Geospatial Health 2022; volume 17(s1):1067

A geographical information system model to define COVID-19 problem areas with an analysis in the socio-economic context at the regional scale in the North of Spain

Olga De Cos-Guerra,[1,2] Valentín Castillo-Salcines,[1,2] David Cantarero-Prieto[2,3]

[1]Department of Geography, Urban and Regional Planning, Universidad de Cantabria; [2]Research Group on Health Economics and Health Services Management - Marqués de Valdecilla Research Institute (IDIVAL); [3]Department of Economics, Universidad de Cantabria, Santander, Spain

Abstract

The work presented concerns the spatial behaviour of coronavirus disease 2019 (COVID-19) at the regional scale and the socio-economic context of problem areas over the 2020-2021 period. We propose a replicable geographical information systems (GIS) methodology based on geocodification and analysis of COVID-19 microdata registered by health authorities of the Government of Cantabria, Spain from the beginning of the pandemic register (29th February 2020) to 2nd December 2021. The spatial behaviour of the virus was studied using ArcGIS Pro and a 1x1 km vector grid as the homogeneous reference layer. The GIS analysis of 45,392 geocoded cases revealed a clear process of spatial contraction of the virus after the spread in 2020 with 432 km^2 of problem areas reduced to 126.72 km^2 in 2021. The socio-economic framework showed complex relationships between COVID-19 cases and the explanatory variables related to household characteristics, socio-economic conditions and demographic structure. Local bivariate analysis showed fuzzier results in persistent hotspots in urban and peri-urban areas. Questions about 'where, when and how' contribute to learning from experience as we must draw inspiration from, and explore connections to, those confronting the issues related to the current pandemic.

Correspondence: Olga De Cos-Guerra, Department of Geography, Urban and Regional Planning, Universidad de Cantabria. Av. De los Castros s/n. 39005-Santander, Cantabria, Spain.
Tel.: +34.942.202074. E-mail: olga.decos@unican.es

Key words: Geocodification; geographic information systems; emerging hotspots; geoprevention; ArcGIS GeoEnrichment Service; Spain.

Funding: this research was funded by the research project INNVAL20/03 (IDIVAL) entitled '*Stress test in the Health System of Cantabria, development of innovative digital technologies to model scenarios of greater health care utilization and socioeconomic and human impact solutions against COVID-19*'.

Contributions: ODC, conceptualization, methodology, spatial analysis, and supervision; VC, ODC, visualization and publishing (graphs and maps); ODC, DC, writing of original draft and review. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: the authors declare no potential conflict of interest.

Ethical considerations: the use of COVID-19 data was accepted for research work by the Research with Medicines Ethics Committee, Cantabria (CEIm) in June 2020 (ID: 2020.238) and continues from September 2021 (CEIm Minutes: 14/2021).

Received for publication: 28 December 2021.
Revision received: 16 February 2022.
Accepted for publication: 16 February 2022.

Geospatial Health 2022; 17(s1):1067
doi:10.4081/gh.2022.1067

Introduction

In the Christmas period 2021, Spain suffered its sixth wave of the coronavirus disease 2019 (COVID-19) pandemic, which changed for the worse with the new the Omicron variant. With a population close to 47.5 million in 2021 (Register of Residents, National Statistics Institute), Spain reached 5.4 millions of COVID-19 cases in December 2021 despite 37.7 million (79.4%) fully vaccinated people (Health Ministry of the Government of Spain).

The management of the pandemic was originally centralised in Spain but moved to the regional health authorities from the end of the strict lockdown in June 2020. In this so called 'new normal' period, each Spanish region set different rules and restrictions in place to overcome each wave of the pandemic. However, the regions coordinated part of the decisions with the Inter-territorial Council of the Government of Spain, which supported part of the regional management with the declaration of two 'states of emergency' from 14 March 2020 to 21 June 2020 and from 25 October 2020 to 9 May 2021. This was done as the Inter-territorial Council constituted the only possible legal framework to establish stricter rules from regional governments and mobility restrictions between regions.

Considering that the region is the reference area of pandemic management in Spain, scientific research must pay attention to its borders. That means that our contribution of new knowledge to help policymakers to design efficient measures mitigating the spread of COVID-19 would be based on local custom approaches and balancing economy and health as proposed by Campagna (2020). Thus, our research was a collaboration between the

University of Cantabria, the Valdecilla Hospital Research Institute (IDIVAL) and the Department of Health of the Regional Government of Cantabria. In line with this collaborative framework, the study area was the Region of Cantabria (Northern Spain). The main goal of our research was based on spatial statistics and geographic information systems (GIS), implemented by ArcGIS Pro from ESRI-Spain COVID-19 GIS Hub, ArcGIS GeoEnrichment Service. We wished to reveal the key spatial spatiotemporal trend patterns of COVID-19 at the regional scale in 2020 and 2021 and approached the problem areas using their socio-economic context as an exploratory research line to identify variables correlated to the number of COVID-19 cases.

The potential of hotspot analysis based on mapping of COVID-19 microdata, a tabular structure of test results continually recorded by the health authorities with fields giving data such as age, sex and date for onset and cure, was explored to differentiate the problem areas. Testing represents key data in pandemic management (Bergquist *et al*., 2020), while cartography can reveal the modality of viral spread since it includes spatial and temporal trends and distinguishes significant spatial patterns. Focusing on areas with hotspots, we could obtain specific, significant hotspot patterns (new, consecutive and intensifying) that appeared along the timeline, such as in the initial time slices (historical), with many significant repetitions over time (sporadic) or with significant hotspot behaviour in more than 90% of the study time (persistent). This approach showed problem areas with interesting spatial and statistical nuances, which is essential to overcome the pandemic challenges.

We have previously used 3D bin analysis of COVID-19 (2021b) based on geocoded microdata, and others have applied it at the city level (Chunbao *et al*., 2020), the county level (Tokey, 2021) and, recently, the sub-district level (Syetiawan *et al*., 2022). It is true that this approach presents methodological risks with respect to the parameters (spatial and time units) (Kulldorff, 2001), which led us to propose non-subjective criteria to define the relative bin size based on expected distance of cases using nearest neighbour analysis, and the internal time periods by dividing the overall period by time units of 2-4 weeks as reference indicator of cumulative incidence.

The methodological proposal began with the geocodification of daily microdata of COVID-19 positive cases registered by the health authorities of the Government of Cantabria. We accumulated 20-months series covering the years 2020 and 2021 with high spatial and temporal resolutions. This allowed a spatially homogeneous analysis based on a 1x1-km vector grid to compare models in both years contextualising the socio-economic variables supported by the ArcGIS GeoEnrichment Service from ESRI-Spain COVID-19 GIS Hub.

The global body of GIS literature, geostatistical methods and spatial patterns on COVID-19, considers several perspectives and scales from the beginning of the pandemic, and presents an increasing trend focused on virus spread (Fatima *et al*., 2021; Franch-Pardo *et al*., 2021). The spatial analysis of COVID-19 includes counterintuitive data that challenge policymakers, who have to think globally but take decisions locally (Salama, 2020) to reduce the risk in the mitigation stage (Jindal *et al*., 2020). A multi-scale approach is necessary in the spatial analysis of the virus and location technology is essential to learn about spatial and epidemiological frameworks needed to design control of the disease (Gerber, 2009). In addition, maximum spatial resolution of geocoding cases can reveal local patterns avoiding distortions of aggregated data from health administrative units, census sections or postal code areas (Cromley, 2019). This approach provides other possible foci not only related to COVID-19 incidence and context variables, but also to socio-demographic characteristics of patients and hotspots (Mohammad Ebrahimi *et al*., 2021).

The main aim of research on COVID-19 in a social sciences context is related to the relation between COVID-19 incidence and demographic, socio-economic and environmental variables. Health geography highlights the importance of the environment in COVID-19 spatial behaviour, especially in urban areas (Das *et al*., 2021) with higher incidence in vulnerable areas, such as slums, especially during the first waves (Ferreira, 2020). Strong evidence links virus distribution and population density in urban areas (Dhaval, 2020; Hamidi *et al*., 2020; Niu *et al*., 2020; Desmet and Wacziarg, 2021). Additionally, some research considers the concentration of economic activities (Huang *et al*., 2020; Perles *et al*., 2021a), socio-economic characteristics (Bamweyana *et al*., 2020; Whittle and Díaz-Artiles, 2020) or environmental conditions, especially pollution (Sera *et al*., 2021). Some authors have linked population density with economic activities as key indicators of 'spatial concurrence' which is directly related to virus spread in urban areas (Buffalo and Rydzewski, 2021). Some interesting research on concurrence has used a spatial interaction index (Al Kindi *et al*., 2021) or applied space-time path and prism to study the movements of new cases in time and space (Yin *et al*., 2021).

