Tourism Management 68 (2018) 101–115

Contents lists available at ScienceDirect

Tourism Management

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

Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources

Filipe Batista e Silva*, Mario Alberto Marín Herrera, Konštantín Rosina, Ricardo Ribeiro Barranco, Sérgio Freire, Marcello Schiavina

European Commission, Joint Research Centre, Italy

HIGHLIGHTS

- Online booking services are emerging sources of detailed information on tourism accommodation supply.
- Data from online booking services and official statistics were combined to assess spatiotemporal patterns of tourism.
- The relative impact of tourism and its seasonality show marked uneven spatial distributions.
- Potential to further exploit seasonal information from online booking services exists.

ARTICLE INFO

Article history:
Received 27 October 2017
Received in revised form
23 February 2018
Accepted 27 February 2018
Available online 15 March 2018

Keywords:
Big data
Mapping
Tourist density
Tourism intensity
Seasonality
Europe

ABSTRACT

Available statistics on tourism from official European sources are limited in terms of both the spatial and temporal resolutions, curbing potential analyses and applications relevant for tourism management and policy. In this study, we produced a novel, complete and consistent dataset describing tourist density at high spatial resolution with monthly breakdown for the whole of the European Union. This is achieved thanks to the integration of data from conventional statistical sources with big data from emerging sources, namely two major online booking services containing the precise location and capacity of tourism accommodation establishments. The produced dataset allowed us to uncover key spatiotemporal patterns of tourism in Europe at unprecedented detail, showcasing the usefulness of complementing official statistical data with emerging big data sources.

1. Introduction

Tourism is a phenomenon with increasing social and economic importance but which has characterised human behaviour for centuries (Butler, 2015). The recent boom in tourism made it an important economic sector in the European Union (EU), but also in other parts of the world. In 2016, the EU had an estimated 40.5% market share of global international tourist arrivals, or around 500 million (UNWTO, 2017). According to available estimates, the total contribution (direct + indirect + induced) of the travel and tourism sector to the EU's GDP in 2016 was 10.2%, but with strong variation between countries, ranging from more than 20% in Malta, Croatia or Cyprus to about 5% in Poland, Netherlands or Romania (World Travel and Tourism Council, 2017). Besides, the importance of tourism as a factor of economic growth has also been demonstrated in many countries by recent studies (Brida, Cortes-Jimenez, & Pulina, 2016; Ohlan, 2017; Perles-Ribes et al., 2017; Salmani, Hossein, & Somayeh, 2014; Seghir et al., 2015).

Tourism has an important territorial dimension, with uneven spatial distribution between and within countries, and delivering localized impacts. The importance of the spatial dimension of tourism is also underscored by findings indicating that tourism growth in one region influences positively tourism in neighbouring regions (Romão, Guerreiro, & Rodrigues, 2017), or that public policy can impact on the spatial patterns of tourism demand (Kang, Kim, & Nicholls, 2014). Seasonality is another distinctive feature of this economic sector, with significant socioeconomic and environmental implications (Butler, 2001; Chung, 2009). Seasonality itself

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

https://doi.org/10.1016/j.tourman.2018.02.020
0261-5177/

Table 1
Data and sources used.

Ref.	Variable/dataset description	Spatial resolution	Temporal resolution	Reference year	Source(s)
a	Nights spent at tourist accommodation establishments	NUTS-2	Annual	2016	Eurostat
b	Number of bed-places	NUTS-3	Annual	2011	Eurostat
c	Nights spent or arrivals at tourist accommodation establishments	NUTS-2/3	Quarterly or monthly	2011	National Statistical Offices
d	Location and capacity (no. of rooms) of tourism accommodation facilities	Lat. and long. coordinates	Not applicable	2017	Online booking services

has a marked geographical structure, varying considerably from region to region, depending on climate and type of destination (e.g. city, sea-side, mountain) (Butler, 2001). Together, these two dimensions of tourism, i.e. the spatial and the temporal, are fundamental to characterise and study tourism in a given territory. And the more countries or regions the area of study encompasses, the more diverse it is likely to be, and the higher the need for sufficiently detailed and comparable spatiotemporal data on tourism.

Consistent tourism data for the EU are primarily assembled and published by Eurostat. However, currently available data from Eurostat have limited spatial and temporal resolutions, hindering EU-wide characterization of tourism at fine spatial and temporal scales. Unconventional, big data sources are emerging, with the potential to improve our knowledge of tourism at unprecedented detail for vast world regions. But, to the best of our knowledge, there are still only a few examples of the use of such emerging sources of data to characterise spatiotemporal patterns of tourism and typically for limited study areas.

The main aim of this study and, simultaneously, its main contribution to international literature is to improve the existing knowledge base of current spatiotemporal distribution of tourism in the EU-28 to enable new insights and applications relevant to tourism management and policy. This main objective can be broken down in four intermediate objectives or tasks, each leading to a tangible output: (i) increase the geographical detail of existing statistics on spatial distribution of tourism demand down to regional level; (ii) derive regional temporal profiles (monthly) of tourism demand; (iii) generate tourist density maps at high spatial resolution on a monthly basis and (iv) exploit the produced information to assess relevant dimensions of tourism regionally such as tourism intensity, seasonality and vulnerability.

To accomplish these objectives, we combined data from two distinct sources: European official statistical bodies, namely Eurostat and National Statistical Offices (NSOs) and online booking services. From Eurostat, we collected nights spent and accommodation capacity at regional level. From NSOs we assembled nights spent or arrivals at tourist accommodation establishment per quarter or month and per region. Finally, from online booking services, geographic coordinates and other descriptors of accommodation establishments were mined, totalling ca. 843 thousand individual records. The datasets were then combined using a predefined protocol to produce multi-temporal grid maps of tourist density at high spatial resolution (100×100 m).

In the following section, we briefly review the current state-of-the-art concerning existing official tourism statistics and examples of the use of unconventional, big data sources for the study of tourism. In the Data and Methods section, we describe in more detail the various input data and the methodology applied to combine them. In the Results section, we show maps of tourist density for Europe and report findings concerning tourism prevalence, seasonality, and intensity, which we finally combine to assess regional vulnerability to shocks in the tourism sector. The last section of the paper wraps-up and discusses the work done and sets out areas that would benefit from further development.

2. Statistical and big data for tourism

When looking at tourism for a territory as large as the EU, the primary source of data is Eurostat.[1] Official statistical bodies such as Eurostat assemble and publish an important set of tourism-related statistical data with regional breakdown. Eurostat usually dedicates a chapter to tourism in its regional statistical yearbooks (e.g. Eurostat, 2016). Statistical data from Eurostat with regional breakdown include, on the demand side, arrivals and nights spent at tourist accommodation establishments, while, on the supply side, capacity of tourist accommodation establishments. All the regional data provided by Eurostat is available on a yearly basis (figures per region and per reporting year). Although relevant to characterising tourism demand and supply density in Europe at the regional level, these statistics do not permit uncovering the spatiotemporal patterns at fine resolution.

While the spatiotemporal resolution offered by official statistical data sources might remain limited, other non-conventional data sources are emerging. These new sources of information, often called 'big data' sources, for their variety, volume and velocity (Katal, Wazid, & Goudar, 2013), are enabling new opportunities for research and analysis in a myriad of domains, including tourism (Benjelloun, Lahcen, & Belfkih, 2015; Rodríguez-Mazahua et al., 2016). In fact, the applications of big data for tourism analytics seem to be growing by the day, and are now numerous and diverse. Social media has been used as a source of user-generated content (e.g. user/customer reviews, posts, photos) to assess international mobility patterns (Hawelka et al., 2014), estimate visitation rates of specific attractions (Wood et al., 2013), identify tourist hot-spots in cities (Garcia-Palomares, Gutierrez, & Minguez, 2015), or to fine-tune tourism marketing strategies (Marine-Roig & Anton Clavé, 2015). Other studies have used web search engine queries to forecast tourism demand for specific destinations (Li et al., 2017, pp. 57–66), or scraped online booking services to monitor hotel prices (Goni et al., 2017).

Mobile network operator (MNO) data is another emerging input for tourism analytics and a particularly promising one for mapping and monitoring patterns of presence of tourists at high spatial and temporal resolutions. Data derived from the use of mobile phones and geo-located to antennas already enabled researchers to assess spatiotemporal visitation patterns of tourist destinations in Estonia (Ahas et al., 2008; Raun, Ahas, & Tiru, 2016). Following these early advances, statistical bodies are conducting pilot studies to test the use of MNO data in the production of official tourism statistics (Dattilo & Sabato, 2017; Demunter & Seynaeve, 2017). However, the use of this data source in a systematic fashion is still hurdled by data access constraints, as profit-driven MNOs are still reluctant to release their data, as proper business models are not yet well established (Debusschere, Wirthmann, & De Meersman, 2017). In addition, there are several methodological challenges associated with the use of MNO data. These include incomplete penetration rates and lack of data for 'roaming' users (Dattilo & Sabato, 2017), heterogeneous market shares of MNOs across regions and

[1] http://ec.europa.eu/eurostat.

