Ecological Economics 200 (2022) 107502

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Analysis

Hot in Twitter: Assessing the emotional impacts of wildfires with sentiment analysis

Maria L. Loureiro [a,*], Maria Alló [b], Pablo Coello [a]

[a] ECOBAS-Universidade de Santiago de Compostela, Departament of Economic Foundations, Spain
[b] ECOBAS-Universidade da Coruña, Department of Economics, Spain

ARTICLE INFO

JEL codes:
D6
Q5

Keywords:
Environmental Impacts
Natural Language Processing
Happiness
Hedonometer
Wildfires

ABSTRACT

Social media generates a significant amount of information in terms of perceptions, emotions, and sentiments. We present an economic analysis using the information provided by Twitter messages, describing impressions and reactions to wildfires occurring in Spain and Portugal. We use natural language processing techniques to analyze this text information. We generate a hedonometer estimate on how sentiments about wildfires vary with exposure, measured via Euclidean distance from the catastrophic event, and air quality. We find that direct exposure to wildfires significantly decreases the expressed sentiment score and increases the expressions of fear and political discontent (protest). Economic valuation of these losses has been computed to be between 1.49€–3.50€/year/Kilometer of distance to the closest active fire. Welfare losses in terms of air quality have been computed as 4.43€–6.59€/day of exposure.

1. Introduction

The effects of extreme environmental events are usually difficult to quantify. In general, a complete environmental and social impact assessment study is required for each of these episodes. However, due to the significant amount of resources required and time needed for these analyses, the impacts of large extreme events are often significantly under-documented. Costs related to these events are multiple and can be broadly classified between direct and indirect costs. Direct costs include biodiversity losses, tourism losses, and forest production losses, for example; whereas indirect losses may entail a much wider array, including fear, anxiety and psychological impacts that affect the population. Furthermore, the indirect effects could go also beyond national boarders, adding a significant spillover effect to the already complex domestic assessments.

It is challenging to identify the specific causal effects of extreme events on individuals (Metcalfe et al., 2011), since there are sometimes no good comparable counterfactuals. As a result of this, a popular way of eliciting well-being consequences and evaluating the indirect negative externalities is through a stated preference study, which estimates a direct willingness-to-pay (WTP) for a reduction in the risk of a particular event (Bateman et al., 2002). The negative effects of wildfires via stated preferences have been analyzed widely in the literature (Englin et al., 2001; Richardson et al., 2012; among others).

A growing branch of the literature has investigated the impact of extreme events on life satisfaction, subjective well-being (SWB), and happiness. The approach has been used to value the residual impacts of climate (Rehdanz and Maddison, 2008), urban air pollution and air quality (Welsch, 2006; Luechinger, 2009), airport noise nuisance (van Praag and Baarsma, 2005), terrorism (Metcalfe et al., 2011), flood hazards (Luechinger and Raschky, 2009), nuclear accidents (Rehdanz et al., 2015), and more recently wildfires in Australia (Ambrey et al., 2017), among others. On a conceptual note, SWB is compared to the standard non-market valuation techniques in Frey et al. (2009), and Dolan and Metcalfe (2008).

We base our research on this SWB and happiness literature in order to assess the impacts of catastrophic events on the emotional state of individuals. Analyzing Twitter messages and using Natural Language Processing (NPL) tools, we construct indicators of sentiments or levels of expressed happiness. Sentiment analysis is a data mining technique that allows for the interpretation and classification of emotions (positive, negative and neutral) with text data. These emotions and sentiments are latent in the spontaneous conversations. Our evaluation method relies on the fact that expressed sentiments are related to experienced utility. We relate these sentiment indicators to the distance, air quality, as well as other socio-economic and climatic conditions, and cultural indicators.

* Corresponding author.
E-mail address: maria.loureiro@usc.es (M.L. Loureiro).

https://doi.org/10.1016/j.ecolecon.2022.107502
Received 10 February 2021; Received in revised form 2 April 2022; Accepted 24 May 2022
Available online 22 June 2022

This study present causal evidence of wildfire occurrence on emotional states and sentiments, using social media. We use Twitter data to examine how Spanish and Portuguese wildfires have detrimental effects on the emotional state of residents with respect to their exposure. These data allow us to estimate the effects in a quasi-experimental setting, using social media data retrieved before, during, and after the catastrophic fires. We find that, among other factors, proximity to wildfires and bad air quality decreased expressed sentiments of happiness significantly.

2. Relevant literature

Natural disasters are widely covered in social media channels. A branch of the economic literature assesses the impact of disasters and natural hazards using social media in human behavior, particularly text. To name a few, Baylis (2020) studies how climate change impacts the way individuals communicate on Twitter, finding consistent and statistically significant declines in expressed sentiment, from both hot and cold temperatures. Kryvasheyeu et al. (2016) study the impact of hurricane Sandy on Twitter conversations, finding a relationship between the proximity to hurricane Sandy's path and economic losses and social media activity. Sisco et al. (2017) analyzed Twitter conversations to assess how extreme weather events generate attention to climate change, finding that the financial damage linked with these events is a good predictor of the attention paid to climate change in the USA.

A different stream of literature analyzes issues related to sensitivity towards disasters and risk perceptions and emotions in the context of hazards mentioned in social media. Theja Bhavaraju et al. (2019) assessed the sensitivity of social media communication to different natural disasters, finding that in comparison to other extreme events, wildfires are less likely to shift tweet frequencies or impact negative sentiment with respect to tornadoes and floods, for example. This is possibly justified by the fact that on most occasions forest fires do not occur near highly populated areas. Kibanov et al. (2017) have shown how mining social media can be useful to understand changes in behavior of users, and thus, design and improve peatland fire and haze disaster management in Indonesia. With the same goal, Tavra et al. (2021) uses social media for filling the gaps in authoritative data and improve crisis mapping in the context of wildfires. Other topics, such as the social amplification of risk (or the notion that identifies certain risks considered as insignificant by experts and end up generating massive public reaction), have been also addressed in the context of social media by Wirz et al. (2018). Analyzing Zika conversations in Twitter, it was found that the attribution of blame or certain incompetence when dealing with the outbreak contributed to the social amplification of risk in social media.

More recently, crowdsourced data from social media are gaining momentum for ecosystem service valuation studies. For example, Ghermandi et al. (2020) investigate using geotagged photos cultural services accrued to local, domestic, and international recreational visitors to the Usumacinta floodplain, a coastal region in Mexico with one of the highest biological and cultural diversities. They find that locals and internationals have different recreational preferences, with international tourists being partially restricted to well-known and accessible sites; while locals post a larger number of pictures with lagoons, natural spaces and traditional landscapes. Sinclair et al. (2020) use geotagged pictures uploaded to social media as a potential source of information substitute of survey data to estimate a travel cost model to German national parks. They find, as expected, downward sloping demand curve, with consumer surplus for access to the parks ranging between €16.54 and €34.90 (2016 prices). In summary, the previously existing literature highlights the importance and validity of social media data reflecting that "human sensors" may anticipate economic impacts (Kirilenko et al., 2015). These types of studies establish citizens as important providers of useful information for science and decision making.

Our work contributes to this literature by providing an analysis of sentiments and perceptions obtained from Twitter conversations during the 2017 Autumn wildfires in Spain and Portugal, going beyond previous wildfire studies, and being one of the first that incorporates the expressed sentiment into a utility framework. The literature on sentiment analysis is growing rapidly, with recent applications to the COVID-19 pandemic (Chandra and Krishna, 2021), the financial crisis (Wan et al., 2021), and vaccines (Khakharia et al., 2021), to name a few. We believe that the immediacy of this source of data can be quite useful to obtain a first approximation of social impacts derived from different events. Furthermore, social data can be a good complement of surveys for stated preferences. In addition, these types of assessments can be carried out at international level in almost real time in order to understand additional geographical spillover effects of the impacts.

