ARTICLE IN PRESS

Annals of Tourism Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

Research Note

A new European regional tourism typology based on hotel location patterns and geographical criteria

Filipe Batista e Silva *, Ricardo Barranco, Paola Proietti, Cristian Pigaiani, Carlo Lavalle

European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra, Italy

ARTICLE INFO

Associate editor: Haiyan Song

Keywords:
Regional tourism typology
Big data
Coastal zones
Cities
Mountains
Rural

ABSTRACT

The development of tourism typologies allows the identification of various dimensions of this important sector and can be used in tourism planning and management. In this study we developed a novel and systematic classification of European Union (EU) regions according to the predominant location of hotels, distinguishing among four distinct geographical areas: coastal zones, mountain and natural areas, cities and rural areas. For this purpose, we use Geographical Information Systems to integrate regional boundaries and the geographical zoning with a detailed dataset of hotel location and capacity. We then classified the EU regions into five typologies with the assistance of a hierarchical clustering algorithm. Further exploration of the results shows very distinct tourism profiles across the identified typologies in terms of tourism intensity, seasonality and other traits.

Introduction

The use of 'big data' to study tourism and support tourism management is an emerging trend thanks to the increasing availability of data, computational power and analytical methods (Li, Xu, Tang, Wang, & Li, 2018). However, the combination of big data and official statistics to produce insights with a European-wide perspective and detailed spatial and temporal dimensions remains mostly an unexplored opportunity (Eurostat, 2017). The work from Batista e Silva et al. (2018) is a contribution in this respect. It produced a geospatial dataset describing the location and capacity of tourism accommodation at high spatial resolution, and combined it with official statistics to reveal the uneven spatial and temporal distribution of tourism in Europe.

A number of tourism typologies have been proposed to segment the tourism market, allowing the recognition of important dimensions of this activity and with applications in tourism planning. These typologies can be grouped in two broad categories: the ones centred on the demand side (e.g. the visitors and their characteristics) and those centred on the supply side (e.g. the destinations and their characteristics) (Coccossis & Constantoglou, 2008). Candela and Figini (2012) argue that tourism destinations are at the core of the tourism system and highlight the importance of characterizing them for decision making. However, to the best of our knowledge, there is no systematic classification of regions according to the type of tourism destination.

To help fill this gap, here we describe a first attempt to classify European Union (EU27) regions. Our approach is based on the location of hotels within regions, and distinguishing among four basic types of geographical areas: coastal zones, mountain and natural areas, cities and rural areas. This approach builds off the assumption that hotel location reflects where, within a region, tourism is

* Corresponding author.
E-mail address: filipe.batista@ec.europa.eu (F. Batista e Silva).

https://doi.org/10.1016/j.annals.2020.103077
Received 16 June 2020; Received in revised form 27 August 2020; Accepted 4 September 2020

actually taking place among the various geographical areas, thus being a useful proxy for describing the type of tourism destination. This is coherent with Hernández-Martín et al. (2016), who support the use of the location of tourism accommodation in spatial segmentation, as well as with the findings of spatial relationship between hotels and amenities and their regional clustering (Lee, Kang, Terry, & Schuett, 2018).

Data and methods

The spatial units chosen for the creation of the new classification were the NUTS3 (version 2016), comprising 1165 regions within the EU27, with a median size of 1875 km^2, typically corresponding to country provinces or districts. The NUTS is an official system of nested territorial units comprising 4 hierarchical levels used for statistical data reporting in Europe.

The method to classify NUTS3 regions involved three steps, employing available geospatial data, Geographical Information Systems and machine learning techniques:

1. Spatial delineation of key geographical areas within each NUTS3;
2. Integration of hotel location and capacity data in the resulting geographical zoning;
3. Classification of the NUTS3 regions using a clustering algorithm.

In the first step we considered four types of geographical areas: coastal zones, mountain and natural areas, cities and rural areas. We delineated 'coastal zones' by applying a 10 km-straight line buffer to the coastline (EuroBoundaryMap[1]). 'Mountain and natural areas' include the areas above 800 m of altitude (EU-DEM v1.1[2]) and a 2 km-straight line buffer from the protected areas listed in the Natura 2000.[3] For 'cities', we considered 683 'urban centres' across the EU27 delineated by Eurostat based on criteria related to urban morphology to consistently obtain city limits irrespective of national definitions (Eurostat, 2018). Finally, all areas that do not belong to any of the latter were deemed 'rural'. In case of overlapping areas, we assumed the following order of priority: 1) cities, 2) coastal zones, 3) mountain and natural areas and 4) rural areas. This order was defined based on our judgment of the touristic prevalence of the different geographical areas. For example, if an area is both a city and a coastal zone (e.g. Barcelona, Copenhagen), then we assume the city is the main driver of visitors. Similarly, if an area is both part of a coastal area and a mountain (not common, but may occur in, for example, Crete, Liguria and Sardinia), then we assume the coastal traits have higher prevalence in driving visitors to the area (see adopted zoning in Supplementary Fig. 2).

