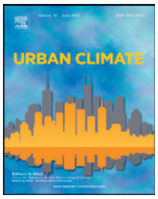
Urban Climate 19 (2017) 170–191

Contents lists available at ScienceDirect

Urban Climate

journal homepage: http://www.elsevier.com/locate/uclim



Crowdsourcing air temperature from citizen weather stations for urban climate research

Fred Meier [*], Daniel Fenner, Tom Grassmann, Marco Otto, Dieter Scherer

Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Germany
Rothenburgstr. 12, D-12165 Berlin, Germany

ARTICLE INFO

Article history:
Received 5 July 2016
Received in revised form 5 December 2016
Accepted 22 January 2017

Keywords:
Urban climate observations
Crowdsourcing air temperature
Data quality assessment
Urban heat island
Netatmo weather station
Berlin

ABSTRACT

Provision of accurate air temperature data in urban environments with high spatial and temporal resolution over long time periods remains a challenge in atmospheric research. Crowdsourcing, i.e., collection of atmospheric data from non-traditional sources like citizen weather stations (CWS), is an alternative and cost-efficient method for exploration and monitoring of urban climates. This study examines the suitability of crowdsourced air temperature (T_{crowd}) measurements from CWS by comparing T_{crowd} from up to 1500 stations with reference air temperature (T_{ref}) in Berlin and surroundings for a period of twelve months (Jan–Dec 2015).

Comprehensive quality assessment of T_{crowd} reveals that erroneous metadata, failure of data collection, and unsuitable exposure of sensors lead to a reduction of data availability by 53%. Spatially aggregated raw data of T_{crowd} already provide a robust estimate of hourly and daily urban air temperature in the study area. Quality-checked T_{crowd} observations show spatio-temporal characteristics of the urban heat island in Berlin with higher spatial variability than T_{ref} in built-up areas. Spatial density of T_{crowd} in Berlin exceeds that of the reference monitoring network by far. However, rigorous data quality assessment is the key challenge in order to fully benefit from this novel data set for urban climate research.

---

[*] Corresponding author.
E-mail address: fred.meier@tu-berlin.de (F. Meier)

http://dx.doi.org/10.1016/j.uclim.2017.01.006
2212-0955

1. Introduction

Provision of atmospheric data from observational networks at high spatial resolution and over long time periods remains a challenge in urban climate research, since at least one of these requirements is commonly not fulfilled (Grimmond, 2006; Muller et al., 2013). Classical meteorological observational networks are designed for detection of synoptic atmospheric conditions, and thus are rarely suitable for city-specific and intra-urban analyses. The development and progression of low-cost and quality-checked sensors with wireless data transmission provides new opportunities for environmental monitoring networks (e.g. Kumar et al., 2015; Whiteman et al., 2000; Young et al., 2014,). This generation of sensors fosters the implementation of high-resolution measurement networks in cities like the Helsinki Testbed (Koskinen et al., 2011) or the Birmingham Urban Climate Laboratory (BUCL) (Chapman et al., 2015; Warren et al. 2016). However, long-term operation of such dense networks is hardly possible given the maintenance costs for the large amount of devices (Chapman et al., 2015).

Beside these comprehensive urban meteorological networks, citizens as data providers offer huge potentials, especially in urban areas due to high population density (e.g. Bell et al., 2013; Castell et al., 2015; Steeneveld et al., 2011; Wolters and Brandsma, 2012). In the field of ecology the concept of involving citizens in science is not new (Dickinson et al., 2010). This concept relies on active participation of citizens to contribute to research. In recent years, a number of efforts have also been made concerning atmospheric applications, e.g. mapping of atmospheric aerosols with smartphones (Snik et al., 2014) or involving citizens in observational meteorological networks such as “CoCoRaHS” (Community Collaborative Rain, Hail and Snow Network, http://www.cocorahs.org) or the UK citizen rainfall network (Illingworth et al., 2014).

Another approach to acquire huge amounts of data is the concept of crowdsourcing, defined by Dickinson et al. (2010) as “getting an undefined public to do work, usually directed by designated individuals or professionals”. A recent comprehensive review expanded this definition, stating that crowdsourcing includes the collection of atmospheric data from public sensors connected to the internet (Muller et al., 2015). For instance, Overeem et al. (2013) took battery-temperature records from smartphones to derive urban air temperatures by using data from the Android application OpenSignal for smartphones. Mass and Madaus (2014) exploited air-pressure measurements from another smartphone application called pressureNET to simulate an active convection event in the United States of America. Other web-based projects collect data from citizen weather stations (CWS), e.g. the Weather Observations Website (WOW, http://wow.metoffice.gov.uk/), the Citizen Weather Observer Program (CWOP, http://wxqa.com) or Weather Underground (WU, https://www.wunderground.com/). More examples of current atmosphere, weather, and climate-related crowdsourcing

Table 1
Data quality levels, criteria for data filtering, potential error sources for crowdsourced air temperature (T_{crowd}) measurements, and data availability at each level.

Quality level	Description criteria for data filtering	Potential error sources	Percent of raw data
A0	Crowdsourced air temperature (T_{crowd}) raw data with correct timestamp	Netatmo API and server limits	100.0
A1	Netatmo stations with valid metadata (latitude, longitude)	User-specific operating error	97.9
A2	80% hourly data per day	Intermittent failure of wireless network, loss of battery power, server failure	91.7
A3	80% daily data per month	Intermittent failure of wireless network, loss of battery power, server failure	70.1
B	Indoor station filter, monthly average and standard deviation of daily minimum air temperature (TN)	User-specific installation error (misuse), netatmo outdoor module set up indoors	59.7
C1	Systematic radiative error filter, positive and significant correlation between global radiation and air temperature difference ($T_{crowd\_ID} - T_{ref}$)	Netatmo outdoor module set up in a sunlit location (no radiation shield)	52.0
C2	Single value radiative error filter, flagging day-time values when air temperature difference ($T_{crowd\_ID} - T_{ref}$) > 3 $*$ SD in T_{ref}	At times the netatmo outdoor module received direct short wave radiation	47.3
D	Outlier filter based on spatial average of $T_{crowd} \pm 3 * SD$ in T_{crowd}	netatmo outdoor module temporarily moved, other measurement errors	47.1