Spatial policies and strategies not only focusing on mitigation (Ye and Hu, 2020), but also on containment measures (Coccia, 2021), constitute efficient approaches. Although important, the potential of GIS in simulations and forecasting for short-term predictions of pandemic evolution are not yet fully developed (Ahasan and Hossain, 2021). One of the most extended methods is geographically weighted regression (GWR) based on socio-demographic multi-variant approaches (Almendra *et al*., 2021). Nevertheless, scale effects are essential in the forecasting model by Dhaval (2020) and related to this, the use of administrative boundaries and aggregated data seems one of the major methodological risks in the spatial analysis of the pandemic (Buffalo and Rydzewski, 2021).

In this context, we aimed to analyse COVID-19 problem areas using microdata and implementing 3D bins and emerging hotspots models. It is essential to avoid aggregated data distortions (applying our method to microdata points instead of aggregated cases by administrative boundaries) and, on the other hand, our proposal contribute to make models comparable over time (using a regular 1x1-km vector grid in 2020 and 2021).

Materials and methods

Study site

This research focuses on the Region of Cantabria, located in the north of Spain. This region has a population of nearly 583,000 inhabitants in 2021 (Register of Residents of the National Statistics Institute) and a surface area of 5300 km^2. The Region of Cantabria presents an important contrast between mountainous features (orography) and population distribution. The coastal area shows higher population densities in contrast to the inland valleys that have higher elevation and less population (Figure 1). Santander, the capital city of Cantabria, is situated in the central coastal area with 173,000 inhabitants near Torrelavega, the second city with about 51,000 inhabitants. According to Batista and Poelman (2016), these two cities, together with over twenty municipalities with a total population of 380,000 (about 65% of the region), define a polycentric functional urban area (FUA). This dynamic area presents a prominent concentration of economic activities, population, transport infrastructures and, additionally, strong commuting between central areas and the surroundings. The intra-regional area has three coastal units and nine inland valleys in a dominant open landscape (BOE-A-2015-682, 2015).

Data

The daily recorded data of epidemic diseases are collected by the health authorities of the Government of Cantabria and kept in what is called the Fast Action Territorial Information System (SITAR). These so called microdata are presented as a tabular structure where each positive tested case of COVID-19 is an anonymised row and many fields give data for address, age, sex as well as dates for onset and cure/decease. The microdata table also includes two binary fields to identify if a positive case is in a care home or if it corresponds to a health professional. In July 2020 the research team implemented SITAR to study the spatial patterns of COVID-19 in the Region of Cantabria using ArcGIS Pro and ArcGIS Online. We geocoded the records of cases testing positive for COVID-19 from the beginning of the official register (29th February 2020) until 2nd December 2021.

The data structure of SITAR includes three thematic geo-databases (GDB) that concern health, socio-demographic context and buildings (De Cos *et al*., 2020). Health GDB is necessary for this work because it includes administrative health areas and location of fundamental facilities (health centres, care homes and pharmacies). Socio-demographic GDB integrates data about demographic structure and socio-economic variables from official sources (Census and Register of Residents of the National Statistics Institute) and, additionally, data from ESRI based on big data (De Cos *et al*., 2021a). Finally, SITAR contains a cadastral GDB at the building level with data referred to economic activities by residential areas, numbers of floors, *etc*.

Approach

We designed a method in three stages for the 20 month study period as shown in Figure 2: i) data entry and layer preparation; ii) identification of COVID-19 problem areas; and iii) socio-economic variables in the identified problem areas.

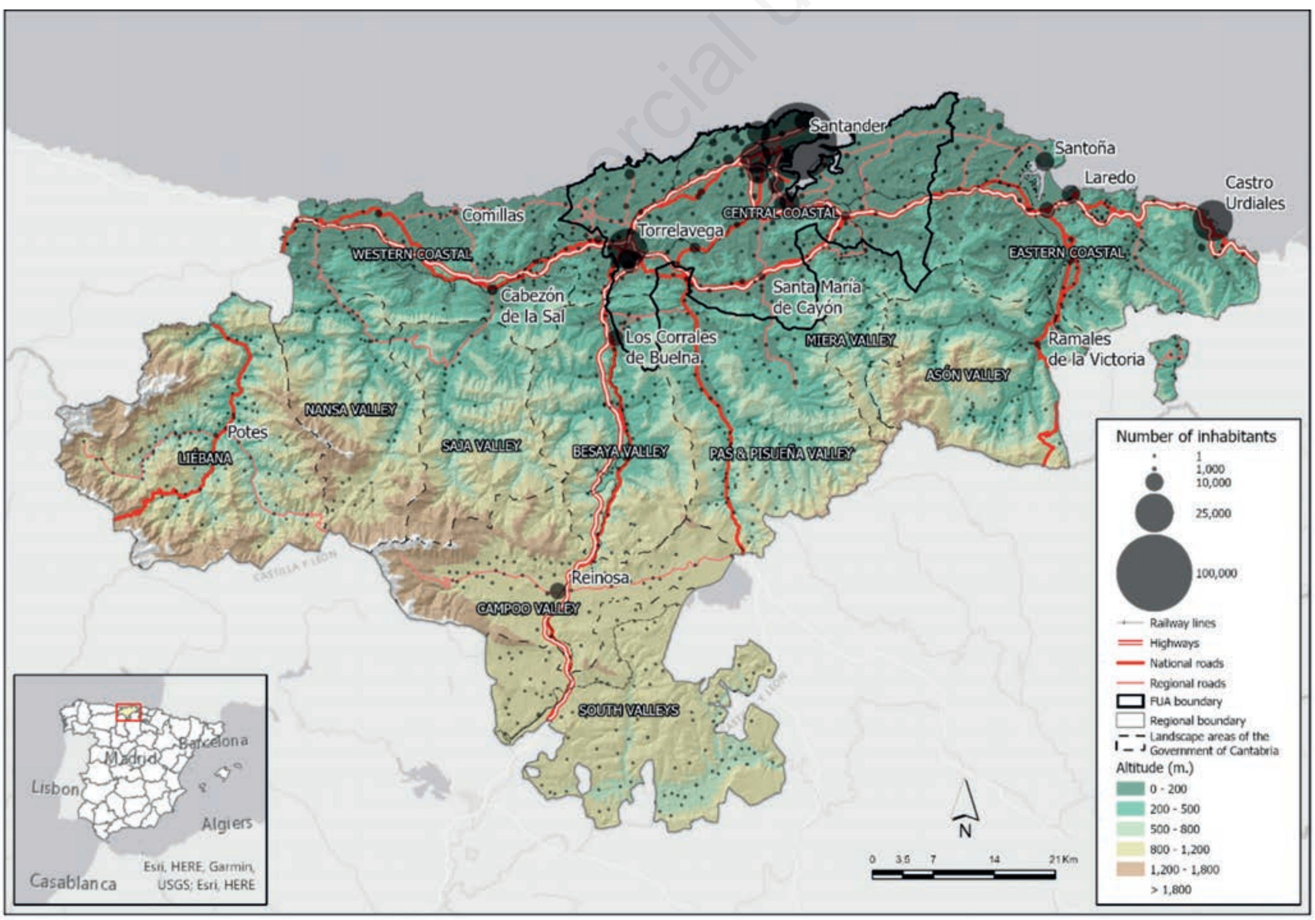

Figure 1. Territorial frame of the Region of Cantabria (North Spain). ***Sources: ESRI (Administrative Base map), National Geographic Institute (National Cartographic Base 200), Government of Cantabria (Digital Elevation Model), National Institute of Statistics (Data from Register of Residents, 2021) and the European Union (Urban Atlas, 2018).***

Data entry and layer preparation

We geocoded the microdata of the COVID-19 cases with ArcGIS based on multiple-field entries (address, locality, municipality and postal code). The new point layer of cases was revised and filtered to delete cases in care homes because the pattern of spread there is different from the rest of the conventional neighbourhoods (De Cos *et al.*, 2021b; Perles *et al.*, 2021a). The geocoded cases were divided in two different layers filtering by the starting year continuing the next stages in parallel for 2020 and 2021. We prepared a regular structure of units with the Region of Cantabria as a base map using a 1x1-km pixel vector grid to analyse and compare the spatial patterns in the COVID-19 problem areas.

