

# Scalable Analysis of COVID-19 Spatiotemporal Patterns Based on Data Mining Tools: Using 3D Bins to Predict Short-time Focus Locations

Olga de Cos (✉ [olga.decos@unican.es](mailto:olga.decos@unican.es))

University of Cantabria <https://orcid.org/0000-0002-2245-5378>

Valentín Castillo

University of Cantabria: Universidad de Cantabria

David Cantarero

University of Cantabria: Universidad de Cantabria

---

## Research

**Keywords:** Spatial patterns, COVID-19 Hotspots, Microdata, Space-time bins, ArcGIS, Geoprevention

**Posted Date:** January 11th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-141410/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** An interesting research line is related to COVID-19 behavior from a territorial and temporal perspective. Hence, the use of 3D space-time bins is a useful tool to contrast limitations of visual assessment and reveal the detailed areas most at risk for the pandemic or even more the emergency hotspots can be useful to not only study but also predict spatial pattern of the COVID-19 at an intra-urban scale.

**Methods:** We developed the SITAR Fast Action Territorial Information System using ESRI Technologies Ecosystem. More specifically, we used ArcGIS Pro (desktop) and ArcGIS Online (cloud). Therefore, our general research methodology is based on Geographic Information Technologies from a multiscalar perspective and based on detailed entities (geocoded COVID-19 cases for the region of Cantabria, Spain). The main research method is related to data mining tools using 3D bins and analysing emerging hotspots.

**Results:** The spatial autocorrelation analysis of the COVID-19 reveals that the distribution of the cases is not random. Otherwise, the Moran's Index confirms that the spatial pattern of COVID-19 cases is statistically significative, and it presents a clustered distribution. And in the cases of elderly homes, COVID-19 outbreaks and spatial focus are linked while in the rest of the cases there is not this spatial association. The analysis of 3D bins and emerging hotspots is revealing from the point of view of geoprevention in that it significantly limits the territory on which it would be important to focus the analysis. In fact, of the 1,414 starting cubes, focusing on the 602 remaining cubes (with statistical significance), all correspond to a hotspot pattern.

**Conclusions:** Our results evidence the existence of significant space-temporal trends that it can serve as support of emerging hotspots of COVID-19 that it can be used as a prelude to what will happen in the next future. To our knowledge, this is the first study for Spain that demonstrates the interest of the 3D space-time cubes method to engage the prevention measures proposed by policy makers with a scalar perspective. 3D bins can therefore be used as a proxy to assess the spatiotemporal patterns in public health studies.

## Background

This paper is addressed at a time when the third wave of the COVID-19 pandemic has challenged the socioeconomic and health structures worldwide. In this context, Spain with a population close to 47.5 million inhabitants and after one draconian confinement and many rules to maintain social distance and reduce mobility has already exceeded 1,800,000 COVID-19 cases, which implies a cumulative incidence of practically 40 cases per 1,000 inhabitants, compared to the 11 cases per 1,000 inhabitants of world average. This issue is due to the fact of the relaxation in compliance with the social distance measures by the population and another reason is that the vaccines did not arrive in time to temper this third wave of the COVID-19 pandemic.

Furthermore, the Spanish case is especially relevant for two reasons. On the one hand, the collapse of their National Health Service due to full capacity of hospitalizations and Intensive Care Units and the severity of the condition in the first wave of March 2020. On the other hand, the difficulties derived from the multiple outbreaks that have occurred since the beginning of the second wave in August. To solve it many health policy makers applied perimeter confinements adapted to neighborhood limits, but there are not enough diagnosis related to these solutions.

The fact is that after the strict confinement (from March to June, 2020) in the new normal stage characterized by the “coexistence with the virus” health authorities need studies and strategic reports based on spatial analyst to take spatial decisions, such as: mobility reduction, disinfection of facilities or intensification of security and vigilance, among others.

In this context, this research is located within the collaboration framework established by University of Cantabria, IDIVAL Valdecilla and the Department of Health of the Regional Government of Cantabria. Our research is based on Geographic Information Technologies and the main goal is focused on analyzing the space-temporal trend of COVID-19 focus in the Autonomous Community of Cantabria (North of Spain). It is necessary to difference the status of the focus with a temporal and multiscale perspective to help policy makers to take decisions in real time to control outbreaks and hotspots.

Two are the main pillars of our methodological contribution: geocoding and analysing daily microdata register of COVID-19 cases and implementing a Fast Action Territorial Information System -desktop and cloud- called SITAR, using ESRI Technologies [1].

To get our research goals, after considering the statistical significance of spatial patterns of pandemic, we use data mining tools in two stages. Firstly, we analyse 3D bins with standard dimensions and multiscale perspective and secondly, we calculate emerging hotspots that is necessary to focus policy makers attention in problematic areas, differencing trend of hotspots, which is necessary to plan the pandemic control from the intra-urban scale to the regional one. It is important to point out that in the theories relating to the space-temporal trend of COVID-19 focus the scale is fundamental, in that a valid theory at the country or department level could not be important at a scale of detail with an intra-urban perspective (at the neighborhood level). Therefore, we defend that the knowledge of the spatial behavior of COVID-19 requires a multi-scalar analysis [2].

Therefore, the research goals are centered on one of the two phases identified by Jindal et al., (2020) [3] for the risk mitigation (control the spread and decrease the disease severity) because using emerging hotspots analysis of COVID-19 could contribute to reduce the spatial sprawl, with scalable and temporal perspective. The application of this method to COVID-19 hotspots is very innovative. Factual, there are very few publications about the application of 3D bins analysis to the pandemic. In this field an interesting precedent is the analysis of spatiotemporal pattern in China based on data from cities [4]. Although the scale is very different to our research and the authors concluded in their paper the consideration of their large scale as a limitation, the cited paper establishes an interesting research frame related to COVID-19 spatial behavior. Nevertheless, related to other diseases analysis with territorial and

temporal perspective, the use of 3D space-time bins has an interesting research frame. Many studies consider this methodology as a way to contrast limitations of visual assessment. Data mining tools, applied to health topics, can reveal the detailed areas that are most at risk for the pandemic during the period under consideration [5] or even more the emergency hotspots can be useful -depending on spatial and temporal granularity- to not only analyze but also to predict spatial pattern of the virus at an intra-urban scale [6]. In these cases, higher detailed scale is based on geocoding of cases as a coordinates pair, as other researchers analyzed the pulmonary TB cases using about 1,000 registers as a key study to implement proper prevention programs [7].

Moreover, at this moment, to reduce the spread it is essential to distinguish a problematic focus from the others answering to the key question “where” [8] and related to this, the role of geotechnologies is strategic to shorten response times for social management [9]. Even more, disciplinary knowledge about territory, sociodemographic content or health advances could take advantage of the geotechnologies opportunities in the era of “circulation stage” materialized in GIS-Cloud technologies [10, 11]. Facing challenges of the second and near third wave of the pandemic we must prioritize the spatial and temporal evolution of focus and, in this context, the use of geotechnologies with an interdisciplinary perspective is an interesting way to overcome the global pandemic of the 21 Century: the COVID-19 [12, 13].

Indeed, the predictions about COVID-19 spread are more fuzzies when spatial scale is more detailed and when the future prediction time is longer. Therefore, there are many studies that predict the pandemic evolution in global or national scales [14, 15] or even researches that try to analyse indirectly the sprawl of the virus using the characteristics of main affected areas, such as: rent, economic activities, density or mobility among others [16–20].

The predictive models at a detailed scale are less developed, being deeper if it refers to incidence geocoded microdata. Additionally, we must consider the difficulty of focus and outbreaks relations. In some cases, new focus are identified with outbreaks (as it occurs in elderly homes, where the location cause and the place of homes of infected people are the same locations) but in other cases, the most important proportion, outbreaks have an spatial incidence in many locations (because of different places where infected people live). Or even more, there are a high proportion of positive cases where the origin of the contagious is unknown.

In this regard, 3D bins analysis is crucial for identifying and evaluating the different risk associated to the active focus or even to analyse with an historical perspective the role of neighbourhoods in the pandemic sprawl.

## Methods

The research is based on the daily microdata record from the Government of Cantabria (Spain) with the permission to be managed only in the context of our research line that was conceded by the Research with Medicines Ethics Committee from Cantabria (CEIm) in June 2020 (ID: 2020.238). The authors want

to clarify that microdata record presents an anonymous format, due to the protection data rules at an international level (European Union Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016) but also at a national level (Organic Law 3/2018 Protection of Personal Data and Guarantee of Digital Rights of 5 December 2018).

