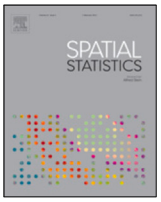
Spatial Statistics 45 (2021) 100535

Contents lists available at ScienceDirect

Spatial Statistics

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



Spatio-temporal clustering: Neighbourhoods based on median seasonal entropy

Miguel Ángel Ruiz Reina

University of Málaga, Department of Theory and Economic History (Staff of Fundamentals), PhD Program in Economics and Business, s/n, Plaza del Ejido, 29013 Málaga, Spain

ARTICLE INFO

Article history:
Received 5 February 2021
Received in revised form 16 July 2021
Accepted 9 August 2021
Available online 24 August 2021

Keywords:
Spatial time series
Seasonal clustering
Entropy
Tourism economics
Neighbourhoods
Information theory

ABSTRACT

In this research, a new uncertainty clustering method has been developed and applied to the spatial time series with seasonality. The new unsupervised grouping method is based on Neighbourhoods and Median Seasonal Entropy. This classification method aims to discover similar behaviours for a time series group and find a dissimilarity measure concerning a reference series *r*. The Neighbourhood's Internal Verification Coefficient criterion makes it possible to measure intra-group similarity. This clustering criterion is flexible for spatial information. Our empirical approach allows us to measure accommodation decisions for tourists who visit Spain and decide to stay either in hotels or in tourist apartments. The results show the existence of dynamic seasonal patterns of behaviour. These insights support the decisions of economic agents.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Time series grouping is based upon data clustering when dynamic behaviour similarities are found. These time series insights can be grouped to help practitioners or researchers make informed decisions. This knowledge can be obtained through patterns based on values, functional forms, correlations, or mathematical characteristics similar to each other (Disegna et al., 2017). The classification of temporal data to obtain knowledge is of interest to researchers from different scientific, computational, health, or environmental disciplines. Since 1960, clustering analysis has tried to solve three main questions: how many clusters there are, what the best clustering algorithm

E-mail address: ruizreina@uma.es.

https://doi.org/10.1016/j.spasta.2021.100535
2211-6753/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

is, and what we should do with the outliers. The main goal of clustering is to find groups with similar characteristics among them, as well as differences among the rest of the data groups studied. In this research, clustering and classification techniques play a central role (Luna-Romera et al., 2018).

The time series study focused on classifying according to difficulty consists in finding similarities—algorithms for clustering are used for this purpose. Thus, different algorithms have been recently developed: connectivity-based clustering (Scotto et al., 2010), centroid-based clustering (Maharaj et al., 2015), distribution-based clustering (Alonso et al., 2006), and density-based clustering (Stuetzle, 2003).

The following relevant approaches for time series clustering and the methodological classification are worth mentioning (Warren Liao, 2005; Caiado et al., 2015; Aghabozorgi et al., 2015; Alonso et al., 2020):

- Time series clustering by features—feature-based in the time domain (Kakizawa et al., 1998; Alonso and Maharaj, 2006; D'Urso and Maharaj, 2009,?; Vilar et al., 2018; Alonso and Peña, 2019; Alonso et al., 2020), the frequency domain (Caiado et al., 2006, 2009; Maharaj and D'Urso, 2010, 2011), or the characteristics of the wavelet decomposition (Durso and Maharaj, 2012; D'Urso et al., 2014).
- Clustering approach based on stochastic time series models (Piccolo, 1990; Maharaj, 1996; Kalpakis et al., 2001; Vilar and Pértega, 2004; Caiado and Crato, 2010; Otranto, 2010; D'Urso et al., 2013, 2016), density function and forecast density (Alonso and Maharaj, 2006; D'Urso et al., 2017), functional approach (James and Sugar, 2003), splines (García-Escudero and Gordaliza, 1999), and, last but not least, copulas, measures of association, and tail dependence (de Luca and Zuccolotto, 2011; Durante et al., 2014; Durante and Sempi, 2015; De Luca and Zuccolotto, 2017; Di Lascio et al., 2017).
- Methods based on time series transformations (Coppi et al., 2010).

Moreover, models being used in copula-based spatial statistics have been developed in recent years (Disegna et al., 2017). Time series copula models applied to economics have been used in Risk Management, Derivate contracts, Portfolio decision problems, Time-varying copula models, and High-dimension copula applications, among others (Patton, 2012, 2013). In the literature, the clustering criteria applied have been developed using the novel COpula-based FUzzy clustering algorithm for Spatial Time series (COFUST), finding applications in other fields. The application of the real case study shows that the COFUST algorithm may help select groups that are both dependent and spatially close, making them more appealing for the cluster analysis results (Disegna et al., 2017).

In our work, we will use the concept of Shannon's Entropy (Shannon, 1948) to make classifications according to the recognition of seasonal patterns. In our context, we can differentiate two types of clustering approaches: (1) supervised, where there is previous information in use; (2) unsupervised, where there is no prior information. In this second approach, we use the Entropy algorithm without prerequisites in our modelling for time series. The aim is to obtain similarities between the classification groups and their subsequent analysis (Aldana-Bobadilla and Kuri-Morales, 2015). The clustering criteria have been proposed based on sets of k centres, as well as on the closest distances. Examples of these are the k-means (MacQueen, 1967) or the fuzzy c-means (Bezdek, 1981; Dunn, 1973; Sripada, 2011). We can mention other methods used as (1) hierarchical clustering methods (Zhang et al., 1996; Ruppert, 2004), density clustering methods (Ester et al., 1996), and (3) meta-heuristic clustering methods (Caruana et al., 2006; Das et al., 2009; Aldana-Bobadilla and Kuri-Morales, 2015).

In the field of Data Mining, comparison clustering between Entropy measures based on the well-known K-means clustering and fuzzy C means clustering has been used. The studies' conclusions indicate that the lower the Entropy, the better the clustering (Sripada, 2011). Considering the above information, we suggest a clustering model based on uncertainty measurement through the Shannon's Entropy algorithm (Delgado-Bonal and Marshak, 2019).

In particular, we propose a new method which classifies time series according to repetitive cyclical flows, based on the distance to a reference series. We measure distances across the Neighbourhoods Seasonal Entropy based on Median Seasonal Clustering Entropy ($MdSCEN_{srjt}$). We

also introduce an internal validation criterion of the cluster called "coefficient of internal verification of the neighbourhood"—thus, we can quantify the similarities between the seasonal Entropy time series. This contribution allows knowing the differences between the dynamic behaviour of an aggregate series and the different series that compose it. The intention is to determine how far away an Entropy series is compared to the general reference series, and later classify it according to disorder with non-Gaussian characteristics. The immediate result is to find the most ordered sequence in the sense of Entropy, and later organise its constituents in groups according to their distance to the reference series. Besides, it allows us to know the distance of the reference series at each moment of the cycle—therefore, we can classify situations of uncertainty depending on the time cycle point. Research related to uncertainty from probabilistic models and decision theory is extracted from the scientific literature (Ruiz-Reina, 2019). This Clustering System contributes to Thermoeconomics and Physical Statistics with application to Economics in Information Theory in relation to dynamic time series structures.

The motivation of the technique and its empirical application. The tourism sector has experienced growth worldwide in recent years, except for the COVID-19 crisis. This crisis has meant the loss of 850 million to 1.1 billion international tourists—at the same time, it has put between 100 and 120 million jobs at risk (UNWTO, 2021). For our empirical section, we will develop the theoretical model on tourist accommodation for data from Spain. The contribution of tourism reached 154,487 million euros in 2019, representing 12.4% of Spanish GDP. According to official statistics, before the global health crisis, the industries related to tourism generated 2.72 million jobs—12.9% of total employment. All tourists who spend the night at their destination must make accommodation decisions, meaning this is a business opportunity for the destination markets characterised by networks o clusters (Zhang et al., 2009). The final decision is quantified and modelled so that economic agents can predict future seasonal behaviour. The dichotomy in accommodation – between tourist apartments and hotels – by country of origin provides insight for primary and secondary market agents (Vlachos and Bogdanovic, 2013; Adamiak, 2018). In this sense, the offer can be adjusted to the demand, creating efficiency in the tourist markets.