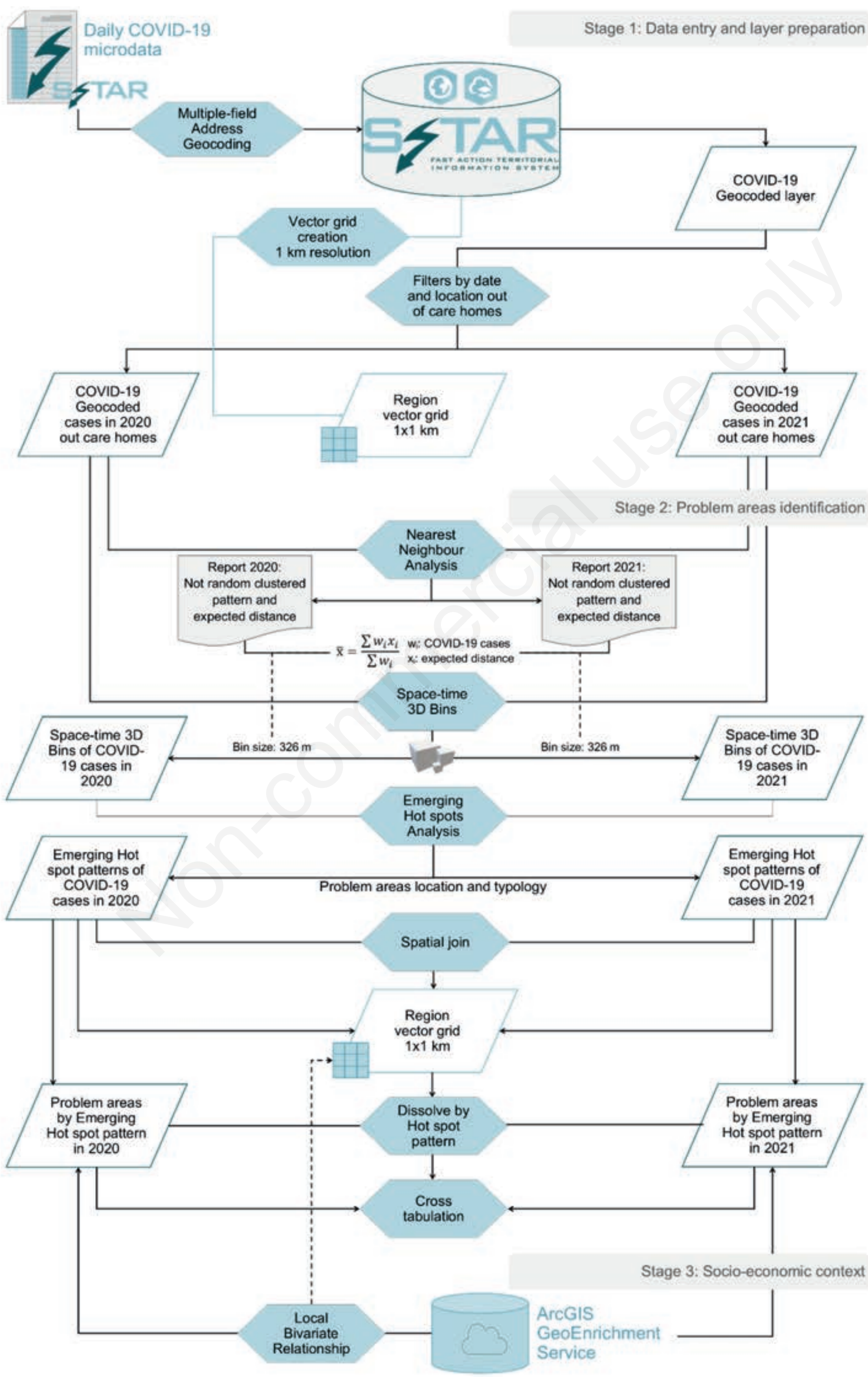

Figure 2. Methodological stages based on microdata and geographical information system analysis.

Identification of the problem areas

The second stage started with a preliminary study of the statistical pattern of COVID-19 cases. We explored the non-randomness of the distribution to guarantee valid spatial pattern results. To achieve this, we applied the extended nearest neighbour analysis that confirmed the statistically significant and clustered distribution of COVID-19 cases for both years. On this basis, our method highlighted the use of space-time cluster analysis using the ArcGIS 3D bin approach with each bin including the two dimensions of the pandemic (where and when) based on the ArcGIS regular space-time cube structure where each location and time step have precise IDs. Bins sharing the same location ID represent a time series, while bins sharing the same time-step ID comprise a time slice. In this way we could represent each place and time with a specific COVID-19 situation. We based the method parameters (spatial size and time slices) on non-subjective criteria and standardised approaches to ensure 'exportability'.

In our approach, the spatial size of each bin was determined by the expected distance obtained in the exploratory nearest neighbour analysis. We used the weighted average of expected distance in 2020 and 2021 (*i.e*. 326 m), while the internal time slices for each year (2020 and 2021) required a minimum of ten periods in the ArcGIS tool for the bin creation. Additionally, we considered common periods used by the health authorities to calculate cumulative incidences (two weeks) and, finally, we simplified the scheme further by using internal periods of four weeks.

The identification of problem areas was based on hotspot analysis. Here, we should point out that the space-time trend could result in a maximum of 16 significant patterns (8 hotspot or 8 coldspot types). Furthermore, the method also identifies a type called 'no pattern detected', which defines areas where the virus existed, but without a significant trend. Problem areas correspond to hotspot patterns where the space-time trend of COVID-19 cases increases significantly. In this stage we compared and cross-tabulated hotspots in 2020 and 2021 to identify the space-time process by pattern. Here, it is essential to adjust the hotspot analysis to the 1x1-km vector grid as homogeneous reference for spatial units.

Socio-economic context of problem areas

Areal interpolation and weighting by population are essential. To conclude our GIS methods proposal, we enriched the polygons representing the problem areas by ESRI's GeoEnrichment Service Data that adapts data from original areas (census sections, for instance) to the areas that users provide. These areas are spatially related to the original data as overlapped areas (total or partially), then each selected variable was estimated by the area overlapped and, finally, it considered the presence of buildings (and indirectly the population living there).

In relation to the variables, we considered the household characteristics at this stage (average size; total spending; and percentage of unemployment); socio-economic conditions (poverty risk by age group; average income per capita; average selling price per m^2; and average rental charges per m^2); demographic structure (population age rate; proportion of young people; and proportion of adults); and residential environment (type of residential property; number of family homes; and number of economic activities). Here, we joined the variables at the spatial 1x1-km grid level and analysed local bivariate relationships in order to get a deeper understanding of the COVID-19 incidence relation with the socio-economic context. This would also avoid distortions caused by the non-stationary behaviour of COVID-19 distribution (Mou *et al*., 2017). Local bivariate relationship is a useful approach to analyse COVID-19 distribution in relation to socio-economic variables for two main reasons: non-linearity and non-stationarity. In contrast to other valid and frequently used methods, such as ordinary least square (OLS) and GWR, local bivariate analysis is not only focused on linear relationships. Indeed, we used local bivariate analysis as a useful method to provide not only spatial intensity of correlation, but also the pattern of the relationship (positively linear, negatively linear, concave, convex, undefined complex or not significant). Furthermore, in previous research on the spatial patterns of COVID-19 at detailed scales, we have demonstrated the non-stationarity of COVID-19 distribution with a statistical significance ($P<0.01$) of the Koenker Index (De Cos *et al*., 2021b). Although GWR is appropriate to analyse non-stationary data, in our research, we do not pretend a predictive or multivariate model in contrast to other studies (Jardim de Figueiredo *et al*., 2022).

Results

The Region of Cantabria recorded 49,556 COVID-19 cases and 623 direct deaths during the 20-month study period. As shown in Table 1, our results are based on the study of 45,392 cases (91.6% of all) due to two main reasons: the geocodifying success and the need to avoid cases from care homes that would have added spatial distortions due to the different spread modality there (De Cos *et al*., 2021a). Thus, this research focused on the spatial pattern of COVID-19 cases in general neighbourhoods.

With regard to the evolution of the pandemic between 2020 and 2021 as shown in Figure 3A, six waves with many peaks and short valley periods can be noted but there are no clear differences between the years. However, as shown in Figure 3B, control measures and the vaccination process had a clear effect on the number of severe cases both with respect to hospitalised patients and those in Intensive Care Units (ICU) as the peaks became increasingly low in 2021 in contrast to the high values in 2020. This clearly shows the recent reduction of virus severity (Jindal *et al*., 2020), which is especially relevant when the same annual dates are compared year by year (first wave front compared to the fourth wave

Table 1. Registered positive COVID-19 cases.

	Global register		Geocoded cases in Cantabria	
	General	In care homes	Total	Not in care homes
COVID-19 positive cases	49,556	2118	47,510*	45,392
Percentage	100.00%	4.27%	95.87%	91.60%

*Geocodification notes: 1175 cases had unknown address in the microdata register, 729 cases had no matched address, and 142 cases were not Cantabria residents and were therefore deleted. *Source: open data from the Government of Cantabria, Spain and COVID-19 microdata daily records from the health authorities (Government of Cantabria, Spain).*

and especially the second wave front compared to the sixth wave). Here, we wish to point out that the higher values of the fourth and fifth waves (2021) with respect to 2020 do not contradict our interpretation. Truly, that period in 2020 (from March to June) corresponds to the strict confinement with a clear valley between the first and second waves.