These daily files have a tabular structure with data related to each person with positive test in COVID-19 in Cantabria. Moreover, microdata includes many fields related to different thematic areas: location (address, locality, municipality and postal code), demographic profile (sex, age and professional category if the case is a sanitary employee) and, finally, health structure and virus details (start and end dates, health area adscription, COVID-19 status -positive in case of active virus, cured or deceased-, test type, binary fields related to hospitalization, UCI cares and elderly homes -if the infected person lives there-. Apart from necessary fields about each COVID-19 case, we point out that some characteristics are related to specific circumstances of the pandemic in Spain: on the one hand, the category data if the infected person works in health sector is connected with the high incidence of the virus in the health care providers (more than 83.000 infected, about 5% of COVID-19 patients in the country) and, on the other hand, the field about elderly homes is related to high incidence of the first wave of the virus in Spain in retirement homes. In our research we separate COVID-19 cases if they live in elderly homes because their geoprevention strategies are different from the rest of focus and if we include it in our methodology we could misread statistical and cartographic results.

Related to temporal perspective, the daily microdata register cumulates at the moment of this research more than 15,000 cases (rows in the microdata table) since the beginning (March 2020) and nowadays more than 4,000 cases are in positive status (actives in the virus, not cured). Attending to the number of records of the study, we have to clear that from the 15,168 initial cases in tabular format we got the corresponding geocoded record in 14,938 cases. Therefore, geocoding process only missed 230 cases (1.5%) because of the absence of adequate content in address fields. Additionally, taking into account that it could be misleading to consider the location of cases in elderly homes, as we explained before, we filtered these cases out of the base layer (1,031 cases, almost 7%). Therefore, finally, our study is based on 13,907 geocoded cases (coordinate pairs or points).

Regarding to the research methodology, it is based on Geographical Information Technologies. More specifically, we implemented SITAR Fast Action Territorial Information System using two environments from ESRI Technologies Ecosystem, accessed by the use license from the University of Cantabria.

More specifically, we use ArcGIS Pro, as a desktop Geographic Information System (GIS) and ArcGIS Online, as a GIS-Cloud, with Operations Dashboard for ArcGIS, Web AppBuilder and Experience Builder. The main advantage of SITAR is the normalization and integration of geodatabases than include data from many different sources. Aside from sources provided by official institutions the research includes geographic, demographic and socio-economic data from ESRI Spain COVID-19 GIS Hub, ArcGIS GeoEnrichment Service based on big data technologies and, finally, Web Map Services to connect to global cartography (satellite images, or similar). Furthermore, our territorial information system has a

multiscalar perspective from detailed scale from buildings (cadastral layer) and point COVID-19 cases until municipalities and region of Cantabria.

Regarding to specific tools and methods, this research is based on daily multiple-field geocodification that convert original microdata tables into point shapes and data mining tools from ArcGIS Pro (ESRI). Regarding to this, we designed a methodology (Fig. 1) based on an exploratory stage to verify the statistical significance of the spatial pattern of COVID-19 microdata and subsequently we develop a data mining analysis stage focused on the creation of space-time 3D bins in a NetCDF layer where points are accumulated into a bigger structure with spatial and temporal dimensions (ESRI). Each bin represents a regular location (with specific area and volume) where cases are counted and aggregated over time in a 3D structure. It includes in each bin location (spatial component) the number of cases over time-steps as slides (temporal component). Therefore, the space-time 3D bins model overcomes the excessive detail of each location to reveal a simplified crosstab model that integrates spatial and temporal components. This new data model is essential to next stage, where we introduce an emerging hotspots analysis to identify trends. The tool is based on Getis-Ord  $G_i^*$  statistics to identify hotspots and Mann-Kendall statistics to determine trends [21]. The method is based in the key field count (aggregated cases) of each bin recorded over time. The pairwise comparisons of each bin value with the subsequent time-step value are essential to identify an incrementing, decrementing or unchanging trend. Related to this, the method includes other parameters to examine current trends (expected and observed sum of cases, z-score to know the trend sign and p-score to check the statistical significance) [6]. On this base, the Mann-Kendall method produces an interesting trend typology about decreasing bins (cold spots) and increasing bins (hot spots) with eight categorized results based on trends (in both directions: increasing and decreasing). Regarding to geoprevention framework, hot spots are more important than the cold ones. Even more, considering the eight typologies of trends, we can disentangle different levels of importance: the more worrying are new, consecutive, intensifying persistent and oscillating hotspots, followed in a second level of importance by: diminishing, sporadic, and historical hotspots. Regarding to this, the SITAR GIS system implemented allows to analyse the spatial patterns of COVID-19 over time at detailed scales and to produce reports in real time related to sources of contagion that may arise at a certain moment.

## Results

We present the results with a multiscale perspective. It is necessary to highlight the importance of analysing spatial patterns of COVID-19 from each case to the regional scale. Therefore, our research is based on two result levels: at a regional scale and at an intra-urban scale in the cities of Santander and Torrelavega (Community of Cantabria, Spain).

Before explaining the results, it is important to introduce our case study. The Autonomous Community of Cantabria (north region of Spain) has just over 580,000 inhabitants and a surface area of 5,300 Km<sup>2</sup> (2,055 square miles), which represents an average regional density close to 110 inhabitants / Km<sup>2</sup> (285 inhabitants/squared meters). However, the distribution of the population presents internal disparities between the coastal municipalities and the interior valleys ones (Fig. 2).

The capital city Santander is located in the central coastal area with 172,000 inhabitants. Santander, the biggest city of the Community of Cantabria, heads the functional urban area (FUA, identified at a European level) that includes Torrelavega, which is the second most important city. Indeed, the Santander FUA extends to 25 municipalities where live 380,000 inhabitants (just over 65% of the population of Cantabria). This polycentric hinterland Santander-Torrelavega highlights for several factors, such as: number of inhabitants, density, activities concentration, existence of main transport infrastructures and a prominent role in daily pendulum movements (commuting) between central areas and the surroundings [22].

Regarding to the general evolution of the pandemic in Cantabria, the accumulated data exceeded recently the 15,000 cases. As the Fig. 3 shows we are in the second or even third wave. Indeed, nowadays the accumulated incidence in the last fourteen days is very high, above 500 cases per 100,000 inhabitants.

Presented the general framework, the research is focused on spatial patterns of COVID-19. Therefore, next sections present the results based on geostatistical analysis from SITAR. Beginning with the statistical significance of spatial distribution of infected cases, we will continue with 3D bin results from both regional and intra-urban scales.

### **Exploratory results of spatial autocorrelation**

The spatial autocorrelation analysis of the COVID-19 reveals that the distribution of the 13,907 cases is not random. Otherwise, the Moran's Index confirms that the spatial pattern of COVID-19 cases is statistically significative, and it presents a clustered distribution (Fig. 4). Indeed, the z score of 6.56 (up to 2.58) implies a probability lower than 1% of a random distribution of the COVID-19 cases.

It is interesting to mention that we explored the Moran's Index including elderly homes cases obtaining that, although the spatial pattern was not random, nevertheless the p value increased from 0 to 0.19 and the probability of not random distribution was less clear (under 5% instead of under 1% without retirement homes cases). These comparative results (with and without considering elderly homes) confirm the importance of excluding retirement homes cases to avoid the distortion of statistical results. Moreover, in the cases of elderly homes, we point out that COVID-19 outbreaks and spatial focus are joined or linked, while in the rest of the cases, the spatial association does not exist.

Additionally, other analysis has been applied to contrast the Moran's Index results. Specifically, the nearest neighbour distance confirms the nonrandom spatial pattern of COVID-19, which is coincident with a clustered model. In the nearest neighbour analysis, as the Table 1 shows, the average observed distance among cases is 38.7 metres at a regional level and the Z score (standard deviation) is -201.58 (under - 2.58), therefore the spatial pattern is clustered and not random with a confidence again greater than 99%.

Table 1

The spatial average nearest neighbour: COVID-19 cases and COVID-19 locations counts (number of overlapped points).

Nearest neighbour Results	Cases (points)	Locations counts
Observed distance	38.7 m	126.5 m
Expected distance	408.2 m	538.5 m
Ratio nearest neighbour	0.095	0.217
Z Score	-201.58	-122.07
P value	0.00	0.00
Source: Authors' elaboration based on COVID-19 Microdata Daily Register from the Health Authorities (Government of Cantabria, Spain).		

