Environmental discourse in hotel online reviews: a big data analysis

Marcello Mariani & Matteo Borghi

Routledge
Taylor & Francis Group

Environmental discourse in hotel online reviews: a big data analysis

Marcello Mariani 🆔 and Matteo Borghi

Henley Business School, University of Reading, Reading, UK

ABSTRACT

The purpose of this study is to investigate if there is a trend in online consumers' environmental discourse, and whether online consumers' environmental discourse differs across different types of online review platforms, i.e., transaction-based vs. community-based platforms. To achieve this purpose, first we define the concept of online consumers' environmental discourse and operationalize the measures of online consumers' environmental discourse *presence* and *depth*. Second, we retrieve more than 5.5 million online reviews related to hotels located in leading tourism destinations in the Americas and Europe, over the period 2003–2018. The online reviews, collected and analyzed using big data analytical techniques, are sourced from two different types of platforms: Booking.com and Tripadvisor. We find that while environmental awareness (i.e., the *presence* of online environmental discourse) is relatively high but declining over time, the *depth* of the environmental discourse is rather marginal but increasing over time. We also observe that both the *presence* and *depth* of environmental discourse, as well as other text analytics (subjectivity, diversity, length, sentiment, readability), related to the environmental discourse differ across platforms. The relevant theoretical contributions and managerial implications for tourism and hospitality research are also discussed.

Introduction

Environmental sustainability and sustainable development have become increasingly relevant topics in the agenda of tourism policy makers, destination managers and tourism researchers alike (Hall, 2019). The importance of these topics is witnessed by the presence of a growing body of research that has been published by an academic journal entirely dedicated to sustainable tourism and a number of other academic outlets. Despite the (rhetorical) emphasis on environmental sustainability that the United Nations World Tourism Organization (UNWTO) has developed in its official documents, a number of scholars have underlined that, at the global scale, tourism is less sustainable than ever (e.g., Oklevik et al., 2019; Rutty et al., 2015; Scott et al., 2016).

Partially drawing on the eight Millennium Development Goals (MDGs) put forward by the Millennium Summit held in 2000 – the 2030 Agenda for Sustainable Development adopted in

---

2015 by all United Nations member states (Bricker et al., 2013) entails 17 Sustainable Development Goals (SDGs), which represent long-term strategic goals aimed at eradicating poverty, reducing inequality and deprivations, improving education and health, and safeguarding and preserving the planet and natural resources (United Nations, 2019). Despite the UN 2030 Agenda for Sustainable Development mentions tourism only three times – in relation to use and preservation of natural sources, generation of employment and promotion of local culture, and use of marine resources (Hall, 2019) – it is interesting to observe how the SDGs have become focal points for the study of the contribution of tourism to sustainable development and the sustainability of tourism (e.g., Saarinen & Rogerson, 2014).

So far, tourism researchers have mostly relied on traditional media coverage of environmental issues (most notably, climate change) to capture and describe the public opinion and discourses about environmental sustainability issues in tourism (e.g., Gössling & Peeters, 2007), thus neglecting the use of big data analytics from user generated content (UGC) to gain knowledge about consumers' perceptions of environmental sustainability issues. The few recent studies that have used online reviews (ORs) to describe corporate social responsibility (e.g., Brazytė et al., 2017; D'Acunto et al., 2020; Ettinger et al., 2018) or capture public opinion about environmental concerns (e.g., Londoño & Hernandez-Maskivker, 2016; Saura et al., 2018), have failed to define, conceptualize, operationalize and empirically examine what we term as *online consumers' environmental discourse,* its dimensions, its evolution over time, and the way it differs across different platforms. Accordingly, this study is the first to 1) define online consumers' environmental discourse and operationalize the measures of online consumers' environmental discourse *presence* and *depth*; 2) track both the *presence* and *depth* of *online consumers' environmental discourse* using ORs for multiple leading tourism destinations, over an extended period of time and across different types of OR platforms. This methodological approach allows us to address an overarching research question: "To what extent has *online consumers' environmental discourse* changed over time and does it differ across different types of online review platforms?" We break down this overarching question into two more specific sub-research questions: (1) Is there a trend of the *presence* and *depth* of online consumers' environmental discourse? 2) Does online consumers' environmental discourse differ across different types of online review platforms – transaction-based vs. community-based? Our work is distinctive as it is the first to track longitudinally both the *presence* and *depth* of online consumers' environmental discourse by using electronic word-of-mouth (eWOM) covering millions of online conversations across different types of OR platforms and across different destinations, countries and continents. Accordingly, this work makes a relevant contribution to the area at the intersection between big data analytics, eWOM, and sustainable tourism research.

To achieve its goal, the paper is structured as follows. Section 2 reviews the relevant tourism and hospitality literature in the fields of big data analytics, eWOM and online reviews, and sustainable tourism. Section 3 illustrates the methodology adopted. In the fourth section we report and discuss the findings. Section 5 elucidates the theoretical contributions and managerial contributions. The sixth and last section draws the main conclusions and identifies the limitations of the study and avenues for future research.

Literature review

Using big data analytics to analyze electronic word-of-mouth

Beyond representing a technological paradigm part of the 4th industrial revolution (Mariani and Borghi, 2019), per se, big data (BD) has been defined as "the enormous volume of both unstructured and structured data generated by technology developments and the exponentially increasing adoption of devices allowing for automation and connection to the internet" (Mariani et al., 2018: p. 3515). BD displays three major features that have been epitomized as the 3 Vs: volume

(i.e., the large size of data, that might be in the order of petabytes), velocity (i.e., the rapidity of data creation, transfer and modification) and variety (i.e., data can take numerous forms, including images, videos, sounds and text). Subsequent definitions of the concept of BD have included further elements such as veracity, (i.e., the completeness and reliability of data) and value (i.e., the processes aimed at extracting valuable insights from data by means of BD analytics) (Bello-Orgaz et al., 2016).

Interestingly, a wide number of economic actors and researchers are finding it particularly useful to deploy BD analytics to identify patterns in data and derive knowledge that can generate competitive business intelligence (Davenport, 2014; Mariani and Fosso Wamba, 2020) and scientific knowledge (Lycett, 2013). The tourism and hospitality sector is not an exception (Li et al., 2018; Mariani et al., 2018) – increasingly, companies and researchers active in these industries are making use of BD by generating BD analytics from a number of different sources, such as: (a) users in the form of user-generated content (UGC); (b) devices in the form of device data; and (c) operations in the form of transaction data. However, the most popular source of data is UGC. For instance, ORs have been extensively leveraged to understand more about online customer satisfaction, experience and engagement with tourism and hospitality services (Guo et al., 2017; Mariani and Predvoditeleva, 2019; Mariani et al., 2020; Mariani & Matarazzo, 2020; Xiang et al., 2015), and tourism destinations (Mariani et al., 2016; Mariani et al., 2018). ORs include both structured and unstructured data. The rating of a review is a number (structured format) but the written text of the review is unstructured and scholars in hospitality and tourism have deployed text analytics (e.g., Xiang et al., 2015) to gain a better understanding of specific features of the review and the reviewer.

ORs allow current, former and prospective consumers to elaborate and share their perceptions and opinions about products, services and brands on the internet (Hennig-Thurau et al., 2004). In marketing literature, they represent a paramount component of the so-called electronic word-of-mouth (eWOM). An expanding group of scholars, in disciplines such as marketing, information management and computer science, is examining both antecedents and consequences of eWOM (Fang, 2014; Rosario et al., 2016). Within the tourism and hospitality domains, eWOM is continuously generated by online consumers writing ORs on major online travel review websites such as Tripadvisor and CTrip, and online travel agencies (OTAs) such as Booking.com and Expedia.com. These ORs have been found to drive consumers' purchasing and booking intentions (Ghosh, 2018), and ultimately sales (Ye et al., 2009), and firms' financial performance (Mariani and Visani, 2019; Mariani and Borghi, 2020; Yang et al., 2018). Travellers rely on eWOM from ORs as a relevant information source for purchase decisions as, typically, the quality of tourism and accommodation service is not known before consumption, therefore it is difficult to assess before purchase (Filieri & McLeay, 2014; Litvin et al., 2008) and online reviews can be a relevant proxy of other customers' satisfactions with the service (Mariani and Borghi, 2018; Mariani et al., 2019).