In the second step, we spatially intersected the 4-class geographical zoning with the NUTS3 boundaries in order to 'split' each NUTS3 region per any of the occurring zones. The resulting intersected GIS layer was then overlaid with the 'hotel layer' using a 'zonal statistics' function to obtain the tourism capacity within each geographical zone within each NUTS3 (Supplementary Fig. 3a–d). The 'hotel layer' is a raster dataset at 100 m × 100 m resolution recording the number of rooms in tourism accommodation based on information available from a combination of online booking platforms (see Batista e Silva et al. (2018) for details).

Based on the referred overlays, we produced a matrix describing each NUTS3 according to the share of the accommodation capacity (no. of rooms) per geographical zone. Using the shares of accommodation capacity in each geographical zone as input variables, we employed a hierarchical clustering to classify each NUTS3 region into a specific typology reflecting the geographical context of the accommodation capacity (see Supplementary Note 1 for a description of the hierarchical clustering procedure).

For the subsequent characterization of the clusters we used tourism statistics from Eurostat for the latest available reporting year, 2018. Data available only at NUTS2 level were downscaled to NUTS3 level proportionally to the accommodation capacity (i.e. nights-spent), or assumed to have equal distribution in all NUTS3 within a NUTS2 (i.e. share of foreign tourists). The seasonality was derived from a combination of data from National Statistical Institutes (monthly, 2011, per NUTS2 or NUTS3) and the number of reviews in TripAdvisor (quarterly, 2017, per tourism accommodation establishment).

Results

We obtained five clusters of NUTS3 regions. By plotting the median share of accommodation capacity in each geographical area for each cluster separately, it is possible to characterize the clusters' profiles (Fig. 1a–d). We labeled the clusters or typologies as follows:

- *Cities*: regions with tourism capacity located predominantly in cities;
- *Coastal*: regions with tourism capacity located predominantly in coastal zones;
- *Mountains and Nature*: regions with tourism capacity located predominantly in mountainous or protected areas;
- *Rural*: regions with tourism capacity located predominantly in rural areas;
- *Urban mix*: regions with tourism capacity located in cities in combination with rural and/or mountainous and protected areas.

The geographical distribution of the classification is represented in Fig. 2, whereas Fig. 3 shows several key tourism indicators across the identified typologies. Coastal regions stand-out as the ones accounting for the majority of nights-spent in the EU (42%),

[1] Eurogeographics, EuroBoundaryMap, https://eurogeographics.org/maps-for-europe/ebm/.

[2] Copernicus EU-DEM, https://land.copernicus.eu/imagery-in-situ/eu-dem.

[3] Natura 2000 (reporting year 2019), https://www.eea.europa.eu/data-and-maps/data/natura-11.