3. Autumn wildfires 2017

In October 2017, the Iberian Peninsula suffered an intensive wave of wildfires, concentrated mostly in the north-west. Although some wildfires were registered on the 4th of October, the worst period was between 13-18th October. In that week more than 7900 forest fires affected Northern Portugal and several regions of Northwestern Spain (Galicia, Asturias, and Castile-León). These wildfires were mainly intentional and claimed the lives of at least 49 individuals, and dozens more were injured. Fires started in Galicia (Spain) on October 13th, and by October 15th, they grew out of control, due to different factors, including the impact of the winds from hurricane Ophelia (New York Times, 2017).

During 2017, according to the data published by the Spanish Ministry of Agriculture and Fisheries, Food and Environment (MAPAMA, 2017), the total number of hectares burned in Spain (174,788) represented the second highest record in the last decade; while in terms of mega-fires (or forest fires above 500 ha), 2017 was the worst year in record, with more than 52 mega-fires. For Portugal, 2017 was the worst year on record so far, in terms of fatalities (more than 120 human lives lost during the year) and 500,000 ha burned (Turco et al., 2019). As far as we know, no comprehensive study has assessed how these dramatic events have affected social well-being.

4. Theoretical background and NLP Methodology

4.1. Economic model

We define the indirect utility function *(V)* as a function of income *(y)* and environmental quality *(h)* as well as other personal and social characteristics (x), whereby:

$$V(y,h,x) \tag{1}$$

Consider a reduction of environmental quality due to a hazard from h_0 to h_1, where $h_0 > h_1$ ceteris paribus. A change in utility derived from a change in hazard levels is then defined as:

$$\Delta V = V(y,h_0|x) - V(y,h_1|x) \tag{2}$$

The compensating variation (CV) is the amount of money that an agent needs to be equal in terms of utility to the prior condition (or condition without the change in environmental quality), and it is given by:

$$\Delta V = V(y|h_0,x) = V(y+CV|h_1,x) \tag{3}$$

Using this theoretical framework, the empirical sentiment equation is denoted as follows:

$$V_{it} = \beta_0 + \beta_1 y_{it} + \beta_2 h_{it} + \sum_{j=1}^{J} \gamma_j X_{j,it} + \varepsilon_{it}, \tag{4}$$

where V_{it} is the utility of individual *i* at time t. Given that utility is non observable, we use as a proxy the sentiment or happiness of each individual *i* extracted from Twitter conversations at time *t*; y_{it} is the income of each individual, h_{it} is the distance to the hazard (active wildfire); and X_{it} are socio-demographic, cultural, and other environmental conditions, including air quality, that may affect overall wellbeing; whereas the betas and gammas are coefficients to be estimated. ε_{it} is a composite error term for individual time invariant characteristics and other environmental time-varying characteristics for each individual i at time t.

Using the empirical specification depicted in eq. (4) to solve for CV in eq. (3), we find that:

$$CV = \frac{\beta_2}{\beta_1} \tag{5}$$

The estimation of eq. (4) and the corresponding CV estimate in eq. (5) involves several important steps. In this particular case, this CV measures the amount of money to be equal in term of utility from being a kilometer closer to the fire. In addition to those related to the econometric modeling, discussed later in this paper, the estimation of the sentiment variable via Twitter employing natural language processing (NLP) algorithms requires important steps.

4.2. NLP methodology

Sentiment analysis is an area of study within the NLP that measures subjectivity and opinions in text. It usually captures an evaluative factor (positive or negative) and its corresponding strength (degree to which the word or phrase is positive or negative (Bhadane et al., 2015)). Although in this application we are using digital text, earlier applications of sentiment analysis used newspapers and printed texts. There are different ways to extract sentiments from text. We used a lexicon-based approach, by which general orientation from a tweet comes from the semantic orientation of words contained in the various lexica used.

To analyze the information retrieved from Twitter, the Hedonometer algorithm has been used (Cody et al., 2015). It should be noted that although this Hedonometer algorithm has been identified with the use of specific lexica and a particular mathematical formulation, from an economic point of view, we consider the term hedonometer broadly, and identical to what was initially referred to as a "hedonimeter" by Edgeworth that describes an instrument for "registering the height of pleasure experienced by an individual" (Colander, 2007). The following analysis is based on different lexica in order to check the robustness of our findings.

The hedonometer algorithm consists of the analysis of a text through its fragmentation into phrases, and subsequently the phrases into words. Words are associated with scores of positive and negative feelings, whereby a total score for the sentence is obtained by different aggregation procedures for the overall topic. Table A1 (Appendix) shows the scores of happiness provided to examples of single words in the LabMT lexicon, including "laughter" (score 8.5), or "deaths" (score 1.64). In order to get an aggregate sentiment score for the entire tweet we employ a weighted average score, considering the frequencies of the different words in each Tweet (as in Eq. (6)).

Multiple lexica can be used for sentiment analysis. In the present study, we use the LabMT lexicon (Dodds et al., 2011), assigning scores of "happiness" to words from 1 to 9 as in Table A1; the VADER (Valence Aware Dictionary and *sEntiment Reasoner*) lexicon (Hutto and Gilbert, 2014), which assigns scores from −1 to 1 (interesting for polarity classification); and the AFINN lexicon (Nielsen, 2011), which assigns happiness scores from −5 to 5, all three of which are quite popular worldwide. The words contained in these three lexica have been classified according to the subjective values of "happiness" given by different research samples of respondents.

The average wellbeing or magnitude of happiness obtained from the hedonometer per text message using the LabMT lexicon has been calculated following Eq. (6)[1]:

$$S(M) = \frac{\sum_{k=1}^{N} h_{avg}(w_k) f_k}{\sum_{k=1}^{N} f_k} = \sum_{k=1}^{N} h_{avg}(w_k) p_k, \tag{6}$$

where *S(M)* refers to the sentiment obtained from a particular text message (M), f_k is the frequency of the kth word w_k for which we have an estimate of average happiness, and $p_k = f_k / \sum_{r=1}^{N} f_r$ is the corresponding normalized frequency of use of this word.

Once the sentiment scores have been obtained per message, then an average of sentiment is obtained by aggregating all scores, where A_{it} are the resulting vectors of sentiments with a total number of m tweets, from the same individual i at a given time t, so that:

$$A_{j,it} = \{S(M)_1, \dots, S(M)_m\} \tag{7}$$

Being estimated the mean sentiment for each individual i at a given time t (*Vit*),

$$Vit = \frac{\sum_{j=1}^{m} A_{it}}{m}; \forall_i; \forall_t \tag{8}$$

5. Datasets description

5.1. Twitter data

Twitter is a social network that was created in 2006, used to exchange messages with people around the world. These messages are known as "tweets" and are limited to 280 characters. A dataset containing all tweets about wildfires in Spain and Portugal in October 2017 was previously purchased from Twitter Corporation, with the aim of carrying out this study. Our data contained all geo-tagged messages retrieved from a Boolean search string containing the words "arson", OR "fires", OR "forest fires", OR "wildfires". By inclusion, any other tweet where the terms "wildfire" OR "fire" were present (such as brush fires, campfires, intentioned fires, among others) were also included. The original dataset contained a total of 14,790 geotagged tweets[2].

As a first step, it was necessary to clean and process the data. The main purpose of the cleaning process was to remove from the analysis all the information that was not related to the specific topic of interest. For this objective, the following exclusion criteria were defined:

- Tweets Retweet (Rt): a Rt was not valid for a text analysis, since it was a repeat of another tweet
- Response Tweets: If a tweet was a response to another, the full tweet was included, not just the response
- Monosyllable tweets: tweets with 1 or 2 syllables were considered invalid, given that the probability that these tweets provide irrelevant information is 85% (after performing an analysis on 100 tweets with these characteristics)
- Tweets about songs, sayings and others, that may include keywords or interest, but may not be related to the topic at hand
- Tweets with unclear geotags.

The data cleaning process was carried out sequentially, which means that each tweet was analyzed independently of the rest. If any of the previous premises were fulfilled, the tweet was automatically discarded for further analysis. This resulted in a clean dataset of 11,248 tweets.

Regarding the processing, each tweet that was selected as valid for analysis had to undergo a second process, which depending on the

[1] For the AFFIN lexicon, S(M)=$\sum_{k=1}^{N} h_{avg}(w_k)$, and for VADER a broad set of rules is being used, as described in **Hutto and Gilbert (2014)**.

[2] About 3% all al tweets are geotagged.

library[3] used, may have required the elimination of urls, emojis, emoticons, symbols, etc. After that, the tweets were translated into English. The translation facilitates the use of multiple tweets written in local languages, and other foreign languages. In that way, it facilitates the application of multiple libraries initially created only in English. Accuracy of translation was assessed through the verification of multiple tweet samples, taking specially into account the adequate representation of minority languages in the dataset, such as Galician or Catalan.

The final dataset used for statistical analysis contained the scores of sentiments, the date and time of each tweet, as well as the location from where the tweet had been written. In an additional processing step, we retrieved the gender of the user by employing the GenderAPI (*Gender-API*, 2019). This API can detect gender from social media by employing the usernames of the registered accounts. A dictionary of names is associated with the gender of the user, which is assigned in a probabilistic way. When the name is unknown or not registered into the database of this API, the unknown category is assigned. This type of identification method is quite popular in gender identification studies, and its large database makes it very reliable with respect to other identification methods (Santamaría and Mihaljevic, 2018). Based on its output, we generated the "expected female" variable.

Additional information retrieved from the Twitter dataset contemplates the geographical origin of the tweet (Spain or Portugal), as well as whether the tweet contained words referring to specific names of political parties, or popular politicians ("PP", "Popular Party", "PSOE", "Socialist Party"…) to mostly show political discontent or institutional blaming with respect to the occurrence or management of the catastrophic events. The variable is named as "political content", and in a similar way to other preference studies, this latter variable may reflect "protest" attitudes, affecting the sentiment scores significantly.

5.2. Wildfires dataset and meteorological variables

In this study we take advantage of a unique dataset that has been constructed by matching the geo-positioning of each tweet with other physical, climatic, and socio-economic data. We merge the tweets database by individual and time with the wildfires, the climate-related data, air quality data, and other socio-economic factors. The wildfires dataset was obtained from the European Forest Fire Information System Database (EFFIS, 2017). This dataset contains a significant amount of satellite information in terms of the propagation of each wildfire per day, reporting the date of occurrence, location, and affected hectares for each wildfire. This detailed dataset containing all wildfires geo-tagged per day was merged with the daily meteorological conditions registered by the meteorological stations retrieved from the *WeatherAPI* (2019). For the purposes of our analyses, we selected the meteorological stations reporting information which are closer to each of the active wildfires. The WeatherAPI provides data on minimum, average, and maximum daily temperatures, as well as on the precipitation probability, among others. These meteorological conditions may favor or disfavor the risk of wildfire occurrence and facilitate their propagation beyond human control. We consider these meteorological conditions to be quite relevant in order to understand the concerns that users may experience when tweeting about the actual risks of propagation or initiation of new fires. This is particularly relevant for the experienced temperature and the existence of extremes in temperature. In addition, data quality variables representing the presence of particles from fires has been retrieved as well from AEMET stations.

5.3. Additional socio-economic variables

As income is a crucial variable for our analysis, data on daily personal income per province in 2017 were collected from the Eurostat dataset (Eurostat, 2021(a), GDP per capita expressed in PPP and divided by 365 days). From the same source (Eurostat, 2021(b)), data measuring the population density of each province were also collected.

6. Results

6.1. Descriptive content analysis and emotion analysis

The final Twitter dataset contains 11,248 tweets, all geo-tagged and suitable for analysis from Spain and Portugal. Looking at the words with higher frequency, and as expected, we find that "wildfire", "fire" and "Galicia" are the most repeated words in Spain, which also refers to the conditions being suffered in the neighboring country. This discourse corresponds with the fact that the region of Galicia was the most affected in Spain, and in particular, certain relatively large cities such as Vigo were in danger. Other popular words refer to messages of encouragement and support to victims, and impact of the economic losses. Furthermore, protest messages referring to the lack of rapid action by the government are also remarked. In the case of Portugal, the word "fire" is also the one that is most repeated. Additional topics referring to the damage suffered by the entire country (Portugal), the sadness of the situation, and the discouragement reflected by profane language is also reflected, as well as discontent with government and politicians at the time of the event.

The NRC lexicon (Mohammad and Turney, 2013) was used for analyzing the correspondence between the words employed and evoked emotions. Out of the eight emotions evaluated (anger, anticipation, disgust, fear, joy, sadness, surprise and trust), the most predominant emotions are negative, with fear being the most common, with 31.7% of the total words evoking it, while sadness is present in 10.7% of the words, anger in 10.5%, and disgust 6.9%.

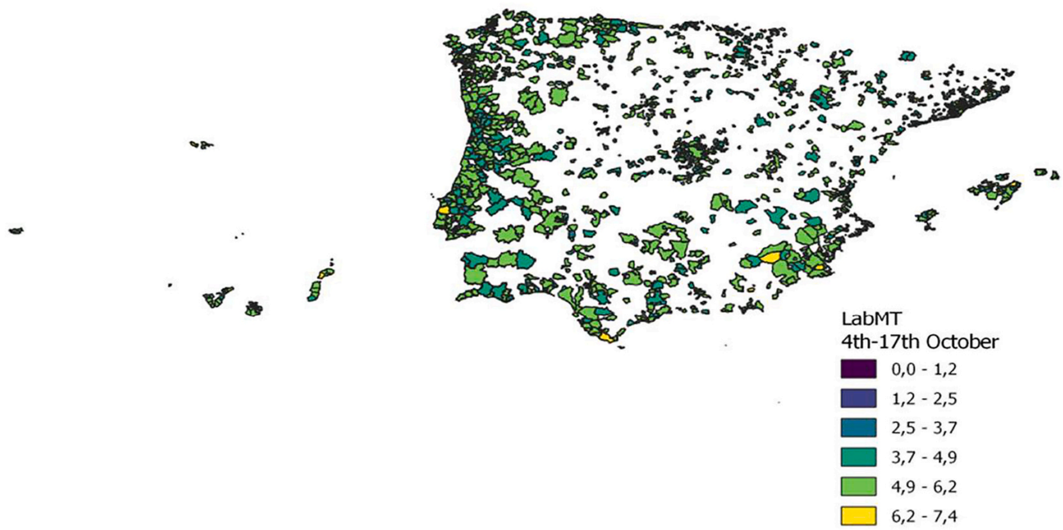
6.2. Regression analysis: The impact of wildfires on sentiments and happiness

The main aim of our work is to assess how the closeness to wildfires affected the individual's wellbeing, as expressed by their sentiment. For the entire month, the average expressed sentiment obtained from the LabMT library had a score of 5.12, revealing on average an ambivalent type of sentiment. For VADER and AFINN, these scores were -0.180 and -1.114, respectively, lying on the negative side. Our findings were consistent across the different lexica, with correlations of mean sentiment scores that were fairly high, considering the methodological differences across libraries, corr(LabMT, AFINN) = 0.424, ($p = .000$); corr (LabMT,VADER) = 0.479, ($p = .000$); corr(AFINN, VADER) = 0.765, ($p = .000$). Maps 1–3 illustrate the geographical variations of expressed sentiment scores measured by the LabMT, AFINN, and VADER, respectively during the days of 4th–17th October, dates on which a major wave of wildfires was recorded. The average negative sentiments tend to correlate quite well with direct exposure to fire in Galicia (north-west Spain) and north Portugal, mainly. Strong negative sentiments are also reported in Madrid and Barcelona, the largest cities of Spain which count with migrant labor force from the rest of the country and abroad.

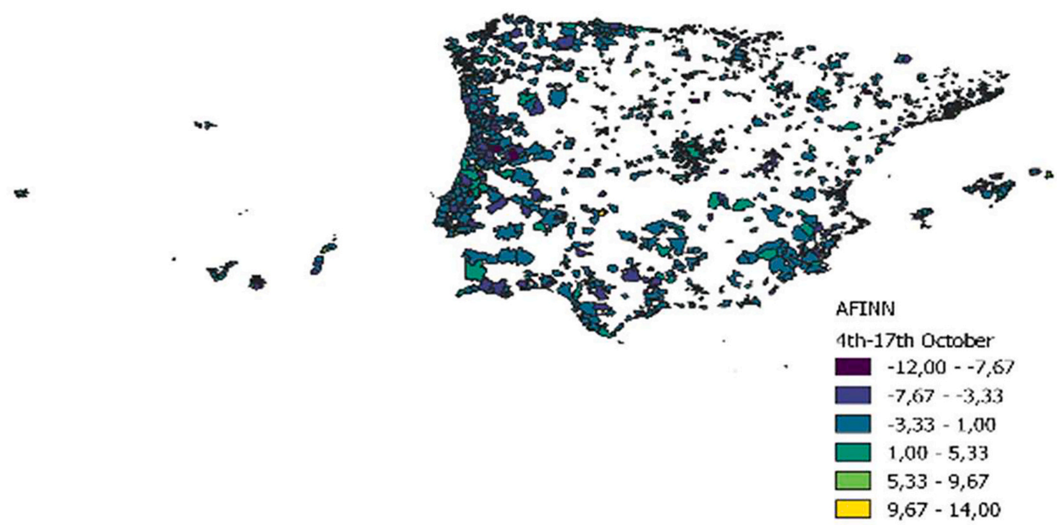
Summary statistics of relevant variables are displayed in Table 1. Based on the fact that LabMT sentiment scores range between the interval 1–9 by definition, in order to show robustness of results we employ several tobit models (See Table 2), controlling for robust errors, including temporal splines, and also exploiting the panel structure of our dataset, including fixed effects by region and random effects. Our main results show that the linear Euclidean distance from where the tweets were written to the closest active wildfires is increasing the sentiment score in a statistically significant way.[4] Other variables such as the

[3] VADER considers emojis, emoticons and symbols.

[4] Additional empirical specifications have been tested for non-linear distant effects, carrying non-significant coefficients, and not increasing the R-adjusted.



Map 1. Sentiments for wildfires. Color gradient ranges from dark blue (very negative) to yellow (not so negative). LabMT results 4-17th October 2017



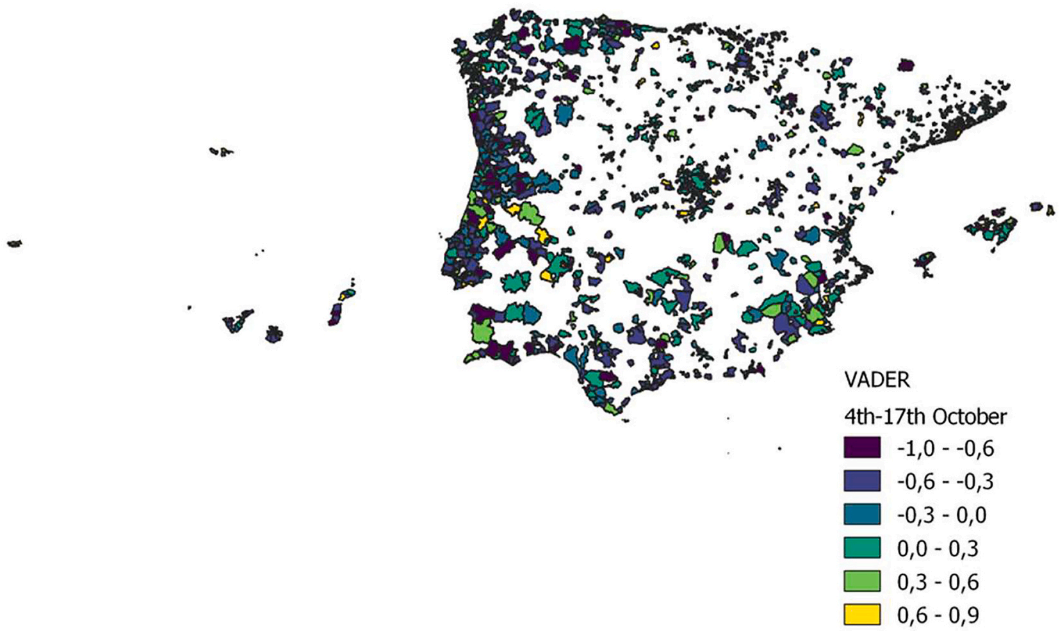
Map 2. Sentiments for wildfires. Color gradient ranges from dark blue (very negative) to yellow (not so negative). AFINN results 4-17th October 2017.

personal income is statistically significant and positive, as expected. In the case of the LabMT sentiment regression, the political protest variable indicates that the tweets referring to the political governance of the hazard decrease the sentiment score. A negative effect is also associated with tweets coming from highly populated areas, and those registered in the worst period of arson fires containing the worst days of the hazards (13th–17th October), or those occurring in some weekends. Gender differences are also found statistically significant in several LabMT regressions, with female users being more positive or increasing the average expressed sentiment. In terms of the meteorological variables, we find that precipitation probability increases the expressed sentiment in many empirical regressions. This may be understood as a sense of relief from the situation of risk generated by the wildfires. As for the temperature, we find that average temperature increases sentiment scores, while the weekend effect is negative. This may be justified by the fact that during the weekend users are more conscious about the existence of wildfires and the potential limitations this may have on outdoor activities, among others. Finally, tweets written from Portugal carry a positive effect on the overall expressed sentiment with respect to those from Spain. Such result may be explained by stronger resilience as well as adaptation processes due to earlier and more destructive wildfires occurring in Portugal the same year.[5]

6.3. Computation of welfare effects

Econometric results can be translated into welfare effects related to the negative emotional impact of wildfires, computing the CV as denoted in eq. (5) and displayed in Table 3. For this particular purpose, we directly consider the estimated coefficients for the distance and income variables in the LabMT. Using the results from Table 2, we find that welfare increases between 1.49 and 3.50€/year/Kilometer of distance away to the closest active fire, depending on the empirical specification used. This is a sign of clear discomfort with closest wildfires.

[5] Lopes et al. (2014) assess the impact of the 2008 financial crisis on SWB in Europe. They find that, paradoxically, SWB increases in Portugal after the crisis aftermath. They explain this puzzling result in Portugal denoting that “not only income but social relationships count in life”.



Map 3. Sentiments for wildfires. Color gradient ranges from dark blue (very negative) to yellow (not so negative). VADER results 4-17th October 2017.

Table 1
Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Twitter data					
LabMT (sentiment score)	11,248	5.127	0.428	1.480	8.000
VADER (sentiment score)	11,248	−0.180	0.360	−0.968	0.940
AFINN (sentiment score)	10,655	−1.114	1.825	−5.000	5.000
Female (Expected)	11,248	0.384	0.486	0.000	1.000
Political content (content related to political parties or politicians)	11,248	0.061	0.239	0.000	1.000
Wildfires and meteorological variables					
Distance to fires(km)	11,248	677.177	816.797	0.000	2000.000
Temperature (Celsius)	11,248	20.412	4.811	0.820	35.360
Precipitation probability	11,248	0.037	0.149	0.000	1.000
Very good-Good air quality* (0 < PM2.5 < 20)	11,248	0.763	0.425	0.000	1.000
Moderate air quality (20 < PM2.5 < 25)	11,248	0.048	0.214	0.000	1.000
Bad/very bad air quality (25 < PM < 50)	11,248	0.187	0.323	0.000	1.000
Other socio-economic variables					
Daily GDP per capita (€)	11,248	74.78	16.010	42.192	106.027
Population density (inhabitants/Km^2)	11,248	386.008	342.080	8.700	1006.200
Portugal	11,248	0.201	0.401	0.000	1.000
Weekend-Holidays	11,248	0.368	0.482	0.000	1.000

* In the empirical analysis and due to the distribution of the original data, we model the effect of having very good/good air quality over the rest of the categories.

6.4. Robustness checks

6.4.1. Using different libraries to compute sentiment scores

For robustness purposes, we re-estimated the sentiment regressions for the tree sentiment scores (LabMT, VADER and AFINN scores), although in this particular case our specification is based on an ordered logit. Given the differences of ranges and means across the dependent variables, scores were aggregated into intervals, depending on their original distribution, from the most negative to the most positive intervals (See Table 4). Previous results obtained with the LabMT regressions are re-enforced in terms of the effects of distance on wildfires, as well as other relevant variables, such as income. We acknowledge, however, that direct comparison of coefficients across models is not plausible due to original differences in the dependent variables. To facilitate the interpretation of results, our welfare estimates are obtained for the LabMT scores only analyzed with the Tobit model.

6.4.2. Air quality

Given the high correlation between outdoor air quality and distance to wildfires, we opt for assessing the impact of air quality on expressed sentiments with a different empirical specification. According to Kelly and Fussell (2020), particulate mas (PM) is the most common pollutant from wildfires smoke, and specifically PM2.5 concentration, which has been used to represent the air quality variable in our empirical model. According to *Agencia Estatal de Meteorología, AEMET* (2022), PM2.5 concentration below 20 μg/m^3 can be considered as indicators of good air quality. Due to the structure of our original data, we consider how the change from good air quality to the rest of levels below affects the expressed sentiments. We find a positive and statistically significant relationship between good air quality and expressed sentiments, while controlling for the variables previously included, as shown in Table 5. Welfare estimates imply that individuals require between 4.43 and 6.5€/day (equivalent to 1616€-2372€/year) to be as well off for the wildfire losses in terms of air quality (See Table 6).[6]

[6] We thank a reviewer for suggesting this additional check to control for air quality effects.

Table 2
LabMT Sentiment Regressions (with distance to wildfires).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Tobit	Tobit Robust SE	Tobit Spline	Tobit	Tobit Spline	Tobit RE	Tobit Spline RE
Distance to fires (km)	0.00020***	0.00020***	0.00030***	0.00007***	0.00011***	0.00017***	0.00028***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Daily GDP per capita	0.05021***	0.05021***	0.03153***	0.06298***	0.05504***	0.05408***	0.03553***
	(0.00041)	(0.00046)	(0.00046)	(0.00076)	(0.00082)	(0.00051)	(0.00059)
Population density	−0.00172***	−0.00172***	−0.00108***	−0.00041***	−0.00035***	−0.00182***	−0.00120***
	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Expected (Female)	0.07923***	0.07923***	0.03994***	0.01089	0.00686	0.12347***	0.06463***
	(0.01393)	(0.01383)	(0.01196)	(0.00948)	(0.00927)	(0.01805)	(0.01535)
Political content	−0.04451	−0.04451*	−0.04183*	−0.08767***	−0.07936***	−0.06976**	−0.06077**
	(0.02832)	(0.02581)	(0.02428)	(0.01924)	(0.01880)	(0.03130)	(0.02908)
Temperature	0.08047***	0.08047***	0.05693***	0.01581***	0.01659***	0.06618***	0.05621***
	(0.00127)	(0.00129)	(0.00118)	(0.00107)	(0.00106)	(0.00154)	(0.00137)
Precipitation probability	0.60391***	0.60391***	0.40373***	0.14132***	0.16384***	0.49127***	0.40243***
	(0.04645)	(0.04507)	(0.04074)	(0.03203)	(0.03183)	(0.04717)	(0.04487)
Weekend-Holidays	−0.12161***	−0.12161***	−0.10611***	−0.03331***	−0.04536***	−0.09244***	−0.08478***
	(0.01501)	(0.01510)	(0.01302)	(0.01029)	(0.01018)	(0.01547)	(0.01454)
Portugal	0.92824***	0.92824***	0.57118***			1.02563***	0.65727***
	(0.01959)	(0.01991)	(0.01776)			(0.02576)	(0.02307)
Interval1 (day 1 to 8)			0.28297***		0.09708***		0.23778***
			(0.00545)		(0.00476)		(0.00613)
Interval2 (day 9 to 17)			−0.34962***		−0.12530***		−0.29348***
			(0.00834)		(0.00700)		(0.00924)
Interval3 (day 18 to 23)			0.06675***		0.03497***		0.05892***
			(0.00640)		(0.00498)		(0.00694)
Interval4 (day 24 to 31)			−0.01734*		−0.00221		−0.02127**
			(0.00969)		(0.00750)		(0.01023)
var(e.labmt)	0.50982***	0.50982***	0.37447***	0.23476***	0.22394***		
	(0.00692)	(0.00747)	(0.00509)	(0.00320)	(0.00305)		
/sigma_u						0.55300***	0.38042***
						(0.01090)	(0.01296)
/sigma_e						0.49925***	0.52044***
						(0.00822)	(0.00825)
F	63,302.020	68,090.360	59,973.020	44,680.990	41,001.610		
Prob>F	0.000	0.000	0.000	0.000	0.000		
Wald chi2						332,442.880	469,165.000
Prob>chi2						0.000	0.000
Log likelihood	−12,193.353		−10,521.044	−8016.787	−7758.9563	−9567.380	−8608.399
Log pseudolikelihood		−12,193.353					
Observations	11,248	11,248	11,248	11,248	11,248	8906	8906
Pseudo R^2	0.64119	0.64119	0.69040	0.76410	0.7717		
Regional Fixed Effects				Yes	Yes		

Standard errors are in parentheses.
***< .01, ** p< .05, *p< .1.

Table 3
Compensating Variation (in terms of distance to wildfires).

Compensated variations WTP €/km/day							
	Tobit	Tobit	Tobit	Tobit Regional Fixed effects	Tobit Spline Regional Fixed effects	Tobit	Tobit
		Robust SE	Spline			RE	Spline RE
WTP	0.0041	0.0041	0.0096	0.0011	0.0019	0.003	0.0077
Ll	0.0044	0.0044	0.0103	0.0013	0.0022	0.0034	0.0084
Ul	0.0037	0.0037	0.0089	0.0009	0.0016	0.0027	0.007
Compensated variations WTP €/km/year							
	Tobit	Tobit	Tobit	Tobit Regional Fixed Effects	Tobit Spline Regional Fixed Effects	Tobit	Tobit
		Robust SE	Spline			RE	Spline RE
WTP	1.4965	1.4965	3.504	0.3900	0.7033	1.095	2.8105
Ll	1.606	1.606	3.7595	0.4640	0.8122	1.241	3.066
Ul	1.3505	1.3505	3.2485	0.3177	0.5981	0.9855	2.555

Confidence intervals have been obtained by bootstrapping (Hole, 2007).

7. Conclusions and limitations

The economic consequences of natural hazards can be quite diverse. However, few studies have assessed the economic impact of these hazards on human wellbeing. In this paper, we take advantage of Twitter text data in order to analyze spontaneous messages related to wildfires. By using sentiment analysis, we estimate sentiment scores from spontaneous conversations, which can be understood as measurements of changes in happiness.

We employ a quite unique dataset that combines sentiment scores with wildfire proximity data and other meteorological and socio-economic conditions, including air quality, during October 2017 in

Table 4
Sentiment Ologit Regressions (with distance to wildfires).

	(1) Ologit LabMT	(2) Ologit Spline LabMT	(3) Ologit VADER	(4) Ologit Spline VADER	(5) Ologit AFFIN	(6) Ologit Spline AFFIN
Distance to fires (km)	0.00010***	0.00008***	−0.00001	−0.00001	−0.00000	−0.00001
	(0.00002)	(0.00003)	(0.00002)	(0.00003)	(0.00002)	(0.00003)
Daily GDP per capita	0.00435** *	0.00402**	0.00665***	0.00667***	0.00467**	0.00476***
	(0.00177)	(0.00178)	(0.00176)	(0.00176)	(0.00183)	(0.00183)
Population density	−0.00028***	−0.00027***	−0.00025***	−0.00024***	−0.00008	−0.00007
	(0.00008)	(0.00008)	(0.00008)	(0.00008)	(0.00009)	(0.00009)
Female (Expected)	0.01873	0.01865	−0.01150	−0.01470	0.01102	0.00636
	(0.03494)	(0.03498)	(0.03440)	(0.03443)	(0.03567)	(0.03570)
Political content	−0.37028***	−0.36338***	−0.19923***	−0.19205***	−0.23084***	−0.22376***
	(0.06794)	(0.06798)	(0.07006)	(0.07007)	(0.07404)	(0.07402)
Temperature	−0.00886**	−0.00814**	−0.01125***	−0.00821***	−0.01657***	−0.01254***
	(0.00399)	(0.00409)	(0.00391)	(0.00401)	(0.00405)	(0.00414)
Precipitation probability	0.02295	0.14490	−0.12894	−0.11752	−0.09296	−0.12119
	(0.11655)	(0.11876)	(0.11689)	(0.11920)	(0.12193)	(0.12426)
Weekend-Holidays	−0.05877	−0.08541**	−0.03746	−0.04049	−0.03491	−0.02715
	(0.03776)	(0.03824)	(0.03719)	(0.03767)	(0.03850)	(0.03897)
Portugal	−0.02301	−0.03926	−0.18982***	−0.19440***	−0.34776***	−0.35001***
	(0.05655)	(0.05671)	(0.05531)	(0.05549)	(0.05758)	(0.05775)
Interval1 (day 1 to 8)		0.04049**		0.03755**		0.01488
		(0.01944)		(0.01856)		(0.01907)
Interval2 (day 9 to 17)		−0.07647***		−0.03276		0.01447
		(0.02815)		(0.02677)		(0.02761)
Interval3 (day 18 to 23)		0.04880**		0.00011		−0.02358
		(0.01906)		(0.01856)		(0.01924)
Interval4 (day 24 to 31)		0.04540		0.01296		0.00250
		(0.02883)		(0.02805)		(0.02930)
/cut1	−1.88419***	−1.82743***	−1.48032***	−1.08853***	−1.69171***	−1.28635***
	(0.15024)	(0.19175)	(0.14727)	(0.18713)	(0.15306)	(0.19263)
/cut2	−1.05578***	−0.99937***	−0.29988**	0.09371	−0.50453***	−0.09568
	(0.14910)	(0.19092)	(0.14645)	(0.18675)	(0.15206)	(0.19210)
/cut3	0.34682**	0.40628**	0.29398**	0.68883***	0.56876***	0.98168***
	(0.14888)	(0.19078)	(0.14646)	(0.18693)	(0.15213)	(0.19243)
/cut4	1.35886***	1.42236***	1.33268***	1.72894***	2.33269***	2.74804***
	(0.14957)	(0.19131)	(0.14704)	(0.18756)	(0.15491)	(0.19483)
Log-likelihood	−17,312.493	−17,295.587	−19,638.935	−19,628.433	−16,384.288	−19,365.630
LRchi2	94.340	128.160	96.500	117.500	138.290	175.610
Prob>chi2	0.000	0.000	0.000	0.000	0.000	0.000
Observations	11,248	11,248	11,248	11,248	10,655	10,655
Pseudo R^2	0.00272	0.00369	0.00245	0.00298	0.00420	0.00534

Standard errors are in parentheses.
All intervals created considered the distribution of sentiment scores. For the LabMT dependent variable: 1, if the sentiment score ≤4.6; 2, if the sentiment score > 4.6 and ≤4.9; 3, if the sentiment score > 4.9 and ≤5.2; 4, if the sentiment score > 5.2 and ≤5.4; 5, if the sentiment score > 5.4.
Intervals created for the VADER dependent variable: 1, if the sentiment score ≤ − 0.5; 2, if the sentiment score > −0.5 and ≤ − 0.3; 3, if the sentiment score > −0.3 and ≤ − 0.2; 4, if the sentiment score > −0.2 and ≤0; 5, if the sentiment score > 0 and ≤0.3 and 6, if the sentiment score > 0.3.
Intervals created for the AFINN dependent variable: 1, if the sentiment score ≤ − 3; 2, if the sentiment score > −3 and ≤ − 2; 3, if the sentiment score > −2 and ≤ − 0.5; 4, if the sentiment score > −0.5 and ≤1.5 and 5, if the sentiment >1.5.
*** < .01, ** < .05, * < .1.

Spain and Portugal. We find that Twitter activity is highly correlated with the propagation of the biggest fires. In addition, most of these tweets reveal sentiments of fear, sadness, and political discontent that decrease the sentiment of happiness. Relevant results from our regression analysis show that distance away from the active wildfires increases sentiment scores significantly, as better air quality does. Following previous studies, we employ the sentiment scores to understand the impact the wildfires on wellbeing, estimating the compensating variation, or amount of income required to leave individuals as well off as not having fires around. We acknowledge that this is a stylized simplification that allows for the use of utility theory in this setting, and as such, results show be taken with care. Taking our sentiment estimates into a standard utility framework, the economic valuation of these impacts was calculated to be between 1.49€-3.50€/Kilometer of distance to the closest active fire, depending on the empirical specification used for the LabMT sentiment score. Ordered regressions for the three estimated sentiments (LabMT, VADER and AFINN) reinforce conclusions in qualitative terms, although a quantitative comparison of coefficients is not recommended to the differences in the original scales of the three sentiment scores. Additional specifications show that air quality is a relevant variable to measure welfare externalities caused by wildfires, computing welfare loses associated with air quality equivalent to 4.43€-6.59€/day.

