



**HAL**  
open science

# What can we learn from Airbnb data on tourist flows? A case study on Iceland

Alexandre Cebeillac, Yvette Vaguet

## ► To cite this version:

Alexandre Cebeillac, Yvette Vaguet. What can we learn from Airbnb data on tourist flows? A case study on Iceland. *Espace Populations Sociétés*, 2021, 2020/3-2021/1, 10.4000/eps.10452. hal-03148135

**HAL Id: hal-03148135**

**<https://hal.science/hal-03148135>**

Submitted on 4 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



---

## What can we learn from Airbnb data on tourist flows? A case study on Iceland

*Que nous apprennent les données Airbnb sur les flux touristiques ? Application à l'Islande*

**Alexandre Cebeillac and Yvette Vaguet**

---



**Electronic version**

URL: <https://journals.openedition.org/eps/10452>  
DOI: 10.4000/eps.10452  
ISSN: 2104-3752

**Publisher**

Université de Lille

Provided by Université de Caen Normandie



**Electronic reference**

Alexandre Cebeillac and Yvette Vaguet, "What can we learn from Airbnb data on tourist flows? A case study on Iceland", *Espace populations sociétés* [Online], 2020/3-2021/1 | 2021, Online since 21 January 2021, connection on 04 June 2024. URL: <http://journals.openedition.org/eps/10452> ; DOI: <https://doi.org/10.4000/eps.10452>

---

This text was automatically generated on February 16, 2023.



The text only may be used under licence CC BY-NC-ND 4.0. All other elements (illustrations, imported files) are "All rights reserved", unless otherwise stated.

---

# What can we learn from Airbnb data on tourist flows? A case study on Iceland

*Que nous apprennent les données Airbnb sur les flux touristiques ? Application à l'Islande*

Alexandre Cebeillac and Yvette Vaguet

---

## AUTHOR'S NOTE

This work was supported by French National Research Agency [grant: ANR PUR-15-CE22-0006].

The authors would like to thank the anonymous reviewers for valuable comments and advice.

## Introduction

- 1 Tourism flows are increasingly important – at least up to the covid-19 crisis - with more than 1.4 billion international trips estimated for 2018, and boreal regions are no exception, despite relatively low volumes [UNWTO, 2019]. However, when it comes to data, reliable and regionally comparable statistics across the Arctic are hard to combine and compile. There are incompatible data sources; some countries count visitors, others count overnight stays, making the gauging of inbound tourism highly difficult, especially with multiple entry points and the fact that not all visitors are tourists. Accommodation statistics offer a good proxy and are often used [Fay & Karlsdóttir 2011].

- 2 Iceland offers a double advantage for quantitative studies on polar tourism: the country is entirely within the geographical area of interest and presents an excellent national statistical system which distributes data free of charge on the Internet. In addition, tourism-related transformations that Iceland has experienced in the last decade are by coincidence, concomitant with the development of online platforms such as *Airbnb*<sup>1</sup>. This article, with an emphasis on methodological aspects, will first come to facts and figures about Arctic and Icelandic tourism, as well as the *Airbnb* platform. Then, it will develop the data and the method that relies on the travel routes of people registered on *Airbnb* as a proxy for more global tourist flows, especially in the Arctic regions where data could be scarce or too spatially aggregated. To do so, we will compare the tourism flows observed in Iceland in recent years with institutional data from the various Icelandic census and statistical offices, which we will compare with the data collected from the *Airbnb* platform in September 2019.
- 3 By collecting and aggregating information about the places visited by users of the platform and the people who visited or hosted them, we will be able to establish flow matrices. This could be used for future studies on the interactions between Arctic regions and the rest of the world through the lens of world tourism.

## 1. Tourism in Arctic

### 1.1 Polar regions in global tourism

- 4 There were 1.4 billion international tourist arrivals (ITA) in 2019 and more than 25 million cruise passengers worldwide [UNWTO, 2019]. The Arctic counts around 10 million tourists within the boundaries of the *Arctic Human Development Report*, every year (AHDR, Larsen & Fondahl 2014). That is to say, that the tourist flows at high latitudes remain relatively low, essentially for reasons of high cost. Indeed, the boreal regions welcome less than 1% of the total ITA (to compare: Europe and France welcome respectively 22% and 5%). It is also low when compared to the most visited cities worldwide: Bangkok, London and Paris welcome about 20 million visitors every year.
- 5 Nevertheless, the figure is significant when compared to the boreal populations, as the number of tourists in the Arctic is more than the double of the number of residents within the AHDR limit. For instance, Svalbard welcomed over 130,000 tourists in 2015<sup>2</sup> while its population is about 2,670 inhabitants. Recent evolution is another aspect that needs to be highlighted. On one hand, as a matter of fact, the top 5 countries accounted for 71% of arrivals in 1950 against only 31% in 2010. Polar regions take part in this diversification of destinations. On the other hand, global tourist flows have increased considerably since the 1950s. The annual growth rate of global tourist flows has been steady at around +5% for three decades. The only exceptions are crises such as the Asian economic crisis in 1997, the Russian financial crisis in 1998, the global crisis 2007-2010 and currently, the covid-19 crisis 2019. Against this backdrop, Polar regions are witnessing an even greater increase than the global trend, especially in Iceland (fig. 1). Overnight stays in Svalbard increased by 62% between 2001 and 2015.
- 6 Nevertheless, growth is uneven, sometimes considerable as in Iceland, sometimes stagnating in other Arctic areas, but there are no examples of decrease in the long-term [Lasserre and Têtu 2015]. Iceland feeds the thinking of its neighbours, notably Greenland, on the interests and disadvantages of developing the tourism industry

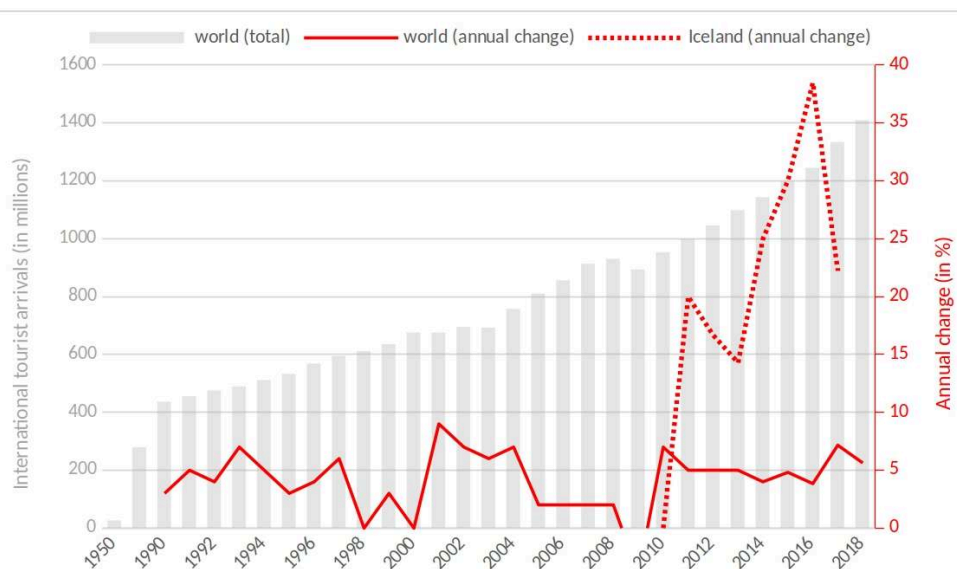
[Steenholdt and Chimirri, 2018]. Indeed, this intensification of flows at the edges of the ecumene is therefore likely to lead to changes in the environment and influence local lifestyles through cultural and economic exchanges and the resulting new sources of income [Müller, 2013, 2015]. It also redraws the interactions between relatively isolated areas that are seen as peripheral, and the major urban centres. Tourism can thus be seen as a global power force that is transforming the “north” into a “global north” that is increasingly connected to the world system [Veijola and Strauss-Mazzullo, 2019, Smith, 2012]

- 7 The 3A combination is of utmost importance for tourism activity: *Attraction*, *Access*, and *Accommodation*. About *Attraction*, climate change has exacerbated the desire for an experience on the icy margins of the ecumene [Grenier, 2009]. The high latitudes have become a destination to go see and experience before it's too late, the ice pack, icebergs, the megafauna (whales), Aurora Borealis as well as to meet indigenous people [Hall and Saarinen 2010, Huddart and Stott, 2020]. Therefore, there is indeed something of a paradox given that in return, tourism is a significant contributor to climate change. About *Access*, it's well known that the Arctic offers in general poor access and the most visited areas are the most accessible ones which are Finland, Norway, Sweden, Alaska and Iceland. To finish, *Accommodation* is a challenge everywhere especially in the tiny remote boreal villages. Cruises are a growing success as they offer an answer to both *Access* and *Accommodation*. However, cruising is increasingly being debated as ships grow in number and size and penetrate further north. Nowadays, platforms such as Airbnb are also flourishing in the Arctic and we will study this below.

