

Crowdsourcing Urban Air Temperatures through Smartphone Battery Temperatures in São Paulo, Brazil

A. M. DROSTE,^a J. J. PAPE,^{b,c} A. OVEREEM,^{c,d} H. LEIJNSE,^c G. J. STEENEVELD,^a A. J. VAN DELDEN,^b
AND R. UIJLENHOET^d

^a *Meteorology and Air Quality Section, Wageningen University, Wageningen, Netherlands*

^b *Institute for Marine and Atmospheric Research (IMAU), Utrecht University, Utrecht, Netherlands*

^c *Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*

^d *Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, Netherlands*

(Manuscript received 3 August 2016, in final form 27 June 2017)

ABSTRACT

Crowdsourcing as a method to obtain and apply vast datasets is rapidly becoming prominent in meteorology, especially for urban areas where routine weather observations are scarce. Previous studies showed that smartphone battery temperature readings can be used to estimate the daily and citywide air temperature via a direct heat transfer model. This work extends model estimates by studying smaller temporal and spatial scales. The study finds the number of battery readings influences the accuracy of temperature retrievals. Optimal results are achieved for 700 or more retrievals. An extensive dataset of over 10 million battery temperature readings for estimating hourly and daily air temperatures is available for São Paulo, Brazil. The air temperature estimates are validated with measurements from a WMO station, an Urban Flux Network site, and data from seven citizen weather stations. Daily temperature estimates are good (coefficient of determination ρ^2 of 86%), and the study shows they improve by optimizing model parameters for neighborhood scales ($<1 \text{ km}^2$) as categorized in local climate zones (LCZs). Temperature differences between LCZs can be distinguished from smartphone battery temperatures. When validating the model for hourly temperature estimates, the model requires a diurnally varying parameter function in the heat transfer model rather than one fixed value for the entire day. The results show the potential of large crowdsourced datasets in meteorological studies, and the value of smartphones as a measuring platform when routine observations are lacking.

1. Introduction

The need for high-resolution urban meteorological measurements is ever increasing. Numerical weather prediction models for cities are improving continuously, which require more accurate measurements in both time and space for data assimilation (Ronda et al. 2017). Ongoing and projected global urbanization (United Nations 2012) makes a thorough understanding of the urban atmosphere vital for urban planning, as well as for reliable forecasts of air quality, energy demand, and heat stress. The urban heat island (UHI)—that is, the difference in canopy air temperature between the rural background and the urban core—has been widely studied (e.g., Oke 1982; Arnfield 2003; Steeneveld et al. 2011; Heusinkveld et al. 2014). Cities experience enhanced radiation uptake during the day as a result of their lower albedo and

high heat storage capacity. Because of the slow nocturnal heat release from the urban fabric to the atmosphere, cities cool down more slowly than their surroundings, which creates the UHI. This may amount to 8 K on hot and calm summer days (Oke 1982). The UHI effect can exacerbate the degree of heat stress experienced by residents (Reid et al. 2009), which is projected to increase as a result of the combination of climate change and global urbanization (United Nations 2012; Miralles et al. 2014). Moreover, Hajat and Kosatky (2010) show that mortality increases with 2% per 1°C increase in high temperature. Oleson et al. (2015) and Molenaar et al. (2016) project a drastic increase of future heat stress days caused by climate change for the United States and Canada, and the Netherlands, respectively. Both studies show that future heat stress is amplified in urban areas, underlining the need for knowledge of the temperature within the urban fabric. To understand these developments, we require urban temperature observations.