Problem areas patterns based on space-time trends

As we stated previously, our methodological approach suggests that the problem areas correspond to hotspot patterns, where the space-time trend of COVID-19 cases is significant and increasing. On this basis, our cartographic results show several differ-

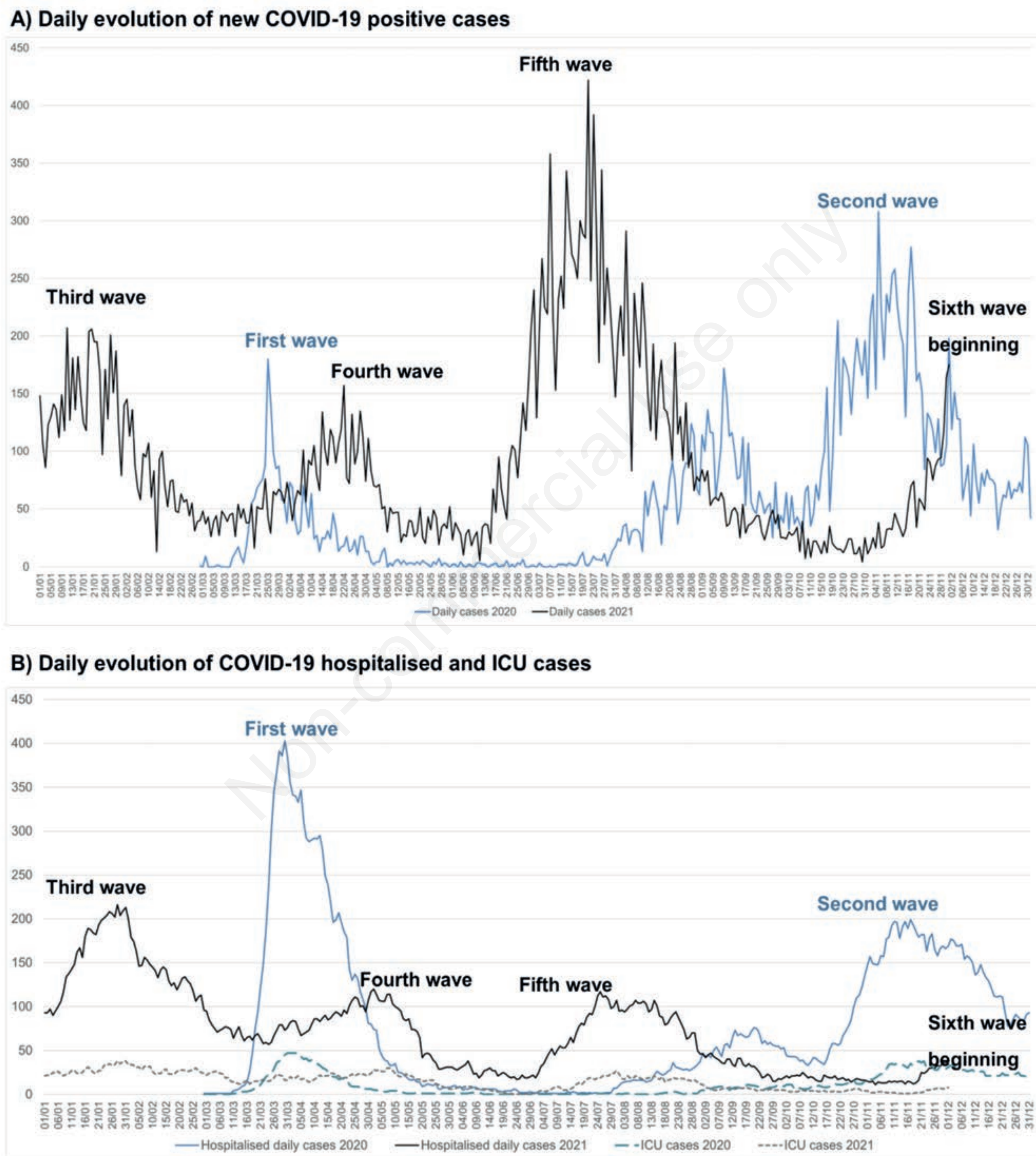

Figure 3. A-B) Evolution of COVID-19 series: a comparison between 2020 and 2021. ***Source: open data on COVID-19 in Cantabria from the Government of Cantabria, Spain.***

ences with respect to distribution and patterns in the COVID-19 problem areas. As shown in Figure 4, the space-time trend of the pandemic in 2020 reveals a process of increasing spread with different patterns of hotspots:

- City centres of the most populated cities (Santander and Torrelavega) show an increasing trend of cases with intensifying hotspots in more than 90% of the time slices during 2020 without presence of coldspots.
- Close peri-urban arcs of sporadic hotspots in the Santander-Torrelavega FUA and the main urban areas along the oriental coast. This picture is part of a repetitive pattern of increasing trends of hotspots in 2020.
- The border areas of the Santander-Torrelavega FUA and service centres in the rural inland valleys show consecutive hotspots with a repetitive pattern in the problem areas, especially in the last time slices of 2020.

In contrast, the spatial patterns in 2021 were more controlled and simpler (Figure 5). The clustered persistent hotspots in the Santander-Torrelavega FUA included significant hotspots without coldspots in more than 90% of the temporal slices in 2021. Other similar problem areas appeared along the oriental coast near Bilbao, a close metropolitan area with over 1 million inhabitants.

Our quantitative results clearly demonstrate the process of a spatial contraction of the pandemic (Table 2). After 432 km^2 of significant hotspot patterns in 2020, their total expanse shrank back to 126.72 km^2 in 2021. Furthermore, our results showed absence of viral transmission in an increasing percentage of areas with cases lacking statistical space-time significance: areas without a pattern occupied 10.1% of Cantabria in 2020 and 17.7% in 2021. It is important to highlight that 41.8% of the cases in 2021 were in areas without a space-time significant trend in contrast to 15.0% in 2020. Another conspicuous aspect in Table 2 is the decrease of the average age of positive cases in 2021: 43.4 years in 2020 and 36.6 in 2021. The highest such age (46.7) appeared in 2020 in an area without patterns belonging to the old, rural part of Cantabria, where low population density and low mobility contained the spread, thereby limiting the development of significant hotspots.

Comparing the results of the cross-tabulation of emerging hotspot patterns in 2020 and 2021, we identified 17 combinations grouped in six classes: disappearing problem areas, evolution of significant patterns, new significant problem areas in 2021, no pattern detected any time, no evident trends and, finally, no cases (Table 3). More than three quarter of the region (76.8%) remained unchanged without cases in both years.

