SOCLIMPACT

Downscaling climate impacts and decarbonisation pathways in EU islands, and enhancing socioeconomic and non-market evaluation of Climate Change for Europe, for 2050 and beyond

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

WORKING PACKAGE 5

MEASURING MARKET AND NON-MARKET COSTS OF CLIMATE CHANGE AND BENEFITS OF CLIMATE ACTIONS FOR EUROPE

Deliverable 5.3. Data Mining from Big Data Analysis.

Coordinated by UNIBO (Anastasia Arabadzhyan, Paolo Figini and Laura Vici) with the participation of ULPGC and other SOCLIMPACT partners of WP5.

Revised version – Version v2.0 (reviewed in line with comments outlined in the letter: Ref. Ares(2021)4638303 - 18/07/2021)

Type of deliverable: R (Report)

Confidentiality level: PU

Revision History

Version	Date	Author	Description
v0.1	21.12.2019	A. Arabadzhyan	Internal first draft
v0.2	07.01.2020	L. Vici	Internal second draft
v0.3	12.01.2020	P. Figini	Final draft to be reviewed according to the quality review internal process
v1.0	31.01.2020	A. Arabadzhyan, P. Figini, L. Vici	Intermediate version
v1.1	21.02.2020	A. Arabadzhyan, P. Figini, L. Vici	Final version to be submitted and published
v2.0	13.08.2021	P. Figini, L. Vici	1. An executive summary has been added at the beginning of the report (pp.4-5) to facilitate reading and better link D5.3 with the other deliverables. 2. A paragraph has been added in the Introduction (p. 6) to explain better that the report does not consider the other Blue Economy sectors but that the extension of this approach to other sectors is promising. 3. Three paragraphs have been added at the end of Section 2 (p. 10) to defend our approach and to firmly locate it within the boundaries of the socio-economic analysis of Big Data. 4. A new section (Section 5.3) has been added at the end of the report (pp. 49-50) to wrap up the scientific discussion and recommend actions for future research.

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

0. Executive Summary

This Report is the main outcome of Task 5.3, which aimed at experimenting with different types of large datasets related to the tourism sector and characterised by velocity, variety, and volume, hence sharing the typical properties of Big Data. The task's main goal was to complement findings of WP5 collected through more established techniques (surveys and questionnaires) and methodologies (choice experiments). Within the allocated time, financial, and human resources, this task selected and pursued three different lines of investigation, aiming at tackling the following research questions:

- How do climate events (like heat waves or storms, proxied by weather conditions and forecasts) impact on accommodation prices? This analysis considered Corsica, Sardinia and Sicily.
- How do climate events impact on destination image and on the activities carried out by tourists in the destination? This analysis considered four islands of the Canary archipelago (Gran Canaria, Fuerteventura, Lanzarote and Tenerife) and four Mediterranean islands (Crete, Cyprus, Malta and Sicily).
- How do climate events (like forest fires) impact on hotel performance? This analysis considered Gran Canaria.

Once estimations of impacts of specific events on prices, performance, and destination image were generated, climate change scenarios prepared by WP4 were used as inputs, thus allowing for the computation of the economic impact of climate change on the tourism sector (measured in terms of variation in yearly tourism expenditure), to be later transferred to WP6.

Findings show that:

- A one-degree increase in expected average temperature is associated with an increase in prices of 4.7% in Sardinia, 3.2% in Corsica, while prices are not significantly affected in Sicily. Similarly, a one-mm increase in daily rain is associated with a decrease in prices of 1.7% in Sardinia, 1.9% in Corsica; again, prices in Sicily are not affected. Such spatial heterogeneity is likely due to the specialization of the destination (a partially cultural destination like Sicily is the least affected by weather) and different revenue management strategies adopted by local hotels.
- By assuming that prices instantly adjust to balance demand and supply, and hence keep constant occupancy rates, projections from WP4 (RCP2.6 and RCP8.5 for 2050 and 2100) can be fed into our estimates to forecast that summer prices (and tourism expenditure) will rise by 0.5 – 2.4% in Corsica because of future changes in human comfort index, by 0.8 – 3.7% in Sardinia, while they will drop by 0.1 – 0.5% in Sicily because of the increase in the thermal stress associated with climate conditions.
- In the second line of investigation, machine learning software of visual object recognition scanned hundreds of thousands of pictures posted by tourists on Instagram to develop metrics able to quantify the concept of destination image, as it is perceived by visitors. The change in the metrics when specific climate events hit a destination imply a change in the relative quality of the holiday experience, with important consequences in terms of tourists' behaviour and expenditure patterns.

We applied this approach to the case-study concerning the huge, devastating fires that took place in Gran Canaria in Summer 2019, finding that tourists who posted fire-related content were (about 5%) more likely to express negative sentiment, as well as tourists who happened to be closer to the epicentre of the fires. Since the literature shows that negative emotions are diminishing satisfaction, willingness to revisit, and to recommend the destination, wildfires generate a negative impact on the destination that can be precisely measured. For example, we estimated that wildfires decreased demand for nature-based activities on Gran Canaria of about 10%.

- Although forest fires negatively impact the destination image, the third line of investigation showed that they did not significantly affect hotel performances (always considering the case-study of fire outbreak in Gran Canaria in summer 2019). This is somehow consistent with the literature, suggesting that tourists are only marginally affected by forest fires in beach destinations. A more careful investigation into the robustness of this result is needed, as there are confounding effects: fires just occurred at the peak of the summer season, and accessibility to remote islands as the Canaries implies very high costs of rescheduling the holiday.

This report explains the methodologies used in the different lines of investigation, the estimation of parameters, and how to connect them with scenarios of climate change produced by WP4 to produce estimations of changes in tourism expenditure in the different islands under scrutiny. Finally, the limitations of this work and the potentialities of Big Data analysis for research on the socio-economic impact of climate change are presented and critically discussed.

1. Introduction

SOCLIMPACT project is built upon the concept of impact chains. Climate Change (CC) physically affects the local environment (in terms of air and water temperature, beach extension, marine and land biodiversity, etc), consequently changing the value of the environmental services exploited by different sectors of the blue economy. Within the project, WP5 mainly aims at estimating the (non-market) value of these environmental services for the islands under investigation, in order to assess, through a stated preference approach, the costs and benefits of adaptation policies. The present Report D5.3, outcome of task T5.3, complements the WP5 analysis by taking a different approach. Rather than considering the hypothetical choices of citizens in front of different scenarios (in a "what if" paradigm, as done in T5.4), we look at real behavioural changes of tourists when events that can be associated to CC appear. In other words, we look at real reactions of individuals, in a sort of revealed preference approach. To undertake this task, we have required the help of Big Data (BD) in a sort of massive "field experiment" implemented during the 2019 tourism season. Data mining and Big Data analysis are among the most recent tools attracting increasing attention of both academic scholars and policymakers in various areas of study. This is particularly promising in the field of tourism research, where we can turn to Big Data to construct and analyze the dynamics of indicators of tourism behavior and expenditure that are not available as conventional ready-to-use quantitative measures.

Given the relevant importance of environmental services for tourism, task T5.3 (and hence this report) only investigates this sector, for many reasons. Firstly, it is an extremely important contributor to the islands' economies and a driver, to a large extent, for demand in other sectors under study in the project (e.g. transport). Secondly, in all its stages (before, during and after consuming the tourism product) the very nature of tourism transactions leaves a massive online trace, which naturally makes it a great candidate for being studied through the prism of Big Data. Thirdly, this is a sector where demand (and not only supply) is directly affected by CC. In a nutshell, CC modifies the environment and thus the quality of the holiday experience directly perceived by tourists at the destination. A shift in quality can trigger a change in the physical volume of tourism (tourists will not arrive or modify their length of stay because of CC), a change in price (as demand and supply would have to readjust to the new quality level) or a change in both, which is the likely outcome in the real world. Should we succeed in estimating the impact of different CC events on the quantity and the price of tourism, we would be able to provide WP6 with a (partial) answer to a key question for the whole SOCLIMPACT project: **how much is tourism expenditure changing because of CC in the destination?** Although this task does not focus on the other Blue Economy sectors, the extension of BD analysis to Maritime transport, Aquaculture and, especially, the Energy sector is possible and likely to produce innovative and interesting findings, particularly if advanced BD analytics are applied. We will return on this in the final considerations.

Moving from the intuition of this approach to the real evaluation, many caveats apply. One, tourism is a composite bundle, including many non-market services (environmental and cultural resources) together with market services (accommodation, leisure activities, food & beverage, etc.): a comprehensive analysis of all activities would be a cumbersome task to carry out: hence, we focus only on few specific services, selected according to data availability. Two, CC impacts different territories differently, each one responding according to its own specificity and the type of tourism hosted. A particular care would hence be needed in the act of transferring values herein estimated to other territories. Three, different events linked to CC (storms, heat waves, forest fires, etc.) affect tourism differently, hence all the impact chains unfolded in SOCLIMPACT should be investigated.

It is easy to understand that a full coverage of all the CC impacts for all the islands and all the tourism services goes well beyond the time and human resources allocated to SOCLIMPACT.

Accordingly, the main target of this task is to experiment for specific cases / scenarios / activities in order to complement results for those channels of transmission from CC to tourism where stated preferences applied in T5.4 hardly work. When possible, this task can also provide robustness checks to Choice Experiment estimations carried out in task T5.4. Given that analyses of T5.3 are not vital for the advancement of the project, this Report has had the freedom to experiment a few lines of investigation which are not only useful for SOCLIMPACT but also methodological innovative and challenging for the characteristics and amount of data to be considered.

The report is structured as follows. Section 2 introduces and classifies Big Data, highlighting advantages and disadvantages of their use, together with their relevance for tourism. Section 3 introduces the (three) lines of research that we have decided to investigate for SOCLIMPACT project: for each one, we unfold research questions and aims, data that have been collected and analysed, and methodologies. Section 4 keeps the same structure (one sub-section for each line of investigation) and summarises the main findings. Section 5 recalls and discusses the main conclusions and sets the table for the estimation of values to be transferred to WP6, something that will be anyway the subject of another report (D5.6).

2. Big Data: definition, classification, and applicability to the tourism sector

The term Big Data (hereafter BD) relates to all types of data, including structured and unstructured unconventional data, which can be collected, stored, and processed following the rapid growth of information technologies and computing power. BD are everywhere, and they are currently exploited by a plethora of actors (individuals and organizations) for many reasons: from companies for managerial purposes to local governments for the development of smart cities; from intelligence agencies for security reasons to households for purchasing decisions. BD are the consequence of the development of Information and Communication Technologies (ICTs) and the by-product of the adoption of networked devices (such as smartphones) by a growing number of consumers and businesses on a global scale.

Their properties are often described by 3 Vs: *Velocity*, implying that they are available at very high frequencies, often in real time; *Variety*, as they may come not only in numeric form, but also as text, images, audio and video content; and as a result, *Volume* in terms of terabytes but also in terms of quantity of observations available for statistical analysis. BD merge data generated in real time by interconnected devices such sensors recording climate information, satellite imagery, digital pictures, texts and videos, prices and other economic variables recorded in electronic markets.

BD are the object of analysis of Data Science, a multidisciplinary field involving mathematics, statistics and computer science. Data science builds on data retrieval, preparation, statistical analysis, modelling and machine learning to investigate issues in several settings. However, BD are also a prominent challenge for business and economic research, and social sciences have recently acknowledged their importance, embracing their use in current research. In fact, BD enable to address novel research questions, to develop innovative research designs and to use new methodologies and approaches.

Methodologically, BD allow researchers to overcome the difficulties of inferring general conclusions when working with samples of the population. In fact, BD virtually allow working with the whole population under scrutiny, without the lags stemming from the availability of census data. Consequently, it is possible to design large-scale "in-vivo randomized" experiments (Aral & Walker, 2012). At the same time, BD permit a more efficient detection of patterns and trends, triggering the identification of new statistical and econometric tools (Varian, 2014; Einav & Levin, 2013, Doornik & Hendry, 2015), not without controversies (Annals of the American Academy of Political Science, special issue, May 2015).

As regards contents, the use of BD in social sciences has recently surged and triggered investigation in many fields of study, such as behavioural economics, spatial economics, revenue management and web marketing. Relevant literature tackles, among other issues: how information and product adoption spread over a network; what is the relevance of social influence bias in the distribution and the dynamics of opinions; the rise of online rating systems as reputation-based authorities; the use of BD analytics for forecasting purposes; the effectiveness of web marketing strategies and their impact on the regulation of markets.

Overall, BD are one of the most interesting frontiers and challenges for economic and business research. However, these advantages come at costs. Most importantly, there is need to apply specific techniques to clean and extract value from the large amount of data at hand, as some datatypes may contain a lot of noise or features that are redundant for the questions of interest. Additionally, large data volumes require storage repositories with sufficient capacity. Finally,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

econometric methods for high dimensional problems should be applied, and this also requires employing advanced human skills and machines with significant computational power.

The Tourism and Travel sector is undoubtedly one with a large variety of applications where Big Data could provide new insights. For instance, tourism is the economic sector with the largest share of e-commerce in OECD countries, and is the one where some of the most important ICT innovations in terms of business models, peer-to-peer (P2P) platforms and revenue management strategies were experimented (among the many, consider businesses such as Booking.com, TripAdvisor.com, BlaBlaCar, Uber, Airbnb). It is then not surprising that most of the economic and managerial research with BD revolves around tourism and travel, the way travellers access information, plan and book trips, and subsequently share their travel experience through social media and review systems. The table below summarizes different datatypes that are used in tourism literature (Li et al., 2018), their properties and potentialities. First, internet users' activity generates a lot or relevant content, such as reviews, travel blogs, and posts in social networks like Twitter or Instagram. This type of data is pulled together under the term User-generated content (UGC). Secondly, very useful information may be obtained from devices, such as GPS and mobile phone tracking data. Finally, the last category is transaction data, which comprises web search data and similar types. All these data have their strength and weaknesses.

Table 2.1. Data types, their properties and potential use.

	Data types	Data Sources	Privacy concerns	Costs	Quality	Applicability
User-generated	textual, photo, video, etc	Social networks: Twitter Instagram Facebook Reviews: Tripadvisor Booking AirBnB	no	low	low	sentiment ⇒ destination image
Device	spatio-temporal	GPS, mobile roaming, Wi-Fi, Bluetooth, etc	yes	high	high	tourist movements and revisits ⇒ tourist loyalty
Transaction	online booking data, hotel costs data	service providers	yes	high	high	good proxies for demand and supply
	websearch data	Search engines (Google, Baidu, etc)	no	low	medium	proxy for tourist demand

As regards UGC, they are openly available (unless restricted by privacy settings) and relatively easy to obtain (some services provide API – Application Program Interface, thus facilitating data collection). Note, however, that generally the easier to get the data, the more effort is required to clean and process them. From social networks and review websites, a wide range of useful information can be retrieved: sentiment or attitude (if a person writes positive or negative reviews/comments) towards a destination, accommodation, attraction, or event. It is also possible to assess which dimensions or features of a destination are most valued by visitors. Platforms that combine different datatypes are of particular interest: for instance, Instagram provides imagery content, complemented by captions to posts and comments from other users, which are textual data, and even geographic coordinates if users chose to geotag the post, therefore, allowing for spatial and temporal analysis.

As opposed to UGC data, device data are more problematic to obtain. Specifically, mobile phone data must deal with privacy protection issues, requiring an agreement with mobile phone companies. Potentially, it could allow pinning down individuals and their movements on a very fine-grained scale, which is of importance in the context of islands, allowing tracking tourists' routes and the choice of activities and attractions visited within a destination.

Transaction data combine the properties of UGC and device data as they are quite heterogeneous. If we consider search engines data, they are very easy to obtain, but their use is limited to serving as a demand proxy, and quality drops as search terms become more specific. At the same time, online booking websites can provide high quality data on prices of accommodation in different locations, but obtaining them requires developing specific web scraping routines, and data collection process is likely to exhibit interruptions and require timely adjustments in case of changes in website structure.

In this *mare magnum* of data and research possibilities, time, financial and human resources available in SOCLIMPACT drove us to make important choices as regards the data selected and the methodologies used, always justified by the research questions of interest. In this respect, it is important to highlight that our main contribution was not to develop new techniques, or use cutting-edge methodologies (e.g. advanced machine learning, network analytical methods), but identify the combination of data and methodology to advance the analysis of D5.3 in a useful way for the project as a whole.