Corresponding author: Arjan Droste, arjan.droste@wur.nl

DOI: 10.1175/JTECH-D-16-0150.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Traditional measurements in urban areas are scarce, and are usually organized as intensive measurement campaigns (Heusinkveld et al. 2014) or cover just a small area (Kotthaus et al. 2012). This data scarcity can be harmful for, for example, megacities in developing countries, where knowledge on urban temperature is crucial for mitigating urban heat and maintaining residents' health. Part of this data scarcity can be overcome by crowdsourcing: utilizing data that are routinely collected by residents or public sensors, and transferred over the Internet (Muller et al. 2015; Warren et al. 2016), most notably by smartphones. Examples in the atmospheric sciences include the Meteorological Phenomena Identification Near the Ground (mPING) app, where users can share information about precipitation (Elmore et al. 2014); the Spectropolarimeter for Planetary Exploration (SPEX) for iPhone (iSPEX) smartphone add-on, which allows users to measure optical thickness (Snik et al. 2014); estimating rain employing microwave links from cellular telecommunication networks (Overeem et al. 2013a); mapping forest fires by using voluntary observations sent by smartphones (Sosko and Dalyot 2015); and using the built-in pressure sensor in many smartphone models to improve surface pressure forecasts (Mass and Madaus 2014). Also, crowdsourced data from citizen weather stations uploaded to, for example, Weather Underground (Wunderground) and the Weather Observation Website (WOW) project (MetOffice 2011) have proven to be valuable in urban research (e.g., Steeneveld et al. 2011; Bell et al. 2013; Warren et al. 2016; Meier et al. 2017). Using these stations asks for strict control of the quality of station and data, in terms of site setup, measurement accuracy, and data gaps. A thorough overview of crowdsourcing projects in atmospheric sciences is given by Muller et al. (2015).

An innovative way of estimating urban air temperatures from smartphones was presented by Overeem et al. (2013b, henceforth O13). Using the OpenSignal application, O13 employ 6-month datasets of smartphone battery temperature readings from eight cities (including São Paulo, Brazil), with on average 844 selected battery temperature readings per city per day (1383 per day for São Paulo alone). They use a straightforward heat transfer model between phone, human body, and air temperature T_{air} (Fig. 1, left) to translate the temperatures of smartphone batteries into a daily averaged, city-averaged air temperature. These daily temperature estimates (T_{est}) are shown to correspond well with measurements taken in the respective cities.

This paper builds upon the study of O13. Here, we employ a much longer (2 yr) and denser (12×10^3 readings per day) dataset for São Paulo. This study

explores the potential of the O13 heat transfer model at refined spatial and temporal scales. The current massively extended availability of battery temperature readings per day facilitates hourly air temperature estimates that have not been possible in previous research. We also investigate whether the model performance for daily average temperatures improves when applied only to selected neighborhoods with their characteristic morphology, that is, the so-called local climate zones (LCZs) (Stewart and Oke 2012). By using validation data obtained from both certified sources (WMO; Urban Flux Network, <http://ibis.geog.ubc.ca/urbanflux/index.html>) and crowdsourced weather stations (Wunderground and Netatmo), we provide a more robust representation of the actual city temperature on both daily and hourly time scales. Section 2 deals with the background of the heat transfer model and the urban heat island, data and methodology are discussed in section 3, results are discussed in section 4, the discussion follows in section 5, and we end with conclusions and perspectives in section 6.

2. Background

a. Heat transfer model

O13 showed that smartphone battery temperatures can be used to obtain a daily average urban T_{air} , using a linear heat transfer model (Fig. 1, left). In the O13 model, the phone battery temperature T_p ($^{\circ}\text{C}$) is regulated by the environmental air temperature T_e ($^{\circ}\text{C}$), the human body temperature T_b ($^{\circ}\text{C}$), and the thermal energy generated by the phone P_p (W) [Eq. (S-2) in O13],

$$T_e = \left(1 + \frac{k_b}{k_e}\right) T_p - \left(\frac{k_b}{k_e} T_b + \frac{P_p}{k_e}\right), \quad (1)$$

where the coefficients k_b and k_e ($\text{W}^{\circ}\text{C}^{-1}$) are determined by the thermal insulation between body and phone, and between phone and environment, respectively [Eq. (S-2) in O13]. Assuming independence among values of T_p , k_e , k_b , P_p , and T_b over the set of measurements, and equilibrium between P_p and the heat flow to the body and environment, leads to [Eq. (S-8) in O13]

$$T_e = m_j (T_p - T_0) + T_0 + \varepsilon. \quad (2)$$

Here, m_j is the average of $(1 + k_b/k_e)$ for a set of observations for city j , ε is a random error, and T_0 is interpreted as the human body temperature [T_b in Eq. (1)], plus a constant, under the assumption that the heat transfer from a phone to the environment is approximately zero. The O13 study found $T_0 = 39^{\circ}\text{C}$ for its eight cities

