

# Mineable or Messy? Assessing the Quality of Macro-Level Tourism Information Derived from Social Media

---

## **Roman Tilly**

Dept. of Information Systems  
and Information Management,  
University of Cologne,  
Germany

[tilly@wim.uni-koeln.de](mailto:tilly@wim.uni-koeln.de)

## **Kai Fischbach**

Chair in Information Systems  
and Social Networks,  
University of Bamberg,  
Germany

[kai.fischbach@uni-bamberg.de](mailto:kai.fischbach@uni-bamberg.de)

## **Detlef Schoder**

Dept. of Information Systems  
and Information Management,  
University of Cologne,  
Germany

[schoder@wim.uni-koeln.de](mailto:schoder@wim.uni-koeln.de)

## **Abstract**

Social media are a primary means for travelers to connect with each other and plan trips. They can also help tourism suppliers (e.g., by providing relevant information), thus overcoming the shortcomings of traditional information sources. User-generated content from social media has already been used in many studies as a primary information source. However, the quality of information derived thus far remains largely unclear. This study assesses the quality of macro-level information on the spatio-temporal distribution of tourism derived from online travel reviews in social media in terms of completeness, timeliness, and accuracy. We found that information quality increased from 2000 to 2009 as online travel reviews increasingly covered more countries, became available earlier than statistics reported by the United Nations World Tourism Organization (UNWTO), were highly correlated with the UNWTO statistics. We conclude that social media are a good information source for macro-level spatio-temporal tourism information and could be used, for example, to estimate tourism figures.

## **Keywords**

online travel reviews; global tourism; social media; virtual travel community; UNWTO; information quality

## Introduction

A growing number of research studies have investigated and documented the impact of social media on travelers and tourism suppliers (Leung, Law, van Hoof, & Buhalis, 2013; Zeng & Gerritsen, 2014). Travelers use user-generated content (UGC) as an information source prior to making purchase decisions (Cox, Burgess, Sellitto, & Buultjens, 2009; Fotis, Buhalis, & Rossides, 2011; Xiang, Wang, O'Leary, & Fesenmaier, 2014; Yoo & Gretzel, 2012) and to report on their experiences, thus contributing to other travelers' information sources (Kang & Schuett, 2013; Leung & Bai, 2013; Munar & Jacobsen, 2014). Tourism suppliers such as hoteliers, restaurant owners, destination marketing organizations, and travel agencies use various types of social media to support different business functions such as, for example, promotion and product distribution (Kasavana, Nusair, & Teodosic, 2010; Leung et al., 2013; Litvin, Goldsmith, & Pan, 2008; Schmallegger & Carson, 2008). They also collect, aggregate, and analyze UGC to learn determinants of customer satisfaction from online travel reviews (e.g., Li, Ye, & Law, 2013) or customer perceptions of a destination (e.g., Inversini, Marchiori, Dedekind, & Cantoni, 2010; Law & Cheung, 2010; Schmallegger & Carson, 2009), thus overcoming some of the shortcomings of traditional information sources (Wöber, 2003, p. 242).

There is, however, a research gap with respect to the quality of information from tourism-related social media. This should be an issue because it is only one-third of all travelers, at best, who actually produce UGC and whose behavior and preferences can thus be studied (Cox et al., 2009; Munar & Jacobsen, 2013; Xiang et al., 2014; Yoo & Gretzel, 2012). Nevertheless, sample representativeness, generalizability of results, interpretability of data, and reliability are seldom addressed in content studies of tourism-related UGC (Lu & Stepchenkova, 2014). Hence, the aim of this study is to contribute to a better understanding of the quality of information from this source. Specifically, we investigate the distribution of online travel reviews over countries and years and ask how their completeness, timeliness, and accuracy compares to a traditional source of such information in tourism, namely the United Nations World Tourism Organization's (UNWTO) annual "Compendium of Tourism Statistics" (UNWTO, 2014). If information quality is sufficient, tourism suppliers may want to compile statistics on global tourism from online travel reviews, eliminate some of the shortcomings of traditional sources, and substitute them with social media.

This article is organized as follows. The following section illustrates the role of social media for travelers and tourism providers, reviews existing studies that have analyzed tourism-related UGC with an emphasis on spatio-temporal analysis, and deduces from the literature appropriate criteria for information quality. In Sections 3 and 4, we describe the datasets and methodology applied to assess information quality and present the results. The final section presents our findings on the potential of tourism-related social media as a source for macro-level tourism information, the implications for tourism suppliers and researchers, limitations of this study, and possibilities for further research.

## Literature Review

### Social Media Use By Tourism Suppliers and Travelers

Tourism suppliers promote products and services on social media websites (Davidson & Keup, 2014; Gupta & Ramachandran, 2013; Lagrosen & Grundén, 2014; Munar, 2012; Tse, 2013), increase the attractiveness of existing websites and services by incorporating social media functionality, and encourage customers to give feedback and ratings (Chan & Guillet, 2011; Chen & Xie, 2008; Huang, 2012; Munar, 2011). Social media also provide opportunities for product distribution by seamlessly offering booking possibilities and responding directly to travelers' demands expressed through UGC (Inversini & Masiero, 2014; Lanz, Fischhof, & Lee, 2010; Noone, McGuire, & Niemeier, 2011). In

general, many researchers see new possibilities for better customer relationship management and even customer-driven innovation and new product and business model testing (Huang, Yung, & Yang, 2011; Kasavana et al., 2010; Kim & Hardin, 2010; Pantelidis, 2010; Rosman & Stuhura, 2013; Schmallegger & Carson, 2008). Nevertheless, suppliers still have to learn how to handle the new opportunity for customers to evaluate their service quality publicly and reach a large audience (Gu & Ye, 2014; Ong, 2012; Park & Allen, 2013).

Before and during a trip, travelers are in need of information because tourism products and services are often intangible experience goods, difficult to evaluate prior to consumption, are often costly, and involve a high level of emotional involvement and risk (Jeng & Fesenmaier, 2002; Lewis & Chambers, 2000; Mills & Law, 2004; Schmallegger & Carson, 2008). Hence, prospective travelers typically gather information from a variety of online and offline sources to reduce uncertainty and increase the quality of their tourism experiences (Buhalis & Law, 2008; Fodness & Murray, 1997; Jeng & Fesenmaier, 2002; Xiang et al., 2014). Information from other consumers (i.e., word of mouth) is particularly important and affects purchasing behavior (Davis & Khazanchi, 2008) because consumers hope to get advice independent of commercial influence and more credible than marketing messages (Allsop, Bassett, & Hoskins, 2007; Litvin et al., 2008). Social media and UGC facilitate electronic word of mouth (Litvin et al., 2008, p. 461; Tham, Croy, & Mair, 2013) and complement or even replace travelers' traditional information sources and communication channels (Allsop et al., 2007; Bieger & Laesser, 2004; Cox et al., 2009; Dickinger, 2011; Fotis et al., 2011; Litvin et al., 2008; Munar & Jacobsen, 2014; Xiang et al., 2014; Yoo & Gretzel, 2012). During or after a trip, travelers may share their thoughts through social media to reciprocate positive experiences to the suppliers, altruistically help other consumers, or because they enjoy it (Kang & Schuett, 2013; Munar & Jacobsen, 2014; Yoo & Gretzel, 2008, 2011).

Several studies have sought to quantify the use of social media among travelers (Cox et al., 2009; Fotis et al., 2011; Xiang et al., 2014; Yoo & Gretzel, 2012, p. 196). Some 20–45 percent of travelers use social media to search for information, evaluate alternatives, and plan trips. Only 5–30 percent of travelers, however, also create content to share experiences. Even though numbers on travelers' use of social media vary, individuals who create UGC are surely not in the majority (compared to all travelers). This raises important questions regarding the quality of information with respect to representativeness and reliability which content studies on tourism-related social media have seldom addressed (Lu & Stepchenkova, 2014).

### **Spatio-Temporal Information from Tourism-Related UGC**

Lu and Stepchenkova (2014) identified five broader topics with which content studies related to social media in tourism and hospitality have been concerned: (1) distilling consumers' perceptions and evaluations of service quality from UGC (mainly from travel reviews); (2) monitoring a destination's image as perceived by different market segments of travelers and in comparison to other destinations' images; (3) exploring the influence of UGC from consumers and suppliers on other travelers' decisions, evaluations of helpfulness, and issues of trust; (4) analyzing travelers' experiences and behaviors, discovering trends, and gaining insight into customers; and (5) deriving spatial and temporal mobility patterns from UGC (meta-) data.

However, spatio-temporal studies are very few. Most present approaches to derive tourists' movement patterns in cities from geo-tagged pictures in social media (e.g., Ardizzone, Di Miceli, La Cascia, & Mazzola, 2012; Girardin, Calabrese, Fiore, Ratti, & Blat, 2008; Girardin, Dal Fiore, Blat, & Ratti, 2007; Popescu, Grefenstette, & Moëllic, 2009). Thus, important points of interest and typical tourist paths in a city can be identified to help learn more about tourist flows. In addition, some studies have developed probabilistic models of tourist movement or have applied machine learning methods to identify recurring tourist movement paths from UGC data and recommend paths to new tourists (Lu,

Wang, Yang, Pang, & Zhang, 2010; Sun, Fan, Bakillah, & Zipf, 2013; Zheng, Zha, & Chua, 2012). The quality of information from UGC is not explicitly addressed in these studies, nor are analyses conducted at the macro-level—two facts that call for further research.

## Information Quality

Information quality is often defined as a multi-dimensional concept (Knight & Burn, 2005; Wang & Strong, 1996). In a review of 12 frameworks of information quality, Knight and Burn found 20 different dimensions of information quality (Knight & Burn, 2005). Which of them should be considered when assessing information quality depends largely on the context of information use and the information consumer (Knight & Burn, 2005; Strong, Lee, & Wang, 1997). For the purpose of this study, we empirically assess the quality of spatio-temporal macro-level information from tourism-related UGC with respect to the following dimensions of information quality (definitions according to Knight & Burn, 2005):

- Accuracy: The extent to which data are correct, reliable, and certified to be error-free
- Timeliness: The extent to which the information is sufficiently up to date for the task at hand
- Completeness: The extent to which information is not missing and is of sufficient breadth and depth for the task at hand

## Approach

The utility of social media as a source of spatio-temporal macro-level tourism information depends largely on how they compare to traditional sources. If the information quality of social media is equal to or superior to traditional sources, information consumers should consider (also) using social media; otherwise, they should stick with the traditional sources. We chose to assess the quality of spatio-temporal macro-level information with respect to the global distribution of travelers over time—that is, the number of travelers per year in each country. We derived this information from user-generated online travel reviews from a virtual travel community (VTC). However, before one develops models to estimate figures of travelers per year and country from travel reviews, information quality should be assessed. Hence, we want to know what quality of information on the global distribution of tourism is provided by online travel reviews compared to traditional sources? We assess information quality in terms of completeness (i.e., how many countries covered), timeliness (i.e., how long until information is accessible), and accuracy (i.e., how accurately is the global distribution of online travel reviews associated with figures from traditional sources).

We used the UNWTO's annual "Compendium of Tourism Statistics" (UNWTO, 2014) as a traditional source for information on the global distribution of tourism. It is perhaps the most comprehensive traditional source for information on the number of travelers per year and country.

## Data Sources

In March 2010, we collected data available in multiple languages, including Chinese, from a large international VTC; we assumed the data covered worldwide tourism. VTCs are arguably an important example of a comprehensive social media application in tourism. From a technological perspective, VTCs typically combine several Web 2.0 technologies such as RSS, AJAX, and Web services, as well as applications such as forums, chats, blogs, wikis, and picture slideshows (Dippelreiter et al., 2008). Hence, they facilitate the convenient and swift exchange of hyper-media UGC and connect many people worldwide. Their success in terms of user acceptance is generally attributed to their correspondence to the functional, social, and psychological needs of their members (Dippelreiter et al., 2008; Wang, Yu, & Fesenmaier, 2002). VTCs are also a common research object and data source in

tourism-related social media research (e.g., Bulchand-Gidumal, Melián-González, & González López-Valcárcel, 2013; Lee, Law, & Murphy, 2011; Park & Allen, 2013; Scott & Orlikowski, 2012, 2014).

We used self-developed software to browse the VTC website. The software began from the cover page, followed every internal link recursively, and collected data from travel reviews (e.g., timestamp, destination, author/user). The resulting dataset comprised data on approximately 7.89 million travel reviews for 232 countries for the years 1999 to 2010.

When assessing the quality of spatio-temporal macro-level information from travel reviews, it is important to have a dataset of travel reviews that is as comprehensive as possible. At the time of data collection, the VTC stated it provided about 30 million reviews and opinions, which presumably covered travel reviews as well as postings in forums. Hence, we also collected data on postings to be able to assess data completeness for postings and reviews. We counted 18.9 million postings, which means we found a total of 26.79 million, or 89.3 percent of travel reviews and postings. It is unclear whether proportionally more reviews or postings are missing, but one can assume that missing travel reviews and postings are about equally distributed.

The UNWTO's annual "Compendium of Tourism Statistics" includes data for more than 200 countries on inbound, domestic, and outbound tourism as well as on tourism industries, employment, and complementary indicators (UNWTO, 2012a). The UNWTO kindly provided us with an electronic copy of the 2012 Compendium covering the period 1995 to 2010 (as of April 4, 2012).

The UNWTO data are gathered according to several indicators. Inbound tourism is measured by "arrivals"—that is, "the flows of international visitors to the country of reference: each arrival corresponds to one inbound tourism trip" (UNWTO, April 2011, p. 13). Domestic tourism is measured by the number of "trips," defined as "activities of a resident visitor within the country of reference (either as part of a domestic tourism trip or part of an outbound tourism trip)" (UNWTO, April 2011, p. 15). "Departures"—namely, "the flows of resident visitors leaving the country of reference" (UNWTO, April 2011, p. 15)—are an indicator for outbound tourism. The total numbers of inbound, domestic, and outbound visitors are broken down again according to the length of stay ("overnight" and "same-day" visitors).

These indicators of the UNWTO's "Compendium of Tourism Statistics" are based on the broader conceptual framework of the "International Recommendations for Tourism Statistics" (IRTS) (UNWTO, April 2011, p. 21), which defines "travel" generally as any activity of a "traveler"—that is, "someone who moves between different geographic locations for any purpose and any duration" (Department of Economic and Social Affairs, 2010, p. 9, para. 2.4). "Trips" are temporally and spatially delimited travel activities (Department of Economic and Social Affairs, 2010, p. 9, para. 2.7).

Further, "tourism" is a subset of travel (Department of Economic and Social Affairs, 2010, p. 10, para. 2.12): a traveler is a "visitor" if he/she takes "a trip to a main destination outside his/her usual environment, for less than a year, for any main purpose (business, leisure or other personal purpose) other than to be employed by a resident entity in the country or place visited." In this case, the trip is a "tourism trip" (Department of Economic and Social Affairs, 2010, p. 10, para. 2.9, 2.12). Visitors may also be sub-classified into (1) tourists/overnight visitors if the respective trip includes a night stop or (2) same-day visitors/excursionists (Department of Economic and Social Affairs, 2010, p. 10, para. 2.13). Furthermore, the IRTS differentiates between inbound, domestic, and outbound travel/travelers/trips/tourism/visitors/tourism trips/tourists/same-day visitors (Department of Economic and Social Affairs, 2010, p. 10, para. 2.5, 2.6, 2.8, 2.10, 2.11, 2.32). A tourism trip may include one or multiple "tourism visits" (i.e., "stay[ing] in a place visited during a tourism trip") (Department of Economic and Social Affairs, 2010, p. 10, para. 2.33).



According to the IRTS, inbound, domestic, and outbound tourism can be combined to derive other forms of tourism (Department of Economic and Social Affairs, 2010, p. 10, para. 2.40):

Internal tourism:	Inbound and domestic tourism
National tourism:	Domestic and outbound tourism
International tourism:	Inbound and outbound tourism

## Methodology

For a common unit of measurement for online travel reviews and the UNWTO, we aggregated travel reviews temporally and geographically into a measure of travel reviews per year and country, which is compatible to the UNWTO statistics. A review always has a timestamp of the date it was written (henceforth called “date of review”). The actual travel date to which the review refers (henceforth called “date of travel”) is optional. We used only reviews that indicated a date of travel. Otherwise, the aggregation to years would not be reliable because the date of the actual trip would be unknown. Reviews were aggregated temporally according to the year given in the date of travel. Further, since the first and last years (1999 and 2010, respectively) were not covered completely, we limited our analysis to the 10-year period from 2000 to 2009.

Travel reviews always relate to a certain facility or accommodation (e.g., hotel, restaurant, museum) located at and associated with a certain destination. The VTC also maintains a geographical hierarchy of destinations (i.e., continents, countries, federal states, regions, cities), which we also collected. Thus, reviews could be aggregated easily to different geographical levels in the hierarchy. The resulting indicator gave the number of travel reviews for a country of reference in a specific year. We termed this indicator VTC.TravelReviews. It is temporally and geographically comparable to the UNWTO indicators, and we were able to construct a list of 210 countries that could be matched by their names between both datasets.

After restricting the set of reviews to those written about one of the 210 countries that could be matched and that indicated a date of travel within the specified 10-year period, there were still approximately 4.76 million reviews, or about 60.2 percent of all available reviews. This restricted dataset is the basis for VTC.TravelReviews.

As mentioned above, the IRTS distinguishes between domestic, inbound, and outbound tourism as well as internal, national, and international tourism. For the online travel reviews in our dataset, we were not able to distinguish whether a certain review referring to a certain country is domestic, inbound, or outbound tourism because the reviewer’s origin cannot be reliably inferred in every case. However, the destination and thus the country of reference of the tourism trip were given. Hence, we can assume that VTC.TravelReviews represents internal tourism—that is, tourism in a country of reference irrespective of whether visitors are residents or non-residents (Department of Economic and Social Affairs, 2010, p. 10, para. 2.40). We were able to assess internal tourism by totaling the respective indicators for inbound and domestic travel when both were provided. However, since the indicators thus constructed were not part of the UNWTO’s or IRTS’s conceptual framework, we termed them “overall tourism.”

Table 1 is an overview of the six UNWTO indicators, the three derived indicators for overall tourism, and the corresponding concepts from the IRTS’s framework.

--- Table 1 about here ---

To assess *completeness*, we compiled the following statistics measuring the coverage of countries for VTC data (date used for aggregation are in parentheses):

Countries (date):	Total number of countries for which online travel reviews could be found that indicate a date of travel (date of travel)
Countries (date, UNWTO):	Number of countries for which online travel reviews could be found that indicate a date of travel, and which could be matched to UNWTO countries (restricted dataset; date of travel)
New Reviews:	Total number of new reviews published on the VTC (date of review because not all reviews indicate a date of travel)
New Reviews (date):	Number of new reviews published on the VTC that indicate a date of travel (date of travel)
New Reviews (date, UNWTO):	Number of reviews published on the VTC that indicate a date of travel and could be matched to UNWTO countries (date of travel; used for VTC.TravelReviews)

We also calculated the number of countries covered by the selected UNWTO indicators.

To assess the *timeliness* of travel review data in the restricted dataset, we calculated the delay in months between the date of travel and the date of review. If a review was written in the same month of the travel, the delay is zero; in the month after the travel, the delay is one; and so forth.

For every year, we also calculated the following statistics from online travel reviews (based on the restricted dataset, aggregated according to date of travel):

Countries (date, UNWTO, 2 months): Number of countries (of those matched to UNWTO) that would be covered if only those travel reviews written with a delay no greater than two months were used.

Countries (date, UNWTO, 6 months): Number of countries (of those matched to UNWTO) that would be covered if only those travel reviews written with a delay no greater than six months were used

In addition to covering many countries in a timely manner, statistics from tourism-related social media data—in this case, online travel reviews—must also adequately reflect the distribution of travelers per year and country to be useful for compiling statistics on global tourism. Hence, we compared the measurements from VTC.TravelReviews to the UNWTO indicators.

The indicators best suited for comparison with VTC.TravelReviews are the derived indicators for overall/internal tourism (O.1 to O.3). Unfortunately, the UNWTO data on the indicators of domestic tourism were very sparse. Hence, we were able to calculate overall tourism for only a few countries. To assess whether online travel reviews are a good information source for global tourism (i.e., tourism for a large number of countries), we also compared the indicators of inbound tourism to VTC.TravelReviews. Nevertheless, one would expect VTC.TravelReviews to reflect better the indicators of overall tourism because they are conceptually more similar.

In the case of online reviews, one cannot be certain whether a complete lack of reviews for a country in a given year indicates a missing value (i.e., null) or no tourism in that country (i.e., 0). Since we wanted to assess how accurately tourism-related social media represent global tourism, we decided to treat no reviews as indicative of no tourism. The perspective taken here is that if one had to rely exclusively on online travel reviews to assess global tourism, how much would that person's perception differ from the UNWTO data (and, presumably, actual tourism)? Consequently,

VTC.TravelReviews can provide measures for each of the 210 countries and for every year—although it may often be zero.

To assess the *accuracy* of online travel review data on global tourism with data from the UNWTO, we calculated Pearson correlations for VTC.TravelReviews on the one hand and for each of the UNWTO indicators on the other hand. Specifically, for a list of countries and a certain year, we correlated the measures of VTC.TravelReviews to O.1, O.2, O.3, 1.1, 1.2, and 1.3, respectively. Missing values were either treated as 0 (VTC.TravelReviews), or respective countries were left out pairwise (UNWTO indicators). This was done for every year from 2000 to 2009.

Since measures of the UNWTO indicators as well as of VTC.TravelReviews differ very strongly in absolute numbers, we also calculated correlations using logarithmic scales for both datasets. Further, we checked scatter plots for every pair of UNWTO indicators (O.1 to O.3; 1.1 to 1.3) and VTC.TravelReviews.

## Comparability of Data Sources

Data from the UNWTO and travel reviews are compatible according to the geographical and temporal data structure. However, there are differences between the two data sources of which one must be aware when comparing them and that could account for differences in the measurements.

Although we use a large number of online travel reviews, the measurement of tourism by travel reviews risks evaluating the travel activities of a few travelers if several reviews were written by one traveler and hence did not reflect independent travel. Since every review indicates a pseudonymized user, one can determine how many users were involved in the creation of a certain set of reviews. Hence, it can be determined whether multiple reviews (e.g., reviews for a year or country) are independent—that is, created by different users.<sup>1</sup> To assess the independence of the travel activities measured by VTC.TravelReviews, we calculated the following statistics (all aggregated to years according to the date of travel of the respective reviews):

Users:	Number of users per year who wrote the reviews in the restricted dataset
Reviews per User (Avg.):	Average number of reviews in the restricted dataset written by one user per year
Reviews per User (Std. Dev.):	Standard deviation of the number of reviews in the restricted dataset written by one user per year

The unit of measurement for the UNWTO indicators are arrivals of non-resident visitors at national borders (inbound) and trips of resident visitors (domestic) in the country of reference. Thus, a visitor is counted each time he/she enters a foreign country or travels within his/her country of residence. As mentioned above, the IRTS recommends that indicators on tourism distinguish between tourism trips and tourism visits. According to the methodological notes, the UNWTO indicators used in this analysis should correspond to (inbound/domestic) tourism trips (UNWTO, April 2011, pp. 13–15). However, for the VTC, single travel reviews are measured that refer to certain facilities. A traveler might make multiple tourism visits during one tourism trip and review multiple facilities. A tourism trip to a foreign or home country may thus be reflected in no reviews, one review, or multiple reviews.

---

<sup>1</sup> Six dummy users (e.g., “a VTC member”) with an extraordinarily large number of reviews (up to several thousands in one year) had to be excluded from the calculation of the statistics of reviews per user. Their reviews (326,614 out of 4.76 million) were, however, included in the restricted dataset and were used for the further analysis.



Moreover, if people travel in groups (e.g., families), the UNWTO data would count each group member individually, whereas in many cases, only one group member would likely write a review. To assess the robustness of our results regarding accuracy—that is, whether these issues related to one trip/multiple-reviews and one review/multiple visitors affect the association of the UNWTO indicators and our social media measure—we also calculated measures of trips per year and country and visitors per year and country.

Regarding the first issue, we counted a travel review as a trip only if no other travel review from the same VTC member to the same country and within the same month of the year has yet been counted, and derived trips from travel reviews accordingly. For example, if VTC member M wrote two travel reviews regarding facilities in country C with April 2007 as the date of travel, this will be counted as one trip of M to C in April 2007. If one review had not included country C but D or if one review indicated May 2007 as the date of travel, it would be counted as two trips, respectively. Trips thus derived were then aggregated by year and country analogous to the travel review measure VTC.TravelReviews. Surely, the resulting trips do not match perfectly with visitors' actual trips, but they are arguably a good approximation in the absence of a more sophisticated approach.

With respect to the problem that one review might represent multiple visitors at a time, we utilized information from the travel reviews regarding the travel group. Most of the reviews (4.04 million, or 84.9%) in our restricted dataset indicated one out of 11 possible answers to the (closed) question “traveled with” (e.g., “solo traveler,” “with spouse,” or “family with teenagers”). We assigned each of these answers a specific travel group size, e.g. 1, 2, or 4, respectively. Similar to trips and VTC.TravelReviews, travel group sizes were then aggregated by year and country. Hence, we could also calculate an approximate measure of visitors per year and country.

Another issue is that the global proportionality of data between countries may be affected by an unequal global distribution of the VTC's user group. For instance, if the VTC enjoyed more frequent use in some countries and less in others, the former countries' tourism preferences would presumably be systematically overrepresented (due to more reviews from these countries) compared to the UNWTO data. What applies to different usage levels of the VTC in different countries might also apply to different levels of Internet infrastructure availability. Further studies should evaluate approaches to account for these issues, for example, by combining data from multiple VTCs.

## Results

### Completeness

Table 2 and Figures 1 and 2 are a detailed overview of the geographic coverage of the UNWTO indicators as well as the statistics calculated from the travel review data.

--- Table 2 about here ---

--- Figures 1 and 2 about here ---

Data were much more complete for inbound tourism in the UNWTO statistics than for domestic tourism and, consequently, also more complete than for overall tourism. The number of countries for which online travel reviews can provide data more than doubled within the 10-year period. This development is presumably due to the enormous increase in the number of reviews written each year. From 2005 to 2009, online travel reviews covered even more countries than the selected UNWTO indicators.

### Timeliness

As Figures 3 and 4 show, more than 85 percent of the reviews were written no more than two months after the travel took place. Some 95 percent of the reviews refer to tourism activities no more than 12 months ago. The annual data published by UNWTO are typically about 18 months old. For example, data covering the year 2010 were first published in the Compendium in June 2012 (UNWTO, 2012b), and data for 2011 are in the 2013 Compendium (UNWTO, 2013).

--- Figures 3 and 4 about here ---

Table 2 also shows the coverage of countries when using only reviews published with a delay of six or two months (“Countries (date, UNWTO, 2 months)” and “Countries (date, UNWTO, 2 months)”). For 2006, for example, if one used only travel reviews written no more than six months after the travel to calculate an indicator like VTC.TravelReviews, one would still have travel review data on 197 countries (97% of the 203 countries for which data were available when not limiting delay, i.e. indicator “Countries (date, UNWTO)”). For a lag of no more than two months, VTC.TravelReviews would still cover 196 countries (96.6%). Over the 10-year period, country coverage for delays of six and two months continually increased. From 2005 on, one would miss only a few countries if one collected travel review data six months after a given period (e.g., a year). From 2006 on, country coverage for VTC.TravelReviews (for a delay of either six or even two months) is better than for any of the UNWTO indicators.

## Accuracy

Table 3 and Figures 5 and 6 show the results of correlating the UNWTO indicators on global tourism with VTC.TravelReviews. Values for the derived indicators on overall tourism can be reasonably evaluated only from 2006 on because UNWTO data from before that date are too sparse. The coefficients for the derived indicators (0.1 to 0.3) from 2006 to 2009 indicate that for the relatively small group of countries that report numbers on inbound and domestic tourism, these measures are highly and significantly correlated with the number of online travel reviews regarding these countries. This is confirmed by the scatter plots for VTC.TravelReviews and the UNWTO derived indicators (e.g., see Figure 7).

--- Table 3 about here ---

--- Figures 5, 6, and 7 about here ---

Correlations for the overall tourism indicators could also be higher because the countries in the smaller group are more similar and not because the derived measures are conceptually better suited. However, that is unlikely since, for example, the group of countries with the highest correlation for a derived measure in 2009 (0.3) can be regarded as quite diverse, comprising Canada, France, Kazakhstan, Latvia, Lithuania, Malaysia, New Zealand, Saudi Arabia, Spain, and Venezuela.

The drop of the correlation for 0.1 in 2009 can be explained by the fact that it was the first year China reported total domestic tourism (2.1) to the UNWTO. That value was much larger than the number of online travel reviews for China in 2009 compared to the relationship of the UNWTO data and travel review data for other countries (see also Figure 7). This indicates that (for 2009 but presumably also before) Chinese tourists wrote far fewer travel reviews when traveling in their home country than did people in other countries—at least to the VTC used in this analysis.

The results for the measures of inbound tourism for a much larger group of countries indicate that even if both data sources differ with respect to whether they include domestic tourism, there are strong and significant correlations between the UNWTO indicators and online travel reviews. Despite occasional drops in some years, all correlations of online travel reviews to the UNWTO indicators by and large increased over time.

With respect to the independence of reviews in the restricted dataset, statistics on reviews per user (see Table 2) indicate that one user contributed between 1 and 1.5 reviews on average (standard deviation: 0.5 to 1.5). This means that the creation of reviews is fairly distributed among users/travelers.

## **Robustness**

To assess the robustness of the results regarding accuracy, we repeated the analysis with measures of the number of trips per year and country as well as the number of visitors per year and country derived from online travel reviews as mentioned above. The results were very similar to those from online travel reviews: from 2006 on, correlation coefficients for trips and visitors to the UNWTO indicators, respectively, differed not more than .03 (in absolute) from those obtained from the analysis of travel reviews. The significance levels of the coefficients were identical to those from travel reviews.

## **Discussion**

Our findings indicate that tourism-related social media can be used as a source of somewhat representative macro-level tourism information. For those countries that provided appropriate data, we found very strong correlations between the overall number of visitors (both inbound and domestic) per year from the UNWTO and the number of online travel reviews. We also saw that measures on global tourism in social media and indicators by the UNWTO have a weaker (but still strong) correlation when we compared reviews to indicators of inbound tourism alone. Put differently, for those countries that do not provide data on domestic tourism, online travel reviews likely provide an accurate assessment of overall tourism (or internal tourism, according to IRTS).

With respect to timeliness and completeness, online travel reviews can report on global tourism much sooner than traditional sources, such as statistics from the UNWTO, and they cover more countries. From 2006 on, online travel reviews covered more countries than the UNWTO even if they had been collected only six months after the end of a year. This offers the opportunity to substitute the UNWTO data with statistics on global tourism derived from travel reviews in cases of missing data or until the UNWTO data are published. Overall, the information quality of online travel review data is equal to or even higher than data from the UNWTO. We conclude that tourism-related social media are a good information sources for global tourism distribution.

These findings contribute to the research on tourism flows and ways to measure them (e.g., using data from new technologies), which are topics of ongoing research at the UNWTO and the associated International Network on Regional Economics, Mobility and Tourism (INRouTe). Moreover, this study contributes to the research on social media in tourism, travel, and hospitality, specifically the spatio-temporal analysis of content from social media. Our study is novel in that it evaluates the information quality of UGC with respect to specific criteria, while most content studies on tourism-related social media have thus far focused on analyzing and drawing conclusions from UGC. In other words, we assessed the quality of information that has been used in other analyses that have generally not explicitly questioned that quality.

## **Limitations and Outlook**

We assessed information quality in terms of completeness, timeliness, and accuracy with respect to a specific application case (macro-level spatio-temporal tourism distribution). Further studies should likewise investigate information quality for other application cases and the quality criteria relevant to them. Furthermore, given the positive results for information quality, further research should develop approaches to estimate global tourism figures quantitatively, for example from online travel reviews. These figures can be used when information is missing from countries or when information has not yet been published. They may also be used to predict future levels of tourism.

To estimate or predict tourism based on longitudinal analysis, however, one would have to overcome some methodological problems. First, longitudinal analyses of the UNWTO and travel review data for single countries can make use of only a few measurement points for one country—at most only 10 in our case (from 2000 to 2009). For domestic and internal tourism from the UNWTO data, this number drops to zero for most of the countries. Second, the measurements for global tourism in both datasets (i.e., the UNWTO indicators and travel reviews) are influenced by multiple time-dependent factors: (1) the general development of global tourism over time (i.e., are there more or fewer visitors, trips, travel reviews, and so on for all countries in one year than in the previous or following years?); (2) the country-specific development of tourism over time (i.e., are there more or fewer visitors, trips, travel reviews, etc., for one country in relation to other countries and in a certain year than in the previous or following years?); and (3) for travel reviews, the development of the VTC's user base over time (i.e., more users will very likely write more reviews independent of the first two factors). These (and probably other) factors would have to be accounted for in a longitudinal analysis.

The comparison of the better-suited derived indicators O.1 to O.3 could be carried out only for a comparatively small set of countries because of limited data from the UNWTO. Despite this limitation, the results were very promising. Using travel reviews from a single VTC creates the risk of systematically underestimating travel in and from some countries in which other VTCs are used more often. Both problems of data coverage could be overcome by adding other sources of data on global tourism and by combining data from multiple travel review websites. This could, though, increase the problems that arise from differences among multiple data sources, which we discussed above. Techniques to account for these differences and assess information quality in terms of consistency should be evaluated.

Our robustness checks for differences when counting visitors or trips instead of travel reviews did not reveal any significant differences but replicated the findings from using a number of travel reviews. However, more sophisticated approaches may be developed to identify travel reviews that refer to the same trip as well as draw content conclusions from the review on the size of the travel group.

When investigating the association of travel reviews and the UNWTO data, one may also ask whether they are correlated because the former reflects the latter or because the former has influenced the latter. Both may be true. However, the effects of valence and online travel review ratings on travel decisions and thus on global tourism (Vermeulen & Seegers, 2009; Ye, Law, & Gu, 2009) are presumably more important than the number of reviews. Moreover, tourism is much older than social media, and people still rely on other online and offline information sources to help them make travel decisions (Fotis et al., 2011; Xiang & Gretzel, 2010; Xiang et al., 2014; Yoo & Gretzel, 2012).

Fake content is, of course, a great challenge to most studies dealing with UGC. Fake reviews may be written by facility owners to promote their own business or to weaken competitors or they may be written by paid third parties. Travelers may also create fake reviews to express exaggerated satisfaction or dissatisfaction with a facility or for other reasons. Travel review websites seek to detect and remove fake reviews but cannot prevent them completely (BBC News, 2013; Fox, 2010). Existing approaches should be evaluated to detect and sort out fake reviews in the context of this study.

## References

- Allsop, D. T., Bassett, B. R., & Hoskins, J. A. (2007). Word-of-Mouth Research: Principles and Applications. *Journal of Advertising Research*, 47(4), 398–411. doi:10.2501/S0021849907070419
- Ardizzone, E., Di Miceli, F., La Cascia, M., & Mazzola, G. (2012). Extracting Touristic Information from Online Image Collections. In *Eighth International Conference on Signal-Image Technology & Internet-Based Systems (SITIS 2012)* (pp. 482–488).
- BBC News. (2013). *TripAdvisor removes fake Brixham restaurant Oscar's*. Retrieved from <http://www.bbc.co.uk/news/uk-england-devon-23504081>
- Bieger, T., & Laesser, C. (2004). Information Sources for Travel Decisions: Toward a Source Process Model. *Journal of Travel Research*, 42(4), 357–371. doi:10.1177/0047287504263030
- Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet - The state of eTourism research. *Tourism Management*, 29(4), 609–623. doi:10.1016/j.tourman.2008.01.005
- Bulchand-Gidumal, J., Melián-González, S., & González López-Valcárcel, B. (2013). A social media analysis of the contribution of destinations to client satisfaction with hotels. *International Journal of Hospitality Management*, 35(0), 44–47. doi:10.1016/j.ijhm.2013.05.003
- Chan, N. L., & Guillet, B. D. (2011). Investigation of Social Media Marketing: How Does the Hotel Industry in Hong Kong Perform in Marketing on Social Media Websites? *Journal of Travel & Tourism Marketing*, 28(4), 345–368. doi:10.1080/10548408.2011.571571
- Chen, Y., & Xie, J. (2008). Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. *Management Science*, 54(3), 477–491. doi:10.1287/mnsc.1070.0810
- Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing and Management*, 18(8), 743–764. doi:10.1080/19368620903235753
- Davidson, R., & Keup, M. (2014). The Use of Web 2.0 as a Marketing Tool by European Convention Bureaux. *Scandinavian Journal of Hospitality and Tourism*, 14(3), 234–254. doi:10.1080/15022250.2014.946232
- Davis, A., & Khazanchi, D. (2008). An Empirical Study of Online Word of Mouth as a Predictor for Multi-product Category e-Commerce Sales. *Electronic Markets*, 18(2), 130–141. doi:10.1080/10196780802044776
- Department of Economic and Social Affairs, S. D. (2010). *International Recommendations for Tourism Statistics 2008. Studies in methods. Series M: no. 83/Rev.1*. New York: United Nations. Retrieved from [http://unstats.un.org/unsd/tradeserv/tourism/08-40120%20IRTS%202008\\_WEB\\_final%20version%20\\_22%20February%202010.pdf](http://unstats.un.org/unsd/tradeserv/tourism/08-40120%20IRTS%202008_WEB_final%20version%20_22%20February%202010.pdf)
- Dickinger, A. (2011). The Trustworthiness of Online Channels for Experience- and Goal-Directed Search Tasks. *Journal of Travel Research*, 50(4), 378–391. doi:10.1177/0047287510371694
- Dippelreiter, B., Grün, C., Pöttler, M., Seidel, I., Berger, H., Dittenbach, M., & Pesenhofer, A. (2008). Online Tourism Communities on the Path to Web 2.0: An Evaluation. *Information Technology & Tourism*, 10(4), 329–353. doi:10.3727/109830508788403132



Fodness, D., & Murray, B. (1997). Tourist information search. *Annals of Tourism Research*, 24(3), 503–523. doi:10.1016/S0160-7383(97)00009-1

Fotis, J., Buhalis, D., & Rossides, N. (2011). Social Media Impact on Holiday Travel Planning: The Case of the Russian and the FSU Markets. *International Journal of Online Marketing*, 1(4), 1–19. doi:10.4018/ijom.2011100101

Fox, L. (2010). *Exclusive: 'Fake review' firms are not an issue, says TripAdvisor*. Retrieved from <http://www.travolution.co.uk/articles/2010/01/22/3184/exclusive-fake-review-%EF%AC%81rms-are-not-an-issue-says-tripadvisor.html>

Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive Computing*, 7(4), 36–43. doi:10.1109/MPRV.2008.71

Girardin, F., Dal Fiore, F., Blat, J., & Ratti, C. (2007). Understanding of tourist dynamics from explicitly disclosed location information. In *International Symposium on LBS & TeleCartography*.

Gu, B., & Ye, Q. (2014). First Step in Social Media: Measuring the Influence of Online Management Responses on Customer Satisfaction. *Production & Operations Management*, 23(4), 570–582. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=95576045&site=ehost-live>

Gupta, S., & Ramachandran, J. (2013). The Park Hotels: Designing Communications. *Asian Case Research Journal*, 17(2), 199–224. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=94659232&site=ehost-live>

Huang, L., Yung, C.-Y., & Yang, E. (2011). How do travel agencies obtain a competitive advantage?: Through a travel blog marketing channel. *Journal of Vacation Marketing*, 17(2), 139–149. doi:10.1177/1356766710392737

Huang, L. (2012). Social Media as a New Play in a Marketing Channel Strategy: Evidence from Taiwan Travel Agencies' Blogs. *Asia Pacific Journal of Tourism Research*, 17(6), 615–634. doi:10.1080/10941665.2011.635664

Inversini, A., Marchiori, E., Dedekind, C., & Cantoni, L. (2010). Applying a Conceptual Framework to Analyze Online Reputation of Tourism Destinations. In U. Gretzel, R. Law, & M. Fuchs (Eds.), *Information and Communication Technologies in Tourism 2010* (pp. 321–332). Vienna: Springer Vienna.

Inversini, A., & Masiero, L. (2014). Selling rooms online: the use of social media and online travel agents. *International Journal of Contemporary Hospitality Management*, 26(2), 272–292. doi:10.1108/IJCHM-03-2013-0140

Jeng, J., & Fesenmaier, D. R. (2002). Conceptualizing the Travel Decision-Making Hierarchy: A Review of Recent Developments. *Tourism Analysis*, 7(1), 15–32.

Kang, M., & Schuett, M. A. (2013). Determinants of Sharing Travel Experiences in Social Media. *Journal of Travel & Tourism Marketing*, 30(1-2), 93–107. doi:10.1080/10548408.2013.751237

Kasavana, M. L., Nusair, K., & Teodosic, K. (2010). Online social networking: redefining the human web. *Journal of Hospitality and Tourism Technology*, 1(1), 68–82. doi:10.1108/17579881011023025

- Kim, J., & Hardin, A. (2010). The Impact of Virtual Worlds on Word-of-Mouth: Improving Social Networking and Servicescape in the Hospitality Industry. *Journal of Hospitality Marketing & Management*, 19(7), 735–753. doi:10.1080/19368623.2010.508005
- Knight, S.-A., & Burn, J. (2005). Developing a Framework for Assessing Information Quality on the World Wide Web. *Informing Science Journal*, 8, 159–172. Retrieved from <http://www.inform.nu/Articles/Vol8/v8p159-172Knig.pdf>
- Lagrosen, S. O., & Grundén, K. (2014). Social media marketing in the wellness industry. *TQM Journal*, 26(3), 253–260. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=95390665&site=ehost-live>
- Lanz, L., Fischhof, B., & Lee, R. (2010). How Are Hotels Embracing Social Media in 2010: Examples of How to Begin Engaging. *HVS International Journal*, 1–28. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=56567871&site=ehost-live>
- Law, R., & Cheung, S. (2010). The Perceived Destination Image of Hong Kong as Revealed in the Travel Blogs of Mainland Chinese Tourists. *International Journal of Hospitality & Tourism Administration*, 11(4), 303–327. doi:10.1080/15256480.2010.518521
- Lee, H. “., Law, R., & Murphy, J. (2011). Helpful Reviewers in TripAdvisor, an Online Travel Community. *Journal of Travel & Tourism Marketing*, 28(7), 675–688. doi:10.1080/10548408.2011.611739
- Leung, D., Law, R., van Hoof, H., & Buhalis, D. (2013). Social Media in Tourism and Hospitality: A Literature Review. *Journal of Travel & Tourism Marketing*, 30(1-2), 3–22. doi:10.1080/10548408.2013.750919
- Leung, X. Y., & Bai, B. (2013). How Motivation, Opportunity, and Ability Impact Travelers' Social Media Involvement and Revisit Intention. *Journal of Travel & Tourism Marketing*, 30(1-2), 58–77. doi:10.1080/10548408.2013.751211
- Lewis, R. C., & Chambers, R. E. (2000). *Marketing leadership in hospitality: Foundations and practices* (3rd ed.). New York: John Wiley & Sons.
- Li, H., Ye, Q., & Law, R. (2013). Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. *Asia Pacific Journal of Tourism Research*, 18(7), 784–802. doi:10.1080/10941665.2012.708351
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468. doi:10.1016/j.tourman.2007.05.011
- Lu, W., & Stepchenkova, S. (2014). User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. *Journal of Hospitality Marketing & Management*. doi:10.1080/19368623.2014.907758
- Lu, X., Wang, C., Yang, J.-M., Pang, Y., & Zhang, L. (2010). Photo2Trip: Generating Travel Routes from Geo-Tagged Photos for Trip Planning. In A. del Bimbo, S.-F. Chang, & A. Smeulders (Eds.), *Proceedings of the international conference on Multimedia* (p. 143).
- Mills, J. E., & Law, R. (2004). *Handbook of consumer behavior, tourism, and the Internet*. New York: Haworth Hospitality Press.

- Munar, A. M. (2011). Tourist-created content: rethinking destination branding. *International Journal of Culture, Tourism and Hospitality Research*, 5(3), 291–305. doi:10.1108/17506181111156989
- Munar, A. M. (2012). Social Media Strategies and Destination Management. *Scandinavian Journal of Hospitality and Tourism*, 12(2), 101–120. doi:10.1080/15022250.2012.679047
- Munar, A. M., & Jacobsen, J. K. S. (2013). Trust and Involvement in Tourism Social Media and Web-Based Travel Information Sources. *Scandinavian Journal of Hospitality and Tourism*, 13(1), 1–19. doi:10.1080/15022250.2013.764511
- Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism Management*, 43(0), 46–54. doi:10.1016/j.tourman.2014.01.012
- Noone, B. M., McGuire, K. A., & Niemeier, H.-M. (2011). Social media meets hotel revenue management: Opportunities, issues and unanswered questions. *Journal of Revenue and Pricing Management*, 10(4), 293–305. doi:10.1057/rpm.2011.12
- Ong, B. S. (2012). The Perceived Influence of User Reviews in the Hospitality Industry. *Journal of Hospitality Marketing & Management*, 21(5), 463–485. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=76607111&site=ehost-live>
- Pantelidis, I. S. (2010). Electronic Meal Experience: A Content Analysis of Online Restaurant Comments. *Cornell Hospitality Quarterly*, 51(4), 483–491. doi:10.1177/1938965510378574
- Park, S.-Y., & Allen, J. P. (2013). Responding to Online Reviews: Problem Solving and Engagement in Hotels. *Cornell Hospitality Quarterly*, 54(1), 64–73. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=84424022&site=ehost-live>
- Popescu, A., Grefenstette, G., & Moëllic, P.-A. (2009). Mining tourist information from user-supplied collections. In D. Cheung, I.-Y. Song, W. Chu, X. Hu, & J. Lin (Eds.), *Proceeding of the 18th ACM conference* (p. 1713).
- Rosman, R., & Stuhura, K. (2013). The Implications of Social Media on Customer Relationship Management and the Hospitality Industry. *Journal of Management Policy & Practice*, 14(3), 18–26. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=94484542&site=ehost-live>
- Schmallegger, D., & Carson, D. (2008). Blogs in tourism: Changing approaches to information exchange. *Journal of Vacation Marketing*, 14(2), 99–110. doi:10.1177/1356766707087519
- Schmallegger, D., & Carson, D. (2009). Destination Image Projection on Consumer-Generated Content Websites: A Case Study of the Flinders Ranges. *Information Technology & Tourism*, 11(2), 111–127. doi:10.3727/109830509789994838
- Scott, S. V., & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, Organizations and Society*, 37(1), 26–40. doi:10.1016/j.aos.2011.11.005
- Scott, S. V., & Orlikowski, W. J. (2014). Entanglements in Practice: Performing Anonymity Through Social Media. *MIS Quarterly*, 38(3), 873–893. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=97267800&site=ehost-live>
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103–110. doi:10.1145/253769.253804

- Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2013). Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems*, (0). doi:10.1016/j.compenvurbsys.2013.07.006
- Tham, A., Croy, G., & Mair, J. (2013). Social Media in Destination Choice: Distinctive Electronic Word-of-Mouth Dimensions. *Journal of Travel & Tourism Marketing*, 30(1-2), 144–155. doi:10.1080/10548408.2013.751272
- Tse, T. S. M. (2013). The Marketing Role of the Internet in Launching a Hotel: The Case of Hotel ICON. *Journal of Hospitality Marketing & Management*, 22(8), 895–908. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=91722151&site=ehost-live>
- United Nations World Tourism Organization. (2011). *Methodological Notes on UNWTO Statistics*.
- United Nations World Tourism Organization. (2012a). *Compendium of Tourism Statistics* (2012 ed., 2nd printing.). Madrid: World Tourism Organization.
- United Nations World Tourism Organization. (2012b). *Compendium of Tourism Statistics, 2012 Edition: June 2012*. Retrieved from <http://statistics.unwto.org/en/news/2012-06-11/compendium-tourism-statistics-2012-edition>
- United Nations World Tourism Organization. (2013). *Compendium of Tourism Statistics, Data 2007 – 2011, 2013 Edition: 2013*. Retrieved from <http://publications.unwto.org/en/publication/compendium-tourism-statistics-data-2007-2011-2013-edition>
- United Nations World Tourism Organization. (2014). *Compendium of Tourism Statistics*. Retrieved from <http://statistics.unwto.org/en/content/compendium-tourism-statistics>
- Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123–127. doi:10.1016/j.tourman.2008.04.008
- Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33. doi:10.2307/40398176
- Wang, Y., Yu, Q., & Fesenmaier, D. R. (2002). Defining the virtual tourist community: implications for tourism marketing. *Tourism Management*, 23(4), 407–417. doi:10.1016/S0261-5177(01)00093-0
- Wöber, K. W. (2003). Information supply in tourism management by marketing decision support systems. *Tourism Management*, 24(3), 241–255. doi:10.1016/S0261-5177(02)00071-7
- Xiang, Z., Wang, D., O'Leary, J. T., & Fesenmaier, D. R. (2014). Adapting to the Internet: Trends in Travelers' Use of the Web for Trip Planning. *Journal of Travel Research*. doi:10.1177/0047287514522883
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179–188. doi:10.1016/j.tourman.2009.02.016
- Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180–182. doi:10.1016/j.ijhm.2008.06.011
- Yoo, K.-H., & Gretzel, U. (2008). What Motivates Consumers to Write Online Travel Reviews? *Information Technology & Tourism*, 10(4), 283–295. doi:10.3727/109830508788403114

Yoo, K.-H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation: Web 2.0 in Travel and Tourism: Empowering and Changing the Role of Travelers. *Computers in Human Behavior*, 27(2), 609–621. doi:10.1016/j.chb.2010.05.002

Yoo, K.-H., & Gretzel, U. (2012). Use and Creation of Social Media by Travellers. In M. Sigala, E. Christou, & U. Gretzel (Eds.), *New directions in tourism analysis. Social media in travel, tourism and hospitality. Theory, practice and cases* (pp. 189–205). Farnham, Surrey, Burlington, VT: Ashgate Pub.

Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10(0), 27–36. doi:10.1016/j.tmp.2014.01.001

Zheng, Y.-T., Zha, Z.-J., & Chua, T.-S. (2012). Mining Travel Patterns from Geotagged Photos. *ACM Transactions on Intelligent Systems and Technology*, 3(3), 1–18. doi:10.1145/2168752.2168770



*Table 1 Original and derived indicators of UNWTO with corresponding concepts from IRTS (only inbound and domestic)*

Length of Stay		Country of reference in relation to visitors' country of residence				UNWTO: overall tourism IRTS: internal tourism	
		Inbound tourism		Domestic tourism			
UNWTO	IRTS	UNWTO	IRTS	UNWTO	IRTS	UNWTO	IRTS
Total	visitors taking tourism trips	1.1 Arrivals – Total visitors	inbound visitors taking inbound tourism trips	2.1 Trips – Total visitors	domestic visitors taking domestic tourism trips	O.1 Overall – Total visitors (1.1 + 2.1)	inbound visitors taking inbound tourism trips and domestic visitors taking domestic tourism trips
Overnight visitors	Tourists taking tourism trips	1.2 Arrivals – Overnight visitors	inbound tourists taking inbound tourism trips	2.2 Trips – Overnight visitors	domestic tourists taking domestic tourism trips	O.2 Overall – Overnight visitors (1.2 + 2.2)	inbound tourists taking inbound tourism trips and domestic tourists taking domestic tourism trips
Same-day visitors	Same-day visitors taking tourism trips	1.3 Arrivals – Same-day visitors	inbound same-day visitors taking inbound tourism trips	2.3 Trips – Same-day visitors	domestic same-day visitors taking domestic tourism trips	O.3 Overall – Same-day visitors (1.3 + 2.3)	inbound same-day visitors taking inbound tourism trips and domestic same-day visitors taking domestic tourism trips

Table 2 Data coverage for countries in UNWTO data set as well as statistics on data coverage, timeliness, and independence of online travel reviews

Indicator	2000	2001	2002	2003	2004
<b>COUNTRIES COVERED BY UNWTO INDICATORS</b>					
Inbound tourism					
1.1 Arrivals – Total visitors	124	127	120	123	122
1.2 Arrivals – Overnight visitors	189	190	184	184	185
1.3 Arrivals – Same-day visitors	103	107	103	104	106
Domestic tourism					
2.1 Trips – Total visitors	1	1	1	3	4
2.2 Trips – Overnight visitors	3	3	4	7	9
2.3 Trips – Same-day visitors	1	1	1	3	4
Overall tourism (derived)					
O.1 Overall – Total visitors	1	1	1	2	3
O.2 Overall – Overnight visitors	3	3	4	6	8
O.3 Overall – Same-day visitors	1	1	1	2	3
<b>ONLINE TRAVEL REVIEWS</b>					
Countries (date)	83	91	123	166	182
Countries (date, UNWTO)	82	91	120	160	175
Countries (date, UNWTO, 2 months)	0	0	30	53	141
Countries (date, UNWTO, 6 months)	0	0	28	47	127
New Reviews	59	1,446	14,965	107,658	253,754
New Reviews (date)	788	1,186	3,641	16,384	63,752
New Reviews (date, UNWTO)	786	1,186	3,636	16,365	63,711
Users	337	536	1,860	8,910	28,362
Reviews per User (Avg.)	1.110	1.099	1.120	1.202	1.286
Reviews per User (Std. Dev.)	.425	.469	.560	.789	.803
Indicator	2005	2006	2007	2008	2009
<b>COUNTRIES COVERED BY UNWTO INDICATORS</b>					
Inbound tourism					
1.1 Arrivals – Total visitors	124	127	127	126	121
1.2 Arrivals – Overnight visitors	189	188	189	182	178
1.3 Arrivals – Same-day visitors	110	114	114	111	108
Domestic tourism					
2.1 Trips – Total visitors	5	13	15	18	22
2.2 Trips – Overnight visitors	12	19	21	27	30
2.3 Trips – Same-day visitors	3	9	11	14	14
Overall tourism (derived)					
O.1 Overall – Total visitors	4	9	13	15	18
O.2 Overall – Overnight visitors	11	17	19	24	27
O.3 Overall – Same-day visitors	2	6	10	10	10
<b>ONLINE TRAVEL REVIEWS</b>					
Countries (date)	205	218	223	226	228
Countries (date, UNWTO)	194	203	208	208	209
Countries (date, UNWTO, 2 months)	181	197	206	208	209
Countries (date, UNWTO, 6 months)	178	196	204	207	209
New Reviews	424,549	643,898	1,148,968	1,985,314	2,774,361
New Reviews (date)	313,925	537,505	876,824	1,170,243	1,775,431
New Reviews (date, UNWTO)	313,781	537,175	876,297	1,169,610	1,774,451
Users	114,635	304,907	588,627	796,764	1,220,656
Reviews per User (Avg.)	1.508	1.471	1.473	1.443	1.427
Reviews per User (Std. Dev.)	1.299	1.286	1.385	1.389	1.344

Table 3 Correlations between UNWTO indicators and VTC.TravelReviews (online travel reviews); \* significant at .05 (two-sided), \*\* significant at .01 (two-sided)

UNWTO indicator	2000	2001	2002	2003	2004
<b>Inbound tourism</b>					
1.1 Arrivals – Total visitors	.481**	.522**	.486**	.473**	.392**
log <sub>10</sub>	.510**	.534**	.579**	.580**	.608**
n	124	127	120	123	122
1.2 Arrivals – Overnight visitors	.631**	.645**	.594**	.552**	.538**
log <sub>10</sub>	.613**	.630**	.692**	.717**	.757**
n	189	190	184	184	185
1.3 Arrivals – Same-day visitors	.337**	.413**	.395**	.392**	.299**
log <sub>10</sub>	.415**	.419**	.408**	.372**	.434**
n	103	107	103	104	106
<b>Overall tourism (derived)</b>					
O.1 Overall – Total visitors	n/a	n/a	n/a	1.000	.991
log <sub>10</sub>	n/a	n/a	n/a	1.000	.498
n	1	1	1	2	3
O.2 Overall – Overnight visitors	.997*	.997	.637	.982**	.917**
log <sub>10</sub>	.873	.929	.845	.670	.468
n	3	3	4	6	8
O.3 Overall – Same-day visitors	n/a	n/a	n/a	1.000	.998*
log <sub>10</sub>	n/a	n/a	n/a	1.000	.806
n	1	1	1	2	3
UNWTO indicator	2005	2006	2007	2008	2009
<b>Inbound tourism</b>					
1.1 Arrivals – Total visitors	.386**	.663**	.646**	.664**	.649**
log <sub>10</sub>	.605**	.676**	.653**	.640**	.649**
n	124	127	127	126	121
1.2 Arrivals – Overnight visitors	.531**	.577**	.597**	.624**	.648**
log <sub>10</sub>	.764**	.804**	.800**	.803**	.811**
n	189	188	189	182	178
1.3 Arrivals – Same-day visitors	.288**	.734**	.709**	.704**	.658**
log <sub>10</sub>	.447**	.568**	.547**	.531**	.513**
n	110	114	114	111	108
<b>Overall tourism (derived)</b>					
O.1 Overall – Total visitors	-.748	.951**	.951**	.973**	.437
log <sub>10</sub>	-.991**	.643	.766**	.843**	.819**
n	4	9	13	15	18
O.2 Overall – Overnight visitors	.891**	.861**	.994**	.993**	.940**
log <sub>10</sub>	.726*	.735**	.830**	.873**	.859**
n	11	17	19	24	27
O.3 Overall – Same-day visitors	1.000	.996**	.986**	.998**	.983**
log <sub>10</sub>	1.000	.888*	.898**	.933**	.895**
n	2	6	10	10	10

Figure 1 Number of countries covered by UNWTO indicators and online travel reviews (VTC.TravelReviews)

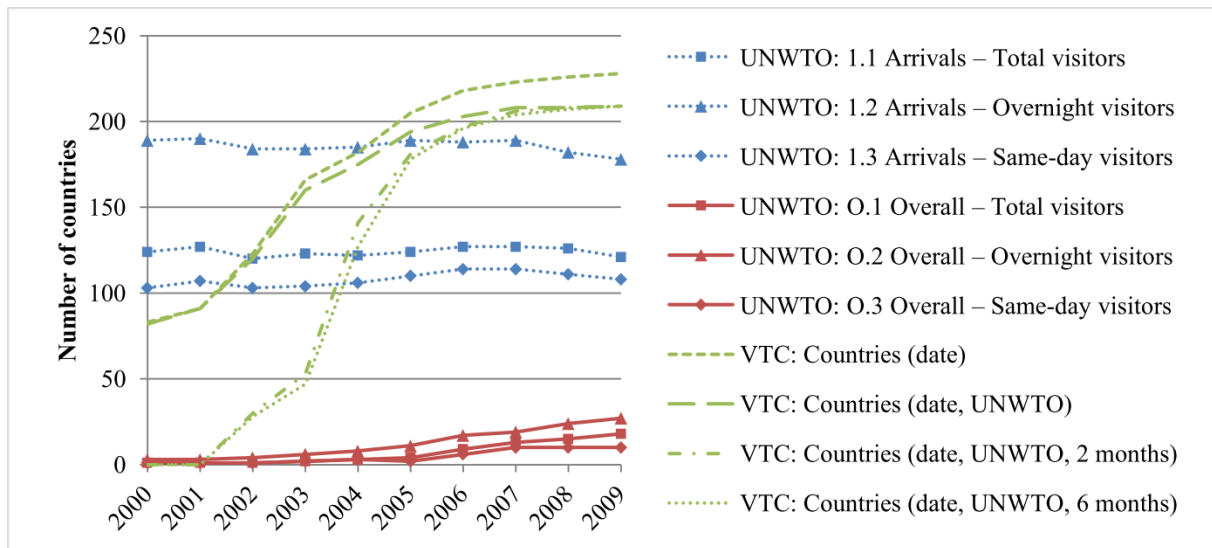


Figure 2 Number of new online travel reviews per year

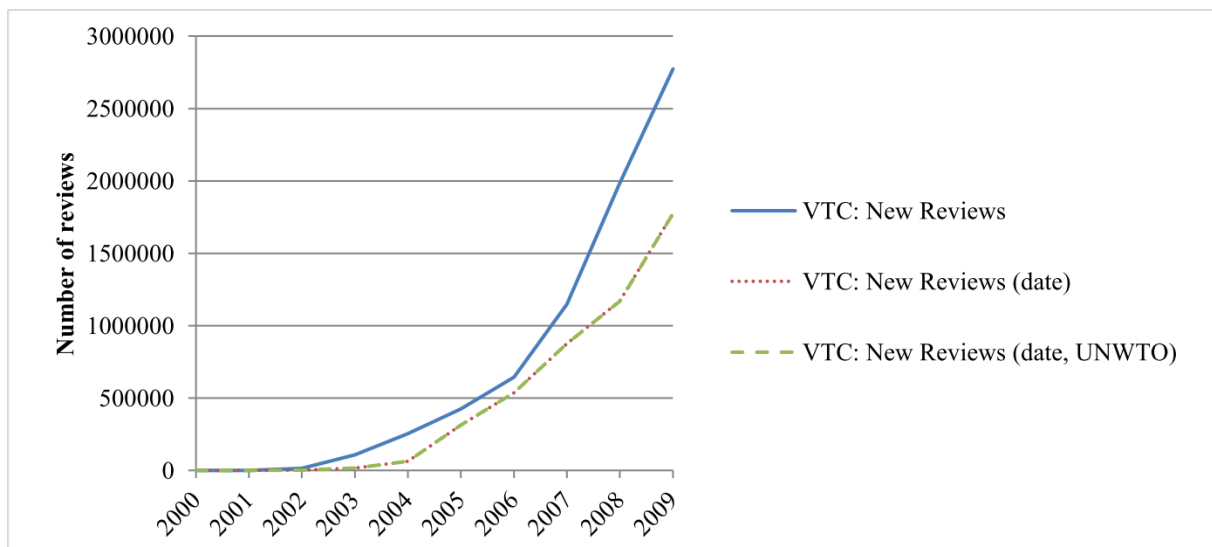


Figure 3 Number of reviews written with a specific delay, i.e., time between date of travel and date of review in months

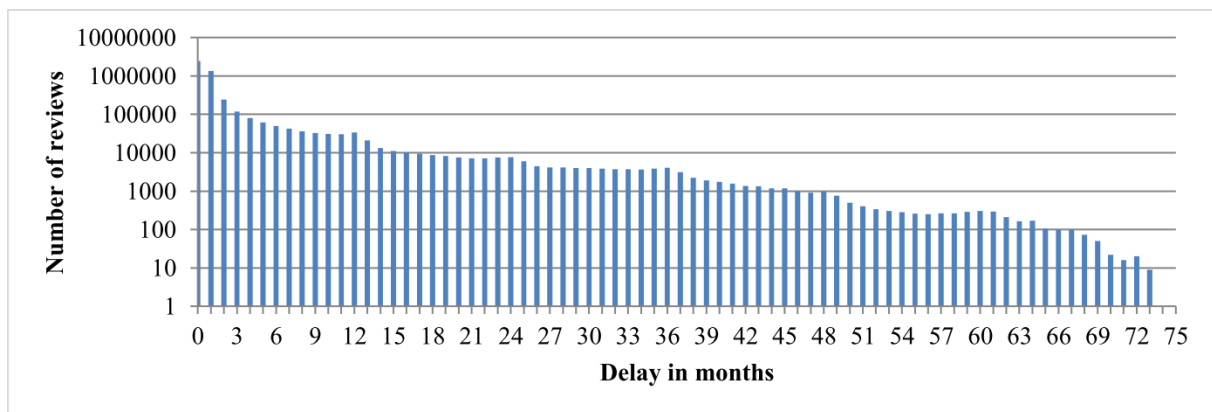


Figure 4 Percentage of reviews (cumulative) written with a specific delay, i.e., time between date of travel and date of review in months

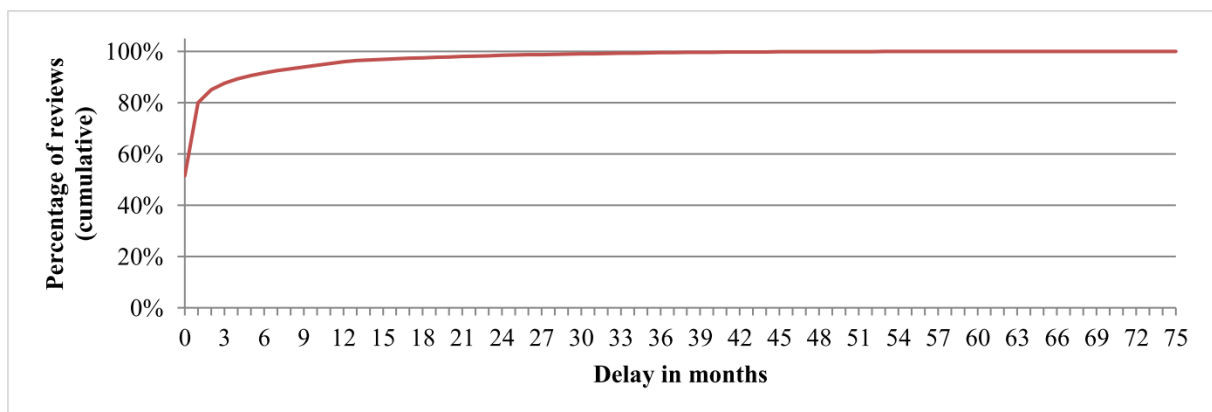




Figure 5 Correlation between VTC.TravelReviews (online travel reviews) and different UNWTO indicators (linear-linear scales)

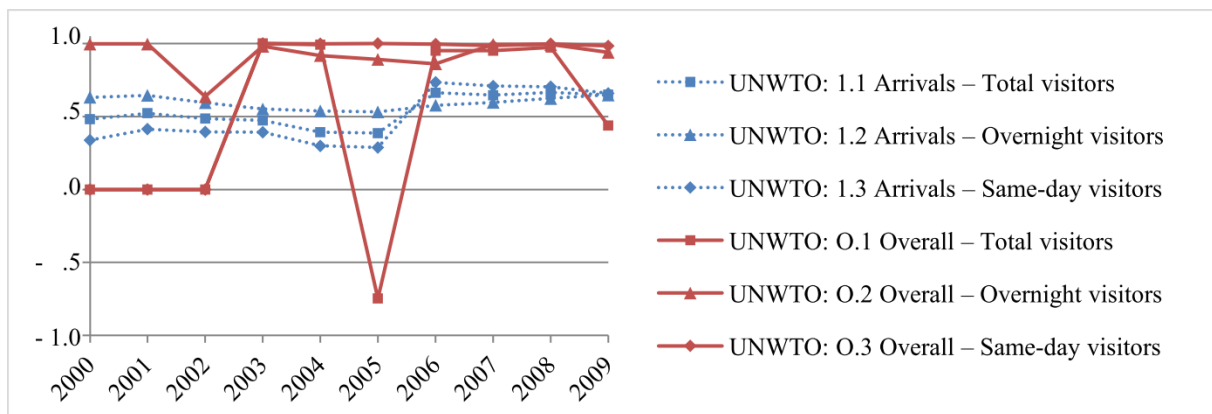


Figure 6 Correlation between VTC.TravelReviews (online travel reviews) and different UNWTO indicators (log-log scales)

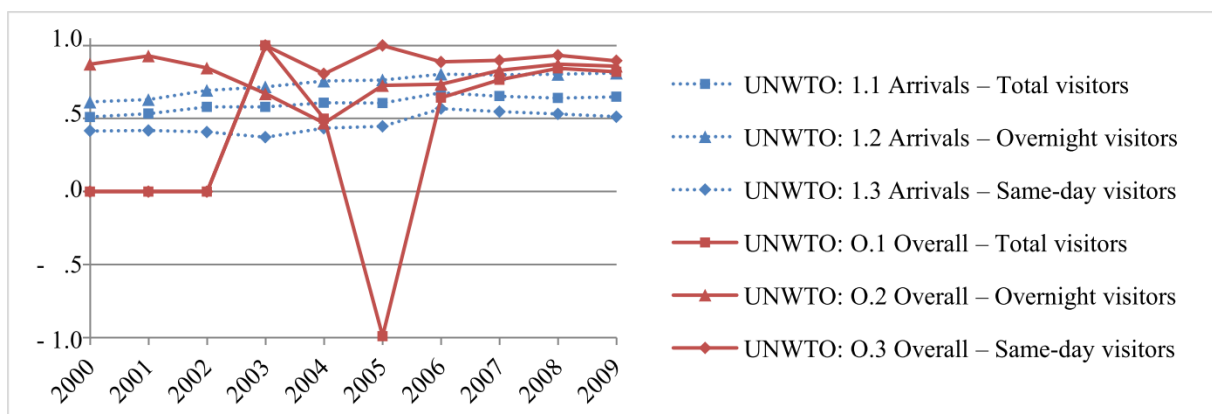


Figure 7 Scatter plot of UNWTO derived indicators of overall tourism and VTC.TravelReviews (online travel reviews) for the year 2009 (log-log scales)