In this paper, our theoretical model will be applied empirically to the tourist accommodation market in Spain. Specifically, we will answer questions about seasonal grouping in the decision-making of economic agents. This model is an unsupervised dynamic cluster development improving the static analysis that has traditionally been carried out in the tourism industry (Yong-Jin et al., 2020). In Economics, clusters are defined as geographic concentrations of interconnected companies, specialised suppliers and consumers, and associated institutions (Porter, 1998). The contribution of the decomposition of tourist accommodation demand into seasonal and spatial groups based on similarity/dissimilarity represents a solution to a non-Gaussian clustering problem and is a development related to seasonal clustering techniques (Inniss, 2006). With clustering, we seek to find common spatial behaviour patterns – by country of origin – according to seasonality criteria for choosing tourist accommodation in hotels or apartments from Jan. 2005 to Aug. 2019. The following section called "Theoretical Analysis framework" will outline a theoretical development for its application in the Tourism field, with its subsequent practical application in the empirical section. The results of the analysis presented in this paper provide economic agents with useful knowledge for the sake of better decision-making. In general, this approach can be applied to non-Gaussian spatio-temporal seasonal data sequences.

The remainder of this investigation is as follows: Section 2 provides new theoretical development of unsupervised clustering based on Entropy algorithms, whereas Section 3 is dedicated to the empirical area—in particular, we apply and develop a theoretical framework to a case study measuring and grouping accommodation in Spain according to nationalities that visit this country as a tourist destination. Once the applied empirical results have been obtained, Section 4 discusses theoretical and practical implications for tourism, and Section 5 discusses the main conclusions reached after using the proposed methods. We also recommend new lines of research for this clustering system. Finally, bibliographic references are listed.

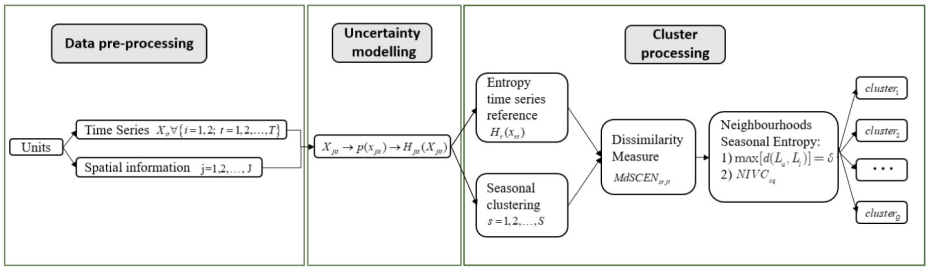


Fig. 1. Measurement of seasonal spatial uncertainty based on Entropy. Three steps: (1) data pre-processing, (2) uncertainty modelling, and (3) cluster processing. The sample period for empirical results goes from Jan. 2005 to Aug. 2019.

2. Methodology

Unsupervised learning is the usual method when there are no predefined processes. In our work, the treatment of groups is hierarchical where each starting point is a cluster, and then they are successively included to create the clustering structure, also known as the agglomerative method (Inniss, 2006). In this methodological section, we will describe the consecutive steps for classifying information based on Entropy. In Fig. 1, we outline the three steps that are subdivided into activities: (1) data pre-processing, (2) uncertainty modelling, and (3) cluster processing.

This clustering process will allow us to obtain knowledge about clusters concerning seasonal spatial time series. The first step contains information on the temporal variables under study, as well as on spatial data. In this context, every time series is a unique cluster. In the second step, the series are treated, and the uncertainty measures are obtained—we will work with Shannon's Entropy concept (Shannon, 1948) in particular. In the third step, we find the contribution of this paper—the previously organised data are clustered based on spatial, temporal, and proximity criteria according to a reference series. Finally, we will obtain the clusters for their treatment and interpretation in the decision-making process. This last step allows science to be provided with knowledge for the particular treatment of the groups that have finally been chosen. At the same time, it allows to make more efficient decisions. In the next subsections, we will develop the phases of seasonal time–space clustering in greater detail.

2.1. Data pre-processing

To obtain consistent information among the grouping of time series for x_{jit} $\forall \{j = 1, 2, \ldots, J; i = 1, 2; t = 1, 2, \ldots, T\}$. The series are column vectors for each period t in such a way that we find $x_{1t} = [x_{11}, x_{12}, \ldots, x_{1T}]^T$y $x_{2t} = [x_{21}, x_{22}, \ldots, x_{2T}]^T$. From a theoretical point of view, we are interested in measuring the influence between the variables $x_{1t} \rightleftarrows x_{2t}$. Causal relationships may create this bidirectional relationship, so selecting one variable conditions the non-choice of the other. In the uncertainty model, we will explain this choice formally.

On the other hand, we have spatial information to examine. We will give the name the j to the spaces to analyse: $\forall\, j = 1, 2, \ldots, J$. For the purposes of this research, we are interested in investigating two time series for the tourist accommodation market in Spain—that is to say, the demand for both hotels and tourist apartments. Spatially, our attention will come from the countries of origin of the tourists who visit Spain during the period under review. Finally, in a generic notation, we will denote the space–time series as follows: x_{jit}.

2.2. Uncertainty modelling

The starting point is the definition of the random variables of interest x_{jit}. To do this, we will define the proportion measures in this way: $p(x_{j1t}) + p(x_{j2t}) = 1$, where it is verified $p(x_{j1t}) =$

$\left(x_{j1t}\right)/\left(x_{j1t}+x_{j2t}\right)$ and $p(x_{j2t})=1-p(x_{j1t})$. The Entropy time series, whose information is denoted, can be expressed in discrete time series as follows (Delgado-Bonal and Marshak, 2019):

$$H_{jt}(X_{jt})=-\sum_{i=1}^{2}p(x_{jit})\log_2 p(x_{jit}) \tag{1}$$

Our article will use the logarithm in base two expressed in bits—however, it is possible to apply other logarithmic bases. For example, the literature indicates the possibility of using natural logarithms. The measurement of Entropy among time series pays attention to the points of maximum and minimum Entropy (Ruiz-Reina, 2019). In this way, we will find the most significant uncertainty when $p(x_{j2t})=p(x_{j1t})=0.5$ and the least uncertainty when $p(x_{j1t})=1$ or $p(x_{j2t})=1$ (see Appendix A). Finally, the Entropy matrix for a spatial series j is defined as follows:

$$p(x_{jit})=\begin{pmatrix}p(x_{j11}) & p(x_{j21})\\ p(x_{j12}) & p(x_{j22})\\ p(x_{j13}) & p(x_{j23})\\ \vdots & \vdots\\ p(x_{j1T}) & p(x_{j2T})\end{pmatrix}\quad H_{jt}(X_{jt})=\begin{pmatrix}-p(x_{j11})\log_2 p(x_{j11})-p(x_{j21})\log_2 p(x_{j21})\\ -p(x_{j12})\log_2 p(x_{j12})-p(x_{j22})\log_2 p(x_{j22})\\ -p(x_{j13})\log_2 p(x_{j13})-p(x_{j23})\log_2 p(x_{j23})\\ \vdots\\ -p(x_{j1T})\log_2 p(x_{j1T})-p(x_{j2T})\log_2 p(x_{j2T})\end{pmatrix} \tag{2}$$

Entropy analysis allows us to obtain order information based on the series values and the dimensionless relationship. Prior knowledge from the data set is frequently unknown—cluster ordering allows grasping disaggregated information from the clustering series for spatial data. Indeed, its use has been widely recognised (Zhang et al., 1996). In our empirical analysis, we have already indicated that we are analysing two types of time series—the decision to stay either in a hotel or in a tourist apartment for the nationalities who were visiting Spain. The Entropy algorithm's measurement allows us to know and measure the uncertainty of what has occurred historically. Thus, when $H_{jt}(X_{jt})=0$, we have total certainty that tourists of the nationality j stay in hotels when $H_{jt}(X_{jt})\simeq 1$, tourists of the nationality j are indifferent between staying in hotels or apartments (Ruiz-Reina, 2019). Once the uncertainty measurement algorithms have been defined, we proceed to the spatio-temporal classification with similar characteristics.

2.3. Cluster processing

The previous phases of data pre-processing and uncertainty modelling provide a framework for the clustering process. The grouping process allows us to reduce the information for analysis. This last phase of cluster processing is divided into sequential parts—part 1, the reference series and seasonal grouping are determined; part 2, the dissimilarity is measured; part 3, the grouping criteria are defined, and part 4 is the final phase of clustering and group interpretation.