The caveats and limitations of the current findings include those related to data sources, such as the fact that we cannot properly account for relevant socio-economic variables at the individual level, since Twitter data provide limited information to preserve the fundamental right of privacy of their users. The only socio-economic variables we were able to retrieve are the expected gender of the Twitter user and location of tweeting. With the location, we estimate the average income and population density. In future research, we expect to be able to account for individual effects in more detail, controlling for education and potential political orientation of tweeter users. Other limitations include the fact that our results may be somewhat biased by the fact that the social media platform Twitter is not fully universal, and users tend to be more educated and wealthier than the average population. Related to that, there may be other potential gender bias, since female are less

Table 5
LabMT Sentiment Regressions (with air quality).

	(1) Tobit	(2) Tobit Robust SE	(3) Tobit Spline	(4) Tobit Regional Fixed Effects	(5) Tobit Spline	(6) Tobit RE	(7) Tobit Spline RE
Good air quality	0.22305***	0.22305***	0.21844***	0.14597***	0.14361***	0.24056***	0.24981***
	(0.01728)	(0.01794)	(0.01545)	(0.01333)	(0.01327)	(0.01914)	(0.01814)
Daily GDP per capita	0.05024***	0.05024***	0.03344***	0.06168***	0.05526***	0.05363***	0.03774***
	(0.00047)	(0.00052)	(0.00051)	(0.00079)	(0.00084)	(0.00056)	(0.00063)
Population density	−0.00174***	−0.00174***	−0.00114***	−0.00041***	−0.00035***	−0.00182***	−0.00127***
	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Female (Expected)	0.08027***	0.08027***	0.04628***	0.01110	0.00804	0.12614***	0.07439***
	(0.01419)	(0.01411)	(0.01248)	(0.00948)	(0.00930)	(0.01834)	(0.01610)
Political content	−0.03599	−0.03599	−0.02643	−0.08716***	−0.07764***	−0.05402*	−0.04050
	(0.02884)	(0.02597)	(0.02532)	(0.01924)	(0.01887)	(0.03097)	(0.02935)
Temperature	0.07921***	0.07921***	0.06194***	0.01555***	0.01709***	0.06446***	0.05944***
	(0.00130)	(0.00132)	(0.00123)	(0.00106)	(0.00108)	(0.00155)	(0.00143)
Precipitation probability	0.72455***	0.72455***	0.60898***	0.17415***	0.21743***	0.56145***	0.55083***
	(0.04713)	(0.04541)	(0.04207)	(0.03195)	(0.03175)	(0.04744)	(0.04569)
Weekend-Holidays	−0.13481***	−0.13481***	−0.13305***	−0.03516***	−0.04876***	−0.10033***	−0.10429***
	(0.01529)	(0.01554)	(0.01357)	(0.01028)	(0.01021)	(0.01557)	(0.01488)
Portugal	0.90830***	0.90830***	0.56845***			1.00807***	0.67282***
	(0.02024)	(0.02107)	(0.01874)			(0.02641)	(0.02441)
Interval1 (day 1 to 8)			0.23054***		0.07234***		0.18231***
			(0.00545)		(0.00440)		(0.00595)
Interval2 (day 9 to 17)			−0.27400***		−0.09081***		−0.21409***
			(0.00841)		(0.00657)		(0.00904)
Interval3 (day 18 to 23)			0.09560***		0.03864***		0.07402***
			(0.00662)		(0.00498)		(0.00712)
Interval4 (day 24 to 31)			−0.06460***		−0.01142		−0.05231***
			(0.00999)		(0.00747)		(0.01039)
/var.(e.labmt)	0.52888***	0.52888***	0.40744***	0.23479***	0.22566***		
	(0.00718)	(0.00767)	(0.00554)	(0.00320)	(0.00307)		
/sigma_u						0.56973***	0.42905***
						(0.01071)	(0.01216)
/sigma_e						0.49738***	0.51657***
						(0.00813)	(0.00828)
Log likelihood	−12,396.731		−10,980.940	−8018.7291	−7801.7696	−9664.627	−8891.927
F	60,986.300	65,343.880	55,057.450	44,677.120	40,685.660		
Prob>F	0.000	0.000	0.000	0.000	0.000		
Log pseudolikelihood		−12,396.731					
Wald chi2						321,094.240	424,224.880
Prob>chi2						0.000	0.000
Observations	11,248	11,248	11,248	11,248	11,248	8906	8906
Pseudo R^2	0.63521	0.63521	0.67687	0.76404	0.77042	.z	.z

***p<0.001, ** p<0.05, *p<0.1

Table 6
Compensating Variation (in terms of air quality).

Compensated variations WTP €/day	Tobit	Tobit Robust SE	Tobit Spline	Tobit Regional effects	Tobit Spline Regional effects	Tobit RE	Tobit Spline RE
WTP	4.4397	4.4397	6.5328	2.3668	2.5986	4.4599	6.5935
ul	5.1710	5.2075	7.5602	2.8344	3.1236	5.2335	7.7000
ll	3.7064	3.6784	5.5087	1.9159	2.0952	3.7007	5.5473

Compensated variations WTP €/year	Tobit	Tobit Robust SE	Tobit Spline	Tobit Regional effects	Tobit Spline Regional effects	Tobit RE	Tobit Spline RE
WTP	1620.506	1620.506	2384.502	863.8720	948.5047	1627.885	2406.631
ul	1887.415	1900.737	2759.473	1034.5466	1140.1196	1910.227	2810.500
ll	1352.836	1342.616	2010.675	699.3073	764.7611	1350.755	2024.764

visible in Twitter than male; or topic bias,[7] since we only analyzed tweets related to this particular topic, and some individuals may be more prompt to tweet about environmental issues than others. Other potential biases may be related to the dataset used, which although relatively large for a geo-tagged dataset of a particular issue, larger datasets may allow to control for additional aspects, such as seasonal variations. We acknowledge that it would be very interesting to check the robustness of our results with additional datasets in future exercises.

However, and in spite of all these limitations, we believe that this work may be quite relevant in order to show the applicability of social media data to environmental valuation and hazard valuation impacts. In summary, we conclude that social media can be considered a complementary source of information to other traditional data sources, such as

[7] We thank a reviewer for raising this potential issue.

the direct information obtained in field work and stated preference surveys. By no means we imply that at this point, social media data can substitute the one gathered by impact assessment surveys, but it may provide additional information useful for economic analysis.

Funding

M. Loureiro acknowledges funding from Agencia Estatal de Investigación, RETOS program, grant number PID2019-111255RB-I00.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Appendix

Table A1
Examples of word average happiness scores and standard deviations for LabMT.

Word	Happiness_average	Happiness_standard_deviation
Laughter	8.50	0.9313
Happiness	8.44	0.9723
Love	8.42	11.082
Excellent	8.18	11.008
Successful	8.16	10.759
Killing	1.70	13.590
Arrested	1.64	10.053
Deaths	1.64	11.386
Died	1.56	11.980
Killed	1.56	12.316
Death	1.54	12.811
Murder	1.48	10.150
Terrorism	1.48	0.9089
Terrorist	1.30	0.9091

References

Agencia Estatal de Meteorología, AEMET. http://www.aemet.es/en/eltiempo/prediccion/composicion_quimica_atmosfera/indice_previsto, 2022 (accessed on 2/01/2022).

Ambrey, C.L., Fleming, C.M., Manning, M., 2017. Forest fire danger, life satisfaction and feelings of safety: evidence from Australia. Int. J. Wildland Fire 26, 240–248.

Bateman, I., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Özdemiroglu, E., Pearce, D., Sugden, R., Swanson, J., 2002. Economic Valuation with Stated Preference Techniques: A Manual. Edward Elgar Publishing.

Baylis, P., 2020. Temperature and temperament: evidence from twitter. J. Public Econ. 184 https://doi.org/10.1016/j.jpubeco.2020.104161.