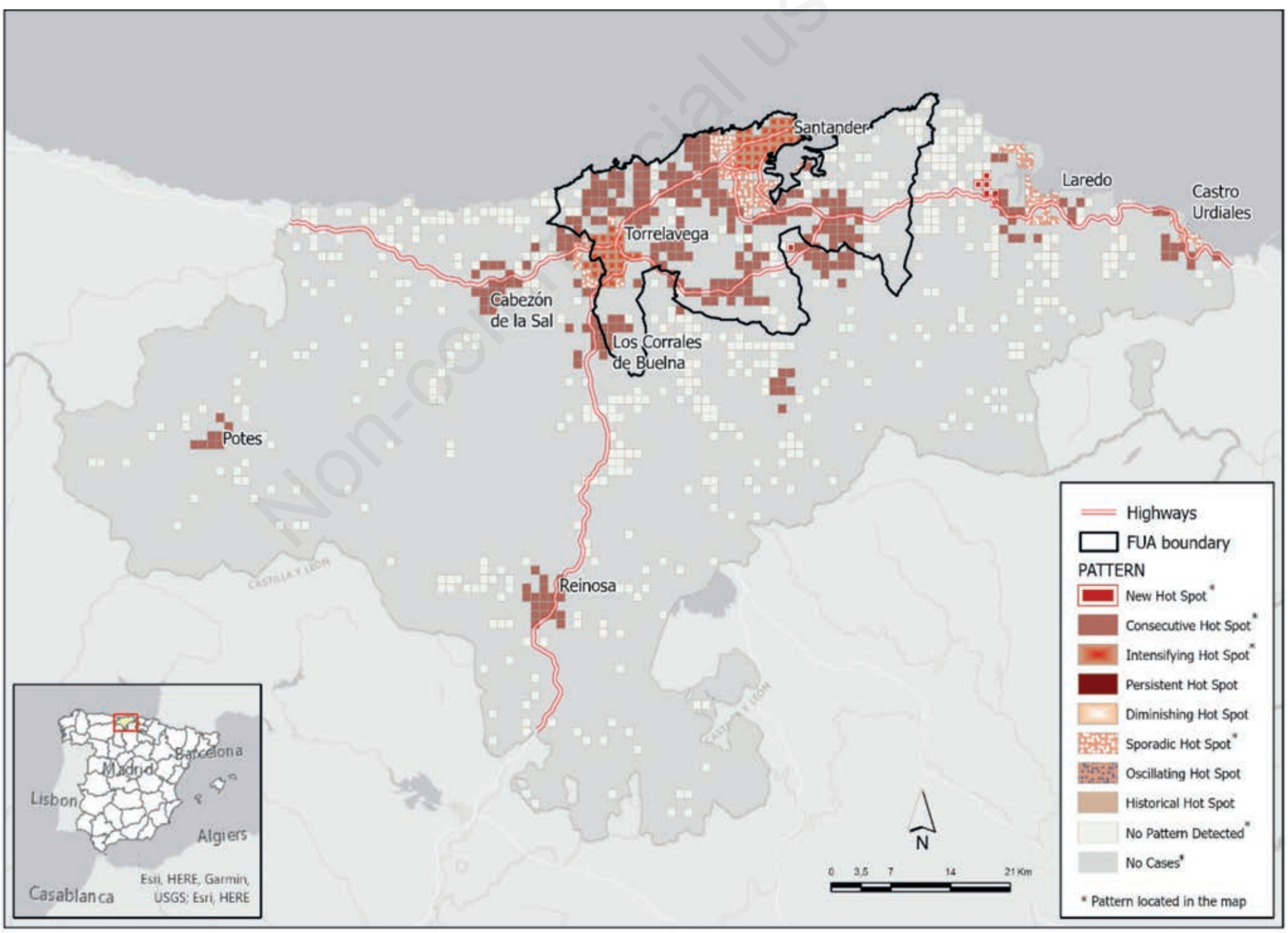

Figure 4. Space-time COVID-19 emerging hotspots in 2020. ***Sources: ESRI (Administrative Base map), National Geographic Institute (National Cartographic Base 200), Copernicus FUA layer and COVID-19 microdata daily records from the health authorities (Government of Cantabria, Spain).***

Table 2. Space-time COVID-19 emerging hotspots dimensions in 2020 and 2021.

COVID-19 space-time patterns in 2020	Area (km^2)	Area (%)	Cases (number)	Cases (%)	Average age
Statistically significant patterns	432.00	8.11	14,521	85.05	42.84
Consecutive hotspots	299.40	5.62	4182	24.49	43.75
Intensifying hotspots	54.68	1.03	7042	41.24	42.13
New hotspots	7.00	0.13	45	0.26	42.16
Sporadic hotspots	70.91	1.33	3252	19.05	44.76
No pattern detected	536.98	10.08	2553	14.95	46.68
No cases	4357.87	81.81	0	-	-
Total	5326.85	100.00	17,074	100.00	43.41
COVID-19 space-time patterns in 2021	**Area (km^2)**	**Area (%)**	**Cases (number)**	**Cases (%)**	**Average age**
Statistically significant patterns	126.72	2.38	16,480	58.20	36.07
Persistent hotspots	119.71	2.25	16,337	57.69	37.28
Sporadic hotspots	4.01	0.08	97	0.34	29.98
Historical hotspots	3.00	0.06	46	0.16	33.12
No pattern detected	942.83	17.70	11,838	41.80	37.24
No cases	4257.29	79.92	0	-	-
Total	5326.85	100.00	28,318	100.00	36.55

Source: COVID-19 microdata daily records from the health authorities (Government of Cantabria, Spain).

Table 3. Cross-tabulated patterns of COVID-19 hotspots 2020-2021.

Hotspot pattern 2020	Hotspot pattern 2021	Area km^2	Area %	Cases Number	Cases %
Disappearing problem areas		316.31	5.94	12,420	27.36
Consecutive hotspots	No pattern detected	259.94	4.88	9767	21.52
Sporadic hotspots	No pattern detected	18.37	0.34	2,460	5.42
New hotspots	No pattern detected	7.00	0.13	127	0.28
Consecutive hotspots	No cases	25.00	0.47	52	0.11
Sporadic hotspots	No cases	5.00	0.09	12	0.03
Intensifying hotspots	No cases	1.00	0.02	2	0.00
Evolution of significant patterns		115.69	2.17	25,996	57.27
Intensifying hotspots	Persistent hotspots	53.68	1.01	18,561	40.89
Sporadic hotspots	Persistent hotspots	43.54	0.82	6,553	14.44
Consecutive hotspots	Persistent hotspots	11.45	0.21	636	1.40
Sporadic hotspots	Historic hotspots	3.00	0.06	88	0.19
Consecutive hotspots	Sporadic hotspots	3.01	0.06	85	0.19
Sporadic hotspots	Sporadic hotspots	1.00	0.02	73	0.16
New significant problem areas in 2021					
No cases	Persistent hotspots	11.04	0.21	39	0.09
No pattern detected any time					
No pattern detected	No pattern detected	400.80	7.52	5953	13.11
No evident trends		392.91	7.38	984	2.17
No pattern detected	No cases	136.18	2.56	306	0.67
No cases	No pattern detected	256.72	4.82	678	1.49
No cases					
No cases	No cases	4090.11	76.78	0	0.00
	Total	5326.85	100.00	45,392	100.00

Source: COVID-19 microdata daily records from the health authorities (Government of Cantabria, Spain).

Focusing on problem areas we interpreted three main patterns that changed between 2020 and 2021:

- Some significant problem areas in 2020 disappeared in 2021 by changing into 'no pattern' or 'no cases'. This process affected 5.9% of the area, where 27.4% of the cases occurred during the study period.
- In a small part of the region (2.2%) significant patterns appeared during both years. Here, the main change was from intensifying and sporadic hotspots in 2020 to a persistent pattern in 2021. Importantly, this small area showed a significant hotspot trend and a high concentration of cases (25,996), which corresponds to 57.3% of all cases detected.
- As the pandemic contraction took hold when moving into 2021, a paradoxical retrograde development was observed in a limited part of the study area (0.2%). Although the new significant problem areas had changed from no cases in 2020 to only 39 in 2021, it still appeared as the persistent hotspot type.

Problem areas in socio-economic context

We analysed the socio-economic context of problem areas using data from the GeoEnrichment Service of ESRI-Spain COVID-19 GIS Hub. Although we accessed and analysed a high number of variables, we focused on the more relevant results. The preliminary analysis of linear bivariate correlations revealed that there was little or no association between the number of COVID-19 cases as the dependent variable and many explanatory variables using the 1x1-km pixel vector grid. However, we detected two exceptions: the population density correlated positively (0.78) with the number of COVID-19 cases, while the average selling price per m^2 correlated negatively (–0.95). The main difference between these explanatory variables is that density presented a homogeneous correlation from the spatial perspective and local bivariate relationships could therefore not be implemented. On the other hand, as shown in Figure 6A, local bivariate relationships between average selling price per m^2 and COVID-19 demonstrated local spatial differences. Globally, however, the spatial pattern showed linear negative relationships in the inland areas, rural service centres and part of coastal areas. Nevertheless, the relationships changed in the Santander-Torrelavega FUA, especially in the persistent problem area of Santander City and its surroundings. Additionally, south of Torrelavega and at the borders of the FUA, the relationships turned into undefined.