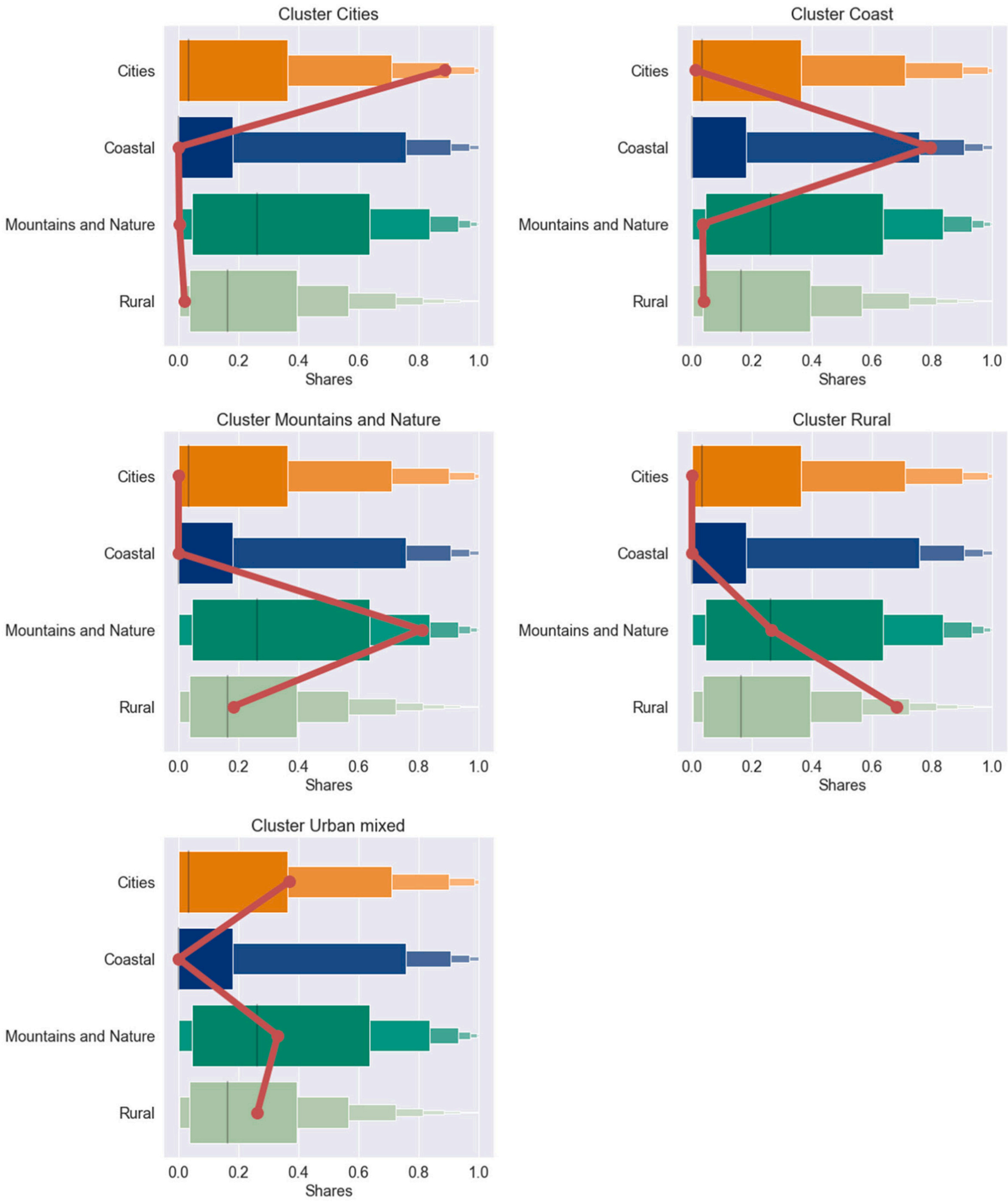

Fig. 1. Cluster profiles. The red line in each plot represents the median value of the share of rooms in each geographical area for the regions of the respective cluster. The 'boxen plots' in the background represent the statistical distribution of each variable for the whole set of NUTS3 regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

followed far behind by cities (19%) and mountains and nature (16%). Coastal regions are also those with the highest tourism intensity, scoring 12.3 nights-spent per inhabitant – substantially higher than mountains and nature, and cities with 7.3 and 5.3 nights-spent per inhabitant, respectively. As for the share of foreign tourists, coastal regions score the highest average value, around 45%. On the other hand, the urban mix and the mountains and nature regions are, on average, characterized by much higher shares of domestic tourists.

The coastal typology is the most seasonal of all typologies, showing a very marked peak of affluence in the summer months. All the remainder regions show a more flattened seasonal profile, although the summer months remain the most popular for tourism across all