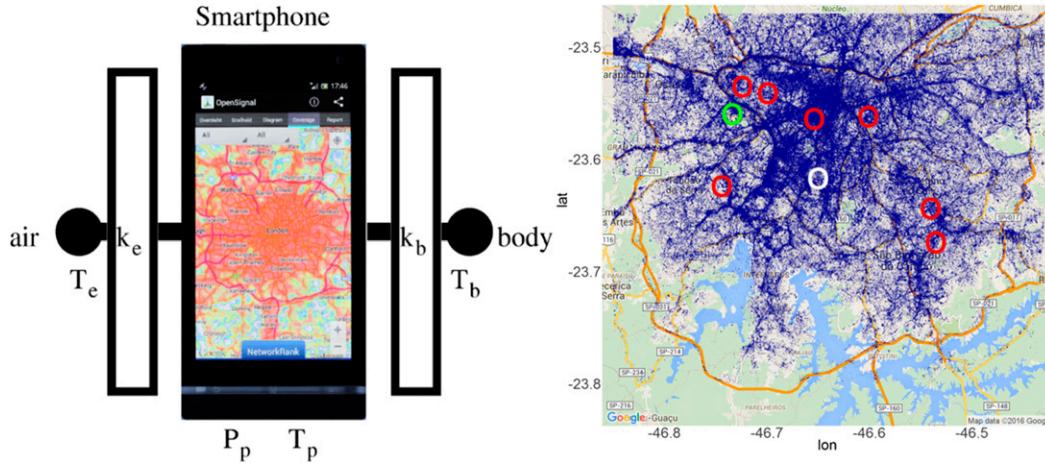


FIG. 1. (left) Conceptual diagram of the heat transfer model (from Overeem et al. 2013b). The $T_{e,p,b}$ is the temperature of the environment (e), phone (p), and body (b); P_p represents the thermal energy produced by the phone (or power); $k_{e,b}$ represents heat transfer rate between phone and environment (e) and body and phone (b). (right) Readings of smartphone battery temperature (in 2013, after selections) taken in São Paulo. Symbols show locations of the hobby meteorology stations (red circles), the FluxNet station (green circle), and WMO station Congonhas (white circle). For the exact coordinates of the stations consult Table 1.

and calibrates m_j separately per city. Both constants are calibrated over the entire dataset (i.e., a constant value for both m_j and T_0). Note that the heat transfer model initially uses daily (and hourly later in this paper) and spatially averaged battery temperatures, rather than instantaneous battery readings. For the full derivation, we refer to the supporting information of O13.

b. UHI and LCZs

The UHI is usually defined as the difference in canopy air temperatures between urban and rural sites. Urban areas differ from their rural surroundings by the high prevalence of impervious surface and buildings, and little vegetation. Building materials have a high heat capacity, storing radiative energy during the day for subsequent slow nocturnal release. Additionally, the low sky-view factor induces efficient heat trapping inside the urban canopy (Oke 1982). These effects cause the city to cool more slowly at night than the countryside, where energy is released much faster by virtue of the high sky-view factor and low heat capacity of vegetation. This creates the UHI, which peaks a few hours after sunset, when rural air temperatures have dropped and urban air temperatures can still be high.

Defining the UHI can be highly subjective: a clear definition of *urban* and *rural* is lacking (Stewart and Oke 2012). Many UHI studies lack proper metadata, making comparisons between cases difficult (Stewart 2011; Stewart and Oke 2012). Defining the UHI as a temperature difference between LCZs can increase objectivity. The LCZ framework classifies land use into 10 urban

and 7 rural zones, each with its distinct surface properties (e.g., impervious fraction, vegetation cover) and building properties (e.g., building height, aspect ratio). The UHI can thereby also be defined as the difference in temperature between a rural and an urban LCZ, or even between two urban LCZs. In this study we define the UHI as a difference in canopy air temperature between two urban LCZs (section 4b).