To further clarify, we followed a bottom-up strategy, first identifying those large datasets that share the main properties of BD (the 3V recalled above – volume, variety, velocity) and that could provide relevant insights into the relationship between CC and tourism. Secondly, since BD can be mined and analysed in many ways and using standard and/or novel approaches, the choice of the methodology was driven by the research questions to tackle. In this perspective, since there is a lot of variety in the approaches taken by different disciplines (economics, computer science, marketing) and with different aims (research, business intelligence, dissemination), we point out that a mix of techniques and approaches was used. When the aim was to analyse cause-effect relationships between socio-economic variables the good old econometric techniques were still considered the best option. Moreover, machine learning algorithms of visual object recognition (although not developed in-house) were at the base of Section 3.2, investigating the content of images posted on Instagram. Again, text mining and sentiment analysis were used in the same section, when investigating the reaction of tourists to climate events. Finally, different econometric techniques (standard and advanced) were used throughout the deliverable in other sections (3.1 and 3.3).

Overall, the aim of this deliverable was not to develop new techniques but to apply existing techniques to relevant sources of BD. Hence, cutting-edge BD analytics, which are the core in Data Science, were not used and our investigation departs from what has been recently called "big data analytics" but is clearly within the boundaries of the analysis of big data. Having said that, the specific lines of investigation on which we decided to focus, the datasets, and the methodologies applied are described in Section 3.

3. Methodology and Data

As discussed above, given the wide variety of data types, destinations, climate events and tourism activities, in this Task we decided to focus on specific issues that would be useful for WP6 and, at the same time, innovative for their contents and methodologies. The choice of data to collect and analyse was driven by several considerations. First, we aimed at those tourism indicators which are the drivers of impacts pinned down by Impact Chains analysis conducted within WP3. Second, the weather and climate variables to be included in the econometric modelling were determined based on which Impact Chains were operationalized and for which climate indicators projections were provided by WP4. Third, given the experimental nature of this Task, we studied different methodologies, data and islands in order to identify what are the most promising lines of investigation to be applied in future extensions of the project.

Accordingly, we decided to proceed with four lines of investigation, one of which was aborted due to the impossibility of getting the data.

- **How do climate events impact accommodation prices**? When climate conditions are expected to change, tourists adjust their behavior: they might switch destination, or they might stay longer or shorter depending on the type of climate event. In turn, these behavioral changes will modify the market equilibrium, pushing hotels to adjust their prices to re-establish the equilibrium between demand and supply. The goal of this sub-project was to complete the following statement: "**when climatic conditions are expected to change by X, prices change by Y**", where X can be, for example, a heat wave, a storm and Y can be expressed in absolute or percentage terms depending on the estimation technique.

- **How do climate events impact destination image?** As evident from Impact Chains development (D3.2), out of 9 specific impact chains for tourism, 5 explicitly mention deterioration of destination image among the demand-side impact. Deterioration of destination image affects the quality of the overall tourism experience and triggers different sentiments, implying behavioral changes. The goal of this sub-project was to complete the following statement: "**when the climate event X happens, sentiment changes by Y**", where X can be, for example, a forest fire, an outbreak of a disease, and Y can be expressed in absolute or percentage terms depending on the estimation technique. This estimation can be later matched with other information (coming from other Tasks or from the literature) to learn how the sentiment triggers changes in tourism expenditure and in demand for different types of activities.

- **How do climate events impact hotel performance?** Similar to the first research line, we focus on the accommodation sector, in order to have a more comprehensive view of the impact of CC events on the hotel management. As shifts in demand affect not only prices but also physical quantities (arrivals and overnight stays) the aim of this line of investigation was to complete the following statement: "**when the climate event X happens, hotel performance changes by Y**", where X can be, for example, a forest fire or an algae invasion, and Y can be one of the main indicators of hotel performance: average daily price, revenue per available room, occupancy rate.

- **How do climate events impact tourists' mobility?** The idea behind this line of investigation was to apply Mizzi et al. (2018) to monitor where tourists stay, and how they modify their movement when the destination is hit by specific climate events (storms, heat waves, etc.). This is fundamental knowledge for destination management as it is likely that tourists would substitute cultural with beach activities when, for example, there is a heat wave.

In turn, pressure on specific sites and on the mobility system could be forecasted and efficiently managed. In order to carry out this research, we needed a specific type of mobile tracking data from phone companies: unfortunately, the company we collaborate with (TIM) is not using this specific system in the two Italian islands which are focal points of SOCLIMPACT project (Sardinia and Sicily), but only for some regions in Northern Italy. Hence, we had to abort this line of research, although the methodology we adopted for destination image analysis can partially make up for the absence of mobile data.

A summary representation of the lines of research is reported in Table 3.1. Once estimations of impacts of specific events on prices and quality are generated, climate change scenarios prepared by WP4 can be used as inputs, thus allowing for computation of the economic impact of climate change on the tourism sector (mainly in terms of variation in tourism expenditure, but also in arrivals), to be later transferred to WP6. However, this is not the aim of this specific task (and report) but an indication of how Task 5.3 can be useful to the whole project. A thorough work of transferring values to other islands and to macroeconomic models will be undertaken in Task 5.5 (and Report D5.6).

Table 3.1 – A summary of the research lines investigated

Line of research	Impact chain analysed	Research Question	Data Sources	Islands under investigation
1	2.1 Impact of thermal stress; 3.2 Impact of storms	When climatic conditions are expected to change by X, how much do prices change?	Booking.com Accuweather.com Weather.com Ilmeteo.it	Corsica Sardinia Sicily
2	1.2 Impact of fires 2.1 Impact of thermal stress	When the climate event X happens, how much does sentiment change?	Instagram Official weather archives	Canary Islands Crete* Cyprus* Malta* Sicily*
3	1.2 Impact of fires 2.1 Impact of thermal stress	When the climate event X happens, how much does hotel performance change?	STR-Share	Canary Islands Mallorca* Malta*
4	1.2 Impact of fires 2.1 Impact of thermal stress 3.2 Impact of storms	When the climate event X happens, how is the mobility of tourists affected?	Mobile phone tracking data	Aborted due to impossibility to get the data

* To be analysed in future elaborations: results are not available for this report.

A more detailed description of the data and the methodologies for the first three lines of research now follows.

3.1. How do climate events impact accommodation prices?

The first question relates to the supply side of the tourism sector and aims at estimating the impact of climate events on hotel prices. The proxy to measure climate event is the weather condition, as weather is typically considered the short-term manifestation of climate (Becken, 2012). As many other services, tourism is characterised by a pattern of advanced booking: services are put on sales before the actual consumption of the service (check-in date) and hence for each date of consumption we have to consider demand and supply for each point in time in the advance booking window. This simple consideration opens a set of related decisions to take as regards data to collect.

First, the current standard in accommodation industry is to use revenue management strategies which imply, among other things, the use of temporal price discrimination (i.e., the price of the same service changes according to the time when it is booked), such as first-minute or last-minute pricing. Accordingly, prices have been obtained through the Booking.com platform, which is the largest accommodation booking website worldwide[1]. A scraping routine was developed to collect all prices posted on Booking.com for all the available rooms of selected hotels for a one-night stay for lead times (i.e. the difference between the day when the price is posted and the check-in date) from 0 to 15 days. Moreover, all the other relevant characteristics of the booking option were recorded: room type, room characteristics, reservation features (free cancellation option, board type, number of available rooms, etc).

Second, if our goal is to check whether climate has an impact on prices, we have to merge Booking.com prices with data on weather. Consistently with Figini et al. (2019), we consider weather forecasts: Figini et al. (2019) intuition is based on a theoretical model where forecasts are common knowledge and can be observed by both accommodation service providers and consumers. As the quality of the leisure experience is affected by weather conditions, weather forecasts (as a signal for expected quality) affect the customers' willingness to pay and hence equilibrium prices: an improvement in weather forecasts would increase the expected quality of the experience, and drive the supply to instantaneously adapt the price, the effect being stronger the higher the level of accuracy of the forecast and the larger the degree of ex ante uncertainty in weather conditions.

Building and extending upon Figini et al. (2019) theoretical framework, we collected data from 0 up to 15 days in advance from three main weather forecast providers: the international providers *Weather.com* and *Accuweather.com*, which feed data for the default weather apps on iOS and Android-based devices respectively, and the local provider *Ilmeteo.it*, which is the most popular private weather forecast provider in Italy. We chose to use several weather data providers to check robustness of results and to ensure that in case of different quality of forecasts both international and domestic tourist segments are represented.

Third, the collection of data for all the islands would have been beyond the human and technical resources available within the project, hence we have narrowed this line of investigation to three islands: Sicily (more precisely, its south-east coast), Sardinia (its north-east coast) and Corsica (its southern coast); the choice of these specific destinations was motivated by several factors. Sicily and Sardinia are among the islands with the largest number of hotel establishments registered on *Booking.com*, whereas Corsica, on the contrary, is a region with a relatively low number of establishments appearing on *Booking.com*; therefore, we aim to contrast evidence from different types of destinations, to be able to make more robust inference when extrapolating results for other islands. At the same time, the three destinations belong to two different countries, thus facing different institutional and legal frameworks, but on the other hand are close to each other geographically, which implies that they are not facing systematically different weather conditions in a given period of time; this would allow attributing any variation in results between destinations

[1] By number of website visits in August 2018, booking.com was the most visited booking website after TripAdvisor (https://www.geckoroutes.com/travel-insights/most-popular-travel-booking-websites-apps/). However, among websites focusing specifically on accommodation, booking.com is the world leader.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

to factors other than weather parameters. In addition, apart from these islands, we selected several destinations which are qualitatively different from the islands in terms of tourism mix and accessibility conditions: two leisure mainland destinations (Riccione and Jesolo-Cavallino), a business destination (Milan), and a cultural destination (Venice). This allows testing sensitivity of findings to other important characteristics of a destination and make well-grounded generalizations later on. At the second step of selection, from the pool of 3, 4 and 5-star hotels a random sample of about 1,300 accommodation structures was constructed.

Overall, we built a dataset consisting of rooms available for bookings in 1,313 hotels in the period between April 29th, 2019 to June 30th, 2019.[2] Out of 1313 properties in the sample 743 are located at the seaside, and 570 in the city area. Figure 3.1 depicts location of hotels in the city areas and over beach mainland destinations (Milan, Venice city, Jesolo/Cavallino, Riccione), while in Figure 3.2 location of hotels in Sicily, Sardinia and Corsica are plotted.

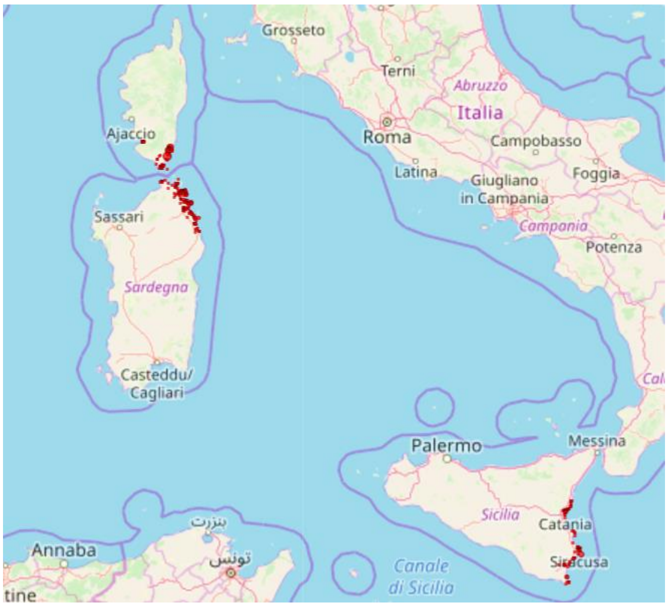
Prices are those posted by hotels for a one-night stay and check-in date from 0 to 15 days in advance from the search date. Extensive information on various room and booking characteristics is used to construct a set of control variables. This dataset is then merged with the information about weather forecasts provided by the three providers for temperature, precipitation and any other relevant weather parameters. We use the *Weatherchannel.com* forecasts as baseline, while *Ilmeteo.it* and *Accuweather.com* forecasts are used for robustness checks.

Figure 3.1. Location of hotels in mainland cities and seaside destinations

[2] In a future analyses, the whole summer period (29th of April to 31st of October, 2019) will be considered.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Figure 3.2. Location of hotels in islands



The final dataset is a panel consisting of about 8 million observations, and the question of interest is then investigated via regression analysis, consistently with the hedonic price model with dynamic characteristics developed by Figini et al. (2019). Such a large dataset not only allows for robust inference and high precision of estimates, but also makes it feasible to include a very large set of independent variables. For this reason, we include a wide range of covariates, most of which in the fixed effect form. The baseline specification of the hedonic model is the following:

$$lnPrice_{rt\Delta} = f(hotel_{FE}, room_{FE}, \Delta_{FE}, search_{day_{FE}}, checkin_{day_{FE}}, breakfast_{FE}, FreeCanc, roomsleft, \#sleeps_{FE}, view, weather_{t\Delta}) + \varepsilon_{rt\Delta} \quad [1]$$

where $lnPrice_{rt\Delta}$ is the natural logarithm of price for room of type *r* observed at date *t* for a check-in date $t + \Delta$. Controls include hotel fixed effect ($hotel_{FE}$), room type fixed effect[3] ($room_{FE}$), booking lead fixed effect (Δ_{FE}), day of week fixed effects for search date t ($search\_day_{FE}$) and check-in date $t + \Delta$ ($checkin\_day_{FE}$), board type[4] ($breakfast_{FE}$), indicator of free cancellation option available ($FreeCanc$), number of available rooms[5] ($\#roomsleft$), number of sleeps fixed effect ($\#sleeps_{FE}$), indicator of "room with a view" ($view$). The main variable of interest in this model is $weather_{t\Delta}$, which is the forecast of the weather parameter observed at date *t* for the date $t + \Delta$.

3.2. How do climate events impact destination image?

As previously recalled, it is evident from Impact Chains development (D3.2) that, out of 9 specific impact chains for the tourism sector, 5 explicitly mention deterioration of destination image among the demand-side impact. Given the importance of this factor, we chose to study in detail how destination image may be affected by weather and climatic events.

[3] 61 room types were manually coded.
[4] 4 options were possible: breakfast not included, breakfast included, half-board or full board.
[5] If contains "only x rooms left on our website", with $1 \leq x \leq 5$.

The concept of destination image has received a lot of attention from the literature, and various studies may define it differently. It is generally accepted to consider it as a composite multidimensional concept represented by a set of image-forming characteristics of the destination which visitors find the most valuable and/or intrinsic to a given location. It is usually split into three different spheres: the cognitive, the affective and the conative sphere (Baloglu and McCleary, 1999; Pike and Ryan, 2004; Agapito et al., 2013). The cognitive image refers to knowledge and beliefs related to the attractions to be seen, expected experiences to remember, and to the general environment of the destination (weather conditions, accommodation structures, attractions, security, health conditions, accessibility, etc.). The affective sphere is related to feelings and emotions that can be triggered by the different characteristics of the destination (Beerli and Martin, 2004). The conative image is consequential to the previous two spheres and refers to the behavioral intentions of tourists regarding future activities.

The majority of studies have been using survey data to analyze destination image. Survey data, however, do suffer from several drawbacks, such as framing effects, limited number of observations, and most importantly, they usually relate to a single destination at a single point in time (at best, only a couple of destinations or two time periods are represented in a single study). The availability of large amounts of user-generated data makes now possible to study destination image from a novel and different perspective. Most of the current research on the impact of social media on destination image formation has focused on textual messages by using content analysis (Költringer and Dickinger, 2015; Chua et al., 2016; Xiang et al. 2015). On the contrary, just a few works analyse the content of tourists and DMOs' photos (Jiang et al., 2013; Stepchenkova and Zhan, 2013; Miah et al., 2017; Deng and Li, 2018). Moreover, the analysis of pictures posted online has mainly aimed at detecting tourists' behavior, tourism recommendations, perceptions, travel routes and trip duration (Kurashima et al. 2013; Okuyama and Yanai, 2013; Lee et al. 2014); only in a very few cases the focus is the destination image. One of the contributions in this direction is a study by Stepchenkova and Zhan (2013), which refers to the case of Peru to investigate differences in the image promoted by the DMO and perceived by tourists in their pictures posted on Flickr. Their approach, however, relies on manual inspection of about 1000 images, focusing on a single destination, and destination image is not studied in dynamics.