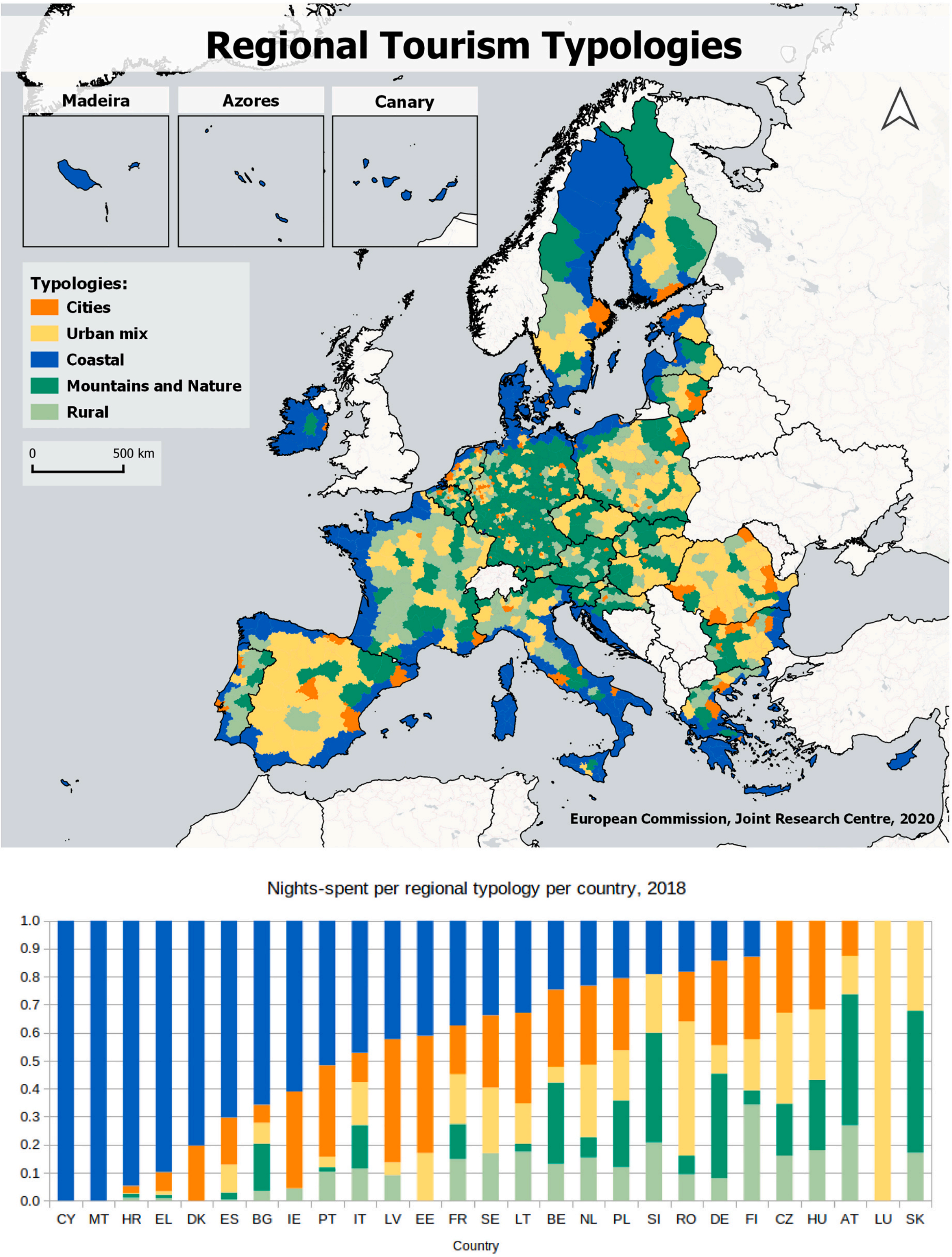

Fig. 2. Geographical distribution of the regional typologies (top) and proportion of each typology in terms of total nights-spent per country in 2018 (bottom).

