



## Public crowdsensing of heat waves by social media data

Valentina Grasso<sup>1,2</sup>, Alfonso Crisci<sup>1</sup>, Marco Morabito<sup>1</sup>, Paolo Nesi<sup>3</sup>, and Gianni Pantaleo<sup>3</sup>

<sup>1</sup>Institute of Biometeorology, Italian National Research Council, Via G. Caproni 8, Florence, Italy

<sup>2</sup>LaMMA Consortium, Via Madonna del Piano 10, Sesto Fiorentino (FI), Italy

<sup>3</sup>DISIT Lab, Distributed Systems and internet/Data Intelligence and Technologies Lab, Dep. of Information Engineering (DINFO), University of Florence, Italy Via S. Marta, Florence, Italy

Correspondence to: Valentina Grasso (v.grasso@ibimet.cnr.it)

Received: 24 January 2017 – Revised: 4 May 2017 – Accepted: 7 June 2017 – Published: 11 July 2017

**Abstract.** Investigating on society-related heat wave hazards is a global issue concerning the people health. In the last two decades, Europe experienced several severe heat wave episodes with catastrophic effects in term of human mortality (2003, 2010 and 2015). Recent climate investigations confirm that this threat will represent a key issue for the resiliency of urban communities in next decades. Several important mitigation actions (Heat-Health Action Plans) against heat hazards have been already implemented in some WHO (World Health Organization) European region member states to encourage preparedness and response to extreme heat events. Nowadays, social media (SM) offer new opportunities to indirectly measure the impact of heat waves on society. Using the crowdsensing concept, a micro-blogging platform like Twitter may be used as a distributed network of mobile sensors that react to external events by exchanging messages (tweets). This work presents a preliminary analysis of tweets related to heat waves that occurred in Italy in summer 2015. Using TwitterVigilance dashboard, developed by the University of Florence, a sample of tweets related to heat conditions was retrieved, stored and analyzed for main features. Significant associations between the daily increase in tweets and extreme temperatures were presented. The daily volume of Twitter users and messages revealed to be a valuable indicator of heat wave impact at the local level, in urban areas. Furthermore, with the help of Generalized Additive Model (GAM), the volume of tweets in certain locations has been used to estimate thresholds of local discomfort conditions. These city-specific thresholds are the result of dissimilar climatic conditions and risk cultures.

### 1 Introduction

Use of social media (SM) during emergencies to communicate timely information has become a practice in the last years. Social media have been used for disaster detection, risk prevention, communication situational awareness, and scientific knowledge. Scholars investigated the use of SM, and particularly Twitter, in different natural disaster situations: earthquakes (Yates and Paquette, 2011; Smith, 2010; Bossu et al., 2015), wild fires (Sutton et al., 2008; Merrifield and Panechar, 2012), floods (Starbird et al., 2010; Vieweg et al., 2010; Bruns and Burgess, 2014), hurricanes (Procopio and Procopio, 2007; Hughes et al., 2014).

SM enable data collection at an unprecedented scale, allowing to record public attention and reactions to events unfolding in both virtual and physical worlds deep involving

social science research (Watts, 2013). However not all social media data are effectively available for research purposes due to platform policies which limit the access to messages. A platform like Twitter, for instance, make available to researchers and analysts only a sample of the public data stream. Access is usually provided by mean of APIs (Application Programming Interfaces) but only to a sampling data set, which for Twitter is around 1 % of the current public data stream (Boyd and Crawford, 2012). Furthermore, APIs structure produces different data retrieval outcomes: Search APIs and Streaming APIs may produce different sample data sets (González-Bailón et al., 2012). Other limitations in SM use for research purposes are related to queries. By searching for messages containing selected keywords the retrieved data set is, in fact, only a sample of the whole Twitter conversations

around a topic. Nevertheless, scholars have been using social media for research purposes on a variety of domains.

Among the different fields, scholars have begun to use SM to extract information about disaster events (Kryvasheyeu et al., 2016; Preis et al., 2013). On social media we may find digital traces of a disaster which can be used to derive the strength and impact of an event. For instance, Herfort et al. (2014) used this approach to verify the link between the spatio-temporal distribution of tweets and the physical extent of floods; Kryvasheyeu et al. (2016) derived Sandy Hurricane per-capita damages by analyzing Twitter activity. Until now little attention has been reserved to the role of SM during heat-waves-related crisis (Watson and Finn, 2014). Extreme temperatures are in fact recognized as critical and this is also confirmed by the recent report produced by the U.S Global Change Research Program (Kim, 2016) which reported that heat waves revealed the highest 10-year estimates of fatalities and represented the second (after hurricanes) estimated economic damages among the main weather and climate disaster events in United States from 2004 to 2013. Heat waves are a silent killer, mostly affecting vulnerable people like the elderly (Morabito et al., 2012) and the very young (Xu et al., 2014). Heat waves can be considered a crisis as they affect the environment, the infrastructure and also the security of citizens. The impact of extreme high temperatures on mortality is particularly high in Europe, accounting for over 80% of the total heat-wave-related deaths worldwide (source EM-DAT, <http://www.emdat.be/>). From previous studies we know that the information shared on Twitter varies greatly from one crisis to another (Olteanu et al., 2015, 2014). Information contents and sources are in fact impacted by various dimensions, like the hazard type (natural hazard or human-induced), the temporal development (instantaneous – like earthquake – or progressive – like hurricane or heat wave), and the geographical spread (localized or diffused). For instance, tweets published from Governments are more diffused in crisis related to natural hazards, which are progressive and diffused. Those are the case where warning are issued.

This work presents a contribution on the use of Twitter during heat waves, a domain not fully explored so far. Most of the studies that investigated the effects of heat waves on health have used sanitary-type indicators (i.e. general and case-specific mortality, hospital admissions, emergency room visits) while the use of SM to evaluate the effect of heat waves on population is poorly studied. In this work we explore the use of SM data as an indicator of heat waves impact on urban population. We considered Tweets as a form of crowdsensing. We refer to crowdsensing as a form of participatory sensing (Burke et al., 2006) where citizens voluntarily collaborate to data collection and sharing using their devices. Twitter mining represents an indirect form of crowdsensing with no explicit engagement of people into data collection. In this case user-generated contents, like tweets, are used for a second purpose in what may be seen

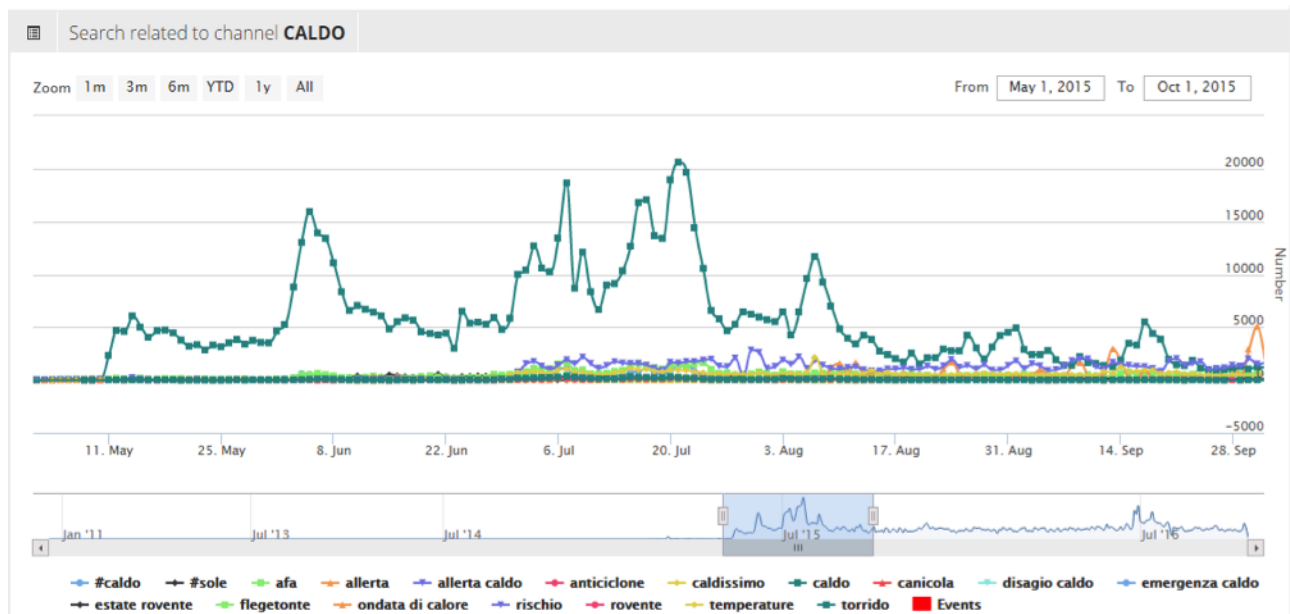
as a mobile Crowdsensing (Guo et al., 2014). The derived data may enable better events detection, timely trend and anomaly analysis and a faster response. The advantage of this approach is connected to the timely information delivered by SM and the possibility to identify the most vulnerable areas as “hot-spots” thanks to the geographical insights obtained by tweets. Heat waves have in fact their maximum impact in urban environments (McCarthy et al., 2010), characterized by the aggravating phenomenon of the urban heat island. Furthermore, the majority of the population is concentrated in cities, like also are the most vulnerable subjects (i.e. the elderly living alone). In particular, based on the Eurostat database ([http://ec.europa.eu/eurostat/statistics-explained/index.php/Population\\_and\\_social\\_conditions](http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_and_social_conditions)), while a higher proportion of the elderly population of the EU-28 countries lived in rural regions, those who were in urban regions were more likely to be living alone. The latter is a well-known risk factor for heat-related mortality (Naughton et al., 2002; Semenza et al., 1996).

The study aims to attest the reliability of SM data as a form of crowdsensing of heat waves impact in urban environment. In particular the following aspects are investigated: verify the usefulness of SM as social indicators of thermal impacts on the population; perform a data-driven estimation of city-specific thresholds of apparent temperature associated with a peak in volumes of tweets, that may be used as a quantitative risk assessment for each location.

## 2 Research design and methodology

Compared to many biometeorological studies, this work proposes an innovative way to investigate and assess the impact of heat wave episodes by using social media data. The analysis considered heat waves occurred in Italy during the very hot summer of 2015. In particular, according to data of the Institute of Atmospheric Sciences and Climate of the National Research Council (<https://www.cnr.it/en/news/6284/estate-2015-la-terza-per-temperatura-dal-1800>), the summer of 2015 was for Italy the third hottest summer since 1800 (after the 2003 and the 2012 summers). Furthermore, the positive temperature anomaly observed during July 2015 was the highest ever observed in Italy. Twitter data in reason of its publicity may be a real-time informative source to monitor impacts of extreme temperatures during heat waves. The methodological approach is based on the exploitation of crowdsensed data semantically linked to heat perception. The main research questions of this study may be summarized as follow:

- Are heat waves actually associated to social media streams semantically related to “heat”?
- Does the social media activity “follow” the spatial and temporal pattern of heat waves?



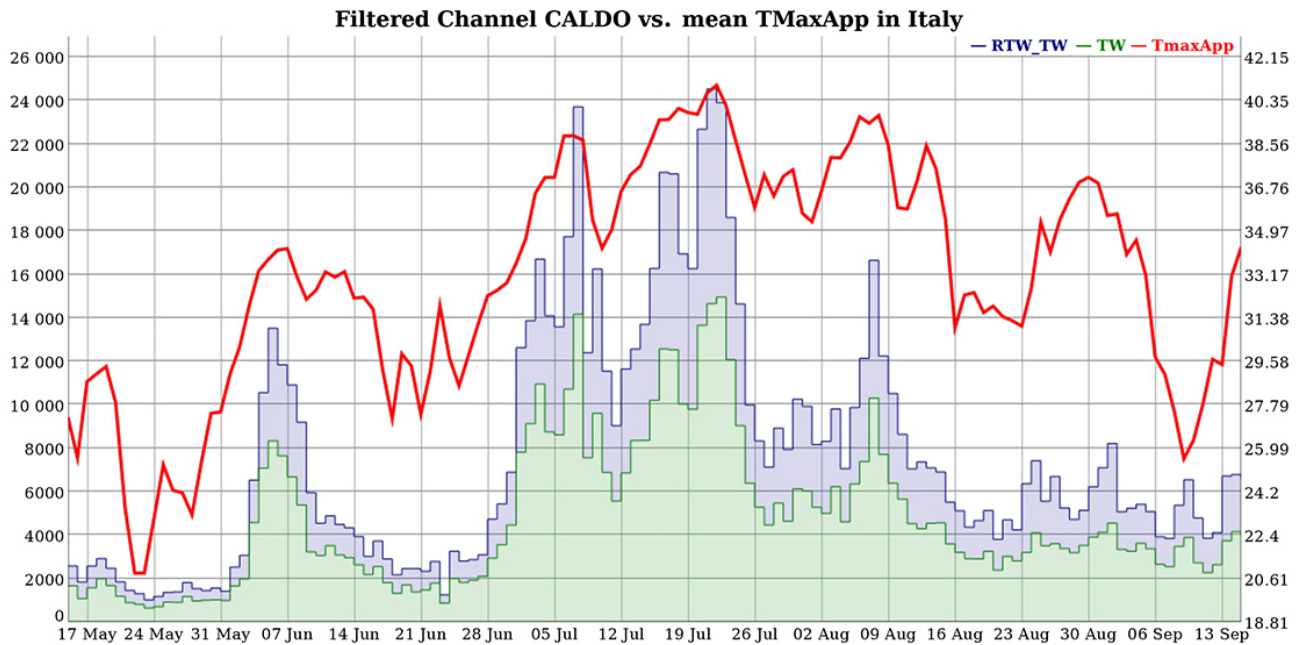
**Figure 1.** Daily volumes of tweets collected by TwitterVigilance in the channel “Heat” (caldo) during May–September 2015.

- Is a daily climatic classification of heat wave (i.e. heat wave days) able to discriminate different levels social media activity?
- Is it possible to use social media to identify local thresholds of heat discomfort?

We used the heat wave definition as provided by EuroHEAT Project (Improving Public Health Responses to extreme weather/heat waves). Following D’Ippoliti et al. (2010), heat wave is defined as a period equal to or longer than two Heat Critical day. This one is defined as a day with maximum apparent temperature exceeding the 90th percentile of the monthly distribution or the ones which minimum temperature exceeds the 90th percentile and maximum apparent temperature exceeds the median monthly climatic referenced value. In our study, the climatic monthly references were assessed for any location considered by using a 30-days filtered daily normals by using a running mean along the year span. Heat waves occurring in Italy during the summer of 2015 (from 15 May to 15 September) were assessed by using meteorological data obtained by the NOAA GSOD – Global Surface Summary of the Day (<https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>) for 21 locations corresponding to the most important Italian cities (see Appendix for location list). Daily maximum apparent temperature was assessed by using the Steadman approach (Steadman, 1984) and critical and heat wave days were calculated for the period 15 May–15 September 2015 ( $N = 124$ ). The regions of the North Tyrrhenian were the most impacted in July. On the other hand, the August’s episode involved the southern eastern areas (mainly Apulia and Calabria).

For the same time interval, tweets filtered for semantically relevant criteria were collected and stored, with the aim to compare the volume of messages and the spatial and temporal pattern of heat waves. Twitter retrieval and storage was performed with the help of the TwitterVigilance, a web-platform developed by the DISIT Lab of the University of Florence. (<http://www.disit.org/6693>). TwitterVigilance is a multi-user tool for Twitter analysis that allows to create and manage multiple parameters for Twitter API querying, to store the tweets data into channels, defined by users. The TwitterVigilance is also a dashboard for fast visualization of main analytics of each channel, as shown in Fig. 1.

To monitor Twitter activity related to extreme temperature conditions we created a “Heat” monitoring channel on TwitterVigilance platform based on a set of keywords and hashtags semantically related to heat conditions. From the total retrieved messages, the original tweets were filtered following most occurring words in Italian language like: caldo (hot), afa (very hot), canicola, sudo, sudato, sudore (sweat), caldissimo (very hot), torrido (scorching), record, allarme (alarm), emergenza (emergency), bollino (mark), bere (drink), anziani (senior), sete (thirsty, dry), umidità (humidity), anticiclone (anticyclone), disagio (discomfort), umido (umid, weat), Caronte/Flegetonte (by media naming of high-pressure systems). By choosing Italian language for queries, we excluded tweets in foreign languages published by tourists or immigrants living in the country, but in this first study we preferred to concentrate on reactions coming from native Italian citizens. To reasonably correlate daily volumes of tweets with daily temperatures we had to rely on tweets containing a geographic reference. Location estimation is



**Figure 2.** Number of Tweets and Retweets and the maximum apparent temperature (average of all stations) for Italy during the period 15 May and 15 September of 2015.

one of the main problem to approach working with Twitter data set. Even if it is possible on Twitter to communicate geo-location information by enabling Global Positioning System (GPS) when using the App, in practice only few messages include this information. Scholars indicate that the percentage of tweets that has geo-location meta-data is about 2% (Burton et al., 2012). When geo-tagging is not provided, it is possible to derive the location of a user by examining his profile description or tweets in his stream; otherwise it is possible to infer location by examining tweet content with specific algorithms based on entity recognition and Natural Language Processing. We inferred location by using this last technique. “Heat” related tweets were partitioned into city/regional streams through the geographical key-terms linked to the locations considered; both the “city name” and “region name” (see Appendix for full locations list). For each local SM data subset, main Twitter metrics were computed as to be compared with relative thermal data. Metrics used were: daily number of tweets (which includes both native tweets and retweets), daily number of native tweets, daily number of retweets, daily number of unique users. Data are summarized in Table 1. The significance of social media metrics vs temperatures association ( $T^\circ$  and Apparent  $T^\circ$  max) was tested by using a linear correlation scheme. To test if SM metrics are significantly different during heat waves a  $t$ -Student test ( $t$ -test) was performed by using daily heat wave status as stratified factor. The data set of retrieved tweets has to be considered as a sample of the whole Twitter conversations, in Italian, mentioning heat during the monitored period.

**Table 1.** Main features of the data set.

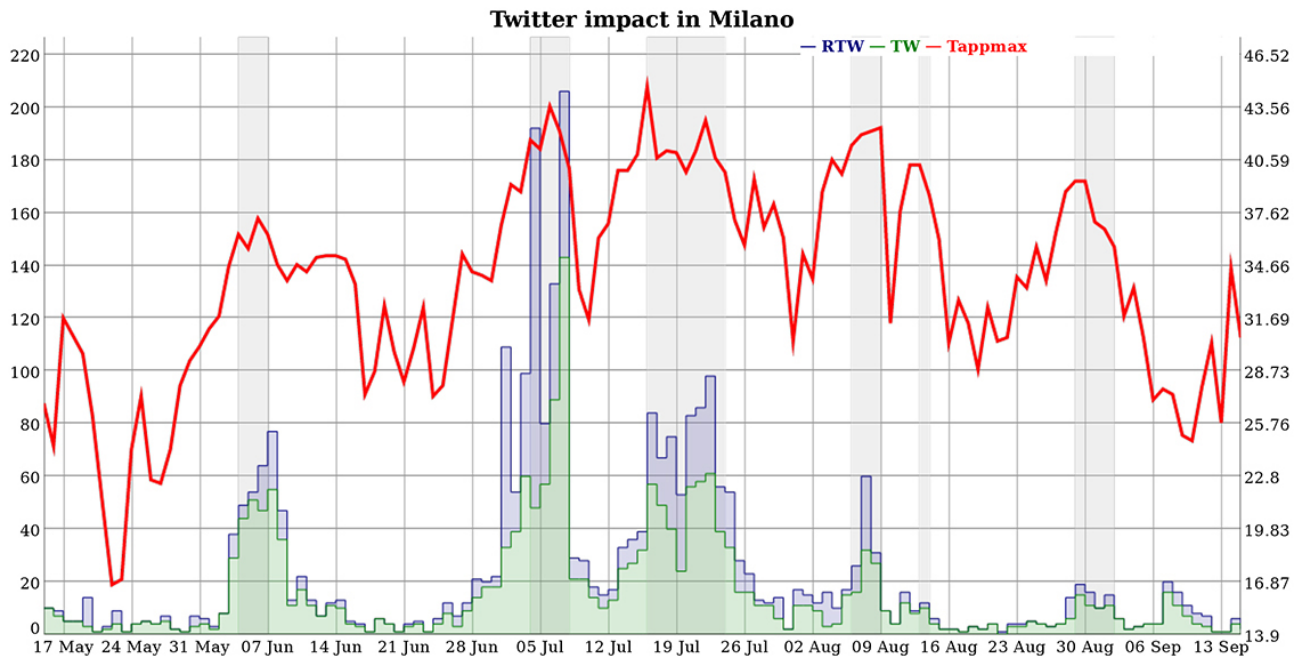
Areas*	Tweets	Ntw	RT	Users
Data set	940 123	585 286	354 837	233 553
Lombardy	2944	1949	995	2152
Lazio	2705	1834	871	2060
Sicily	1785	874	911	1449
Tuscany	1662	1098	564	760
Liguria	1320	910	410	519
Piedmont	1145	698	447	758
Apulia	934	473	461	716
Calabria	916	228	688	859
Sardinia	648	457	191	452
Emilia R.	606	461	145	433
Veneto	589	463	126	392
Umbria	353	247	106	144
Abruzzo	240	181	59	138
Trentino AA	221	171	50	157
Friuli VG	218	134	84	140
Marche	106	77	29	73
Molise	61	53	8	37

\* Areas are the results of messages containing the name of the local cities considered and/or the name of the region they belong to (see Appendix for full location list). Ntw = Native tweets; RT = Retweets.

### 3 Results and discussion

From 15 May to 15 September 2015 through the Twitter-Vigilance channel and by applying semantic filters explained above, we collected 940 123 tweets sent by 233 553 unique





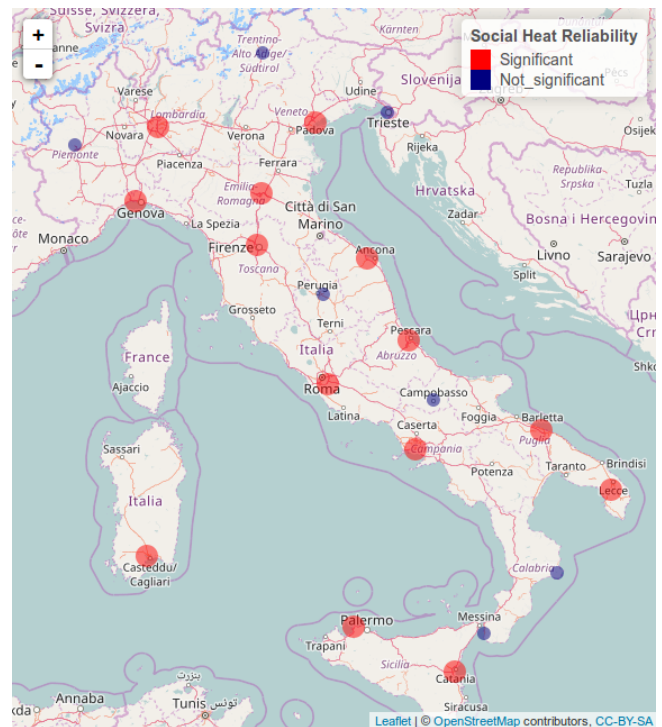
**Figure 3.** The pattern of the daily number of tweets (green histogram) and retweets (blue histogram) and the daily maximum apparent temperatures (red line) and the heat-waves periods (transparent blue bars) for Milan.

users. The data set is composed of 585 286 native tweets (62 % of the whole data set), tweets originally published by users, and 354 837 retweets (38 %), tweets written by other users that are reposted, typically starting with *RT: @user-name*. Table 1 shows main metrics for the whole data set and for main areas considered. The areas with greater number of tweets and users are those more impacted by heat wave events. Some regions impacted by heat wave episodes do not show however similar volumes of tweets, reasonably because they are less populated rural areas where Twitter users are few and few are the tweets mentioning heat related contents.

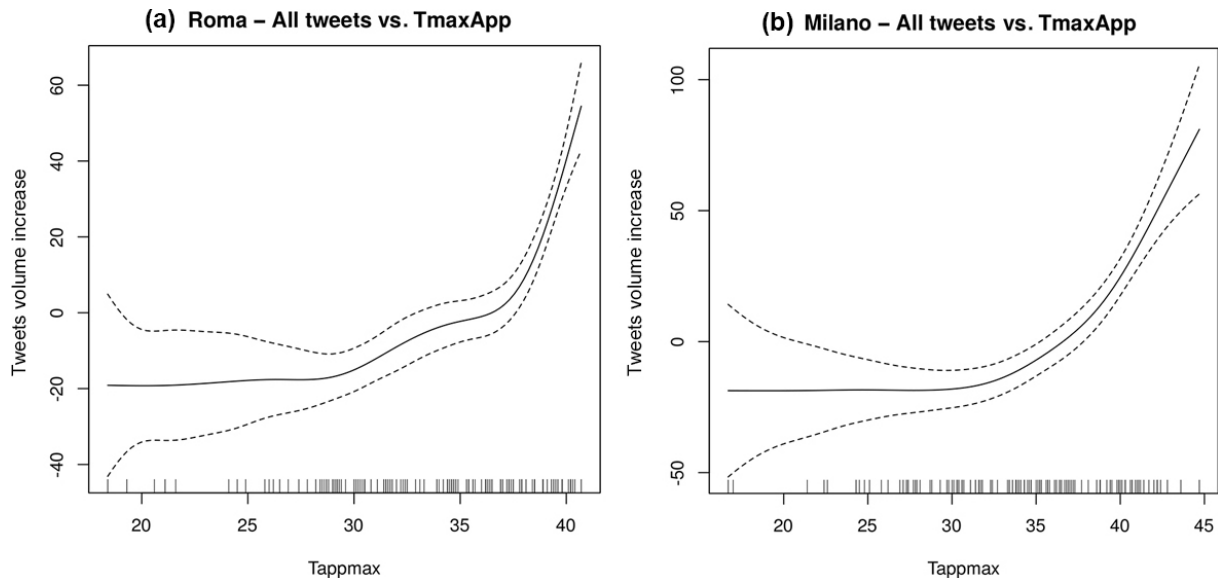
Daily Twitter data were compared to weather data and a clear association was observed, at national level, between the number of Tweets and Retweets and the maximum apparent temperatures (the average of all stations), as it is shown in Fig. 2.

In particular, five main heat waves episodes were detected in Italy during the summer of 2015. The statistical association among daily volumes of tweets and apparent temperatures (and the difference of the ones during heat wave days) is stronger in cities with high population density, like Milan, than in other cities. As an example Fig. 3 shows the graph of the patterns of the daily number of tweets and retweets related to heat (green and blue bars), the daily maximum apparent temperatures (red line) and the heat wave occurrences in Milan. In this case the association among tweets volumes and heat temperature is highly statistically significant.

Except for Northeast, Piedmont, Apennines areas and Calabria, the great part of locations and regions revealed signif-



**Figure 4.** Social heat reliability in several Italian cities. Red spot: significant association; blue spot: not significant association.



**Figure 5.** GAM Predicted daily Tweets volumes by Maximum Apparent Temperatures as a function of the  $x$  axis in Milan (b) and Rome (a). Values are centered on 0.

icant associations between Twitter metrics and the increase in apparent temperatures (Fig. 4). Red spot in the map represents locations where the association among tweets daily volumes and maximum temperatures is significant, blue spot are locations with no significant association. Association is reliable in locations where tweets are sufficiently numerous. In more rural regions with lower population density Twitter users are fewer. The small volume of tweets related to these locations is not suitable for a significant association.

Other interesting results were obtained by using Generalized Additive Models (GAM) analyses (Hastie and Tibshirani, 1990) in different cities. GAMs are a non-linear extension of generalized linear models that perform well in this kind of predictive analysis. By modeling the daily volume of tweets mentioning heat, it is possible to estimate city-specific thresholds of local maximum apparent temperature. These thresholds correspond to breakpoints in the range, where the number of tweets suddenly increases. In southern cities the growth of heat-related twitter activity shows higher thermal threshold ( $37^{\circ}\text{C}$  in Rome) than in northern ones ( $33^{\circ}\text{C}$  in Milan) (Fig. 5). These thresholds correspond to the levels of maximum apparent temperature where people begin to perceive heat discomfort. City-specific thresholds imply a local heat risk perception that could be linked to urban climatic factors of the city investigated. These thresholds are also a product of local risk cultures and are, in fact, higher in locations with warmer climate or closer to the sea, like Rome, respect to Milan (see Fig. 5). In this respect, tweets prove to be a valuable form of crowdsensing to detect heat waves impact in urban areas. When Apparent Maximum Temperature reaches a city-specific threshold, people start to comment or complain about the heat. Knowing local thresholds for main

urban areas may be important to improve preparedness measures at the regional and local level and reduce local vulnerability. Lexical analysis of tweets contents could potentially offer further understanding about specific impacts, local discomforts or health symptoms.

#### 4 Conclusions

The analysis of summer 2015 heat wave episodes through social media analytics showed that sudden growth of Twitter activity related to heat conditions seems to identify correctly the peak days of heat wave episodes and also allows a geographical identification of high impact situations. During heat related crisis this may facilitate response efforts at local level, especially if more geolocated tweets are available. Following a crowdsensing approach, daily volumes of tweets related to heat may thus be considered as a further indicator to assess heat waves impact at national level and even local. Compared to the most used sanitary-type indicators (i.e. general and case-specific mortality, hospital admissions, emergency room visits), Twitter volumes may be obtained faster and easier, with the help of data retrieval and storage platforms like TwitterVigilance. In this first contribution, authors did not considered the nature and contents of tweets, which could instead provide further information and feedbacks about local perceptions and impacts of heat waves. This could be the subject of a following research work.

Responsive tools to monitor the impacts of heat waves are few. Social media analytic shows a twofold usefulness for emergency/disaster management. Firstly, SM activity metrics give a quantitative and reliable feedback from large urban areas where the heat risk is higher and where vulnerable people

generally live. Secondly, SM monitoring may also provide at the same time an alternative communication channel to reach urban population and increase situational awareness on heat related risks.

**Code availability.** The code of the work is available at the public repository [https://github.com/alfcrisci/EMS\\_heat\\_twitter/tree/master/code](https://github.com/alfcrisci/EMS_heat_twitter/tree/master/code) (Crisci et al., 2016a).

**Data availability.** All data used in the work are available in RDS file format [https://github.com/alfcrisci/EMS\\_heat\\_twitter/tree/master/data](https://github.com/alfcrisci/EMS_heat_twitter/tree/master/data) (Crisci et al., 2016b).

**Appendix A: Location filtering**

Regions name used for filtering: Marche, Puglia, Emilia Romagna, Trentino Alto Adige, Sardegna, Molise, Calabria, Toscana, Liguria, Lombardia, Sicilia, Umbria, Abruzzo, Lazio Piemonte, Friuli Venezia Giulia, Veneto.

Locations name used for filtering: Ancona, Bari, Bologna, Bolzano, Cagliari, Campobasso, Catania, Crotone, Firenze, Genova, Lecce, Milano, Napoli, Palermo, Perugia, Pescara, Reggio Calabria, Roma, Torino, Trieste, Venezia.



**Competing interests.** The authors have declared that no competing interests exist.

**Disclaimer.** The article reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

**Special issue statement.** This article is part of the special issue "16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC)". It is a result of the 16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC), Trieste, Italy, 12–16 September 2016.

**Acknowledgements.** The article represents the results of the CARISMAND project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 653748.

Edited by: Antti Makela

Reviewed by: Laure Fallou, Hayley Watson, and one anonymous referee

## References

- Bossu, R., Mazet-Roux, G., Roussel, F., Steed, R., and Etivant, C.: The EMSC tools used to detect and diagnose the impact of global earthquakes from direct and indirect eyewitnesses' contributions, in: Information Systems for Crisis Response and Management (ISCRAM) 2015 Conference Proceedings, 24–27 May 2015, Kristiansand, Norway, 24–27, 2015.
- Boyd, D. and Crawford, K.: Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, *Information, Communication & Society*, 15, 662–679, 2012.
- Bruns, A. and Burgess, J.: Crisis communication in natural disasters: The Queensland floods and Christchurch earthquakes, *Twitter and society*, 89, 373–384, 2014.
- Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., and Srivastava, M. B.: Participatory Sensing, in: World Sensor Web Workshop, ACM Sensys, ACM, 31 October–3 November, Boulder, Colorado, USA, 2006.
- Burton, S. H., Tanner, K. W., Giraud-Carrier, C. G., West, J. H., and Barnes, M. D.: "Right time, right place" health communication on Twitter: value and accuracy of location information, *J. Med. Internet Res.*, 14, e156, <https://doi.org/10.2196/jmir.2121>, 2012.
- Crisci, A., Grasso, V., Morabito, M., Nesi, P., and Pantaleo, G.: EMS\_heat\_twitter/code/, Repository of "Public crowdsensing of heat waves by social media data" – EMS conference contribution, CNR – IBIMET, available at: [https://github.com/alfcrisci/EMS\\_heat\\_twitter/tree/master/code](https://github.com/alfcrisci/EMS_heat_twitter/tree/master/code) (last access: 2 January 2017), 2016a.
- Crisci, A., Grasso, V., Morabito, M., Nesi, P., and Pantaleo, G.: EMS\_heat\_twitter/data/, Repository of "Public crowdsensing of heat waves by social media data" – EMS conference contribution, CNR – IBIMET, available at: [https://github.com/alfcrisci/EMS\\_heat\\_twitter/tree/master/data](https://github.com/alfcrisci/EMS_heat_twitter/tree/master/data) (last access: 2 January 2017), 2016b.
- D'Ippoliti, D., Michelozzi, P., Marino, C., de'Donato, F., Menne, B., Katsouyanni, K., Kirchmayer, U., Analitis, A., Medina-Ramón, M., Paldy, A., Atkinson, R., Kovats, S., Bisanti, L., Schneider, A., Lefranc, A., Iñiguez, C., and Perucci, C. A.: The impact of heat waves on mortality in 9 European cities: results from the EuroHEAT project, *Environ. Health*, 9, <https://doi.org/10.1186/1476-069X-9-37>, 2010.
- González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., and Moreno, Y.: Assessing the bias in communication networks sampled from Twitter, *Social Networks*, 38, 16–27, 2012.
- Guo, B., Yu, Z., Zhou, X., and Zhang, D.: From participatory sensing to mobile crowd sensing, in: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on, 593–598, IEEE, 24–28 March 2014, Budapest, Hungary, 2014.
- Hastie, T. J. and Tibshirani, R. J.: Generalized additive models, vol. 43, CRC Press, London, UK, 1990.
- Herfort, B., Albuquerque, J. P. d., Schelhorn, S.-J., and Zipf, A.: Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the River Elbe Flood in June 2013, in: International Conference on Information Systems for Crisis Response and Management, 11, 18–21 May 2014, The Pennsylvania State University, University Park, Pennsylvania, USA, 2014.
- Hughes, A. L., St Denis, L. A., Palen, L., and Anderson, K. M.: Online public communications by police & fire services during the 2012 Hurricane Sandy, in: Proceedings of the 32nd annual ACM conference on Human factors in computing systems, 26 April–1 May 2014, Toronto, Canada, 1505–1514, ACM, 2014.
- Kim, E. J.: The impacts of climate change on human health in the United States: A scientific assessment, by uS Global change research program, *J. Am. Plann. Assoc.*, 82, 418–419, 2016.
- Kryvasheyev, Y., Chen, H., Obradovich, N., Moro, E., Van Henteryck, P., Fowler, J., and Cebrian, M.: Rapid assessment of disaster damage using social media activity, *Science advances*, 2, e1500779, <https://doi.org/10.1126/sciadv.1500779>, 2016.
- McCarthy, M. P., Best, M. J., and Betts, R. A.: Climate change in cities due to global warming and urban effects, *Geophys. Res. Lett.*, 37, L09705, <https://doi.org/10.1029/2010GL042845>, 2010.
- Merrifield, N. and Panechar, M.: Uncertainty Reduction Strategies via Twitter: The 2011 wildfire threat to Los Alamos National Laboratory, in: Proceedings from AEJMC Annual Conference, 9–12 August 2012, 2012. Chicago, IL, USA, 2012.
- Morabito, M., Profili, F., Crisci, A., Francesconi, P., Gensini, G. F., and Orlandini, S.: Heat-related mortality in the Florentine area (Italy) before and after the exceptional 2003 heat wave in Europe: an improved public health response?, *Int. J. Biometeorol.*, 56, 801–810, 2012.
- Naughton, M. P., Henderson, A., Mirabelli, M. C., Kaiser, R., Wilhelm, J. L., Kieszak, S. M., Rubin, C. H., and McGeehin, M. A.: Heat-related mortality during a 1999 heat wave in Chicago, *Am. J. Prev. Med.*, 22, 221–227, 2002.
- Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S.: CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises, in: ICWSM, 8th International Conference on Weblogs and Social Media, ICWSM 2014, 1–4 June 2014, Ann Arbor, USA, Code 114731, 2014.