The preliminary geostatistical analysis is fundamental to support the following research stages. The fact that the COVID-19 spatial pattern is statistically significant allows a deeper analysis based on data mining and 3D space-time bins.

### COVID-19 bins at a regional scale with a global temporal period

This section presents the analysis of 3D bins and emerging hotspots for the Autonomous Community of Cantabria throughout the complete period of the data series, that is, from March 1 to November 20, 2020. Hence, the sequence includes the COVID-19 cases corresponding to waves 1 and 2 facing the third wave too.

In the absence of established standard thresholds of time and distance (size of the cubes) for the creation of 3D COVID-19 bins, we set time periods of 4-week and based the distance on preliminary statistical analysis, considering as a reference the expected distance threshold 538.5 m (3.3 miles) derived from exploratory spatial average nearest neighbour analyses (as indicated in Table 1). The 4-week interval is suitable since it includes 2 periods of the usual reference time for the study of accumulated incidents (that is, 14 days) and fulfils the condition of the method to be applied that establishes a minimum of 10 moments of time for development of the bins.

The result provided by the visualization of the cubes at the regional level (Fig. 5) is revealing in that it simplifies the information of the starting points or their corresponding heat map and brings to the fore the hierarchies of the pandemic affectation in the analysed territory. 1,414 bins are identified that respond to differentiated intensities and distributions. The cubes show two outstanding levels of spatial segregation: firstly, the one that refers to the inland coast differences, with a clear articulation in the coastal municipalities, and secondly, its organization is concentrated in areas of high density and mobility, as occurs in the Santander-Torrelavega sector of the FUA of Santander and especially in the western arc of Santander Bay, which is where the peri-urbanization processes of the city of Santander are most intense.



The activity of the cubes of the eastern coastal region is much more prominent than that of the western coast. The main reason of this disparity West-East is the proximity of Bilbao, the tenth most important city in Spain (near 350,000 inhabitants in the city but more than 1 million inhabitants in the metropolitan area). Bilbao is located 100 Km (62 miles) from Santander and has an important role as a pole of economic attraction in the North of Spain and its influence on the eastern municipalities of Cantabria, as well as the outstanding exchange of flows with the neighbouring community are factors to take into account in these results. Furthermore, at a European level, the FUA of Bilbao exceeds its Autonomous Community borders (Basque Country) and it includes the oriental part of Cantabria. It is an important sign of the intense inter-relation of the East part of Cantabria and Bilbao.

On the other hand, inland Cantabria has a layout based on small cubes in most of the territory except for the regional headwaters of the interior valleys.

The first expressive result of emerging hotspots is that of the 1,414 cubes identified in the region, 812 (57%) do not present a pattern that can be associated with a specific hot or cold spot. In fact, from the COVID-19 distribution and the consideration of its spatial pattern over time do not derive any cold spots.

The analysis of emerging hotspots is revealing from the point of view of geoprevention in that it significantly limits the territory on which it would be important to focus the analysis. In fact, of the 1,414 starting cubes, 57% do not present any identifiable pattern such as cold or hot spot. Despite its importance in the number of cubes, it should be noted that only 22% of the cases occurred in these areas. Focusing on the 602 remaining cubes (with statistical significance), all correspond to the hotspot pattern and according to the number of cases the results obtained are presented as follows:

- 251 bins (42%) are new hot spots, so it is a hot spot with statistical significance only in the final part of the period analyzed (October-November). These cubes house 2,190 cases (16% of the series analyzed).
- 178 bins (30%) are oscillating hot spots. These are significant hot spots in the final period (October-November) but with a previous trend in which it has been a significant cold spot. In this typology less than 90% of the time intervals have been significant hot spots. In Cantabria this typology is the second not only number of bins but also in number of cases, since it accumulates 3,163 (23%). This type is also where the lowest mean age of the cases is detected (41.7 years).
- 98 bins (16%) are sporadic hot spots. This typology corresponds to locations that are and are no longer a hot spot several times in the time considered. Less than 90% of the time intervals have been significant hot spots and have never behaved as significant cold spots. This typology is striking in that in a limited number of cubes (16%) it is the one with the most COVID-19 cases, with a total of 3,988 cases (30%). Furthermore, possibly related to the large number of cases in this kind of hot spot is where more deaths have been concentrated without considering those corresponding to residences (42 deaths, equivalent to 31%).
- 75 bins (12%) are consecutive hot spots. These are areas with significant hot spots in a single run without interruption in the final time intervals considered. These cubes were not before the last run

significant hot spots. These hotspots are also the least common in terms of number of cases, accumulating 1,105 cases (8%).

On the other hand, in addition to the general statistical guidelines, the spatial pattern is very interesting (Fig. 6). The new hot spots establish the areas of new cases that, in urban areas already hit in the first wave by the pandemic, present a peripheral configuration in the form of haloes around areas with a high intensity of cases previously. Hence, it would respond to the space process COVID-19 sprawl in the central and eastern coastal area. On the other hand, in rural areas in most cases they correspond to new locations in the headwaters of the region.

Considering the typologies with alternating cases over time, these are concentrated in the municipalities of the two main cities Santander and Torrelavega, as well as in the periphery of Santander in the arc of the bay. It is noteworthy that these typologies do not intermingle, but rather are segregated in an east-west direction in Santander and north-south in Torrelavega, with sporadic hotspots coinciding with areas with the highest social content and oscillating hotspots with more modest areas.

### **Second COVID-19 wave using 3D and emerging hotspots analysis at an intra-urban scale**

The methodological adaptations of the scale from the spatial point of view may also be accompanied by specific changes in the temporal resolution and the bin distance parameter. Thus, in this section, the emerging hotspots in the municipality of Santander are analysed at the intra-urban level and with a period corresponding to the second wave, so the cases are filtered from August 1 to November 20. The analysis is based on 3,477 COVID-19 cases that are analysed in their space-time cubes at 7-day intervals.

This study more spatial and temporally limited yields results of interest and expressive geospatial information for decision-making from the geoprevention perspective. Considering the average expected distance (based on nearest neighbour analysis: 102.05 m -0.06 milles-) as the distance in the construction of the cubes, almost 700 bins are obtained for the whole of the municipality of Santander (Fig. 7), which present an important development in the entrance streets to the city as well as in central areas, while in the north with a lower population density and to the east with a predominantly high status social gradient the configuration of the cubes is very discreet in number and size.

The analysis of emerging hotspots at this scale results in a filtering of non-significant cases for 278 bins (40%, a much lower proportion than at the regional scale) in which 830 cases (24%) are located. Therefore, we will focus on the typologies of the rest of the cubes (418, 60%), all of them hotspots except for one that gives a cold spot profile.

Table 2  
Emerging Hotspots of COVID-19 in the municipality of Santander.

Emerging hotspot typology	Bins	Cases	Deceased*	Average age
Oscillating hotspots	244	1,485	12	42.0
Sporadic hotspots	88	789	9	44.2
Consecutive hotspots	58	277	2	41.4
New hotspots	27	95	1	43.0
<b>Total</b>	<b>417</b>	<b>2,646</b>	<b>24</b>	<b>42.7</b>
*Without considering residents in elderly homes.				
Source: Authors' elaboration based on COVID-19 Microdata Daily Register from the Health Authorities (Government of Cantabria, Spain).				

Regarding the spatial pattern of emerging hotspots, we highlight the existence of significant patterns in the center and east of the city. Areas of modest social content and high density are identified as sporadic hotspots, areas in which hot spots occur in various periods of time in response to a spatial repetition factor. Oscillating hotspots stand out in neighborhoods of medium and medium-low social content that will be areas of attention for upcoming geoprevention actions since they are identified as significant hotspots in this last period.

Moreover, the consecutive hotspots present a peripheral distribution and correspond to areas of lower population density. Finally, the existence of points that respond to new hotspots with a concentrated distribution in the area of contact with the eastern sector with high social content and modest neighborhoods to the west, both focus of peripheral and opposite position, deserves a mention.

## Discussion

Our findings reveal that the use of COVID-19 microdata provides a useful way of interpreting trends behind emerging hotspots with a multi-scalar perspective. Moreover, the method based on space-temporal bins has a high potential to obtain relevant spatial patterns of focus with both spatial and temporal perspective.

This is essential since access to COVID-19 data with a high level of spatial disaggregation (points as coordinates after geocoding) and with a high granularity (daily) must be properly analyzed in order that they can be transformed into strategic spatial information. In other words, the optimal data sequences could be useless - beyond expressive heat maps or cumulative incidence rates for districts or health areas - if they are not analyzed using expressive data mining techniques.