## 1.2 The Icelandic tourism boom

- 8 Iceland is the most advanced case to illustrate the tourism boom in the Arctic (fig. 1). In 2017, mass tourism brought 2.69 million visitors, *i.e.* more than 7 times the country's population (348,000 inhabitants). Moreover, it increased rapidly: 4.13 times between 2009 and 2017 [Statistics Iceland, 2019, 2020a].

Figure 1: Evolution of tourism in the world and in Iceland.



Source: UNWTO, 2019; Statistics Iceland, 2020a.

- 9 This tourism boom in Iceland can be attributed to the attractions mentioned below for the whole of the Arctic, but in addition, Iceland offers a breathtaking natural environment due to its unique geological features. Moreover, it is also due to a juxtaposition of one-off events specific to Iceland. The worst financial crisis in modern memory occurred in 2008 after the failure of the three main banks. The collapse of the krona made the country a cheap destination as well as the country happy to develop the tourism sector [Jóhannesson and Huijbens, 2013]. Then, the eruption of the Eyjafjallajökull volcano in 2010 reminded the whole world that Iceland did exist with its tremendous landscapes. There were subsequent deluge of budget flights making the country a quick and cheap plane ride<sup>3</sup>. Last but not least, ‘set-jetting’ has reinforced the trend with the successful series *Game of Thrones* (GoT). The experience of ‘set-jetting’ was defined as a rapidly growing form of tourism [Joliveau, 2009]. It is about transforming film locations into tourist attractions and dates back to the early-2000s with the film *The Lord of the Rings*. New Zealand, the set of the saga, had then seen its number of tourists double between 2000 and 2010.
- 10 Since 2011, the series GoT probably contributed to increase the number of visitors in Iceland (fig. 1). The number of tourists jumped by 252% from 500,000 in 2013 to 1.77 million in 2016. Between 2012 and 2016, the number of visitors to the volcanic sand beach of Vik jumped by 1300%<sup>4</sup>. In the same way, in late 2015, the famous popstar Justin Bieber also contributed to the Iceland tourism boom, with his clip *'I'll show you'*. This clip set in the Fjadrargljúfur Canyon, has been seen over 460 million times on the internet. This tourist site has since been closed to the public<sup>5</sup>.
- 11 Finally, tourism has been an important factor in the country's recovery from the crisis of 2008. It is now a key sector of the Icelandic economy. It currently accounts for about 14% of employment and 8.6% of GDP in 2017 compared to 3.5% in 2009 [Statistics Iceland, 2017, 2020b]. This shift from an economy historically based on fishing to one based more on tourism is illustrated by the country's renunciation of whaling for the second consecutive year. In fact, the cetacean is more profitable with whale-watching as an activity. Tourism has become a great source of income and revival for some villages but also a source of problems. Often, urban space is reorganised to fit the practices and expectations of tourists (e.g. renewal of types of shops, opening hours...) frequently at the expense of the locals [Mermet, 2017, 2019].
- 12 Among the 2.69 million arrivals in 2017 in Iceland, ‘only’ 465,000 are cruise passengers with *Accommodation* provided. Air travel accounted for 85% of the arrivals [Huddart and Stott, 2020 ; Statistics Iceland, 2020a]. Thus, the need of accommodation is tremendous.
- 13 The official data in Iceland are well informed concerning the inflows. One of the objectives of this paper is to find other alternative data sources allowing us to observe the same trends. The latter should ideally be transferable to other geographical areas, at a sufficiently fine scale to enable trends at sub-national scales to be appreciated. We will then be able to better understand and assess tourist flows at a global scale, and to assess the place of the Arctic in the global tourism system.

## 2 Data and Methods

### 2.1 A quick literature review on databases of global tourist flows

- 14 A large number of data sources have been mobilized in recent years to study tourist flows and practices in a quantitative way. This section presents a quick review of the literature, ranging from the use of institutional data to individual digital traces.

#### Institutional data

- 15 The World Tourism Organization (UNWTO<sup>6</sup>) compiles annual data on incoming flows of people, including tourists, based on censuses in more than 150 countries. This paid dataset is more or less complete, accurate and updated depending on the country and has served as the basis for many studies on tourist flows. Among the most recent studies, Lozano and Gutiérrez [2018] analysed the main incoming and outgoing flows between each country. In particular, they studied the overall structure of the network formed and the levels of centrality of each country according to different indicators and algorithms (degree of centrality, pagerank, *etc.*). They also identified sub-networks, or patterns of 3 to 4 countries with the closest links.

#### Air flows as a proxy for tourist flows

- 16 As the increase in global tourism flows goes hand in hand with the increase in air traffic, Gabrielli *et al.*, [2019] used data provided by a private company that provided information on the number of monthly passengers between 239 countries and regions between 2010 and 2018. By decomposing the time series to extract seasonality, they were able to show that some countries are linked by different forms of tourism, whether mass tourism or more individual tourism. However, the data used is not free and the scale studied (the country) does not make it possible to show links between sub-regions.
- 17 Over the past ten years, we have observed the emergence of a new type of data: digital traces. The following paragraphs present a review of their use in the study of tourist flows and practices at different scales.
- 18 Telephony data are widely used in the study of national and local mobility [see Barbosa *et al.*, 2018 ; Blondel, Decuyper and Krings, 2015]. However, these data are provided by national operators and do not allow the identification of flows at the supranational level [Ahas, Silm and Tiru, 2018]. Conversely, photo and content sharing sites (*e.g.* online social networks) while not fully representative of the population, are more ubiquitous as they are potentially usable anywhere in the world [Longley, Adnan and Lansley, 2015]. Users of these platforms can generate geolocalized data that can be used in the study of mobility and migration at any scale.

#### Photo sharing websites

- 19 To our knowledge, the first studies on tourist places based on data from the Internet were carried out using data from photo-sharing sites on platforms such as *Flickr*<sup>7</sup> [Girardin *et al.*, 2008 ; Sun *et al.*, 2013] or *Panoramio* [Schlieder and Matyas, 2009]. These studies were more aimed at highlighting the attractive areas or temporal dynamics of

different tourist sites, although Girardin and his colleagues insisted on the possibility of reconstructing preferential paths and local flows. Still using data from *FlickrR*, Junker, Akbar and Cuquet [2017] created a network where the nodes are photos posted in a given geographical mesh and where the links between two nodes are weighted by the number of users having posted images in the two areas considered. Their paper does not detail the geographical centralities and focuses only on the characteristics of the resulting network, in particular on its scale-free nature, *i.e.* a small number of locations are highly connected to the others, acting as hubs.

### Twitter as a proxy to study tourist flows

- 20 However, for reasons of easier access to individual public and geolocalized data, most studies on international flows have been carried out using data collected on *Twitter*. We can thus note the work of Hawelka *et al.*, [2014] based on approximately one billion geolocated messages (or tweets) collected worldwide in 2012. By analysing the number of tweets sent by each user per country, they were able to establish flow matrices between countries of residence and countries visited occasionally, notably for tourism purposes. They defined mobility profiles for each country as well as the temporal flows linking each country. By comparing with various data from other sources, they were able to establish that *Twitter* was a good proxy for analyzing and quantifying global mobilities, including tourism flows.
- 21 Other studies, such as the one conducted by Bassolas *et al.*, [2016], quantified and analysed the attractiveness and the links between the 20 tourist sites that generated the greatest number of digital traces, taking into account the countries of residence of the people behind the tweets. Provenzano, Hawelka and Baggio [2018] carried out a comparative study between the UNWTO data and the geolocalized tweets used in the paper by Hawelka *et al.*, [2014] for intra-European flows. Their approach, based on graph analysis, first shows a tendency to preferentially visit neighbouring countries, but points some differences in the spatial distribution of visitors between the two datasets.
- 22 The development of online platforms allowing short-term accommodation rentals between individuals and the resulting online data may also seem very promising for studying tourism practices.