We build upon Stepchenkova and Zhan (2013) to investigate the images that tourists post when they are on holiday in a given destination. Our aim is to proxy the concept of destination image with the features that are most popular in the "aggregate picture" shot by the community of tourists. Accordingly, we can also investigate how these features change when some specific climate events hit the destination, and how tourists react. To operationalize this perspective, we first have to come out with a measure of destination image, and then look at how this measure changes overtime with climate events. Secondly, we have to find a way to translate pictures into relevant data to be statistically analyzed.

To do so, we propose **using the universe of travel-related content from Instagram pictures to construct destination image proxies**[6] for a given set of destinations. First, we collect posts published on Instagram in the period June - September, 2019, containing the photo(s), the name of the destination in the hashtag, as well as travel-related keywords in the caption (this is done to focus on posts of tourists and thereby excluding residents). Since we aim at obtaining a vector of feature occurrences for each of the destinations on a very fine-grained scale, we need to process several hundred thousand of photos, which, of course, cannot be done manually. We therefore resort to Google's Cloud Vision API (GCV), a tool based on powerful machine learning

[6] Our approach, therefore, captures the cognitive dimension of destination image, as it is perceived by tourists.

algorithms such as convolutional neural networks which perform various image processing tasks: label recognition, facial detection, object localization and others. From the available GCV options we selected label recognition: as opposed to object localization, which yields a list of objects that appear in the image, label detection is more general and returns a set of a broader group of categories: image labels can identify objects, locations, phenomena, activities, specific plant and animal species, products, and even more generic and abstract categories (e.g. *love, friendship, tourism*) which appear in the picture. GCV processes single images and returns an output with a list of up to ten labels and related scores, which take values from 0.5 to 1 and indicate the degree of accuracy matching the label with the given image. After obtaining a set of labels for each picture, we then construct the collective image of each destination, and analyze how it differs across destinations, and how it changes in response to weather conditions and climate events.

More specifically, we define the image of a given destination *i* as a time-varying vector (with a time span from 1 to T) of label frequencies for a fixed set of K labels. In matrix notation, we write down a *K* x *T* matrix as in [2]:

$$Image_i = \begin{pmatrix} freq\_lab_1^1 & \cdots & freq\_lab_T^1 \\ \vdots & \ddots & \vdots \\ freq\_lab_1^K & \cdots & freq\_lab_T^K \end{pmatrix} \qquad [2]$$

In the most general case, *K* may include all possible labels that have ever appeared in the imagery during the given time span. Alternatively, labels can also be grouped in broader categories. Frequencies may refer to the ratio of the number of occurrences of label *k* to either the total number of labels, or to the number of pictures, or to the number of posts at time *t*. Finally, the degree of time disaggregation has to be set, to determine whether the time periods from 1 to *T* are hours, days, weeks, or months. We can tailor these parameters in accordance with the question of interest.

An advantage of our approach is that this measure is flexible and can in principle be constructed for any destination at any point in time, thus resolving the main drawbacks of conventional survey data method, which rarely covers several destinations and several periods in a single study. Such a property of the proxy we construct allows for a rigorous comparison of images of different destinations, as well as for studying short-term image dynamics driven by variation in activities undertaken by tourists in a destination. This, in turn, can help understanding which risks are more important for a given destination.

Once we compute the labels frequency, we can also rank labels for each island and use the ranking to construct a metrics of "destination image distance", that is, to which extent the image of one destination differs from another. Whereas many different concepts of distance can be considered, the simplest metrics is proposed in this analysis and reported in [2a]: the average squared rank distances of the set of top *N* labels (we chose *N*=20 in our analysis). The Index of Distance in Destination Image (IDDI) is hence introduced as:

$$IDDI_{ij} = \frac{\sum_{k_i}(rank_{ik_i} - rank_{jk_i})^2 + \sum_{k_j}(rank_{jk_j} - rank_{ik_j})^2}{2} \qquad [2a]$$

The proposed index measures distance in images of destinations *i* and *j*, based on the first *N* labels sorted by number of occurrences, so that $k_i = \{label_{1i}, \ldots, label_{Ni}\}$. An alternative index may consider absolute instead of squared distances, as using squared rank differences inflates the resulting distance between images if a pair of labels has very different ranks.

Note that such a metrics becomes meaningful only when there are three or more destinations in comparison. We further suggest normalizing the absolute distance between destinations to the smallest in the sample, so that:

$$\widetilde{IDDI_{ij}} = \frac{IDDI_{ij}}{\min\{\overrightarrow{IDDI}\}} \qquad [2b]$$

The absolute and relative frequency of labels in Instagram posts can also be used to determine how external events (for example linked to extreme weather conditions stemming from climate change) affect behaviors and activities undertaken by tourists, possibly influencing the destination image.

As regards the choice of the social network to scrape data from, we opted for Instagram for several reasons. First, Instagram is one of the fastest growing social networks: while Snapchat's daily user count and Facebook's monthly count grew by 2.13% and 3.14% in Q1 2018 respectively, Instagram is growing by about 5% per quarter and from 300 million monthly active users recorded at the beginning of 2015 it has reached 1 billion by June 2018 (https://techcrunch.com/2018/06/20/instagram-1-billion-users/). Second, as opposed to various text-based platforms, the core of Instagram are pictures, which not only have a more profound impact on people's opinions, preferences and perceptions (Hirschman, 1986), but are also capable of capturing those features of a destination which travelers find attractive and valuable, but typically do not describe them in tweets or reviews. The reviews, in turn, are a valuable source of information about tourist satisfaction and opinion about very specific sites, hotels, restaurants, etc., but they are not able to fully represent the image of destination as a whole.

Apart from pictures, other types of data are available for retrieval from Instagram posts: image caption, comments, geotag (in case the user opted for attaching it). Geotagging is a feasible option also in other networks, but users are much more prone to indicate their location when posting on Instagram, thus making its data a more precise tool for geospatial investigation, for example to analyze visitors' activity. Tenkanen et al. (2017) find that social media activity is a good proxy for visitor counts, especially when Instagram data are used, as opposed to Twitter and Flickr.

Moreover, from a purely technical point of view, it is preferable to use Instagram because its data are less of a black box than, for example, Twitter: while the Twitter API returns a sample of tweets that match the criteria outlined in the query (without giving an indication of the total number of tweets of interest), data obtained from Instagram are the universe of posts from open accounts which match specific search criteria. Panoramio, another image-based social media, has the same problem as Twitter and filters posts in an undocumented way. Flickr shares the same characteristics of Instagram, but is far less popular: the main reason why almost all the studies on photographs use Flickr is because its API easily allows to get metadata and pictures (Li et al., 2018). Finally, given that the primary aim of this work is to develop a time-varying vector of the features composing the perceived image of the destination by travelers, Instagram data appear to fit the purpose better than online reviews, which are mostly tied to a specific hotel, attraction, restaurant, etc.

The islands for which Instagram data were collected are the four largest islands of the Canary archipelago (Tenerife, Gran Canaria, Fuerteventura, Lanzarote), Cyprus, Crete, Malta and Sicily. By analogy with the first research question, two destinations with different characteristics (Athens, a cultural mainland destination and Salento, a leisure mainland destination) were added, in order to test robustness of results and applicability of the approach to a wider pool of

SOCLIMPACT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

destinations.[7] Canary Islands are of a particular interest as an archipelago and were selected with intention to investigate whether images of individual islands can substantially differ from each other. This allows for more refined predictions for the archipelago as a whole. Another reason to focus on the Canaries is the possibility to study the impact of the massive forest fires which occurred on the islands in the summer of 2019, with most devastating wildfire on Gran Canaria. This is done not only using imagery data, but also textual data from captions and comments. Given the importance of the Forest fires impact chain, which is among the three operationalized impact chains for the tourism sector, our analysis is hoping to provide valuable insights on which effects may a forest fire similar to that of Gran Canaria have, if it occurs in other, similar destinations.

The built dataset consists of all retrievable posts that contain a hashtag with the island name and one of the travel-related keywords (*travel, visit, vacation, holiday, trip* and cognate words) in the caption: this way we target posts that focus on the islands as a tourism destination. The timespan is from 8th of June 2019 to 28th of September 2019, hence covering the Summer season 2019. Table 3.2 provides some descriptive statistics of posts for each of the islands analyzed in the first wave, from which we see that there are not important differences in terms of average number of comments, likes or shared geotags.

Table 3.2 – Descriptive statistics of Instagram posts, by destination

	Island				
Indicator	*Tenerife*	*Gran Canaria*	*Fuerteventura*	*Lanzarote*	*Cyprus*
Num. of posts (total)	49,197	30,042	25,442	20,817	63,561
Avg. num. of pictures per post	1.78	1.68	1.6	1.83	1.75
Avg. num. of comments per post	2.22	2.56	2.25	2.08	2.32
Avg. num. of likes per post	67.4	68.9	74.46	74.73	79.65
Share of geotagged posts	66.5%	66.4%	67.6%	64.4%	70%

Although Big Data tools can provide evidence for cases where conventional data are insufficient, for the research questions outlined in 3.1 and 3.2 we exploit those types of user-generated and transaction data which exhibit high velocity. This property allows obtaining valuable insights but has two important drawbacks. First, it requires continuous real-time data collection, which implies that occasional data loss due to changes in website structure or scraping interruptions is almost inevitable. Second and again, since the data are collected in real time, the meaningfulness of the analysis in which it will be subsequently utilized depends on which weather and climate events will actually be observed during the scraping period. For example, it would not be possible to provide any evidence of the impact of heatwaves on outcomes of interest, if during the observation period no heatwaves took place. In other words, we would be able to provide insights regarding impacts of those weather conditions and climatic events, which will have occurred in selected destinations when the data was being collected.

3.3. How do climate events impact hotel performance?

The third question focuses on the supply side of the tourism sector and aims at estimating the impact of climatic events on hotel performance indicators. To this aim, we cannot resort to public data scraped from the Internet, as performance is private and valuable information for

[7] Similar to the other line of research, due to limit in human and computing resources, this report only analyses the four Canary Islands and Cyprus. Future analyses will include Malta, Crete, Sicily, Athens and Salento.

hotels, and we use data provided by STR-SHARE instead. STR is a private company collecting data with daily frequency from hotels worldwide and providing benchmarking, analytics and marketplace insights for global hospitality sectors. In particular, STR collects data on hotel supply (in terms of number of rooms) and demand (OR, occupancy rate – the share of available rooms that are sold over a period of time), ADR (average daily rate), total revenue and REVPAR (revenue per available room). Through their SHARE Centre, they collaborate with academia and research institutes, by providing their data for free for research purposes.[8]

In this preliminary assessment of the effect of climate events on hotel performance we focus on the impact chains of forest fires, by studying the case of wildfires occurred in the Canary Islands in August 2019. As already described in Section 3.2, dry and windy weather contributed to the spread of a number of forest fires in Gran Canaria, Tenerife, and Lanzarote, with the most severe and devasting ones that broke out in Gran Canaria (it resulted in the damage of large forest areas and in the evacuation of thousands of residents). To assess the effects of these fires we focused on two of the Canary Islands: Gran Canaria and Tenerife. The reason why we focus only on these two islands and not on all the islands of the archipelago is that the methodology we apply needs to consider similar and comparable destinations, but only one severely hit by fires. Hence, we have Gran Canaria (the treatment group), and one island not (or slightly) damaged by fires, Tenerife (the control group), as the analysis carried out in Section 4.2 reveals that these two islands have a very similar destination image, hence suggesting that Tenerife should be the best control group for Gran Canaria.

The estimation is undertaken using an Interrupted Time Series model (ITSA) that allows to compare treated and control groups after an event that is expected to interrupt the level and trend of a variable, while controlling for autocorrelation. The model can be specified as follows:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 X_t T_t + \beta_4 Z + \beta_5 Z T_t + \beta_6 Z X_t + \beta_7 Z X_t T_t + \varepsilon_t \qquad [3]$$

where Y_t is one of the hotel performance variables analyzed at day t (ADR, REVPAR, OR), for hotels in Tenerife and in Gran Canaria. T_t is a trend variable, X_t is a dummy variable representing the occurrence of a forest fire (before the fire it takes value 0, after wildfire outbreak it takes value 1), and X_tT_t is an interaction term. The dummy variable Z denotes the treated group. We consider Tenerife as the control group and Gran Canaria, where wildfires mainly spread and damaged, as the treatment.

While β_0 and β_1 represent, respectively, the starting level of the dependent variable Y_t (intercept) and its trajectory before the occurrence of forest fires for the control group, the change in the level and in the trajectory of Y_t due to the fire outbreak after the occurrence of forest fires for the control group are represented, respectively, by β_2 and β_3. β_4 and β_5 indicate, respectively, the intercept difference and slope difference between treatment and control before fire outbreak, whereas β_6 and β_7 embody the intercept and slope difference respectively between treatment and control groups after fire occurrence.

Given that fires occurred in mid-August 2019, we only considered data for Summer 2019 (to be more precise, from 1st June to 31st August, 2019. Table 3.3 presents some descriptive statistics of ADR, REVPAR and OR for both the islands. The table includes also descriptive statistics for the same period in 2018.

[8] We thank STR-SHARE team, in particular Steve Hood, for the prompt provision of data about hotels in the investigated regions.

Table 3.3 – Descriptive statistics of hotel performance measures for the islands Gran Canaria and Tenerife

		Gran Canaria				Tenerife			
Variable	**Period 2019**	**mean**	**sd**	**min**	**Max**	**mean**	**sd**	**min**	**Max**
Average daily rate (ADR)	Jun-19	79.02	3.12	71.31	84.15	85.18	3.61	79.66	91.85
	Jul-19	87.24	4.46	76.23	93.71	106.80	3.35	99.71	112.37
	Aug-19	90.51	5.38	80.46	98.70	118.59	6.13	101.79	127.43
	Summer 2019	85.66	6.52	71.31	98.70	103.72	14.56	79.66	127.43
Revenue per available room (REVPAR)	Jun-19	59.29	5.66	48.97	71.65	61.96	6.64	49.59	74.17
	Jul-19	70.05	6.91	57.98	84.87	78.11	6.92	65.09	93.87
	Aug-19	72.48	5.80	62.34	83.29	96.17	9.16	75.10	110.46
	Summer 2019	67.36	8.36	48.97	84.87	78.93	15.93	49.59	110.46
Occupancy rate (OR)	Jun-19	75.05	6.78	61.87	86.83	72.61	5.67	61.47	84.04
	Jul-19	80.22	5.64	69.27	90.57	73.07	5.26	63.16	85.59
	Aug-19	80.04	3.45	73.71	87.43	80.97	4.85	71.01	90.42
	Summer 2019	78.47	5.90	61.87	90.57	75.58	6.48	61.47	90.42
Variable	**Period 2018**	**mean**	**sd**	**min**	**max**	**mean**	**sd**	**min**	**max**
Average daily rate (ADR)	Jun-18	71.86	5.15	62.94	82.01	87.98	4.45	81.42	96.12
	Jul-18	82.19	7.75	67.12	93.17	111.09	6.93	102.19	139.55
	Aug-18	92.86	7.39	78.67	108.51	119.87	7.55	104.31	129.15
	Summer 2018	82.42	10.96	62.94	108.51	106.51	14.90	81.42	139.55
Revenue per available room (REVPAR)	Jun-18	53.66	6.24	41.35	70.02	63.94	5.05	56.52	76.61
	Jul-18	66.06	8.78	52.34	80.46	82.20	9.15	64.16	100.35
	Aug-18	74.33	6.32	61.69	85.63	95.77	8.59	78.16	109.45
	Summer 2018	64.80	11.11	41.35	85.63	80.82	15.19	56.52	109.45
Occupancy rate (OR)	Jun-18	74.64	6.34	58.24	86.42	72.67	4.20	60.98	84.11
	Jul-18	80.17	4.62	69.04	87.42	73.88	5.59	61.86	84.68
	Aug-18	80.08	3.31	74.07	86.22	79.80	3.46	73.24	87.79
	Summer 2018	78.34	5.49	58.24	87.42	75.48	5.45	60.98	87.79

For a robustness check, we also conduct a difference-in-difference analysis to quantify the effects of wildfire outbreak on hotel performance measures. By focusing on the same sample, the following model is then estimated:

$$Y_t = \beta_0 + \beta_1 Z + \beta_2 X_t + \beta_3 X_t Z + \varepsilon_t \quad [4]$$

where Y_t is one of the outcome variables analyzed for each island, daily measured, X_t is an indicator of the post-treatment period equal to 1 starting from the 14th of August[9] and 0 otherwise, $X_t Z$ is an interaction term equal to 1 if an observation belongs to Gran Canaria and is observed in the post-fire period.