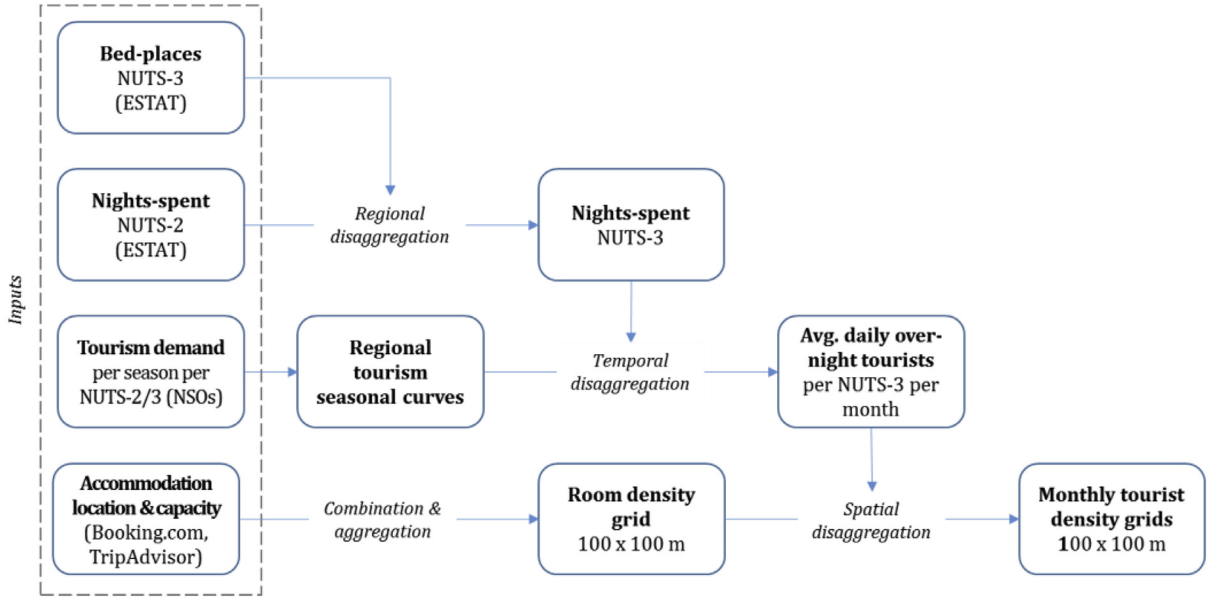


Fig. 1. Tourism data disaggregation workflow.

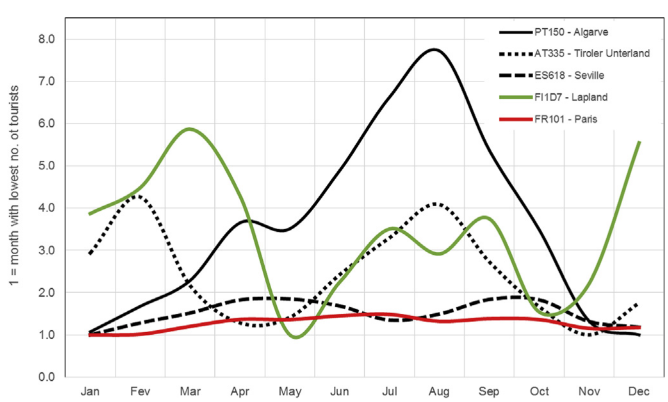


Fig. 2. Seasonal curves for selected regions in Europe, 2011. Source: NSOs. Own elaboration.

socioeconomic groups and issues with mobile phone usage patterns by different users, all leading to selection biases (Demunter & Seynaeve, 2017).

For the reasons reviewed above, the potential of official statistics and big data sources when used in isolation is still limited, especially for applications requiring high spatial and temporal resolutions for vast study areas. But their combination, as proposed and applied in this paper, can yield what is still lacking in international literature: a comprehensive, consistent assessment of current spatiotemporal patterns of tourism for the EU-28 at high resolution.

3. Data and methods

To advance the spatiotemporal mapping of tourist density in Europe, we aimed at producing a set of tourist density grids at high spatial resolution on a monthly basis, i.e. 12 tourist density grids at 100×100 m resolution. Herein, 'tourist density' is short for 'average daily number of overnight tourists' per given spatial reporting unit. In other words, an approximation to the number of tourists that can be found at a given location in a typical day of the month. Here, locations refer to accommodation establishments, i.e. where tourists stay predominately during the night-time for shelter and rest, and tourist density encompasses all types of visitors regardless of the motivation of the visit (e.g. business, leisure or personal purpose),[2] and includes both domestic (i.e. national) and non-domestic (i.e. international) visitors, while excluding same-day visitors.

To produce the tourist density grids, we resorted to variables sought from various sources, and with different characteristics regarding the spatial and temporal detail, data structure and format. Table 1 summarises the main data inputs used and their characteristics. The data were then integrated following a set of operations conducted using statistical software and Geographical Information Systems (GIS). The methodological workflow is outlined below in five main steps, with the letters in brackets referring to the datasets listed in Table 1. Fig. 1 illustrates the workflow.

1. Downscale yearly nights spent (a) from NUTS-2 to NUTS-3[3] level proportionally to the number of bed-places available per NUTS-3 (b);
2. Breakdown the resulting yearly nights spent at NUTS-3 level by months using the share of nights spent (or arrivals) per month derived from NSOs data (c);
3. Transform the resulting monthly nights spent per NUTS-3 to 'average daily number of overnight tourists' by dividing the total nights spent in a month by the total number of days of the corresponding month (e.g. January has 31 days, so every 31 nights spent correspond to an average of 1 tourist on a daily basis);
4. Aggregate the number of rooms from point data (d) to a grid system of 100×100 m cells;
5. Disaggregate the average daily number of overnight tourists per month and per NUTS-3 from step 3 to grid level proportionally to the accommodation capacity as derived according to step 4.

[2] Consistent with the definition of 'tourism' from UNWTO (2010).
[3] NUTS (Nomenclature of Territorial Units for Statistics) is the Eurostat's official regional subdivision for collection and reporting of statistical data. It is structured in four hierarchical levels, from NUTS-0 (countries) to NUTS-3 (sub-regions).

Table 2
Number of accommodation establishments, rooms and bed-places by data source for EU-28.

	Establishments	Rooms	Bed-places
Booking.com	532,346	7,528,249	n/a
TripAdvisor	310,958	9,818,732	n/a
Combined	716,103	13,218,804	n/a
Eurostat	597,358	n/a	30,850,722

Notes.
1) All values refer to the territory of EU-28, excluding Atlantic islands of Portugal and Spain and French overseas territories.
2) Figures from Eurostat refer to the year 2016, except for Ireland and Portugal (2015).
3) Figures regarding Booking.com and TripAdvisor as of February 2017 and August 2017, respectively.

The procedure described in step 1 assumes a good fit between demand and supply at the NUTS-3 level. While that correlation cannot be assessed at that level with available data, the NUTS-2 level data indicates a strong positive correlation of 0.88 between total nights spent and accommodation capacity measured as the available number of bed-places. The reason, however, for an imperfect correlation between demand and supply relates to regional variability in accommodation occupancy rates, related to regional-specific drivers and typologies of tourism, difficult to assess EU-wide.[4]

As for step 2, the monthly breakdown of yearly nights spent was supported by data collected from every NSO in EU-28. Data were either extracted directly from the NSO website or delivered upon prior written request. From each NSO, we used nights spent or no. of arrivals per NUTS-3 or NUTS-2, and per month or quarter, depending upon data availability. The extracted data were used to derive region- and month-specific shares of nights spent or arrivals, which were then applied to the yearly nights spent per NUTS-3 obtained in step 1. When only quarterly data were available, a mean-preserving smoothing interpolation (Rymes & Myers, 2001) was applied to generate monthly shares. Greece was the only country for which temporal data were not available at sub-national level. One issue concerning this step relates to the mismatch between the years of observation of the regional monthly data (2011) and the regional yearly tourism demand figures that we wanted to break down (2016). This issue should, however, be fairly minor because seasonal patterns of tourism demand tend to be stable due to climate, institutional reasons and inertia (Butler, 2001).[5]

The monthly share of nights spent or arrivals, as per the above-mentioned NSO data, can be plotted to reveal rather distinct seasonal curves amongst regions, as shown in Fig. 2. The curves show the ratio between each month's average daily number of overnight tourists and the minimum value observed in the series so that 1 = month with the lowest number of tourists. Several patterns can be identified: strong unimodal distribution for the Algarve region (predominantly a beach destination) with a strong peak in August; very marked bimodal distribution for the Tirol region (mountain tourism); multimodal distribution for Lapland; the relatively flat curve and bimodal distribution for Seville, and the even flatter curve for Paris, indicating a constant inflow of tourist throughout the year.

The procedure described in step 3, whereby total nights spent are converted to average daily number of tourists, was introduced in this study for three reasons. First, it corrects for the different number of days of each month, which slightly distorts the total number of nights spent per month. Second, number of tourists is more tangible than number of nights spent, thus easing the interpretation. Third, it allows for a more straightforward comparison with other regional socioeconomic figures such as on residents or employment.

The aggregation of the point-based number of rooms to 100×100 m grids cells (step 4) took into account two major online booking services, Booking. com and TripAdvisor. [6]Table 2 reports basic statistics regarding the volume of each of the two datasets for the EU-28. To ensure maximum coverage of accommodation establishments, we considered information from both datasets, while minimising potential double counting. A procedure was therefore designed to identify and remove overlapping accommodation establishments between the two datasets. The Booking. com dataset was defined as the baseline due to the higher amount of records. Each record from TripAdvisor was then evaluated against records from Booking. com within a 250 m radius. Duplicates were identified by applying a degree of similarity that took into account positional proximity as well as the difference in the number of rooms between records. As a result, about 127 thousand records from TripAdvisor dataset were considered to have a duplicate in the Booking. com dataset and were discarded. The mean separation distance between duplicates was 45 m, and over 94% of the removed pairs had equal room count in both datasets.

The final combined dataset included more than 716 thousand points, nearly 35% more than in the Booking. com dataset (see Table 2). Official figures from Eurostat (2016) report about 20% fewer establishments than the number of records in the combined dataset for the same geographical area. Reasons for this apparent mismatch could be related to the definition of establishments (e.g. one establishment in official statistics corresponding to more than one record in online booking services), reporting biases (lack of certain accommodation categories in official statistics) and under-detection of duplicates in the combined dataset. Notwithstanding, the ratio between the total number of bed-places from Eurostat and the number of rooms from our combined point dataset yields a plausible 2.3 bed-places per room. Fig. 3 shows the resulting room density per grid cells of 10×10 km.

Finally, in step 5, overnight tourists per month in each NUTS-3 were disaggregated to 100 m cells containing accommodation rooms proportionally to the number of rooms in the cell. The resulting grid map allocates tourists to accommodation establishments, thus being a plausible representation of tourist density during the night-time when most tourists are assumed to be located in their rooms for shelter and rest.

In summary, the above-described workflow can be put as a set of sequential variable downscaling or disaggregation steps, first from NUTS-2 to NUTS-3 level, then from annual to monthly ('temporal disaggregation') and finally from NUTS-3 to fine grid cell level. The spatial disaggregation steps are often referred in the literature to as dasymetric interpolation or dasymetric mapping (Mennis, 2003). Batista e Silva, Gallego, and Lavalle (2013) defined it as "cartographic technique whereby ancillary thematic data is

[4] Regression analysis was employed to assess additional factors determining tourism demand at regional level. Although factors such as total regional GDP, capital region, and distance to the most relevant airport were found significantly associated with nights spent at the NUTS-2 level, the contribution to explanatory power was negligible when compared to a model using only the number of bed-places.