## 2.2 Use of *Airbnb* data in the literature

- 23 *Airbnb*, founded in 2008, allows owners willing to sublet part or all of their residence to people looking for accommodation for a relatively short period of time. In return, the platform collects a percentage of the transaction. The growth in the number of *Airbnb* users has been very rapid, even exponential. Nowadays, there are about 7 million accommodation offers in more than 100,000 cities<sup>8</sup> worldwide, ranging from boat cabins to villas and rooms in private homes.
- 24 Work based on the data retrieved from *Airbnb* began in mid-2010, facilitated by a freely available online python-based script<sup>9</sup> [Slee, 2015]. This script enables the retrieval of "listings", *i.e.* the available accommodation and all associated information (from the number of rooms to the rental price and the type of accommodation) in a given geographical area. That said, just because the code is available does not mean that



researchers unfamiliar with programming can use it. Thus, some websites use this script and put some information online for free, but only for a few large cities. We can note here for example *Inside Airbnb*<sup>10</sup> for large international cities, or the *Airbnb Observatory*<sup>11</sup> which focuses on large French cities. While these two projects are based on a rather militant approach that aim for transparency, paying alternatives also exist, for example through the services of the Airdna website<sup>12</sup>, whose slogan is « *Analyze Easier. Invest Wiser. List Smarter. Succeed Faster* ».

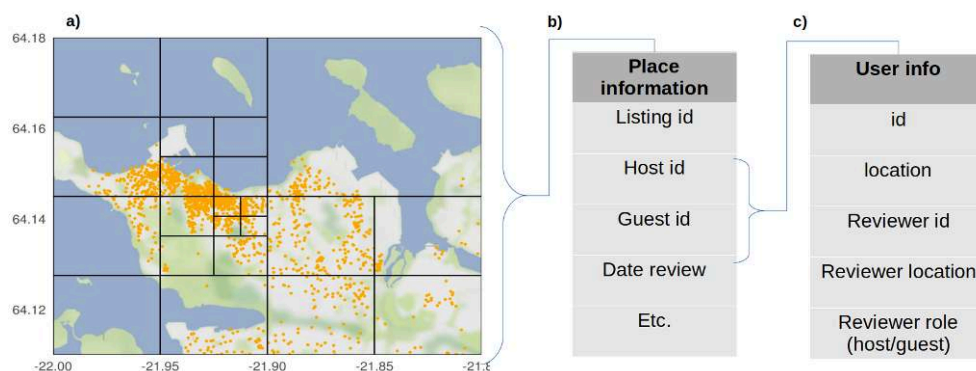
- 25 A review of the literature, well supported by other recent reviews on *Airbnb* [Dann, Teubner and Weinhardt, 2019 ; Guttentag, 2019 ; Prayag and Ozanne, 2018] shows that a large number of scientific articles questions the company's business model. Thus, though *Airbnb* does not lease any of its own real estate and is said to belong to the “*sharing economy*”, the company would be worth between \$31 to \$37 billion in 2019<sup>13</sup>. To compare, the Marriott Group, the world's leading hotel group, which owns more than 6,900 hotels (including the Hilton Group) and 1,317,168 rooms around the world<sup>14</sup>, has a market capitalization of \$42.8 billion in 2019<sup>15</sup>. Other scientific articles question *Airbnb* impact on local society like competition with the traditional hotel industry, tax evasion, difficulty in finding long term accommodation [Adamiak, 2018 ; Bouquet, Vacher and Vye, 2019 ; Eliasson and Ragnarsson, 2018 ; Gutiérrez *et al.*, 2017]. The phenomenon is probably accentuated in Iceland where 99% of the population had access to the Internet in 2018 [World Bank, 2020]. In Reykjavík, the development of *Airbnb* has led to a turmoil in the real estate market, where more than 5% of housing were available on the rental platform in 2016<sup>16</sup>. Thus, home rentals can therefore be seen as additional income, especially in the context of the crisis of 2008, and this in turn helps to welcome the flow of tourists [Mermet, 2017].
- 26 Among all the studies done using *Airbnb* data, none of them address - to our knowledge - the potential of these data in the analysis of flows (tourist or not). Teubner [2018] nevertheless questions the relations between hosts and guests in the form of an analysis of a-spatial graphs obtained from *Airbnb* data in 16 American cities. The nodes are hosts, guests (or both) linked together by a validated transaction (one person left a comment after hosting). The author then compares the average number of links according to the type of node (host or guest) and concludes that people offering hosting have higher levels of centrality than others in the graph he has created. But the spatial aspects are unfortunately discarded to the detriment of a mathematical description of the network that does not shed light on the relationships between the cities studied.
- 27 *Airbnb* data has the advantage of allowing the study of the most frequented tourist places over the years, and also to appreciate the flows between different geographical areas. Therefore, focusing on temporary accommodation in Iceland that today is mainly available on *Airbnb* and occupied by tourists, could contribute to the picture of tourism flows in the Arctic. From this information, we will be able to assess our methodology by comparing our *Airbnb* data recorded with the tourist flows recorded by the Icelandic official statistics.

## 2.3 Collecting Airbnb information

### 2.3.1 Retrieve the accommodations available in a given area

- 28 As Tom Slee points out, his script sometimes encounters difficulties in adapting to the change in the organization of the *Airbnb* website<sup>17</sup>. As we are not very familiar with the python language, we have developed a script in R language allowing us to perform a similar task. This script is public and accessible to the programming community for improvement<sup>18</sup>. Its working principle is relatively simple: (1) we first enter the geographical boundary of a polygon. (2) If the number of responses is greater than 300 (the limit of results displayed by *Airbnb*), the polygon is subdivided and the search continues until all rental offers are retrieved. This approach is similar to the *INSEE* natural level grid method<sup>19</sup>, where the territory is divided into polygons of different sizes so that no tile contains more than  $n$  items (here 300 *Airbnb* dwellings) (fig. 2.a). (3) All the information relating to the accommodation is then collected: its location, its number of beds available, its average rating, the number of comments, the owner's identifier, etc. Information on 7,523 dwellings (for 4,398 owners) has thus been recorded as of 12 September 2019 for the whole of Iceland. Most quantitative studies on *Airbnb* stop at the analysis of such a dataset. In our case, we would like to collect more information in order to be able to create flow matrices and have a more precise view of the link between Iceland and the rest of the world.

Figure 2 : Airbnb data acquisition process.



Data collection framework, example for Reykjavik. The geographical boundary of the city is subdivided in such a way that each bounding box contains a maximum of 300 accommodations, here represented by the orange dots (a). From this information about the accommodations, we retrieve the comments left by the guests (b). These comments are associated with the identifier of the place of accommodation, the owner and guest IDs and a date. From the owner and guest IDs, we then retrieve their declared place of residence, and all comments they have received or left from Airbnb users (c).

Source: Airbnb/Stamen/Authors

### 2.3.2 Data collection for a spatialized network

- 29 *Airbnb* users can rate and leave comments on the accommodations they have been to - although this is not always the case. This information appears on the web page of the place of accommodation (fig. 2.b). From the list of dwelling, we queried the API to retrieve all comments, focusing on the person's ID, their role (guest or host), the date of the comment, and the declared place of residence (fig. 2.c). We were able to retrieve around 470,000 comments left by nearly 300,000 people. From this data collection, it is

already possible to create an a-spatial relational network of hosts/hosts like Teubner [2018]. However, knowing the domicile indicated, it is possible to spatialise this network and focus on the relations between the places of origin of the people accommodated in relation to the places they have visited, and thus to assess the flows to Iceland *via* origin-destination matrices.

- 30 By collecting the comments left and received by the 300,000 people directly linked to Iceland *via Airbnb*, it is possible to obtain the places they actually visited and the people they accommodated (repeating the step described in fig 2.c). In the end, we get more than 3.1 million dated comments left by 1.94 million people, all concerning 100,000 locations around the world (the homes of hosts and guests).
- 31 The coordinates of the localities were then retrieved using free geographic directories such as *Geoname*<sup>20</sup> and *Datasciencetoolkit*<sup>21</sup> or proprietary directories such as *Foursquare*<sup>22</sup>. In order to validate the correct location of the declared domicile, we kept only the places appearing in at least two of these bases with distances of less than 20 kilometers. The missing places were then searched in *Google*, using the place API<sup>23</sup>, which only allows a few free daily queries. Note that these places of residence are declarative, sometimes approximate (region or country) and may have changed over time - only the last place of residence declared by the user is filled in.
- 32 The data collected could be used to analyse the flows linked more or less directly to Iceland and to assess the centrality, which can be interpreted as a relative level of preference for this country among the other visited destinations.

