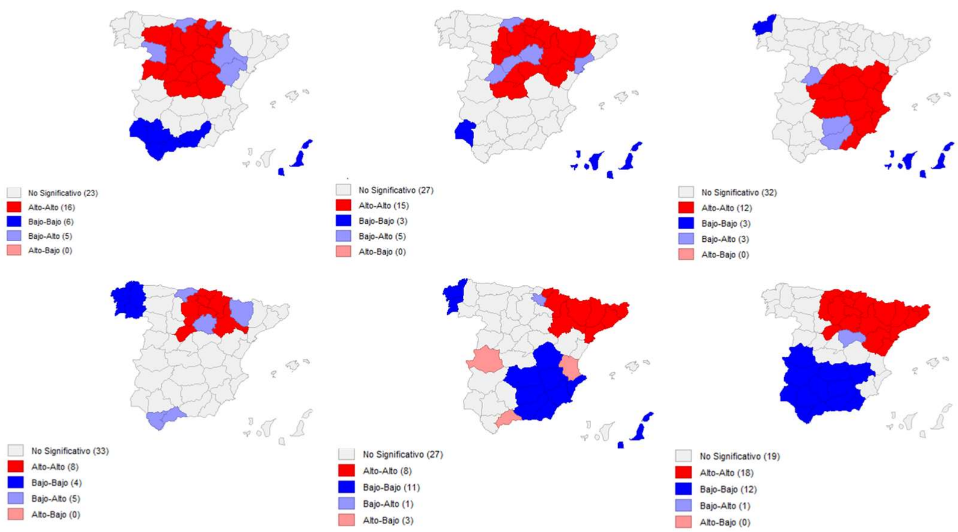


Figure 8. Local Moran's I. Number of deaths by waves. Physical contiguity matrix.

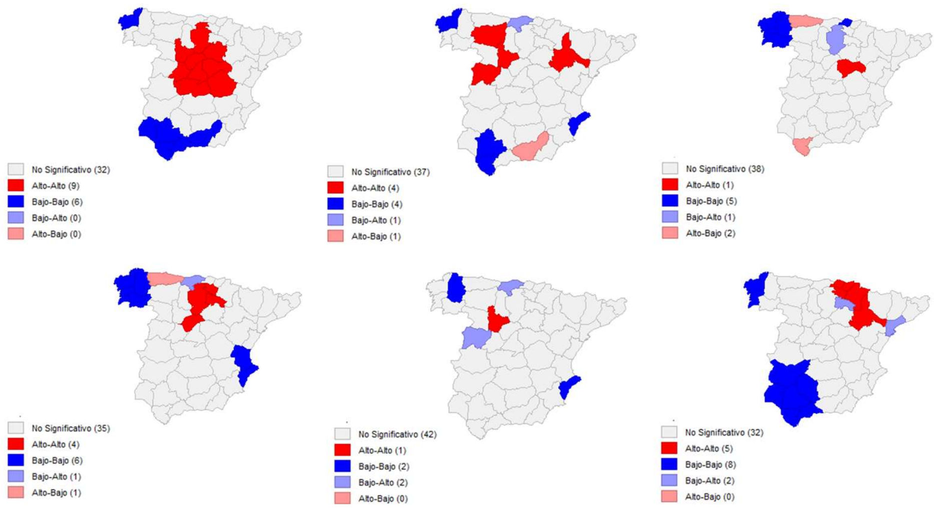
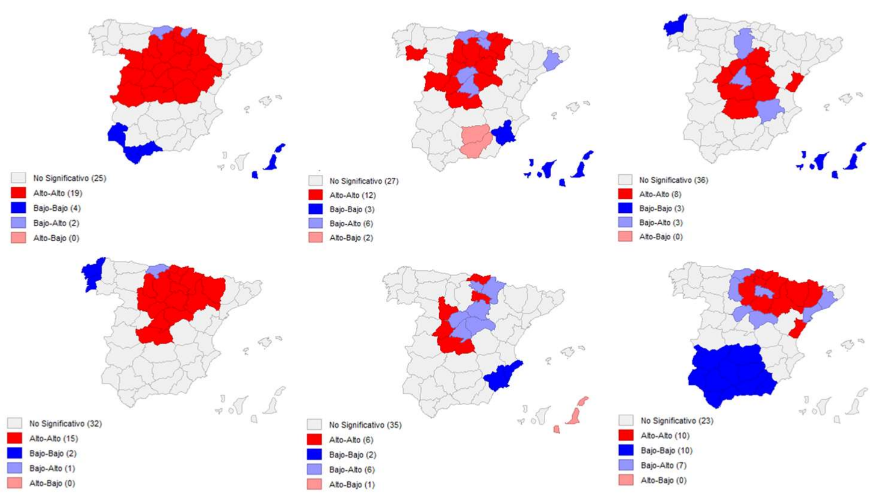


Figure 9. Local Moran's I. Number of deaths by waves. Inverse Distance matrix.



Regardless of the matrix used, we observe that there are different clusters in the distribution of the number of cases. During the first wave, centric provinces represented clusters with the highest values of COVID-19 incidence, whereas southern provinces represented low values clusters. In the second wave, north-eastern provinces constituted clusters with the greatest number of cases, whereas southern and western provinces represented clusters with the lowest ones. Interestingly, two provinces, Soria and Segovia, presented relatively low values in contrast to the provinces around. During the third wave, the center-eastern provinces were clusters with the highest incidence values, whereas some provinces in the north comprised clusters with the lowest ones. In the fourth wave, the north-eastern provinces conformed clusters with the highest number of cases, whereas the north-western and center-eastern provinces composed a low-value cluster. Curiously, three provinces did not belong to the high-values cluster: Huesca, Soria and Malaga, whose values were rather small. During the fifth wave, the north-eastern provinces had the greatest incidence, while the north-western and south-eastern provinces had the lowest one. Interestingly, two provinces did not belong to the surrounding clusters: Madrid and Alava. Finally, in the sixth wave, north-eastern provinces represented clusters with high number of cases, and southern provinces were clusters of low cases. All in all, COVID-19 expanded unevenly between Spanish provinces over the six observed waves. In particular, both in the first and last wave, the southern provinces appear as clusters with low values as far as the number of cases is concerned, while the opposite occurs with the central provinces in the first wave and the northeastern provinces in the last. In turn, at various moments in time, the Galician provinces and their contiguous ones appear as clusters with low incidence values.