[5] This has been confirmed for regions in selected countries (i.e. Spain, France and The Netherlands), where series of regional monthly nights spent for different years between 2011 and 2016 were compared, often showing correlations of nearly 1.

[6] The data were obtained using ad-hoc routines for web data extraction. The data collected consisted of the geographical coordinates of accommodation establishments and the respective number of rooms, and used for the sole purpose of this non-profit, non-commercial research.

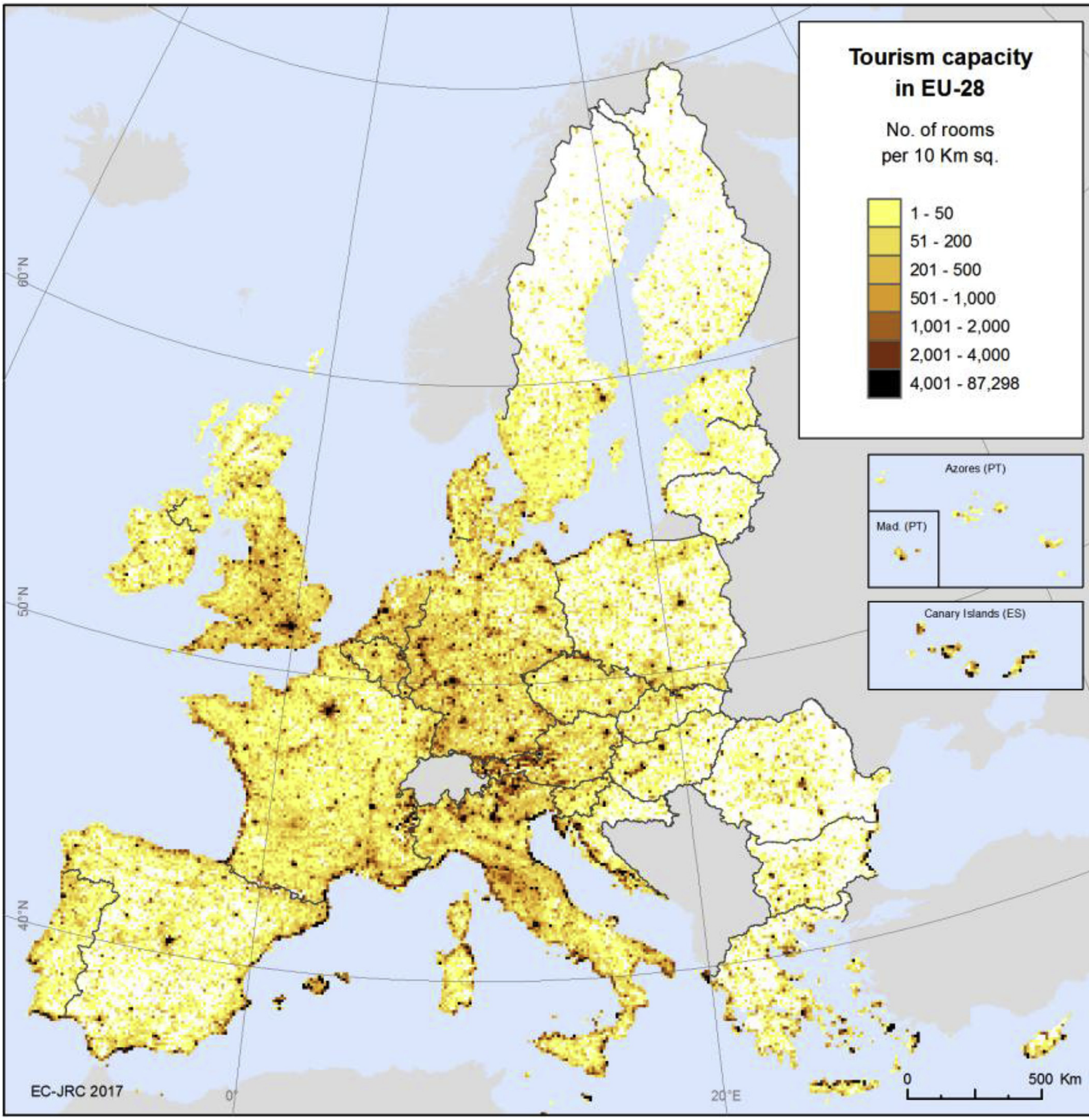


Fig. 3. Tourism capacity in EU-28 in 2017. Source: Booking. com and TripAdvisor. Own elaboration.

used to refine the geographical representation of a quantitative variable reported at coarse spatial aggregations". Although dasymetric mapping is mostly applied to produce fine-grained maps of residential population distribution from coarser statistical zoning systems, applications to other variables have been reported such as mobile phone users (Järv, Tenkanen, & Toivonen, 2017), crime events (Mennis, 2016) or even tourist density (Vaz & Campos, 2013).

In addition to the obvious gains in spatial detail and accuracy, another key advantage of fine-grained tourist density grids is that they allow for flexible variable aggregations to any other zoning system, including alternative regular grid systems. Furthermore, contrary to administrative boundaries, regular grid systems allow for comparability of variables across spatial units due to their homogenous size, thus reducing the zone effect of the so-called Modifiable Areal Unit Problem (MAUP) (Openshaw, 1983).

4. Results

4.1. Main spatiotemporal patterns of tourism in Europe

To illustrate the rich spatial granularity of the produced dataset,

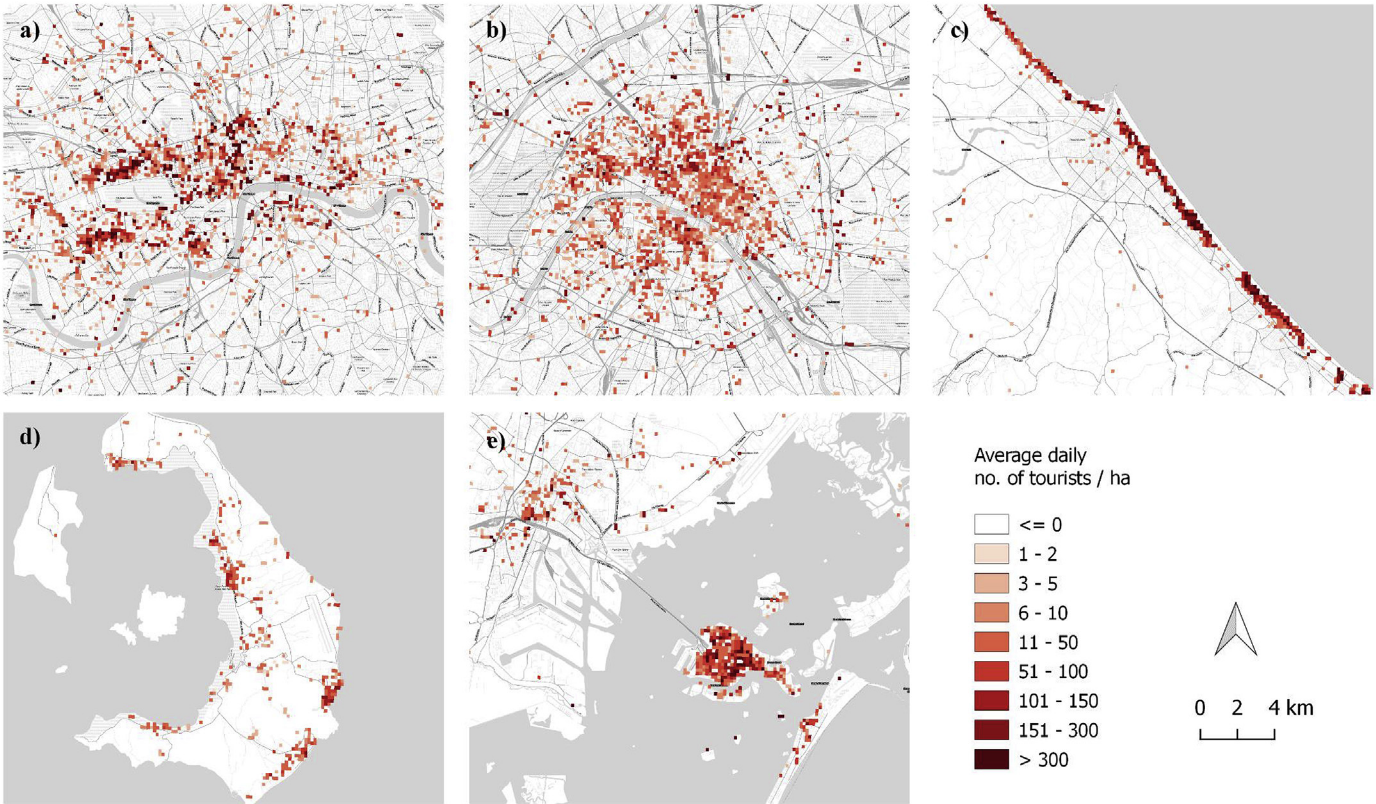


Fig. 4. Tourist density in August 2016 in selected locations: a) London, b) Paris, c) Rimini d) Santorini, e) Venice. Source of background map: Stamen Design with data by OpenStreetMap.

Fig. 4 shows tourist density in the month of August at 100 × 100 m resolution for a selection of different popular tourist destinations across Europe. The dataset reveals significant differences in the spatial distribution of tourism, including sprawled patterns (London and Paris), clustered (Santorini), concentrated (Venice) and linear (Rimini), owing to local geography and typology of tourism.

To facilitate the visualisation of tourist density at European scale, we aggregated the grids originally produced at 100 × 100 m to a larger cell size. The maps in Fig. 5 show tourist density per 10 km^2 cells for selected months (i.e. January, May, August and November). The maps' legends were kept invariant across months to allow comparing the spatiotemporal variation of tourist density.