3. Data and methodology

a. Smartphone battery temperature data

The study region is São Paulo, which is located just south of the Tropic of Capricorn, at roughly 23.55°S, 46.63°W, at 760 m MSL. São Paulo is characterized by a subtropical maritime climate with mild dry winters and humid summers. The study area is confined to a rectangle around the city center, between 23.47° and 23.80°S, and 46.43° and 46.85°W (Fig. 1, right).

The battery temperatures are obtained from the OpenSignal app, a smartphone application that measures network signal strength from available providers. This app also logs T_p from the temperature sensor present in smartphone batteries. A T_p reading is taken when 1) the phone is being plugged into or removed from the power source and 2) when the phone is turned on or off. The selection procedure in this study follows that of O13. Only those readings made 1) at the time the phone is being plugged into the power source or 2) when the phone is turned on or off and the battery is discharging are considered. To avoid spurious data in the

analyses, an additional selection removes those battery temperatures outside the range between 10° and 47°C, since these readings are likely to be erroneous (because of, e.g., battery charging or intensive processor use), as battery temperature values are typically around 30°C (O13). The battery temperature dataset covers 1 January 2013 up to 31 December 2014. During this period an average of 16×10^3 battery readings per day are left after filtering, though this number is significantly lower ($\approx 1 \times 10^3 \text{ day}^{-1}$) at the start of 2013 and rises to as much as $40 \times 10^3 \text{ day}^{-1}$ for several months in 2014.

b. Weather station air temperature data

Three main sources of weather station T_{air} data are employed for calibration and validation, that is, WMO station Congon (WMO station 837800), the 17-m-tall urban FluxNet tower of the Micrometeorology Group of the University of São Paulo (IAG-USP), and a set of seven citizen weather stations. Congonhas is located at an airport, south of the city center, in the middle of a built-up environment (see Table 1 for station metadata, including classification into LCZs). The WMO data fully cover 2013 and 2014, with very few hours missing (less than 3 h per month) and with 1.5-m T_{air} (°C) measured at the full hour, available as rounded integers. The FluxNet T_{air} measurements are taken every 5 min and subsequently averaged into hourly values around the hour, at 0.1°C accuracy.

The data from the citizen stations are freely available for download from the Netatmo and Wunderground platforms (www.netatmo.com and www.wunderground.com, respectively), where weather enthusiasts can share their station data. First, we selected only stations with fewer than 100 missing days per year. A day is considered as missing if it contains less than 21 h of data. Very few stations meet these criteria in 2013 but seven stations remain in 2014, one of which also has a sufficient record length in 2013 (see Table 1). Measurement accuracy is variable between brands of weather stations, since the more expensive stations tend to measure at a higher degree of accuracy, for example, as a result of better radiation shielding and sensor quality (Bell et al. 2015). Typically, the better citizen stations have temperature measurement errors during daytime of around 0.5°C (Steenefeld et al. 2011; Bell et al. 2015).

Since the instrument placement and setup of these citizen stations are not bound to strict rules, we have applied a series of filters, to ensure quality, accounting for the recommendations made by Stewart (2011). Data entries with sudden large temperature jumps ($>2^\circ\text{C}$ increments between two consecutive hourly measurements) that are not confirmed in either the WMO or the FluxNet site data are removed. The temporal resolution of the measurements varies between the stations but lies

mainly between 5- and 10-min intervals. For comparison to the other stations, we have averaged the measurements to an hourly mean temperature. Past studies (e.g., Steeneveld et al. 2011; Bell et al. 2013; Bell et al. 2015; Meier et al. 2017) have demonstrated the value of these citizen data to good effect.

c. Calibration and validation

To create independent calibration and validation sets, data from 2013 are designated to the calibration set and data from 2014 to the validation set. Both the WMO station and the urban FluxNet station are fully active throughout these years and are used for both calibration and validation (Table 1). The majority of the citizen stations have only a sufficient number of measurement days in 2014 and will therefore be used for validation purposes only. Hence, the model is calibrated and validated against the best possible representation of the average urban T_{air} , rather than just one fixed station, in order to ensure the most robust results. The original validation method in O13 may suffer from autocorrelation between calibration and validation datasets, since the authors alternately assign days to the calibration and validation sets. To avoid autocorrelation problems, this study uses a statistically independent calibration and validation set, to ensure that positive model outcomes are not artificial.

The number of selected battery readings for 2013 totals nearly 3×10^6 readings; the battery dataset for 2014 reaches on average 24×10^3 selected readings per day for 8.8×10^6 readings in total. Battery readings are averaged into hourly and daily values. Days with fewer than 200 readings (i.e., six days in 2013, none in 2014) are excluded from the analysis. For the hourly analysis, July 2013 and July 2014 are set as calibration and validation datasets, respectively. All days in July 2013 and 2014 have more than 200 measurements per day.

The T_0 parameter [Eq. (2)] has been determined by O13 as an average over eight cities, rather than separately for each city under its consideration. In this work, the T_0 parameter is optimized for São Paulo using a least squares approach, based on the 1-yr calibration (2013) dataset of battery readings (section 4a). The value of m_j [Eq. (2)] is likewise determined, separately for the daily and the hourly calibration datasets. Parameter T_0 can be interpreted as the approximate human body temperature, which is not expected to fluctuate, whereas m_j represents a ratio of insulation coefficients. Factors influencing insulation (such as clothing) will be more variable over time. Therefore, m_j is calibrated separately for the analysis of hourly temperatures, resulting in two calibration datasets—one for the daily dataset and one for the hourly dataset—which are used to train the model.

TABLE 1. Metadata of the stations used in São Paulo. LCZs, distinction between balcony and garden stations, and the street aspect ratio (Height/Width) have been visually determined by using Google Streetview. Greenness fraction has been determined in a radius of 250 m around the stations using Google Earth. Note that begin and end of the record indicate the length of the record at our disposal: “active” means the station is still running as of 31 Nov 2015. Citizen stations have missing days indicated for only 2014, since they are not used in the calibration process. Netatmo is the name of stations and is the brand name, i.e., hardware name (Netatmo).

Station code/name	Begin/end	No. missing days in 2014	Lat (°S), Lon (°W)	Local Climate Zone of the station	Citizen station hardware	H/W; balcony/garden station	Greenness (%)
Congonhas (WMO 837800)	1 Jan 2013/active	0 (0 in 2013)	23.617, 46.650	3		— (airfield)	10
FluxNet station at USP	1 Jan 2013/31 Oct 2014	3 (13 in 2013)	23.560, 46.734	6		Rooftop. H/W ≈ 0.3	40
ISPSOPAU3	2 Sep 2013/20 May 2015	4	23.624, 46.744	4	DIY Arduino Funnel input/output (Fio), BMP085, DHT22	Balcony. H/W ≈ 1–2	14
ISPSANTO2	5 Aug 2013/active	91	23.674, 46.535	3	Ambient Weather WS-2080	Garden. H/W ≈ 0.5	8
ISOPAUOLO61	3 Mar 2014/active	63	23.562, 46.601	2	AcuRite Pro Weather Station	Garden. H/W ≈ 1–1.5	4
ISAOPAU11	1 Jan 2013/active	0	23.541, 46.699	4	Davis Vantage Pro2	Balcony. H/W up to 3	12
ISOPAUOLO5	1 Jan 2013/5 Dec 2014	44	23.564, 46.654	4	HOB0 U30	Balcony. H/W up to 3	4
Aifenas	7 Apr 2014/31 Mar 2015	96	23.644, 46.540	3	Netatmo	Garden. H/W ≈ 0.5–1	6
Casinha	14 Sep 2013/31 Mar 2015	0	23.535, 46.724	3	Netatmo	Garden. H/W ≈ 0.5–1	12

Since the dataset used for analysis is extensive, we can determine the minimum number of battery readings \bar{N} needed for a stable model result. Employing random sampling (in space and time), \bar{N} measurements are selected for every day and averaged into one daily value. This procedure is repeated 100 times per chosen value of \bar{N} to capture the mean battery temperature as accurately as possible, so every day has 100 mean battery temperatures. Each of these temperatures is validated against the city average air temperature (section 4a).

d. Daily air temperature modeling for a single neighborhood

A point of interest is the role of the environment on the T_p reading. An analysis of model performance as function of the distance between battery reading and validation station yielded no significant relation (not shown). Instead, we study the influence of the urban fabric on the environment, as measured by battery temperatures. Muller et al. (2015) write that “the utility of smartphones for higher resolution UHI analysis . . . is still to be explored.” To this end we utilize the LCZ classification for São Paulo, which was constructed using a GIS algorithm (Mills et al. 2015) and is freely available. The location of each battery reading is coupled to the corresponding location on the LCZ map (Fig. 2); the battery readings are subsequently grouped by LCZ and are used to validate the heat transfer model per LCZ. São Paulo mainly consists of low-rise buildings: LCZ3 in the center (compact low-rise buildings), a wide spread of LCZ6 (open low-rise buildings) closer to the city border, and several clusters of LCZ8 (large low-rise buildings) (Fig. 2).

e. Hourly air temperature estimation

For determining hourly T_{est} , we use July 2013 as calibration data and July 2014 as validation data. Model parameter m_j is calibrated to the diurnal temperature course in July. Term T_0 is set at the optimal value for the entire year, determined using the methods described in section 3c. For this analysis, July is the preferred month because of high data availability, and because July is one of the driest and cooler months, limiting possible effects of data distortion as a result of weather conditions (e.g., more people staying inside during precipitation events). Additional data selections, such as selections on LCZs and smartphone series, are not feasible with the hourly averaged data because of the strong reduction in available measurements, especially during nighttime. Figure 3 shows the availability of smartphone readings against time of day (UTC). Around 0800 UTC (0500 LT) the number of measurements is at its minimum, at less than 10% of the daytime data density. Removing these data will strongly reduce the applicability of the

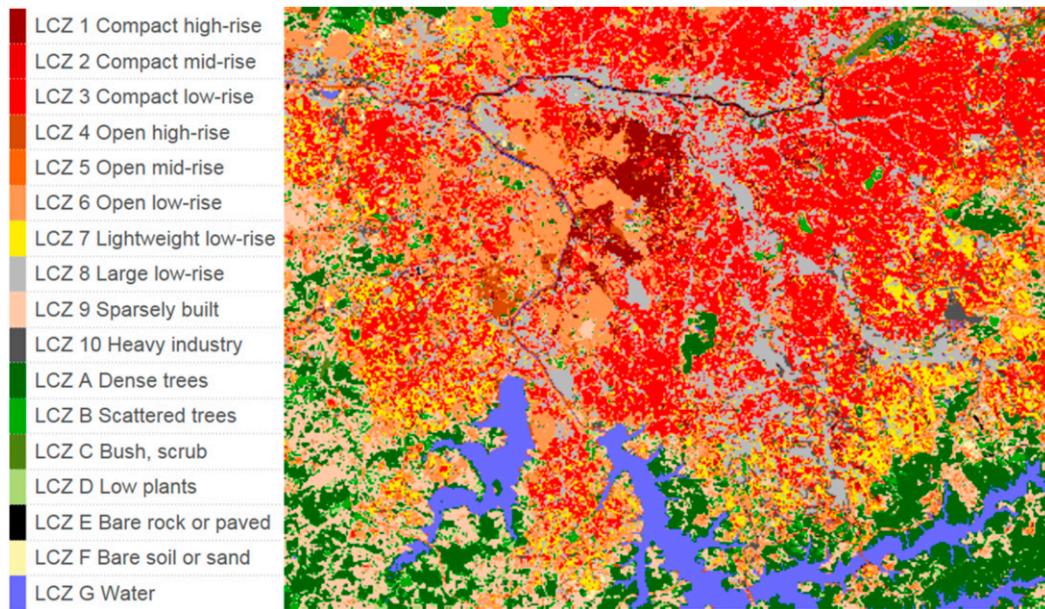


FIG. 2. LCZ in the research area (São Paulo) at a resolution of 120 m^2 . The city is mainly made up of: LCZ3, compact low-rise buildings (22%); LCZ6, open low-rise buildings (15%); and LCZ8, large low-rise buildings (14%). [Image and legend are obtained from Geopedia, the World Urban Database and Access Portal Tools (WUDAPT) for visualizing LCZ data (<http://geopedia.world/>).]

dataset; however, excluding the nighttime hours will lead to an unreliable calibration of m_j and to missing hours in the resulting validation. In addition, we will explore the effect of using 24 hourly m_j constants, to better capture T_{air} variation. By this methodology the average diurnal variability of human behavior (different clothing, being inside/outside, etc.) in July will be accounted for through m_j .

4. Results

a. Estimation of daily air temperatures

Figure 4 shows a validation of time series of daily T_{est} against observed average city T_{air} , computed as the average of the various temperature measurements available (WMO, FluxNet, and the citizen weather stations). In general T_{est} compares very well with the observed air temperature, as was also concluded by O13. The analysis uses optimized values of m_j and T_0 : optimizing T_0 for São Paulo only slightly changes its value in comparison to the standard value in O13 (from 39° to 39.8°C). The coefficient of determination ρ^2 is 0.87, with a mean error (ME; or bias) of -0.53°C . This bias is largest in January and February 2014. The quantity T_{est} is consistently up to 2°C lower than the actual measured temperature. We hypothesize that this is related to the number of battery measurements available in the calibration data. The number of measurements per day in

the period January–May is roughly 12 times lower than in the rest of the year ($\approx 1 \times 10^3$ vs $\approx 12 \times 10^3$). This could affect the calibration, since the months with the highest temperature peaks are underrepresented in the

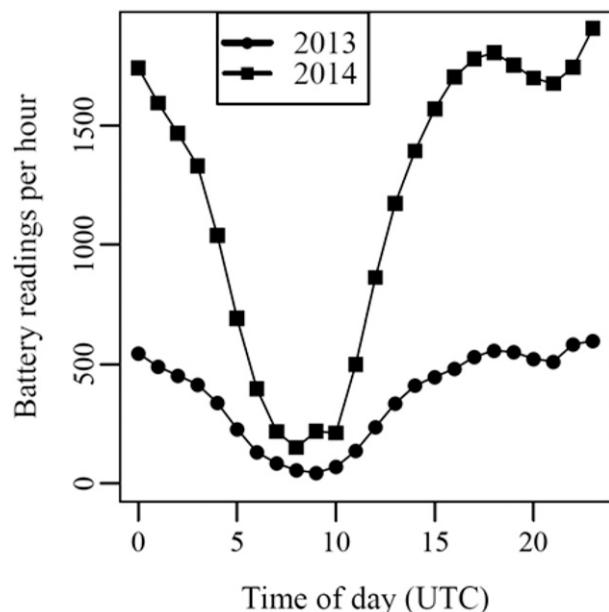


FIG. 3. Number of available battery readings after selection in July 2013 (circles) and July 2014 (squares) per hour as function of time of day. São Paulo is in time zone UTC $- 3 \text{ h}$ (summertime).

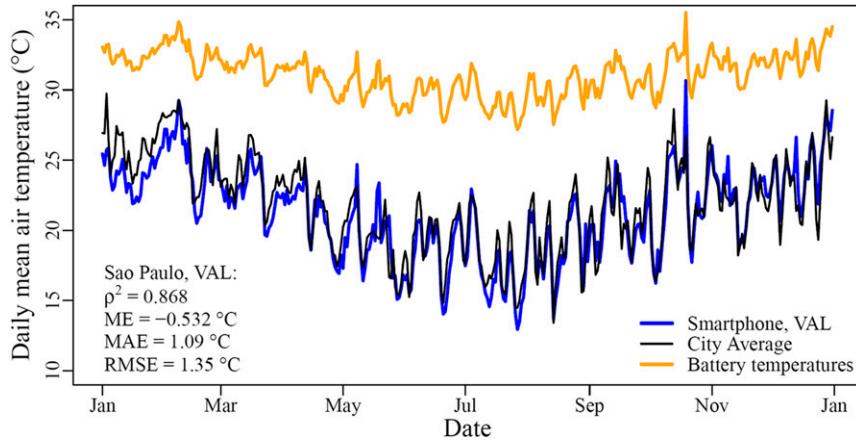


FIG. 4. Average daily air temperature in São Paulo during 2014 from T_{est} (blue line) validated against city average (WMO, FluxNet, and hobby stations) air temperature observations (black line) and daily averaged battery temperatures used as input (orange line).