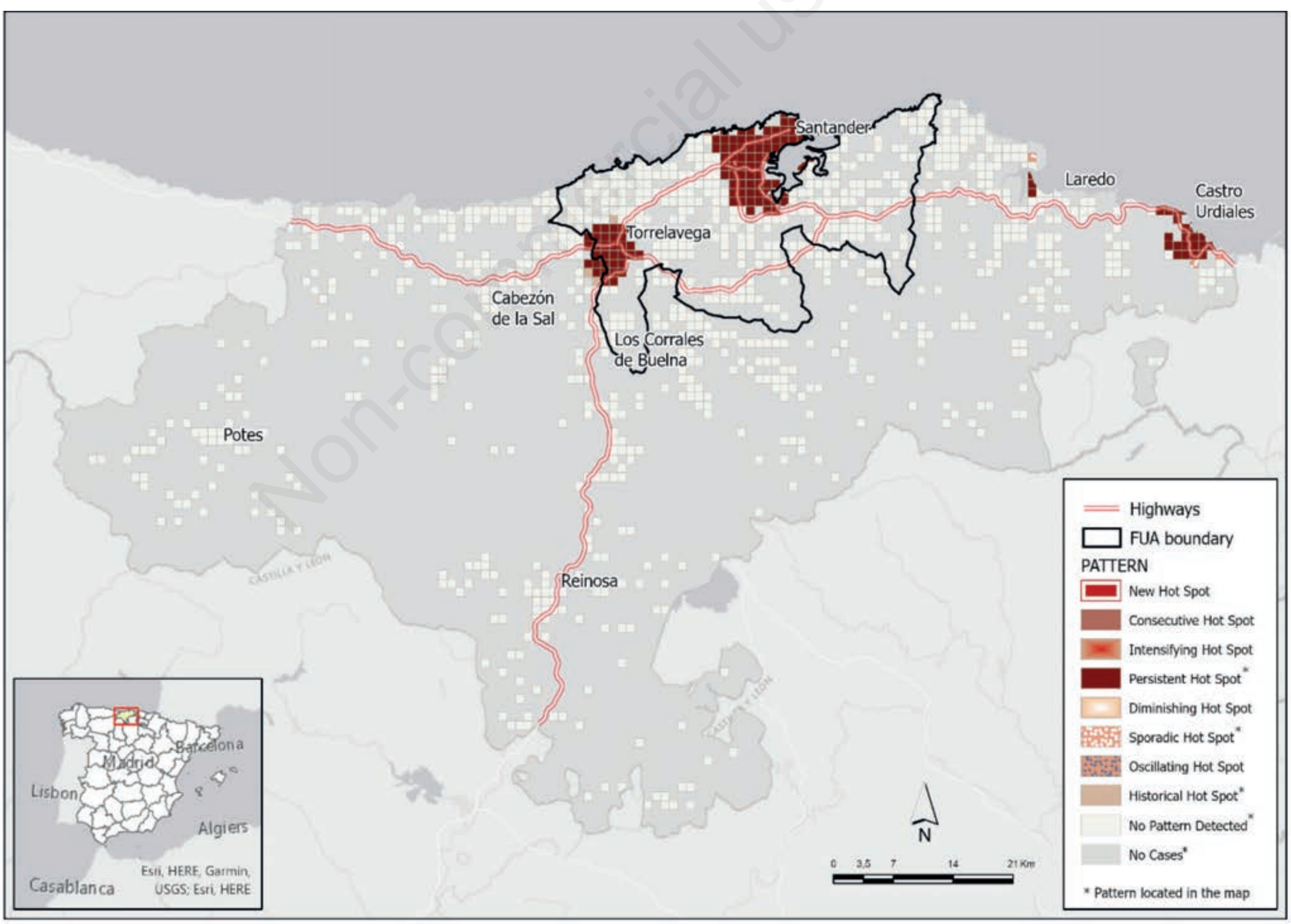

Figure 5. Space-time COVID-19 emerging hotspots in 2021. ***Sources: ESRI (Administrative Base map), National Geographic Institute (National Cartographic Base 200), Copernicus FUA layer and COVID-19 microdata daily records from the health authorities (Government of Cantabria, Spain).***

Different local relationships with COVID-19 appeared in other correlated variables, such as the percentage of aged people (mainly negative linear, except in the Santander-Torrelavega FUA and the oriental coastal area) and the income per capita that resulted in a fuzzy pattern where an undefined complex type was predominant and only some small cluster of pixels appeared in the oriental and occidental coastal areas. As shown in Figure 6B, the total spending by household variable showed an opposite relationship with COVID-19 cases depending on the dynamism of the area. Peri-urban areas with prominent commuting and rural service centres, *i.e.* places where people and economic activities were concentrated, presented positive linear relationships. Total spending by household can be understood as an indirect indicator of lifestyle that increases the risk of COVID-19 contagion. In contrast, low density areas and inland valleys both showed a negative linear correlation with the number of COVID-19 cases.

Discussion

Our method highlights the use of cluster analysis based on space-time and 3D bins and emerging hotspots pioneered by Hägerstand (1970) and Gatalsky *et al*. (2004) and developed into a commercial product by ESRI. The resultant cartography is revealing as it includes spatial and temporal trends and can detect significant spatial patterns. To our best knowledge, this method was applied to COVID-19 analysis by Chunbao *et al*. (2020) using data at the city level, then by De Cos *et al*. (2021b) using geocoded microdata, by Tokey (2021) adapted to the county level and, recently, by Syetiawan *et al*. (2022) at the sub-district level. Our models answer 'where', a key question according to Kamel and Geraghty (2020), while we also focus on 'when', while we aim deeper to solve the question in relation to control. 'How' and 'why' in future studies.

Geospatial sciences play a strategic role in shortening response times for pandemic management (Zhou *et al*., 2020) and planning measures according to spatial models of key topics, such as incidence ratios, hotspot setting and even accessibility to healthcare centres (Mohammadi *et al*., 2021). Along with this thinking, we previously developed a geoprevention approach based on models to control the virus spread (De Cos *et al*., 2021b). The results aimed to facilitate the alignment of restrictions and measures with specific patterns of each area in the region. Nevertheless, here we have to consider difficulties linked to Spanish territorial and administrative organisation that can result in mismatched hotspots, administrative health units and management units (Andrés *et al*., 2021). In doing so, we observed significant differences within municipalities that encouraged acting at local scales (Salama, 2020), in this case under the municipal boundary, a focus suitable for our methodological approach. By applying a high spatial disaggregation, *i.e.* coordinate pairs and temporal granularity (daily), our multi-scale approach (Dhaval, 2020) not only enabled a detailed space-time analysis, but also identification of new territorial boundaries (the epidemiological problem areas) without constraints from administrative or management units.

Importantly, the context conditions (the socio-economic variables, building characteristics, economic activity, *etc*.) remained largely unchanged during the study period, while other variables necessarily changed in real time introducing some fuzziness in the GIS models. This can, for example, occur with pollution levels that can change by the hour in urban areas, which necessitates a fine scale chronogeography framework (Parkes and Nigel, 1980) or it can lead to spatial concurrence (Buffalo and Rydzewski, 2021).

The changing spatial pattern of COVID-19 is related to the regulatory framework (varying from almost normality to strict confinement), which can produce different results in the spread of the virus depending on timely decisions (Seong *et al*., 2021). Many researches confirm the important role of containment measures to mitigate viral spread (Alcântara *et al*., 2020; Li *et al*., 2020). We agree that it plays an important role and admit that it is a limitation that changes in regulatory context were not included in our model, although we did consider it as background in the interpretation of results. Combining changing factors (both mobility and regulatory framework), Buffalo and Rydzewski (2021) state that models of pedestrian mobility and spatial concurrency models are less reliable in non-mobility restriction periods, while the spread of the virus is linked to proximate areas in restrictive frameworks as demonstrated in a case study carried out in Malaga, Spain using a dynamic model based on the confluence-transmission concept (Perles *et al*., 2021b).

Our study showed a contraction and simplification of the COVID-19 problem areas during the study period. Although hotspots affected urban, peri-urban areas and rural service centres in 2020, the spread of the virus was more controlled, and reduced to the main cities as in 2021. This is in accordance with Buffalo and Rydzewski (2021) who concluded that when there is community transmission, the spatial spread of the virus does not only affect the main cities but also those of medium size (as seen in Figure 4). On the other hand, the problem areas in 2021 were more common in densely populated areas as they concentrate the risk factors, a fact also noted by Fernández *et al*. (2021) in their case study in Asturias, a region bordering Cantabria. Thus, population density and COVID-19 spatial patterns are linked (Bamweyana *et al*., 2020), which is also evident from the 'effective local density' (Desmet and Wacziarg, 2021), the inhabitants to area ratio as shown on the 1x1-km pixel vector grid.

Some recent studies dissociate socio-economic vulnerability and COVID-19 cases as only being punctual associations (Buffalo and Rydzewski, 2021), a result that we also obtained when applying local bivariate relationship of income per capita that resulted in a predominant fuzzy pattern of undefined complex type. Another dissociation is related to ageing. We obtained a mainly negative linear relation except in the Santander-Torrelavega FUA and the oriental coastal area. Therefore, ageing may be related with severity but not with spread. Similarly, a recent study of USA counties found that areas highly affected by COVID-19 mostly had a low proportion of aged people (Tokey, 2021). As also stated by Almendra *el al.* (2021), our results suggest that areas primarily inhabited by a dynamic mix of middle-age and young with high spending habits per household and an active lifestyle decisively contribute to the genesis of significant COVID-19 problem areas.