Although not in every location, tourist density is generally the highest in August. The largest cities in Europe tend to be hotspots of tourism throughout the year. Coastal areas and islands are also popular year-round but peak significantly in summer months. Alpine areas display high tourist densities in both summer and winter but are comparatively less dense in mid-season (spring and fall). Many parts in the centre and west of Europe, particularly the Netherlands, Germany, as well as Britain have typically very high tourist densities throughout the year. A possible explanation is the high population density of these countries, possibly combined with a high prevalence of business- and/or cultural-related tourism which are less affected by climate conditions. Conversely, the northern and eastern European countries show generally lower tourist densities.

To further ease visualisation and analysis, the monthly gridded tourist densities were aggregated to regular hexagons each 25 km wide and 541 km^2 in size. We found that among the top ten locations (i.e. hexagons) by tourist density, seven correspond to capital cities (London, Paris, Berlin, Madrid, Rome, Prague and Vienna). The remainder corresponds to beach-tourism destinations in Spain. Table 3 shows the average daily number of overnight tourists in the top ten locations in Europe in 2016 according to our results.

The maps in Fig. 6 highlight the top 5% and 10% most popular locations (regarding the number of tourists) for each of the four seasons. It is noticeable that many locations are persistently popular across the various seasons; in particular, the largest cities in Europe, the Alps, many parts of the Netherlands, Britain, west Germany and centre-north of Italy, but also many seafront areas. However, the coast of the Black Sea, the Greek, Italian and French islands and the Croatian coast are only among the most popular in the warm months. In the mid-season, some sparse locations in Ireland and Scandinavia are among the most attractive as well.

Fig. 7 indicates the season with the highest number of overnight tourists per region, confirming that summer is the most popular season for almost every region in Europe. This is explained by two important, correlated facts: the summer months, and particularly August, are those when most people traditionally go on holidays, and when many activities are closed (e.g. education) or have reduced activity (e.g. manufacturing). In addition, the warm temperatures are a very important pull factor for holidays in the majority of regions. Nonetheless, there are some exceptions. The winter season is the most popular in some alpine and Scandinavian regions due to favourable natural conditions for winter sports/activities. Autumn seems to be very popular in Ireland as well as in some inland regions of Romania, Bulgaria and Croatia. Finally, spring is the most popular in some city-regions such as Rome, Brussels, Madrid, Bucharest, Milan, or Linz, plus areas in Andalusia, Bulgaria, east Croatia and north of Paris.

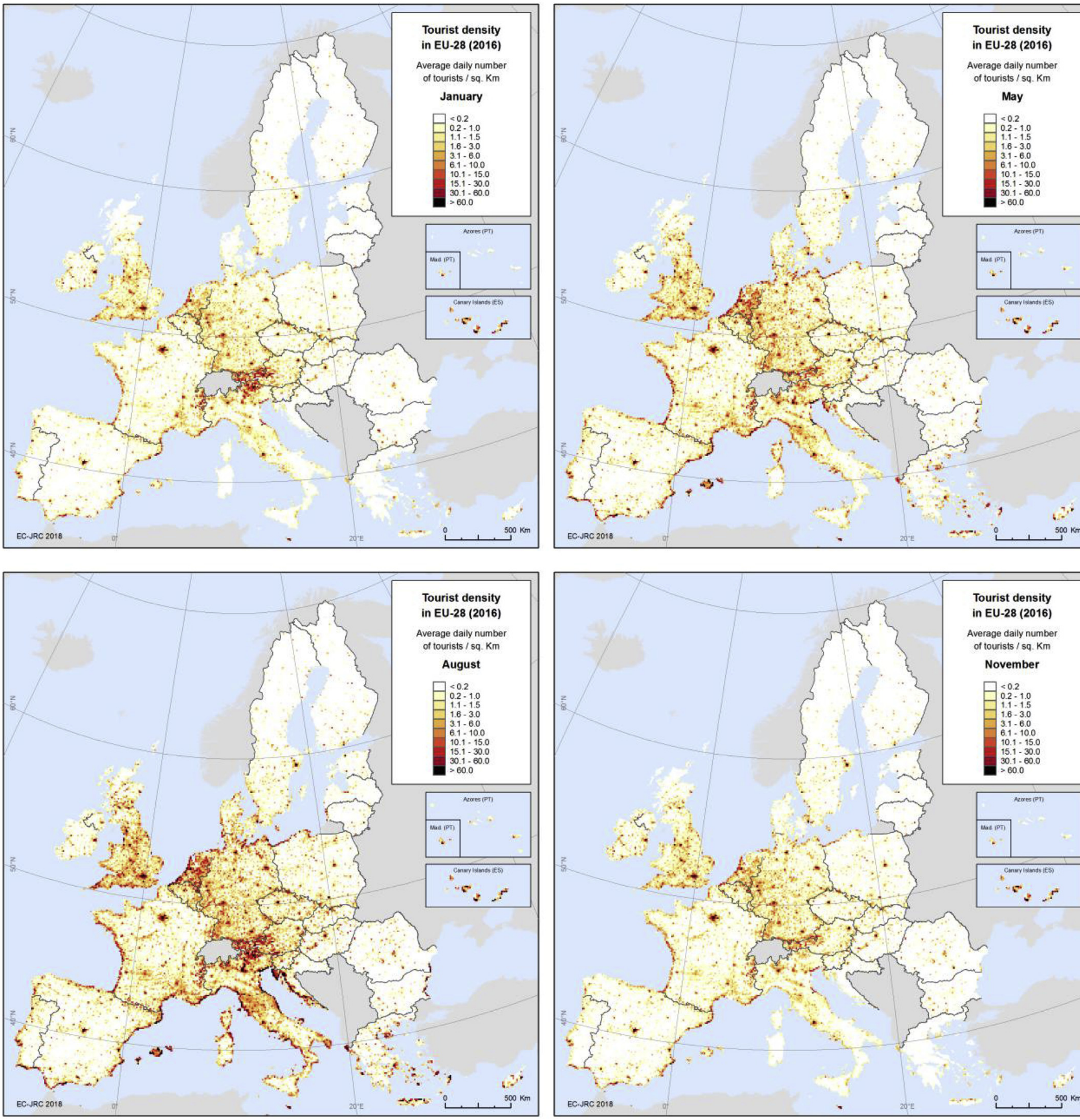


Fig. 5. Tourist density in selected months of the year in EU-28, 2016.

4.2. *Tourism intensity, seasonality and vulnerability*

In the previous section, spatiotemporal patterns of tourism in Europe were examined using the herein constructed tourist density dataset, revealing substantial uneven distribution of tourism demand both in space and time. It is thus evident that regions are affected by tourism very differently: some are little or not touristic at all, while others are very touristic ('tourism intensity'); some receive fairly steady tourist inflows year-round, while inflows are particularly uneven throughout the year in others ('tourism seasonality').

Tourism *intensity* could be defined as the relative importance of

Table 3
Top 10 locations by tourist density in Europe in 2016.

Rank	Location	Average daily number of overnight tourists ('000)
1	London	111.4
2	Paris	99.9
3	Berlin	73.5
4	Gran Canaria	58.9
5	Madrid	49.7
6	Tenerife	47.8
7	Rome	44.6
8	Prague	43.3
9	Vienna	37.6
10	Palma de Mallorca	35.6

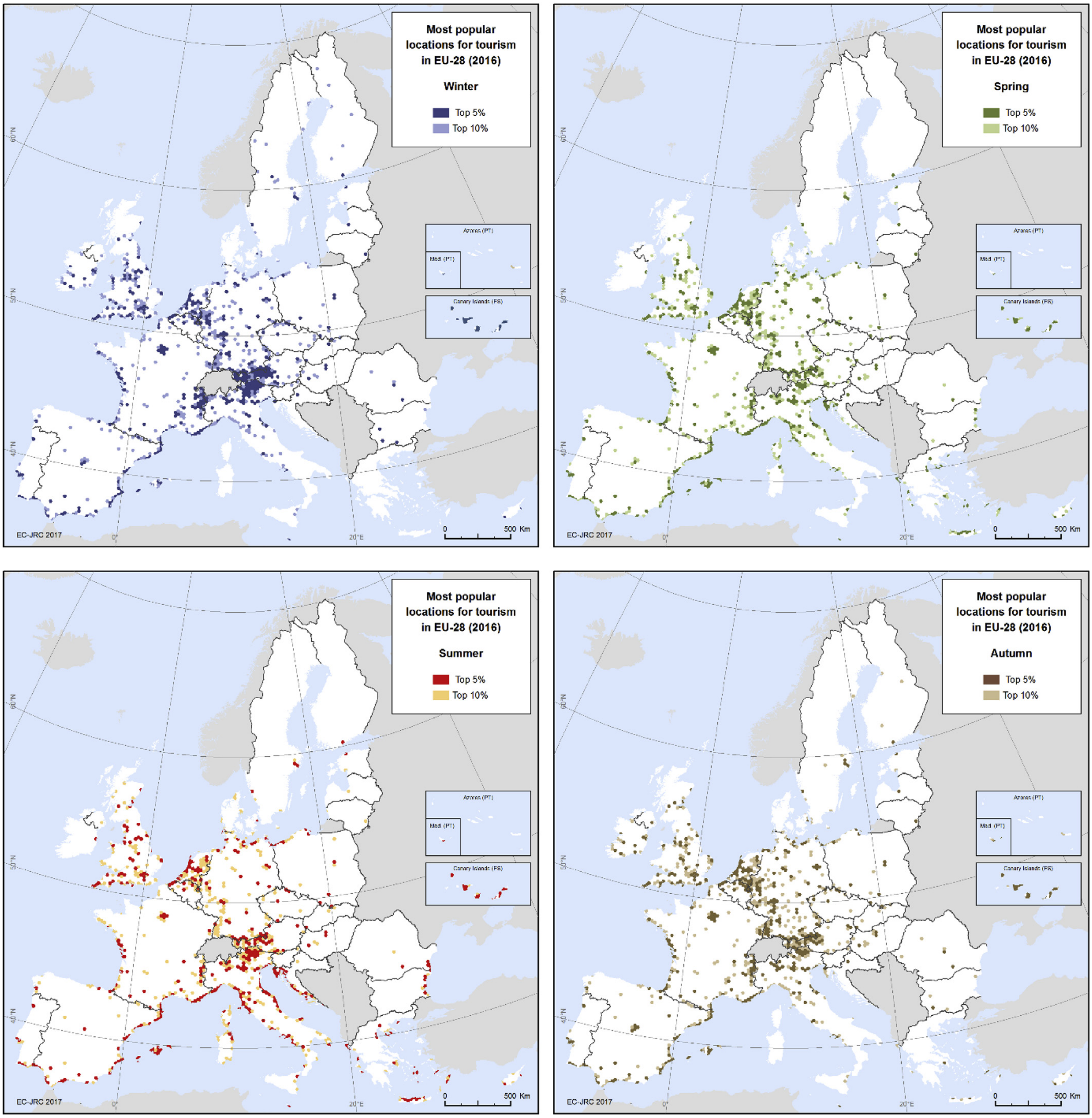


Fig. 6. Most popular locations for tourism per season in EU-28, 2016.