A number of studies have used tourism- and hospitality-related ORs and eWOM to generate big data and analytics (Mariani, 2019). For instance, Xiang et al. (2015) have used big data and text analytics from Expedia ORs to identify the text-derived factors that hospitality service customers associate with their experience and the extent to which these factors are associated with customer satisfaction, operationalized by means of OR ratings. Wood et al. (2013) have deployed social media (namely, the locations of Flickr photographs) to quantify visitation rates at more than 800 recreational sites around the globe, thus concluding that crowd-sourced information can be utilized as a reliable proxy for empirical visitation rates and to understand how changes in ecosystems could alter visitation rates. Xiang et al. (2017) use text analytics from Tripadvisor, Expedia and Yelp, to make sense of how the entire hotel population in Manhattan, New York City, is represented on these platforms. Guo et al. (2017) use latent Dirichlet analysis on 266,544 ORs of hotels to identify 19 key dimensions of customer service mentioned by reviewers in their ORs. To summarize, UGC and ORs (in the context of travel, tourism and hospitality) are critical to gain insights about tourists' opinions,

perceptions and behaviors. In the next subsection we discuss how research has used eWOM to make sense of online users' environmental awareness, perceptions and concerns.

Environmental discourse in tourism

The United Nations 2030 Agenda for Sustainable Development resolution mentions tourism in relation to the use and conservation of natural resources (Hall, 2019). Several stakeholders operating in the tourism, travel and hospitality industries are increasingly aware that natural resources' use and consumption is growing at an uncontrollable speed and will likely have a detrimental environmental impact on the planet (DiPietro et al., 2013;; Gössling & Peeters, 2015). For this reason, policy makers (Intergovernmental Panel on Climate Change, 2013), tourism and hospitality firms (Ettinger et al., 2018), and also a portion of consumers/travelers (Liu et al., 2013) are becoming increasingly concerned for the environment and voice these concerns using eWOM. This is the reason why UGC and eWOM have been used to make sense of online users' opinions about environmental and public health issues (e.g., Chisholm & O'Sullivan, 2017; Palomino et al., 2016; Reyes-Menendez et al., 2018).

Individuals voice their environmental concerns via UGC in the guise of social media posts (e.g., Chisholm & O'Sullivan, 2017; Palomino et al., 2016; Reyes-Menendez et al., 2018) and also through online reviews (e.g. Londoño & Hernandez-Maskivker, 2016; Saura et al., 2018). As far as social media posts are concerned, Reyes-Menendez et al. (2018) have analyzed 5,873 tweets using the hashtag #WorldEnvironmentDay and adopted textual analysis to cluster the tweets across the Sustainable Development Goals (SDGs), thus identifying the key environmental and public health issues that most concern Twitter users (these include climate change, water pollution, global warming, etc.).

As far as online reviews are concerned, Londoño and Hernandez-Maskivker (2016) gathered Tripadvisor ORs pertaining to the hotels within the Tripadvisor Green Leaders program in six destinations (Berlin, Boston, Chicago, Copenhagen, Paris and Toronto) and analyzed them using sentiment analysis to detect managerial practices aimed at environmental sustainability. Interestingly, the authors reveal two major findings: first, most of the hotels that implemented green practices, did it for commercial imperatives and to enter the new niche market of green consumers. Second, hotel customers do not recognize green practices yet and do not consider environmental issues when writing a review.

Based on a small sample of 2,487 ORs of 30 Costa Rican hotels possessing a Certification for Sustainable Tourism, Brazytè et al. (2017) find that 31.7% of them mention implicitly sustainability indicators. Moreover, with descriptive statistics they show that the OR ratings, where sustainability indicators are explicitly used, are higher than review ratings of hotels where sustainability indicators are not mentioned. By deploying qualitative content analysis on a small sample of 1,383 ORs related to 47 Austrian hotels (and mentioning CSR aspects), Ettinger et al. (2018) find that a large majority of the review (92.89%) was of a positive or neutral nature. However, the scholars do not articulate what positive or neutral implies in terms of actual ratings. Saura et al. (2018) collected data about the top 25 Swiss hotels based on the Tripadvisor Traveler's Choice ranking 2018 and implemented sentiment analysis on them. Based on a sample of 8,331 ORs, they found that there are some key factors related to environmental management detected by travelers.

To sum up, a few researchers (Brazytè et al., 2017; D'Acunto et al., 2020; Ettinger et al., 2018; Lee et al., 2016; Londoño & Hernandez-Maskivker, 2016; Saura et al., 2018; Yu et al., 2017) have attempted to analyze, in different ways, eWOM to gain some understanding of tourists and hotel guests' opinions about environmental sustainability issues. However, this body of research displays relevant limitations. First, while most of the studies assume that some forms of UGC can be used to make sense of online users' opinions about environmental issues and concerns (e.g.,

Reyes-Menendez et al., 2018), none of them develop a clear definition of *online consumers' environmental discourse*; we instead define this as "eWOM in the form of ORs directly related to online consumers' evaluations of environmental issues". Second, existing research does not define and distinguish clearly the *presence* of online consumers' environmental discourse (i.e., online consumers' environmental awareness), from the *depth* of online consumers' environmental discourse (i.e., the extent to which online consumers dig in-depth about environmental aspects and concerns). Third, extant literature has mainly analyzed the data in a static fashion, without taking a longitudinal perspective that allows identifying a trend in online consumers' environmental discourse. However this is critical as some scholars have found that there has been a trend towards higher environmental consciousness (Cho, 2015; Flammer, 2013), while others have found that many consumers are becoming increasingly skeptical about environmental initiatives (Carrigan & Attalla, 2001), or have no environmental awareness (Londoño & Hernandez-Maskivker, 2016). Fourth, existing studies have built on a relatively small amount of online reviews (e.g., Saura et al., 2018) which is far from representative of entire populations (and big data) of ORs and can seriously limit the generalizability of any findings. Fifth, none of the previous studies have endeavoured to examine online consumers' environmental discourse across different types of OR platforms by means of a cross-platform analysis. Indeed, all of the studies using ORs (see D'Acunto et al., 2020; Londoño & Hernandez-Maskivker, 2016; Saura et al., 2018) focus on a single community-based platform (Tripadvisor) and disregard transaction-based platforms such as Booking.com . This is rather surprising, as not all eWOM is equal as online review platforms differ in their structure and functionalities (Marchand et al., 2017), and this influences how ORs are produced and consumed (You et al., 2015). Accordingly, Gligorijevic (2016) argues that OR platforms in tourism and hospitality can be categorized into two types: 1) transaction-based OTAs like Booking.com and Expedia.com; 2) community-based sites such as Tripadvisor, Yelp and OpenRice. With transaction-based OTAs, ORs provide information to prospective buyers and to platform company managers on product popularity; whereas community-based sites were created with the aim of facilitating exchange of opinions, through ORs, of like-minded travelers.

In our work, we try to address these gaps and develop a cross-platform study allowing us to capture to what extent online consumers' environmental discourse has been changing over time and whether that discourse differs across different types of OR platforms.

Methodology

Data and sample

This paper adopts a quantitative approach to the analysis of online reviews. Online review data for this research was collected in January 2019 based on the research design framework illustrated in Appendix 1. In more detail, ORs of hotels were collected from two different platforms: the *community-based* platform Tripadvisor and the *transaction-based* platform Booking.com . Tripadvisor was chosen because it is the largest online travel review site worldwide, while the OTA Booking.com was selected as it embeds the largest number of certified hotel reviews worldwide (Revinate, 2017).

We developed two different web crawlers (i.e., applications that retrieve data from the web automatically) developed in Python, a general-purpose programming language. First, we retrieved the entire list of reviewed hotels located in four of the top 10 tourism destinations in the Americas (i.e., New York City, Miami, Orlando and Las Vegas) and four of the top 10 tourism destinations in Europe (i.e., London, Paris, Rome and Barcelona), derived from tourist arrivals according to Euromonitor International's 2018 report (Geerts, 2018). Secondly, based on a second web scraping module, we retrieved the entire populations of ORs pertaining to the hotels located in all these destinations in both the platforms.

Table 1a. Sample of Tripadvisor ORs (2003–2018).

Tourism destination	n of reviews (English)	% on the total	n hotel
Barcelona	257,597	6.3%	481
Las Vegas	688,527	16.7%	234
London	1,031,412	25.0%	1,057
Miami	88,432	2.2%	118
New York	759,614	18.4%	466
Orlando	516,107	12.5%	333
Paris	465,899	11.3%	1,763
Rome	313,977	7.6%	1,120
Total	**4,121,565**	**100**%	**5,572**

Table 1b. Sample of Booking.com ORs (2017–2018).