Part 1: Determination of the reference series and seasonal clustering. In unsupervised clustering, we must take a series of reference Entropy $H_r(x_{rt})$, since it is a dimensionless concept and it is a "reference" for the rest of the series of Entropy to be analysed. For example, in the study of Spanish national tourism, we can build our model according to the aggregate behaviour patterns for each series to be explored. Seasonal grouping makes it possible to simplify common cyclical behaviour patterns with a periodicity of less than one year. In this way, the seasonal interpretation of the series fluctuations is expected in the Tourism field (Baron, 1984). For our empirical work, the reference series $H_r(x_{rt})$ will be Spain's total Entropy to compare with the data disaggregated by nationalities j of tourists visiting Spain. From the seasonal point of view for data with monthly periodicity, we will regroup the data in the twelve months of the year $s\in[1,12]$. Therefore, after the seasonal grouping, the series will be called $H_{sj}(x_{jt})$, and for the seasonal reference series, $H_{sr}(x_{rt})$. This simplification of data every month allows analysing the seasonal behaviours of the Entropy series.

Part 2: Dissimilarity Measure. The concept described here is the Median Seasonal Clustering Entropy ($MdSCEN_{srjt}$), and it measures the Euclidean distance between the seasonal Entropy series j

and the seasonal reference series. Mathematically, the measure of seasonal difference according to the j series is defined as follows:

$$MdSCEN_{srjt} = Median\left(\sqrt{\left(H_{sr}(x_{rt}) - H_{sj}(x_{jt})\right)^2}\right) \tag{3}$$

The $MdSCEN_{srjt}$ presents two characteristics of grouping: (a) it is not affected by extreme values, either seasonal variability or atypical values; (b) series with seasonality usually show extreme values in practice, so the median is preferable to the concept of mathematical expectation. We will group by monthly data for our empirical development and calculate the values of the series j compared with the reference series r.

Part 3: Neighbourhoods Seasonal Entropy. Once we have information about the distances to the reference series, we proceed to order the uncertainty measures based on past due dates by proximity to the reference series. The number of groups has been widely discussed in the literature, and there have been contributions based on the minimum descriptive distance (Lafuente-Rego and Vilar, 2016). In our model, we must include an *ad-hoc* criterion, and subsequently establish an internal evaluation criterion for the cluster:

- The *ad-hoc* criterion refers to the distance of the maximum Entropy interval of the cluster. The maximum distance value is constant δ—in our case, we decided $\max[d(L_u, L_l)] = 0.05$, where d is the distance between two Entropy series, the upper (u) and lower (l) limits are (L_u, L_l).
- The following expression will determine the Neighbourhood's Internal Verification Coefficient:

$$NIVC_{sq} = \frac{\max[d(MdSCEN^{u}_{srjt}, MdSCEN^{l}_{srjt})]}{\max[d(L_u, L_l)]} \tag{4}$$

  Verifying $0 \leq NIVC_{sq} \leq 1$, the most remarkable intra-group similarity will occur when $NIVC_q = 0$, otherwise we will find greater diversity with $NIVC_q = 1$. Under the internal verification criterion, there may be limit values that meet the conditions—however, we consider this study a valid tool to establish clusters and eliminate acyclicity so as to make objective decisions.

This clustering $NIVC_{sq}$ allows the series to be grouped objectively based on seasonal Entropy. Likewise, it is possible to identify the outliers of interest in different fields of research. In our case, we will use a definition that is easy to interpret, and it is the use of divided intervals in neighbourhoods for non-grouped data. There are optimisation arguments and more in-depth readings in the literature, but it is not within the scope of this paper (Honarkhah and Caers, 2010).

Part 4: Clustering and interpretation. Finally, the neighbourhood-based clusters are obtained: $cluster_q = cluster_1, cluster_2,\ldots, cluster_Q$. This last phase allows us to recognise common patterns of uncertainty. This way, we can study the internal similarity among the groups and analyse their differences. In our case, the intention is to establish a cluster of behaviour in demand for tourist accommodation based upon the nationalities that visit Spain.

In sum, we have proposed a robust cluster analysis. With this methodology, it is possible to establish similarities and differences concerning a reference Entropy series. The criterion based on medians allows overcoming statistical problems based on outliers that could distort the cluster's seasonal interpretation. This methodology proves its value so as to understand the spatial uncertainty between two time series for different series j. In the next section, we will study the real use of this analysis method in uncertainty models applied to the tourism sector.

3. A case study: Entropy in decision-making regarding the tourist demand for Spanish accommodation

This third section of this article will deal with the empirical application of the previously described methodology. The subdivision is as follows: Section 3.1 briefly explains the theory and reasons why this type of grouping applies for tourism markets; subsection number 3.2 describes the empirical development with data applied to an economy such as Spain, with a tourism industry representing around 12 percent of the Gross Domestic Product and international repercussions for the rest of economies.

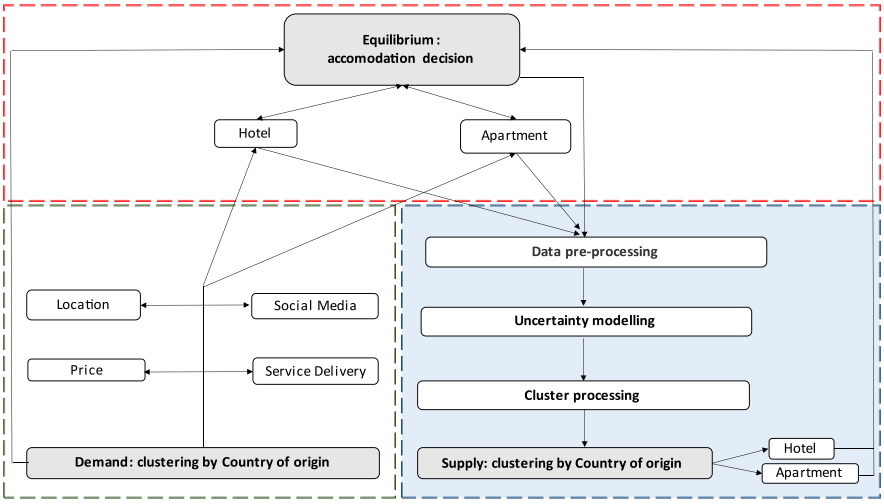


Fig. 2. Measurement of spatial uncertainty based on Entropy. Demand, supply, and equilibrium in the tourist market for accommodation in hotels vs. apartments.

3.1. Theoretical analysis framework for tourism demand

Not infrequently decisions are made in situations of uncertainty—measuring randomness between time series is relevant to the economic agents. Certain Environmental, Physical, Social, or other Scientific areas present uncertainties that must be measured and classified. Their measurement allows to know the core of the phenomenon to be analysed, and the classification allows the information to be grouped so as to discover common characteristics of the phenomena, which simplify the reality to be diagnosed (Delgado-Bonal and Marshak, 2019). Learning about processes implies gains in computational efficiency in the study area of interest. We will develop a criterion for measuring uncertainty and seasonal spatial clustering with applications in different origin regions in our work. In particular, we will create the working scheme of Fig. 2, where we will deal with decision-making under uncertainty for tourism accommodation markets. We are interested in measuring temporary accommodation used by tourists visiting Spain between hotels and tourist apartments.

The core of Fig. 2 is to cluster measurements of spatial uncertainties (by country of origin) and thus make decisions efficiently for hotels or apartments. To this effect, the study is based on three large blocks: equilibrium, accommodation decision, and demand and supply, in relation to tourists from a particular country of origin. Some variables, such as location, price, social media, and service delivery determine the demand-side accommodation decision. These variables are the core of decision-making, and based on this, uncertainty demand for tourist accommodation must be measured. Once time–space values are known, we proceed to the modelling of the uncertainty. The steps are as follows: first, demand data by country of origin are collected and chronologically ordered. Second, the data are organised to measure the uncertainty under the concept of Entropy—the core is modelling the decision to book a hotel or apartment accommodation. Third, after collecting the data, we proceed to data processing and distance measurement, taking a reference series under the criteria of the researcher (in this work we propose a new method). Finally, a clustering criterion – the similarity among series – is established.

The steps described in the clustering process allow a truthful and interpretable approach to the decision-making process by the market agents. In this way, we can find equilibrium in the market based on spatio-temporal clustering criteria.