When the analysis is repeated for the variable number of death people, the following results are observed. First, the number of provinces that do not end up showing a significant autocorrelation scheme is greater than in the number of cases, a result that is also in line with the lower intensity of autocorrelation at a global level obtained with the Moran's I test. Second, the spatial scheme changes along time. It can be seen that while, in the first wave, the clusters of high and low values detected when analyzing the number of cases are very similar to those detected in the number of deaths, in the sixth wave this is not the case. Thus, while in the number of cases, practically all the provinces of Andalusia together with Badajoz were clusters with low values, for the number of deaths, the clusters are centered on the Andalusian provinces most located in the southwest, also highlighting the Galician provinces of La Coruña and Pontevedra. Similarly, while practically all the provinces of Catalonia, Aragon, Navarra and neighboring Basque Country stood out for being clusters with high values in the case of COVID cases, the number of clusters with high values of deaths was reduced exclusively to Zaragoza, Navarra and the Basque Country. In turn, Tarragona and La Rioja were spatial outliers, with significantly lower values of deaths than their neighbors.

It is interesting to note that, more markedly than in the case of the incidence variable, in practically all the waves, the Galician provinces (especially Pontevedra and La Coruña) appear as clusters with significantly low values of deaths. Finally, it should be noted that when the inverse distance matrix is used and the concept of neighborhood is extended, it is observed that the territorial extension of the clusters and outliers is greater than with the first order contiguity matrix.

5. Method and main findings

5.1. Method

Our regression analysis is based on a set of ordinary least square regressions, in which the dependent variable is one of the two variables proxying for the incidence (number of cases and deaths per 100000 inhabitants) and the explanatory variables are the main determinants (climate, agglomeration, ageing, wealth and connectivity). Our full model is the following:

$$\begin{aligned} Incidence_i = {} & \beta_0 + \beta_1 Climate_i + \beta_2 Agglomeration_i + \beta_3 Ageing_i + \beta_4 Wealth_i \\ & + \beta_5 Connectivity_i + u_i \end{aligned}$$

where *i* refer to the Spanish provinces in a given wave, so that we run six different regressions, one for each wave. Given that multicollinearity problems appear among regressors, we end up with a different model in each wave in order to avoid such problem and find the model that fits the best the variability of the two dependent variables. Additionally, since we have found that the distribution of both the number of cases and deaths in the Spanish provinces suffer from the presence of a spatial autocorrelation process, we check to what extent this process is also present in our regressions. We test it through the use of the Moran's I test for the existence of a spatially correlated disturbance term in the regressions, and if this is the case, we run the robust Lagrange Multiplier tests to infer which spatial model is more appropriate in each case, either the spatial lag or the spatial error models (Anselin, 1988).

5.2 Main findings

As a step prior to the regression analysis, we proceed with a brief comparative analysis of the existing correlation between our two incidence variables and the different potential determinants. Figures 10 and 11 show the results of the aforementioned correlation matrices.

Figure 10. Correlation matrix among Number of cases and its potential determinants by waves.

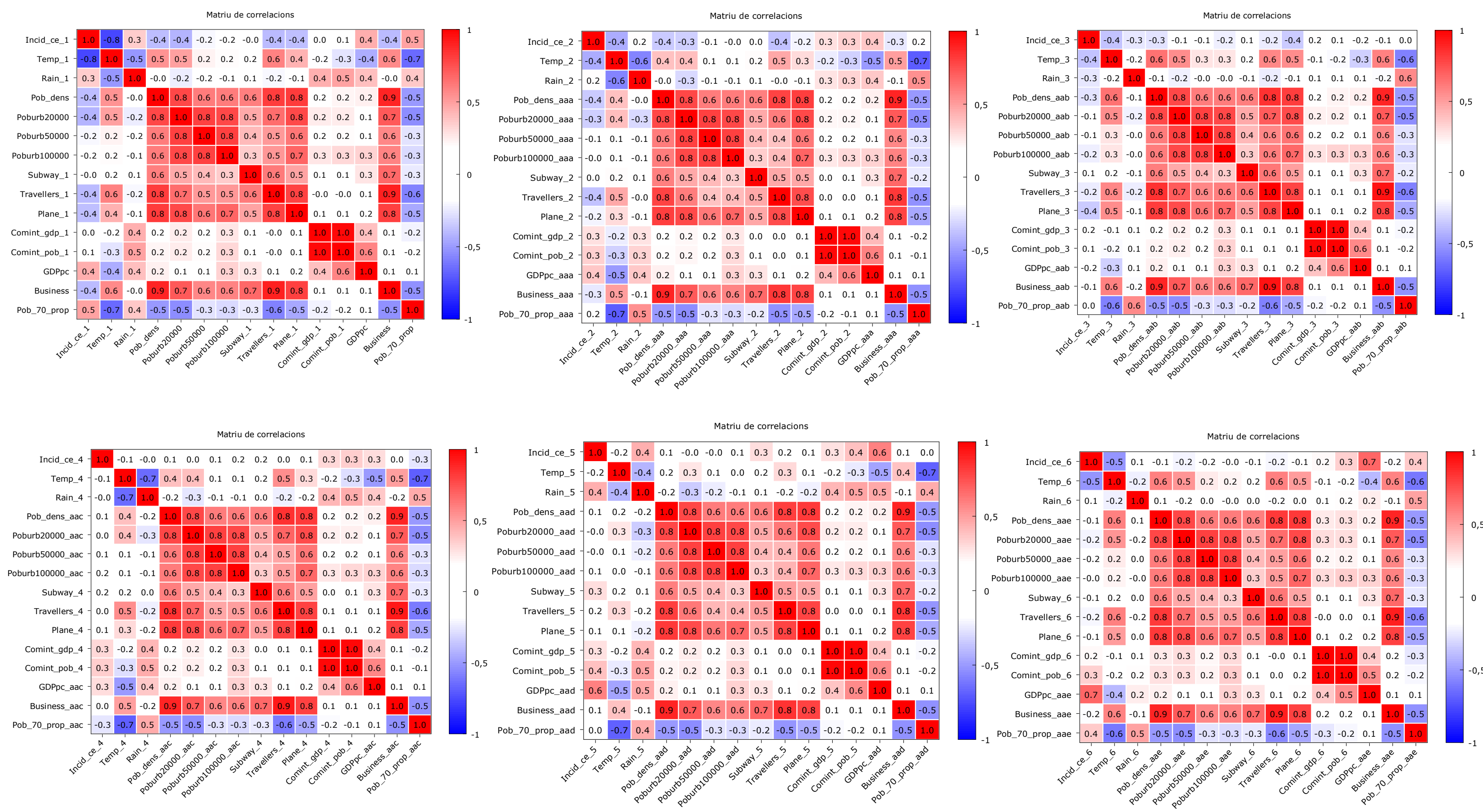
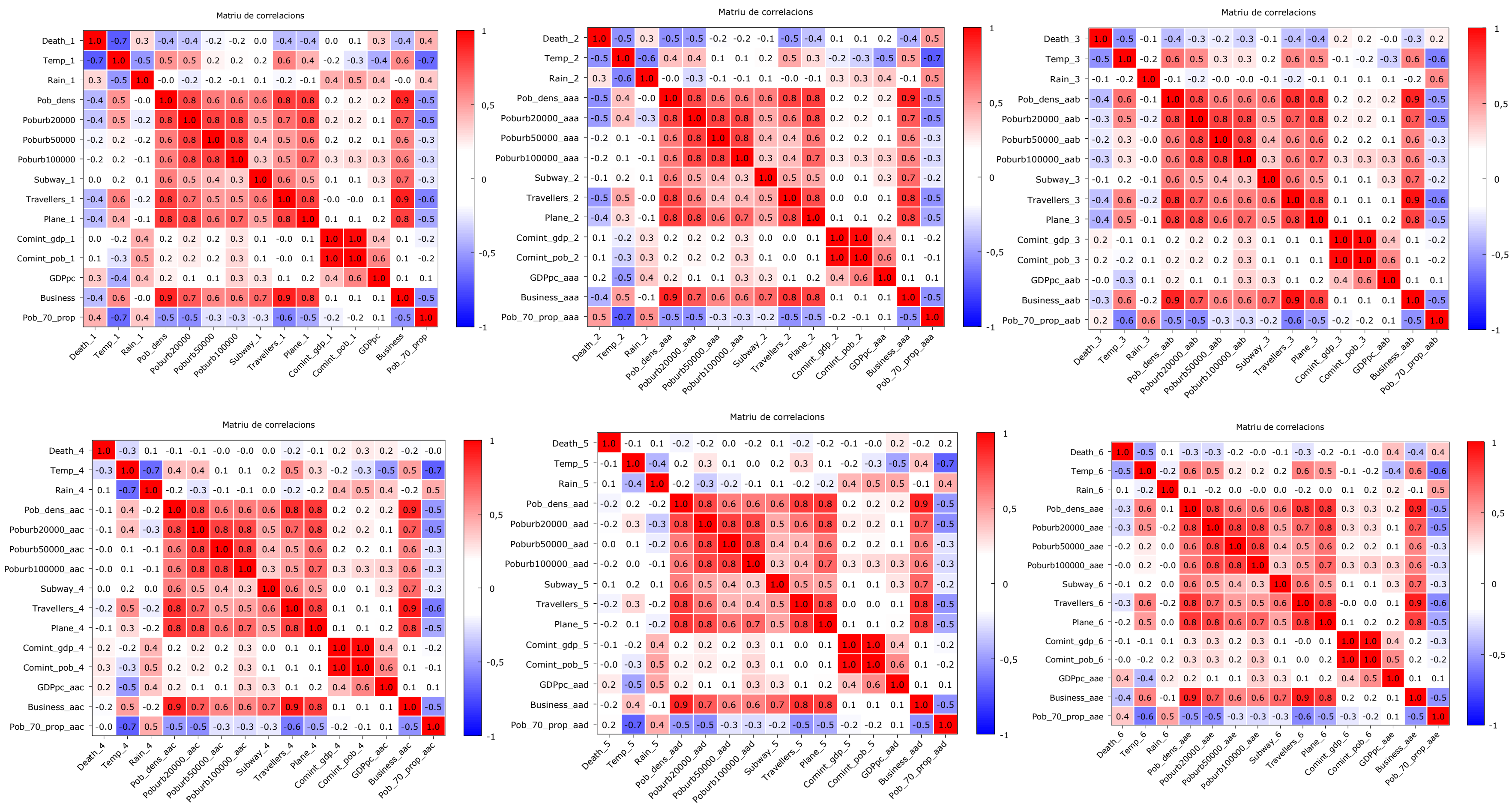