In this regard, our study demonstrates the diagnostic contribution provided by the combined exploratory analysis of statistical significance together with the calculations of space-time bins and, finally, the emerging hotspots. A recursive filtering is achieved from the cubes that are not significant to those that are statistically significant and on these a typology that would reveal the specific areas on which it would be useful to act. Therefore, we support the high potential of this kind of analysis to plan periodic geoprevention actions in affected areas.

However, we understand that a limitation attributable to our research is that we cannot compare our results with those obtained in other studies at similar scales of the same subject, as has already happened to other authors with the patterns detected in other respiratory diseases [7]. However, it is possible to identify thematic affinity insofar as similarities are observed in the approach to applying 3D space-time bins in other investigations that, like ours, would seek to contribute to geoprevention plans, and this concept would go beyond the health research and could be extended to other topics of interest such as security [23, 24], which would keep the similarity of the methodology used by us for point layers analysis, in our case with a novel approach towards geoprevention in health and more precisely in relation to a priority research object, such as COVID-19.

It is true that our results have expiration dates and that they could undergo modifications as a function of evolution over time and parameters linked to the spatial scale of analysis. In fact, when adding long periods of time, which with a daily series and with the COVID-19 virus in full activity, could be considered above the month, there are certain typologies of hotspots that are omitted or masked in internal periods of the set of time considered. This is what happens with growing hotspots (locations with clustering intensity of high counts in each time-step) or with persistent hot spots, that is, a significant hot spot in 90% of the time intervals and without a distinguishable tendency to increase or decrease. These two cited categories -that would be of high interest for geoprevention actions- have not been obtained in our analysis, which could be due to the long period considered. This crucial aspect is pointed out at a methodological level for future research, although it does not invalidate or make us question the results obtained. We simply propose the opportunity to use this method continuously over time and at shorter intervals or time-steps (1–2 days) as a follow-up measure for policy makers in terms of geoprevention.

On the other hand, we have to clarify that although the method may have parameters that could introduce a quota of subjectivity, in our study we have opted for the use of distance thresholds based on the nearest neighbor analysis that allows the methodological design to be objectified. This also acts as a standard element that could be extensive for the replication of the method in other case studies or even at other scales of analysis.

This line of research is likely to be linked with other approaches addressed in works cited in the background, specifically those aimed at analyzing the socioeconomic, demographic and functional framework of the areas that accumulate COVID-19 cases [16–20]. Thus, we will work to study the conditions and environment variables that occur in the different kinds of hotspots in order to find social patterns that can be correlated or ultimately explain the spatial patterns in this paper presented.

# Conclusions

To our knowledge, this is the first study for Spain that demonstrates the interest of the 3D space-time bins method to engage the prevention measures proposed by policy makers with a scalar perspective. Thus, it provides support to decision makers to help making decisions related to geoprevention, from the perspective of the necessary linkage of the measures to the particularities of the spatial pattern of the COVID-19 in each territory.

It should be remembered that the method requires at least 10 moments of time for the analysis of emerging points, with which, having a series of COVID-19 cases daily, we understand that it could be applied as a diagnosis every 20 or 30 days at the intra-urban level. Moreover, it could help decision-making with strategic information as it brings to the fore a diagnosis of emerging points, which will be of special interest when hot spots are identified in some situations that may be associated with outstanding levels of danger and dynamism (new, growing or persistent hotspots, among others).

Finally, our results report significant space-temporal trends. It can serve as support due to their short-term predictive nature in that when considering both the intensity due to the accumulation of cases, as well as their tendency and also due to associated processes of spatial diffusion, the typologies of emerging hotspots can be interpreted as a prelude to what will happen in the coming days and weeks.

# Abbreviations

ESRI: Environmental Systems Research Institute; FUA: Functional Urban Area; GIS: Geographic Information System; SITAR: Fast Action Territorial Information System.

# Declarations

## Acknowledgements

The main source is the daily microdata record from the health authorities (Government of Cantabria - Spain), information that it is allowed to be managed only for research work “Implementación de SITAR-COVID19 Cantabria. Sistema de Información Territorial de acción rápida en relación al COVID-19 en la Comunidad Autónoma de Cantabria” by the Research with Medicines Ethics Committee from Cantabria (CEIm) in June 2020.

## Authors' contributions

ODC designed and developed spatial analysis operations. VC designed and prepared thematic cartography. ODC, VC and DC made a substantial contribution to the interpretation of data analyzed. ODC and DC were involved in drafting and revising the manuscript for important intellectual content. ODC and DC critically revised drafts of the work. ODC, VC and DC approved the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

## Fundings

This research was funded by Government of Cantabria (Spain) grant number UC: 10.3834.64001 called “*Asistencia en la adecuación de Cantabria al plan para la transición hacia una nueva normalidad en tiempos del Covid-19: aportaciones socioeconómicas*” University of Cantabria - IDIVAL Valdecilla - Government of Cantabria (Spain). And the APC was funded by IDIVAL support program.

## Availability of data and materials

The use permission of COVID-19 microdata conceded by the CEIm do not let us to share or publish original data.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## References

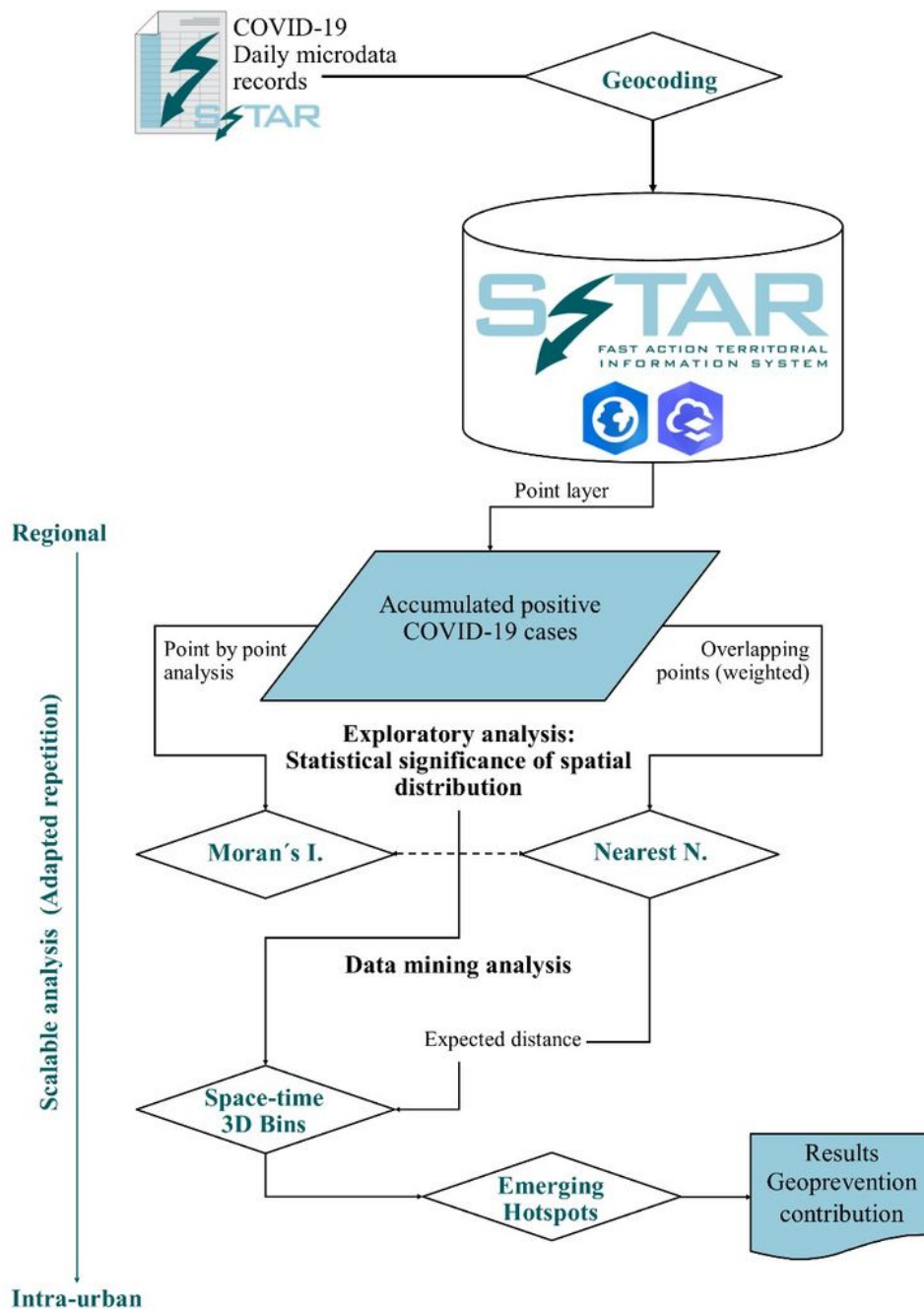
1. De Cos O, Castillo V, Cantarero D. Facing a Second Wave from a Regional View: Spatial Patterns of COVID-19 as a Key Determinant for Public Health and Geoprevention Plans. *J. Environ. Res. Public Health*. 2020;17:8468. <https://www.mdpi.com/1660-4601/17/22/8468> Accessed 1 Dec 2020.
2. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the Total Environment*. 2020;728:138884.
3. Jindal C, Kumar S, Sharma S, Choi YM, Efird JT. The Prevention and Management of COVID-19: Seeking a Practical and Timely Solution. *J. Environ. Res. Public Health*. 2020; 17:3986. <https://doi.org/10.3390/ijerph17113986> Accessed 10 Oct 2020
4. Chunbao M, Dechan T, Tingyu M, Chunhua B, Jian Q, Weiyi P, Zhiyong Z. An analysis of spatiotemporal pattern for COVID-19 in China based on space-time cube. *Journal of Medical Virology (J Med Virol.)*. 2020;92:1587-1595.
5. Abdrakhmanov SK, Mukhanbetkaliyev YY, Korennoy FI, Karatayev BS, Mukhanbetkaliyeva AA, Abdrakhmanova AS. Spatio-temporal analysis and visualization of the anthrax epidemic situation in livestock in Kazakhstan over the period 1933-2016. *Geospatial Health*. 2017;12(589):316-329.
6. Youlin Z, Liang G, Junwei L, Honghui L, Lei Y, Ning W, Yijun Z, Xu D. Analyzing hemorrhagic fever with renal syndrome in Hubei Province, China: a space-time cube-based approach. *Journal of*