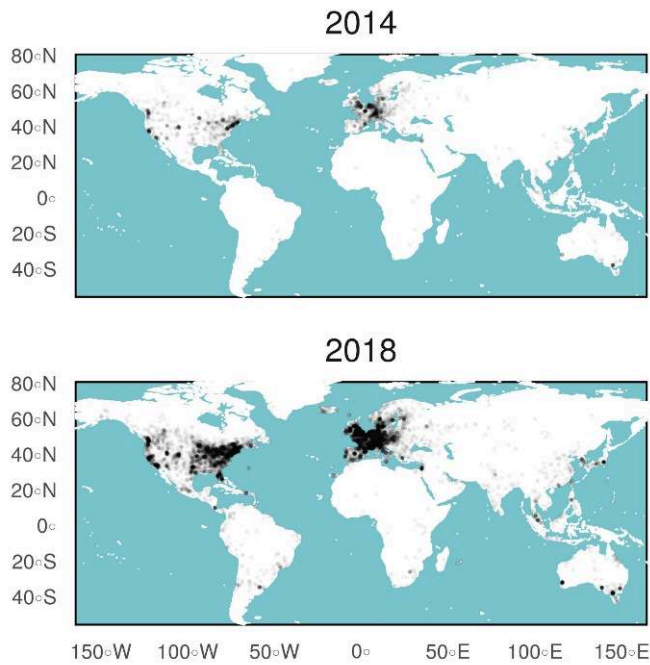
## 3 Results

### 3.1 Comparison with institutional data

#### 3.1.1 Impact of *Airbnb* on the classic hotel business

- 33 As can be seen in figure 3, the spatial reach of people visiting Iceland *via Airbnb* has expanded significantly between 2014 and 2018, although the majority of people appear to come from North America and Europe, and increasingly from East Asia (China, Japan, Taiwan).

Figure 3: Maps of the places of residence of Airbnb users in Iceland.

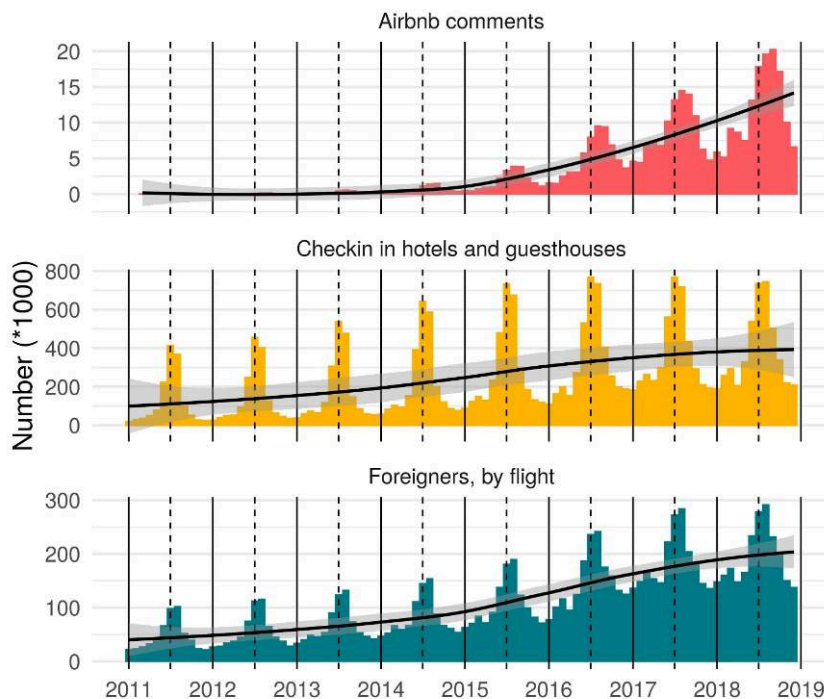


Declared residence of Airbnb users who have visited Iceland between 2014 (~9000) and 2018 (~150 000). The darker the dot, the more users there are.

Source : Airbnb 2019 / Authors

- 34 While this increase in *Airbnb* users in Iceland probably follows a mechanical effect of increased use of this platform on a global scale, its share in real tourism needs to be estimated. To do so, we will mobilize two types of institutional data from the Icelandic Statistical Office<sup>24</sup>, namely the number of passengers arriving at Keflavik International Airport (more or less the gateway to Iceland) and the number of check-ins at hotels or guest houses. This information is available by month and by nationality. Figure 4 below shows the number of registrations by month according to these three sets of data, between 2010 and 2018.

Figure 4: Comparison of Airbnb, hotels and guesthouses and international visitors, by month.

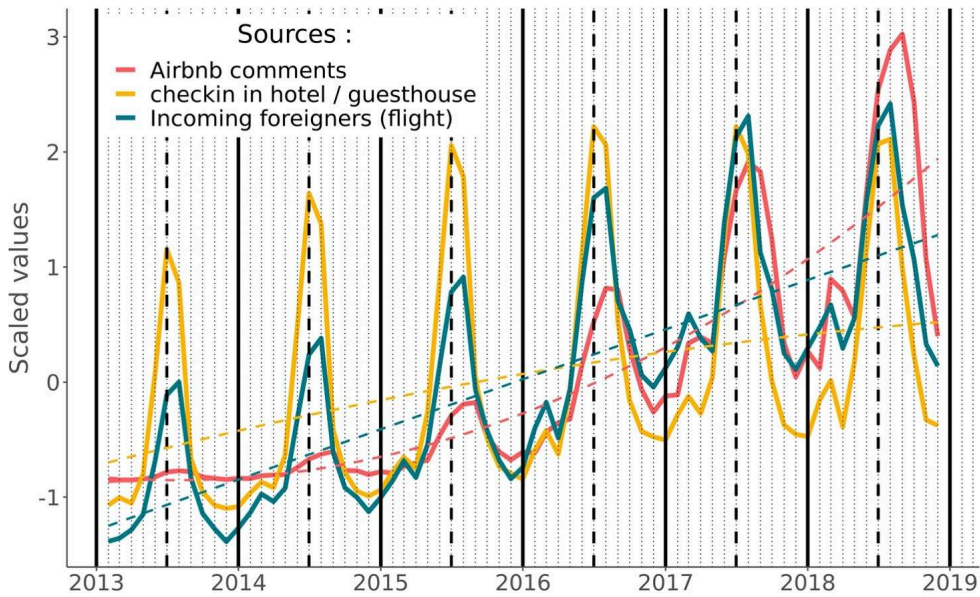


Monthly evolution of the number comments on Airbnb (top), check-in in hotels and guesthouses (middle) and number of international visitors arriving by air (bottom). Plain vertical lines stand for January, the dashed lines for July. The trend, in black, is calculated using the loess regression method.

Source : Airbnb 2019 / Iceland Statistic 2020 / Authors

- 35 First of all, we can note the importance of the summer period on tourism activity, with the appearance of a second peak in February/March starting in 2016. Overall, we observe an upward trend, even more marked among *Airbnb* users. This last point can be explained by the fact that we only collected information on accommodation that was online in September 2019, hiding those that have been removed from the platform. The introduction of restrictions on *Airbnb* rentals in Iceland from 2017 (e.g. 90 days maximum rental period)<sup>25</sup> does not appear to have slowed the growth in the use of the platform. This can be explained by the fact that the tourist season is mainly concentrated in summer, over two or three months.
- 36 It should be noted, however, that there was stagnation in the number of hotel registrations between 2017 and 2018, concomitant with a growth reduction of the number of entries for the same period. The number of *Airbnb* users is well below the number of hotel registrations, probably because traditional hotels remain the preferred offer. Furthermore, the institutional data counts all members of the same group where our approach only takes into account the owner of the *Airbnb* account, even if the reservation was made for several people. Moreover, users of the platform do not necessarily leave comments, therefore no trace of their visit. Nevertheless, this threefold observation (increase in the use of *Airbnb*, stagnation in the hotel industry and a slowdown in the number of entries) suggests a strong link between these three indicators, as shown in figure 5, which presents this data, re-scaled.