Figure 11. Correlation matrix among Number of deaths and its potential determinants by waves.



As a first conclusion, we observe a significant and negative correlation between the number of cases and the temperature in the first three waves and in the sixth, being especially intense in the first wave (in fact, this is the variable with the highest correlation with the number of cases). In this way, especially at the beginning of the pandemic, clearly the provinces with a higher average temperature (especially Andalusian provinces, the islands and Murcia and Alicante) presented clearly low case rates and vice versa, a fact that would be in line with the arguments that virus inactivation occurs at high temperatures. On the other hand, neither in the fourth nor in the fifth wave is this correlation significant (as can be seen by observing that the southern provinces of Spain began to show high incidence values). However, there is no clear pattern between the number of cases and the other climatological variable analysed, rainfall. Thus, the correlation is low and, moreover, changing in terms of sign.

Second, with respect to the variables that capture the potential level of agglomeration (Pob_dens, Poburb20000, Poburb50000 and Poburb100000), only the significant and negative correlation between the variable population density and the number of cases is noteworthy, and only during the first three waves (especially in the first two and not significant in the last three). In this way, from the beginning of the pandemic (March 2020) to the end of the third wave (March 2021), the provinces with the highest (lowest) population density were those that, contrary to expectations, had the lowest (highest) number of cases presented in average terms. This result would be in line with the fact, as was the case with temperature, the areas with the least incidence in the first two waves correspond to the areas with the highest population density located in the areas of the Mediterranean coast of Valencia and Murcia, the Andalusian Mediterranean area, the Andalusian South Atlantic area and the coast of the Canary and Balearic Islands.

The demographic variable of the percentage of the population over 70 years of age offers only a significant and positive correlation observed in the first wave and in the sixth is noteworthy, a result closely linked to the lower incidence and lower age population.

With regard to the mobility of people (travellers, plane and subway), a significant and negative correlation is detected with the number of cases, but only in the first three waves, not in the last three. Again, this apparently counterintuitive result (the greater the number of travelers, the lower the potential spread of the virus), would be explained by the concentration of the greatest number of travelers in hotels and rural accommodation in coastal areas (areas of the Mediterranean coast of Valencia and Murcia, the Andalusian Mediterranean area, the Andalusian South Atlantic area and the coast of the Canary and Balearic Islands), all regions with lower levels of incidence in the first waves. The variables related to international trade and merchandise mobility (comint_gdp and comint_pob) do not show any conclusive results or notable correlation with the incidence variable.

Regarding GDPpc, a significant and positive correlation is detected in the first two waves but especially in the last two. In the case of the first two waves, this correlation would be reflecting both the fact of the lower incidence in the areas with the lowest GDPpc (provinces of Andalusia, Extremadura and the Canary Islands) and the higher incidence in regions with high GDPpc such as Madrid, Navarra, La Rioja and the Basque Country. In the case of the last two waves, the lowest incidence would be repeated in the southern part of the country and the highest incidence concentrated in Catalonia, Valencia, Navarra and the Basque Country (high GDP pc values). With regard to the business variable, there does not seem to be a clear behavior beyond a discreet and negative correlation in the first two waves.

As a summary, we can conclude that the variables that seem to have a greater correlation with the number of cases are, on the one hand, temperature (negative correlation in the three waves and in the last one, being especially intense in the first one) , GDP pc (positive correlation in the first two waves and in the last two, especially in the sixth), Population density (negative correlation in the first three waves) and Proportion of population older than 70 years (positive correlation in the first and last wave). These results would be evidencing above all the lower incidence in the southern part of the country and on the islands (areas of high temperature, high population density, low levels of GDP pc and low proportion of elderly population) and the highest levels of incidence in the first and sixth wave in the area of Catalonia, Valencia, Navarra and the Basque Country (areas of high GDP pc). Finally, and reviewing the determinants by waves, the number of cases seems to have little correlation with all the potential determinants analysed.