- International Medical Research. 2019;47(7):3371-3388.
7. Sayed MH, Masoud P, Masoud B, Manoochehr K, Meysam O. Pattern mining analysis of pulmonary TB cases in Hamadan province: Using space-time cube. *International Journal of Epidemiologic Research*. 2017;4(2):111-117.
  8. Kamel MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21<sup>st</sup> century GIS technologies are supporting the global fight against outbreaks and epidemics. *Int J Health Geogr*. 2020;19:8. <https://doi.org/10.1186/s12942-020-00202-8> Accessed 1 Sep 2020.
  9. Zhou C, Su F, Pei T et al. COVID-19: Challenges to GIS with Big Data. *Geography and* 2020;1:77-87.
  10. Buzai GD. De Wuhan a Luján. *Evolución especial del COVID-19. Posición*. 2020;3:1-21. <https://tinyurl.com/yxdch6am> Accessed 5 Sep 2020.
  11. Sarwar S, Waheed R, Sarwar S, Khan A. COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions? *Science of the Total Environment*. 2020;730:139089.
  12. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. *Sci. Total Environ*. 2020;739:140033.
  13. Hemant B, Anjum S, Kumar S, et al. Understanding COVID-19 transmission through Bayesian probabilistic modeling and GIS-based Voronoi approach a policy perspective. *Environment, Development and Sustainability*. <https://doi.org/10.1007/s10668-020-00849-0> Accessed 10 Oct 2020.
  14. Cássaro F, Pires L. Can we predict the occurrence of COVID-19 cases? Considerations using a simple model of growth. *Science of the Total Environment*. 2020;728:138834
  15. Sánchez-Villegas P, Daponte-Codina A. Modelos predictivos de la epidemia de COVID-19 en España con curvas de Gompertz. *Gaceta Sanitaria*. 2020. <https://doi.org/10.1016/j.gaceta.2020.05.005> Accessed 15 Oct 2020.
  16. Whittle RS, Díaz-Artiles A. An ecological study of socioeconomic predictors of detection of COVID-19 cases across neighborhoods in New York City. *BMC Medicine* 2020;18: 271. <https://doi.org/10.1186/s12916-020-01731-6> Accessed 15 Sep 2020.
  17. Dhaval DD. Urban Densities and the COVID-19 Pandemic: Upending the Sustainability Myth of Global Megacities. *ORF Occasional Paper*. 2020;244:1-42.
  18. Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 Pandemic? Early findings and lessons for planners. *Journal of the American Planning Association*. 2020. DOI: 10.1080/01944363.2020.1777891
  19. Bamweyana I, Okello DA, Ssengendo R. Socio-Economic Vulnerability to COVID-19: The spatial case of Greater Kampala Metropolitan Area (GKMA). *Journal of GIS*. 2020;12:302-318 <https://doi.org/10.4236/jgis.2020.124019> Accessed 5 Oct 2020.

20. Borjas GJ. Demographic determinants of testing incidence and COVID-19 infections in New York city neighborhoods. National Bureau of Economic Research 2020; Working Paper: 26952.
21. Ogneva-Himmelberger Y. Spatial analysis of drug poisoning deaths and Access to substance-use disorder treatment in the United States. Proceedings of the 5th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2019):315-321. DOI: 10.5220/0007828703150321
22. Batista F, Poelman H. Mapping Population Density in Functional Urban Areas. A Method to Downscale Population Statistics to Urban Atlas Polygons. JRC Technical Reports. 2016. European Commission. <https://tinyurl.com/y9r2mbkt> Accessed 10 May 2020.
23. Kang Y, Cho N, Son S. Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation. PLoS ONE. 2018;13(5) <https://org/10.1371/journal.pone.0196845> Accessed 2 Nov 2020.
24. Hashim H, Wan-Mohd WMN, Sadek ESSM, Dimyati KM. Modeling urban crime patterns using spatial space time and regression analysis. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 2019;XLII-4/W16:247–254. <https://doi.org/10.5194/isprs-archives-XLII-4-W16-247-2019> Accessed 15 Oct 2020.