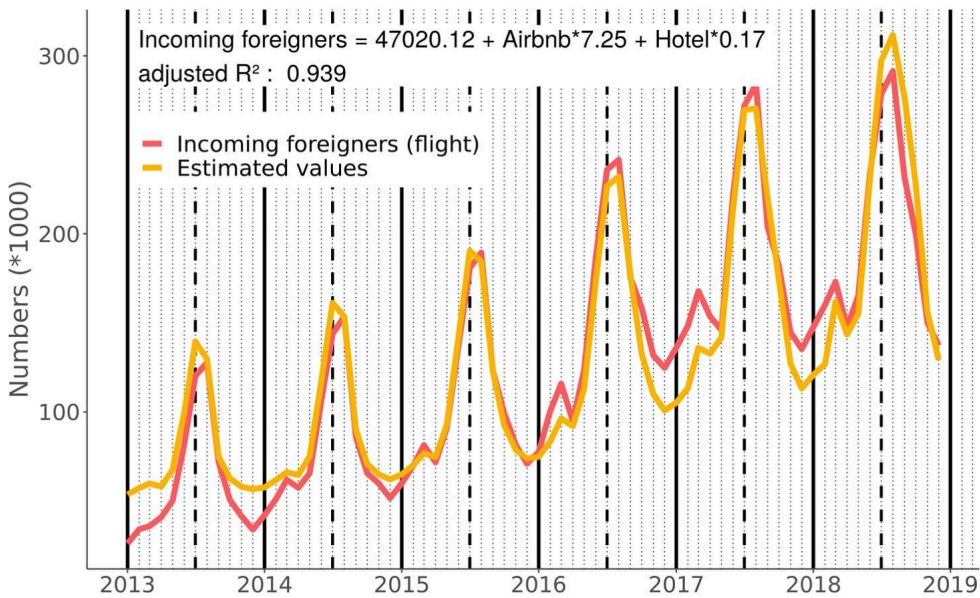
Figure 5 : Monthly evolution of comments on Airbnb, check-in in hotels and guesthouses and number of foreign visitors (scaled).



Bold vertical lines stand for January, the dashed line for July.

Source : Airbnb 2019 / Iceland Statistic 2020 / Authors.

Figure 6 : Observed and estimated number of passengers per month.



Bold vertical lines stand for January, the dashed line for July.

Source : Airbnb 2019/ Iceland Statistic 2020 / Authors.

- 37 We then applied a multiple linear regression of type  $y = a + b \cdot x_1 + c \cdot x_2$ , where  $y$  is the number of incoming passengers,  $x_1$  the number of monthly comments on *Airbnb* and  $x_2$  the number of check-ins in hotels,  $a$ ,  $b$  and  $c$ , coefficients. The underlying hypothesis that foreign visitors are distributed mainly in hotels and *Airbnb* rentals seems to be

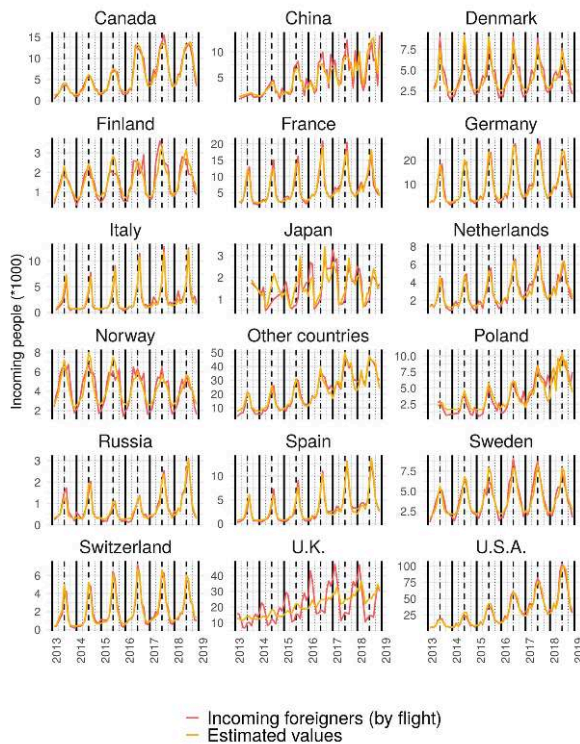
validated, as shown by the result of the multiple linear regression (Figure 6) - despite an underestimation of the number of winter visitors from 2017 onwards.

- 38 *Airbnb* users constitute a growing and important part of tourism in Iceland, although due to the biases mentioned above it is not exactly quantifiable. The next section will assess the extent to which this relationship is verified according to the countries of origin of foreign visitors.

### 3.1.2 Various tourist profiles depending on the country of origin

- 39 We can also apply the multiple regression model between the number of passengers, the number of registrations in conventional hotels and the number of *Airbnb* rentals per country (Figure 7), in order to assess whether the national practices of the *Airbnb* platform correspond to the global model applied previously. In the majority of cases, it is indeed possible to estimate the number of incoming visitors extremely accurately from the hotel and *Airbnb* data. With the exception of people coming from the United Kingdom, the evolution of multi-year trends, *i.e.* the order of visitor magnitudes per year, is systematically replicated. Summer and early spring peak visitor numbers coincide for most countries.

Figure 7: observed and estimated number of passengers per month and per country of origin.

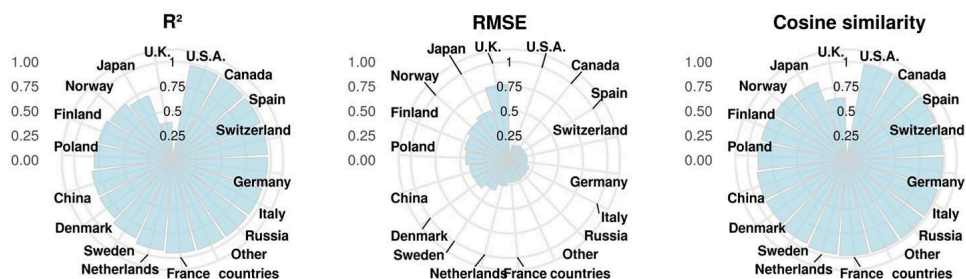


Bold vertical lines stand for January, the dashed line for July.

Source : Airbnb / Iceland Statistic 2020 / Authors

- 40 Different indices commonly used to compare two time series, namely the cosine similarity [Provenzano and Baggio, 2019], the root mean square error and the  $R^2$  determination coefficient, were all applied to each country to synthesize the model deviations (figure 8).

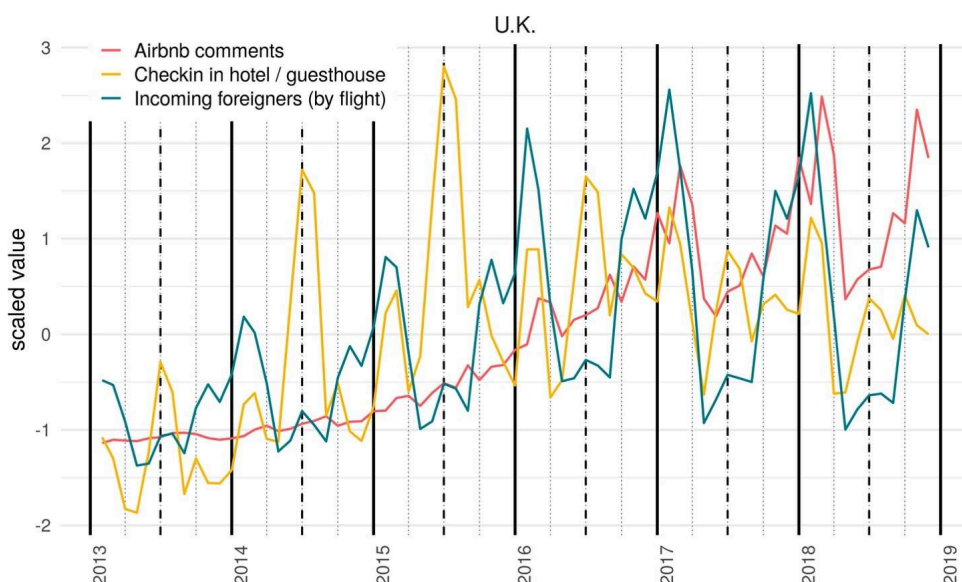
Figure 8: Values of different indicators of proximity between observed and estimated time series by country.



From left to right, the correlation coefficient, the root mean square error and the cosine similarity.

- 41 The differences between the predictive model and the observed data are overall very small for most countries. Nevertheless, the R<sup>2</sup>s for the United Kingdom, Japan and Norway are less than 0.75. For Japan, this relatively low correspondence could be explained by a low number of users reporting on *Airbnb* in Iceland. For Norway, it might be due to the fact that users of *Airbnb* tend to visit Iceland during the summer while official data shows also two peaks of visit in spring and fall, that may point to different temporal practices between *Airbnb* users and followers of classic hotels. In the case of the United Kingdom, the discrepancy could be due to anomalies in the institutional data up to 2017. Indeed, as shown in figure 9, peaks in hotel occupancy are observed during the summer period, while peaks in entry by air to the territory are recorded during the spring. Other origins, such as entry into the territory by cruise ships have been sought, but without convincing results.

Figure 9: Number of comments on Airbnb, check-in in hotels and visitors from the United Kingdom (2013-2019).