When the analysis is repeated for the Death variable, the conclusions are very similar. The only aspect worth noting is the slightly lower intensity of the positive correlation with GDPpc (especially in the last wave) and the practically non-existence of a significant correlation between the number of deaths and the potential determinants neither in the fourth nor in the fifth wave.

Following the conclusions of the analysis above on the correlation between our main determinants and the two variables of incidence, we have considered only one proxy for each determinant, in particular the one that was observed to have a higher correlation with the incidence of the pandemic. This implies the introduction of temperature as a proxy for climate conditions, population density as a proxy for agglomerations, ageing for demographic factors, GDPpc for economic wealth and the existence of subway as a proxy for connectivity.

Tables 5 and 6 present the results of the regression analysis for the number of cases and deaths of COVID-19, respectively. The bottom part of the tables presents the different tests which allow to check for the presence of spatial autocorrelation in the regressions, namely the Moran's I test as well as the Lagrange Multiplier tests. The tests are applied using a weight matrix based on the inverse of the distance between the centroids of each pair of provinces, which present the advantage of allowing to take the islands into consideration. The following conclusions can be drawn.

Table 5. Regression of number of cases of COVID-19 per 100,000 inhabitants over main determinants

	(1) 1st wave	(2) 1st wave	(3) 2nd wave	(4) 2nd wave	(5) 3rd wave	(6) 3rd wave	(7) 4th wave	(8) 4th wave	(9) 5th wave	(10) 5th wave	(11) 6th wave	(12) 6th wave
Pob_70	0.052*						-0.059**		-0.008		0.027***	
	(0.030)						(0.024)		(0.008)		(0.009)	
Pob_dens	-0.399***		-0.233***		-0.166***		-0.078		-0.056*			
	(0.108)		(0.056)		(0.057)		(0.074)		(0.029)			
GDPpc	2.148***		1.016***				1.027**		0.655***		1.099***	
	(0.493)		(0.493)				(0.391)		(0.136)		(0.171)	
Subway	0.062*		0.035*		0.044***				0.018**			
	(0.032)		(0.019)		(0.021)				(0.009)			
Temperature		-4.475***		-1.585***		-0.4691***		-0.509		-0.314		-0.698***
		(0.531)		(0.454)		(0.167)		(0.508)		(0.251)		(0.164)
Constant	-14.87***	17.643***	-1.367	12.593***	8.597	8.924***	-2.049	8.427***	1.426	8.674	-1.988	11.150***
	(4.813)	(1.381)	(2.830)	(1.332)	(0.234)	(0.351)	(3.781)	(1.368)	(1.326)	(0.775)	(1.717)	(0.382)
N	50	50	50	50	50	50	50	50	50	50	50	50
Akaike	96.94	83.31	43.61	52.09	46.32	45.06	73.88	79.49	-32.02	-13.14	-3.73	16.60
Moran's I	3.25***	3.05***	5.56***	5.19***	5.98***	8.18***	5.92***	6.27***	1.75*	3.65***	4.35***	8.49***
LM-Lag	7.91***	0.46	26.08***	10.03***	18.68***	0.95	0.00	0.55	1.06	0.65	15.76***	9.70***
LM-Error	0.97	0.71	0.26	15.18***	5.91***	0.97	2.09	1.39	0.83	8.28***	1.54	0.01

Number of cases as dependent variable. Standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < 0.01$

Table 6. Regression of number of deaths of COVID-19 per 100,000 inhabitants over main determinants

	(1) 1st wave	(2) 1st wave	(3) 2nd wave	(4) 2nd wave	(5) 3rd wave	(6) 3rd wave	(7) 4th wave	(8) 4th wave	(9) 5th wave	(10) 5th wave	(11) 6th wave	(12) 6th wave
Pob_70	0.048		0.039				-0.036		0.008		0.020	
	(0.040)		(0.024)				(0.033)		(0.022)		(0.019)	
Pob_dens	-0.384***		-0.349***		-0.169***		-0.171*		-0.196**		-0.121**	
	(0.124)		(0.087)		(0.056)		(0.102)		(0.078)		(0.058)	
GDPpc	2.329***		0.806**				1.168**		0.675**		1.035***	
	(0.651)		(0.397)				(0.536)		(0.358)		(0.304)	
Subway			0.046*						0.050*			
			(0.027)						(0.025)			
Temperature		-4.929***		-2.729***		-0.826***		-1.446		-0.255		-0.854***
		(0.708)		(0.621)		(0.188)		(0.639)		(0.578)		(0.233)
Constant	-18.83***	16.602***	-3.635	11.705***	4.715	5.714***	-7.92	6.445***	-3.611	3.222	-6.712	5.514***
	(6.303)	(1.840)	(3.881)	(1.823)	(0.245)	(0.393)	(5.193)	(1.721)	(3.491)	(0.787)	(2.939)	(0.542)
N	50	50	50	50	50	50	50	50	50	50	50	50
Akaike	124.99	112.01	75.35	83.44	64.72	56.44	105.61	102.44	64.76	70.40	48.70	51.40
Moran's I	4.31***	3.86***	3.02***	4.99***	6.45***	5.68***	4.31***	4.49***	-0.13	1.05	1.91*	2.52*
LM-Lag	7.90***	0.624	7.42***	0.01	16.25***	4.15**	0.66	0.30	0.01	2.21	0.68	0.36
LM-Error	1.01	0.928	1.57	1.54	5.78***	0.01	0.01	1.44	0.13	2.17	0.12	0.01

Number of cases as dependent variable. Standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < 0.01$

First, temperature is a clear determinant for both variables proxying for the incidence of the COVID-19, in particular in the first three waves as well as in the sixth. In those waves, the inclusion of temperature would take away the relevance of the other variables, so that we decided to run two separate models: a first one with the socio-economic determinants and a second one with the temperature variable alone. According to the Akaike criteria, in the first and third waves, the model including only the average temperature in the province provides a better model than the one considering the socio-economic variables. These findings show that virus is more active in low temperatures, although the relevance and magnitude of the influence is much lower in the waves in which most people are vaccinated. This pattern is observed both in the number of cases and deaths due to the virus.

As for the effect of demographic conditions, we observe that it is only significant and positive during the first and last wave in the case of the number of cases. It seems, therefore, that only in those waves, a relatively large population of older adults translate into higher diffusion rates of the virus but not in their severity. In the rest of waves, the parameter is not significant or even negative, such as in the 4th wave both for the number of cases and deaths, which could be related to two facts. First, the stringent social distance restrictions that Spanish provinces suffered were more strict in the case of elderly people living in old-care residences, while the rest of the population started having much lower restrictions since the vaccination in the 4th wave was highly extended. Second, the fact that adults over 70 tend not to work and move less that younger adults, with the lower likelihood to be infected.