[9] The date with a few days after the first fire outbreak (August 10th) was selected so that time would have passed for the effect to be captured. Our results are robust to setting August 10th or August 17th as a cutoff day (August 17th being the date of the third fire outbreak, the most devasting one).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

4. Results

4.1. The impact of weather on prices

As recalled in the previous section, the intuition suggests that, when the weather forecast is bad, tourists are less willing to book a trip; being aware of that and monitoring a lower-than-expected number of bookings, hotels reduce the offered price, anticipating the lower quality of the tourism experience. However, what stands behind "bad" weather may vary across destinations. Tables 4.1 to 4.4 contain results of estimating equation [1] alternatively using two main weather parameters: temperature[10] (*t*) and rain in forecast[11] (*rain*) for different destination subsamples. Table 4.1 provides results for the full sample and for the city and seaside areas separately. According to coefficients obtained for the full sample of hotels, rain in forecast is associated with 4.4% lower price offered by the hotels, whereas 1℃ higher forecasted temperature increases prices offered by 1%. However, important differences are hidden behind this average result. As evident from subsample analysis, sensitivity of seaside destinations is much higher than that of cities: the impact of rain in forecast amounts to about 10% decrease in the price offered for seaside destinations and only 1.1% for the cities. As regards temperature, a 1 ℃ warmer forecast results in 3% higher price offered in seaside hotels, while no significant relationship between temperature and prices is found in cities. We further investigate regional differences by estimating equation [1] separately for each destination. Several findings emerge here, presented in Table 4.2[12].

Table 4.1 – Full sample, city areas, seaside areas

	Full b/se	Full b/se	City b/se	City b/se	Sea b/se	Sea b/se
rain	-0.044***		-0.011***		-0.104***	
	(0.003)		(0.002)		(0.006)	
t		0.010***		-0.001		0.028***
		(0.001)		(0.001)		(0.002)
Controls and FEs	✓	✓	✓	✓	✓	✓
NT	8206067	7895000	4432099	4264358	3773968	3630642
Nclust	1313	1313	570	570	743	743
rmse	0.27	0.27	0.27	0.27	0.26	0.24

S.e. clustered on hotel level; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[10] Measured in ℃.
[11] A dummy variable taking value 1 if rain or storm is in forecast (in verbal description and in the icon), zero otherwise. While it may seem more appealing to use the probability of rain as metrics of interest, for the baseline analysis we selected the rain dummy to ensure comparability between different data providers (ilmeteo.it does not provide probability of precipitation). Our results, however, are robust to using precipitation probability and volume instead of "rain" dummy.
[12] Here and elsewhere NT is a total number of observations, Nclust stands for the number of hotels, rmse is the root mean square error, a tick in the corresponding row reflects that control variables and fixed effects (FE) mentioned in the baseline equation are included. Given the "large N small T" data structure, in all cases standard errors are clustered on hotel level, which allows accounting for potential autocorrelation and heterosketasticity.

SOCLIMPACT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

It appears that hotel establishments in Sicily do not change their price in response to weather forecasts. This is not due to the fact that Sicily was facing substantially different weather conditions throughout the period of study: precipitation was very similar to that of Corsica, although temperatures were slightly higher on average. At the same time, as compared to other islands, hotel prices in Sicily appear to be the lowest. We further find little evidence of dynamic pricing strategy being applied in this region, and this is probably the reason lying behind the non-significance of estimations. Next, the most sensitive destinations are Corsica, Sardinia and Riccione. Although Riccione is a mainland destination and we assume that, due to its higher accessibility,

Table 4.2 – Regional heterogeneity

	Milan b/se	Venice b/se	Venice Sea b/se	Riccione b/se	Sicily b/se	Sardinia b/se	Corsica b/se
rain	-0.024***	0.005	-0.109***	-0.161***	0.016	-0.158***	-0.125***
	(0.002)	(0.004)	(0.007)	(0.013)	(0.009)	(0.008)	(0.008)
Controls and FEs	✓	✓	✓	✓	✓	✓	✓
NT	2744924	1687175	1108688	545441	797577	1036643	285619
Nclust	301	269	249	142	132	154	66
rmse	0.28	0.24	0.25	0.31	0.18	0.26	0.21
t	0.002**	-0.006***	0.031***	0.041***	-0.003*	0.047***	0.032***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
	(0.002)	(0.000)	(0.000)	(0.000)	(0.049)	(0.000)	(0.000)
Controls and FEs	✓	✓	✓	✓	✓	✓	✓
NT	2638082	1626276	1065871	521987	767327	1002173	273284
Nclust	301	269	249	142	132	154	66
rmse	0.28	0.23	0.22	0.28	0.18	0.21	0.18

S.e. clustered on hotel level; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

demand should be more elastic, the effects of rain and temperature are similar to those for Sardinia and amount to about 16% price decrease when rain is in forecast and 4% increase for each degree higher forecasted temperature.

Very similar but slightly lower estimates are found for Corsica, as well as for Venice seaside, where the effects amount to 11 and 3% respectively. Lower sensitivity of hotel prices on the seaside of Venice can be explained by proximity of Venice itself, which is a cultural destination and offers a wide range of activities to tourists apart from beach leisure. For Venice itself, we find no impact of forecasted rain on prices, while the relationship with temperature is negative. In contrast, a very different picture appears in Milan, where the effects of forecasted rain and temperature have the same direction as in leisure destinations, but are much smaller in magnitude. This may seem counterintuitive since Milan is predominantly a business destination. We further investigate these two findings for Venice and Milan by testing for nonlinearity with respect to temperature, as well as whether sensitivity to rain in forecast is driven by weekend days. Columns (1) and (2) of Table 4.3 contain estimation results for slightly modified equation [1], which are reported in [5] and in [6]:

$$lnPrice_{rt\Delta} = f(hotel_{FE}, room_{FE}, \Delta_{FE}, search_{day_{FE}}, checkin_{day_{FE}}, breakfast_{FE}, FreeCanc, roomslef, \#sleeps_{FE}, view, rain_{t\Delta}, weekend\_rain_{t\Delta}) + \varepsilon_{rt\Delta} \quad [5]$$

Here, apart from the forecasted rain indicator, its interaction with the weekend indicator is added. Columns (3) and (4) present results for the following specification:

$$lnPrice_{rt\Delta} = f(hotel_{FE}, room_{FE}, \Delta_{FE}, search_{day_{FE}}, checkin_{day_{FE}}, breakfast_{FE}, FreeCanc, roomsleft, \#sleeps_{FE}, view, temp_{t\Delta}, temp^2_{t\Delta}) + \varepsilon_{rt\Delta} \quad [6]$$

The results confirm both our hypotheses. First, we see that the negative impact of rain on prices in Milan is indeed driven by weekend travellers, who probably go there mainly for shopping and for cultural activities. The corresponding decrease in prices amounts to 4.6%. Meanwhile, hotel prices in Venice remain insensitive to forecasted precipitation. As regards nonlinear impact of temperature in city areas, we find evidence both in Milan and in Venice. Indeed, while for seaside towns high temperatures are favourable, as tourists prefer to spend time in the water and on the beach, the same temperature becomes a negative factor in a city, which substantially decreases tourists comfort. For this reason, the relationship with temperature in cities follows an inverse U-shape: higher temperatures are positively correlated to prices up to a certain point, after which the relationship becomes negative. For seaside destinations, we do not find any evidence of nonlinear relationship between prices and temperatures.

Table 4.3 – City areas : weekend effects and non-linearity

	Milan b/se	Venice b/se	Milan b/se	Venice b/se
rain	-0.008**	0.005		
	(0.003)	(0.004)		
weekend_rain	-0.046***	0.002		
	(0.005)	(0.005)		
t			0.076***	0.027***
			(0.004)	(0.004)
t^2			-0.002***	-0.001***
			(0.000)	(0.000)
Controls and FEs	✓	✓	✓	✓
NT	2744924	1687175	2638082	1626276
Nclust	301	269	301	269
rmse	0.28	0.24	0.27	0.23

S.e. clustered on hotel level; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, we explore heterogeneity in price sensitivity to weather forecasts with respect to hotel categories. We focus on seaside areas and estimate equation [1] for 3, 4 and 5-star hotels separately. As follows from results in Table 4.4, there is little difference in sensitivity of hotels of different categories to forecasted weather. Prices for higher class hotels, 5-star in particular, appear to be slightly more responsive to both rain and temperature: rain in forecast is associated with about 12% lower prices, while 1 °C higher forecasted temperature leads to 4% higher prices offered for 5-star hotels, whereas the corresponding effects for their 3-star counterparts amount to 10 and 3% respectively.

Table 4.4 – Heterogeneity by number of stars (seaside areas)

	3★ b/se	4★ b/se	5★ b/se	3★ b/se	4★ b/se	5★ b/se
rain	-0.097***	-0.111***	-0.117***			
	(0.008)	(0.010)	(0.022)			
t				0.025***	0.029***	0.037***
				(0.002)	(0.002)	(0.006)
Controls and FEs	✓	✓	✓	✓	✓	✓
NT	1684823	1844312	244833	1620594	1773132	236916
Nclust	444	268	31	444	268	31
rmse	0.24	0.27	0.26	0.23	0.25	0.22

S.e. clustered on hotel level; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To summarize, our analysis revealed that hotel prices are sensitive to weather forecasts in most destinations. In particular, on island seaside destinations, with exception of Sicily, rain in forecast leads to a price decrease of about 13-16%, while 1 ℃ warmer temperature is associated with 3-5% higher prices. As expected, impacts in cultural and business destinations are less pronounced or absent. These findings are important with respect to extrapolation of results to other islands. For example, taking the estimates of Sardinia as a benchmark, an indicative assessment for Cyprus can take into account the fact that, given the higher weight of the cultural component in its image, hotel prices in Cyprus are likely to be less sensitive to precipitation. In a similar fashion, if an island has a large share of high-class hotels (in our seaside hotels subsample only about 4% are 5-star establishments), the average impact on hotel prices is likely to be slightly higher.

4.2. The impact of climate events on tourism quality and destination image

The methodology introduced in Section 3.2 is applied to two types of analysis: in the first one we disregard the time dimension and consider the whole pool of labels for the destinations under investigation in a cross-section framework. In the second analysis we study the frequency dynamics for certain labels and groups of labels, by referring to specific rows of the image matrix.[13]

Figure 4.1 presents word clouds of label frequencies for each island under investigation and considering the ten labels provided by GCV for each picture. We used the full range of labels in their raw form (i.e. without any grouping of labels): it is true that in this way a massive number of features has to be analyzed but, on the other hand, aggregating characteristics to more general topics (e.g. pooling together all food-related labels into a "Food" category) is likely to smoothen differences and might lead to overlooking specific characteristics.

[13] As reported before, this Report investigates the four islands of the Canary archipelago and Cyprus. Future elabortions will also report findings for Crete, Malta and Sicily.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Figure 4.1 – Word clouds based on image labels, Canary Islands and Cyprus

Gran Canaria

Tenerife

Fuerteventura

Lanzarote

Cyprus

At first glance of Figure 4.1, it may seem that all destinations are extremely similar which, perhaps, is of little surprise given that they could all be considered sea and sun destinations: hence, labels like *Sky, Sea, Vacation, Tree, Beach* are among the most frequent for all islands (Table A.1 in the Appendix shows the top-20 labels for each island). Nonetheless, some differences can be spotted: *Mountain* appears relatively more frequently in Tenerife's image than in other islands'; *Sea* and *Ocean* have relatively more weight in Fuerteventura's image; *Art* and *Architecture* are of more importance in Cyprus than in all Canary Islands. To study more deeply the features that distinguish one destination from each other, we compare the rankings of words across islands and determine which labels are relatively more or less important in each island's image.

Cyprus clearly differs from the archipelago of the Canary Islands by the density of cultural heritage, which is immediately visible from the frequency of labels such as *Greek, Mosaic, Amphitheatre, UNESCO heritage*, which are unique for this destination. All the labels representing architectural, religious and historical sites (*History, Historic, Ruins, Site, Ancient, Arch, Building, Courtyard, Dome, Mosque, Holy, Medieval*, etc.) also have higher ranks than on the Canaries. Landscapes and nature are also relevant features of its image, but relatively less than other islands: all nature-based terms are relatively less frequent. Another feature that is more prominent for Cyprus is shopping: terms like *Fashion, Market, Shop, Dress, Clothes, Leather, Formal wear, Accessories, Gown, Bag, Handbag* are more frequent than in any other island.

Gran Canaria and Tenerife have similar images (this is important for the third line of investigation, see Section 3.3 and 4.3) but they also reveal some distinct features. Gran Canaria appears as the most "urban", with highest ranks for labels such as *City, Town, Urban, Settlement, Residential, Metropolitan, Neighborhood*, second only to Cyprus for *Architecture* and *Building*. Interestingly, Gran Canaria has the highest frequency of flora in the posted pictures, and the specificity of labels can even give an idea of the most typically represented genus or species: *Arecales, Elaeis, Bougainvillea*. The latter one contributes to the high frequencies of color-related labels: *Red, Magenta* and *Pink*. Labels *Desert, Sahara, Dune* and *Erg* are almost unique for Gran Canaria and refer to the famous Dunas de Maspalomas. Maspalomas is also a popular destination for the LGBT communities, and this fact is evident from the relatively high frequency of labels such as *Male,* *Barechested, Beard, Muscle*. Finally, the relative frequency of selfies is highest on Gran Canaria than in the other islands.

Tenerife's most famous attraction is Teide National park, crowned by the Teide Peak. This is confirmed by the high frequency of labels *Mountain, Ridge* and *Range*. Unique labels such as *Dolphin, Whale, Bottlenose* point to the fact that whale and dolphin spotting is a distinctive feature of the island's image. Relatively high frequency of *Golf, Adventure*, and *Park* confirm that the island is also famous for golf and theme parks. As it comes to nightlife, we find that Tenerife is leading the ranking with labels such as *Party, Crowd, Alcohol, Drink, Event* that are relatively more frequent than in other islands. Even food (*Dish, Food, Cuisine* labels) appears to have relatively higher importance on Tenerife's image than elsewhere. Overall this evidence suggests that Tenerife's image is, perhaps, the most heterogeneous and, consequently, it appears to be the most generic. This is also manifested in the fact that very general labels such as *Tourism* and *Recreation* are more frequently attached to pictures from Tenerife than from elsewhere.

Lanzarote stands out with its arid landscapes which immediately appear from the high frequency of labels *Landform, Volcanic, Rock, Formation, Soil, Landscape, Geology*, and even *Crater*, which is unique for this island. Two other very distinctive features of Lanzarote's image are *Cactus* which mainly arises from the cactus garden in Guatiza, and *Camel* stemming from a popular camel riding attraction at the Timanfaya National Park. Unlike previous cases, in Lanzarote visitors have a more pronounced preference for active tourism, diving and cycling in particular (*Dive, Scuba, Underwater, Bicycle, Cycle* labels). Interestingly, although Tenerife is usually considered as the most child-friendly island, we find that children (*Child, Toddler*) appear relatively more frequently in Lanzarote images than elsewhere. Another peculiarity is that although Tenerife is the leader in frequency rankings for generic alcohol-related labels (*Alcohol, Drink*), Lanzarote has a higher frequency of *Wine* label relative to other islands: in fact, Lanzarote is ripe for wine production and wine tasting is one of the specific activities in this island.

Fuerteventura appears as the most distant from other destinations. The core of its image are vast sandy seashores and turquoise waters: Fuerteventura leads the frequency of labels *Beach, Shore, Sand, Coast, Turquoise, Ocean*. Beaches are the primary attraction for any sea & sun destination, including Tenerife and Gran Canaria but, for both of them, the frequency ranking for the beach-related group of labels is the lowest. Strong winds that gave the name to the Fuerteventura island are also another constituent pillar of the island's image. Winds are highly valued by visitors, as they allow doing a variety of sports, and this is shown by the relatively high frequency rankings of labels like *Wind, Wave, Sport, Fitness, Equipment, Surf, Kitesurfing, Boardsport, Bodyboard, Windsport*. Another activity that is more typical in Fuerteventura is exploring the island by car, as evident from relatively high frequency of labels *Transport, Automotive, Highway, Road, Tire, Wheel*. On the contrary, the island is last when it comes to food, alcohol, urban sites, partying and big events, with all the related labels ranking lower than in all other destinations. It then appears that Fuerteventura's visitors prefer active pastimes and appreciate the untouched landscape of the island, which makes it a destination with the least consumerist type of tourism.