Tourism destination	n of reviews (English)	% on the total	n hotel
Barcelona	136,767	8.8%	450
Las Vegas	51,755	3.3%	154
London	595,146	38.2%	1,027
Miami	46,515	3.0%	154
New York	247,471	15.9%	418
Orlando	92,804	6.0%	246
Paris	265,206	17.0%	1,662
Rome	122,102	7.8%	933
Total	**1,557,766**	**100**%	**5,044**

Regarding the selected time frame, we retrieved the entire history of ORs for hotels listed in Tripadvisor. However, since the company launched its platform in 2000 in the US and only two years later in Europe, we selected 2003 as the starting point of our analysis in order to make the samples consistent and comparable for both geographical areas. With Booking.com, the platform rolls over its review data every two years, so we were only able to scrape reviews over the latest two years (2017 and 2018) for each of the retrieved hotels. As with other studies adopting text mining techniques (e.g., Xiang et al., 2017), we kept only ORs written in English in our final database.

As a result, the final sample of ORs retrieved for Tripadvisor consists of 4,121,565 ORs, while for Booking.com it consists of 1,557,766 ORs – overall 5,679,331 ORs were retrieved and constitute the analyzed UGC. By considering multiple continents, destinations, countries and platforms in our research design, we aim at generating findings that are robust and generalizable, thus overcoming several of the deficiencies of extant studies focusing on a single destination or a single platform (e.g., Brazytė et al., 2017; D'Acunto et al., 2020; Ettinger et al., 2018; Lee et al., 2016; Londoño & Hernandez-Maskivker, 2016; Saura et al., 2018; Yu et al., 2017).

Tables 1a and 1b show the sample of ORs analyzed in the study by destination and platform.

Techniques adopted

As we use large volumes of data to address our overarching research question, we have adopted a number of data science techniques. Descriptive statistics and parametric and non-parametric tests of (mean and median) differences were deployed.

Variables

In our study we focus on online consumers' environmental discourse defined as "eWOM in the form of ORs directly related to online consumers' evaluations of environmental issues". As online consumers' environmental discourse is a multidimensional construct, we operationalize it into

Table 2. Variables description.

Variable	Description
Environmental_Presence	It is a dummy variable that is equal to 1 if the review includes at least one word in the environmental dictionary developed by Pencle and Mălăescu (2016), and zero otherwise.
Environmental_Depth	It is a ratio equal to the number of environment-related words (words present in the Pencle and Mălăescu (2016) environmental dictionary) over the total number of words in the review, multiplied by 100.
Review Diversity	It refers to lexical diversity and it is operationalized as the ratio of unique words to total words in the online review text (Lahuerta-Otero & Cordero-Gutiérrez, 2016; Zhang et al., 2016). It ranges from 0 to 1, whereby 1 equates to a text with an absence of redundancies and lexically diverse.
Review Length	It represents the number of words included in each online review (Chevalier & Mayzlin, 2006; Zhang et al., 2016).
Review Polarity	The *polarity* (also known as sentiment score) was operationalized using a continuous variable ranging from −1 to +1 respectively equating to extremely negative and extremely positive content and emotions. To create this measure we used the Valence Aware Dictionary for Sentiment Reasoning (VADER), which exploits a set of heuristics along with a specific lexicon dictionary for this particular task (Hutto & Gilbert, 2014).
Review Readability	It refers to the simplicity of a text for a reader's understanding. It consists of presentation and content and it is measured on a numeric scale. We operationalized readability by means of the Automated Readability Index (ARI) (Senter & Smith, 1967).
Review Subjectivity	It refers to the degree of subjectivity of the review and it has been measured using a variable ranging from 0 to 1, whereby the lower end (i.e. 0) equates to the highest possible use of objective words to describe the service consumed. In particular, we leverage on the TextBlob Python library (Loria, 2014) which uses deep learning techniques to extrapolate the subjectivity of a given text.

two distinct variables: the *presence* of online consumers' environmental discourse which proxies online consumers' environmental awareness, and the *depth* of online consumers' environmental discourse which proxies the extent to which online consumers dig in-depth about environmental aspects and concerns. Accordingly, our two focal variables are *Environmental_Presence* and *Environmental_Depth*. The first, captures the presence of at least one environment-related word based on the environmental dictionary developed by Pencle and Mălăescu (2016). Originally developed to cover two dimensions of corporate social responsibility (CSR) – the social and environmental dimension – the overarching dictionary includes a specific dictionary for environmental aspects that includes 451 terms. This content analytic dictionary has been developed using a technique that captures the meaning embedded in a text (computer-aided text analysis) and has been validated in the context of US initial public offerings in relation to four dimensions of corporate social responsibility (CSR) (Pencle & Mălăescu, 2016) and in the context of CSR in the hospitality setting (D'Acunto et al., 2020). The full list of terms can be found in the article by Pencle and Mălăescu (2016).

The latter variable (i.e., *Environmental_Depth*) measures the share of environment-related words in an OR out of the overall amount of word in the same review and, as such, it captures the depth of the environmental discourse within the focal OR. Table 2 contains the description of the focal variables embedded in the study.

In addition to these two variables, and in line with extant literature in the marketing and data and computer science field, we focused also on a set of text analytics including *Review Diversity*, *Review Length*, *Review Polarity*, *Review Readability* and *Review Subjectivity*. *Review Diversity* captures the lexical diversity of the review and it is operationalized as the ratio of unique words to the overall number of words in the OR text (Lahuerta-Otero & Cordero-Gutiérrez, 2016; Zhang et al., 2016). *Review Length* is the count of the number of words included in the online review (Chevalier & Mayzlin, 2006; Zhang et al., 2016). *Review Polarity* (also known as sentiment score) has been operationalized using a continuous variable ranging from −1 to +1 and is computed – in line with Alaei et al. (2019) – using the Valence Aware Dictionary for Sentiment Reasoning (VADER), which exploits a set of heuristics along with a specific lexicon dictionary for this particular task (Hutto & Gilbert, 2014). The VADER-based sentiment analysis technique and related

Table 3. Online environmental discourse presence and depth in eWOM.

Platform	2003–2018		2017–2018	
	Presence	Depth	Presence	Depth
Tripadvisor	58.26%	0.973%	53.80%	0.998%
Booking.com			24.76%	1.047%

measure has been deployed because recent research (Alaei et al., 2019) has found that it outperforms other sentiment analysis classifiers used in the tourism and hospitality domain, while being consistent with them. *Review Readability* refers to the simplicity of a text for a reader's understanding. Consistently with extant literature (Korfiatis et al., 2012), we operationalized readability by means of the Automated Readability Index (ARI). ARI has a long tradition in text mining analysis; it was developed in the 1960s to estimate the number of years of formal education a person needs in the US education system to understand an (English) text on the first reading (Senter & Smith, 1967). *Review Subjectivity* relates to the degree of subjectivity of an OR and it has been measured by leveraging the Python TextBlob library which uses deep learning techniques to extrapolate the subjectivity of a given text (Loria, 2014; Zhao et al., 2019).

Findings and discussion

The analyses conducted reveal that a relevant number of online consumer reviews on Tripadvisor are environmentally related, with 58.26% of Tripadvisor ORs including at least one environmental-related word over the period 2003–2018 (see Table 3). The share is slightly lower (53.80%) over the more recent duration 2017–2018. Overall, this seem to suggest that online Tripadvisor reviewers are environmentally concerned. The situation is slightly different for Booking.com ORs, as 24.76% of these include at least one environmental-related word over the period 2017–2018. Taken together, these figures suggest that the *presence* of environmental-related eWOM is relevant in both platforms, representing one fourth of Booking.com ORs and one half of Tripadvisor ORs. However, the inter-platform comparison over the timespan 2017–2018 suggests that the *presence* of environmental-related eWOM is higher for community-based OR platforms such as Tripadvisor, than it is for transaction-based platforms such as Booking.com (see column four of Table 3). This result might be explained in light of the different aims and functionalities of the two types of platforms (Gligorijevic, 2016) and the most pronounced social media connotation of Tripadvisor, as it encourages the development of a proper conversation between like-minded travelers (rather than a mere evaluation of services that is typically the aim of ORs on transactional platforms and e-commerce websites). Moreover, Tripadvisor (unlike Booking.com) has a long tradition in recognizing the green practices of hotels through an ad hoc initiative called the GreenLeaders Program (Font & Tribe, 2001; Londoño & Hernandez-Maskivker, 2016). As community-based OR platforms typically encourage online environmental conversations, it is interesting to observe that, instead, users of transaction-based platforms write fewer green reviews than their community-based platform counterparts.

The actual share of environmental-related words in each OR is, on average, relatively low. Over the period 2017–2018 the share is equal to 0.998% for Tripadvisor ORs and 1.047% for Booking.com ORs. This interesting finding suggests that while only one fourth of Booking.com users mention environmental concerns in their ORs, those who do discuss them, do so with a depth similar to that of Tripadvisor's users. In other words, Booking.com reviewers are more motivated to express their environmental concerns as they have made an added effort within a platform that is not conceived for community interaction but mostly for transactions.