The goal is to understand the decision-making behaviour the economic agents show in uncertainty situations, and this article previously develops a theoretical framework based on the scheme

in Fig. 1. This clustering system allows grouping the seasonal spatio-temporal information, whereas the modelling enables efficient decision-making. In the methodological section, we will develop the theoretical analysis, and later we will observe the empirical results with the application to the tourist accommodation market in Spain by means of the study of 20 nationalities that visit Spain.

The flows of international tourists with Spain's destination imply socio-economic relationships creating information networks that provide value to public and private economic agents (Reina, 2020). Consequently, the time–space analysis is crucial in promoting the tourism sector in the countries of origin. The agglomeration of tourist groups and their research promote the destination country's economic activity (Disegna et al., 2017).

Here we make use of the Entropy-based cluster analysis to measure uncertainty, finding typical/similar behaviour among tourist flows according to the country of origin for the period analysed. Specifically, given the information on the demand for tourist accommodation in the past, we are interested in modelling the uncertainty in decision-making for tourist accommodation between hotels and apartments. We focus on identifying seasonal behaviours among the nationalities that visit Spain each month. In this way, we find the opportunity to (1) identify seasonal behaviour of the demand for accommodation; (2) recognise the nationalities that demand each type of accommodation (by country of origin), and (3) find clusters of foreign nationalities with seasonal demand patterns. To do this, we will apply the methodology outlined in the previous sections with the empirical results set out in the following subsection.

3.2. Empirical results

The empirical results show a work scheme described in the method of Fig. 1. Next, we will detail the data pre-processing, uncertainty modelling, and cluster processing for the sample period from January 2005 to August 2019.

Data pre-processing. In this work, the temporal database used comes from the Spanish National Institute of Statistics (INE[1])—it contains data from overnight stays in hotels and tourist apartments according to the country of origin from January 2005 to August 2019. 20 countries of origin have been studied, with four large geographical groupings (Other European Countries, Rest of the EU, Rest of the world, Africa). The tourists who visit Spain during the selected time span are from Germany, Austria, Belgium, Denmark, USA, Finland, France, Greece, Ireland, Italy, Luxembourg, Norway, Netherlands, Portugal, United Kingdom (UK), and Sweden. According to the INE, the nationalities that generate a greater flow of tourists entering and leaving Spain are the UK and Germany. The distances of each nationality j concerning the total number of overnight stays in Spain will be the reference measure r (denoted as Total in the rest of the document).

The individual time series of overnight stays shows the seasonal, trend, and cycle components (Reina, 2020). Considering the diagram in Fig. 2, we will describe the modelling of uncertainty in the next section.

Uncertainty modelling. Fig. 3 represents the spatio-temporal uncertainty modelling for the $j =$ *Total, Germany, France, Italy, Netherlands, and United* series analysed. We denote hotel accommodations as $i =$ *hotels, apartments* for $t =$ *Jan 2005, …, Aug 2019*. We model the decision of hotel accommodation vs. apartment for the 20 series of nationalities of origin that are studied here (see Appendix B). The Entropy functions results show a cyclical behaviour with a periodicity less than one year, compatible with the seasonality of the time series for the nationalities of origin described above. Based on our study, we have selected the reference series r as the one called Total. This series shows the global behaviour of all series and the measurement/comparison of this series with the rest will allow us to cluster the information before decision-making.

Cluster processing. In the previous sections, we indicated that tourist accommodation time series shows seasonal behaviour s. The factors that dictate seasonal fluctuations are determined by both

[1] INE: Instituto Nacional De Estadística.
Data downloaded by country of origin for hotels (https://www.ine.es/jaxiT3/Tabla.htm?t=2038) and apartments (https://www.ine.es/jaxiT3/Tabla.htm?t=1998&L=0).

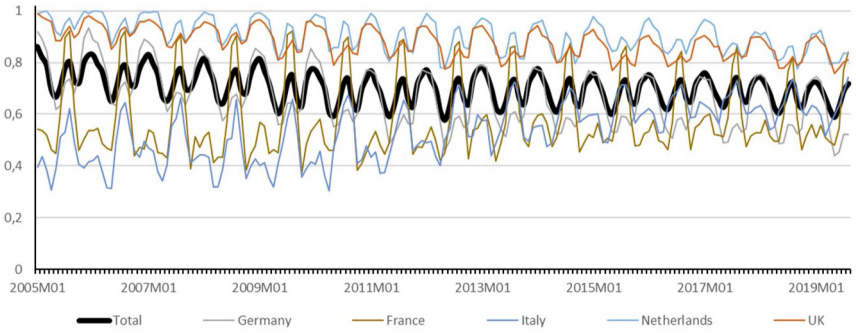


Fig. 3. Measurement of spatial uncertainty based on Entropy from Jan. 2005 to Aug. 2019. Here the reference series r is called Total. The figure represents the Entropy series of the main nationalities of tourists visiting Spain (Germany, France, Italy, Netherlands, and United Kingdom) for the period analysed. In Appendix B, we study the Entropy series clustered according to the methodology described in this paper (j = *Total, Africa, …, Rest of the World*).

the weather and the calendar (Lim et al., 2010). A consequence of these effects on decision-making is the cyclical fluctuations in accommodation decision-making estimated by the Entropy function of the time series. The set of seasonal changes in the previous section gives an idea of how different they are. The series of Fig. 3 show common patterns of intra-annual behaviour. For this reason, taking the so-called *Total* as the reference series, we indicate that we group seasonally the values s = *January, February, March, … December*. Considering $MdSCEN_{srjt}$ as a dissimilarity measure, we obtain the medians of total differences, and the seasonal grouping allows the classification of the series.

The results of the methodology show a dynamic differentiated conglomeration. For a straightforward interpretation of the results, the spatial elements (nationalities that visit Spain) are shown in Figs. 4 and 5. We can observe different dynamics in the clusters formed based on Eq. (2). The graphs show inter-group and intra-group differences. In Fig. 4, we can see eight groups for January and February, whereas the months of April and June have 7 groups, and March and May consist of six groups. One aspect to highlight is that the seasonal presence in a cluster is not static, a clear example being the case of Germany. For the first six months of the year, it presents a dynamic behaviour, moving away from the reference series. In January, it is the closest nationality to the reference series, while in June it belongs to the third cluster closest to the reference series. For the second part of the year (Fig. 5), in July, it is in the second cluster, and in December, it is the closest to the reference series.

As expected, the dynamic seasonal classification allows differentiating behaviours and grouping them satisfactorily. Focusing on a reference series, we can state that different groupings are created. From an administrative perspective, this partition allows us to classify nationalities according to their decision-making ability from an objective perspective for the sake of analysis. In the next discussion sections with theoretical and practical implications, we will expose this tool's usability to make data-driven decisions in a digital environment. This will always be done under the central hypothesis of groupings: intra-group similarity and dissimilarity between groups. The robust criteria developed in this article illustrate measurable behaviours that create value for public or private institutions.

The intragroup variability measures in Table 1 reveal relevant details—the seasonal clusters present more significant intra-group dissimilarity in group 4 (except in April and November), cluster 5 (except May and July), and cluster 6 (except February, May, and August). For the rest, it is worth highlighting more remarkable similarity in clusters 1, 2, and 3. The number of groups finally varies among 6 (March, May, and September), 7 (January, February, April, June to August, and October to December), and 8 clusters (January, February, and November). Given the results obtained in Table 1, we can say that the groups with the most remarkable diversity are numbers 4 and 5.