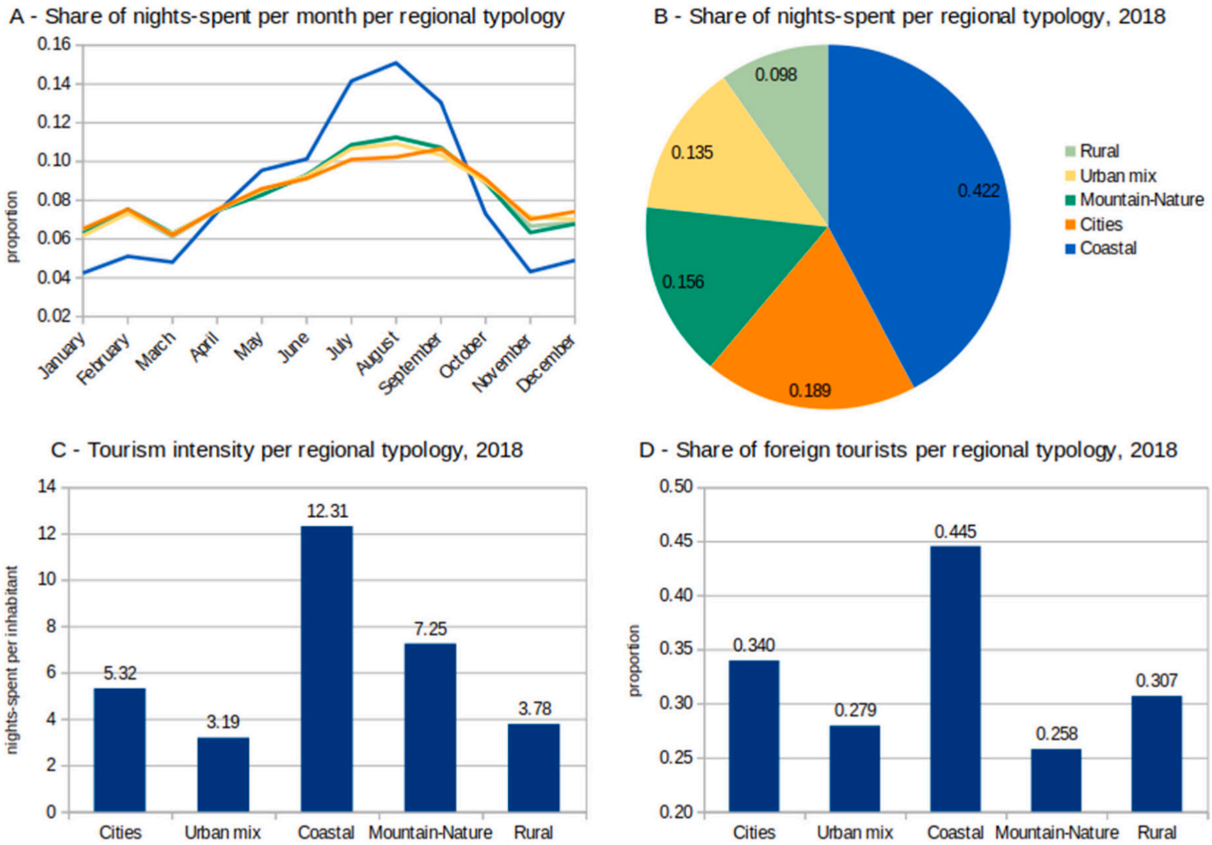

Fig. 3. Key tourism indicators per regional typology.

typologies (Fig. 3a). Cities appear with the lowest seasonality, with slightly higher lows in the winter and lower highs in the summer. Table 1 summarizes these findings.

Discussion and conclusions

In this paper we classified EU NUTS3 regions in 5 typologies. To determine the predominant type of tourism destination of each NUTS3, we looked at the location of tourism accommodation within each NUTS3 in relation to four distinct geographical areas: coastal zones, mountain and natural areas, cities, and rural areas. Some NUTS3 can be very sizeable and encompass a variety of different geographical features. Our approach focuses on where tourism supply is actually located within each region, and is therefore compelling compared to approaches that ignore the location of hotels. A similar outcome could be attained by inquiring tourists on their reasons for travelling. However, a survey covering the whole EU, and that is statistically representative for each NUTS3, would be impractical.

Evidently, the geographical criteria and decision rules applied can be object of debate. However, the characterization of the clusters indicate sensible and plausible differences, corroborating the approach overall. Possibly, a higher breakdown to include other geographical features (e.g. presence of snow, or type of coast) would be desirable, but require further reflection in view of feasibility and usability.

This classification provides new insights on the geography of tourism in Europe and could be used as an additional input to study

Table 1
Qualitative characterization of the regional typologies according to four tourism indicators. Scores ('+++' high, '++' moderate, '+' low) were assigned to each regional typology based on their indicator values as per Fig. 3 (more details in Supplementary Note 2).

Regional typology	Nights-spent	Tourism intensity	Share of foreign tourist	Seasonality
Cities	++	++	++	+
Coastal	+++	+++	+++	+++
Mountain and nature	++	++	+	++
Rural	+	+	++	++
Urban mix	++	+	+	++

territorial patterns of tourism (Majewska, 2015), its spatiotemporal dynamics and relation with natural, cultural or other assets (Romão, Guerreiro, & Rodrigues, 2017). Since tourism demand and destination characteristics are intertwined, our typology could also be used in tourism demand forecasting, a key tool for decision-makers (Song, Qiu, & Park, 2019). Finally, we believe that the profiling of regions is especially relevant in the context of the response to COVID-19 pandemic and its impacts on tourism (Gössling, Scott, & Hall, 2020).