tourism for a region. This can be measured, for example, by calculating the ratio between tourism demand and residential population. According to Eurostat (2010), this simple indicator is a "better guide to the economic significance of tourism for a region than the absolute number of overnight stays. Furthermore, in the context of the sustainability of tourism, it can also be seen as an indicator of the possible tourism pressure". In turn, tourism *seasonality* is the fluctuation or variation of tourist inflows during the year in a given territory. According to Butler (2001), seasonality is influenced by factors related to both the demand and supply sides. Demand-side factors include response to climate variation between seasons, institutionalised holidays and vacation tradition/inertia. Supply-side factors include climate conditions, physical attractions, opportunities for activities and socio-cultural events. Although very common, seasonality is usually seen as an undesirable aspect of tourism, as it determines fluctuation of revenue, employment, as well as under- and over-utilisation of infrastructure, services and resources. Seasonality can, however, also have positive effects as it provides a period of rest for the regeneration of natural resources or reestablishment of socio-cultural features (Bender, Schumacher, & Stein, 2005; Chung, 2009; Grizane, 2016).

Although these two properties, tourism intensity and

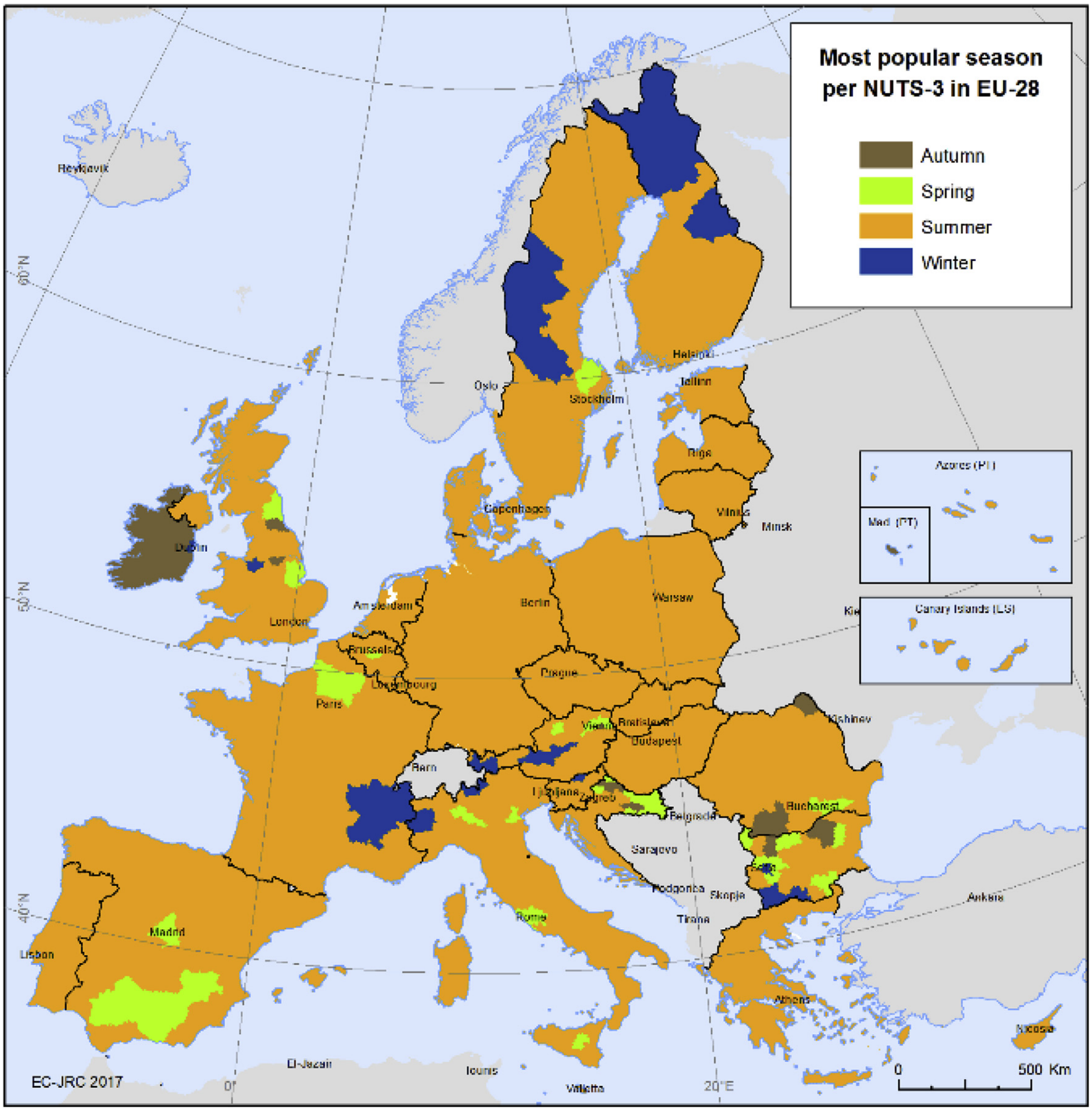


Fig. 7. Most popular season per NUTS-3 in EU-28, 2011.

seasonality, are in themselves interesting to characterise tourism at regional level, we argue that their combination can reveal a third, policy-relevant, property of tourism in regions: regional *vulnerability* to shocks in the tourism sector, or regional vulnerability to tourism, for short. We define *regional vulnerability to tourism* as the susceptibility of a region to be affected in case of shocks or disruptions in the tourism sector.

Regions with both high tourism intensity and high seasonality are deemed to be more vulnerable to the tourism sector and any shocks that may affect it (e.g. economic crises reducing overall demand for tourism, terrorism events, environmental or socioeconomic disruptions reducing demand for, or transport access to, certain destinations). Conversely, a region with low tourism intensity and low seasonality is less vulnerable to shocks affecting the tourism sector. Indeed, recent studies point out high seasonality as a factor of vulnerability of tourism destinations (van der Veeken et al., 2015).

Based on the dataset developed in this study, we propose a *regional vulnerability to tourism index* that relies on the two pillars described above. The implemented strategy is threefold. First, we measure the degree of tourism intensity and seasonality in each region using suitable quantitative and continuous variables. Second, we chunk each series based on quartiles, with each region scoring from 1 to 4 for both intensity and seasonality. Third, tourism vulnerability is calculated as the product of the respective scores for intensity and seasonality, thus ranging from 1 to 16.

Tourism intensity is often defined as the ratio between a measure of tourism demand (e.g. tourist arrivals or nights spent) and a measure of the demographic or economic size of a region, as in Eurostat (2010), Dumbrovská and Fialová (2014) or Liu and Pratt (2017). In here, we apply the following location quotient of tourism (*LQtur*) (Voltes-Dorta, Jiménez, & Suárez-Alemán, 2014):

$$LQtur_i = \frac{\frac{tourists_i}{\sum_{i=1}^{n} tourists_i}}{\frac{pop_i}{\sum_{i=1}^{n} pop_i}} = \frac{\frac{tourists_i}{pop_i}}{\frac{\sum_{i=1}^{n} tourists_i}{\sum_{i=1}^{n} pop_i}} \quad (1)$$

where *tourists* corresponds to the average daily number of overnight tourists over the year (i.e. number of nights spent/365), *pop* is