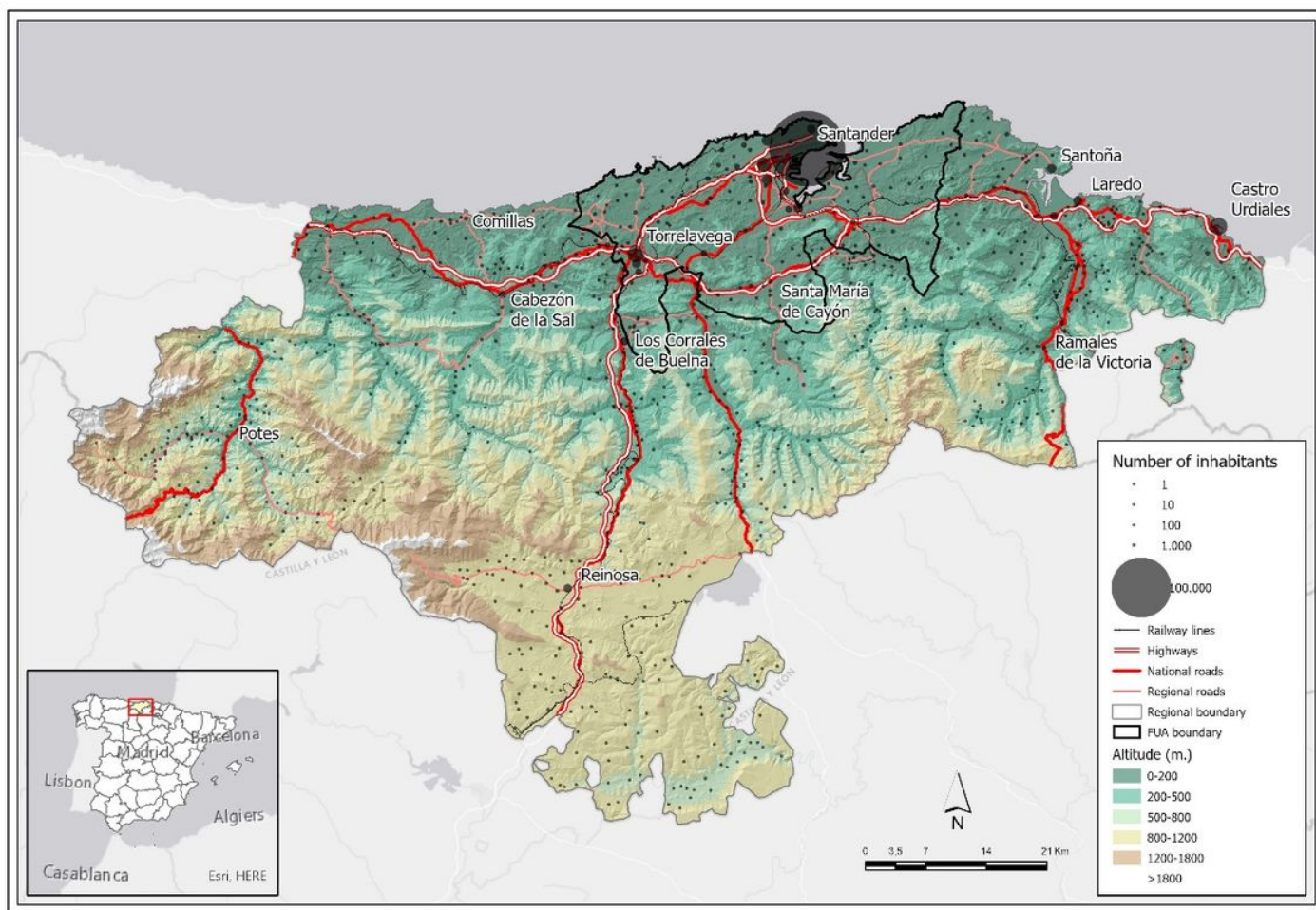
## Figures





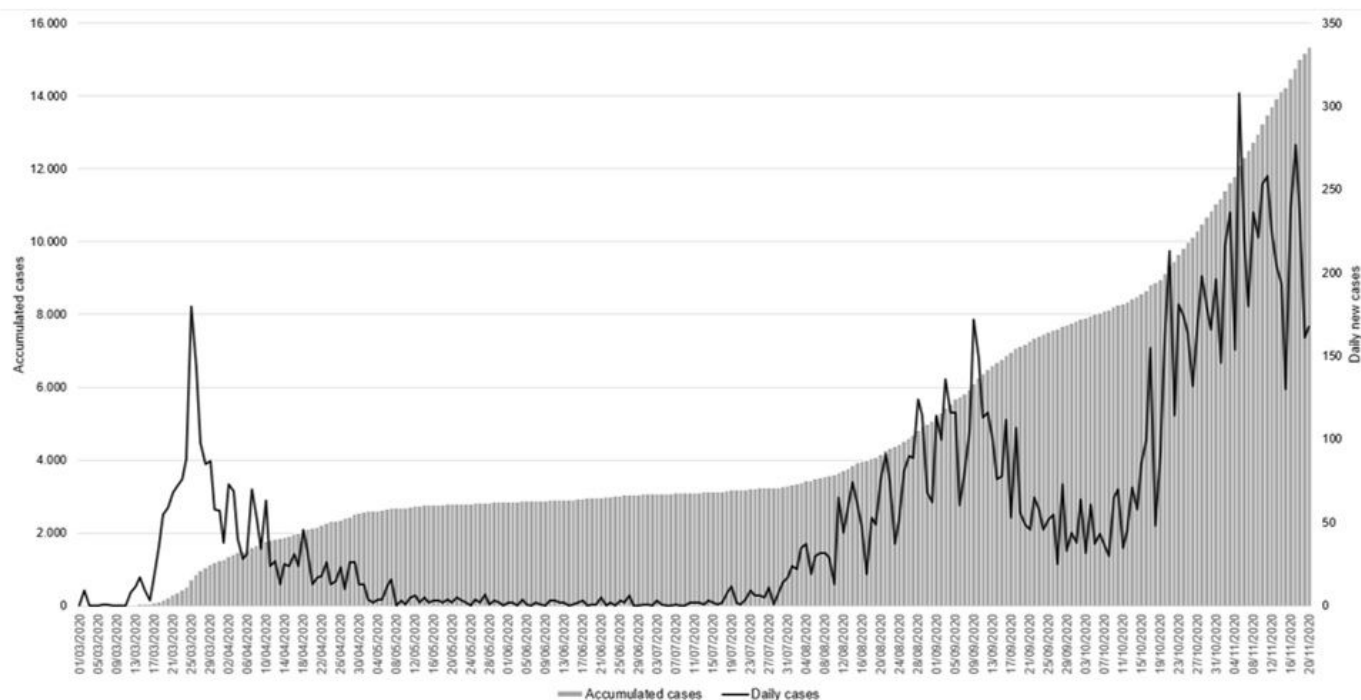
**Figure 1**

Methodological design of the research based on SITAR and 3D bins analysis (data mining tools).



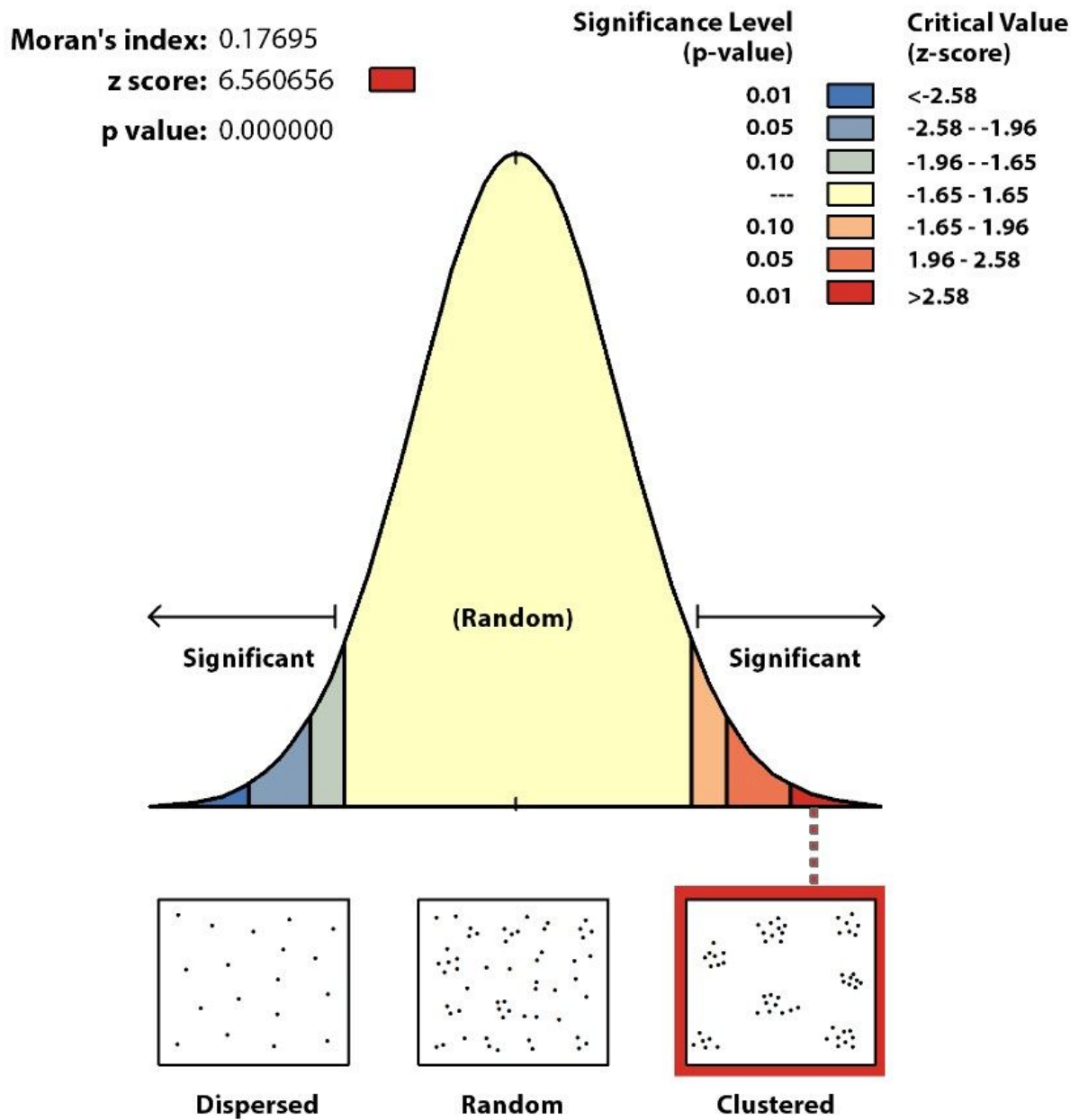
**Figure 2**

Territorial frame of Cantabria in the national context: settlements, FUA and main communication infrastructures. Source: Authors' elaboration based on ESRI (Administrative Base map), National Geographic Institute (National Cartographic Base 200) and Government of Cantabria (Digital Elevation Model). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