If we plot the trend of online consumers' environmental discourse *presence* over the 2003–2018 timespan for Tripadvisor we detect an inverted U-shaped trend for the entire sample of ORs, with an increase from 58.2% in 2003 to 68.0% in 2008, followed by a sharp decline to

53.2% in 2018. This finding seems to reconcile the extant puzzle in the literature where there are contrasting positions: research that has detected a general trend toward higher environmental consciousness over the past decades (Flammer, 2013), and research and industry reports that have recently found the opposite (BBC, 2010; Betz & Peattie, 2012; Koenig-Lewis et al., 2014). Our finding suggests that the trend is mixed and was increasing until 2008 and decreasing thereafter.

Interestingly, ORs related to hotels located in American tourism destination cities are more likely to include at least one environmental-related word than their counterparts in Europe (see the blue lines in Figure 1). There are two possible explanations for this finding. First, the selected European cities are, on average, more sustainable than their American counterparts, based on the Arcadis Sustainable Cities Index Arcadis, 2016 (Arcadis, 2016) – one of the most cited indices to measure city sustainability. More specifically, London, Paris, Rome and Barcelona are in the Top 25 positions of the 2016 ranking. However, the American cities appear in lower positions, with New York in 26th position, Miami 54th, Orlando and Las Vegas not even among the Top 100. As part of the destination infrastructure, it might be that hospitality firms in the US are less environmentally concerned than their European counterparts. A plausible explanation might be that hotel guests staying in US hotels have more to comment on the presence/absence of environmental initiatives as the hospitality infrastructure is perceived less sustainable because it is part of a less sustainable destination.

A second and complementary explanation stems from recent research showing that European workers consider their employers to be "green", while US workers are more critical and believe that their governments and firms are not doing enough in terms of green initiatives (American Management Association, 2019). As most of the hotel guests in US destinations are from the US/America and most of the hotel guests in European destinations are from Europe – as is clear from official statistics and from the distributions of reviewers' countries/continents of origin in our online reviews populations (see Appendix 2) – we expect US consumers to be more likely to write comments concerning environmental issues than European consumers (as they feel that institutions and firms are not doing enough for the environment).

On the other hand, the trend of consumers' environmental discourse *depth* over the 2003–2018 timespan for Tripadvisor has an overall increase from 0.85% in 2003 to 0.99% in 2018, having reached a peak of 1.01% in 2016. Also, in this case, consumers' environmental discourse *depth* is higher in the American subsample than in the European one (see the red lines in Figure 1). The possible explanations mirror those of the findings for the presence variable developed above: first, hotel guests staying in US hotels desire to articulate their discourse on the environmental initiatives of hotels as the hospitality infrastructure is perceived as less sustainable; second, as most of the hotel guests in US destinations are from the US, we expect US consumers to be more likely to write comments on environmental issues than European consumers (American Management Association, 2019) (because they feel that institutions and firms are not doing enough for the environment).

If we read the *Environmental_Presence* and *Environmental_Depth* trends together, overall we observe a decreasing trend of ORs dealing with environmental aspects (i.e., a decline of the *Environmental_Presence*) accompanied by an increasing prominence of environmental-related discourse in those reviews covering environmental topics (i.e., an increase of *Environmental_Depth*). Overall, the combination of these opposite trends might suggest that after an apparent "green fad" taking place between 2003 and 2008, environmental aspects have been covered by a contracting body of ORs that deal with environmental issues in a more elaborated fashion. In other terms, there is a shrinking share of online consumers that explicitly mention environmental aspects in their ORs, but these consumers are increasing the depth of their discourse about environmental concerns. Clearly, this is a novel research result which is not yet covered in the current eWOM literature.
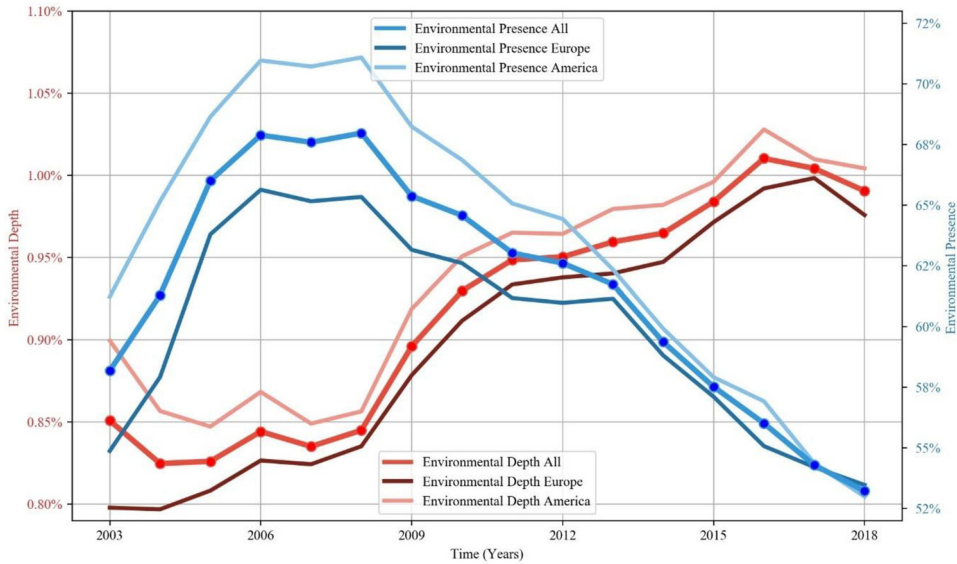
Figure 1. Online environmental discourse presence and depth in Tripadvisor eWOM, 2003–2018.
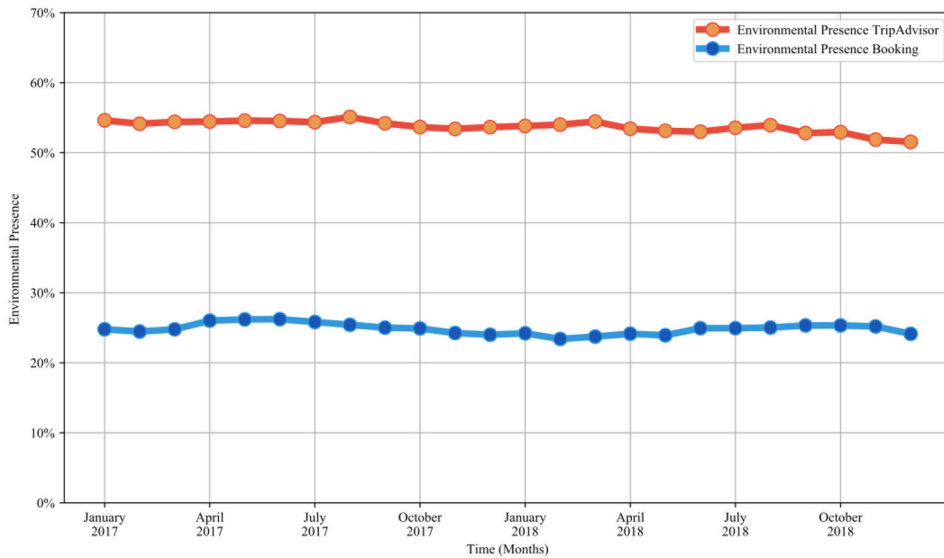

Figure 2. Online environmental discourse presence in Tripadvisor vs. Booking.com eWOM, 2017–2018.

The combined trends might also imply that consumers are increasingly less receptive to companies' ecological claims and are perceived as "greenwashing", i.e., purposefully built claims that (hospitality) firms make for their financial gains, rather than to deliver benefits to the environment (Koenig-Lewis et al., 2014). Accordingly, while we did not count the unique consumers that left a review, the figures seem to show that ORs touching environmental-related aspect are becoming less popular over time but those few ORs dealing with environmental-related aspects are more thoughtful and comprehensive in developing an environmental discourse.

If we focus on the period January 2017 to December 2018 (for which, we also have Booking.com ORs), we observe a decreasing trend in environmental discourse *depth* (see Figure

Table 4. Text analytics comparison between environmental and non-environmental reviews for both Tripadvisor and Booking.com, 2017–2018.

	Tripadvisor				Booking.com			
	Env reviews (n = 615,656)		Non-env reviews (n = 528,805)		Env reviews (n = 385,645)		Non-env reviews (n = 1,172,121)	
	Means		Medians		Means		Medians	
Variable	Δ	t	Δ	W x 10^{11}	Δ	t	Δ	W x 10^{11}
Subjectivity	−0.007	−32.67***	−0.01	1.55***	0.07	176.91***	0.02	2.45***
Length	63.07	359.32***	42	2.27***	37.37	383.86***	30	3.49***
Diversity	−0.06	−314.03***	−0.06	1.11***	−0.09	−404.11***	−0.11	1.31***
Sentiment	0.06	65.49***	0.03	1.97***	0.14	141.79***	0.284	2.78***
Readability	1.59	81.36***	1	1.85***	1.48	96.88***	0.8	2.53***

Note: *** $p < 0.001$.