With respect to the level of agglomeration, population density is a clear determinant of the case and death rates in the Spanish provinces, presenting a significant negative parameter in almost all waves, although more clearly for the first three waves. This result goes against the expectation that contact rates are higher in more dense areas and, as a consequence, it would be positively correlated with the transmission of COVID-19. A possible reasoning would be in line with what was obtained in other works (Noland, 1995; Paez et al, 2021) in which this negative correlation is explained by the so-called risk compensation, that is, a situation where people adapt their behavior according to the perceived level of risk, so that they become more careful when the perceived risk is higher, as in dense areas, and vice versa.

The variable proxying for economic wealth, GDPpc, presents a clear positive relationship with the number of cases and deaths in all the waves except for the third one. This general finding would point to the fact that wealthier provinces may remain more active, in relative terms, even during a lockdown due to the presence of more non-traded activities. This would be in addition to the fact that wealthier provinces can be in networks of global cities with a greater "potential to be further ahead in the trajectory of the pandemic" (Páez et al, 2021, pp 399).

Our last determinant refers to the connectivity of the province. When we introduce the variable that proxies for mass transit systems in a province, the relationship with both the number of cases and deaths is clearly positive in the first three waves and in the fifth one. This would point to the cauldrons of social contact represented by this kind of transportation methods.

The tests presented at the bottom of the tables to conclude on the presence of a process of spatial autocorrelation in the models offer several conclusions. Firstly, we observe that for the models considering the socio-economic determinants of the pandemic, there is a need to consider a spatial model including the spatial lag of the dependent variable. This is especially true in the three first waves for the two variables proxying for the incidence. In the case of the number of deaths, the last three waves do not present any spatial pattern, whereas for the number of cases, the fourth wave should include a spatial error process. Second, in the models that only consider meteorological conditions, we observe that spatial autocorrelation does not seem to be a problem. This is probably due to the fact that the temperature variable is already picking up the spatial pattern observed in the distribution of the incidence of the COVID-19.

All in all, we conclude that even after considering the effect of the different socio-economic determinants of the incidence of the COVID-19 in the Spanish provinces, there seems to exist a process of spatial autocorrelation in the distribution of the disturbance terms in most of the models considering these socio-economic drivers. In other words, there exist interactions across provinces that should be considered explicitly in the model. In addition, a consistent conclusion is that the spatial model that would better consider such interaction across provinces is the spatial autoregressive model (SAR), as pointed out by the Lagrange Multiplier tests. Table 7 presents the results of the estimation of these spatial lag models for the socio-economic determinants, which introduce a spatial lag of the dependent variable as an additional regressor.

Table 7. Spatial regressions of number of cases and deaths of COVID-19 per 100,000 inhabitants over main determinants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Number of cases				Number of deaths		
	1st wave	2nd wave	3rd wave	6th wave	1st wave	2nd wave	3rd wave
Pob_70	0.029			0.013*	0.026	0.025	
	(0.027)			(0.007)	(0.036)	(0.023)	
Pob_dens	-0.311***	-0.144***	-0.072***		-0.266**	-0.299***	-0.109**
	(0.096)	(0.046)	(0.043)		(0.107)	(0.078)	(0.043)
GDPpc	1.376***	0.545**		0.405***	1.338**	0.474	
	(0.491)	(0.247)		(0.143)	(0.614)	(0.368)	
Subway	0.056**	0.018	0.019*			0.034	
	(0.028)	(0.015)	(0.016)			(0.024)	
Spatial lag	0.507***	0.641***	0.733***	0.764***	0.575***	0.422**	0.687***
	(0.146)	(0.127)	(0.111)	(0.095)	(0.148)	(0.168)	(0.126)
Constant	-10.18**	-2.085	2.389***	-2.052	-11.242	-1.842	1.687***
	(4.474)	(2.251)	(0.923)	(1.316)	(5.813)	(3.483)	(0.549)
N	50	50	50	50	50	50	50
Akaike	92.27	30.13	26.72	-21.56	118.43	72.04	46.15
LR test	6.67***	15.48***	21.60***	19.83***	8.56***	5.68***	20.57***

Number of cases as dependent variable. Standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < 0.01$

From the estimation of the spatial models, we observe that the conclusions extracted with respect to the different determinants of the incidence of the COVID-19 are maintained now. In addition, we observe that a higher incidence in the neighbouring provinces has a positive influence on the incidence of the pandemic in a given province. This shows that we can model the incidence of the COVID-19 as an interprovincial contagion process.

6. Conclusions

This paper focuses on the incidence of the COVID-19 in the Spanish provinces along the six waves of the pandemic, with two main objectives. First, we compare the uneven spatial distribution of the incidence, both in terms of the number of cases and the number of deaths per 100,000 inhabitants, across the six waves considering exploratory spatial techniques. Second, we study whether the determinants of the diffusion of the pandemic have evolved along time or if they have maintained across waves.

As for the evolution in time, we observe that the pattern of the spatial distribution of the number of cases and deaths had important changes across waves, with only some similarities between the second and the fourth waves for the number of cases and between the second and sixth in deaths. And, in any case, the correlation across waves is higher in terms of deaths.

As for the spatial distribution of the pandemic across the Spanish provinces, a pattern of spatial association between neighboring provinces is obtained, greater for the rate of cases than deaths, which can be the result of the fact that the contagion of the disease is due to a process of transmission between human beings, with proximity being an essential issue. Also, the pattern of spatial behavior of the rate of deaths is quite similar to the one of the cases during the first three waves, while more differences are observed during the last three waves. A possible explanation is that, throughout the year 2021 and especially in the last wave, the high level of vaccination in the population had a clear incidence in the reduction of the rate of deaths, so that a high number of cases was not necessarily accompanied with a high number of deaths.

Another spatial feature in the distribution of the pandemic in the Spanish provinces is the presence of a clear generalized pattern of positive spatial autocorrelation between nearby provinces, both in terms of cases and deaths, although their intensity is higher in the former. At a local level, it seems that COVID-19 expanded unevenly between Spanish provinces over the six observed waves. In particular, both in the first and last wave, the southern provinces appear as clusters with low values as far as the number of cases is concerned, while the opposite occurs with the central provinces in the first wave and the northeastern provinces in the last. In turn, at various moments in time, the Galician provinces and their contiguous ones appear as clusters with low incidence values, more markedly for the rate of deaths.

As for the determinants of the pandemic, we observe that one of the main determinants is temperature, presenting a negative relationship in the first three waves and in the last one, being especially intense in the first one. We align, therefore, with most of the literature stating that temperature has a certain inhibitory effect on the transmission of the virus. As for the socio-economic determinants, we obtain that GDP pc (positive relation in most of the waves) and population density (always negative correlation) are the main drivers of the incidence of the pandemic in the Spanish provinces. On the contrary, the influence of the proportion of population older than 70 years and the mobility of people are not such a clear determinant.