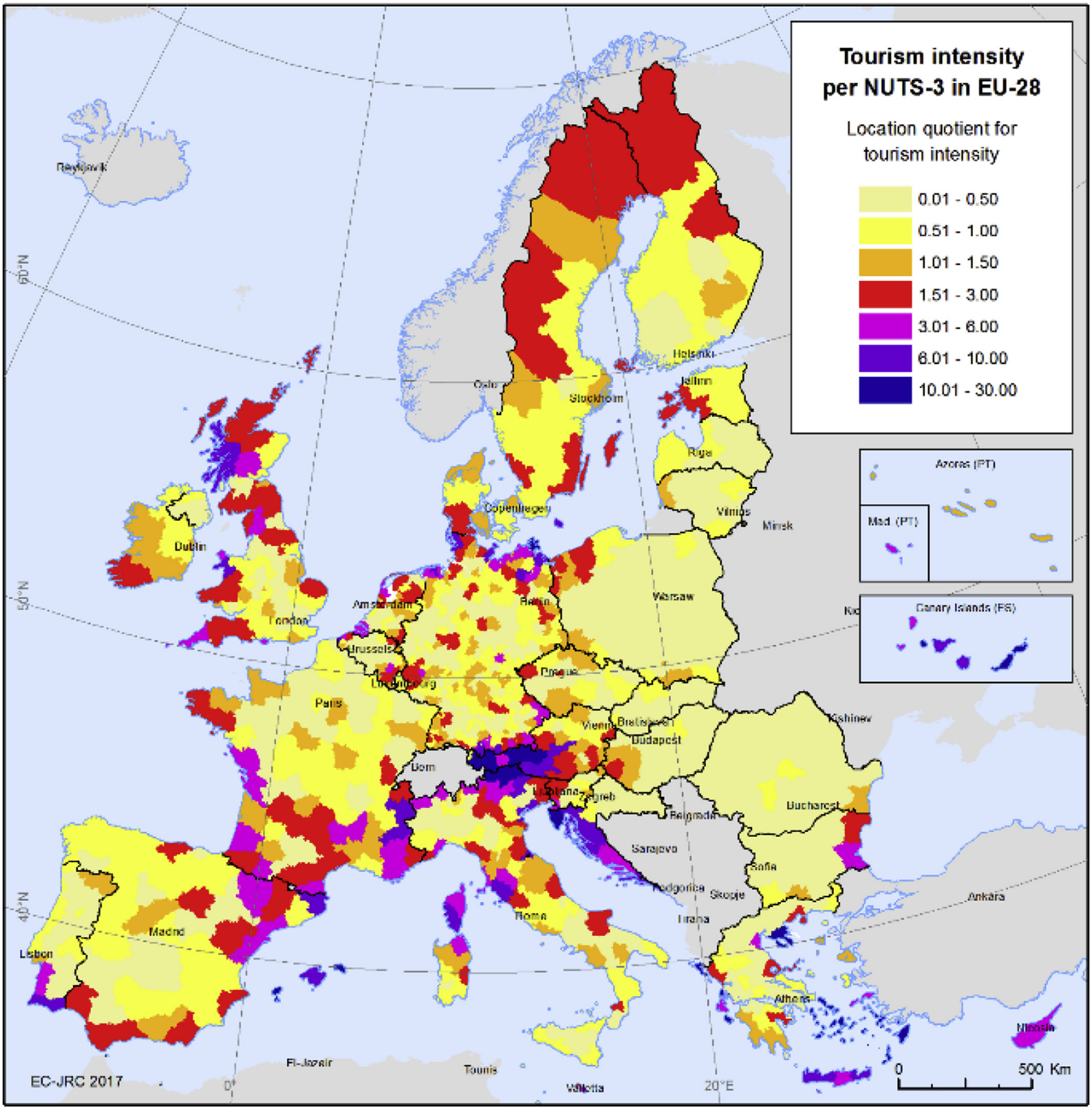


Fig. 8. Tourism intensity per NUTS-3 in EU-28, 2016.

the total resident population, i is a NUTS-3 region and n is the total number of NUTS-3 regions in the study area (EU-28). The location quotient is a way of quantifying how concentrated a particular activity is in a region as compared to a reference territorial unit (e.g. country, continent). In this case, any NUTS-3 region with $LQtur > 1$ is more tourism-intensive than the average tourism intensity in EU-28. In Equation (1), total employment could have been used instead of total resident population. However, we argue that to assess the relative importance of tourism for a region, the most appropriate denominator is total population, as the revenue from tourism may spill, directly or indirectly, to people who are not employed (e.g. through rents, informal economy or family ties). The resulting $LQtur$ is robust to the use of population or employment (correlation of 0.98 between the two options). According to Voltes-Dorta et al. (2014), tourism intensity is possibly better defined as the tourism-related tax revenues relative to the total tax revenue by municipalities. However, such information is not available, nor it would be consistent amongst the countries within our study area. Fig. 8 shows the tourism intensity as herein defined for the EU-28 NUTS-3 regions in 2016. Tourism intensity is highest in the alpine region, Spanish and Greek islands, Algarve, Corsica, central Italy, Croatian and Bulgarian coast, and also parts of Britain.

A possible way to measure seasonality across EU regions is to apply the coefficient of variation (CV) (Bender et al., 2005;

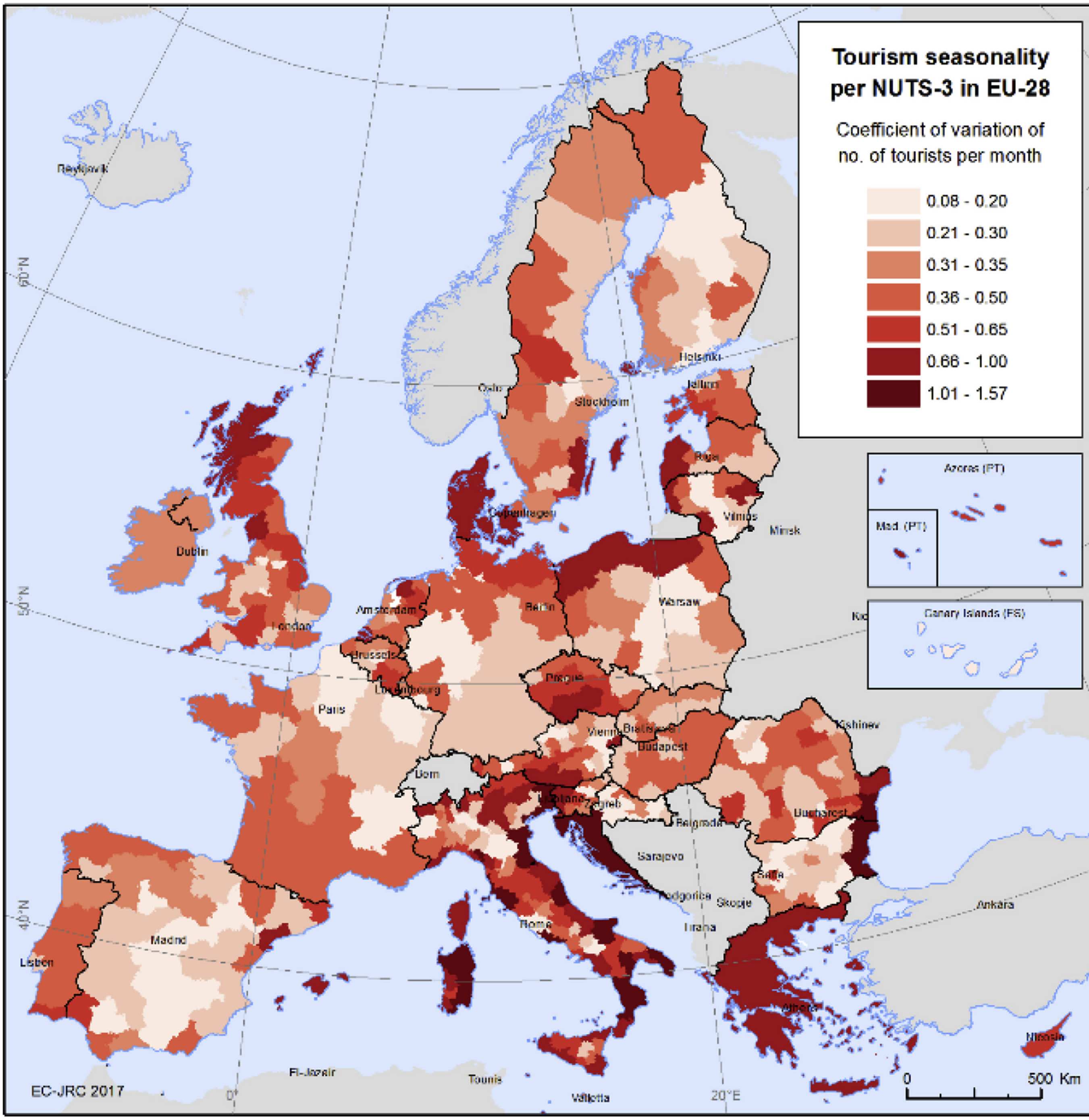


Fig. 9. Tourism seasonality per NUTS-3 in EU-28, 2016.

Yacoumis, 1980) to each region's monthly series of the average daily number of overnight tourists. Because CV is defined as the ratio between the standard deviation and the mean, it can be used for comparing regions regardless of the total number of tourists. A high CV means that the number of tourists varies significantly between months as compared to the annual average, hence high seasonality. Fig. 9 shows the tourism seasonality as herein defined for the EU-28 NUTS-3 regions. Most of the regions highly affected by seasonality are islands and coastal, hence predominantly oriented to beach tourism and thus dependent on climate conditions.

Fig. 10 shows the resulting regional vulnerability index to tourism, as the product of the regional scores for tourism intensity and seasonality. This assessment comes as a useful complement to the tourist density maps from Fig. 4. Although some regions in Northern Europe attract modest numbers of tourists, they are significantly exposed to the tourism sector due to the relative importance of tourism regionally and/or to high seasonality. Conversely, EU capital cities are amongst the most popular destinations, but score low or very low in the regional vulnerability index due to low seasonality and low relative importance of the tourism sector. In general, regions most vulnerable to the tourism sector are coastal (Mediterranean, Atlantic, Baltic and in the Black sea), mountainous (both the Alps and the Pyrenees), plus most of the Mediterranean islands. Italy