Source: Airbnb / Iceland Statistic 2020 / Authors

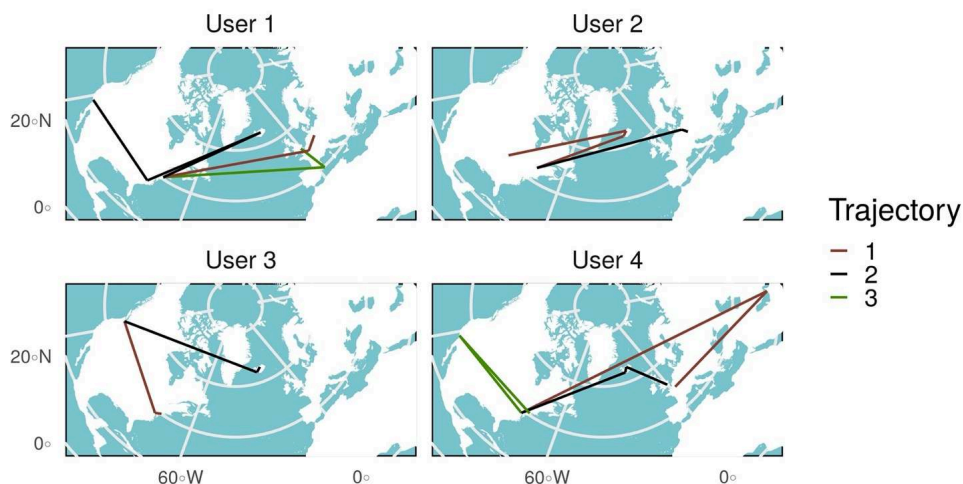


- 42 This section has demonstrated a constant increase in the overnight stays recorded on the *Airbnb* platform in accordance with the official statistics. It has also shown the close link between the number of entrants to Iceland and the number of check-ins in hotels and *Airbnb* accommodation regardless of the country of origin of the travellers. As a result, seasonal variations by country of origin of tourists are well reflected in the *Airbnb* data. Thus, this confirms that the *Airbnb* data may be a good proxy for estimating the number of travellers visiting Iceland annually.
- 43 We are now going to lay down the methodological bases making it possible to assess to what extent the *Airbnb* data collected in Iceland can be used to analyse tourist flows, particularly in the Arctic regions.

### 3.2 Toward an Airbnb users trip analysis

- 44 We have already collected information on all the places visited by people who have been to Iceland. It is then possible to reconstruct individual trajectories when the time difference between two comments is less than a certain period of time, *e.g.* 30 days (figure 10, example for 4 people).

Figure 10: Example of 3 trajectories for 4 users.

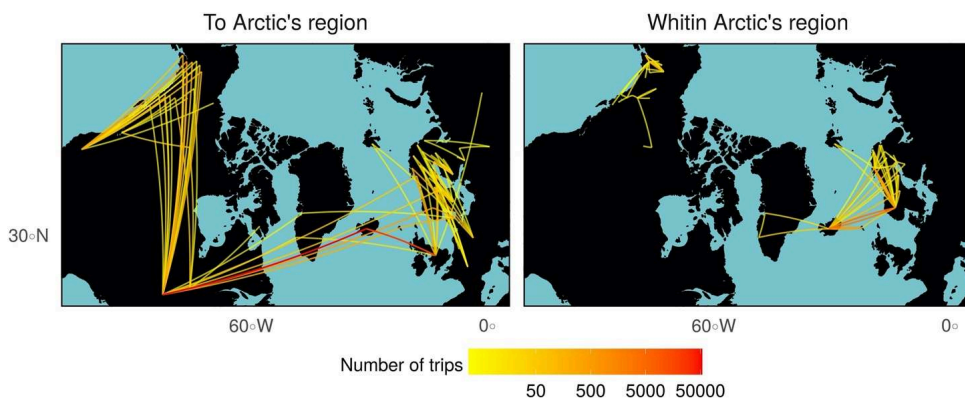


Each color corresponds to a path defined by comments posted within 30 days of each other.

- 45 This possibility of tracking users obviously raises ethical and privacy issues that have already been discussed [Cebeillac and Rey-Coyrehourcq, 2019]. When it comes to this point, it is worth a reminder that the retrieved data are publicly accessible and represent only a sample - we have only the comments while *Airbnb* has all the information about the users. However, from a user privacy perspective, the collection of individual data online for research purposes is permitted by the GDPR.
- 46 Aggregating trajectories makes it possible to obtain origin-destination matrices, which have the particularity of going beyond the "home-destination" flow since they take into account the intermediate localities visited. For display and interpretation purposes, we only select trajectories with at least one location in our study area and the localities are grouped together following the *dbscan* algorithm [Ester *et al.*, 1996], according to a distance of 200km. We then obtain more than 480,000 flows between 1306 destinations.

- 47 Given the difficulties of displaying so many flows in a comprehensive way, we have focused on the two main sources of flows towards the Arctic zones (figure 11, left) and between them (right). Most of the trips to Alaska thus start from the East and West coasts of the United States. The latter is also one of the most emitting areas for Airbnb users towards the Arctic regions with the United Kingdom, especially towards Greenland. Nevertheless, many flows to the Scandinavian Peninsula also come from France, Germany and the Netherlands.
- 48 If we look at the flows within the circumpolar zone, we can note that Alaska is disconnected from the European network which includes Greenland. If by construction it is not surprising that Iceland is one of the main sources or origins of the flows, we can observe a large number of trips within the Scandinavian Peninsula too. The trips to Greenland pass through Iceland, while those to Svalbard leave mainly from Norway. Figure 11 also highlights how Russian Arctic is isolated especially considering that its northern part is highly populated. Indeed, Northeast Canada shows no flow either but its poorly inhabited at the contrary.

Figure 11 : The two main flows toward the Arctic regions (left) and within the circumpolar zone (right).



As Iceland records a lot of flows, we opted for a logarithmic colour scale in order to better highlight the other flows.

Source : Airbnb 2019 / Authors.

- 49 Further network analysis could then be done, such as applying community detection algorithms [Fortunato, 2010]. This could highlight geographical sub-areas where interactions are strongest between them, and therefore shows the position of Arctic regions in a more global tourism network. Nevertheless, it would be necessary to collect the same information for the whole Arctic zone for a more in-depth analysis.

## 4 Discussion & perspective

- 50 The explosion of polar tourism and the Icelandic success have given rise to much reflection in neighbouring countries such as Greenland. Tourism also gradually changes the way of life of local populations who benefit and suffer from it. Indeed, Arctic tourism is a significant source of incomes but its impacts are complex. It has a severe impact on the fragile Arctic environment. However, Arctic tourism allows certain rural and isolated territories to be reborn as its rise has launched a process of

urban exodus. In fact, the main tourist destinations are not cities but the Arctic countryside especially the coastal and mountain edge villages. Unfortunately, the biggest winners are often from the South and not within the polar regions where there is risk to create new mono-activity economic profiles. Indeed, tourism sometimes represents the main economic sector as it is the case of the town of Rovaniemi in Finnish Lapland, up to it was nicknamed "Santa Claus village". This refers to the process of folklorization of the Arctic cultures.

- 51 This paper focuses mainly on Iceland, and we have retrieved the information on housing available in September 2019, which implies that the data collected is underestimated compared to the actual use of the platform. But the method used allows the indirect collection of information on other areas, even if the method should be extended to the whole Arctic zone in order to better appreciate the flows and interactions. Nevertheless, our approach offers the advantage of being easily transferable to other Arctic (and non-Arctic) regions due to the global distribution of Airbnb and the ease to collect the data (the code is available online). This is subject to Airbnb's penetration rate being sufficiently high in the study area. Further studies could also be carried out, and especially the methodology should be able to assess the impact of the coronavirus epidemic on global tourism and its repercussions in the Arctic regions.
- 52 One other limitation of our approach is that relies on *Airbnb* users posting comments, which is not always the case. Also, the use of the platform certainly varies both in tourist destinations - absence or competition between platforms - and among tourists according to their socio-geographical area and individual characteristics.
- 53 Another important point is that not all tourists are counted in the *Airbnb* data. Conversely, not all *Airbnb* users are tourists. This concern is somehow related to the delicate issue of the definition of a tourist which varies from one country to another. Thus, the difficulty to work at international level, is recurrent regardless of data. At least, our approach makes it possible to overcome political boundaries that statistics cannot. Indeed, as the etymology reminds us, statistics are a matter for the states and are rarely, if ever, completely comparable.
- 54 Thus, even in the Russian Arctic, where tourism is poorly developed and statistics about this activity are little or unavailable, Airbnb accommodation is available. For example, Norilsk, a closed city of 100,000 inhabitants, offers about 60 Airbnb accommodations and they are all trackable.
- 55 Therefore, although the data on tourism in the Arctic should always be considered with caution as the figures are low, we hope that this study can open up new avenues for research, like a better understanding in tourists practices by country of origin, the dynamics of tourist flows, etc. It is also possible to explore the impact of the different boundaries of the Arctic, as the figures are currently given in those of the Arctic Human Development Report [Larsen & Fondahl 2014]. It would then be possible to compare the Canadian mid-Arctic and the Norwegian high Arctic, for example.