A main message coming from our findings is that the incidence of COVID-19 is a result of a rather set of factors, both meteorological and socio-economic, whose impact changes along the waves. In this sense, temperature is a key driver in the waves in which population was not vaccinated. Once most of the population is covered by the vaccine, temperature does not have such a clear influence. Overall, our results would be evidencing above all the lower incidence in the southern part of the country and on the islands (areas of high temperature, high population density and low levels of GDP pc) and the highest levels of incidence in the first and sixth waves in the area of Catalonia, Valencia, Navarra and the Basque Country (areas of high GDPpc). While little can be done with respect to the meteorological conditions or changing agglomeration, regional governments can respond to the pandemic with the provision of an efficient vaccination procedure.

References

Ahmadi, M., Sharifi, A., Dorosti, S., Ghoushchi, S. J., & Ghanbari, N. (2020). Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Science of the total environment*, *729*, 138705.

Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses*, *12*(3), 254.

Amdaoud, M., Arcuri, G., & Levratto, N. (2021). Are regions equal in adversity? A spatial analysis of spread and dynamics of COVID-19 in Europe. *The European Journal of Health Economics*, *22*(4), 629-642.

Anselin, L. (1988). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.

Bauer, L., Broady, K., Edelberg, W., & O'Donnell, J. (2020). Ten facts about COVID-19 and the US economy. *Brookings Institution*, *17*.

Bayode, T., Popoola, A., Akogun, O., Siegmund, A., Magidimisha-Chipungu, H., & Ipingbemi, O. (2022). Spatial variability of COVID-19 and its risk factors in Nigeria: A spatial regression method. *Applied Geography*, *138*, 102621.

Bhadra, A., Mukherjee, A., & Sarkar, K. (2021). Impact of population density on Covid-19 infected and mortality rate in India. *Modeling earth systems and environment*, *7*(1), 623-629.

Briz-Redón, Á., & Serrano-Aroca, Á. (2020). A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Science of the total environment*, *728*, 138811.

Cao, Y., Hiyoshi, A., & Montgomery, S. (2020). COVID-19 case-fatality rate and demographic and socioeconomic influencers: worldwide spatial regression analysis based on country-level data. *BMJ open*, *10*(11), e043560.

Carozzi, F., Provenzano, S. & Roth, S. (2020). Urban density and COVID-19. Centre for Economic Performance (CEP) discussion paper, n1 1711. *Available at SSRN 3643204*.

Coccia, M. (2020). Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Science of the Total Environment*, *729*, 138474.

Coelho, M. T. P., Rodrigues, J. F. M., Medina, A. M., Scalco, P., Terribile, L. C., Vilela, B., & Dobrovolski, R. (2020). Global expansion of COVID-19 pandemic is driven by population size and airport connections. *PeerJ*, *8*, e9708.

Dalziel, B. D., Kissler, S., Gog, J. R., Viboud, C., Bjørnstad, O. N., Metcalf, C. J. E., & Grenfell, B. T. (2018). Urbanization and humidity shape the intensity of influenza epidemics in US cities. *Science*, *362*(6410), 75-79.

Ehlert, A. (2021). The socio-economic determinants of COVID-19: A spatial analysis of German county level data. *Socio-economic planning sciences*, *78*, 101083.

Elhorst, J. P. (2014). *Spatial econometrics from cross-sectional data to spatial panels*. Springer.

Fatima, M., O'Keefe, K. J., Wei, W., Arshad, S., & Gruebner, O. (2021). Geospatial analysis of COVID-19: A scoping review. *International Journal of Environmental Research and Public Health*, *18*(5), 2336.

Ficetola, G. F., & Rubolini, D. (2021). Containment measures limit environmental effects on COVID-19 early outbreak dynamics. *Science of The Total Environment, 761. 144432.*

Fonseca-Rodríguez, O., Gustafsson, P. E., San Sebastián, M., & Connolly, A. M. F. (2021). Spatial clustering and contextual factors associated with hospitalisation and deaths due to COVID-19 in Sweden: a geospatial nationwide ecological study. *BMJ Global Health*, *6*(7), e006247.

Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial statistics*, *38*, 100443.

Gullón, P., Cuesta-Lozano, D., Cuevas-Castillo, C., Fontán-Vela, M., & Franco, M. (2022). Temporal trends in within-city inequities in COVID-19 incidence rate by area-level deprivation in Madrid, Spain. *Health & place*, 102830.

Iyanda, A. E., Boakye, K. A., Lu, Y., & Oppong, J. R. (2022). Racial/ethnic heterogeneity and rural-urban disparity of COVID-19 case fatality ratio in the USA: a negative binomial and GIS-based analysis. *Journal of Racial and Ethnic Health Disparities*, *9*(2), 708-721.

Jamil, T., Alam, I., Gojobori, T., & Duarte, C. M. (2020). No evidence for temperature-dependence of the COVID-19 epidemic. *Frontiers in public health*, 436.

Liang, Z., Zhu, H., Wang, X., Jing, B., Li, Z., Xia, X., & Sun, B. (2020). Adjuvants for coronavirus vaccines. *Frontiers in Immunology*, *11*, 589833.

Liu, Y., Mao, B., Liang, S., Yang, J. W., Lu, H. W., Chai, Y. H., & Xu, J. F. (2020). Association between age and clinical characteristics and outcomes of COVID-19. *European Respiratory Journal*, *55*(5).

Ma, S., Lai, X., Chen, Z., Tu, S., & Qin, K. (2020a). Clinical characteristics of critically ill patients co-infected with SARS-CoV-2 and the influenza virus in Wuhan, China. *International journal of infectious diseases*, *96*, 683-687.

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., & Luo, B. (2020b). Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of the total environment*, *724*, 138226.

Maza, A., & Hierro, M. (2022). Modelling changing patterns in the COVID-19 geographical distribution: Madrid's case. *Geographical Research*, *60*(2), 218-231.

Mena, G. E., Martinez, P. P., Mahmud, A. S., Marquet, P. A., Buckee, C. O., & Santillana, M. (2021). Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science*, *372*(6545), eabg5298.

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the total environment*, *728*, 138884.

Morley, J. E., & Vellas, B. (2020). COVID-19 and older adult. *The journal of nutrition, health & aging*, *24*(4), 364-365.

Mota, C. R., Bressani-Ribeiro, T., Araújo, J. C., Leal, C. D., Leroy-Freitas, D., Machado, E. C., & Chernicharo, C. A. (2021). Assessing spatial distribution of COVID-19 prevalence in Brazil using decentralised sewage monitoring. *Water research*, *202*, 117388.