2) on both Tripadvisor and Booking.com . This trend, which is portrayed by connecting monthly rather than annual observations, is consistent with the trend illustrated in Figure. 1 over the years 2017–2018.

It is worthwhile noticing that the environmental presence on Tripadvisor is higher than on Booking.com based on the computation of parametric tests – the Welch t-test equals 500.3864 ($p < 0.001$). Nonparametric tests such as the Mann-Whitney test indicate a value of 63.3×10^{11} ($p < 0.001$). The reported dissimilarity might be due to the differences across the two types of platforms analyzed that have been found to display different features in previous research, with Tripadvisor being more of a social media type of platform (Gligorijevic, 2016), as based on previous empirical comparisons across platforms (Xiang et al., 2017). Indeed, Tripadvisor buoys up the articulation of an online conversation between peer travelers rather than offering a mere tool for the evaluation of a hospitality service, like in the case of the e-commerce platform Booking.com

Overall, the comparison of trends between the two different platforms analyzed (i.e., the transaction-based Booking.com and the community-based Tripadvisor) reveals that there is an overall consistency of the trends of environmental discourse presence across platforms. This suggests that – despite hotel populations being represented differently across platforms in terms of text analytics (see Xiang et al., 2017) – we instead find that OR textual features display similar forms across distinctively different platforms.

By comparing mean and median values of text analytics for ORs covering environmental-related aspects vs. ORs not covering them, we find that Tripadvisor ORs covering environmental-related aspects are longer ($t = 359.26$, $p < 0.001$), less readable ($t = 81.36$, $p < 0.001$), display a more positive sentiment ($t = 65.49$, $p < 0.001$), less subjective ($t = -32.67$, $p < 0.001$), and less lexically diverse ($t = -314.03$, $p < 0.001$) than ORs not covering environmental-related aspects (see Table 4).

On the other hand, Booking.com ORs covering environmental-related aspects are longer ($t = 383.86$, $p < 0.001$), less readable ($t = 96.88$, $p < 0.001$), display a more positive sentiment ($t = 141.79$, $p < 0.001$), more subjective ($t = 176.91$, $p < 0.001$), and less lexically diverse ($t = -404.11$, $p < 0.001$) than ORs not covering environmental-related aspects.

Overall, the findings point to a similarity across platforms, with the only exception of OR lexical diversity. While not controlling for other variables, other than the focal attribute, these basic comparisons illustrate that the observed differences are prevalent in both platforms.

If we dig in more depth and compare the text analytics of the environmental-related online reviews across the two platforms, additional interesting results emerge. As is clear from Table 5, Tripadvisor ORs covering environmental-related aspects are longer ($t = 401.82$, $p < 0.001$), less readable ($t = 67.14$, $p < 0.001$), less subjective ($t = -16.63$, $p < 0.001$), display a more positive sentiment ($t = 250.31$, $p < 0.001$), and less lexically diverse ($t = -424.34$, $p < 0.001$) than environmental-related ORs in Booking.com

Table 5. Text analytics comparison between Tripadvisor and Booking.com for environmental ORs, 2017–2018.

	Env reviews			
	Tripadvisor (n = 615,656)		Booking.com (n = 385,645)	
	Means		Medians	
Variable	Δ	t	Δ	W x 10^11
Subjectivity	−0.005	−16.63***	−0.013	1.13***
Length	74.67	401.82***	52	1.82***
Diversity	−0.101	−424.34***	−0.101	0.64***
Sentiment	0.27	250.31***	0.199	1.77***
Readability	1.401	67.14***	0.4	1.28***

Note: *** $p < 0.001$.

These further analyses suggest that online consumers' environmental discourse on a community-based platform (i.e., Tripadvisor) is more objective, more difficult to read and has a much higher sentiment polarity. This is consistent with the finding stemming from the computation of the presence of the environmental discourse, which indicates that community-based platforms users are more environmentally concerned than transaction-based platform users; they are more knowledgeable about environmental issues and, therefore, comment on them in a more objective way (with technical language that can make reviews less readable) than users of transaction-based platforms. Moreover, as users on community-based platforms are more interested in sharing their opinions than in actual transactions, their sentiment in what they write is more positive.

Theoretical and managerial contributions and implications

Theoretical contributions

This work contributes to big data analytics, eWOM and environmental research streams in tourism and hospitality in multiple ways. First, to the best of our knowledge, this is the first study defining *online consumers' environmental discourse* and operationalizing the measures of *presence* vs. the *depth* of the online discourse.

Second, it represents the first attempt to track the evolution of the *presence* and *depth* of *online consumers' environmental discourse* over time in digital settings (i.e., on multiple and different types of digital platforms). This work complements extant research in the sustainable tourism literature, emphasizing that the environmental discourse is co-shaped by a multitude of stakeholders in the tourism field, including, not only firms (Ettinger et al., 2018; Peeters & Gössling, 2008), NGOs and national and international governments (Font & Lynes, 2018) that have adopted or espoused environmental schemes and practices (Pedersen & Neergaard, 2006), but also tourists and travelers (Oklevik et al., 2019) that might be interested in more sustainable experiences.

Third, this study finds that online consumers' environmental discourse *presence* has recorded a decreasing trend over the 2003–2018 period, while the *depth* has increased. This helps reconcile mixed research findings that indicated that there has been a trend towards higher environmental consciousness (Cho, 2015; Flammer, 2013), yet many consumers are becoming increasingly skeptical about environmental initiatives (Carrigan & Attalla, 2001; Koenig-Lewis et al., 2014) which are sometimes perceived as "greenwashing" (Betz & Peattie, 2012). By decomposing the environmental discourse into *presence* and *depth* we offer a more nuanced appreciation and explanation of how consumers' environmental awareness and the depth of the discourse have evolved over time and, unlike Londoño and Hernandez-Maskivker (2016), we identify a clear trend in online consumers' environmental discourse (*presence* and *depth*) over time.

Fourth, we propose juxtaposing online analytical approaches of the examination of environmental discourse to traditional approaches that have been typically based on archival/official documents and reports (e.g., Gössling, 2013). As such, we propose to shift the scholarly attention to how *environmental discourses* are being shaped and evolve on digital platforms by means of analyzing UGC in the form of online reviews, thus further extending the nascent research stream that has deployed UGC to make sense of users' opinions about environmental and public health issues (e.g., Reyes-Menendez et al., 2018). Accordingly, our study suggests, that in the age of digital platforms, consumers' environmental concerns can be captured effectively by means of data analytics of online consumer reviews.

Fifth, and related to the previous point, we contribute to the marketing and consumer behavior literature revolving around consumers' behaviors and evaluations of environmental and green aspects (e.g., Thøgersen et al., 2010). Rather than relying on small sample surveys asking respondents stated attitudes and behaviors (i.e., perceptions), we look at the real evaluations after consumption by means of big data analytics from a large sample of more than 5.5 million consumers' ORs. As such, not only do we help to bridge the stated-actual behavior gap, but we also find that environmental-related eWOM is relatively high but decreases over time, whereas the depth of environmental discourse in ORs increases over time.

Sixth, we contribute to the big data analytics research stream in tourism management literature (Li et al., 2018; Mariani et al., 2018), by suggesting that extrapolating analytics from different digital platforms (namely, community vs. transaction-based digital platforms) while providing statistically different quantifications of environmental discourse *presence* and *depth*, offers overall consistent results when it comes to identifying the trends of the environmental discourse *presence* and *depth*. At the same time, our results indicate that, *within* platforms, environmental ORs are systematically different from non-environmental ORs in terms of text analytics. Moreover, we detect cross-platform differences as environmental ORs generated on community-based platforms differ systematically from environmental ORs generated on transaction-based platforms. Environmental discourse on community-based platforms is longer, more objective and more positive in terms of sentiment, which suggests that users of community-based platforms are more aware of environmental issues, communicate about it in a more knowledgeable manner and play the function of advocates of the environment. Accordingly, this enhances our knowledge of online consumers' environmental discourse through a cross-platform approach that is missing from previous studies using ORs (see Londoño & Hernandez-Maskivker, 2016; Saura et al., 2018) and suggests that researchers interested in monitoring empirically online environmental discourse should take into account and control for the platform type.

Practical implications

Several practical implications stem from this work, including implications for destination managers and tourism policy makers, tourism and hospitality practitioners, and digital platforms managers and developers.

As the long-term trend of consumers' environmental awareness (i.e., the *presence* of environmental discourse) is declining over time, destination managers and tourism policy makers should invest more time and financial resources in making consumers aware of the evidenced-based scientific consensus about environmental issues (Gössling & Peeters, 2007). Investing in educating tourists and travelers to recognize the importance of environmental issues should be a priority for educational systems, as it has been shown that the level of consumers' education and ecological knowledge plays a relevant role in environmental consumption behaviors (Chan, 2001). Interestingly, our findings show that the *depth* of environmental discourse is low for both Booking.com and Tripadvisor, but it is increasing over time, possibly the result of an increased acculturation on sustainability matters of online consumers that are particularly aware of

environmental issues. It seems, therefore, that destination managers should invest in pro-environmental initiatives and events – including festivals and workshops (Getz & Page, 2016) – to increase the environmental awareness of tourists. This might contribute to make their destination more competitive and attractive for the "green" segments of the market, vis-à-vis other destinations (Mariani et al., 2014). However, they should also recognize the work of hotel managers implementing green practices (e.g., incentivizing customers who save water and electric energy, or reuse towels).

Tourism and hospitality practitioners need to further develop their online strategies to meet the changing needs and wants of green consumers (Gustin & Weaver, 1996). Moreover, their marketing and communication strategies should be tailored to multiple channels (in relation to OR platforms) to comply with the tenets of multichannel marketing (Duffy, 2004) and omnichannel marketing (Verhoef et al., 2015). Moreover, hospitality managers might tailor their online strategies across platforms. For instance, they might gain membership of a green program on a community-based platform that ranks hotels based on their environmental performance (such as the Tripadvisor Green Leaders program). However, this might not be necessary on transaction-based platforms (such as Booking.com) that focus more on transactional aspects rather than environmental values.

Platform managers and developers that operate online community travel review platforms and OTAs can benefit from these findings as their platforms increasingly host an important share of environmental-related ORs. As it has been found that in offline settings the presence of a "green" attribute appears to heighten consumers' involvement and deliberation in decision making (Thøgersen et al., 2012), platform developers could generate a further service attribute for customers' assessment (i.e., "eco-friendliness" of the hospitality service). This might help hotels to obtain ad hoc feedback on their green practices and for green consumers (Pedersen & Neergaard, 2006) to assess the green attributes of hospitality services before online booking and purchase. Overall, this will help reduce green consumers' bounce rates and ultimately enhance conversions and reservations.

Conclusions and limitations

This study contributes to advance our knowledge of consumers' online environmental discourse in the tourism and hospitality sector. To the best of our knowledge, it constitutes the first attempt to define and operationalize environmental discourse *presence* and *depth* by means of a big data analysis on a large sample of ORs across multiple platforms and destinations (countries and continents). As such, it makes a relevant contribution to the area at the intersection between eWOM, big data analytics and sustainable tourism research, as well as to the more narrowly defined research stream of online consumer behavior trying to explore and examine consumers' environmental concerns (e.g., Londoño & Hernandez-Maskivker, 2016; Reyes-Menendez et al., 2018; Saura et al., 2018).

Based on more than 5.5 million ORs retrieved and analyzed with advanced big data techniques from Tripadvisor and Booking.com, this study has found that online consumers' environmental awareness (i.e., the *presence* of environmental discourse) is relatively high but declining over time, whereas the *depth* of the environmental discourse is low but increasing over time. In line with the specific nature of the two platforms analyzed (a community vs. a transaction-based type of platform), we have observed differences in the *presence* and *depth* of the environmental discourse across platforms. Moreover, our results indicate that, *within* platforms, environmental ORs are systematically and statistically different from non-environmental ORs in terms of analytics. In addition, we detect cross-platform differences as environmental ORs generated on community-based platforms differ systematically and significantly from environmental ORs generated on transaction-based platforms – they are longer, more objective and more positive in terms of

sentiment. These findings should inform not only practice, but also future research using different types of platforms.

This study is not without limitations. First, the analyses could be strengthened by adding additional reviewer-level variables (such as age, education, socio-cultural differences and gender) partially analyzed in extant eWOM studies (Mariani et al., 2019) that have been found to affect pro-environmental attitudes and behaviors (Naderi & Van Steenburg, 2018). Additional variables might help us gain a more nuanced picture of the results based on demographics that display many missing values in ORs data. Also OR submission device might be a variable to control for (Mariani et al., 2019) . Secondly, further work might extent the analysis to other relevant destinations in Asia (such as Hong Kong and Bangkok) and perhaps also examine satisfaction with destinations (Guizzardi and Mariani, 2020) through online reviews. Third and last, researchers can consider expanding our studies to other OR platforms, for example, sharing economy platforms (such as Airbnb) that present a hybrid platform type (Ek Styvén and Mariani, 2020) where transactional features are mixed with social and community-based features.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Dr. Marcello Mariani is a Full Professor of Management and Entrepreneurship at the Henley Business School, University of Reading (UK) and member of the Henley Center for Entrepreneurship and the European Institute for Advanced Studies in Management. His current research interests include big data and analytics, eWOM, digital business models, AI, IoT, automation and coopetition strategies in service industries. He has authored more than 100 publications. His researches have been published in Industrial Marketing Management, Journal of Business Research, Journal of Advertising, Industrial and Corporate Change, Psychology & Marketing, Long Range Planning, Technological Forecasting and Social Change, International Journal of Electronic Commerce, European Accounting Review, Production Planning & Control, Tourism Management, Annals of Tourism Research, Journal of Travel Research, International Journal of Contemporary Hospitality Management, International Journal of Hospitality Management, International Studies in Management and Organizations, and more.

Matteo Borghi is a PhD student at the Henley Business School, University of Reading. He earned a MSc in Business Informatics at the University of Pisa (Italy) with summa cum laude. During his bachelor degree in Information Science for Management (University of Bologna) he was selected as one of the best student and, after his graduation, was a research assistant at the Center for Advanced Studies in Tourism (University of Bologna). His research lies at the intersection of data science, management and entrepreneurship, with special reference to the impact of Industry 4.0 technologies on digital business modelling and e-Reputation of tourism and hospitality companies.

ORCID

Marcello Mariani 🆔 http://orcid.org/0000-0002-7916-2576

References

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, *58*(2), 175–191. https://doi.org/10.1177/0047287517747753

American Management Association. (2019). *New survey looks at the 'greening' of Europe vs. the U.S*. https://www.amanet.org/articles/new-survey-looks-at-the-greening-of-europe-vs-the-us

Arcadis. (2016). *Sustainable cities index 2016. Putting people at the heart of city sustainability*. https://www.arcadis.com/en/global/our-perspectives/sustainable-cities-index-2016

BBC News. (2010, February 7). Climate scepticism 'on the rise', BBC poll shows. [Online]. http://news.bbc.co.uk/1/hi/8500443.stm

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *An International Journal on Information Fusion*, *28*, 45–59. https://doi.org/10.1016/j.inffus.2015.08.005

Betz, F. M., & Peattie, K. (2012). *Sustainability marketing: A global perspective* (2nd ed.). Wiley.

Brazytė, K., Weber, F., & Schaffner, D. (2017). Sustainability management of hotels: How do customers respond in online reviews? *Journal of Quality Assurance in Hospitality & Tourism*, 18(3), 282–307. https://doi.org/10.1080/1528008X.2016.1230033

Bricker, K. S., Black, R., & Cottrell, S. (2013). *Sustainable tourism and the millennium development goals*. Jones & Bartlett Publishers.

Carrigan, M., & Attalla, A. (2001). The myth of the ethical consumer – do ethics matter in purchase behaviour? *Journal of Consumer Marketing*, 18(7), 560–578. https://doi.org/10.1108/07363760110410263

Chan, R. Y. K. (2001). Determinants of Chinese consumers' green purchase behavior. *Psychology & Marketing*, 18(4), 389–413.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Chisholm, E., & O'Sullivan, K. (2017). Using Twitter to explore (un)healthy housing: Learning from the #Characterbuildings campaign in New Zealand. *International Journal of Environmental Research and Public Health*, 14(11), 1424.

Cho, Y. N. (2015). Different shades of green consciousness: The interplay of sustainability labeling and environmental impact on product evaluations. *Journal of Business Ethics*, 128(1), 73–82. https://doi.org/10.1007/s10551-014-2080-4

D'Acunto, D., Tuan, A., Dalli, D., Viglia, G., & Okumus, F. (2020). Do consumers care about CSR in their online reviews? An empirical analysis. *International Journal of Hospitality Management*, 85, 102342. https://doi.org/10.1016/j.ijhm.2019.102342

Davenport, T. H. (2014). How strategists use 'big data' to support internal business decisions, discovery and production. *Strategy & Leadership*, 42(4), 45–50.

DiPietro, R. B., Cao, Y., & Partlow, C. (2013). Green practices in upscale foodservice operations. *International Journal of Contemporary Hospitality Management*, 25(5), 779–796.

Duffy, D. L. (2004). Multi-channel marketing in the retail environment. *Journal of Consumer Marketing*, 21(5), 356–359. https://doi.org/10.1108/07363760410549177

Ek Styvén, M., & Mariani, M. M. (2020). Understanding the intention to buy secondhand clothing on sharing economy platforms: The influence of sustainability, distance from the consumption system, and economic motivations. *Psychology & Marketing*, 37(5), 724–739. https://doi.org/10.1002/mar.21334

Ettinger, A., Grabner-Kräuter, S., & Terlutter, R. (2018). Online CSR communication in the hotel industry: Evidence from small hotels. *International Journal of Hospitality Management*, 68, 94–104. https://doi.org/10.1016/j.ijhm.2017.09.002

Fang, Y. H. (2014). Beyond the credibility of electronic word of mouth: Exploring eWOM adoption on social networking sites from affective and curiosity perspectives. *International Journal of Electronic Commerce*, 18(3), 67–102. https://doi.org/10.2753/JEC1086-4415180303

Filieri, R., & McLeay, F. (2014). E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of Travel Research*, 53(1), 44–57. https://doi.org/10.1177/0047287513481274

Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56(3), 758–781. https://doi.org/10.5465/amj.2011.0744

Font, X., & Lynes, J. (2018). Corporate social responsibility in tourism and hospitality. *Journal of Sustainable Tourism*, 26(7), 1027–1042. https://doi.org/10.1080/09669582.2018.1488856

Font, X., & Tribe, J. (2001). Promoting green tourism: The future of environmental awards. *International Journal of Tourism Research*, 3(1), 9–21. https://doi.org/10.1002/1522-1970(200101/02)3:1<9::AID-JTR244>3.0.CO;2-Q

Geerts, W. (2018). Top 100 city destinations 2018. *Euromonitor International*. [Online] https://go.euromonitor.com/white-paper-travel-2018-100-cities

Getz, D., & Page, S. (2016). Progress and prospects for event tourism research. *Tourism Management*, 52, 593–631. https://doi.org/10.1016/j.tourman.2015.03.007

Ghosh, T. (2018). Predicting hotel book intention: The influential role of helpfulness and advocacy of online reviews. *Journal of Hospitality Marketing & Management*, 27(3), 299–322. https://doi.org/10.1080/19368623.2017.1364198

Gligorijevic, B. (2016). Review platforms in destinations and hospitality. In R. Egger, I. Gula, & D. Walcher, (Eds.), *Open tourism* (pp. 215–228). Springer.

Gössling, S. (2013). Urban transport transitions: Copenhagen, city of cyclists. *Journal of Transport Geography*, 33, 196–206. https://doi.org/10.1016/j.jtrangeo.2013.10.013

Gössling, S., & Peeters, P. (2007). It does not harm the environment!' An analysis of industry discourses on tourism, air travel and the environment. *Journal of Sustainable Tourism*, 15(4), 402–417. https://doi.org/10.2167/jost672.0

Gössling, S., & Peeters, P. (2015). Assessing tourism's global environmental impact 1900–2050. *Journal of Sustainable Tourism*, 23 (5), 639–659. https://doi.org/10.1080/09669582.2015.1008500

Guizzardi A, Mariani MM. Introducing the Dynamic Destination Satisfaction Method: An Analytical Tool to Track Tourism Destination Satisfaction Trends with Repeated Cross-sectional Data. *Journal of Travel Research*. September 2020. doi:10.1177/0047287520958205

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Gustin, M. E., & Weaver, P. A. (1996). Are hotels prepared for the environmental consumer? *Hospitality Research Journal*, 20(2), 1–14. https://doi.org/10.1177/109634809602000201

Hall, C. M. (2019). Constructing sustainable tourism development: The 2030 agenda and the managerial ecology of sustainable tourism. *Journal of Sustainable Tourism*, 27(7), 1044–1060. https://doi.org/10.1080/09669582.2018.1560456

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, 18(1), 38–52. https://doi.org/10.1002/dir.10073

Hutto, C. J., & Gilbert, E. (2014). *VADER-sentiment-analysis*. ICWSM.

Intergovernmental Panel on Climate Change. (2013). *Climate change 2013. The physical science basis*. Cambridge University Press.

Koenig-Lewis, N., Palmer, A., Dermody, J., & Urbye, A. (2014). Consumers' evaluations of ecological packaging – Rational and emotional approaches. *Journal of Environmental Psychology*, 37, 94–105. https://doi.org/10.1016/j.jenvp.2013.11.009

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. *Electronic Commerce Research and Applications*, 11(3), 205–217. https://doi.org/10.1016/j.elerap.2011.10.003

Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter. *Computers in Human Behavior*, 64, 575–583. https://doi.org/10.1016/j.chb.2016.07.035

Lee, H., Jai, T. M., & Li, X. (2016). Guests' perceptions of green hotel practices and management responses on TripAdvisor. *Journal of Hospitality and Tourism Technology*, 7(2), 182–199. https://doi.org/10.1108/JHTT-10-2015-0038

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468. https://doi.org/10.1016/j.tourman.2007.05.011

Liu, S., Law, R., Rong, J., Li, G., & Hall, J. (2013). Analyzing changes in hotel customers' expectations by trip mode. *International Journal of Hospitality Management*, 34, 359–371. https://doi.org/10.1016/j.ijhm.2012.11.011

Londoño, M. L., & Hernandez-Maskivker, G. (2016). Green practices in hotels: The case of the GreenLeade'rs Program from TripAdvisor. *Sustainable Tourism VII*, 1, 1–13.

Loria, S. (2014). *Textblob: Simplified text processing*. Retrieved 4 November 2020, from http://textblob.readthedocs.org/en/dev/

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4), 381–386. https://doi.org/10.1057/ejis.2013.10

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, 34(2), 336–354. https://doi.org/10.1016/j.ijresmar.2016.09.003

Mariani, M. M., Borghi, M., & Gretzel, U. (2019). Online reviews: Differences by submission device. *Tourism Management*, 70, 295–298. https://doi.org/10.1016/j.tourman.2018.08.022

Mariani, M. M., Borghi, M., & Kazakov, S. (2019). The role of language in the online evaluation of hospitality service encounters: An empirical study. *International Journal of Hospitality Management*, 78, 50–58. https://doi.org/10.1016/j.ijhm.2018.11.012

Mariani, M. M., Buhalis, D., Longhi, C., & Vitouladiti, O. (2014). Managing change in tourism destinations: Key issues and current trends. *Journal of Destination Marketing & Management*, 2(4), 269–272. https://doi.org/10.1016/j.jdmm.2013.11.003

Mariani, M., Di Fatta, G., & Di Felice, M. Understanding Customer Satisfaction With Services by Leveraging Big Data: The Role of Services Attributes and Consumers' Cultural Background. *IEEE Access.*, 7, 8195–8208. https://doi.org/10.1109/ACCESS.2018.2887300

Mariani, M. M., & Borghi, M. (2018). Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tourism Management*, 66, 47–52. https://doi.org/10.1016/j.tourman.2017.11.006

Mariani, M. M., & Borghi, M. (2020). Online Review Helpfulness and Firms' Financial Performance: An Empirical Study in a Service Industry. *International Journal of Electronic Commerce*, 24(4), 421–449. https://doi.org/10.1080/10864415.2020.1806464

Mariani, M. M., & Visani, F. (2019). Embedding eWOM into efficiency DEA modelling: An application to the hospitality sector. *International Journal of Hospitality Management*, 80, 1–12. https://doi.org/10.1016/j.ijhm.2019.01.002

Mariani, M. M., Mura, M., & Di Felice, M. (2018). The determinants of Facebook social engagement for national tourism organizations' Facebook pages: A quantitative approach. *Journal of Destination Marketing & Management*, 8, 312–325. https://doi.org/10.1016/j.jdmm.2017.06.003

Mariani, M.M., Matarazzo, M. Does cultural distance affect online review ratings? Measuring international customers' satisfaction with services leveraging digital platforms and big data. J Manag Gov (2020). https://doi.org/10.1007/s10997-020-09531-z

Mariani, M. M., Borghi, M., & Okumus, F. (2020). Unravelling the effects of cultural differences in the online appraisal of hospitality and tourism services. *International Journal of Hospitality Management*, 90, 102606https://doi.org/10.1016/j.ijhm.2020.102606

Mariani, M., & Predvoditeleva, M. (2019). How do online reviewers' cultural traits and perceived experience influence hotel online ratings?. *International Journal of Contemporary Hospitality Management*, 31(12), 4543–4573. https://doi.org/10.1108/IJCHM-11-2018-0927

Mariani, M. M., & Fosso Wamba, S. (2020). Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies. *Journal of Business Research*, 121, 338–352. https://doi.org/10.1016/j.jbusres.2020.09.012

Mariani, M., & Borghi, M. (2019). Industry 4.0: A bibliometric review of its managerial intellectual structure and potential evolution in the service industries. *Technological Forecasting and Social Change*, 149, 119752https://doi.org/10.1016/j.techfore.2019.119752

Mariani, M. (2019). Big data and analytics in tourism and hospitality: A perspective article. *Tourism Review*, 75(1), 299–303. https://doi.org/10.1108/TR-06-2019-0259

Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management*, 30(12), 3514–3554. https://doi.org/10.1108/IJCHM-07-2017-0461

Mariani, M. M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional destination management organizations. *Tourism Management*, 54, 321–343. https://doi.org/10.1016/j.tourman.2015.12.008

Naderi, I., & Van Steenburg, E. (2018). Me first, then the environment: Young Millennials as green consumers. *Young Consumers*, 19(3), 280–295. https://doi.org/10.1108/YC-08-2017-00722

Oklevik, O., Gössling, S., Hall, C. M., Steen Jacobsen, J. K., Grøtte, I. P., & McCabe, S. (2019). Overtourism, optimisation, and destination performance indicators: A case study of activities in Fjord Norway. *Journal of Sustainable Tourism*, 27(12), 1804–1821. https://doi.org/10.1080/09669582.2018.1533020

Palomino, M., Taylor, T., Göker, A., Isaacs, J., & Warber, S. (2016). The online dissemination of nature–health concepts: Lessons from sentiment analysis of social media relating to 'nature-deficit disorder. *International Journal of Environmental Research and Public Health*, 13(1), 142. https://doi.org/10.3390/ijerph13010142

Pedersen, E. R., & Neergaard, P. (2006). Caveat Emptor – Let the buyer beware! Environmental labelling and the limitations of 'Green' consumerism. *Business Strategy and the Environment*, 15(1), 15–29. https://doi.org/10.1002/bse.434

Peeters, P., & Gössling, S. (2008). Environmental discourses in the aviation industry: The reproduction of mobility. In P. Burns and M. Novelli (Eds.), *Tourism and mobility: Local-global connections* (pp. 187–203).Elsevier.

Pencle, N., & Mălăescu, I. (2016). What's in the words? Development and validation of a multidimensional dictionary for CSR and application using prospectuses. *Journal of Emerging Technologies in Accounting*, 13(2), 109–127. https://doi.org/10.2308/jeta-51615

Revinate. (2017). *Global hotel reputation benchmark report 2017*. https://learn.revinate.com/hospitality-research-studies/global-hotel-reputation-benchmark-report-2017

Reyes-Menendez, A., Saura, J. R., & Alvarez-Alonso, C. (2018). Understanding #WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach. *International Journal of Environmental Research and Public Health*, 15(11), 2537.

Rosario, A. B., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297–318. https://doi.org/10.1509/jmr.14.0380

Rutty, M., Gössling, S., Scott, D., & Hall, C. M. (2015). The global effects and impacts of tourism: An overview. In C. M. Hall, S. Gössling, & D. Scott (Eds.), *The Routledge handbook of tourism and sustainability* (pp. 36–63).Routledge.

Saarinen, J., & Rogerson, C. M. (2014). Tourism and the millennium development goals: Perspectives beyond 2015. *Tourism Geographies*, 16(1), 23–30. https://doi.org/10.1080/14616688.2013.851269

Saura, J. R., Reyes-Menendez, A., & Alvarez-Alonso, C. (2018). Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels. *Sustainability*, 10(9), 3016. https://doi.org/10.3390/su10093016

Scott, D., Gössling, S., Hall, C. M., & Peeters, P. (2016). Can tourism be part of the decarbonized global economy? The costs and risks of carbon reduction pathways. *Journal of Sustainable Tourism*, 24(1), 52–72.

Senter, R. J., & Smith, E. A. (1967). *Automated readability index*. Aerospace Medical Research Laboratories.

Thøgersen, J., Haugaard, P., & Olesen, A. (2010). Consumer responses to ecolabels. *European Journal of Marketing*, *44*(11/12), 1787–1810. https://doi.org/10.1108/03090561011079882

Thøgersen, J., Jørgensen, A. K., & Sandager, S. (2012). Consumer decision making regarding a 'green' everyday product. *Psychology & Marketing*, *29*(4), 187–197.

United Nations (2019). *The sustainable development goals report 2019*. United Nations Department of Economic and Social Affairs. https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. *Journal of Retailing*, *91*(2), 174–181. https://doi.org/10.1016/j.jretai.2015.02.005

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*(1), 2976–2977. https://doi.org/10.1038/srep02976

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Xiang, Z., Schwartz, Z., Gerdes, J. H., Jr., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120–130. https://doi.org/10.1016/j.ijhm.2014.10.013

Yang, Y., Park, S., & Hu, X. (2018). Electronic word of mouth and hotel performance: A meta-analysis. *Tourism Management*, *67*, 248–260. https://doi.org/10.1016/j.tourman.2018.01.015

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180–182. https://doi.org/10.1016/j.ijhm.2008.06.011

You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, *79*(2), 19–39. https://doi.org/10.1509/jm.14.0169

Yu, Y., Li, X., & Jai, T. M. (2017). The impact of green experience on customer satisfaction: Evidence from TripAdvisor. *International Journal of Contemporary Hospitality Management*, *29*(5), 1340–1361. https://doi.org/10.1108/IJCHM-07-2015-0371

Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, *33*(2), 456–481. https://doi.org/10.1080/07421222.2016.1205907

Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, *76*, 111–121. https://doi.org/10.1016/j.ijhm.2018.03.017

Appendices

Appendix 1 – **Research design framework**
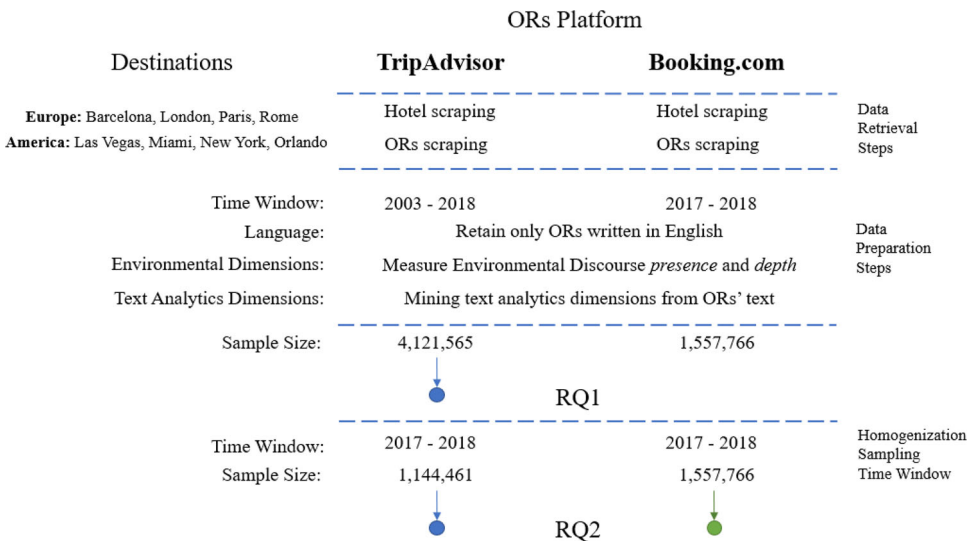
Appendix 2 – **Hotel reviewers' share by continent of origin**

Destination Continent of Origin	Barcelona	Las Vegas	London	Miami	New York	Orlando	Paris	Rome
Africa	0.97	0.39	0.79	0.43	0.66	0.35	1.24	0.70
Asia	6.99	2.70	5.22	2.17	4.09	1.49	7.37	6.37
Europe	64.84	16.25	71.17	18.12	32.79	16.75	59.55	61.05
Non-EU Europe	0.56	0.09	0.26	0.17	0.20	0.06	0.45	0.49
North America	19.68	75.43	17.16	65.72	54.33	74.29	23.41	23.73
Oceania	3.09	3.16	3.84	1.26	3.99	1.29	4.52	4.07
South America	3.87	1.98	1.56	12.13	3.94	5.77	3.46	3.60