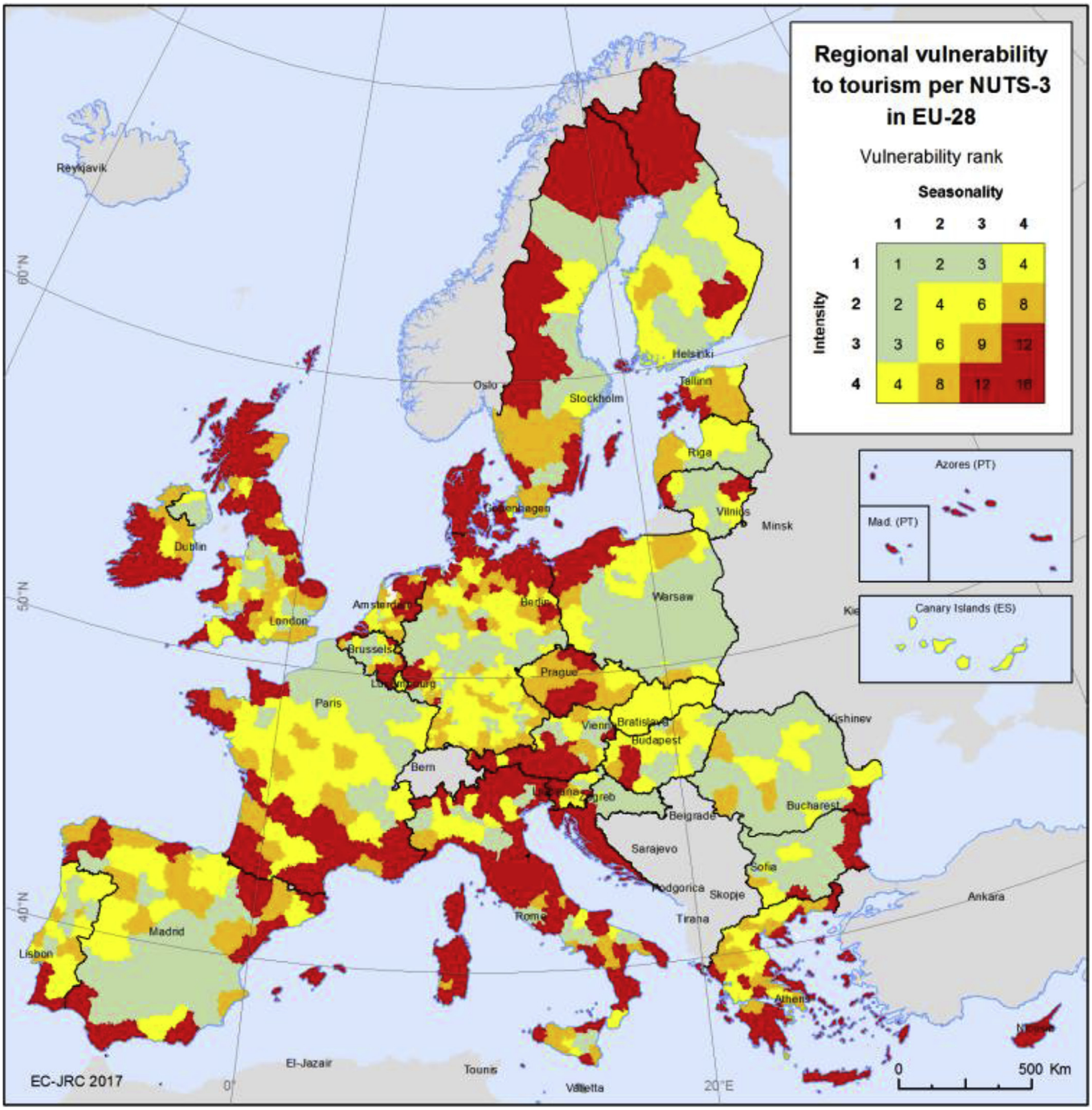


Fig. 10. Regional vulnerability to tourism index per NUTS-3 in EU-28, 2016.

is a remarkable case with a large share of its regions scoring high vulnerability.

5. Discussion, conclusions and way forward

Available statistical data from official European data sources on tourism is limited in terms of both the spatial and temporal resolutions, curbing potential analyses and applications relevant for tourism management and policy. However, combined with emerging, big data sources, conventional statistical data can be enriched by furthering its spatial and temporal granularity. In this paper, we sought, obtained and combined data from multiple data sources to upgrade the state-of-the-art knowledge on tourism for the European Union. We produced a novel, complete and consistent dataset describing the average daily number of overnight tourists per regions and regular grid cells of various shapes and spatial resolutions with monthly breakdown. The produced dataset allowed us to distil key spatiotemporal patterns and characteristics of tourism in Europe at both regional and local scales.

Two main 'ingredients' were integrated with demand-side tourism statistics from Eurostat to achieve the dataset mentioned above: 1) regional seasonal curves derived from data from National Statistical Offices of the EU-28 and, 2) the location and capacity of accommodation establishments from two major worldwide online booking services. The data extraction from NSOs, although indispensable for the refinement of the temporal detail of tourism demand at regional level, proved to be a particularly laborious task due to the number and diversity of databases (28 in this case), posing problems for data extraction automation and scalability. As for the second input used, the completeness of accommodation establishments cannot be fully warranted, even if information from two major worldwide online booking services has been used. Notwithstanding, the information used provides, to the best of our knowledge, the most complete, spatially detailed, and up-to-date picture of the location of accommodation establishments available so far.

Although the produced dataset has value on its own as a contribution to the state-of-the-art of tourism research in Europe, we exploited the novel dataset to deliver new insights concerning current (i.e. 2016) regional patterns of tourism intensity, tourism seasonality and, ultimately, a generic assessment of regional vulnerability to shocks in the tourism sector. In brief, our data and analyses indicate clearly that the relative impact of tourism in Europe and its seasonality vary greatly from country to country and, even more so, from region to region and from locality to locality. Cities, as well as islands, coastal areas and the Alps, tend to be major hotspots for tourism in Europe. Based on the assessment of regional vulnerability to tourism, cities are less susceptible to shocks in the tourism sector as compared to other areas because their dependence on tourism is relatively low and are less affected by seasonality. Although these characteristics are generally acknowledged amongst tourism researchers, the dataset herein produced allows for an inspection of the varying tourist density and intensity levels at unprecedented high spatial and temporal resolutions, consistently for the whole of the EU-28.

The proposed index of regional vulnerability to shocks in the tourism sector integrates the relative importance of tourism (i.e. tourism intensity) and the degree of seasonality per region, demonstrating the analytical potential of the fine-grained tourism data harnessed in this study. Although fairly simple in conception, this indicator's strengths lie on its transparency, consistency and quantitative nature, allowing comparisons between EU's regions. Moreover, it can be updated and improved with a more precise definition of tourism intensity, e.g. a definition based on the actual economic contribution of tourism regionally, should appropriate and comparable data become available for all EU regions. One limitation of the indicator, however, is that it does not account for the capacity or potential of regions to overcome events or conditions that compromise the sector. Finally, as an all-purpose indicator, it is not specific to different types of shocks or to different typologies of tourism destination. Liu and Pratt (2017), for example, have studied the more specific case of vulnerability and resilience of tourism to terrorism at country scale, while Terkenli (2005) argued that in the case of Crete, Greece, impacts of seasonality in the landscape and society may as well vary according to development stage of tourism. A more comprehensive framework to study vulnerability and resilience of tourism destinations has been proposed by Calgaro, Lloyd, and Dominey-Howes (2014).

A follow-up of the herein presented work should focus on the improvement of the seasonal variation of tourism demand regionally by exploiting information from emerging, big data sources. An alternative to resorting to somewhat heterogeneous seasonal data from separate NSOs across Europe could be the use of data available from TripAdvisor. This source provides the number of reviews per accommodation establishment and per season, which could be used to generate consistent and seasonal curves per any desired geographical delimitation. Finally, the accommodation capacity and tourist density maps described in this paper are already being used as input to the production of high-resolution population density grids for Europe that take into account major daily and seasonal population variations (Batista e Silva, 2017).

Disclaimer

The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

Acknowledgements

This study was done in the context of the ENACT project ("ENhancing ACTivity and population mapping") of the European Commission Joint Research Centre. The authors would like to thank Carlo Lavalle for his continuous support, and Eric Koomen, Nicola Pontarollo, Chris Jacobs Crisioni, João Romão and the reviewers for their useful comments and suggestions.

References

Ahas, R., et al. (2008). 'Evaluating passive mobile positioning data for tourism surveys: An Estonian case study'. *Tourism Management, 29*(3), 469–486. https://doi.org/10.1016/j.tourman.2007.05.014.

Batista e Silva, F., et al. (2017). 'Spatiotemporal mapping of population in Europe: The "ENACT" project in a nutshell'. In *57th european regional science association (ERSA) congress*.

Batista e Silva, F., Gallego, J., & Lavalle, C. (2013). 'A high-resolution population grid map for Europe'. *Journal of Maps, 9*(1), 16–28. https://doi.org/10.1080/17445647.2013.764830.

Bender, O., Schumacher, K. P., & Stein, D. (2005). 'Measuring seasonality in central Europe's tourism – how and for what?'. In M. Schrenk (Ed.), *Proceedings of the CORP 2005* (pp. 303–309).

Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2015). 'An overview of big data opportunities, applications and tools. In *2015 intelligent systems and computer vision, ISCV 2015*. https://doi.org/10.1109/ISACV.2015.7105553.

Brida, J. G., Cortes-Jimenez, I., & Pulina, M. (2016). 'Has the tourism-led growth hypothesis been validated? A literature review', Current Issues in Tourism. *Routledge, 19*(5), 394–430. https://doi.org/10.1080/13683500.2013.868414.

Butler, R. W. (2001). 'Chapter 2-seasonality in Tourism: Issues and implications'. In T. Baum, & S. Lundtorp (Eds.), *Advances in tourism research* (pp. 5–21). Oxford: Pergamon. https://doi.org/10.1016/B978-0-08-043674-6.50005-2.

Butler, R. (2015). 'The evolution of tourism and tourism research'. *Tourism Recreation Research, 40*(1), 16–27. https://doi.org/10.1080/02508281.2015.1007632.

Calgaro, E., Lloyd, K., & Dominey-Howes, D. (2014). 'From vulnerability to transformation: A framework for assessing the vulnerability and resilience of tourism destinations', journal of sustainable tourism. *Routledge, 22*(3), 341–360. https://doi.org/10.1080/09669582.2013.826229.

Chung, J. Y. (2009). 'Seasonality in tourism: A review'. *e-Review of Tourism Research, 7*(5), 82–96. Available at http://ertr.tamu.edu/content/issues/volume-7-issue-1-6-2009/volume-7-issue-5-october-2009-applied-research-note-1/.

Dattilo, B., & Sabato, M. (2017). 'Travelling SIM and trips: An approach to make mobile phone data usable in tourism statistics. In *New techniques and technologies for statistics 2017*. https://doi.org/10.2901/EUROSTAT.C2017.001.

Debusschere, M., Wirthmann, A., & De Meersman, F. (2017). 'Official statistics and mobile network operators: A business model for partnerships. In *New techniques and technologies for statistics 2017*. https://doi.org/10.2901/EUROSTAT.C2017.001.

Demunter, C., & Seynaeve, G. (2017). 'Better quality of mobile phone data based statistics through the use of signalling information - the case of tourism statistics. In *New techniques and technologies for statistics 2017. Eurostat*. https://doi.org/10.2901/EUROSTAT.C2017.001.

Dumbrovská, V., & Fialová, D. (2014). 'Tourist intensity in capital cities in central Europe: Comparative analysis of tourism in Prague, vienna and budapest'. *Czech Journal of Tourism, 3*(1), 5–26. https://doi.org/10.2478/cjot-2014-0001.

Eurostat. (2010). *Eurostat regional yearbook 2010 edition*. Luxembourg: Publications office of the European Union. https://doi.org/10.2785/40203.

Eurostat. (2016). *Eurostat regional yearbook 2016 edition*. Luxembourg: Publications office of the European Union. https://doi.org/10.2785/29084.

Garcia-Palomares, J. C., Gutierrez, J., & Minguez, C. (2015). 'Identification of tourist hot spots based on social networks: A comparative analysis of european metropolises using photo-sharing services and GIS. *Applied Geography, 63*, 408–417. https://doi.org/10.1016/j.apgeog.2015.08.002.