Murugesan, B., Karuppannan, S., Mengistie, A. T., Ranganathan, M., & Gopalakrishnan, G. (2020). Distribution and trend analysis of COVID-19 in India: geospatial approach. *Journal of Geographical Studies*, *4*(1), 1-9.

Noland, R.B. (1995) Perceived risk and modal choice: risk compensation in transportation systems. Accident Analysis & Prevention, 27(4), 50-521.

Onder, G., Rezza, G., & Brusaferro, S. (2020). Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *Jama*, *323*(18), 1775-1776.

Paez, A., Lopez,F., Menezes, T., Cavalcanti, R., & Galdino da Rocha, M. (2021) A Spatio-Temporal Analysis of the Environmental Correlates of COVID-19 Incidence in Spain, *Geographical Analysis*, 53, 397–421 doi: 10.1111/gean.12241.

Pequeno, P., Mendel, B., Rosa, C., Bosholn, M., Souza, J. L., Baccaro, F., & Magnusson, W. (2020). Air transportation, population density and temperature predict the spread of COVID-19 in Brazil. *PeerJ*, *8*, e9322.

Pérez, L. I., Millán, P. C., Hontañón, P. C., & de Sabando, V. I. L. (2021). Aplicación de modelos econométricos para el análisis de la incidencia del COVID-19 en España. *Anales de ASEPUMA*, (29), 8.

Rodríguez, I. G., Sáez, M. P., & Prieto, D. C. (2022). *The dynamics of COVID-19: An empirical analysis with a view to spatial health econometrics using macrodata* (No. 2201). Universidade de Vigo, GEN-Governance and Economics research Network.

Rodríguez-Pose, A., & Burlina, C. (2021). Institutions and the uneven geography of the first wave of the COVID‐19 pandemic. *Journal of Regional Science*, *61*(4), 728-752.

Romero, L. D., & Arroyo, M. (2022). Geografías de la pandemia COVID-19 en España: el retorno del medio rural. *Documents d'Anàlisi Geogràfica, 2022, vol. 68, num. 1, p. 139-166*.

Sajadi, M. M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., & Amoroso, A. (2020). Temperature, humidity, and latitude analysis to predict potential spread and seasonality for COVID-19. *Social Science Research Network*.

Shi, P., Dong, Y., Yan, H., Li, X., Zhao, C., Liu, W., & Xi, S. (2020). The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak-evidence from China. *MedRxiv*.

Sun, Y., Hu, X., & Xie, J. (2021). Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England. *Science of the total environment*, *758*, 143595.

Tosepu, R., Gunawan, J., Effendy, D. S., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of the total environment*, *725*, 138436.

Wang, M., Jiang, A., Gong, L., Lu, L., Guo, W., Li, C., & Li, H. (2020). Temperature significantly change COVID-19 transmission in 429 cities. *medrxiv*.

Wong, D. W., & Li, Y. (2020). Spreading of COVID-19: Density matters. *Plos one*, *15*(12), e0242398.

Xie, J., & Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment*, *724*, 138201.

Yang, T. C., Kim, S., Zhao, Y., & Choi, S. W. E. (2021). Examining spatial inequality in COVID-19 positivity rates across New York City ZIP codes. *Health & place*, *69*, 102574.

Appendix A. Main variables definition and sources

Table A1. Definition of main variables

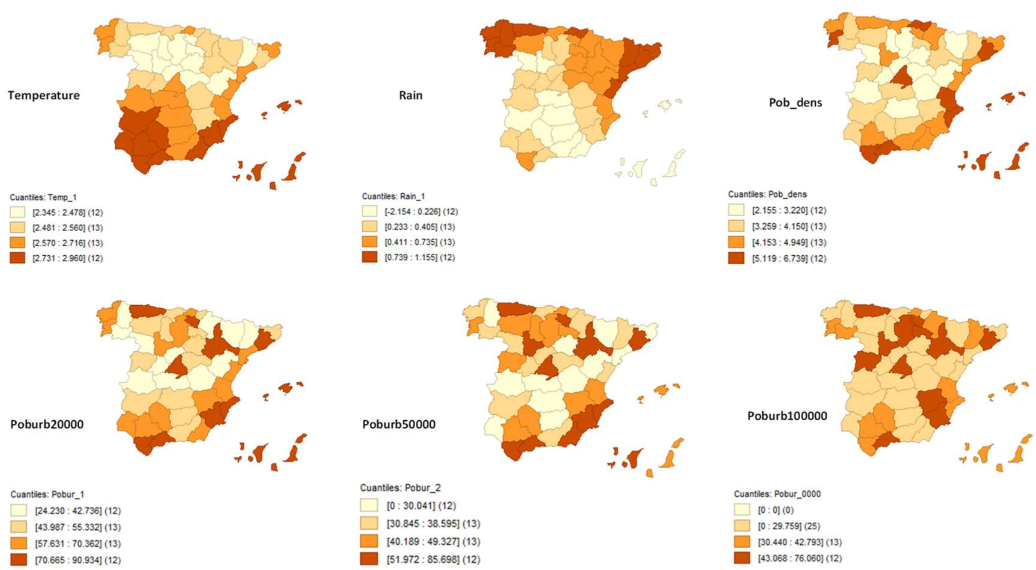
Variable	Definition	Level	Frequency	Last data	Source
Cases	Number of cases per 100.000 inhabitants	Province	Daily	may-22	Instituto de Salud Carlos III
Deaths	Number of deaths per 100.000 inhabitants	Province	Daily	ago-22	Instituto de Salud Carlos III
TempMax	Maximun temperature (average in the month)	Province	Daily	31/12/2021	Agri4Cast JRC
TempMin	Minimun temperature (average in the month)	Province	Daily	01/01/2022	Agri4Cast JRC
TempAvg	Average temperature (average in the month)	Province	Daily	02/01/2022	Agri4Cast JRC
RainAvg	Rain (average in the month)	Province	Daily	03/01/2022	Agri4Cast JRC
GDPpc	GDP pc 2019	Province	Yearly	2019	INE
Business	Number of enterprises	Province	Yearly	2019	INE
PopDens	population density	Province	Yearly	2022	INE
UrbPop_20000	Porcentaje de población (%) que vive en ciudades de más de 20.000 habitantes	Province	Yearly	2021	INE
UrbPop_50000	porcentaje de población (%) que vive en ciudades de más de 50.000 habitantes	Province	Yearly	2021	INE
UrbPop_100000	Porcentaje de población (%) que vive en ciudades de más de 100.000 habitantes	Province	Yearly	2021	INE
Travellers_hot	number of travellers in hotels	Province	Monthly	2022M04	INE
Nights_hot	number of nights in hotels	Province	Monthly	2022M05	INE
Travellers_rur	number of travellers in rural hotels	Province	Monthly	2022M06	INE
Nights_rur	number of nights in rural hotels	Province	Monthly	2022M07	INE
AirPassangers	Número de pasajeros de transporte aéreo	Province	Monthly	2020	AENA
Metro	travellers in metro	City	Monthly	2022M03	INE
Pop_70	share of people over 70 years old	Province	Yearly	2021	INE
IntTrade_GDP	international trade (exports+imports). Share over GDP	Province	Yearly	2021	ICEX
IntTrade_Pob	international trade (exports+imports) in per capita terms	Province	Yearly	2022	ICEX