- Olteanu, A., Vieweg, S., and Castillo, C.: What to expect when the unexpected happens: Social media communications across crises, in: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 14–18 March 2015, Vancouver, Canada, 994–1009, ACM, 2015.
- Preis, T., Moat, H. S., Bishop, S., Treleaven, P., and Stanley, H. E.: Quantifying the digital traces of Hurricane Sandy on Flickr, *Sci. Rep.*, 3, 3141, <https://doi.org/10.1038/srep03141>, 2013.
- Procopio, C. H. and Procopio, S. T.: Do you know what it means to miss New Orleans? Internet communication, geographic community, and social capital in crisis, *J. Appl. Commun. Res.*, 35, 67–87, 2007.
- Semenza, J. C., Rubin, C. H., Falter, K. H., Selanikio, J. D., Flanders, W. D., Howe, H. L., and Wilhelm, J. L.: Heat-related deaths during the July 1995 heat wave in Chicago, *New Engl. J. Med.*, 335, 84–90, 1996.
- Smith, B. G.: Socially distributing public relations: Twitter, Haiti, and interactivity in social media, *Public Relat. Rev.*, 36, 329–335, 2010.
- Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S.: Chatter on the red: what hazards threat reveals about the social life of microblogged information, in: Proceedings of the 2010 ACM conference on Computer supported cooperative work, 6–10 February 2010, Savannah, Georgia, USA, 241–250, ACM, 2010.
- Steadman, R. G.: A universal scale of apparent temperature, *J. Clim. Appl. Meteorol.*, 23, 1674–1687, 1984.
- Sutton, J., Palen, L., and Shklovski, I.: Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires, in: Proceedings of the 5th International IS-CRAM Conference, 4–7 May 2008, Washington, DC, USA, 624–632, 2008.
- Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L.: Microblogging during two natural hazards events: what twitter may contribute to situational awareness, in: Proceedings of the SIGCHI conference on human factors in computing systems, 10–15 April 2010, Atlanta, GA, USA, 1079–1088, ACM, 2010.
- Watson, H. and Finn, R. L.: Social media and the 2013 UK heat wave: opportunities and challenges for future events, in: Proceedings of the 11th International ISCRAM Conference, 18–21 May 2014, University Park, Pennsylvania, USA, 2014.
- Watts, D. J.: The virtual lab, in: Proceedings of the sixth ACM international conference on Web search and data mining, 4–8 February 2013, Rome, Italy, 1–2, ACM, 2013.
- Xu, Z., Sheffield, P. E., Su, H., Wang, X., Bi, Y., and Tong, S.: The impact of heat waves on children's health: a systematic review, *Int. J. Biometeorol.*, 58, 239–247, 2014.
- Yates, D. and Paquette, S.: Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake, *Int. J. Inform. Manage.*, 31, 6–13, 2011.