---

## BIBLIOGRAPHY

- ADAMIAK Czesław, 2018, "Mapping Airbnb supply in European cities," *Annals of Tourism Research*, 71, pp. 67–71.
- AHAS Rein, SILM Siiri, TIRU Margus, 2018, "Measuring Transnational Migration with Roaming Datasets," p. 5.
- BARBOSA Hugo, BARTHELEMY Marc, GHOSHAL Gourab, JAMES Charlotte R., LENORMAND Maxime, LOUAIL Thomas, MENEZES Ronaldo, RAMASCO José J., SIMINI Filippo, TOMASINI Marcello, 2018, "Human mobility: Models and applications," *Physics Reports*, 734, pp. 1–74.
- BASSOLAS Aleix, LENORMAND Maxime, TUGORES Antònia, GONÇALVES Bruno, RAMASCO José J, 2016, "Touristic site attractiveness seen through Twitter," *EPJ Data Science*, 5(1).
- BLONDEL Vincent D, DECUYPER Adeline, KRINGS Gautier, 2015, "A survey of results on mobile phone datasets analysis," *EPJ Data Science*, 4(1).
- BOUQUET Claire, VACHER Luc, VYE Didier, 2019, "Que nous dit l'offre Airbnb sur l'évolution des territoires touristiques ? Le cas de La Rochelle/Île de Ré," *Mappemonde*, 125.
- CEBEILLAC Alexandre, REY-COYREHOURCQ Sébastien, 2019, "Webscraping : enjeux techniques et éthiques," *Données individuelles localisées : enjeux juridiques, éthiques et techniques à l'heure du RGPD*.
- DANN David, TEUBNER Timm, WEINHARDT Christof, 2019, "Poster child and guinea pig – insights from a structured literature review on Airbnb," *International Journal of Contemporary Hospitality Management*, 31(1), pp. 427–473.
- ELÍASSON Lúðvík, RAGNARSSON Öundur Páll, 2018, "Short-term renting of residential apartments: Effects of Airbnb in the Icelandic housing market."
- ESTER Martin, KRIEGLER Hans-Peter, SANDER Jorg, XU Xiaowei, 1996, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *KDD-96 Proceedings*.
- FAY, Gini., KARLSDÓTTIR, Anna., 2011. "Social indicators for arctic tourism: observing trends and assessing data". *Polar Geography*, 34, pp. 63–86, DOI: 10.1080/1088937X.2011.585779.
- FORTUNATO Santo, 2010, "Community detection in graphs," *Physics Reports*, 486(3–5), pp. 75–174.
- GABRIELLI Lorenzo, DEUTSCHMANN Emanuel, NATALE Fabrizio, RECCHI Ettore, VESPE Michele, 2019, "Dissecting global air traffic data to discern different types and trends of transnational human mobility," *EPJ Data Science*, 8(1).
- GIRARDIN Fabien, CALABRESE Francesco, DAL FIORE Filippo, RATTI Carlo, BLAT Josep, 2008, "Digital footprinting: Uncovering tourists with user-generated content," *IEEE Pervasive computing*, 7(4), pp. 36–43.
- GRENIER, Alain A., 2009. "Conceptualisation du tourisme polaire : cartographier une expérience aux confins de l'imaginaire". *Téoros. Revue de recherche en tourisme*, 28, pp.7–19.
- GUTIÉRREZ Javier, GARCÍA-PALOMARES Juan Carlos, ROMANILLOS Gustavo, SALAS-OLMEDO María Henar, 2017, "The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona," *Tourism Management*, 62, pp. 278–291.
- GUTTENTAG Daniel, 2019, "Progress on Airbnb: a literature review," *Journal of Hospitality and Tourism Technology*.

- HALL, C.Mickael., SAARINEN, Jarkko, 2010. "Geotourism and Climate Change. Paradoxes and Promises of Geotourism in Polar Regions." *Téoros. Revue de recherche en tourisme*, 29, pp. 77–86.
- HAWELKA Bartosz, SITKO Izabela, BEINAT Euro, SOBOLEVSKY Stanislas, KAZAKOPOULOS Pavlos, CARLO Ratti, 2014, "Geo-located Twitter as the proxy for global mobility patterns," *Cartography and Geographic Information Science*.
- HUDDART David, STOTT Tim, 2020, "The Arctic Islands: Svalbard and Iceland," in *Adventure tourism: environmental impacts and management*, Liverpool, Palgrave Macmillan, pp. 51–100.
- JÓHANNESSON Gunnar Thór, HUIJBENS Edward H., 2013, "Tourism Resolving Crisis?," in Müller Dieter K., Lundmark Linda, Lemelin Raynald H. (dir.), *New Issues in Polar Tourism*, Dordrecht, Springer Netherlands, pp. 133–147.
- JOLIVEAU Thierry, 2009. "Connecting Real and Imaginary Places through Geospatial Technologies: Examples from Set-jetting and Art-oriented Tourism". *The Cartographic Journal*, 46, pp. 36–45, DOI :10.1179/000870409X415570.
- JUNKER Christian, AKBAR Zaenal, CUQUET Martí, 2017, "The Network Structure of Visited Locations According to Geotagged Social Media Photos," in Camarinha-Matos Luis M., Afsarmanesh Hamideh, Fornasiero Rosanna (dir.), *Collaboration in a Data-Rich World*, Cham, Springer International Publishing, pp. 276–283.
- LARSEN, Joan Nymand, FONDAHL Gail. eds., 2014. *Arctic human development report (AHDR): regional processes and global linkages*. Copenhagen: Nordic Council of Ministers, 500 p.
- LASSERRE Frédéric, TÊTU Pierre-Louis., 2015. "The cruise tourism industry in the Canadian Arctic: analysis of activities and perceptions of cruise ship operators". *Polar Record*, 51, pp. 24–38, DOI : 10.1017/S0032247413000508.
- LONGLEY Paul A., ADNAN Muhammad, LANSLEY Guy, 2015, "The geotemporal demographics of Twitter usage," *Environment and Planning A*, 47(2), pp. 465–484.
- LOZANO Sebastián, GUTIÉRREZ Ester, 2018, "A complex network analysis of global tourism flows," *International Journal of Tourism Research*, 20(5), pp. 588–604.
- MERMET Anne-Cécile, 2017, "Tourism gentrification in time of crisis: the case of 101 Reykjavik," in Gravari-Barbas Maria, Guinand Sandra (dir.), *Tourism & gentrification in contemporary metropolises : international perspectives*, London & New York, Routledge, Contemporary Geographies of Leisure, Tourism and Mobility, pp. 52–74.
- MERMET Anne-Cécile, 2019, "Entendu / Entretien : Airbnb et la gentrification touristique des villes," *Urbanités*. <http://www.revue-urbanites.fr/entendu-entretien-mermet/>
- MÜLLER Dieter K., 2013, "Tourism and the definition of the Arctic," in Lemelin Raynald Harvey, Maher Patrick, Liggett Daniela (dir.), *From Talk to Action: How Tourism is Changing the Polar Regions*, Thunder Bay, Lakehead University Centre for Northern Studies, pp. 9–20.
- MÜLLER Dieter K., 2015, "Issues in Arctic Tourism," in Evengård Birgitta, Larsen Joan Nymand, Paasche Øyvind (dir.), *The new arctic*, Cham Heidelberg New York Dordrecht London, Springer, pp. 147–158.
- PRAYAG Girish, OZANNE Lucie K, 2018, "A systematic review of peer-to-peer (P2P) accommodation sharing research from 2010 to 2016: progress and prospects from the multi-level perspective," *Journal of Hospitality Marketing & Management*, 27(6), pp. 649–678.