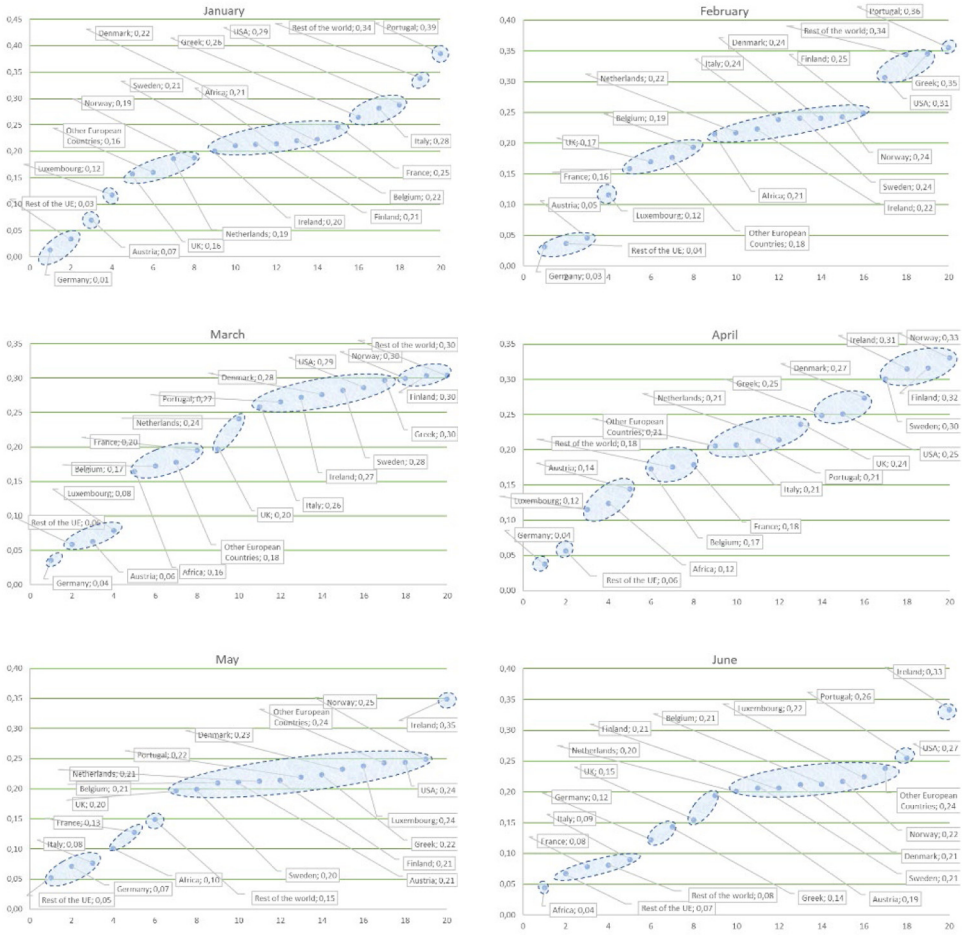


Fig. 4. Clustering seasonal and spatial Entropy based on the differences with the series r between Jan. and Jun. The X-axis represents the order of distance concerning the reference series. The Y-axis represents the distance to the reference series. In position 0, the series r will always appear because $MdSCENsrjt = 0$.

From Table 1, we see that the number of clusters is dynamic. This robust cluster analysis allows us to analyse the advantages and disadvantages of clustering seasonal time series. Among the advantages, it is worth highlighting the possibility of identifying seasonal behaviour of the series and seasonal groupings according to analysis period, as well as scalability and speed with large volumes of data. Among the disadvantages, we can mention the fact that the seasonal collection reveals historical behaviours, and the *ad-hoc* criterion is a creation of the researcher. The latter can be considered an advantage under certain types of data.

To summarise this section, we can argue that accommodation decisions between hotels and apartments by foreigners visiting Spain are dynamic and can be grouped concerning a reference series r. This grouping allows us to establish supply techniques for the demand clusters in relation to tourism markets, according to Fig. 2. These strategies can improve efficient tourist market allocations and management policies for tourism development in the analysed area.

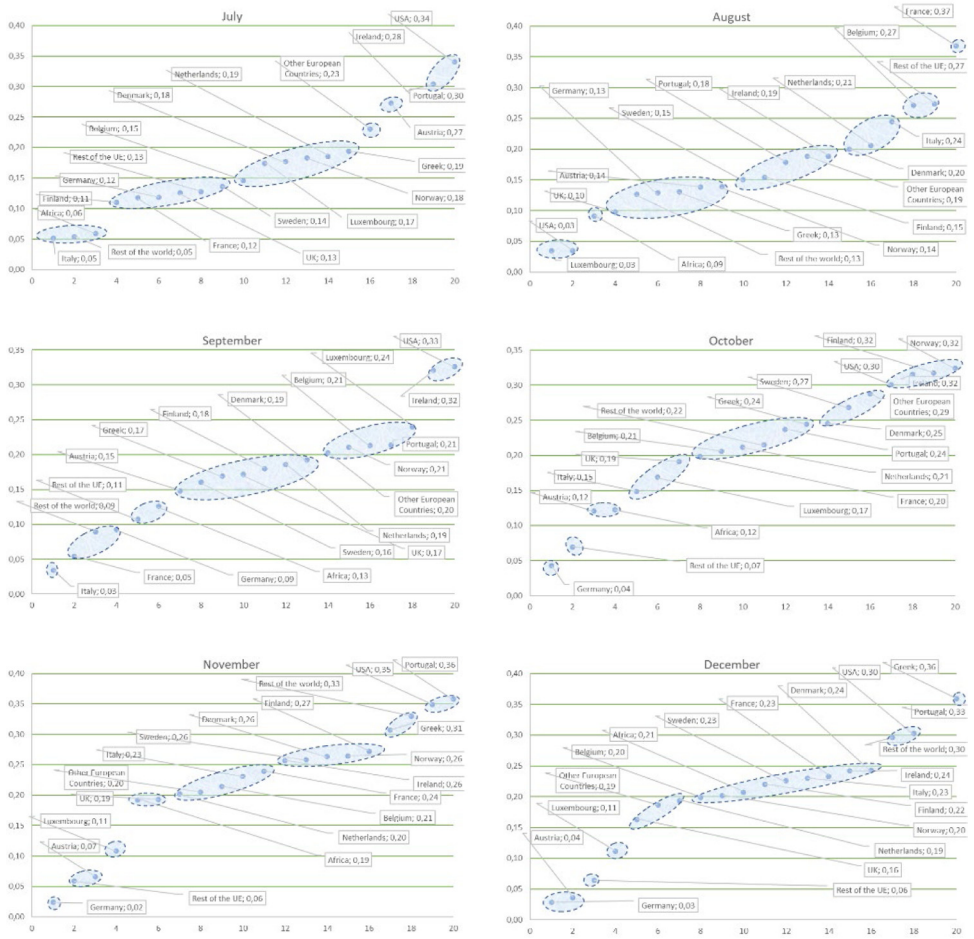


Fig. 5. Clustering of seasonal and spatial Entropy based on the differences with the series r between Jul. and Dec. The X-axis represents the order of distance concerning the reference series. The Y-axis represents the distance with the reference series. In position 0, the series r will always appear because $MdSCENsrjt = 0$.

4. Discussions: theoretical and practical implications in tourism

This study contributes to understanding the importance of grouping tourists in relation to the spatio-temporal hotel or tourist apartments accommodation for the sake of unsupervised decision. Initially, we have observed there is a direct relationship between the two types of accommodation, and we have modelled clusters according to the country of origin of foreign tourists visiting Spain. In this article, by means of the methodology proposed, cyclical behaviours have been identified for the 20 nationalities that visited Spain between January 2005 and August 2019. These results are relevant for researchers and practitioners to use geospatial data and analytical techniques whenever they wish, implement offers, and develop accommodation strategies for tourists. Besides, knowledge about accommodation and clustering allows recognising customer profiles for primary or secondary businesses related to tourism, such as transport, vehicle rental, and museums, among others (Dredge, 1999).

Empirically, we have formally demonstrated different types of demand depending on the months of the year, assuming a contribution to subsequent studies. According to the country of residence,

Table 1
Neighbourhood's Internal Verification Coefficient for the clusters found in the treatment process according to nationalities that visit Spain in the intra-annual period from January to December. Values close to zero indicate maximum similarity among the cluster members—when they are close to one; it indicates maximum dissimilarity.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
January	0.41	0.00	0.00	0.63	0.44	0.44	0.00	0.00
February	0.11	0.00	0.00	0.70	0.57	0.00	0.74	0.19
March	0.00	0.40	0.27	0.92	0.56	0.15	–	–
April	0.00	0.00	0.58	0.12	0.62	0.49	0.60	–
May	0.00	0.10	0.00	0.94	0.00	0.00	–	–
June	0.00	0.47	0.43	0.79	0.74	0.23	0.00	–
July	0.05	0.00	0.52	0.97	0.00	0.18	0.73	–
August	0.01	0.00	0.82	0.76	0.89	0.05	0.00	–
September	0.00	0.77	0.38	0.88	0.76	0.10	–	–
October	0.00	0.00	0.05	0.85	0.91	0.84	0.46	–
November	0.00	0.13	0.00	0.02	0.75	0.31	0.45	0.19
December	0.14	0.00	0.00	0.60	0.89	0.75	0.00	–

behaviours show similarities to neighbouring countries, finding intra-group similarities with the Neighbourhood's Internal Verification.