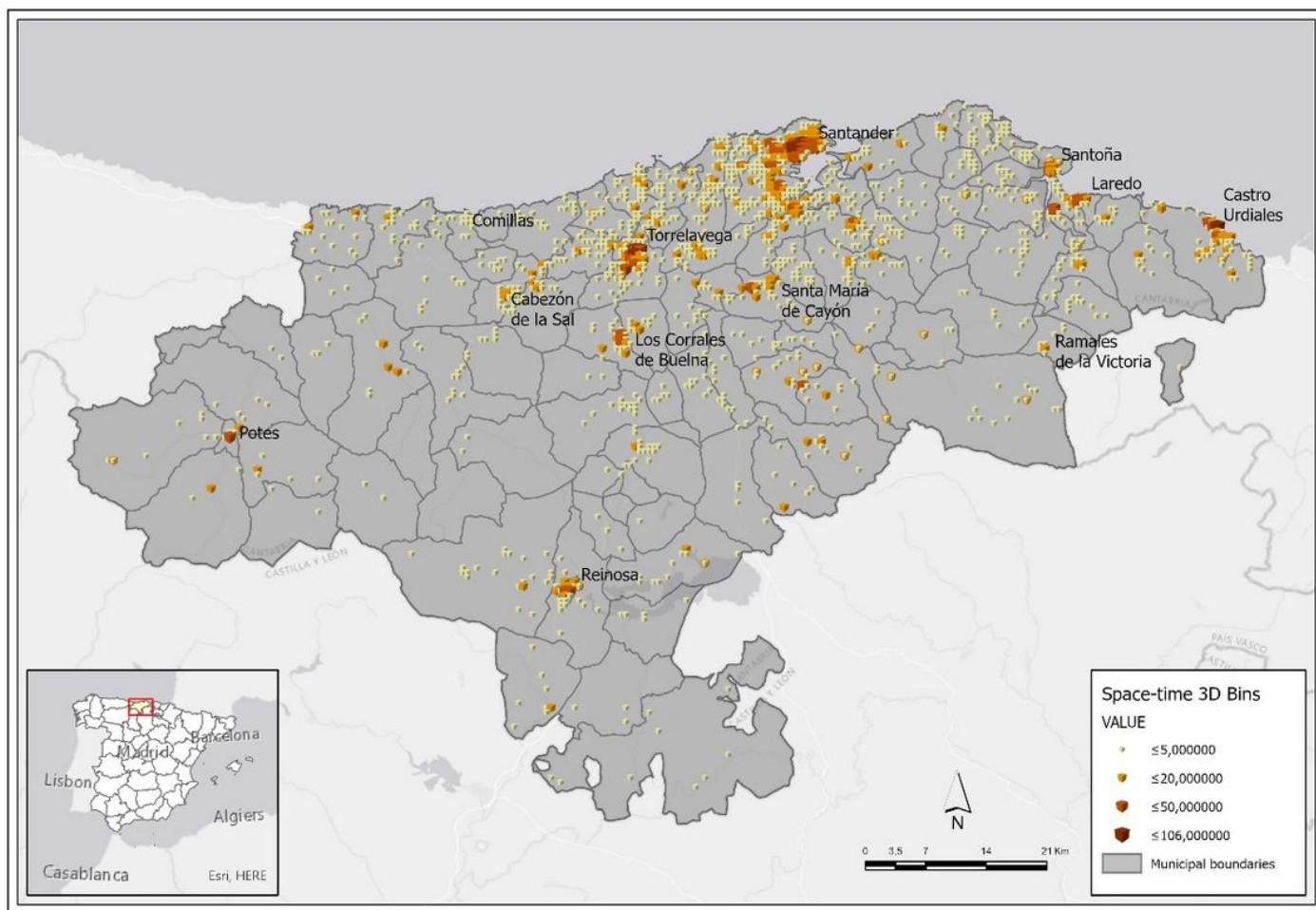
**Figure 3**

Daily and accumulated evolution of COVID-19 cases in the Community of Cantabria (March – November). Source: Authors' elaboration based on Historic register of evolution of COVID-19 in Cantabria (Open Data from the Government of Cantabria).



**Figure 4**

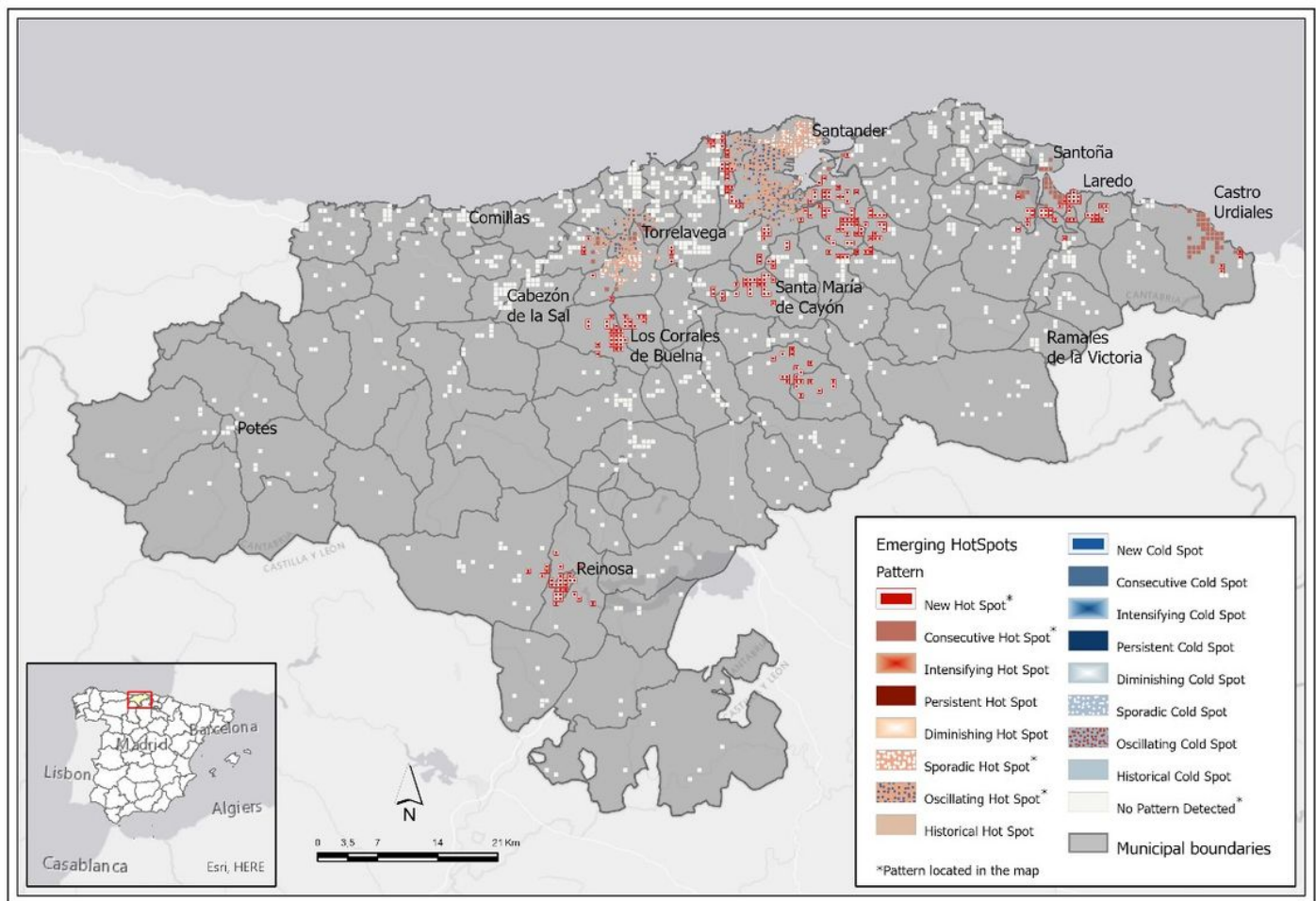
Graphic results of COVID-19 Moran's Index: Normal distribution and typology of spatial pattern. Source: Authors' elaboration based on COVID-19 Microdata Daily Register from the Health Authorities (Government of Cantabria, Spain).