Local bivariate relationships demonstrated that the virus presents a non-stationary spatial behaviour. Type and intensity of relation between explanatory and dependent variables changed across the study area. Nevertheless, deep research along this line of investigation is necessary in the future. Exploratory study of permanent variables is of interest and should be analysed monographically considering also other factors that might affect the COVID-19 spatial patterns. Indeed, a multi-factorial approach contributing to knowledge of the spread of the virus in urban and metropolitan areas (Perles *et al*., 2021b) could be revealing. We say this since chrono geographic analysis at detailed scales and applied to com-

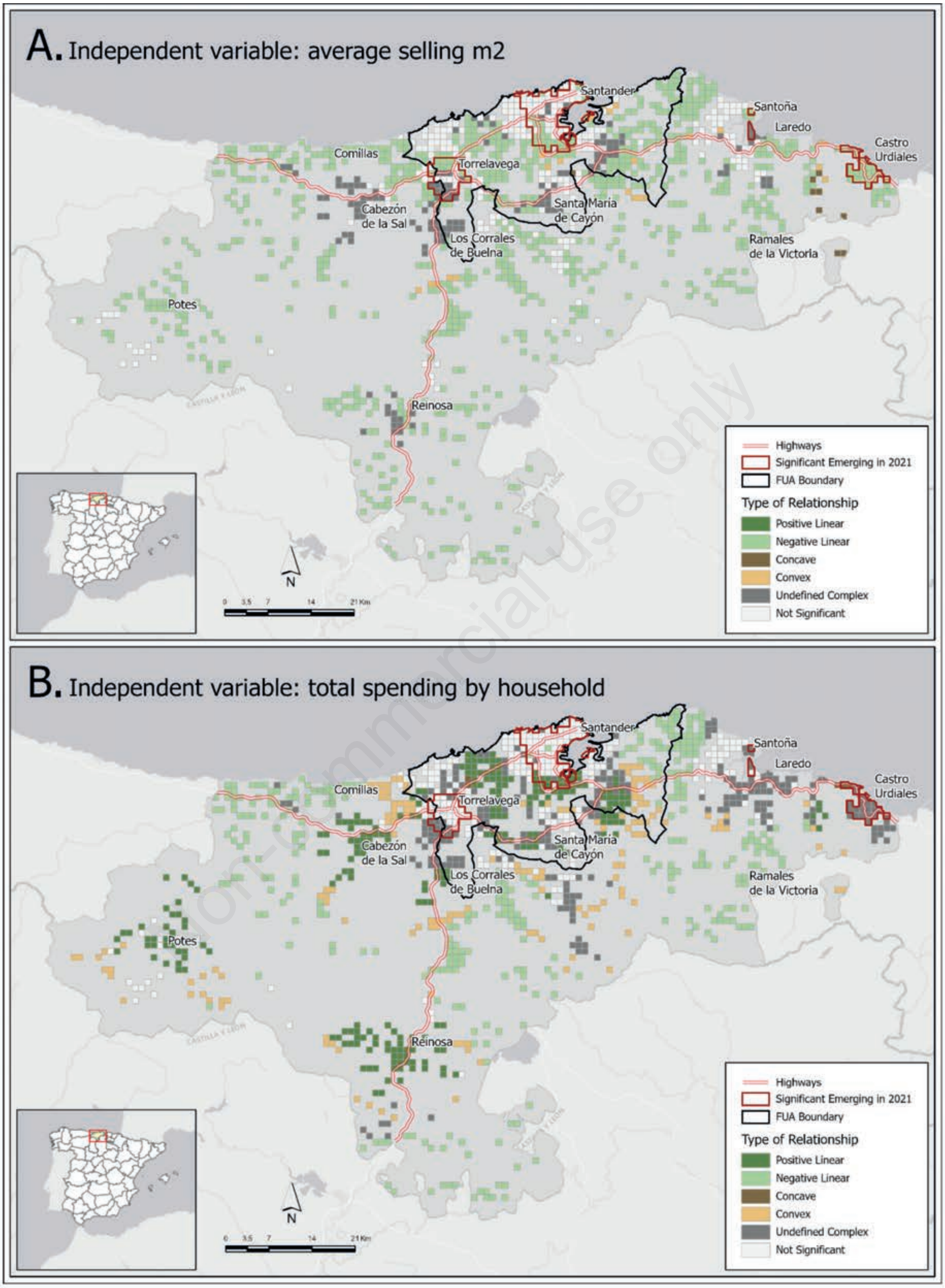

Figure 6. A-B) Examples of local bivariate relationships with the COVID-19 cases as the dependent variable: average selling price per m^2 and total spending by household. ***Sources: ESRI (Administrative Base map), National Geographic Institute (National Cartographic Base 200), Copernicus FUA layer and COVID-19 microdata daily records from the health authorities (Government of Cantabria, Spain).***

plex, dense and dynamic systems, where many factors can contribute to viral spread, are of interest in terms of spatial concurrence (Buffalo and Rydzewski, 2021).

Many rules and measures to reduce COVID-19 transmission were implemented by the national and regional governments in Spain during the two years of the pandemic. The vaccination process that started in January 2021 is one of the most important ones for the work presented here, as we finished this research in 2021, at which time 83.4% of the population of Cantabria had been vaccinated according to health regional data for the Government of Cantabria. This is a very high vaccination ratio considering that many children have not been vaccinated yet, even though vaccination of 5 to 11-year olds started in mid-December 2021.

At the beginning of the pandemic Jindal *et al*. (2020) stated that the two stages for the risk mitigation would be to reduce the spread and decrease disease severity. However, with an effective vaccine available, we hold the idea that controlling spatial patterns (virus spread) can reduce severity indirectly because the probability of infection of non-vaccinated and vaccinated vulnerable people decreases in a scenario of low contagion. This consideration is of special importance at the time this paper was conclude (Christmas, 2021), Cantabria registered the worrying record of near 1000 new cases per day in the context of the sixth wave with the Omicron variant.

Invisible boundaries complicate spatial pattern analysis and virus containment. Examples include new problem areas, administrative boundaries used by policy makers to implement containment and mitigation measures, statistic areas to access to data sources about socio-economic context, *etc*. This is a wide research field with open borders as demonstrated by Mollalo *et al*. (2021) in their recently published scoping review about spatial analysis of COVID-19 vaccination. The study of vaccination from a spatial approach is increasing and more sophisticated techniques than choropleth maps are necessary. When data registers are available from health authorities, our methodological proposal may be applied to vaccination microdata, which could show hotspots as areas where increased vaccination is needed, while coldspots and areas without visible patterns of spread might indicate areas requiring a deeper analysis.

Moore *et al*. (2021) state that vaccination against COVID-19 seems insufficient with regard to containing outbreaks and spread, at least in areas with low-medium proportion of vaccinated people. However, with data from the sixth wave, we reinforce that idea and conclude that spatial models are revealing, innovative and can assist decision-making, *i.e*. it is time to put social sciences, GIS and health geographics at the service of health authorities.

Conclusions

Our research demonstrates the nuances of COVID-19 areas and contributes to an understanding of the pandemic from a spatial point of view. Admittedly, an important varying factor during the 20-month study period were the different proportion of vaccinated people in the second year and the changing decision-making in forecasting, organization and management of health services as the COVID-19 epidemic evolved.

Focusing on the significant hotspots, the effective population density, as indicated by commuting and concentration of economic activities in the urban, peri-urban areas and rural service centres under study can be understood as potential concurrence areas. This reinforced the interest in where, when and how people interact in the same neighbourhood. Hence, GIS-science tools ecosystem and health geographic concepts and methods can contribute to tackle the pandemic from an interdisciplinary approach joining spatial research to other specific and consolidated disciplines in the pandemic, *e.g*., epidemiology, medicine and virology in particular.

Our models revealed problem areas and distinguished several patterns. On this basis, we provide support for decision makers in matters of geoprevention as a way to design measures at regional scale with local targeted strategies. In this sense, virus contention plans can be strengthened with spatial perspective.

References

Ahasan R, Hossain MM, 2021. Leveraging GIS and spatial analysis for informed decision-making in COVID-19 pandemic. Health Policy Technol 10:7-9.

Alcântara E, Mantovani J, Rotta L, Park E, Rodrigues T, Campos F, Souza CR, 2020. Investigating spatiotemporal patterns of the COVID-19 in São Paulo State, Brazil. Geospat Health 15:925.

Al Kindi KM, Al-Mawali A, Akharusi A, Alshukaili D, Alnasiri N, Al-Awadhi T, Charabi Y, El Kenawy AM, 2021. Demographic and socioeconomic determinants of COVID-19 across Oman - A geospatial modelling approach. Geospat Health 16:985.

Almendra R, Santana P, Costa C, 2021. Spatial inequalities of COVID-19 incidence and associated socioeconomic risk factors in Portugal. BAGE 91.

Andrés G, Herrero D, Martínez M, 2021. Cartographies on COVID-19 and functional divisions of the territory: an analysis on the evolution of the pandemic based on Basic Health Areas (BHA) in Castile and Leon (Spain). BAGE 91.

Bamweyana I, Okello DA, Ssengendo R, 2020. Socio-economic vulnerability to COVID-19: the spatial case of Greater Kampala Metropolitan Area (GKMA). J GIS 12:302-18.

Batista F, Poelman H, 2016. Mapping population density in functional urban areas. A method to downscale population statistics to urban atlas polygons. JRC Technical Reports. European Commission. Available from: https://tinyurl.com/y9r2mbkt

Bergquist R, Kiani B, Manda S, 2020. First year with COVID-19: Assessment and prospects. Geospat Health 15:953.