Bhadane, C., Hard, D., Doshi, H., 2015. Sentiment analysis: measuring opinions. Procedia Comp. Sci. 45, 808–814.

Chandra, R., Krishna, A., 2021. COVID-19 sentiment analysis via deep learning during the rise of novel cases. PLoS One 16 (8), e0255615. https://doi.org/10.1371/journal.pone.0255615.

Cody, E.M., Reagan, A.J., Mitchell, L., Dodds, P.S., Danforth, C.M., 2015. Climate change sentiment on twitter: an unsolicited public opinion poll. PLoS One 10 (8), e0136092. https://doi.org/10.1371/journal.pone.0136092.

Colander, D., 2007. Edgeworth's Hedonimeter and the quest to measure utility. J. Econ. Perspect. 21 (2), 215–225.

Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A., Danforth, C.M., 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. PLoS One 6, e26752.

Dolan, P., Metcalfe, R., 2008. Valuing non-market goods: A comparison of preference-based and experience-based approaches. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.194.1714&rep=rep1&type=pdf.

EFFIS, 2017. European Forest Fire Information System. Data available upon request at. https://effis.jrc.ec.europa.eu/.

Englin, J., Loomis, J., González-Cabán, A., 2001. The dynamic path of recreational values following a forest fire: a comparative analysis of states in the intermountain west. Can. J. For. Res. 31 (10), 1837–1844.

Eurostat, 2021. European Commission. Population Density by NUTS3 Regions. https://ec.europa.eu/eurostat/databrowser/view/demo_r_d3dens/default/table?lang=en.

Frey, S., Luechinger, S., Stutzer, A., 2009. The life satisfaction approach to valuing public goods: the case of terrorism. Public Choice 138, 317–345.

GenderAPI. Available at: https://gender-api.com/en/about-us.

Ghermandi, A., Camacho-Valdez, V., Trejo-Espinosa, H., 2020. Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. Tour. Manag. 77, 104002.

Hole, A.R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ. 16, 827–840. https://doi.org/10.1002/hec.1197.

Hutto, C.J., Gilbert, E.E., 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Kelly, J. Frank, Fussell, C. Julia, 2020. Global nature of airborne particle toxicity and health effects: a focus on megacities, wildfires, dust storms and residential biomass burning. Toxicol. Res. 9 (4), 331–345. https://doi.org/10.1093/toxres/tfaa044.

Khakharia, A., Shah, V., Gupta, P., 2021. Sentiment Analysis of COVID-19 Vaccine Tweets Using Machine Learning (June 18, 2021). Available at SSRN: https://ssrn.com/abstract=3869531.

Kibanov, M., Stumme, G., Amin, I., et al., 2017. Mining social media to inform peatland fire and haze disaster management. Soc. Netw. Anal. Min. 7, 30. https://doi.org/10.1007/s13278-017-0446-1.

Kirilenko, A.P., Molodtsova, T., Stepchenkova, S.O., 2015. People as sensors: mass media and local temperature influence climate change discussion on twitter. Glob. Environ. Chang. 30, 92–100. https://doi.org/10.1016/j.gloenvcha.2014.11.003.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., Cebrian, M., 2016. Rapid assessment of disaster damage using social media activity. Sci. Adv. 2 (3), e1500779 https://doi.org/10.1126/sciadv.150077<9.

Lopes, H., Calapez, T., Porto, C., 2014. Does the macroeconomic context influence subjective well-being in Europe and Portugal? The puzzling case of the 2008 crisis. Port. J. Soc. Sci. 13 (1), 3–19. https://doi.org/10.1386/pjss.13.1.3_1.

Luechinger, S., 2009. Valuing air quality using the life satisfaction approach. Econ. J. 119, 482–515.

Luechinger, S., Raschky, P., 2009. Valuing flood disasters using the life satisfaction approach. J. Public Econ. 93 (3–4), 620–633.

MAPAMA, 2017. Los incendios forestales en España. https://www.mapa.gob.es/es/desarrollo-rural/estadisticas/iiff_2017_def_tcm30-446071.pdf.

Metcalfe, R., Powdthavee, N., Dolan, P., 2011. Destruction and distress: using a quasi-experiment to show the effects of the September 11 attacks on mental well-being in the United Kingdom. Econ. J. 121 (550), 81–103. https://doi.org/10.1111/j.1468-0297.2010.02416.x.

Mohammad, S.M., Turney, P.D., 2013. Crowdsourcing a Word–Emotion Association Lexicon. Comput. Intell. 29 (3), 436–465.

New York Times, 2017. Deadly Fires Sweep Portugal and Northern Spain. https://www.nytimes.com/2017/10/16/world/europe/portugal-spain-fires.html.

Nielsen, Finn Årup, 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In: Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big things come in small packages 718 in CEUR Workshop Proceedings 93–98. 2011 May. http://arxiv.org/abs/1103.2903.

van Praag, B.M.S., Baarsma, B.E., 2005. Using happiness surveys to value intangibles: the case of airport noise. Econ. J. 115 (500), 224–246. https://doi.org/10.1111/j.1468-0297.2004.00967.x.

Rehdanz, K., Maddison, D., 2008. Local environmental quality and life-satisfaction in Germany. Ecol. Econ. 64, 787–797.

Rehdanz, K., Welsch, H., Narita, D., Okubo, T., 2015. Well-being effects of a major natural disaster: the case of Fukushima. J. Econ. Behav. Organ. 116, 500–517.

Richardson, L.A., Champ, P.A., Loomis, J.B., 2012. The hidden cost of wildfires: economic valuation of health effects of wildfire smoke exposure in southern California. J. For. Econ. 18, 14–35.
Santamaría, L., Mihaljevic, H., 2018. Comparison and benchmark of name-to-gender inference services. PeerJ. Comp. Sci. 4, e156 https://doi.org/10.7717/peerj-cs.156.
Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A., 2020. Valuing nature-based recreation using a crowdsourced travel cost method: a comparison to onsite survey data and value transfer. Ecosystem Services 45, 101165.
Sisco, M.R., Bosetti, V., Weber, E.U., 2017. When do extreme weather events generate attention to climate change? Clim. Chang. 143 (3), 227–241. https://doi.org/10.1007/s10584-017-1984-2.
Tavra, M., Racetin, I., Peroš, J., 2021. The role of crowdsourcing and social media in crisis mapping: a case study of a wildfire reaching Croatian City of Split. Geoenviron Disasters 8, 10. https://doi.org/10.1186/s40677-021-00181-3.
Theja Bhavaraju, S.K., Beyney, C., Nicholson, C., 2019. Quantiative analysis of social media sensitivity to natural diasters. Int. J. Disaster Risk Reduction 39, 101251. https://doi.org/10.1016/j.ijdrr.2019.101251.
Turco, M., Jerez, S., Augusto, S., et al., 2019. Climate drivers of the 2017 devastating fires in Portugal. Sci. Rep. 9, 13886. https://doi.org/10.1038/s41598-019-50281-2.
Wan, X., Yang, J., Marinov, S., et al., 2021. Sentiment correlation in financial news networks and associated market movements. Sci. Rep. 11, 3062. https://doi.org/10.1038/s41598-021-82338-6.
WeatherAPI. Available at. https://www.meteomatics.com/en/api/overview/.
Welsch, H., 2006. Environment and happiness: valuation of air pollution using life satisfaction data. Ecol. Econ. 58 (4), 801–813.
Wirz, C.D., Xenos, A.M., Brossard, D., Sheufele, D., Chung, J.H., Massarani, L., 2018. Rethinking social amplification of risks: social media and Zika in three languages. Risk Anal. 38 (12), 2599–2624.