Acknowledgements

This study was performed in the frame of the Knowledge Centre for Territorial Policies of the European Commission, with the aim to improve the knowledge base necessary to support European strategies of recovery from the COVID-19 pandemic.

CRediT authorship contribution statement

FBS, CL and PP designed the research, CP, FBS and RB performed the data collection and geoprocessing, RB carried the cluster analysis, FBS analysed the results, FBS and PP wrote the paper with inputs from all authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.annals.2020.103077.

References

Batista e Silva, F., Marín Herrera, M. A., Rosina, K., Ribeiro Barranco, R., Freire, S., & Schiavina, M. (2018). Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management, 68*, 101–115.

Candela, G., & Figini, P. (2012). The economics of tourism destinations. In G. Candela, & P. Figini (Eds.), *The economics of tourism destinations* (pp. 73–130).

Coccossis, H., & Constantoglou, M. E. (2008). The use of typologies in tourism planning: Problems and conflicts. In H. Coccossis, & Y. Psycharis (Eds.), *Regional analysis and policy: The Greek experience* (pp. 273–295).

Eurostat. (2017). Tourism statistics: Early adopters of big data?. In *Sixth UNWTO international conference on tourism statistics measuring sustainable tourism* (2017th ed.).

Eurostat. (2018). *Methodological manual on territorial typologies. 2018 edition*.

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: A rapid assessment of COVID-19. *Journal of Sustainable Tourism*, 1–20.

Hernández-Martín, R., Simancas-Cruz, M. R., González-Yanes, J. A., Rodríguez-Rodríguez, Y., García-Cruz, J. I., & González-Mora, Y. M. (2016). Identifying micro-destinations and providing statistical information: A pilot study in the Canary Islands. *Current Issues in Tourism, 19*(8), 771–790.

Lee, K. H., Kang, S., Terry, W. C., & Schuett, M. A. (2018). A spatial relationship between the distribution patterns of hotels and amenities in the United States. *Cogent Social Sciences, 4*(1).

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323.

Majewska, J. (2015). Inter-regional agglomeration effects in tourism in Poland. *Tourism Geographies, 17*(3), 408–436.

Romão, J., Guerreiro, J., & Rodrigues, P. M. M. (2017). Territory and sustainable tourism development: A space-time analysis on European regions. *Region, 4*(3), 1.

Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research, 75*, 338–362.

Filipe Batista e Silva is a geographer specialized in spatial planning and Geographical Information Systems by the University of Porto, Portugal. He is currently a scientific and policy officer at the European Commission's Joint Research Centre in Ispra, Italy, involved in territorial impact assessment of EU trends and policies using geospatial data and analysis. Filipe is author of various scientific publications dealing with population, land use, tourism and environmental mapping and modelling applications.

Ricardo Barranco is an Environmental engineer from Portugal with 12 years of working experience. He originally worked four years as a GIS analyst in the gas and electricity sector in Genoa, Italy. In the subsequent years he has been a scientific project officer at the European Commission Joint Research Centre. His research topics include urban and territorial planning, data science and geospatial data collection from unconventional sources, visualization and analysis.

Paola Proietti holds a Bachelor and MSc in Economic and Social Sciences from Bocconi University, Italy. Later she specialized in Urban and Regional Science receiving her PhD in 2019. Her research cuts across Human and Economic Geography, Political Economy, Public Policy and it is mainly quantitative. In 2019 she joined the European Commission where she contributes to the Knowledge Centre for Territorial Policies focusing on spatial inequality, poverty, housing, migration, tourism and SDGs.

Cristian Pigaiani graduated in 2003 in Environmental and Land engineering. From 2004 he worked as a GIS analyst and technician for electromagnetic field analysis in the private sector. In 2019 he started working as a GIS consultant at the European Commission Joint Research Centre, supporting geospatial data processing and analysis. His research interests include spatial analysis, terrain modelling and analysis, spatial data structures, big data analysis and management and geospatial data visualization.

Dr. **Carlo Lavalle** has over 25 years of experience in modelling and data analysis for policy applications. Since 1990, he is with the Joint Research Centre of the European Commission. Carlo coordinates the overall development of the LUISA Territorial Modelling Platform and of the European Commission Knowledge Centre for Territorial Policies. His interests are in interdisciplinary research related to urban and regional development.