BOE-A-2015-682 (2015). Ley 4/2014, de 22 de diciembre, del Paisaje. Texto consolidado. Gobierno de Cantabria. Available from: https://www.boe.es/buscar/pdf/2015/BOE-A-2015-682-consolidado.pdf

Buffalo L, Rydzewski AL, 2021. Territorial dynamics of the COVID-19 pandemic in the province of Córdoba, Argentina. BAGE 91.

Campagna M, 2020. Geographic Information and COVID-19 outbreak. Does the spatial dimension matter? J Land Use Mobil Environ 31-44.

Chunbao M, Dechan T, Tingyu M, Chunhua B, Jian Q, Weiyi P, Zhiyong Z, 2020. An analysis of spatiotemporal pattern for COVID-19 in China based on space-time cube. J Med Virol92:1587-1595.

Coccia M, 2021. Pandemic Prevention: Lessons from COVID-19. Encyclopedia 1:433-44.

Cromley EK, 2019. Using GIS to address epidemiologic research questions. Curr Epidemiol Rep 6:162-73.

Das A, Ghosh S, Das K, Basu T, Dutta I, Das M, 2021. Living environment matters: Unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. Sustain Cities Soc 65:102577.

De Cos O, Castillo V, Cantarero D, 2020. Facing a second wave from a regional view: spatial patterns of COVID-19 as a key determinant for public health and geoprevention plans. Int J Environ Res Public Health 17:8468.

De Cos O, Castillo V, Cantarero D, 2021a. Data mining and socio-spatial patterns of COVID-19: geo-prevention keys for tackling the pandemic. BAGE 91.

De Cos O, Castillo V, Cantarero D, 2021b. Differencing the risk of reiterative spatial incidence of COVID-19 using space-time 3D bins of geocoded daily cases. ISPRS Int J Geo-Inf 10:261.

Desmet K, Wacziarg R, 2021. Understanding spatial variation in COVID-19 across the United States. J Urban Econ [In press].

Dhaval DD, 2020. Urban densities and the COVID-19 pandemic: upending the sustainability myth of global megacities. ORF Occasional Paper 244:1-42. Available from: https://tinyurl.com/5n8hwd2j

Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O, 2021. Geospatial analysis of COVID-19: A scoping review. Int J Environ Res Public Health 18:2336.

Fernández F, Herrera D, Fernández C, 2021. Temporal and territorial dimension of the COVID-19 pandemic in Asturias, Spain. BAGE 91.

Ferreira MC, 2020. Spatial association between the incidence rate of COVID-19 poverty in the São Paulo municipality, Brazil. Geospat Health 15:921.

Franch-Pardo I, Desjardins M, Barea-Navarro I, Cerdà A, 2021. A review of GIS methodologies to analyze the dynamics of COVD-19 in the second half of 2020. Transact in GIS 00:1-49.

Gatalsky, P, Andrienko N, Andrienko K, 2004. Interactive analysis of event data using space-time cube. In: E. Banissi (Ed.), Proceedings of the 8th International Conference on Information Visualization, July 2004, London, UK. IEEE Computer Society, Los Alamitos, pp. 145-52.

Gerber TD, Ping D, Armstrong-Brown J, McNutt LA, Cole FB, 2009. Charting a path to location intelligence for STD control. Public Health Rep 124:49-57.

Hägerstrand T, 1970. What about people in regional science? Reg Sci Assoc Pap 24:7-21.

Hamidi S, Sabouri S, Ewing R, 2020. Does Density Aggravate the COVID-19 Pandemic? Early Findings and Lessons for Planners. J Am Plann Assoc 86:495-509.

Huang J, Kwan MP, Kan Z, Wong MS, Tung Kwok CY, Yu X, 2020. Investigating the relationship between the built environment and relative risk of COVID-19 in Hong Kong. ISPRS Int J Geo-Inf 9:624.

Jardim de Figueiredo CJ, de Miranda CM, Ferreira AG, Pimentel A, da Silva SM, 2022. Vulnerability to COVID-19 in Pernambuco, Brazil: a geospatial evaluation supported by multiple-criteria decision aid methodology. Geospat Health 17:1000.

Jindal C, Kumar S, Sharma S, Choi YM, Efird JT, 2020. The prevention and management of COVID-19: seeking a practical and timely solution. Int J Environ Res Public Health17:3986.

Kamel MN, Geraghty EM, 2020. 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 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 19:8.

Kulldorff M, 2001. Prospective time periodic geographical disease surveillance using scan statistic. J R Statist Soc A Ser A Stat Soc 164:61-72.

Li H, Li H, Ding Z, Hu Z, Chen F, Wang K, Peng Z, Shen H, 2020. Spatial statistical analysis of coronavirus disease 2019 (COVID-19) in China. Geospat Health 15:867.

Mohammad Ebrahimi S, Mohammadi A, Bergquist R, Dolatkhah F, Olia M, Tavakolian A, Pishgar E, Kiani B, 2021. Epidemiological characteristics and initial spatiotemporal visualisation of COVID-19 in a major city in the Middle East. BMC Public Health 21:1373.

Mohammadi A, Mollalo A, Bergquist R, Kiani B, 2021. Measuring COVID-19 vaccination coverage: an enhanced age-adjusted two-step floating catchment area model. Infect Dis Poverty 10:118.

Mollalo A, Mohammadi A, Mavaddati S, Kiani B, 2021, Spatial analysis of COVID-19 vaccination: a scoping review. Int J Environ Res Public Health 18:12024.

Moore S, Hill EM, Tildesley MJ, Dyson L, Keeling MJ, 2021. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. Lancet Infect Dis 21:793-802.

Mou Y, He Q, Zhou B, 2017. Detecting the spatially non-stationary relationships between housing price and its determinants in China: Guide for housing market sustainability. Sustainability 9:1826.

Niu X, Yue Y, Zhou X, Zhang X, 2020. How urban factors affect to spatiotemporal distribution of infectious diseases in addition to intercity population movement in China. ISPRS Int J Geo-Inf 9:615.

Parkes DN, Nigel JT, 1980. Times, spaces, and places; a chronogeographic perspective. John Wiley and Sons, New York, NY, USA, 527pp.

Perles MJ, Sortino JF, Mérida MF, 2021a. The neighborhood contagion focus as spatial unit for diagnosis and epidemiological action against COVID-19 contagion in urban spaces: a methodological proposal for its detection and delimitation. Int J Environ Res Public Health 18:3145.

Perles MJ, Sortino JF, Cantarero FJ, Castro H, De la Fuente AL, Orellana-Macías JM, Reyes S, Miranda J, Mérida M, 2021b. Potential of hazard mapping as a tool for facing COVID-19 transmission: the geo-COVID cartographic platform. BAGE 91.

Salama AM, 2020. Coronavirus questions that will not go away: interrogating urban and socio-spatial implications of COVID-19 measures. Emerald Open Res 2:14.

Seong H, Hyun HJ, Yun JG, Noh JY, Cheong HJ, Kim WJ, Song JW, 2021. Comparison of the second and third waves of the COVID-19 pandemic in South Korea: Importance of early public health intervention. Int J Infec Dis 104:742-5.

Sera F, Armstrong B, Abbott S, Meakin S, O'Reilly K, Von Borries R, Schneider R, Royé D, 2021. A cross-sectional analysis of meteorological factors and SARS-CoV-2 transmission in 409 cities across 26 countries. Nature Communic 12:5968.

Syetiawan A, Harimurti M, Prihanto Y, 2022. A spatiotemporal analysis of COVID-19 transmission in Jakarta, Indonesia for a pandemic decision support. Geospat Health 14:1042.

Tokey AI, 2021. Spatial association of mobility and COVID-19 infection rate in the USA: A county-level study using mobile phone location data. J Transp Health 22:101135.

Whittle RS, Díaz-Artiles A, 2020. An ecological study of socioe-

conomic predictors of detection of COVID-19 cases across neighborhoods in New York City. BMC Med 18:271.
Ye L, Hu L, 2020. Spatiotemporal distribution and trend of COVID-19 in the Yangtze River Delta region of the People's Republic of China. Geospat Health 15:889.
Yin Z, Huang W, Ying S, Tang P, Kang Z, Huang K, 2021. Measuring of the COVID-19 Based on Time-Geography. Int J Environ Res Public Health 18:10313.
Zhou C, Su F, Pei T, Zhang A, Yuyan D, Luo B, Zhidong C, Wang J, Yuan W, Zhu Y, Song C, Chen J, Xu J, Li F, Ma T, Jiang L, Yan F, Yi J, Hu Y, Liao Y, Xiao H,2020. COVID-19: Challenges to GIS with Big Data. Geogr Sustain 1:77-87.