Goni, E., et al. (2017). 'Big Data: Automatic hotel prices collection on the internet for the tourism survey in the Basque country. In *New techniques and technologies for statistics 2017. Eurostat*. https://doi.org/10.2901/EUROSTAT.C2017.001.

Grizane, T. (2016). 'Measuring tourism seasonality in regions of Latvia. In

Proceedings of the 2016 international conference "economic science for rural development (pp. 59–64). Available at http://llufb.llu.lv/conference/economic_science_rural/2016/Latvia_ESRD_42_2016-59-64.pdf.

Hawelka, B., et al. (2014). 'Geo-located twitter as proxy for global mobility patterns', cartography and geographic information science. *Taylor & Francis, 41*(3), 260–271. https://doi.org/10.1080/15230406.2014.890072.

Järv, O., Tenkanen, H., & Toivonen, T. (2017). 'Enhancing spatial accuracy of mobile phone data using multi-temporal dasymetric interpolation. *International Journal of Geographical Information Science. Taylor & Francis, 31*(8), 1630–1651. https://doi.org/10.1080/13658816.2017.1287369.

Kang, S., Kim, J., & Nicholls, S. (2014). 'National tourism policy and spatial patterns of domestic tourism in South Korea', journal of travel research. *SAGE Publications Inc, 53*(6), 791–804. https://doi.org/10.1177/0047287514522875.

Katal, A., Wazid, M., & Goudar, R. H. (2013). 'Big data: Issues, challenges, tools and Good practices. In *2013 6th international conference on contemporary computing, IC3 2013* (pp. 404–409). https://doi.org/10.1109/IC3.2013.6612229.

Li, X., et al. (2017). *'Forecasting tourism demand with composite search index'* (Vol. 59). Tourism management Elsevier Ltd. https://doi.org/10.1016/j.tourman.2016.07.005.

Liu, A., & Pratt, S. (2017). 'Tourism's vulnerability and resilience to terrorism'. *Tourism Management, 60*(Supplement C), 404–417. https://doi.org/10.1016/j.tourman.2017.01.001.

Marine-Roig, E., & Anton Clavé, S. (2015). 'Tourism analytics with massive user-generated content: A case study of barcelona. *Journal of Destination Marketing and Management, 4*(3), 162–172. https://doi.org/10.1016/j.jdmm.2015.06.004. Elsevier.

Mennis, J. (2003). 'Generating surface models of population using dasymetric mapping', the professional geographer. *Blackwell Publishing, 55*(1), 31–42. https://doi.org/10.1111/0033-0124.10042.

Mennis, J. (2016). 'Dasymetric spatiotemporal interpolation', the professional geographer. *Routledge, 68*(1), 92–102. https://doi.org/10.1080/00330124.2015.1033669.

Ohlan, R. (2017). 'The relationship between tourism, financial development and economic growth in India'. *Future Business Journal, 3*(1), 9–22. https://doi.org/10.1016/j.fbj.2017.01.003.

Openshaw, S. (1983). 'The modifiable areal unit problem. *Concepts and Techniques in Modern Geography, 38*. Available at https://www.scopus.com/inward/record.uri?eid=2-s2.0-0021054431&partnerID=40&md5=0e215729adaf9cf1f07802af859e11e7.

Perles-Ribes, J. F., et al. (2017). 'Is the tourism-led growth hypothesis valid after the global economic and financial crisis? The case of Spain 1957–2014. *Tourism Management, 61*, 96–109. https://doi.org/10.1016/j.tourman.2017.01.003.

Raun, J., Ahas, R., & Tiru, M. (2016). *'Measuring tourism destinations using mobile tracking data'* (Vol 57, pp. 202–212). Tourism Management. Elsevier Ltd. https://doi.org/10.1016/j.tourman.2016.06.006.

Rodríguez-Mazahua, L., et al. (2016). 'A general perspective of big data: Applications, tools, challenges and trends'. *The Journal of Supercomputing, 72*(8), 3073–3113. https://doi.org/10.1007/s11227-015-1501-1.

Romão, J., Guerreiro, J., & Rodrigues, P. M. M. (2017). 'Territory and sustainable tourism development: A space-time analysis on european regions'. *The Region, 4*(3), 1. https://doi.org/10.18335/region.v4i3.142.

Rymes, M. D., & Myers, D. R. (2001). 'Mean preserving algorithm for smoothly interpolating averaged data'. *Solar Energy, 71*(4), 225–231. https://doi.org/10.1016/S0038-092X(01)00052-4.

Salmani, B., Hossein, P., & Somayeh, R. (2014). 'Assessing the dynamic economic impact of tourism for OIC members. *World Applied Sciences Journal, 32*(6), 1098–1105. https://doi.org/10.5829/idosi.wasj.2014.32.06.1927.

Seghir, G. M., et al. (2015). 'Tourism spending-economic growth causality in 49 countries: A dynamic panel data approach. *Procedia Economics and Finance, 23*, 1613–1623. https://doi.org/10.1016/S2212-5671(15)00402-5.

Terkenli, T. S. (2005). 'Human activity in landscape seasonality: The case of tourism in Crete. *Landscape Research, 30*(2), 221–239. https://doi.org/10.1080/01426390500044408.

UNWTO. (2010). *International recomendations for tourism statistics 2008*. New York: Available at. https://unstats.un.org/unsd/publication/Seriesm/SeriesM_83rev1e.pdf.

UNWTO. (2017). *Tourism highlights - 2017 edition*. Madrid: Available at. http://www.e-unwto.org/doi/book/10.18111/9789284419029.

Vaz, E., & Campos, A. C. (2013). 'A multi-dasymetric mapping approach for tourism'. *Journal of Spatial and Organizational Dynamics, 1*(4), 265–277. Available at https://ideas.repec.org/a/ris/jspord/0018.html.

van der Veeken, S., et al. (2015). 'Tourism destinations' vulnerability to climate change: Nature-based tourism in Vava'u, the kingdom of Tonga', tourism and hospitality research. *SAGE Publications, 16*(1), 50–71. https://doi.org/10.1177/1467358415611068.

Voltes-Dorta, A., Jiménez, J. L., & Suárez-Alemán, A. (2014). 'An initial investigation into the impact of tourism on local budgets: A comparative analysis of Spanish municipalities'. *Tourism Management, 45*(Supplement C), 124–133. https://doi.org/10.1016/j.tourman.2014.02.016.

Wood, S. A., et al. (2013). 'Using social media to quantify nature-based tourism and recreation'. *Scientific Reports, 3*(2976). https://doi.org/10.1038/srep02976.

World Travel and Tourism Council. (2017). *Travel and tourism economic impact 2017-country reports*. Available at https://www.wttc.org/research/economic-research/economic-impact-analysis/country-reports/.

Yacoumis, J. (1980). 'Tackling seasonality: The case of Sri Lanka. *International Journal of Tourism Management, 1*(2), 84–98. https://doi.org/10.1016/0143-2516(80)90031-6.

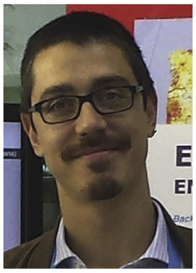
Filipe Batista e Silva graduated in Geography in 2006 and later obtained a MSc in Geographic Information Systems and Spatial Planning, both at University of Porto, Portugal. There, he worked as a research fellow until 2010. He then joined the Joint Research Centre of the European Commission and has since been involved in regional and spatial analyses, land use and population mapping and modelling, integrated modelling and impact assessment of EU policies. He regularly contributes to technical and policy documents and is author of papers in various scientific journals.

Mario Alberto Marín Herrera is a consultant in the field of Geographic Information Systems, Remote Sensing and Urban and Regional Planning. He currently manages large geospatial databases of human activity and infrastructures. He received his MSc in Geographic Information Systems at the University of Edinburgh. His undergraduate studies in Environmental Sciences at the University of Pablo de Olavide have provided him a background in spatial planning and environmental management.

Konštantín Rosina graduated in Human Geography in 2010 and later joined the OECD as a consultant for geographic information systems. Further specializing in spatial analysis and geographic information science, he received PhD at Comenius University in Bratislava in 2015 and worked as a research fellow at the Geographic Institute of Slovak Academy of Sciences until 2016. Currently he is involved in exploratory research in the field of land use mapping and spatial modelling of human population at the Joint Research Centre of the European Commission. He co-authored several technical reports and scientific papers in peer-reviewed journals.

Ricardo Ribeiro Barranco graduated in Environmental Engineering in 2008 in Faro, Portugal. Obtained his MSc in Wageningen University and Research, Netherlands, where he specialized in Geographical Information Systems. Later worked as a consultant at the gas and electricity sector in GISIG association and in the private sector at Iren company, both based at Genoa, Italy. In 2012 joined the Joint Research Centre of the European Commission where he carries spatial analyses focused on cities and urban areas, using Big Data and Machine Learning techniques.

Sérgio Freire holds a Masters in Geography from the University of Kansas (USA). He has previously worked at the National Centre for Geographic Information (Portugal), at the Portuguese Geographic Institute, and at Universidade Nova de Lisboa. Now at the Joint Research Centre of the European Commission, his research focuses on developing applications of the JRC's Global Human Settlement Layer for disaster exposure, risk, and vulnerability analysis, including modelling population distribution at a range of spatial and temporal resolutions. Current activities also include global mapping and characterization of human settlements, and developing satellite-based indicators to support monitoring of Sustainable Development Goals.



Marcello Schiavina graduated in Environmental Engineering in 2008, and later specialized in population dynamics, distribution and migrations receiving his PhD in Quantitative Ecology in 2014. He worked as a research fellow at the Department of Electronics, Bio-engineering and Information Technology of Politecnico di Milano until 2016. He then joined the European Commission Joint Research Centre and has since worked on global spatial analyses, population mapping and modelling and urban/rural dynamics. He regularly contributes to technical reports and he is author of scientific papers in various peer-reviewed journals.FBS designed the study, analysed the results and wrote the manuscript. MAMH undertook data collection and assembly. KR performed data processing and contributed to the writing of the manuscript. RB undertook data collection and assembly. SF contributed to the design of the study and to writing the manuscript. MS assisted with data collection, performed data processing, harmonization, and integration. All authors read and approved the final version of the manuscript.