Theoretical implications in Tourism. This study reveals some implications for geostatistical research and the clustered tourist accommodation market based on the empirical results. The results provide evidence on the spatial relational and seasonal heterogeneity between the tourism industry and the tourist accommodation market, contributing to the tourism theory in the time series cluster (Michael, 2003). The empirical results obtained from the proposed methodology reveal the importance of this information for the economies of both origin and destination in relation to the industry (Yang and Fik, 2014). To measure spatio-temporal relationships, we have developed the Clustering of Seasonal and Spatial Entropy, finding patterns according to the country of origin and the seasonal pattern. The results reveal that differences in accommodation decisions (hotels vs. apartments) among nationalities and seasons of the year highlight the need for more research to identify both the individual and the combined effects of tourism clusters on the performance of accommodation provider (Peiró-Signes et al., 2015).

The specific location findings relate the dynamic clusters of seasonal behaviours relevant to the industry, such as hiring particular profiles to provide services to the visiting nationalities depending on the month. The intra-cluster and neighbourhood tools make a relevant contribution (Majewska, 2015). The Neighbourhood's Internal Verification Coefficient algorithm quantitatively measures the dynamic seasonal similarity between the accommodation decisions that make up the group for theoretical or practical strategies. Finally, we can say that this article's academic contribution contributes to the mapping and understanding to find the supply–demand equilibrium (Fig. 2) of international tourist accommodation markets and secondary industries (Vlachos and Bogdanovic, 2013; Adamiak, 2018).

Practical implications in Tourism. The unsupervised methodology developed in this article has important implications for tourism practitioners. For the tourism accommodation industry, this study suggests that accommodation decisions between hotels and apartments have a complete advantage in tourism clusters both in seasonal analysis and in neighbourhoods that optimise decision-making. Specifically, the existence of a dichotomous accommodation decision is not fixed, but dynamic and seasonal according to the geolocation of origin, which means adjusting the equilibrium (Fig. 2) of demand and supply in the destination market (Jackson and Murphy, 2006). The knowledge about dynamic seasonal decisions made by the groups allows quantifying and measuring actions with potential users (Yong-Jin et al., 2020). This classification system can improve the user experience under the concept of Next Best Activity that provides consumers with products or services depending on their seasonal behaviour.

Furthermore, this study's new empirical results are relevant due to their seasonal dynamics and changes on clustering groups. Let us highlight the two most relevant nationalities according

to the official sources of the INE—Germany and the United Kingdom. For the first four months of the year, Germany's data show a very similar behaviour to the reference series, so it is part of the first cluster. During May, June, July, and August, the seasonal action moves away concerning the reference series for the final months of the year, showing a similar behaviour to that of the beginning of the year. This behaviour reflects that German tourists move away from other nationalities' usual behaviour during the high season in the Spanish tourist market. Tourists from the UK show a similar pattern throughout the year, usually belonging to the third or fourth cluster furthest from the reference series, so their behaviour is dynamic, but we can consider it stable non-Gaussian. In contrast, countries such as the USA present a very dissimilar behaviour to the reference series r, perhaps determined by the distance from the destination country (Spain)—likewise, countries closer to Spain such as Portugal also present differences despite their proximity. The implications of knowing about these seasonal tourist flows allow exploiting infrastructures, roads, museums, services, or any other consumption that tourists make according to origin and time according to seasonal patterns (Ashworth and Page, 2011).

From the tourism policy perspective, this study shows how through the methodological analysis and its implementation of seasonal clustering based on uncertainty in decision-making by country of origin. These space–time interactions of neighbouring groups allow us to understand the tourist accommodation market as a complement to the traditional tourism analysis for either public or private agents, meaning one more tool in the data-driven business.

5. Conclusions

In general, clustering methods refer to the grouping of similar objects, the classification of unobservable elements of decision uncertainty minimises the risk and provides knowledge about decision-making. In this paper, a new seasonal spatio-temporal clustering process based on Entropy measurement has been proposed. The process requires several phases for the grouping-pairing of seasonal time series described in the methodological section (Fig. 2). The goal is to identify entropy behaviour patterns between the series by means of spatial information and a reference series. Cluster processing has been adopted to establish objective criteria for the organisation of seasonal spatio-temporal information. The advantage of using this procedure is that we can construct knowledge objectively, not using fictitious relationships or biased subjective criteria. The seasonal clustering problem can be solved efficiently. Furthermore, the coefficient of internal verification of neighbourhood used allows verifying intra-group similarities, meaning a measure of the local benefit that grouping provides to treat similar behaviour data. The initial hypothesis of cluster based on Entropy and its subsequent development is a dimensionless tool of advantage compared to criteria based on measurement units.

In this work, we have applied the process to real tourist accommodation data in Spain to demonstrate how it works in real cases. In particular, we have classified decision making by economic agents according to their country of origin and their seasonal accommodation decision (hotels vs. apartments). The Entropy cluster analysis allows the initially disordered uncertainty classification as observed in Fig. 3. The exposed method has allowed us to obtain a seasonal dynamic classification of the behaviour in the field of study, reflected in Figs. 4 and 5. The seasonal grouping reveals similarities in the decision-making by economic agents.

The application of this process proves the grouping method suggested for the spatio-temporal series is both useful and effective. In particular, the proximity to the reference series r tends to organise the truth clusters compared to the rest of the nationalities. Furthermore, the inclusion of seasonal information in a classic spatial context allows us to identify similarities among the series analysed. The dissimilarity measures based on the Median play an essential role in the seasonal classification of the extreme values in the series. The empirical section results respond to non-dynamic limitations of dynamic clustering of the literature in the tourism sector (Yong-Jin et al., 2020).

Limitations and future research directions. The application of the unsupervised clustering process exposed in this paper allows the use and application in a grouping of spatio-temporal series with patterns of seasonality. The study is developed in a destination area with more than 20 nationalities

– or groups studied – and new geographical regions can use this clustering methodology. However, recent technical research lines remain open. First, this article has worked with Entropy between two sequences of decisions. We consider the theoretical/practical application of multinomial series or the use of binary distributions (Bernoulli). Second, in our work, we have developed the criterion based on monthly seasonality. This could be applied to weekly, or quarterly seasonality, for instance, depending on the nature of the available data. Third, the seasonality with repetitive patterns studied allows a grouping in twelve months, giving rise to an *ad-hoc* criterion of interval grouping. We encourage future researchers to explore the boundaries between clusters and the possibility that elements can appear simultaneously in two groups. The Neighbourhood Internal Verification Coefficient criterion can verify the benefits of this new approach, resulting in classification algorithms. Finally, numerous real cases in which seasonal and spatio-temporal clustering algorithms are applied – that is, any analysis with time series data volumes – are of interest to researchers and practitioners for Transport, Electricity Markets, Operation Research, or Finance, among others.

Acknowledgements

The author wishes to acknowledge the support given by the University of Malaga. PhD. Program in Economics and Business, effective from July 16, 2013. Especially to Professor Antonio Caparrós Ruiz from the Department of Statistics and Econometrics of the University of Málaga, for reviewing this work. The author acknowledges the anonymous peer reviewers' valuable comments that significantly improved the quality of this paper. This research is associated with the group of Faculty of Economic and Business Sciences at the University of Malaga: "Social Indicators-SEJ157". The research group has funded the professional editing service in English. Research Funders: "Funding for open access charge: Universidad de Málaga/CBUA".