**Figure 5**

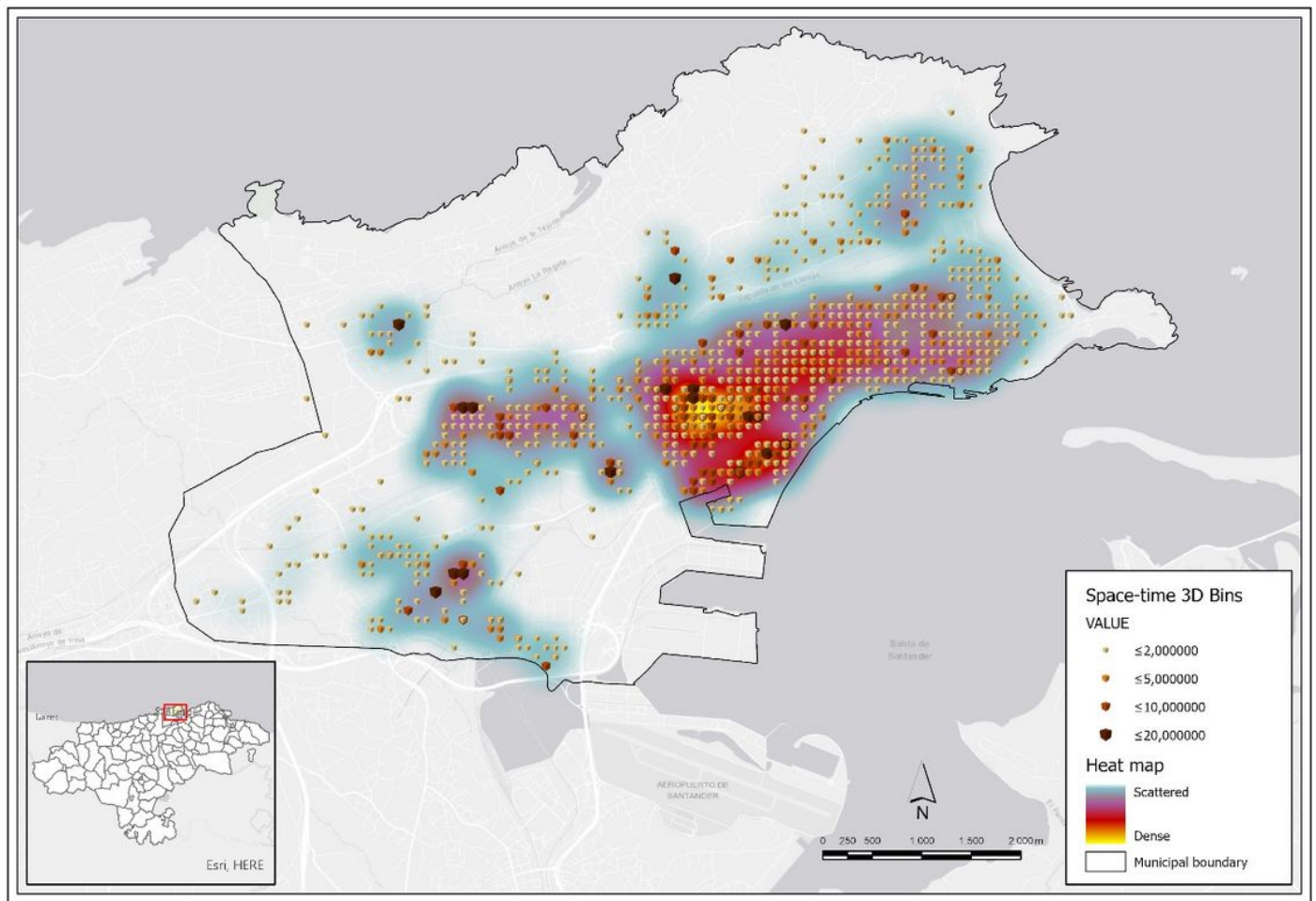
Space-time 3D Bins of COVID-19 in the Autonomous Community of Cantabria (March – November 2020). Source: Authors’ elaboration based on COVID-19 Microdata Daily Register from the Health Authorities (Government of Cantabria, Spain) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.





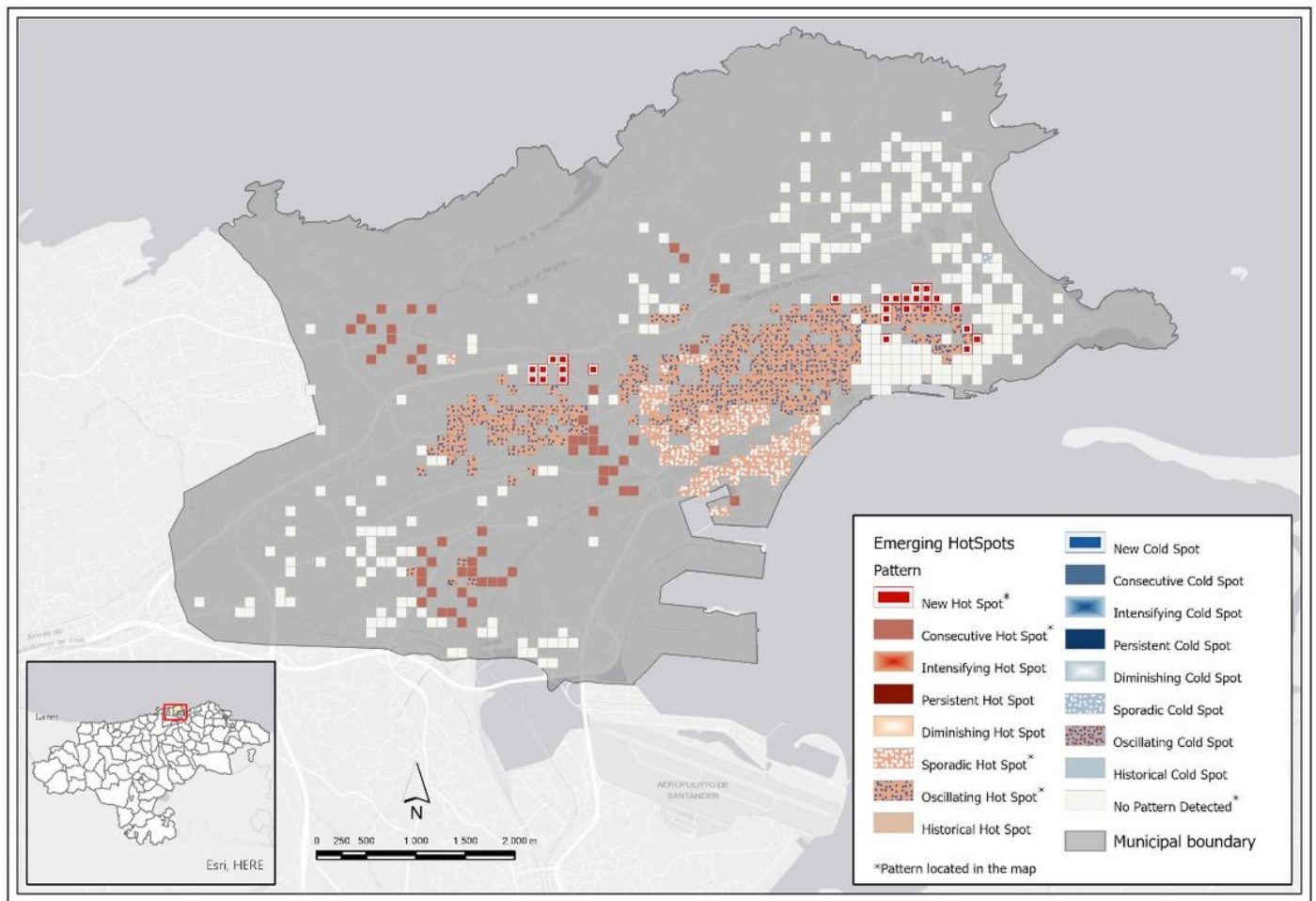
**Figure 6**

Emerging Hotspots of COVID-19 in the Autonomous Community of Cantabria (March – November 2020). Source: Authors' elaboration based on COVID-19 Microdata Daily Register from the Health Authorities (Government of Cantabria, Spain) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 7**

Space-time 3D Bins of COVID-19 in the municipality of Santander (August – November 2020). Source: Authors' elaboration based on COVID-19 Microdata Daily Register from the Health Authorities (Government of Cantabria, Spain). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 8**

Space-time 3D Bins of COVID-19 and heat map in the municipality of Santander (August– November 2020). Source: Authors' elaboration based on COVID-19 Microdata Daily Register from the Health Authorities (Government of Cantabria, Spain). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.