- PROVENZANO Davide, BAGGIO Rodolfo, 2019, “A complex network analysis of inbound tourism in Sicily,” *International Journal of Tourism Research*.
- PROVENZANO Davide, HAWELKA Bartosz, BAGGIO Rodolfo, 2018, “The mobility network of European tourists: a longitudinal study and a comparison with geo-located Twitter data,” *Tourism Review*, 73(1), pp. 28–43.
- SCHLIEDER Christoph, MATYAS Christian, 2009, “Photographing a City: An Analysis of Place Concepts Based on Spatial Choices,” *Spatial Cognition & Computation*, 9(3), pp. 212–228.
- SLEE Tom, 2015, “Airbnb Data Collection: Methodology and Accuracy”. <http://tomslee.net/airbnb-data-collection-methodology-and-accuracy>. [Accessed April 26, 2019]
- STATISTICS ICELAND, 2017, “Main aggregated results, 2009-2017.” [Accessed November 28, 2019]
- STATISTICS ICELAND, 2019, “Population by municipality, sex, citizenship and quarters 2010-2019.” [Accessed November 28, 2019]
- STATISTICS ICELAND, 2020a, “International arrivals by modes of transport, 2009-2017.” [Accessed November 28, 2019]
- STATISTICS ICELAND, 2020b, “Number of employees in activities related to tourism 2008-.” [Accessed November 28, 2019]
- STEENHOLDT Naja Carina, CHIMIRRI Daniela., 2018. “Tourism & Quality of Life in Greenland: Exploration through farm Stays in South Greenlandic Settlements”. In *Arctic Yearbook* Available at: <https://arcticyearbook.com/arctic-yearbook/2018/2018-scholarly-papers/273-tourism-quality-of-life-in-greenland-exploration-through-farm-stays-in-south-greenlandic-settlements> [Accessed June 10, 2020].
- SUN Yeran, FAN Hongchao, HELBICH Marco, ZIPF Alexander, 2013, “Analyzing Human Activities Through Volunteered Geographic Information: Using Flickr to Analyze Spatial and Temporal Pattern of Tourist Accommodation,” in Krisp Jukka M. (dir.), *Progress in Location-Based Services*, Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 57–69.
- TEUBNER Timm, 2018, “The web of host–guest connections on Airbnb: a network perspective,” *Journal of Systems and Information Technology*, 20(3), pp. 262–277.
- VEIJOLA Soile, STRAUSS-MAZZULLO Hannah, 2019, “Tourism at the Crossroads of Contesting Paradigms of Arctic Development,” in Finger Matthias, Heininen Lassi (dir.), *The globalarctic handbook*, New York, NY, Springer Berlin Heidelberg.
- WORLD BANK, 2020, Indicator “Individuals using the Internet (% of population).” Available at: <https://data.worldbank.org/indicator/IT.NET.USER.ZS> [Accessed October 27, 2020].

## NOTES

1. <https://www.airbnb.com/> [Accessed November 28, 2019]
2. Statistics Norway: <https://www.ssb.no/> [Accessed July 28, 2020]
3. <https://www.rcinet.ca/eye-on-the-arctic-special-reports/arctic-tourism/> [Accessed July 20, 2020]
4. <https://bfmbusiness.bfmtv.com/entreprise/ces-6-destinations-touristiques-ont-explose-grace-agame-of-thrones-1218828.html> [Accessed July 20, 2020]
5. <https://edition.cnn.com/travel/article/iceland-canyon-closing-overtourism-justin-bieber/index.html> [Accessed September 28, 2019]

6. <https://unwto.org/fr> & <https://www.e-unwto.org/doi/book/10.18111/9789284420414> [Accessed September 28, 2019]
7. <https://www.flickr.com/>
8. <https://news.airbnb.com/fast-facts/> [Accessed September 28, 2019]
9. <https://tomslee.net/category/airbnb-data> [Accessed April 26, 2019]
10. <http://insideairbnb.com/behind.html> [Accessed June 26, 2019]
11. <http://observatoire-airbnb.fr/> [Accessed June 26, 2019]
12. <https://www.airdna.co/> [Accessed June 26, 2019]
13. <https://www.forbes.com/sites/deniselyohn/2019/04/02/why-the-corporate-valuations-of-rent-the-runway-airbnb-and-lyft-may-be-completely-wrong/#7d5fd0f16453>  
[https://www.lemonde.fr/economie/article/2019/03/08/airbnb-joue-la-carte-de-l-amazonisation\\_5433193\\_3234.html](https://www.lemonde.fr/economie/article/2019/03/08/airbnb-joue-la-carte-de-l-amazonisation_5433193_3234.html) [Accessed August 26, 2019]
14. <https://marriott.gcs-web.com/annual> [Accessed August 26, 2019]
15. <https://www.zonebourse.com/MARRIOTT-INTERNATIONAL-14633490/> [Accessed August 26, 2019]
16. <https://splinternews.com/airbnb-has-made-it-nearly-impossible-to-find-a-place-to-1793856969> [Accessed August 26, 2019]
17. « There are continued problems getting the script to work with new AirBnB page layouts », <https://github.com/tomslee/airbnb-data-collection> [Accessed April 26, 2019]
18. <https://github.com/cebeillac/Rbnb>. Tested last time the 22/09/2020
19. INSEE: National (French) Institute of Statistics and Economic Studies. [https://www.insee.fr/fr/statistiques/fichier/4176281/documentation\\_DonneesCarroyees.pdf](https://www.insee.fr/fr/statistiques/fichier/4176281/documentation_DonneesCarroyees.pdf) [Accessed April 26, 2019]
20. <http://api.geonames.org/searchJSON?q=CITY> [Accessed June 08, 2019]
21. <http://www.datasciencetoolkit.org/maps/api/geocode/json?sensor=false&address=CITY>
22. [https://api.foursquare.com/v2/venues/search?client\\_id=ID&client\\_secret=SECRET&v=DATE&limit=500&intent=browse&near=VILLE&radius=5000](https://api.foursquare.com/v2/venues/search?client_id=ID&client_secret=SECRET&v=DATE&limit=500&intent=browse&near=VILLE&radius=5000). It is necessary to create an account and the number of free requests is limited to 1000 per day. [Accessed June 08, 2019]
23. <https://developers.google.com/places/web-service/search#TextSearchRequests> [Accessed June 08, 2019]
24. <https://px.hagstofa.is/pxen/pxweb/en/> [Accessed April 26, 2019]
25. <https://www.airbnb.com/help/article/2462/responsible-hosting-in-iceland> [Accessed September 10, 2020]

## ABSTRACTS

Tourist flows are increasingly important – at least up to the covid-19 crisis - and the Arctic regions are no exception. In particular, Iceland is experiencing an unprecedented tourism boom that began a decade ago. This goes hand in hand with the development of online short-term rental platforms such as *Airbnb*. This article questions how public information on the *Airbnb* website can be used to study tourist flows, with Iceland as a case study. Using a program we developed, we collected information about the accommodation available on the platform, as well as any comments left by visitors.

Each member's profile also includes the place of residence declared by the user, which makes it

possible to know the country of origin. We then compared this collected information with official statistical data. *Airbnb*, which is growing steadily, seems to be gaining ground on the classic hotel business. The *Airbnb* data reflects the officially recorded flows by country of origin. This makes the case for using this data for the Arctic region, for which statistical data is hard to find at a fine-scale.

Les flux touristiques sont de plus en plus importants et les zones arctiques ne font pas exceptions. Plus particulièrement, l'Islande connaît un boom touristique sans précédents depuis une dizaine d'années. Ce dernier va de pair avec le développement de plateformes de locations de courtes durées en ligne, comme *Airbnb*. Cet article interroge l'intérêt des informations publiques sur le site Airbnb pour les recherches sur les flux touristiques. L'Islande constitue dans ce cadre, le cas d'application. À l'aide d'un programme que nous avons développé sous R, nous avons collecté les informations sur les logements disponibles sur la plateforme, ainsi que tous les commentaires laissés par les visiteurs.

Le profil de chaque membre renseigne également le lieu du domicile déclaré par l'utilisateur, ce qui permet de connaître le pays d'origine. Nous avons ensuite comparé ces informations récoltées aux données statistiques officielles. Airbnb, en forte croissance semble prendre des parts sur l'hôtellerie classique. Les données Airbnb reflètent bien les flux enregistrés officiellement par pays d'origine. Une utilisation dans d'autres zones, notamment en Arctique ou des données statistiques ne sont pas disponibles à échelle fine est alors envisageable.

## INDEX

**Keywords:** Iceland, tourism, touristic flows, Airbnb

**Mots-clés:** Islande, tourisme, flux touristiques, Airbnb

## AUTHORS

### ALEXANDRE CEBEILLAC

Université de Rouen - UMR IDEES 6266

[alexandre.cebeillac@univ-rouen.fr](mailto:alexandre.cebeillac@univ-rouen.fr)

### YVETTE VAGUET

Université de Rouen - UMR IDEES 6266

[yvette.vaguet@univ-rouen.fr](mailto:yvette.vaguet@univ-rouen.fr)