Appendix A

In this section of Appendix A, we will demonstrate the maximum Entropy situation given the definition of methodological Section 2.2. Information Theory: Shannon's Entropy. The problem to solve mathematical optimisation is as follows: ($p_i \in R; i = 1, 2$):

$$\arg\max f(p_i) = \arg\max H(p_i) = \sum_{i=1}^{2} p_i \log_2\left(\frac{1}{p_i}\right) \tag{A.1}$$

$$\text{subject to} \sum_{i=1}^{2} p_i = 1 \tag{A.2}$$

We propose the Lagrangian as follows (where λ represents the Lagrange multiplier):

$$L(\mathrm{H}(p_i)) = \sum_{i=1}^{2} p_i \log_2\left(\frac{1}{p_i}\right) - \lambda \sum_{i=1}^{2} p_i - 1 \tag{A.3}$$

The steps to optimise our problem are solved in the following expressions:

$$\frac{\partial L(\mathrm{H}(p_i))}{\partial p_1} = -\log_2(p_1) - p_1 \frac{1}{p_1}\frac{1}{\log(2)} - \lambda = 0 \tag{A.4}$$

$$\frac{\partial L(\mathrm{H}(p_i))}{\partial p_2} = -\log_2(p_2) - p_2 \frac{1}{p_2}\frac{1}{\log(2)} - \lambda = 0 \tag{A.5}$$

By regrouping terms, we can reach the following expressions of Maximum Entropy in the agents' decisions:

$$\log_2(p_1) = \log_2(p_2) \tag{A.6}$$

$$2^{\log_2(p_1)} = 2^{\log_2(p_2)}$$

$$p_1 = p_2 = 0.5$$

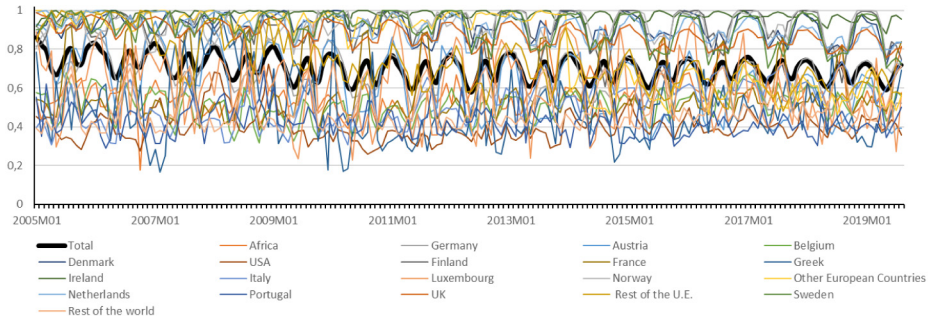


Fig. 6. Measurement of spatial uncertainty based on Entropy from Jan. 2005 to Aug. 2019. Here the reference series r is called Total. The figure represents the Entropy series of all nationalities of tourists visiting Spain for the period analysed. In Appendix B, we study the Entropy series clustered according to the methodology described in this paper.

Appendix B

See Fig. 6.

References

Adamiak, C., 2018. Mapping airbnb supply in European cities. Ann. Tour. Res. 71 (C), 67–71. http://dx.doi.org/10.1016/j.annals.2018.02.008.

Aghabozorgi, S., Seyed Shirkhorshidi, A., Ying Wah, T., 2015. Time-series clustering - A decade review. Inf. Syst. 53, 16–38. http://dx.doi.org/10.1016/j.is.2015.04.007.

Aldana-Bobadilla, E., Kuri-Morales, A., 2015. A clustering method based on the maximum entropy principle. Entropy 17, 151–180. http://dx.doi.org/10.3390/e17010151.

Alonso, A.M., Berrendero, J.R., Hernández, A., Justel, A., 2006. Time series clustering based on forecast densities. Comput. Statist. Data Anal. 51, 762–766. http://dx.doi.org/10.1016/j.csda.2006.04.035.

Alonso, Andrés M., Galeano, P., Peña, D., 2020. A robust procedure to build dynamic factor models with cluster structure. J. Econometrics 216 (1), 3552. http://dx.doi.org/10.1016/j.jeconom.2020.01.004.

Alonso, Andrés M., Maharaj, E.A., 2006. Comparison of time series using subsampling. Comput. Statist. Data Anal. 50, 2589–2599. http://dx.doi.org/10.1016/j.csda.2005.04.010.

Alonso, Andrés M., Peña, D., 2019. Clustering time series by linear dependency. Stat. Comput. 29, 655–676. http://dx.doi.org/10.1007/s11222-018-9830-6.

Ashworth, G., Page, S.J., 2011. Urban tourism research: Recent progress and current paradoxes. Tour. Manag. 32 (1), 1–15. http://dx.doi.org/10.1016/j.tourman.2010.02.002.

Baron, R.R.V., 1984. Tourism terminology and standard definitions. Tour. Rev. 39 (2–4), http://dx.doi.org/10.1108/eb057891.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. http://dx.doi.org/10.1007/978-1-4757-0450-1.

Caiado, J., Ann Maharaj, E., D'Urso, P., 2015. Time-series clustering. In: Handbook of Cluster Analysis. pp. 241–264. http://dx.doi.org/10.1201/b19706.

Caiado, J., Crato, N., 2010. Identifying common dynamic features in stock returns. Quant. Finance 10, 797–807. http://dx.doi.org/10.1080/14697680903567152.

Caiado, J., Crato, N., Peña, D., 2006. A periodogram-based metric for time series classification. Comput. Statist. Data Anal. 50, 2668–2684. http://dx.doi.org/10.1016/j.csda.2005.04.012.

Caiado, J., Crato, N., Peña, D., 2009. Comparison of times series with unequal length in the frequency domain. Comm. Statist. Simulation Comput. 38, 527–540. http://dx.doi.org/10.1080/03610910802562716.

Caruana, R., Elhawary, M., Nguyen, N., Smith, C., 2006. Meta clustering. In: Proceedings - IEEE International Conference on Data Mining, ICDM. http://dx.doi.org/10.1109/ICDM.2006.103.

Coppi, R., D'Urso, P., Giordani, P., 2010. A fuzzy clustering model for multivariate spatial time series. J. Classification 27, 54–88. http://dx.doi.org/10.1007/s00357-010-9043-y.

Das, S., Abraham, A., Konar, A., 2009. Metaheuristic clustering. In: Metaheuristic Clustering. http://dx.doi.org/10.1007/978-3-540-93964-1.

de Luca, G., Zuccolotto, P., 2011. A tail dependence-based dissimilarity measure for financial time series clustering. Adv. Data Anal. Classif. 5, 323–340. http://dx.doi.org/10.1007/s11634-011-0098-3.

De Luca, G., Zuccolotto, P., 2017. Dynamic tail dependence clustering of financial time series. Statist. Papers http://dx.doi.org/10.1007/s00362-015-0718-7.