Once the frequency ranks of labels are available, the IDDI index (Index of Distance in Destination Image) can be computed according to [2b], measuring how much the image of one destination differs from another. Table 4.5 is a symmetric matrix presenting the average squared rank distances of the set of top 20 labels for the islands under investigation. Distances are normalized to the value of two closest destinations, which happen to be Tenerife and Gran Canaria: this last distance is hence set to unity. It appears that Fuerteventura's image is the most specific among the five islands, as it is the most distant from the other ones. It is interesting to highlight that, although belonging to the same institutional and cultural background and being close geographically, Tenerife and Gran Canaria's images seem to be closer to the one of Cyprus, than the Fuerteventura's. We also check robustness of these conclusions using absolute rank differences instead of squared terms; as appears from Table A.2 in the Appendix, the findings are the same as described above.

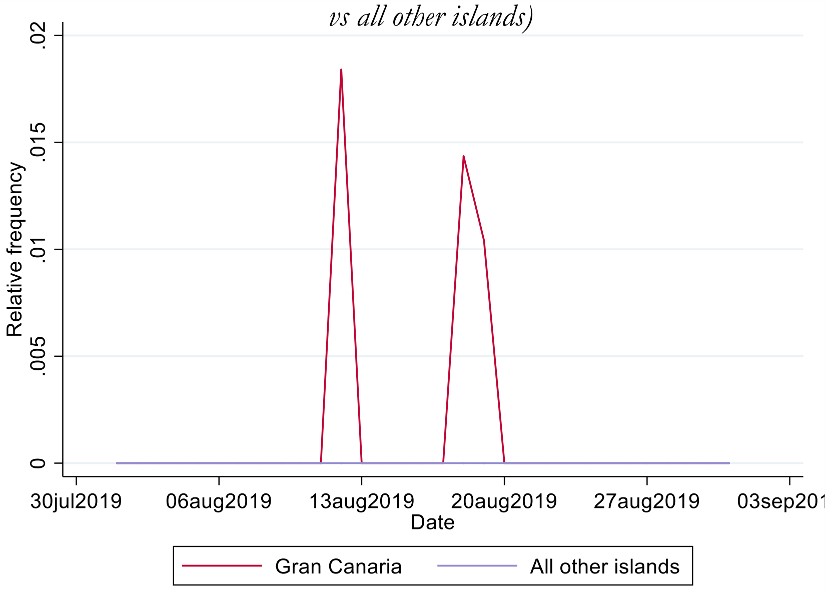
Table 4.5 – Normalized Index of Distance in Destination Image for five islands based on top-20 labels ranks (absolute rank differences)

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus
Lanzarote		8.8	2	4.1	6.9
Fuerteventura	8.8		32.2	10	22.6
Tenerife	2	32.2		1	6
Gran Canaria	4.1	10	1		5.3
Cyprus	6.9	22.6	6	5.3	
Avg. Distance	5.4	18.4	10.3	5.1	10.2

This result suggests that destinations like Fuerteventura may be relatively more vulnerable to climate change and to extreme events, since their visitors seem to weight the natural environment and weather-dependent activities relatively more than elsewhere.

After having explored the panel dimension, comparing the image characteristics of different destinations we now turn to exploit the proposed approach along the time dimension, looking at variations of the destination image overtime. A straightforward application is to analyze if the frequency of specific labels was triggered in reaction to specific events. This would highlight the perceived importance of such events for tourists. To illustrate this point, we provide an example of analyzing tourists' response to extreme events that might hit the destination. In general, tourists are not willing to post anything that makes themselves and others feel negative, and traveling-related posts typically exhibit very little negativity (Deng and Li, 2018). However, when they decide

Figure 4.2 – Daily relative frequencies of wildfire-specific labels in August (Gran Canaria vs all other islands)



to do so, they may indicate that the event is perceived of high importance and, despite the negative mental cost, users prefer to share images and sentiment. Figure 4.2 plots the relative frequency of *wildfire* and *explosion* labels for the island of Gran Canaria in the month of August, compared to all the other islands.

Kinks appear precisely at the time of the two major fires having occurred on the Gran Canaria island during that period, and the relative frequency of labels, although remaining low (around 1.5 - 2% of posts), indicates a change in the image perception. This case is investigated in detail below.

Another application of our approach is exploring substitutability of different activities undertaken by tourists in a destination and their sensitivity to weather conditions. Destinations might be interested to know to what degree visitors can easily substitute one type of activity with another when weather conditions force them to do so. The pressing issue of climate change urges to develop adaptation and mitigation strategies especially on islands, which are more susceptible to negative impacts and more dependent on tourism than most mainland destinations. Hence, this analysis may shed light on the resilience of a destination.

To illustrate this point, we collected daily data on daytime weather in the islands (from openweathermap.org) and analyzed how the frequency of labels associated to different types of activity performed by the tourists change in response to shifts in weather conditions. To keep things simple, the only distinction between beach-related and non-beach-related activities is considered and we focus on the list of the top-20 most frequent labels for posted pictures to find those that are representative of one of the two activities. Accordingly, we selected *beach, sea,* and *ocean* to represent beach-related leisure activities, while *tree* and *plant* were chosen to represent nature-based activity undertaken away from the beach. The following equation [7] is then estimated through OLS:

$$freq_t = \alpha + \beta temp_{10d} + \gamma trend_t + \delta trend_t^2 + \varepsilon_t \qquad [7]$$

where $freq_t$ is the relative frequency at day t of beach-related or nature-based labels as described above (daily observations entering the sample were required to have at least 50 posts per day). The main variable of interest $temp_{10d}$ is used to proxy the weather conditions and stands for 10-day averages of daytime temperature recorded at most visited locations within islands (the last chunk of September is a 13-day average). This way we allow for the possibility that users may post related content a few days later; standard errors are therefore clustered on 10-day chunks level for correct inference. As warmer temperatures, associated to better weather conditions, favors leisure activities on the beach, while lower temperatures make them less comfortable hence forcing tourists to find alternatives (e.g. visit natural parks), we expect that signs of coefficient β are positive for beach-related and negative for non-beach related activities. Possible seasonality effects are controlled through the inclusion of linear and quadratic trends in equation [7], which is estimated separately for each destination.

Results are presented in Table 4.6 and show that the relationship is statistically significant only in a few cases (Gran Canaria and Tenerife for nature-based activities and – weakly – Fuerteventura for beach-related activities, but with the opposite sign). It is important to highlight that effects depend crucially on two factors: the association between specific activities and the set of labels representing them, and the ability of the temperature index to proxy real weather conditions observed at the destination. For example, the 10-day mean temperatures observed in the Canary Islands were between 20 and 25 degrees for all the four islands, while on Cyprus the range was from 27 to 31 degrees. Clearly, tourists were facing consistently hot conditions in Cyprus throughout the whole period and this could explain why in our data we do not find a positive impact of temperature on beach-related labels. It is likely that more variability in temperatures would have shown a different result. Beach-related activities in Fuerteventura turn out to be in a weak negative relationship with temperature, which could be explained by the fact that lower

Table 4.6 – Weather conditions and different types of posted activity, pooled sample (Gran Canaria, Tenerife, Lanzarote), clustered s.e.

	Gran Canaria b/p	Tenerife b/p	Fuerteventura b/p	Lanzarote b/p	Cyprus b/p
			Beach-related labels		
10-day avg temp	0.012	0.019	-0.057*	0.022	0.014
	(0.313)	(0.145)	(0.072)	(0.227)	(0.184)
Linear trend	✓	✓	✓	✓	✓
Quadratic trend	✓	✓	✓	✓	✓
constant	0.332	0.317	2.749***	0.474	1.351***
	(0.203)	(0.190)	(0.001)	(0.190)	(0.000)
			Nature-based labels		
10-day avg temp	-0.047***	-0.017**	0.009	0.006	0.009
	(0.0115)	(0.0070)	(0.0352)	(0.0228)	(0.0064)
Linear trend	✓	✓	✓	✓	✓
Quadratic trend	✓	✓	✓	✓	✓
constant	1.729***	1.040***	0.353	0.530	0.895***
	(0.000)	(0.000)	(0.659)	(0.291)	(0.000)
N	113	113	113	104	112

P-values in parentheses; significance levels: * p<0.1, ** p<0.05, *** p<0.01

temperatures are usually accompanied by stronger winds. Given that Fuerteventura attracts tourists who prefer active sports (such as various types of surfing) to sunbathing, windier (and colder) days may, in fact, attract more visitors to the seaside. As regards nature-based activity, higher temperatures lead to lower frequencies of non-beach related activities in Tenerife and Gran Canaria, which is line with expectations.

This exercise indirectly confirms the validity of the cross-destination image distance index IDDI proposed in the previous subsection. Recall from Table 4.5 that Gran Canaria and Tenerife have the most similar images, followed by Lanzarote, whereas Fuerteventura appears to be more distant. When using the complementary data on temperatures, we see from Table 4.6 that the most similar islands in their association between temperature and activities are again Tenerife and Gran Canaria, followed by Lanzarote, whereas Fuerteventura stands out (we refrain from making any specific conclusions for Cyprus due to the substantially different temperature regime on this island). Since it appears from both Tables 4.5 and 4.6 that Tenerife, Gran Canaria, and Lanzarote can be grouped together, we run equation [7] on the pooled sample to improve estimation efficiency and obtain more robust results, which are presented in Table 4.7. Columns (1) and (3) include destination fixed effects as additional control variables, while in Columns (2) and (4) we add destination-specific linear and quadratic time trends (as opposed to common trends in (1) and (3)), as an additional robustness check.

In line with expectations, pooling the sample allows to estimate the parameters of interest more precisely, providing more robust evidence that warmer temperatures favor beach-based activities. Given that on average the temperatures observed on islands were not extremely high, the positive coefficients in columns (1) and (2) and the negative coefficients in columns (3) and (4) are pointing at weather-driven substitution between beach and nature-based activities.

Table 4.7 – Weather conditions and different types of posted activity, pooled sample (Gran Canaria, Tenerife, Lanzarote), clustered s.e.

	Beach-related labels		Nature-based labels	
	(1)	(2)	(3)	(4)
	b/p	b/p	b/p	b/p
10-day avg temp	0.018**	0.018**	-0.020***	-0.019***
	(0.031)	(0.028)	(0.002)	(0.007)
Linear trend	✓		✓	
Quadratic trend	✓		✓	
Destination FE	✓	✓	✓	✓
Destination-specific linear trend		✓		✓
Destination-specific quadratic trend		✓		✓
constant	0.253	0.216	1.130***	1.155***
	(0.118)	(0.150)	(0.000)	(0.000)
N	330	330	330	330

P-values in parentheses; significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Case study: Wildfires on Canary Islands

There is little doubt that wildfires are a severe threat to natural and human environments. In light of increasing observed and future projected temperatures, forest fires have been growing over time (in fact, already in 2019 there have been more wildfires in the EU than over the past decade), and might have a severe impact on those destinations that are more susceptible to the risk of forest fires. In this part of the report, we focus on the case of Canary Islands and investigate how tourists react to wildfires by describing fire-related contents in their Instagram posts.

The month of August of 2019 was marked by severe forest fires which hit the Canary Islands archipelago. The first fire started in Artenara, on the Gran Canaria island, on the 10th of August, and by the 13th of August, when it was stabilized, had affected about 1500 hectares, forcing evacuation of more than 1000 people. The second fire took place in the Cazadores area on the late evening of the 12th of August. Having lasted for two days, it affected 160 hectares of land by the 14th of August. Finally, the 17th of August the third fire hit the surroundings of Valleseco province of the Gran Canaria and was so severe that it was stabilized only five days after, having affected a huge area of about 10,000 hectares and forced evacuation of 10,000 people. Since the first and the third fires had more severe consequences, they received a lot more attention from the media.

The islands of Lanzarote and Tenerife were also touched by fires, but the damage was much more moderate than that on the Gran Canaria. In the evening on the 18th of August a fire occurred in Villaflor, Tenerife, and late night of the same day in Bosquecillo, Lanzarote, having affected 1.44 and 2 hectares of land respectively. In both cases, the fires were under control within 24 hours. Here, the media coverage of these events was local rather than international. For this reason, we expect these two fires, as well as the second fire on Gran Canaria, to have triggered a less pronounced (if any) response than the two most severe fires on Gran Canaria.

We start by generating binary variables for each of the fire events described above. The variables are set to 1 for the days when the fire was present and 0 otherwise. Note that the events are partially overlapping: for instance, the second fire on Gran Canaria occurred on the day when the first fire was not fully extinguished yet; fires on Tenerife and Lanzarote took place when the largest fire on Gran Canaria was in full swing. Therefore, some corrections are necessary to mitigate the confounding effects stemming from overlapping. In the case of the first fire on Gran Canaria, we restrict the sample so that the date range is from August 1st to August 13th. As the second (and minor) fire on Gran Canaria was largely overlapping with the first one, we exclude it from the analysis. Additionally, we restrict the estimation subsamples to posts related to each separate island. Given that tourists from other islands may also have reacted to such events, in the first approximation we do not make any additional restriction on the tourists' location. Table 4.8 provides results of estimation the equation [8a] below for each separate event, via logistic regression:

$$\mathbb{P}(Caption\_Fire_{it}) = \Lambda\{\alpha + \beta Fire_t + \varepsilon_{it}\} \qquad [8a]$$

Here $Caption\_Fire_{it}$ is a binary variable which takes value 1 if fire-related keywords were mentioned in the post's caption and 0 otherwise. It is important to recall that in this section we analyze text in the post's caption, not labels generated through VOR of posted images as done at the beginning of Section 4.2. This was motivated by the fact that in many instances travelers who decide to talk about wildfires were not witnessing them directly. Hence, even if the posted imagery does not depict the wildfire, the post caption might be dedicated to it. The photo accompanying such a post can be a map of the island with highlighted areas that were affected, or just a picture taken by the user in another location of the island. Given that posts that depict the fire itself are very few, using fire mentions in captions appears as a more reasonable approach. $Fire_t$ is a dummy variable equal to 1 in days when the corresponding fire was active; it is hence equivalent to time fixed effect (this is why time fixed effects per se are not included in the specification). Hereafter $\Lambda\{.\}$ stands for logistic cumulative distribution function (cdf).

Table 4.8 – Logistic regressions for fire indicators, robust s.e.

Fire	GC_1	GC_3	Tenerife	Lanzarote
	b/p	b/p	b/p	b/p
GCfire1	3.003**			
	(0.004)			
GCfire3		1.152***		
		(0.000)		
Tenfire			0.241	
			(0.690)	
Lanzfire				(-)
				(-)
constant	-7.347***	-5.114***	-5.495***	-4.613***
	(0.000)	(0.000)	(0.000)	(0.000)
N	2254	5307	8402	2646

P-values in parentheses; significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. The coefficient of Lanzfire could not be estimated because in the restricted subsample there were no fire-related posts.

Table 4.8 shows that only the fires on Gran Canaria, which resulted in the most severe damage, have significantly impacted the probability of tourists posting fire-related content. The calculation of marginal effects lets estimate that the first fire was associated with a higher probability (1.3%) of a Gran Canaria related travel post mentioning the event than in days without fires. For the third fire, the effect amounts to about 1%. Note that in principle a cumulative impact of all the fires on Gran Canaria might be at play, being embedded in the *GCfire3* coefficient. It is also possible that other events that took place around the same days are contaminating the effects. To disentangle this issue and make the analysis more precise, we then interact the fire dummies with the distance between the poster and the fire location; this also allows to include time fixed effects separately. Table 4.9 contains estimates for equation [8b]:

$$\mathbb{P}(Caption\_Fire_{it}) = \Lambda\{\alpha + \beta Fire_t \times Dist_i + timeFE + \varepsilon_{it}\} \qquad [8b]$$

where the main variable of interest is the interaction of the fire dummy with the distance between the post geotag and the fire location. Table 4.9 shows that when date fixed effects are accounted for, the impact of the first Gran Canaria wildfire appears insignificant, though this may be due to sample reduction, as adding some time fixed effects results in omission of observations for which the time dummy predicts failure perfectly. Therefore, results presented in Table 4.8 could be, in fact, more meaningful. While accounting for distance does not affect conclusions for the Tenerife fire, the effect of distance proves to be significant for the largest fire on the archipelago: when posters were further away from the fire location, they were less likely to mention fire on their Instagram posts. Marginal effects for this estimation amount to 4.7% higher probability of mentioning fire in the post, but the effect diminishes with distance: with each kilometer further from the epicenter the probability decreases by 0.18%.