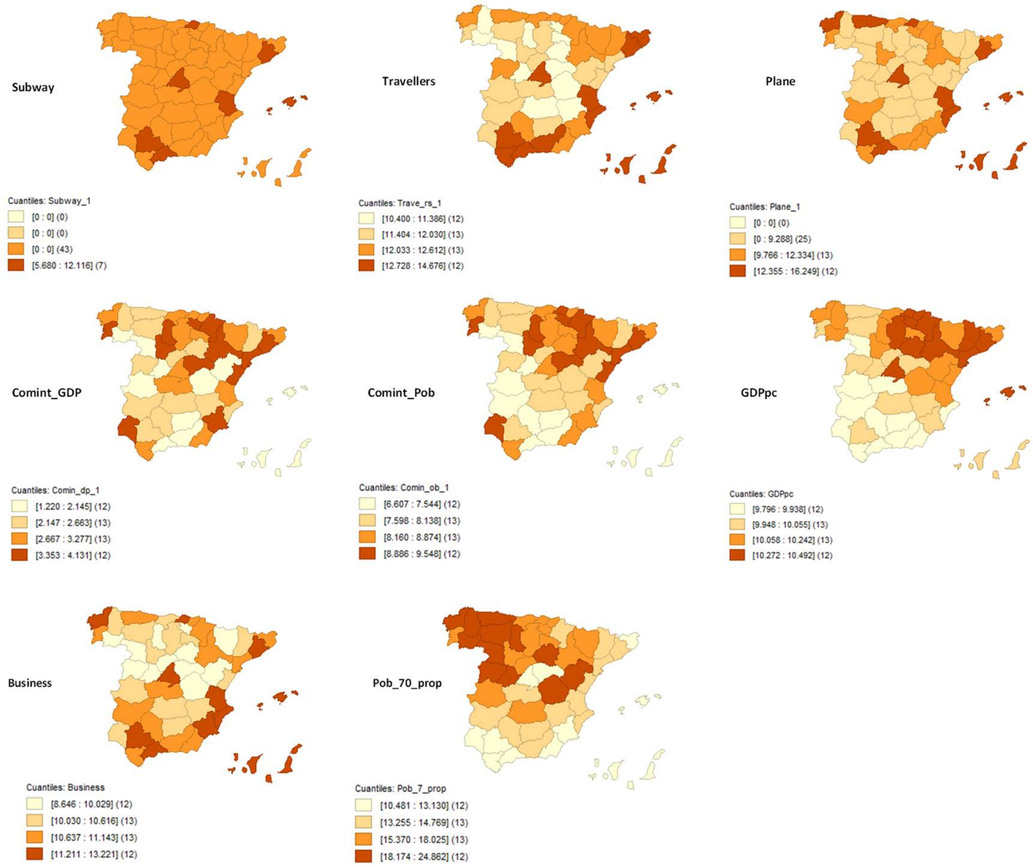
The data for the incidence of the COVID19 can be found in the web page of the Instituto de Salud Carlos III: https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos. As commented in such a webpage, the data presented in this COVID-19 Panel are obtained from the declaration of COVID-19 cases to the National Epidemiological Surveillance Network (RENAVE) through the computer platform via Web SiViES (Spanish Surveillance System) managed by the National Epidemiology Center (CNE). This information comes from the epidemiological case survey that each Autonomous Community completes when a case of COVID-19 is identified. Due to the change in the COVID-19 Surveillance and Control Strategy, as of March 28, 2022, only COVID-19 cases in the population aged 60 and over are shown in this panel.

The COVID-19 Panel presents geographic information on cumulative incidence rates at 14 days and 7 days, for people aged 60 or over and the evolution of the number of COVID-19 cases for this age group since the start of the pandemic.

To calculate all the parameters, the imputed date fecha_imp is used, which is the date of onset of symptoms or, failing that, the date of diagnosis minus 6 days (from the start of the pandemic to May 10, 2020), minus 3 days (from May 11, 2020 to March 28, 2022 and less 2 days from March 28, 2022. For asymptomatic cases, the date of diagnosis is used. In those cases in which there are no symptoms or diagnosis, the key date is used (the date of onset of symptoms and, in its absence, the date of declaration to the AC, up to 10 May 11, 2020); from May 11 onwards, the Key date is the earliest of the consultation or diagnosis dates; Occasionally it can be replaced by the date of sampling). Until May 10, 2020, cases diagnosed by a daily test are included. Positive diagnosis of active infection, as well as all hospitalized cases, ICU admissions and deaths; as of May 11, cases confirmed by PCR, or by antigen tests, are included. As of March 28, 2022, only confirmed cases are reported in the population aged 60 and over. The population used to calculate the incidence rates comes from the official population figures resulting from the revision of the municipal register on January 1 of the National Statistics Institute of 2021.

Figure A1. Quantile maps of potential determinants of Number of cases and deaths




Subway
Cuantiles: Subway_1
[0 : 0] (0)
[0 : 0] (0)
[0 : 0] (43)
[5.680 : 12.116] (7)
Travellers
Cuantiles: Trave_rs_1
[10.400 : 11.386] (12)
[11.404 : 12.030] (13)
[12.033 : 12.612] (13)
[12.728 : 14.676] (12)
Plane
Cuantiles: Plane_1
[0 : 0] (0)
[0 : 9.288] (25)
[9.766 : 12.334] (13)
[12.355 : 16.249] (12)
Comint_GDP
Cuantiles: Comin_dp_1
[1.220 : 2.145] (12)
[2.147 : 2.663] (13)
[2.667 : 3.277] (13)
[3.353 : 4.131] (12)
Comint_Pob
Cuantiles: Comin_ob_1
[6.607 : 7.544] (12)
[7.598 : 8.138] (13)
[8.160 : 8.874] (13)
[8.886 : 9.548] (12)
GDPpc
Cuantiles: GDPpc
[9.796 : 9.938] (12)
[9.948 : 10.055] (13)
[10.058 : 10.242] (13)
[10.272 : 10.492] (12)
Business
Cuantiles: Business
[8.646 : 10.029] (12)
[10.030 : 10.616] (13)
[10.637 : 11.143] (13)
[11.211 : 13.221] (12)
Pob_70_prop
Cuantiles: Pob_7_prop
[10.481 : 13.130] (12)
[13.255 : 14.769] (13)
[15.370 : 18.025] (13)
[18.174 : 24.862] (12)

UBIREA

Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

WEBSITE: www.ub.edu/irea • **CONTACT:** irea@ub.edu

AQR

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona