**Air Pollution and Tourist Sentiment**

# Introduction

空气污染会影响旅游……。数据说明，很多旅游城市有空气污染问题；空气污染带来什么样的影响。很多研究研究了……，但是采用问卷、建模方法，无法……。

社交媒体数据的应用和情感分析的方法，能够……。很多旅游研究用社媒数据和情感分析进行……的研究。

本文利用POI数据，利用情感分析方法分析on-site 和 real-time sentiment，并分析空气污染对sentiment 的影响。

Environmental quality is a prevailing factor in determining competitiveness of tourist destinations, and a crucial issue in the travel decision-making process (Deng, Li, & Ma, 2017; Peng & Xiao, 2018). Air pollution, as one of the greatest environmental issues, its impact on tourism has become a widespread concern, especially in China (Deng et al., 2017). Many tourism destinations are facing to the air pollution issues. For example, estimates have shown that approximately 20% of top-ranked tourist cities are perceived as having heavy air pollution (Hedrick-Wong and Choong 2016). In China, 29% of 5,796 outdoor tourist attractions did not have any clean months in 2016, and the ratio may rise to 34% in 2030 (Sun, Mei, Li, & Shi, 2019). Researchers have examined the effects of air pollution on tourism demand (Wang, Fang, & Law, 2018; Zhou, Qu, Du, Yang, & Liu, 2018), tourism sales revenue (Yoon, 2019), tourist arrivals (Deng et al., 2017; Xu & Reed, 2017), destination image (Becken, Jin, Zhang, & Gao, 2017; Peng & Xiao, 2018), and tourist perception (J. Li, Pearce, Morrison, & Wu, 2016; Peng & Xiao, 2018; Zhang, Hou, Li, & Huang, 2019). However, whether and how the air pollution affects tourist online sentiment is not clear.

Tourist sentiment 很重要。主要用什么方法研究了什么问题，发现……会影响tourist sentiment。However, the effects of air pollution on tourist sentiment is not investigated. Also, those studies estimated the 之后的感受、不在场的感受。Recently, the wide usage of social media data and sentiment analysis technique provide a new way to measure on-site and real-time tourist sentiment ().

Therefore, this study measures the real-time and on-site tourist sentiment, with the help of social media data and sentiment analysis technique, and investigate the impact of air pollution on tourist online sentiment.

Background: effects of air pollution on tourism (in reality; in literature). For example, affect tourist arrivals, destinations etc. ……However, the effects of air pollution on tourists’ online sentiment/experience (on-site and real-time sentiment/experience) need to be investigated.

Social media and the sentiment analysis provide a new way to measure on-site and real-time tourists experience. With the help of sentiment analysis, several studies using social media data to ……

This study aims to confirm the effectiveness of measuring tourists on-site experience using social media and sentiment analysis and investigate the effects of air pollution on tourists’ sentiment. ……

# Literature Review

请同学们协助做两个文献表格

Table1 air pollution on tourism

Table1 sentiment in tourism research（这里列清楚sentiment是怎么测量的）

## Air Pollution and Tourism

## Sentiment in Tourism

# 3. Data and Variables

## 3.1 Data

To measure the real-time and on-site tourist sentiment, check-in social media data was used. Geotagged check-in social media data was collected from Sina Weibo, the most popular microblogging platform in China, using web crawler technology. The check-in Weibo posts contain the users’ location information when posting Weibo in a certain location, and the check-in Weibo posts will be listed in the check-in homepage of this location.

We took China’s AAAAA tourist resorts as our study samples for three concerns. First, China is suffering from increasingly concerning levels of air pollution (Becken et al., 2017). In China, the air pollution issue and its impact on tourism has become a widespread concern (Deng et al., 2017; Peng & Xiao, 2018). Second, the uneven spatial distribution of air pollution in China across provinces and cities makes it easier to capture the impacts of air pollution on different destinations. Third, China’s AAAAA tourist resorts represent the highest level of destinations in China and attract most of the tourists. Therefore, we chose China’s AAAAA tourist resorts as our study samples.

Until November 1st, 2019, 259 tourist destinations were certificated as China’s AAAAA tourist resorts. Some of the destinations do not have a check-in homepage, and some destinations contain multiple sites and have a check-in homepage for each site. For example, the China’s AAAAA tourist resorts “Mount Qingcheng - Dujiangyan Irrigation System Tourist Resort” (青城山-都江堰风景名胜区) contains two destination sites, Qingcheng Mountain (青城山) and Dujiangyan Irrigation System (都江堰水利工程). Therefore, Weibo posts data were collected from Weibo check-in data from 241 check-in homepages of 241 tourist destinations (220 tourist resorts). Finally, a total of 201,653 check-in Weibo posts were collected, and Weibo posts update between August 14th, 2015 and October 14th, 2019.

The limitations of the data must be acknowledged. First, due to the limitation of the application programming interface (API) of Sina Weibo, only Weibo posts that listed in the latest 150 pages of each check-in homepage can be collected. However, since we successfully collected more than 200,000 Weibo posts, we believed it is acceptable when doing big data analysis (e.g., Part et al., 2018). Second, our data may suffer from self-selection bias (Goh, 2012; Hughes, Swaminathan, & Brooks, 2019; Y. Li & Xie, 2019; Meire, Hewett, Ballings, Kumar, & van den Poel, 2019). We discussed this issue in the robustness check section, and used several additional analyses to test the robustness of our results.

## 3.2 Dependent Variable

The wide usage of social media and the sentiment analytic technology provides a new way to measure on-site and real-time expressed sentiment (Zheng, Wang, Sun, Zhang, & Kahn, 2019). Thus, we model the effects of air pollution on tourists’ sentiment using tourists’ Weibo posts. Before sentiment analysis, we cleaned our Weibo posts data following three steps: (1) Weibo posts that seems to be advertisements were excluded; (2) Weibo posts that have too few words (less than 15 Chinese characters) were excluded, as those posts can not reveal the emotional tendency of tourists; (3) the user-level Weibo posts were collapsed into destination/day-level. Finally, a total of 43,506 Weibo posts were included in our analysis.

Sentiment analysis was conducted using API of Tencent sentiment analysis service from Tencent natural language processing platform[[1]](#footnote-1). For each Weibo post, the sentiment analysis first excluded non-Chinese characters (e.g., numbers, English characters, punctuations), then segmented sentences into Chinese characters, and finally get the sentiment scores based on Tencent sentiment dictionary. The sentiment scores range from 0 to 1, with 0 represents an extremely negative sentiment and 1 represents an extremely positive sentiment. We rescale the sentiment score to range from 1 to 100 by multiply it by 100, following Zheng et al. (2019), to make sure the coefficient is not too long. The overall experience index for a tourist site on a given day was measured by the median sentiment of all check-in Weibo posting in this destination on this day. （确认下 measurement of sentiment 是否在旅游里面有创新，如果有把分析过程写的更详细些，以便说contribution；甚至将sentiment的描述分析作为study1）

## 3.3 Independent Variables

Air quality data was collected from China National Environmental Monitoring Centre. This website reports real-time hourly concentrations for the major air pollutants such as PM2.5, PM10, SO2, O3, NO2 and CO at about 1,000 monitoring stations. The monitoring center also releases a composite air quality measure, the Air Quality Index (AQI), which synthesizes information on the six major pollutants.

We collected the air quality records from the nearest monitoring station to each destination site, by calculating the distance between the destination site and monitoring station based on latitude and longitude data. Finally, the real-time air quality records were collected and collapsed to the destination/day-level. Since AQI is a synthesized and widely used index that measures air pollution levels (ref), our first measure of air quality is the synthesized index *AQI*.

Further, China sets 100 as the threshold of “blue sky”, which indicates that when the AQI exceeds the threshold of 100, the air quality is considered as polluted and unacceptable. Thus, we used a dummy variable named as *POLLUTED* to measure whether the air is polluted or not, with 1 represents polluted air (AQI > 100) and 0represents clean air (AQI ≤ 100).

Finally, the Ministry of the Environment ranks the air quality into six levels based on the AQI: excellent (AQI < 50), good (AQI < 100), lightly polluted (AQI < 150), moderately polluted (AQI < 200), heavily polluted (AQI < 300) and severely polluted (AQI ≥ 300). Thus, we also used this AQI ranking as another category variable to measure air conditions, named *LEVEL*.

## 3.4 Control Variables

Factors considered as control variables included destination variable, time variables and weather conditions. First, the destination dummy variables were used to controlled the effects of destination. Destination effects can reflect some unobservable factors like characteristics and heterogeneity of destinations. Thus, the destination (*DESTINATION*) was used as a dummy variable to control for destination effects.

Second, 旅游的季节性。Also, when considering the seasonality and annual or monthly variation of tourism, a better understanding of the impact of climate change and weather condition can be developed (e.g., Becken, 2013; Hewer, Scott, & Fenech, 2016). Therefore, *YEAR* and *MONTH* dummy variables were used to fix the year and month effects. Further, effects of weekends and holidays should be considered (Hewer et al., 2016), as people may be more likely or have more time to travel on these days. Also, people may be happier and expressed more positive sentiment on social media on weekends and public holidays (Zheng et al., 2019). Thus, weekend and public holiday dummies (*HOLIDAY*) were used to show how tourists’ sentiment changes in weekends and holidays.

Finally, it has been generally accepted that weather conditions may affect tourist perception, experience and behavior (McKercher, Shoval, Park, & Kahani, 2015; Wilkins, Urioste-Stone, Weiskittel, & Gabe, 2018) (Simpson et al., 2008). Thus, weather conditions were included in our model. A number of studies have shown that temperature (Falk, 2014; Hewer et al., 2016; Wilkins et al., 2018), precipitation (Falk, 2014; Hewer et al., 2016), humid (Becken, 2013; Chen, Lin, Li, & Liu, 2017; Goh, 2012), wind (Becken, 2013; Goh, 2012) and cloudiness (Lin & Matzarakis, 2011) affect tourist perception and tourism activity. Thus, temperature (*TEMPERATURE*), humidity (*HUMID*), wind level (*WIND*), precipitation (*RAIN*) and cloudiness (*CLOUD*) were considered.

Also, it has been observed that temperature has an inverted U relationship with tourism demand (Falk, 2014), tourist arrivals (Hewer et al., 2016), social media expressed sentiment (Zheng et al., 2019). Therefore, the quadratic term of temperature (*TEMPERATURE^2*) were also included in the model. Weather data of each destination were collected from a widely-used Chinese weather query website. All variables used in this study are listed in Table X.

Table X. Variable definitions and summary statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Definition | Mean | Min. | Max. | Std. | n |
| *SENTIMENT* | Median sentiment index in a destination/day level | 80.35 | 0 | 100 | 33.31 | 43,506 |
| Source: Sina Weibo (http://www.weibo.com) | | | | | | |
| *AQI* | Air quality index in a destination on a day | 58.32 | 7.87 | 500 | 36.05 | 43,506 |
| *POLLUTED* | A dummy variable equals to 1 if AQI exceeds 100 | / | 0 | 1 | / | 43,506 |
| *LEVEL* | A category variable with six levels of air pollution (*Excellent*, *Good*, *Lightly polluted*, *Moderately polluted*, *Heavily polluted* and *Severely polluted*) | / | / | / | / | 43,506 |
| Source: China National Environmental Monitoring Centre (www.cnemc.cn) | | | | | | |
| *TEMPERATURE* | Temperature in a destination on a day (℃) | 25.08 | -23 | 46 | 8.63 | 43,506 |
| *TEMPERATURE^2* | The quadratic term of *TEMPERATURE* | 703.27 | 0 | 2116 | 386.03 | 43,506 |
| *HUMID* | Humidity in a destination on a day ranging from 1 to 100, with 100 represents highly humid. | 58.23 | 4 | 100 | 19.56 | 43,506 |
| *WIND* | Wind level in a destination on a day | 5.13 | 1 | 18 | 1.91 | 43,506 |
| *RAIN* | Precipitation in a destination on a day (mm) | 6.88 | 0 | 636 | 19.45 | 43,506 |
| *CLOUD* | Cloud-covered in a destination on a day ranging from 1 to 100, with 100 represents full cloud. | 41.09 | 0 | 100 | 30.9 | 43,506 |
| Source: a Chinese weather query website (https://tianqi.911cha.com/guoneijingdian.html) | | | | | | |
| *YEAR* | Year dummies | / | 2015 | 2019 | / | 43,506 |
| *MONTH* | Month dummies | / | 1 | 12 | / | 43,506 |
| *HOLIDAY* | A dummy variable equals to 1 if the day is a weekend or a public holiday |  | 0 | 1 |  | 43,506 |
| *DESTINATION* | Destination dummies | / | / | / | / | 43,506 |

# 4. Results

## 4.1 Base Model

We first look at the simple relationship between air pollution and tourist sentiment. First, the data is sorted by AQI and divided into 500 groups. As shown in Figure 1, the mean value of median sentiment index for each group is represented as dot, and fitted by the downwards sloping line with 95% confidence interval. There is a negative correlation between AQI and tourist sentiment as shown in Figure 1. Second, we plotted the median sentiment index across different AQI levels. The line chart in Figure 2 also shows a significant decrease of median sentiment index in higher AQI levels. Considering that the air pollution shows a significant and negative correlation with tourist sentiment, we took AQI as the independent variable and the median sentiment index as the dependent variable and proposed a base OLS model without any control variables:

|  |  |
| --- | --- |
| *SENTIMENTi,j* = β0 + β1*AQIi* + ε*i,j* | (1) |

In this model, *SENTIMENTi,j* indicates the median sentiment index of destination *i* on date *j*, β0 is the intercept term, and *εi* is the random error term.

The result of the base model shows that the *AQI* has a significant and negative effect on tourist sentiment (β1 = −0.0445, *p* < 0.01). Given that the dependent variable may be affected by other factors, we next specify our main model with controlling for destination effects, time effects and weather conditions.

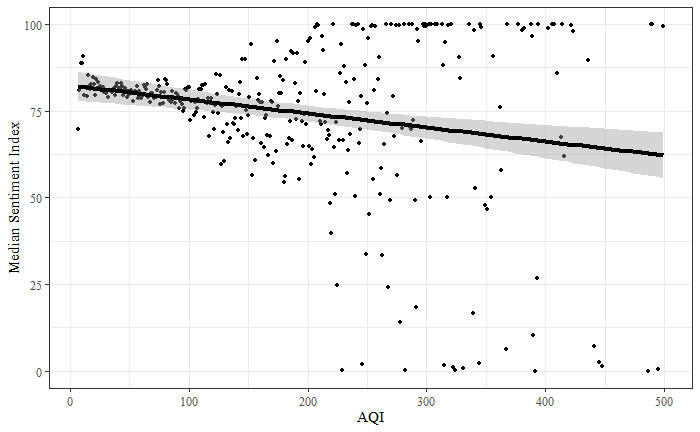


Figure 1. Median Sentiment Index and AQI

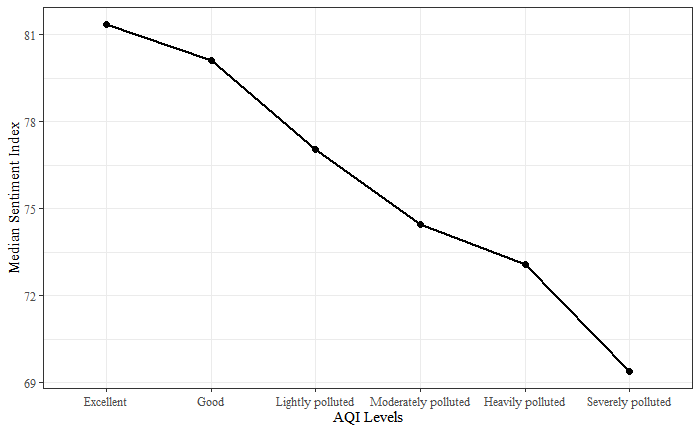


Figure 2. Median Sentiment Index and AQI Levels

## 4.2 Model Specification

To specify the effect of air pollution on tourists’ sentiment, we controlled for destination effects, time effects and weather conditions. Using three measures of air pollution as the key independent variable and median sentiment as the dependent variable, our full model was proposed as follows:

|  |  |
| --- | --- |
| *SENTIMENTi,j* = β0 + β1*POLLUTIONi* + β2*TEMPERATUREi* + β3*TEMPERATURE^2i* + β4*HUMIDi* + β5*WINDi* + β6*CLOUDi* + β7*RAINi* + β8*HOLIDAYi* + Ω*DESTINATION* + Π*YEAR +* ΛMONTH+ ε*i,j* | (2) |

In this model, *POLLUTION* represents the three different measures of air pollution: *AQI*, *POLLUTED*, and *LEVEL*. Thus, β1 is the key coefficient; a negative value of β1 indicates the negative effects of air pollution on tourists’ sentiment.

## 4.3 Effect of Air Pollution on Tourist Sentiment

The results of our full model (Equation (2)) was presented in Table X. After controlling for the control variables, AQI still shows a negative and significant relationship between tourist sentiment. In Full Model 1, the coefficient (β1 = −0.0152, *p* < 0.01) indicates that a one standard deviation increase in AQI is related to a 0.0152 standard deviation decrease in tourists’ median sentiment index.

Then, we estimated the effects of the dummy variable (*POLLUTED*) on tourists’ sentiment. Results in Full Model 2 shown that the median sentiment expressed in the polluted days (AQI ≥ 100) was significantly smaller than in the unpolluted condition (AQI < 100), indicating that the tourists represents more negative sentiment on polluted days (β1 = -1.1995, *p* < 0.05).

Finally, we analyzed the effects of air pollution levels on tourists’ sentiment, using a category variable (*LEVEL*) as the independent variable. Results in Full Model 3 shown a non-linear and negative relationship between air pollution and tourists’ sentiment, which indicates that tourists’ sentiment decreased monotonically and non-linearly with the increase of the air pollution levels. Specifically, the tourists’ sentiment was significantly and negatively affected by air pollution when in *Moderately polluted* (β1 = -2.7213, *p* < 0.05), *Heavily polluted* (β1 = -3.8512, *p* < 0.05) and *Severely polluted* (β1 = -6.4443, *p* < 0.1) conditions.

As for the control variables, after controlled for *YEAR* and *MONTH* effects, weather conditions and *HOLIDAY* and are also correlated with the tourists’ sentiment. First, *TEMPERAURE* shows an inverted U effect on tourists’ sentiment, with a positive coefficient of *TEMPERATURE* and a negative coefficient of the quadratic term (*TEMPERATURE^2*). Second, *HUMID* and *CLOUD* show negative and significant effects, while *WIND* and *RAIN* have nonsignificant effects on tourists’ sentiment. Finally, tourists show a more positive sentiment on weekends and holidays (*HOLIDAY*) than on workdays.

Table X. Effects of Air Pollution on Tourists’ Sentiment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Base Model** | **Full Model 1** | **Full Model 2** | **Full Model 3** |
| *AQI* | -0.0445\*\*\* | -0.0152\*\*\* |  |  |
| *Polluted (AQI ≥ 100)* |  |  | -1.1995\*\* |  |
| Default: *LEVEL* *(Excellent)* | | | | |
| *LEVEL (Good)* |  |  |  | n.s. |
| *LEVEL (Lightly polluted)* |  |  |  | n.s. |
| *LEVEL (Moderately polluted)* |  |  |  | -2.7213\*\* |
| *LEVEL (Heavily polluted)* |  |  |  | -3.8512\*\* |
| *LEVEL (Severely polluted)* |  |  |  | -6.4443\* |
| *Temperature* |  | 0.3374\*\*\* | 0.3391\*\*\* | 0.3396\*\*\* |
| *Temperature^2* |  | -0.0092\*\*\* | -0.0092\*\*\* | -0.0093\*\*\* |
| *Humid* |  | -0.0288\* | -0.0258\* | -0.0249\* |
| *Wind* |  | n.s. | n.s. | n.s. |
| *Rain* |  | n.s. | n.s. | n.s. |
| *Cloud* |  | -0.0179\*\* | -0.0160\*\* | -0.0162\*\* |
| *Holiday* |  | 1.3223\*\*\* | 1.3116\*\*\* | 1.3312\*\*\* |
| *YEAR* | NO | YES | YES | YES |
| *MONTH* | NO | YES | YES | YES |
| *DESTINATION* | NO | YES | YES | YES |
| R^2 | 0.0023 | 0.2486 | 0.2486 | 0.2487 |

Note: \*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01

# 5. Robustness Checks

## 5.1 Robustness of Independent Variable

PM2.5 is one of the main components of air pollution and the primary pollutant in many areas (ref). Also, in our data samples, AQI and PM2.5 show a high and significant correlation (coefficient = 0.9415, *p* < 0.01). Therefore, we used PM2.5 to test the robustness of the independent variable. Results are shown in Table X. In Model 1, PM2.5 also negatively and significantly affect tourists’ sentiment (β = −0.0434, *p* < 0.01) in the simple free model without control variables. After controlling for related variables, PM2.5 still shows a negative and significant effect (β = −0.0153, *p* < 0.01), indicating that a one standard deviation increases in PM2.5 is related to a 0.0153 standard deviation decrease in tourists’ median sentiment index. The results further reinforce the findings in that air pollution negatively and significantly affects the tourists’ sentiment.

Table X. Results of Robustness Check: Effects of PM2.5

|  |  |  |
| --- | --- | --- |
|  | **Model 1** | **Model 2** |
| *PM2.5* | -0.0434\*\*\* | -0.0153\*\*\* |
| *Temperature* |  | 0.3406\*\*\* |
| *Temperature^2* |  | -0.0092\*\*\* |
| *Humid* |  | -0.0256\* |
| *Wind* |  | n.s. |
| *Rain* |  | n.s. |
| *Cloud* |  | -0.0166\*\* |
| *Holiday* |  | 1.3164\*\*\* |
| *YEAR* | NO | YES |
| *MONTH* | NO | YES |
| *DESTINATION* | NO | YES |
| R^2 | 0.0014 | 0.2486 |

Note: \*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01

## 5.2 Robustness of Dependent Variable

In our main models, we used the median sentiment index as the dependent variable. To test the robustness of our results, we also estimated the effects of air pollution on the mean sentiment index. We collapsed the sentiment into destination/day level based on the mean value of the sentiment, and took the mean sentiment index as the dependent variable. Similarly, we estimated the effects of three measure (*AQI*, *POLLUTED* and *LEVEL*) on the mean sentiment index. As shown in Table X, after replacing our dependent variable into the mean sentiment index, the air pollution still negatively and significantly affects the tourists’ sentiment and shows similar results to our main models, again confirming the robustness of our previous findings.

Table X. Results of Robustness Checks on Dependent Variable

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| *AQI* | -0.0256\*\*\* | -0.0156\*\*\* |  |  |
| *Polluted (AQI ≥ 100)* | |  | -1.2527\*\* |  |
| Default: *LEVEL (Excellent)* | | | | |
| *LEVEL (Good)* |  |  |  | n.s. |
| *LEVEL (Lightly polluted)* | |  |  | n.s. |
| *LEVEL (Moderately polluted)* | |  |  | -2.6905\*\* |
| *LEVEL (Heavily polluted)* | |  |  | -4.1251\*\*\* |
| *LEVEL (Severely polluted)* | |  |  | -7.3558\*\* |
| *Temperature* |  | 0.2984\*\*\* | 0.2969\*\*\* | 0.2970\*\*\* |
| *Temperature^2* | | -0.0083\*\*\* | -0.0083\*\*\* | -0.0085\*\*\* |
| *Humid* |  | -0.0239\* | -0.0208\* | n.s. |
| *Wind* |  | n.s. | n.s. | n.s. |
| *Rain* |  | n.s. | n.s. | n.s. |
| *Cloud* |  | -0.0167\*\* | -0.0157\*\* | -0.0159\*\* |
| *Holiday* |  | n.s. | n.s. | n.s. |
| YEAR | NO | YES | YES | YES |
| MONTH | NO | YES | YES | YES |
| DESTINATION | NO | YES | YES | YES |
| R^2 | 0.0008 | 0.2330 | 0.2329 | 0.2330 |

Note: \*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01

## 5.3 Selection Bias

Since we were dealing with an online population, our samples may suffer from self-selection bias (Hughes et al., 2019; Y. Li & Xie, 2019; Meire et al., 2019) (Goh, Heng, & Lin, 2013). For example, tourists travelled to the destination are obviously more likely to express their experience on the social media; also, tourists have an extreme positive or extreme negative experience are more likely to report their on-site experience; finally, the sentiment expressed in the Weibo posts may due to some unobservable factors, which drives the emotion tendency in the Weibo posts. Therefore, to alleviate the endogeneity and selective bias concern, we used a Propensity Score Matching (PSM) method. PSM is an effective way to adjust for the differences in the treatment and control group (Rosenbaum & Rubin, 1983), and is widely-used in related studies using social media data (e.g., Y. Li & Xie, 2019).

In this study, the propensity score is the predicted probability that the observe unit (a certain destination in a certain day) receives a treatment condition (a polluted day) on the value of covariates. When the propensity scores are close enough, the treatment is considered random, and the selection bias are considered eliminated.

First, we specified a logit model to predict the probability of a destination in a polluted day. We used the dummy variable (*POLLUTED*) as the dependent variable, and the predict variables include destination variables, time variables and weather conditions. Different from our main model, we use the latitude (*LATITUDE*) and longitude (*LONGITUDE*) to capture the location of destination. Also, *YEAR*, *MONTH* and *HOLIDAY* dummy variables, and weather variables similar to the main model were used in this logit model. Results were shown in Table X. Results indicate that those variables are good predictors to the probability of a destination in a polluted day. Thus, those variables were used in the next matching step.

Based on the results of the logit model, a 1:1 nearest-neighbor matching algorithm was used to match a destination in a polluted day with a destination in an unpolluted day. The resulting matched sample contains a total 7,590 samples, with 3,975 in the polluted condition and 3,795 in the unpolluted condition.

Matching results were shown in Table X. After matching, the control group (unpolluted condition) and treatment group (polluted condition) are not significantly different in all covariates. When considering the sentiment reported in each destination, the median sentiment index was significantly different between control group and treatment group. Specifically, the median sentiment index in the polluted condition (mean = 76.04) was significantly smaller than the median sentiment index in the unpolluted condition (mean = 79.06, *p* < 0.001), again confirm the negative effect of air pollution on tourists’ sentiment. Therefore, after eliminating the selection bias problem, our results were still robustness.

Table X. Results of the logit model in PSM

|  |  |
| --- | --- |
|  | **Logit** |
| *Latitude* | 0.1799\*\*\* |
| *Longitude* | n.s. |
| *Temperature* | 0.1227\*\*\* |
| *Temperature^2* | -0.0014\*\*\* |
| *Humid* | -0.0138\*\*\* |
| *Wind* | -0.0509\*\*\* |
| *Rain* | -0.0170\*\*\* |
| *Cloud* | n.s. |
| *Holiday* | 0.1797\*\*\* |
| YEAR  MONTH | YES  YES |

Note: \*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01

Table X. Results of PSM

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Unpolluted (AQI < 100)** | **Polluted (AQI ≥ 100)** | **p** |
| *SENTIMENT* | 79.06 (34.51) | 76.04 (36.91) | <0.001 |
| *Latitude* | 33.53 (6.14) | 33.51 (4.35) | 0.865 |
| *Longitude* | 113.17 (7.80) | 113.02 (5.87) | 0.357 |
| *Temperature* | 18.58 (9.54) | 18.47 (9.34) | 0.623 |
| *Temperature^2* | 436.20 (350.85) | 428.56 (372.74) | 0.358 |
| *Humid* | 46.59 (20.19) | 46.51 (17.90) | 0.852 |
| *Wind* | 5.22 (1.93) | 5.15 (1.89) | 0.090 |
| *Rain* | 1.21 (4.67) | 1.13 (4.77) | 0.499 |
| *Cloud* | 29.89 (29.76) | 29.41 (28.05) | 0.466 |
| YEAR | YES | YES | / |
| MONTH | YES | YES | / |
| HOLIDAY | YES | YES | / |
| n | 3795 | 3795 | / |

Note: \*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01. Mean values reported and standard deviation in the parentheses.

# 6. Conclusion and Discussion

Contribution:

1. 从sentiment角度拓展pollution对tourist behavior的影响
2. 从pollution角度拓展对tourist on-site sentiment的认识
3. 从sentiment analysis角度提出测量on-site and real time tourist sentiment的方法
4. 证明利用UGC和sentiment analysis 分析游客在线情感/体验的有效性。a new index to measure the on-site and real-time tourist experience using social media data and sentiment analysis
5. 识别空气污染影响的新因变量（发现空气污染对游客在线情感的影响）identifying the effects of air pollution on tourists’ online sentiment

Implication:

1. Pollution：
2. Sentiment：
3. Measure：
4. for destination, reports or alerts
5. for tourists, protection or reschedule.

# References

References

Becken, S. (2013). Measuring the Effect of Weather on Tourism. *Journal of Travel Research*, *52*(2), 156–167. https://doi.org/10.1177/0047287512461569

Becken, S., Jin, X., Zhang, C., & Gao, J. (2017). Urban air pollution in China: destination image and risk perceptions. *Journal of Sustainable Tourism*, *25*(1), 130–147. https://doi.org/10.1080/09669582.2016.1177067

Chen, C.‑M., Lin, Y.‑C., Li, E. Y., & Liu, C.‑C. (2017). Weather uncertainty effect on tourism demand. *Tourism Economics*, *23*(2), 469–474. https://doi.org/10.5367/te.2015.0513

Deng, T., Li, X., & Ma, M. (2017). Evaluating impact of air pollution on China’s inbound tourism industry: a spatial econometric approach. *Asia Pacific Journal of Tourism Research*, *22*(7), 771–780. https://doi.org/10.1080/10941665.2017.1331923

Falk, M. (2014). Impact of weather conditions on tourism demand in the peak summer season over the last 50years. *Tourism Management Perspectives*, *9*, 24–35. https://doi.org/10.1016/j.tmp.2013.11.001

Goh, C. (2012). Exploring impact of climate on tourism demand. *Annals of Tourism Research*, *39*(4), 1859–1883. https://doi.org/10.1016/j.annals.2012.05.027

Hewer, M., Scott, D., & Fenech, A. (2016). Seasonal weather sensitivity, temperature thresholds, and climate change impacts for park visitation. *Tourism Geographies*, *18*(3), 297–321. https://doi.org/10.1080/14616688.2016.1172662

Hughes, C., Swaminathan, V., & Brooks, G. (2019). Driving Brand Engagement Through Online Social Influencers: An Empirical Investigation of Sponsored Blogging Campaigns. *Journal of Marketing*, *83*(5), 78–96. https://doi.org/10.1177/0022242919854374

Li, J., Pearce, P. L., Morrison, A. M., & Wu, B. (2016). Up in Smoke? The Impact of Smog on Risk Perception and Satisfaction of International Tourists in Beijing. *International Journal of Tourism Research*, *18*(4), 373–386. https://doi.org/10.1002/jtr.2055

Li, Y., & Xie, Y. (2019). Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research*, 1–19.

Lin, T.‑P., & Matzarakis, A. (2011). Tourism climate information based on human thermal perception in Taiwan and Eastern China. *Tourism Management*, *32*(3), 492–500. https://doi.org/10.1016/j.tourman.2010.03.017

McKercher, B., Shoval, N., Park, E., & Kahani, A. (2015). The [Limited] Impact of Weather on Tourist Behavior in an Urban Destination. *Journal of Travel Research*, *54*(4), 442–455. https://doi.org/10.1177/0047287514522880

Meire, M., Hewett, K., Ballings, M., Kumar, V., & van den Poel, D. (2019). The Role of Marketer-Generated Content in Customer Engagement Marketing. *Journal of Marketing*, *83*(6), 21–42. https://doi.org/10.1177/0022242919873903

Peng, J., & Xiao, H. (2018). How does smog influence domestic tourism in China? A case study of Beijing. *Asia Pacific Journal of Tourism Research*, *23*(12), 1115–1128. https://doi.org/10.1080/10941665.2018.1527776

Sun, J., Mei, J., Li, H., & Shi, J. (2019). Reschedule Your Travel Plans: Human Health and Air Pollution. *Journal of Travel Research*, *58*(8), 1408–1410. https://doi.org/10.1177/0047287518806411

Wang, L., Fang, B., & Law, R. (2018). Effect of air quality in the place of origin on outbound tourism demand: Disposable income as a moderator. *Tourism Management*, *68*, 152–161. https://doi.org/10.1016/j.tourman.2018.03.007

Wilkins, E., Urioste-Stone, S. de, Weiskittel, A., & Gabe, T. (2018). Effects of Weather Conditions on Tourism Spending: Implications for Future Trends under Climate Change. *Journal of Travel Research*, *57*(8), 1042–1053. https://doi.org/10.1177/0047287517728591

Xu, X., & Reed, M. (2017). Perceived pollution and inbound tourism in China. *Tourism Management Perspectives*, *21*, 109–112. https://doi.org/10.1016/j.tmp.2016.12.006

Yoon, H. (2019). Effects of particulate matter (PM10 ) on tourism sales revenue: A generalized additive modeling approach. *Tourism Management*, *74*, 358–369. https://doi.org/10.1016/j.tourman.2019.04.008

Zhang, K., Hou, Y., Li, G., & Huang, Y. (2019). Tourists and Air Pollution: How and Why Air Pollution Magnifies Tourists’ Suspicion of Service Providers. *Journal of Travel Research*, *20*, 1-20. https://doi.org/10.1177/0047287519859710

Zheng, S., Wang, J., Sun, C., Zhang, X., & Kahn, M. E. (2019). Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nature Human Behaviour*, *3*(3), 237–243. https://doi.org/10.1038/s41562-018-0521-2

Zhou, B., Qu, H., Du, X., Yang, B., & Liu, F. (2018). Air Quality and Inbound Tourism in China. *Tourism Analysis*, *23*(1), 159–164. https://doi.org/10.3727/108354217X15143857878714

1. https://nlp.qq.com/help.cgi?topic=api#sentiment [↑](#footnote-ref-1)