Table 4.9 – Logistic regressions for fire indicators interacted with distance, robust s.e.

Fire	GC_1	GC_3	Tenerife
	b/p	b/p	b/p
GCfire1	1.567		
	(0.583)		
GCfire1_x_dist1	-0.081		
	(0.519)		
GCfire3		3.260*	
		(0.016)	
GCfire1_x_dist3		-0.130**	
		(0.005)	
Tenfire			-0.265
			(0.878)
Tenfire_dist			-0.023
			(0.572)
Date FE	✓	✓	✓
constant	-5.056***	-4.997***	-4.966***
	(0.000)	(0.000)	(0.000)
N	860	3088	5477

P-values in parentheses; significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$

The citation of the fire in the caption does not say anything about the sentiment triggered by the event. Next, we hence study whether wildfire outbreaks have resulted in higher probability of negative sentiment to occur in the posts' captions. In fact, it could be the case that visitors are more likely to exhibit negative emotions without directly mentioning the fires. Table 4.10 shows results of estimating specifications which are similar to [8a] and [8b], but having a binary indicator related to the negative sentiment as dependent variable. The specifications, which include the two major Gran Canaria fires, are reported in equations [9] and [10]

$$\mathbb{P}(Caption\_Neg_{it}) = \Phi\{\alpha + \beta Fire_t + timeFE + \varepsilon_{it}\} \qquad [9]$$

$$\mathbb{P}(Caption\_Neg_{it}) = \Phi\{\alpha + \beta Fire_t \times Dist_i + timeFE + \varepsilon_{it}\} \qquad [10]$$

$Caption\_Neg_{it}$ is a binary variable taking value 1 if a negative sentiment is expressed in the caption, 0 otherwise. Detection of Table 4.10, interestingly allows to highlight that the first fire now appears to have triggered a more pronounced reaction in terms of negative sentiment, when distances are accounted for: the first fire itself is associated with 5.5% higher probability of expressing negativity in the comments, whereas being each kilometer further away from the epicenter reduces the effect by 0.27%.

Table 4.10 – Logistic regressions for negative sentiment, robust s.e.

	GC_1	GC_3	GC_1	GC_3
	b/p	b/p	b/p	b/p
GCfire1	0.288		2.980*	
	(0.383)		(0.013)	
GCfire3		0.090		0.290
		(0.749)		(0.827)
inter_dist1			-0.146***	
			(0.000)	
inter_dist3				-0.086*
				(0.015)
Date FE			✓	✓
constant	-4.112***	-4.148***	-4.277***	-3.885***
	(0.000)	(0.000)	(0.000)	(0.000)
N	2254	5307	2091	4834

P-values in parentheses; significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Overall, evidence from Tables 4.8 - 4.10 suggests that the two largest fires that occurred on the Gran Canaria island are found to have triggered travelers' reaction which manifested itself in higher probability of mentioning the fires in the posts, as well as higher probability of expressing negative sentiment, linked to the distance between the poster and the wildfire.

To continue in this line of investigation, another aspect of interest is the content of comments that other users leave in response to fire-related posts and to the content related to Canary Islands in general. On the one hand, comments are typically reactions to posts, but it is plausible that commenters learned the extreme events from media coverage and mentioned the fire in their response to pictures from Canary Islands even if the event was not mentioned in the original post. We hence selected the subsamples of posts that neither depicted nor mentioned the fire in the caption and estimated equation [11]:

$$\mathbb{P}(CommFire_{it}) = \Lambda\{\alpha + \beta Fire_t + \varepsilon_{it}\} \qquad [11]$$

As suggested by results in Table 4.11, there is no statistically significant relationship between fire outbreaks and mention in the comments of neutral posts, although the coefficient for the third fire case is close to being significant (p-value = 15.4%). In fact, checking manually the raw data allows to see that indeed there are some comments asking about the fires below the posts that do not mention these events at all. However, these well-aware users are probably relatively low in numbers, and the regression coefficient remains insignificant.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Table 4.11 – Mentioning fires in comments of neutral posts, robust s.e.

Fire	GC_1 b/p	GC_3 b/p	Tenerife b/p	Lanzarote b/p
GCfire1	0.501			
	(0.527)			
GCfire3		0.923		
		(0.154)		
Tenfire			0.530	
			(0.618)	
Lanzfire				0.000
				(.)
constant	-6.346***	-6.539***	-6.881***	-6.260***
	(0.000)	(0.000)	(0.000)	(0.000)
N	5261	5261	8367	2620

P-values in parentheses; significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 4.12 – Comments sentiment in response to fire mentions, robust s.e.

	(1) b/p	(2) b/p
Caption_Fire	2.176***	1.974***
	(0.000)	(0.000)
Caption_Neg		0.834
		(0.078)
constant	-3.453***	-3.453***
	(0.000)	(0.000)
Date FE	✓	✓
N	4976	4976

P-values in parentheses; significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Assessing the reaction to posts that mention wildfire events is important to understand if negativity is also expressed in the comment section. Knowing from having estimated equation [10] that mentioning fire in caption is strongly related to expressing negative sentiment in the post caption, we also include Caption_Neg as a control, to be able to disentangle the reaction of commenters to information about the fires from the propagation of negative sentiment from post caption to the comments. The equation of interest is hence equation [12]:

$$\mathbb{P}(Comm\_Neg_{it}\} = \Lambda\{\alpha + \beta_1 Caption\_Fire_{it} + \beta_2 Caption\_Neg_{it} + timeFE + \varepsilon_{it}\} \qquad [12]$$

which is estimated on the Gran Canaria subsample, column (2) (for comparison we also present results without Caption_Neg as a control variable in column (1)). In accordance with previous findings, we see from Table 4.12 that fire mentions in captions are associated with higher probability of negative sentiment among those who comment the posts of tourists. The average marginal effect implied by the specification represented in column (1) amounts to about 5.1% higher probability of negative sentiment appearing among commenters when fire is mentioned in the caption. From column (2) it appears that the negative emotion in comments is mainly triggered by mentioning the fire itself than from negative captions related to it. However, the latter relationship is (weakly) significant at 10% level (p-value = 7.8%). In this second specification the average partial effect of the fire mentioned in the caption is 4.6%.

To summarize, results of this case study suggest that the huge, devastating fires that took place on Gran Canaria island are of concern for tourists, as they are cited and discussed on Instagram where travelers typically avoid posting any content with negative connotation. Importantly, tourists who posted fire-related content are much more likely to express negative sentiment, as well as tourists who happened to be closer to the epicenter of the fires. Since the literature shows that negative emotions are diminishing satisfaction, willingness to revisit and to recommend the destination (Prayag et al., 2013; Xu et al., 2019), we can conclude that the wildfires may have a negative impact on this conative component of Gran Canaria destination image.

4.3. The impact of climate events on hotel performance

In this section, we estimate the Interrupted Time Series model reported in eq. [3] of Section 3.3 using a multiple-group design to assess the effects of wildfires in Canary Islands on three different hotel performance measures: average daily rates (ADR), revenue per available room (REVPAR) and occupancy rate (OR). Specifically, we compare hotel performances in Gran Canaria, where the most damaging fires occurred, with Tenerife. We start applying the Cumby-Huizinga general test for autocorrelation (Cumby and Huizinga, 1992) to test (and correct) the autocorrelation structure. Accordingly, we consider 5 lags, 2 lags and 2 lags in respectively the ADR, REVPAR and OR specifications.

As shown in Table 4.13, the initial mean level difference between Gran Canaria and Tenerife (Z) is not statistically significant for all the considered measures, whereas the difference in slope before the outbreak of fires was negative and statistically significant for ADR and REVPAR. It is also clear from Figure 4.3 that the trajectories of ADR and REVPAR appear to rise faster in Tenerife than in Gran Canaria before fires occur, whereas the initial mean level for the three variables do not seem to differ. The last panel of this Figure shows a less clear situation for OR, due to its high variability over time.

While the occurrence of fires did not have a significant impact on the mean level of ADR neither in Tenerife nor Gran Canaria (see the coefficients of X and ZX), there is an immediate significant increase in the levels of REVPAR and OR for Tenerife which amplifies the difference in level with respect to the corresponding levels for Gran Canaria. A radical and significant change of direction in trend appears for both the control and treated group. However, the annual reduction after fire occurrence is less pronounced for the treated than for the control group (ZXT vs. XT). Indeed, the impact of fires on the three measures is likely to be masked by the fact that fires broke out in the peak of the season, and thus it is normal that afterwards hotel performances worsen. This is true for both the control and treated groups, but the immediate improvement in the level of performance for Tenerife can indicate a demand shift from the island hit by fires to Tenerife.

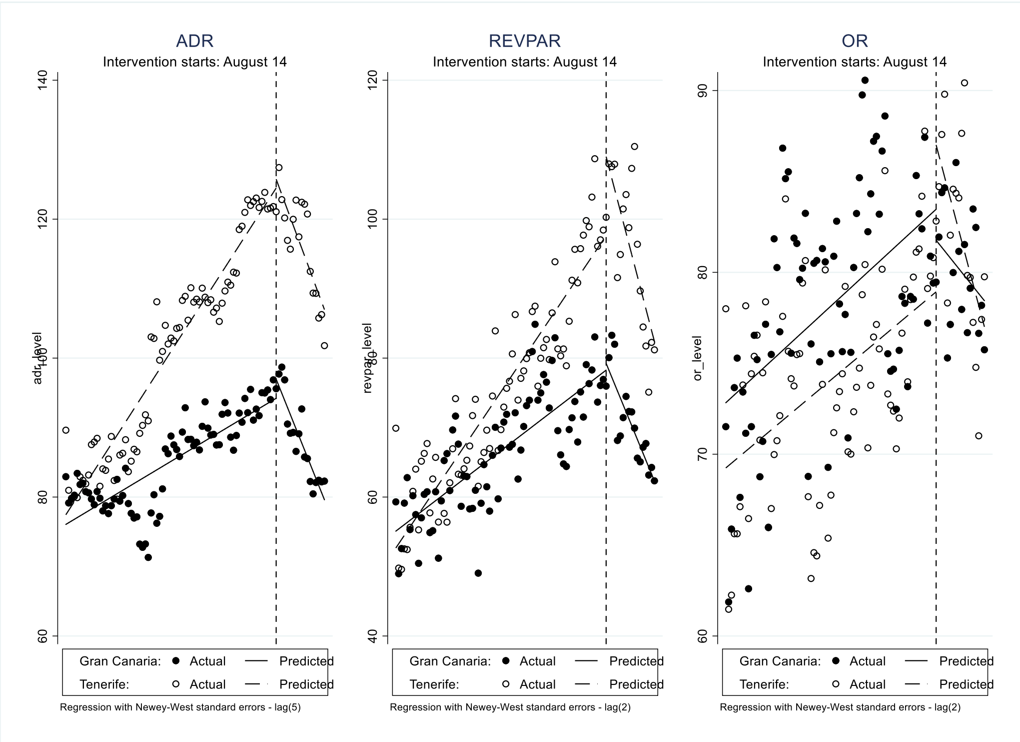
This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Table 4.13 – Multiple-group Interrupted Time Series Analysis with Newey–West standard errors. Effect of forest fires on ADR, REVPAR, and OR in Gran Canaria (treated group) w.r.t. Tenerife (control group)

	(1) ADR (5 lags)	(2) REVPAR (2 lags)	(3) OR (2 lags)
T	0.636***	0.608***	0.131***
	(0.035)	(0.043)	(0.034)
Z	-1.389	2.424	3.586
	(2.595)	(2.018)	(2.252)
Z_T	-0.390***	-0.294***	0.013
	(0.048)	(0.053)	(0.052)
X	1.274	11.365***	8.079***
	(2.268)	(3.499)	(2.086)
X_T	-1.740***	-2.173***	-0.719***
	(0.211)	(0.311)	(0.180)
Z_X	1.482	-10.389**	-9.781***
	(2.606)	(4.325)	(2.975)
Z_X_T	0.473**	0.863**	0.380*
	(0.237)	(0.346)	(0.220)
cons	77.453***	52.669***	69.243***
	(1.734)	(1.614)	(1.537)
N	184	184	184
F test	171.65	94.57	13.98
F(DoF)	(7, 176)	(7, 176)	(7, 176)
p-value	0.000	0.000	0.000

Note: Standard errors in brackets. * p<0.10, ** p<0.05, *** p<0.01

Figure 4.3 – Multiple-group Interrupted Time Series Analysis with Newey–West standard errors.



This result is partially confirmed by the difference-in-difference approach, according to eq. [4] of Section 3.3. In particular, as shown in Table 4.14, the outbreak of fires has significant and negative effects on all the three hotel performance measures for Gran Canaria (X · gran_canaria). Although Gran Canaria and Tenerife are the most similar islands from different perspectives (see Section 4.2), their hotel performance dynamics appear to be driven by some specific features.

Table 4.14 – Difference-in-difference to assess the effects of forest fires on ADR, REVPAR and OR in Gran Canaria (treated group) w.r.t. Tenerife (control group)

	(1) ADR	(2) REVPAR	(3) OR
X	15.719***	20.850***	7.986***
	(2.731)	(2.965)	(1.518)
gran_canaria	-15.629***	-8.324***	4.057***
	(1.709)	(1.855)	(0.950)
X · gran_canaria	-12.445***	-16.597***	-5.960***
	(3.863)	(4.193)	(2.147)
_cons	100.649***	74.850***	74.021***
	(1.208)	(1.311)	(0.672)
N	184	184	184
AIC	1387.664	1417.851	1171.548
R2	0.491	0.357	0.185

To control for this aspect, we replicate the interrupted time series analysis and the difference-in-difference analysis by using different control and treated groups. We concentrate our analysis on Gran Canaria comparing data of summer 2019 (from the 1st of June 2019 to the 31st of August 2019) as treated group, with daily measures recorded in summer 2018 (from the 1st of June 2018 to the 31st of August 2018) as control group. This allows neutralizing the effects due to the specificities of each island.

Results of ITSA model are summarized in Table 4.15, while the visual inspection of Figure 4.4. highlights that the summer seasons 2018 and 2019 are not completely comparable since both the intercepts and the slopes for ADR and REVPAR before 14th of August statistically differ (it is not the case, however, for the OR). Seasons 2018 and 2019 initially had different dynamics, which seem to converge in mid-August. Data confirm that the values of ADR, REVPAR and OR measured in mid-August 2018 and 2019 do not significantly differ. After the fire outbreaks also the slopes of REVPAR and OR do not significantly differ, while the trajectory of ADR significantly differs, although slightly. Difference-in-difference analysis confirms the same findings (see Table 4.16).

It is important to note that in the above analysis a key assumption for correct inference is parallel pre-intervention trends of treated and control groups. As evident from Figures 4.3 and 4.4, this holds only for occupancy rates variable. Common trends assumption does not hold for other variables of interest even when a triple difference model is adopted. As regards occupancy rates, there's some evidence of the negative impact of the wildfire, but only when Tenerife is used as a control group. This result holds in both the diff-in-diff framework, and when ITSA is applied, so the two models confirm each other's conclusions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Table 4.15 – Multiple-group Interrupted Time Series Analysis with Newey–West standard errors. Effects of forest fires on ADR, REVPAR, and OR in Gran Canaria by comparing data in summer 2018 and 2019

	(1) ADR (5 lags)	(2) REVPAR (10 lags)	(3) OR (10 lags)
T	0.431***	0.447***	0.143***
	(0.056)	(0.039)	(0.048)
Z	11.111***	8.485***	0.344
	(3.501)	(2.341)	(3.000)
ZT	-0.186***	-0.134***	0.001
	(0.065)	(0.051)	(0.071)
X	3.180	0.603	-2.690
	(2.262)	(2.374)	(2.165)
XT	-1.672***	-1.372***	-0.073
	(0.101)	(0.134)	(0.166)
ZX	-0.424	0.373	0.988
	(2.601)	(3.151)	(3.136)
ZXT	0.405***	0.062	-0.267
	(0.148)	(0.155)	(0.188)
cons	64.952***	46.609***	72.484***
	(2.920)	(2.066)	(1.955)
N	184	184	184
F test	61.58	71.83	5.71
F(DoF)	(7, 176)	(7, 176)	(7, 176)
p-value	0.000	0.000	0.000

Figure 4.4 – Multiple-group Interrupted Time Series Analysis with Newey–West standard errors. Summer 2018 vs. Summer 2019 for Gran Canaria.

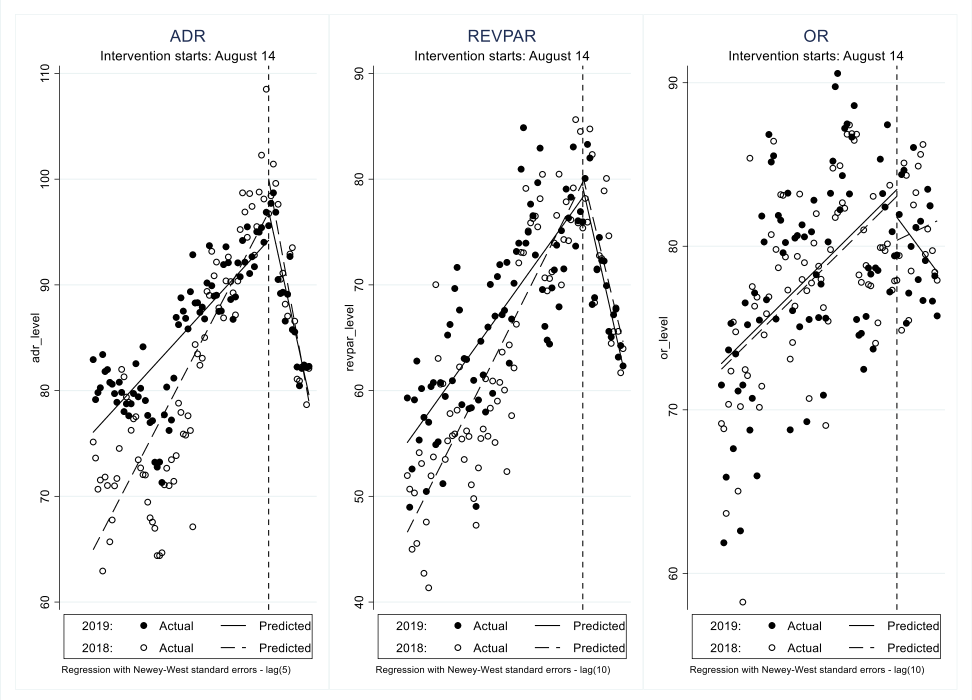


Table 4.16 – Difference-in-difference to assess the effects of forest fires on ADR, REVPAR, and OR in Gran Canaria. Summer 2019 (treated group) w.r.t. summer 2018 (control group)

	(1) ADR	(2) REVPAR	(3) OR
x	8.805***	9.527***	3.260**
	(2.279)	(2.479)	(1.478)
year2019	4.326***	3.585**	0.380
	(1.425)	(1.550)	(0.924)
X_year2019	-5.531*	-5.274	-1.234
	(3.223)	(3.505)	(2.090)
_cons	80.694***	62.941***	77.698***
	(1.008)	(1.096)	(0.654)
N	184	184	184
AIC	1320.973	1351.918	1161.631
R2	0.115	0.105	0.036

To summarise, this block of analysis provided weak evidence of impact of fire on occupancy rates, and overall wildfire outbreaks do not seem to have significantly affected hotel performances in high season in Gran Canaria. This is somehow consistent with previous literature (Hystad and Keller, 2008; Thapa et al., 2013), also reported in the Report of D5.1. In this context, two main aspects may play a significant role. First, fires occurred in the peak of tourism high season, so it is natural to expect declining hotel performances in the following days. This evidence is clear both in Gran Canaria and in Tenerife for both summers of 2018 and 2019. Thus, it is difficult to discern the decline in hotel performances due to fire outbreaks from the one due to the natural end of the peak season. The second issue concerns the geographic location of islands. Reaching and leaving popular islands in high season is not easy (specially for long-haul destinations as the Canaries) and, in general, it requires advanced booking of flights or ferries. So, it is difficult to modify the departure date. Moreover, cancelling holidays in high season for controllable fires (even if widespread) in inland areas can be extremely expensive. The loss for cancelling holidays includes not only the charge for cancelling hotel reservation but also non-refundable travel fees. This may reduce the probability of cancelling pre-booked holidays and allow hotels located in islands to maintain stable performances. Differently, in case of fire outbreaks in the mainland, higher mobility and accessibility are more likely to reduce tourism flows and to affect hotel performances. This likely difference between islands and mainland should be recalled when applying a value-transfer approach.

5. Discussion and Conclusions

This concluding section has two aims. First, we translate the main findings of the analysis (which are expressed in terms of estimated coefficients or other similar metrics) into values that can represent the variation in either arrivals, prices or tourism expenditure stemming from specific climate events. This is done in Sub-Section 5.1. With some caveat (estimations are carried out only for a few islands, for a few activities and for a few types of event), this exercise might provide a guideline for the work of estimating the economic impact of climate change on the tourism sector, to be carried out in D5.6. The second aim of this section is to reassess and discuss the results of Task T5.3 into a kind of executive summary; this is done in Sub-Section 5.2. A few considerations on the limitations and caveats of our work are then expressed in Sub-Section 5.3.

5.1. Estimation of the economic impacts

The impact on accommodation prices

Table 5.1 summarizes the expected impact of temperature and precipitation on hotel prices, built on the coefficients estimated in Section 4.1. Given that the analysis was conducted using data of the summer season only, we suggest that the quantified values should not be extended to the other seasons. It is important to highlight that data on weather forecasts are theoretically more precise than real weather in capturing the expected quality of the experience: if a tourist books today for the next weekend on the forecast that there will be rain, the posted price is consistent with the utility he/she attaches to the stay. A low accuracy in the forecast (e.g. the weekend will turn out to be sunny) will only have the impact of transferring welfare from the hotel to the consumer, as the posted price would not be the equilibrium one. Hence, weather forecasts better capture the reaction of demand and supply than actual weather and climate conditions. Estimations are hence valid under the assumption that forecasts will remain at least as reliable in the future as they are today, and that tourists will keep resorting to forecasts to adjust their expectations about weather conditions and, hence, the quality of the holiday.

Table 5.1 – Implied impacts of climatic variables on prices and revenues

Change in climatic variables	**Implied impact on hotel prices**			**Implied impact on hotel revenues**		
	Sardinia	Corsica	Sicily	Sardinia	Corsica	Sicily
1 ℃ increase in avg temperature	4.7% increase	3.2% increase	No effect	At least 4.7% increase	At least 3.2% increase	Increase by the factor of overnights increase
1mm increase in daily precipitation	1.7% decrease	1.9% decrease	No effect	At least 1.7% decrease	At least 1.9% decrease	Decrease by the factor of overnights decrease

Note: the impact of 1mm increase in daily precipitation was estimated using accuweather forecasts, since the weatherchannel data do not provide forecasts for millimeters of precipitation.

Moreover, note that although the impact on prices could be precisely quantified, the effects on the number of overnight stays, and therefore, revenues, cannot be estimated with precision because the number of sales is not publicly observed. It is possible, however, to provide the lower bound on the impact on revenues, by assuming that occupancy rates are constant: this assumption is consistent with the considered theoretical framework of dynamic pricing, where prices adjust immediately to compensate variations in demand.

The impact on demand for different types of activity

The work carried out in Section 4.2 exploits label frequencies of Instagram pictures and allows to estimate the intensive margin effect: hence, the impact of weather and extreme events on internal demand for different types of activities. Results obtained in Section 4.2 suggest that temperature is driving substitution between nature-based and beach-based activities, and that wildfires on Gran Canaria have a negative impact on the affective dimension of its image. Consistently, we conduct a difference-in-difference analysis to quantify the effect of both temperature and wildfire outbreak on the cognitive aspect of destination image. We focus on the subsample of three islands: Tenerife, Gran Canaria and Lanzarote, and the following seemingly unrelated regression (SUR)[14] model is then estimated:

$$\begin{aligned} LnFreq_{nature_{it}} &= \alpha_n + \beta_n temp_{10d} + \delta_{n1} post_t + \delta_{n2} treat_{it} + \gamma_{n1} trend_t + \gamma_{n2} trend_t^2 \\ &\quad + destFE + \varepsilon_{nit} \\ LnFreq_{seaside_{it}} &= \alpha_s + \beta_s temp_{10d} + \delta_{s1} post_t + \delta_{s2} treat_{it} + \gamma_{s1} trend_t + \gamma_{s2} trend_t^2 \\ &\quad + destFE + \varepsilon_{sit} \end{aligned} \quad [13]$$

where $LnFreq_{j_{it}}$ is the natural logarithm of nature-related and seaside-related labels as described in Section 3.2, $post$ is an indicator of the post-treatment period, which is equal to 1 starting from the date of the first fire (10th of August), whereas $treat$ is an interaction of post and Gran Canaria fixed effect, and therefore it is equal to 1 if an observation belongs to Gran Canaria and is observed in the post-fire period.

[14] This econometric method is chosen because it accounts for correlation of error terms across the two equations, which in the current setting will improve efficiency of the estimates.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Table 5.2 – Impact of temperature and wildfire on demand for beach-based and nature-based activities

	Ln_Freq_Nature	Ln_Freq_Seaside
	b/p	b/p
10-day avg temp	-0.035**	0.036**
	(0.049)	(0.012)
post-fire	0.029	-0.064
	(0.607)	(0.149)
treat	-0.097**	0.035
	(0.043)	(0.359)
Linear trend	✓	✓
Quadratic trend	✓	✓
Destination FE	✓	✓
constant	0.427	-1.239***
	(0.267)	(0.000)
N	330	330

P-values in parentheses; significance levels: * p<0.1, ** p<0.05, *** p<0.01. Covariance matrix estimated using small sample adjustment.

The results of the estimation are presented in Table 5.2. Assuming that the change in label frequencies is an appropriate proxy for demand, we can conclude that a 1 ℃ higher temperature implies a decrease in demand for nature-based activities by 3.5%, at the same time increasing demand for beach-based activities by about the same amount. The wildfires which took place on Gran Canaria led to a decrease in demand for nature-related activities in the island by almost 10%; importantly, in contrast to impact of temperature, there is no evidence that this negative effect was substituted by higher demand for beach-based leisure.

Merging impacts with inputs from WP4 projections

The ultimate goal of SOCLIMPACT project is to estimate the socio-economic impact of climate change. In the organization of the project, this is done by merging the scenarios built by WP4 for future climate conditions in the islands' territories with the estimations built by WP5 about the reaction of the blue economy sectors to climate events. This sub-section starts, for the tourism sector, the task of investigating how to combine these two sets of information to provide reliable values for WP6 socio-economic modelling. It is important to remind that this is not the aim of this Report, and such a work will be fully undertaken in Task 5.5.

The merging, unfortunately, is quite nontrivial and the following computations must be handled with care and to be discussed with all SOCLIMPACT partners. We start by recalling that the very nature of the data employed in this Task investigates the response to very recent weather conditions and events, not to future scenarios. Nevertheless, it can reasonably be assumed that the relationships documented will be maintained also in a larger-scale perspective. We therefore conduct an analysis of possible impacts of projected climate change for different scenarios. In giving projections for near and distant future, WP4 provides average number of days per year when Humidex is above 35°C: a day with Humidex above 35°C describes conditions from discomfort to imminent danger for human beings. In accordance with the formula, high discomfort day indicator is constructed:

$$\mathbb{I}[T(h) \geq 35]$$

$$T(h) = T_{max} + \frac{5}{9}(e - 10)$$

$$e = 6.112 \times \frac{h}{100} \times 10^{\frac{7.5T_{max}}{273.3+T_{max}}}$$

where *e* is the water vapor pressure, *Tmax* is the maximum 2m air temperature (°C) and *h* is the relative humidity (%).

To match WP4 with our investigation, we have to re-estimate equation [1] inserting this discomfort indicator in place of the $weather_{t\Delta}$ variable, to obtain the following point estimates: -0.028 for Sicily[15] (significant at 1% level), 0.221 for Sardinia and 0.151 for Corsica (significant at 0.1% level). This implies that, everything else held constant, for a day when forecasted temperature and relative humidity are such that Humidex is higher than 35°C, hotels in Sicily would offer 2.8% lower prices, whereas those in Sardinia and Corsica would increase prices by 22 and 15% respectively.

We then merge these estimates with inputs provided in D4.3, reporting the number of days with high discomfort levels based on the Representative Concentration Pathways (RCPs) 2.6 and 8.5 for near (2050) and distant (2100) future. Assuming that the relationship between discomfort indicator and prices offered will remain constant over time, it is possible to make a back-of-the-envelope calculation to assess the change in yearly average accommodation price offered based on the expected share of days[16] with Humidex higher and lower than 35°C. Results are presented in Table 5.3[17] and are food-for-thought that we provide to SOCLIMPACT project for future elaboration and discussion.

Table 5.3 – Projected days with moderate and high discomfort levels and implied change in yearly average accommodation price offered (%)

Corsica	Projected number of days per year with Humidex > 35°C		Estimated change in average yearly price (%)
reference	mean	14.4	-
	s.d.	3.8	
RCP2.6 near	mean	27.1	0,519%
	s.d.	5.2	
RCP2.6 distant	mean	27.1	0,519%
	s.d.	5.2	
RCP 8.5 near	mean	28.9	0,592%
	s.d.	5.4	
RCP8.5 distant	mean	72.1	2,357%
	s.d.	8.5	
Sardinia			
reference	mean	49.7	-
	s.d.	7.0	

[15] This appears to contradict the results presented in Section 4.1 where we do not find evidence of hedonic pricing applied by hotels in Sicily. We still present the results for Humidex for the sake of logical coherence, but note that preliminary analysis on the timespan covering May-October 2019 suggests that there is no statistically significant relationship between Humidex and prices in Sicily, thus being consistent with our previous findings.

[16] We assume that at present the reference values are observed.

[17] For a more detailed explanation of calculations see the accompanying Excel file "5.3 Merging impacts with inputs from WP4 projections".

RCP2.6 near	mean	65.3	0,913%
	s.d.	8.1	
RCP2.6 distant	mean	64.0	0,837%
	s.d.	8.0	
RCP 8.5 near	mean	68.6	1,106%
	s.d.	8.3	
RCP8.5 distant	mean	112.2	3,658%
	s.d.	10.6	
Sicily			
reference	mean	52.0	-
	s.d.	7.2	
RCP2.6 near	mean	68.5	-0,127%
	s.d.	8.3	
RCP2.6 distant	mean	70.1	-0,139%
	s.d.	8.4	
RCP 8.5 near	mean	74.1	-0,170%
	s.d.	8.6	
RCP8.5 distant	mean	118.7	-0,514%
	s.d.	10.9	

5.2. Summary and concluding remarks

The main objective of Task 5.3 was to collect and analyze relevant real-time and near-real-time data to provide insights about potential impacts of events linked to climate change on various indicators of the tourism sector. In the *mare magnum* of possibilities, we selected the data to be gathered and the research questions to be addressed on the basis the availability of data, computing power and human resources. We then focused on three main lines of investigation: the impact of weather forecasts (proxy for occurrence of climate events) on accommodation prices, the impact of weather and climatic events on destination image and on preferred activities, and the impact of forest fires on hotel performance. The ultimate goal of Task T5.3 (and of this report) was to experiment with novel and promising approaches, to complement (or to add to) results of Task 5.4. The merging of our estimates with scenarios prepared by WP4 will contribute to WP6 by providing inputs to calculate changes in price and in tourism expenditure when different scenarios for different territories and for different points in time are considered.

To provide an overview of how Task 5.3 relates to other tasks within the Project, Table 5.4 builds on Table 3.1 and summarizes the links between research questions investigated within the task with the corresponding Impact chains set by WP3. Table 5.4 also highlights the focus and the islands for which results are provided, and potential use in bridging WP4 climatic projections with WP6 economic modelling.

Table 5.4 – Linkages between T5.3 and other Work Packages

Line of investigation	Corresponding ICs (link with WP3)	Focus	Islands	Application in value transfer (Task 5.5: bridging WP4 and WP6)
Impact of weather and climate on accommodation prices	2.1: Loss of comfort due to increase of thermal stress	Both demand and supply sides (general equilibrium effect)	Corsica, Sardinia, Sicily	Estimation of impact of Humidex on accommodation prices; using Humidex projections to assess future changes in prices and revenues.
Impact of weather and climate on destination image	1.2: Loss of attractiveness due to increased danger of forest fire in tourism areas	Demand side	Gran Canaria, Tenerife, Fuerteventura, Lanzarote	To be done in Task 5.5
Impact of weather and climate on hotel performance	1.2: Loss of attractiveness due to increased danger of forest fire in tourism areas	Both demand and supply sides (general equilibrium effect) with emphasis on supply side	Gran Canaria, Tenerife, Fuerteventura, Lanzarote	To be done in Task 5.5

The main results of Task 5.3 can be summarized as follows:

- Weather forecasts impact on accommodation prices: a one-degree increase in expected average temperature is associated with an increase in prices of 4.7% in Sardinia, 3.2% in Corsica, while prices are not significantly affected in Sicily. Similarly, a one-mm increase in daily rain is associated with a decrease in prices of 1.7% in Sardinia, 1.9% in Corsica; again, prices in Sicily are not affected.
- We cast a doubt on the generalization of results: in fact, there seems to be a lot of spatial heterogeneity due to the specialization of the destination (cultural and business destinations are the less affected by weather) and different impacts on hotels of different quality (the upper-scale hotels are more heavily affected). Additionally, this report only analyses prices and forecasts data from May to June 2019, therefore, results obtained may be hinging upon the period of observation. Future elaborations will analyze the full summer season and the impact of other climate events.
- As these estimates are based on posted prices, and we do not avail of data on the dynamics of arrivals and overnight stays, we can only assume that tourism expenditure changes by the same percentage as price (consistently with the hypothesis of constant occupancy rates). By feeding different projections from WP4 (RCP2.6 and RCP8.5 for 2050 and 2100) into our estimates for the different islands and for different points in time, we can forecast that summer prices (and tourism expenditure) will rise by 0.5 – 2.4% in Corsica because of future changes in human comfort index, by 0.8 – 3.7% in Sardinia, while they will drop by 0.1 – 0.5% in Sicily because of the increase in the thermal stress associated with climate conditions.
- In the second line of investigation, we use artificial intelligence to scan thousands of pictures posted by tourists on Instagram to develop metrics able to quantify the concept of destination image, as it is perceived by visitors. This way, we can compare destinations and assess how similar or different they are in the eyes of tourists, with nontrivial consequences

when various climatic events occur. A change in the metrics recorded when specific climate events hit a destination implies a change in the relative quality of its holiday experience, with important consequences in terms of tourism behavior and expenditure pattern.

- The application of this approach to the case-study concerning the huge, devastating fires that took place on Gran Canaria in Summer 2019 suggests that tourists who posted fire-related content are (about 5%) more likely to express negative sentiment, as well as tourists who happened to be closer to the epicenter of the fires. Since the literature shows that negative emotions are diminishing satisfaction, willingness to revisit and to recommend the destination (Prayag et al., 2013; Xu et al., 2019), we can conclude that the wildfires may have a negative impact on the destination. Applying our method of approximating dynamics of demand for different types of activities, we also find that wildfires decreased demand for nature-based activities on Gran Canaria by about 10%.
- Although forest fires might negatively impact the destination image, the third line of investigation shows that they seem not to have significantly affected hotel performances (always considering the case-study of fire outbreak in Gran Canaria in summer 2019). This is somehow consistent with previous literature (Hystad and Keller, 2008; Thapa et al., 2013), also reported in D5.1, for which tourists are only marginally affected by forest fires in beach destinations. A more careful investigation into the robustness of this result is needed, as there are confounding effects: fires just occurred at the peak of the summer season, and accessibility to remote islands as the Canaries implies very high costs of rescheduling the holiday. Differently, in case of fire outbreaks in the mainland, an easier mobility and accessibility is more likely to reduce tourism flows and undesirably affect hotel performances. This expected difference between islands and mainland should be recalled when attempting to transfer values.
- The findings presented in this report should be treated as preliminary and are subject to several limitations. First, we highlight again that given the real-time nature of data collection it is only possible to make inference about impacts of weather and climatic conditions that take place in the short run. For example, if during the process of data collection we only observed temperatures in the range between 25 and 34°C, the effect of thermal stress on outcomes may only be inferred assuming that higher temperatures do not have a structurally different impact than recorded temperatures. Second, our tentative value transfer exercise presented in Section 5.1 hinges upon the assumption that the short-term response captured herein would occur with the same magnitude also in the very long-term perspective. Whether this is a valid assumption may be disputable and requires additional research. Third, in this report the analysis of impacts of weather forecasts on prices, the results of which are subsequently used in Section 5.1, is conducted using data covering the time span from May to June 2019, which is not the full summer season. The broader time span from May to October 2019 should be used in future elaborations. Finally, provisional findings of Section 5.1 are built upon investigation of the relationship between weather forecasts and prices, thus precluding from any inference on how this impacts revenues, which would be of higher interest (the only assumption that can be done is that the change in revenues is equal to the change in prices, that is, occupancy rates are constant). This caveat is to be overcome by merging data coming from T5.4 in subsequent work that is carried out in Task 5.5, which specifically focuses on value transfer to GEM-E3 and GINFORS models.

5.3 Scientific discussion and recommendations for future actions

This report investigates the impact of CC on the tourism sector of some of the islands analysed in SOCLIMPACT by using an approach based on the analysis of BD, that can effectively complement more standard approaches (for example based on choice experiments) carried out throughout the rest of WP5. Findings can open promising windows on how to methodologically approach the relationship between climate and tourism, at the same time providing food-for-thought for the project and for researchers willing to pursue this stream of investigation in the future. On this matter, a few considerations are recalled.

First, within this task, we only scratched the surface of BD and their application to the socio-economic consequences of CC. Four years after the start of the project, with different human and financial resources, and given the quick evolution of methodologies, tools, and data storage in the last few years, more sophisticated analytical methods (e.g., network analysis, advanced machine learning) could be applied to these and other sources of data.

Second, most of the findings are island-specific, showing that events linked to the CC impact different territories in a heterogeneous way. Accordingly, future research needs to develop a robust methodology of value transfer, to extend results to destinations differing for key characteristics.

Third, some of the results are partially in contrast with the expected negative role that CC plays for the local economy. Specifically, we find that increases in the temperature are linked to increases in prices, bringing economic benefits to the tourism industry. However, this is not fully surprising and is in line with intuition, as people are more likely to spend time in beach destinations when higher temperatures or heat waves hit the territory. Apparently, this is more likely to happen in the future, due to the widely use of remote working and videoconferences boosted by the COVID-19 pandemic. Thus, the approach followed in this report is likely to provide a useful complement to standard analysis based on non-market value (D5.5), as it focuses and highlights aspects that cannot be easily grasped through choice experiments. This poses a question that must be addressed by future research: how to combine different methodologies and approaches in a unique estimate of changes in tourism expenditure in reaction to CC. In D6.2 we work in this direction, by proposing a rough integration of both approaches, but more refinement is needed.

Fourth, the analysis of BD can deliver much more, when using diverse and more sophisticated techniques. For example, the idea of exploiting GPS data of mobile phones to track tourists' mobility in response to climate events – which was not pursued given the impossibility to access data from phone companies for privacy reasons – appears promising and worth pursuing if privacy concerns can be effectively handled. In the same line of thought, the use of this "Big Data approach" to other Blue Economy sectors (e.g., the energy sector, for which BD could be available) should be considered as an important extension of this work.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

References

Agapito D., Mendes J., Valle P., 2013, Exploring the conceptualization of the sensory dimension of tourist experience, *Journal of Destination Marketing & Management*, 2(2), 62-73.

Aral, S. and Walker, D. 2012. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341.

Baloglu S., McCleary K., 1999, A model of destination image formation, *Annals of Tourism Research*, 26(4), 868-897.

Becken, S., 2012- Measuring the effect of weather on tourism: a destination and activity based analysis, *Journal of Travel Research*.

Beerli A., Martin J.D., 2004, Factors influencing destination image, *Annals of Tourism Research*, 31(3), 657-681.

Chua A., Servillo L., Marcheggiani E., Moere A.V., 2016, Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy, *Tourism Management*, 57, 295-310.

Cumby, R., & Huizinga, J., 1992. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica*, 60(1), 185-195.

Deng N., Li X.R., 2018, Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection, *Tourism Management*, 65, 267-278.

Doornik, J.A. and Hendry D.F. 2015. Statistical Model Selection with 'Big Data' *Cogent Economics & Finance*.

Einav, L., Levin, J.D. 2013. The data revolution and economic analysis, *NBER working paper n. 19035*.

Figini, P., Cicognani, S., Zirulia, L., 2019. Booking in the rain : Testing the impact of public information on prices. *DSE working paper n. 1137*.

Hirschman, E. C. (1986). The effect of verbal and pictorial advertising stimuli on aesthetic, utilitarian and familiarity perceptions. *Journal of Advertising* 15 (2): 27–34.

Hystad, P. W., & Keller, P. C. 2008. Towards a destination tourism disaster management framework: Long-term lessons from a forest fire disaster. *Tourism management*, 29(1), 151-162.

Jiang K., Yin H., Wang P., Yu N., 2013, Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions. *Neurocomputing*, 119, 17-25.

Költringer C., Dickinger A., 2015, Analyzing destination branding and image from online sources: A web content mining approach, *Journal of Business Research*, 68, 1836-1843.

Kurashima T., Iwata T., Irie G., Fujimura K., 2013, Travel route recommendation using geotagged photos, *Knowledge and Information Systems*, 37(1), 37-60.

Lee I., Cai G., Lee K., 2014, Exploration of geo-tagged photos through data mining approaches, *Expert Systems with Applications*, 41, 397-405.

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

Miah S.J., Vu H.Q., Gammack J., McGrath M., 2017, A Big Data Analytics Method for Tourist Behaviour Analysis, *Information & Management*, 54, 771-785.

Mizzi, C. ; Fabbri, A., Rambaldi, S., Bertini, F., Curti, N., Sinigardi, S., Luzi, R., Venturi, G., Micheli D., Muratore, G., Vannelli, A., Bazzani, A., 2018. Unraveling pedestrian mobility on a road network using ICTs data during great tourist events, *EPJ Data Science,* 7(1) :1-21.

Okuyama K., Yanai K. (2013). A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web. In: *The Era of interactive media* (pp. 657e670). New York: Springer.

Pike S., Ryan C., 2004, Destination positioning analysis through a comparison of cognitive, affective, and conative perceptions, *Journal of Travel Research*, 42, 333–342.

Prayag, G., Hosany, S., & Odeh, K. (2013). The role of tourists' emotional experiences and satisfaction in understanding behavioral intentions. *Journal of Destination Marketing & Management*, *2*(2), 118-127.

Stepchenkova S., Zhan F., 2013, Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography, *Tourism Management*, 36, 590-601.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific reports*, *7*(1), 17615.

Thapa, B., Cahyanto, I., Holland, S. M., & Absher, J. D. (2013). Wildfires and tourist behaviors in Florida. *Tourism Management*, 36, 284-292.

Varian, H.R., 2014. Big Data : new tricks for econometrics. *Journal of Economic Perspectives*, 28(2) :3-28.

Xiang Z., Schwartz Z., Gerdes J.H., 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.

Xu, Z., Zhang, H., Zhang, C., Xu, M., & Dong, N. (2019). Exploring the role of emotion in the relationship between museum image and tourists' behavioral intention: The case of three museums in Xi'an. *Sustainability*, *11*(3), 559.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Appendix

Table A.1. Top-20 labels by frequency

	Tenerife		Gran Canaria		Lanzarote		Fuerteventura		Cyprus	
rank	label	# of occur.	label	# of occur.	label	# of occur.	label	# of occur.	label	# of occur.
1	Sky	30177	Sky	19309	Sky	17366	Sky	25652	Sky	33698
2	Vacation	21544	Vacation	15070	Vacation	12353	Sea	18920	Vacation	31342
3	Tree	19964	Tree	13986	Water	11051	Vacation	18686	Water	27804
4	Mountain	18511	Water	10458	Tree	10246	Ocean	18145	Sea	25210
5	Water	17875	Plant	9312	Sea	9883	Water	17784	Ocean	22905
6	Sea	15808	Sea	9140	Ocean	9609	Beach	15584	Tree	22315
7	Ocean	14770	Summer	8057	Mountain	8386	Tree	11442	Summer	18629
8	Plant	13683	Mountain	7760	Plant	7378	Summer	10332	Fun	14970
9	Fun	10766	Ocean	7620	Beach	6046	Coast	10096	Plant	14502
10	Summer	10602	Fun	7304	Summer	5865	Shore	8422	Architecture	14282
11	Tourism	10162	Tourism	6518	Landform	5653	Fun	8028	Beach	14054
12	Landform	9920	Nature	6201	Coast	5564	Sand	7984	Tourism	13287
13	Photography	8817	Photography	6079	Fun	5552	Cloud	7095	Coast	12141
14	Cloud	8711	Landform	5850	Cloud	5531	Plant	6915	Building	10928
15	Nature	8428	Architecture	5620	Tourism	5304	Blue	6470	Photography	10836
16	Beauty	8058	Beach	5586	Landscape	5060	Landform	6360	Beauty	10604
17	Coast	7961	Hair	5235	Rock	4792	Nature	6278	Leisure	9825
18	Beach	7493	Palm	4974	Photography	4715	Body	6110	Nature	9514
19	Leisure	7404	Beauty	4882	Blue	3889	Photography	6069	Blue	9133
20	Hair	7227	Landscape	4757	Shore	3733	Mountain	6016	Hair	9051
Total # of occur.		835306		541522		441131		585693		1064845
Total # of pictures		87570		50469		38095		40707		111232

Table A.2. Normalized Index of Distance in Destination Image for five islands based on top-20 labels ranks (absolute rank differences).

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus
Lanzarote		2.8	1	2	2.1
Fuerteventura	2.8		3.7	3	3.3
Tenerife	1	3.7		1	2
Gran Canaria	2	3	1		2.1
Cyprus	2.1	3.3	2	2.1	
Avg. Distance	2.1	3.2	2.0	2.0	2.4

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Figure A.1. Word clouds based on image labels: Crete, Malta, Sicily.

Crete

Malta

Sicily

SOCLIMPACT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661

Table A.1a. Normalized Index of Distance in Destination Image for eight islands based on top-20 labels ranks (squared rank differences).

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		8.8	2	4.1	6.9	4.8	111.8	27.4
Fuerteventura	8.8		32.2	10	22.6	12.8	187.3	51.9
Tenerife	2	32.2		1	6	8.4	43.9	16.5
Gran Canaria	4.1	10	1		5.3	24	69.4	31
Cyprus	6.9	22.6	6	5.3		3.9	30.1	18.6
Crete	4.8	12.8	8.4	24	3.9		18.1	12.5
Malta	111.8	187.3	43.9	69.4	30.1	18.1		14.2
Sicily	27.4	51.9	16.5	31	18.6	12.5	14.2	
Avg. Distance	23.7	46.5	15.7	20.7	13.3	12.1	67.8	24.6

Table A.1b. Normalized Index of Distance in Destination Image for eight islands based on top-20 labels ranks (absolute rank differences).

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.8	1.4	2	2.1	2.2	7.5	3.7
Fuerteventura	2.8		3.7	3.1	3.3	2.6	9.3	5.2
Tenerife	1.4	3.7		1	1.8	3.2	6.1	3.5
Gran Canaria	2	3.1	1		2.1	4.1	7.4	4.3
Cyprus	2.1	3.3	1.8	2.1		2	4.7	3.8
Crete	2.2	2.6	3.2	4.1	2		7.4	3.1
Malta	7.5	9.3	6.1	7.4	4.7	7.4		2.9
Sicily	3.7	5.2	3.5	4.3	3.8	3.1	2.9	
Avg. Distance	3.1	4.3	3	3.4	2.8	3.5	6.5	3.8

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No776661