Delgado-Bonal, A., Marshak, A., 2019. Approximate entropy and sample entropy: A comprehensive tutorial. In: Entropy. pp. 1–37. http://dx.doi.org/10.3390/e21060541.
Di Lascio, F.M.L., Durante, F., Pappadà, R., 2017. Copula-based clustering methods. In: Copulas and Dependence Models with Applications.
Disegna, M., D'Urso, P., Durante, F., 2017. Copula-based fuzzy clustering of spatial time series. Spatial Stat. 21 (Part A), 209–225. http://dx.doi.org/10.1016/j.spasta.2017.07.002.
Dredge, D., 1999. Destination place planning and design. Ann. Tour. Res. 26 (4), 772–791. http://dx.doi.org/10.1016/S0160-7383(99)00007-9.
Dunn, J.C., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. 3, 32–57. http://dx.doi.org/10.1080/01969727308546046.
Durante, F., Pappadà, R., Torelli, N., 2014. Clustering of financial time series in risky scenarios. Adv. Data Anal. Classif. 8, 359–376. http://dx.doi.org/10.1007/s11634-013-0160-4.
Durante, F., Sempi, C., 2015. Principles of Copula Theory. http://dx.doi.org/10.1201/b18674.
D'Urso, P., De Giovanni, L., Maharaj, E.A., Massari, R., 2014. Wavelet-based self-organizing maps for classifying multivariate time series. J. Chemom. 28, 28–51. http://dx.doi.org/10.1002/cem.2565.
D'Urso, P., De Giovanni, L., Massari, R., 2016. GARCH-based robust clustering of time series. Fuzzy Sets and Systems 305, 1–28. http://dx.doi.org/10.1016/j.fss.2016.01.010.
D'Urso, P., Di Lallo, D., Maharaj, E.A., 2013. Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks. Soft Comput. 17, 83–131. http://dx.doi.org/10.1007/s00500-012-0905-6.
D'Urso, P., Maharaj, E.A., 2009. Autocorrelation-based fuzzy clustering of time series. Fuzzy Sets and Systems 160, 3565–3589. http://dx.doi.org/10.1016/j.fss.2009.04.013.
Durso, P., Maharaj, E.A., 2012. Wavelets-based clustering of multivariate time series. Fuzzy Sets and Systems 193, 33–61. http://dx.doi.org/10.1016/j.fss.2011.10.002.
D'Urso, P., Maharaj, E.A., Alonso, A.M., 2017. Fuzzy clustering of time series using extremes. Fuzzy Sets and Systems 318, 56–79. http://dx.doi.org/10.1016/j.fss.2016.10.006.
Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining.
García-Escudero, L.Á., Gordaliza, A., 1999. Robustness properties of k means and trimmed k means. J. Amer. Statist. Assoc. 94, 956–969. http://dx.doi.org/10.1080/01621459.1999.10474200.
Honarkhah, M., Caers, J., 2010. Stochastic simulation of patterns using distance-based pattern modeling. Math. Geosci. 42, 487–517. http://dx.doi.org/10.1007/s11004-010-9276-7.
Inniss, T.R., 2006. Seasonal clustering technique for time series data. European J. Oper. Res. 175, 376–384. http://dx.doi.org/10.1016/j.ejor.2005.03.049.
Jackson, J., Murphy, P., 2006. Clusters in regional tourism an Australian case. Ann. Tour. Res. 33 (4), 1018–1035. http://dx.doi.org/10.1016/j.annals.2006.04.005.
James, G.M., Sugar, C.A., 2003. Clustering for sparsely sampled functional data. J. Amer. Statist. Assoc. 98, 397–408. http://dx.doi.org/10.1198/016214503000189.
Kakizawa, Y., Shumway, R.H., Taniguchi, M., 1998. Discrimination and clustering for multivariate time series. J. Amer. Statist. Assoc. 93, 328–340. http://dx.doi.org/10.1080/01621459.1998.10474114.
Kalpakis, K., Gada, D., Puttagunta, V., 2001. Distance measures for effective clustering of ARIMA time-series. In: Proceedings - IEEE International Conference on Data Mining. ICDM, pp. 273–280. http://dx.doi.org/10.1109/icdm.2001.989529.
Lafuente-Rego, B., Vilar, J.A., 2016. Clustering of time series using quantile autocovariances. Adv. Data Anal. Classif. 10, 391–415. http://dx.doi.org/10.1007/s11634-015-0208-8.
Lim, G.G., Kim, D.H., Choi, M., Choi, J.H., Lee, K.C., 2010. An exploratory study of the weather and calendar effects on tourism web site usage. Online Inf. Rev. 34 (1), 127–144. http://dx.doi.org/10.1108/14684521011024164.
Luna-Romera, J.M., García-Gutiérrez, J., Martínez-Ballesteros, M., Riquelme Santos, J.C., 2018. An approach to validity indices for clustering techniques in Big Data. Prog. Artif. Intell. 7, 81–94. http://dx.doi.org/10.1007/s13748-017-0135-3.
MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297.
Maharaj, Elizabeth Ann, 1996. A significance test for classifying ARMA models. J. Stat. Comput. Simul. 54, 305–331. http://dx.doi.org/10.1080/00949659608811737.
Maharaj, Elizabeth A., Alonso, A.M., D'Urso, P., 2015. Clustering seasonal time series using extreme value analysis: An application to Spanish temperature time series. Commun. Stat. Case Stud. Data Anal. Appl. 1, 175–191. http://dx.doi.org/10.1080/23737484.2016.1179140.
Maharaj, Elizabeth Ann, D'Urso, P., 2010. A coherence-based approach for the pattern recognition of time series. Physica A 389, 3516–3537. http://dx.doi.org/10.1016/j.physa.2010.03.051.
Maharaj, Elizabeth Ann, D'Urso, P., 2011. Fuzzy clustering of time series in the frequency domain. Inform. Sci. 181, 1187–1211. http://dx.doi.org/10.1016/j.ins.2010.11.031.
Majewska, J., 2015. Inter-regional agglomeration effects in tourism in Poland. Tour. Geogr. 17 (3), 408–436. http://dx.doi.org/10.1080/14616688.2014.997279.
Michael, E.J., 2003. Tourism micro-clusters. Tour. Econ. 9 (2), 133–145. http://dx.doi.org/10.5367/000000003101298312.
Otranto, E., 2010. Identifying financial time series with similar dynamic conditional correlation. Comput. Statist. Data Anal. 54, 1–15. http://dx.doi.org/10.1016/j.csda.2009.07.026.
Patton, A.J., 2012. A review of copula models for economic time series. J. Multivariate Anal. 110, 4–18. http://dx.doi.org/10.1016/j.jmva.2012.02.021.

Patton, A., 2013. Copula methods for forecasting multivariate time series. In: Handbook of Economic Forecasting. http://dx.doi.org/10.1016/B978-0-444-62731-5.00016-6.
Peiró-Signes, A., Segarra-Oña, M. del V., Miret-Pastor, L., Verma, R., 2015. The effect of tourism clusters on U.S. hotel performance. Cornell Hosp. Q. 56 (2), 155–167. http://dx.doi.org/10.1177/1938965514557354.
Piccolo, D., 1990. A distance measure for claasifying ARIMA models. J. Time Series Anal. 11, 153–164. http://dx.doi.org/10.1111/j.1467-9892.1990.tb00048.x.
Porter, M.E., 1998. Clusters and the new economics of competition. Harv. Bus. Rev. 76 (6), 77–90.
Reina, M.Á.R., 2020. Big data: Forecasting and control for tourism demand. In: R, I., Valenzuela, O., Rojas, F., Herrera, L.J., Pomares, H. (Eds.), Theory and Applications of Time Series Analysis. ITISE 2019, Springer, Cham, pp. 273–286. http://dx.doi.org/10.1007/978-3-030-56219-9_18.
Ruiz-Reina, M.Á., 2019. Entropy of Tourism: the unseen side of tourism accommodation. In: Proceedings of the International Conference on Applied Research in Business, Management and Economics.
Ruppert, D., 2004. The elements of statistical learning: Data mining, inference, and prediction. J. Amer. Statist. Assoc. http://dx.doi.org/10.1198/jasa.2004.s339.
Scotto, M.G., Alonso, A.M., Barbosa, S.M., 2010. Clustering time series of sea levels: Extreme value approach. J. Waterway Port Coast. Ocean Eng. 136, 215–225. http://dx.doi.org/10.1061/(ASCE)WW.1943-5460.0000045.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 37, 9–423. http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x.
Sripada, S.C., 2011. Comparison of purity and entropy of K-means clustering and fuzzy c means clustering. Indian J. Comput. Sci. Eng. 2 (3), 343–346.
Stuetzle, W., 2003. Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample. J. Classification 20 (1), 25–47. http://dx.doi.org/10.1007/s00357-003-0004-6.
UNWTO, 2021. The first global dashboard for tourism insights. https://www.unwto.org/unwto-tourism-dashboard.
Vilar, J.A., Lafuente-Rego, B., D'Urso, P., 2018. Quantile autocovariances: A powerful tool for hard and soft partitional clustering of time series. Fuzzy Sets and Systems 340, 38–72. http://dx.doi.org/10.1016/j.fss.2017.03.006.
Vilar, J.A., Pértega, S., 2004. Discriminant and cluster analysis for Gaussian stationary processes: Local linear fitting approach. J. Nonparametr. Stat. 16, 162–443. http://dx.doi.org/10.1080/10485250410001656453.
Vlachos, I., Bogdanovic, A., 2013. Lean thinking in the European hotel industry. Tour. Manag. 36, 354–363. http://dx.doi.org/10.1016/j.tourman.2012.10.007.
Warren Liao, T., 2005. Clustering of time series data - A survey. Pattern Recognit. 38, 1857–1874. http://dx.doi.org/10.1016/j.patcog.2005.01.025.
Yang, Y., Fik, T., 2014. Spatial effects in regional tourism growth. Ann. Tour. Res. 46, 144–162. http://dx.doi.org/10.1016/j.annals.2014.03.007.
Yong-Jin, A.L.J., Jang, S., Jinwon, K., 2020. Impacts of peer-to-peer accommodation use on travel patterns. Ann. Tour. Res. 83, 102960. http://dx.doi.org/10.1016/j.annals.2020.102960.
Zhang, T., Ramakrishnan, R., Livny, M., 1996. BIRCH: An efficient data clustering method for very large databases. http://dx.doi.org/10.1145/235968.233324, SIGMOD Record (ACM Special Interest Group on Management of Data).
Zhang, X., Song, H., Huang, G.Q., 2009. Tourism supply chain management: A new research agenda. Tour. Manag. 30 (3), 345–358. http://dx.doi.org/10.1016/j.tourman.2008.12.010.