model calibration (fewer measurements are available). The model seems to perform well for temperatures in the middle of the range; however, for temperatures close to the upper and lower limits, the model response underestimates the amplitude. When solely WMO data are used for calibration, the results deteriorate as a result of the coarser resolution (1°C) of the WMO data. Interestingly, calibrating m_j for separate seasons does not improve the performance (not shown), which indicates that variability in the (daily averaged) heat transfer is not very strong over the year. Though São Paulo experiences seasonal variation in temperature,

daily average temperature variability is smaller than, for instance, that of continental climates.

Figure 5 shows the model performance (ρ^2 and RMSE) as a function of \bar{N} used per day. It appears that above ≈ 700 measurements, the performance quality converges to a constant value. Apparently adding more data does not raise the quality beyond a certain threshold, but rather opens up more options for detailed analyses. This justifies stratifying the large dataset at our disposal into subsets for individual LCZs and even into hourly time intervals. The number of measurements left in these selections should still produce reliable results.

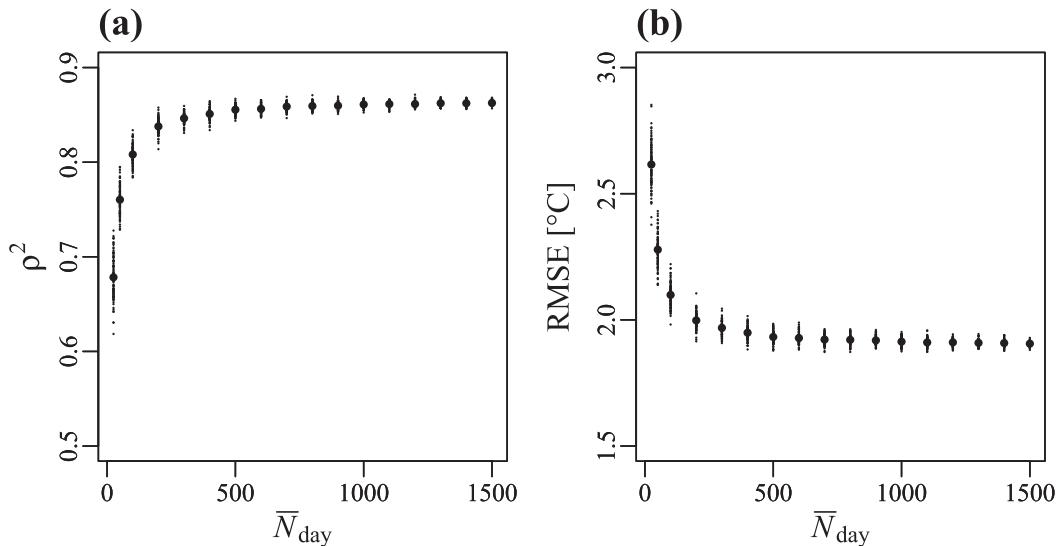


FIG. 5. Relation of the mean \bar{N} used per day (horizontal axis) against model performance: (a) ρ^2 and (b) RMSE. Average of 100 random samples of \bar{N} readings (large black dots) and the 100 samples (small black dots) to indicate the spread. Calibration of the data is performed on the entire dataset, using FluxNet as an air temperature reference. Validation is done against the average city air temperature (all stations combined).

b. Daily temperature estimates per LCZ

The next step is to explore the model potential for the selected LCZ data, and whether spatial temperature differences can be identified and quantified by the smartphones. Whereas the city surface cover is mainly LCZ3 and LCZ6 (Fig. 2), a disproportionately large number of measurements (19%) originate from LCZ1 (3% of surface cover): a compact high-rise building, which is typically found in the city center (Fig. 6). Datasets of battery readings from the LCZs with the most measurements (LCZ1, LCZ3, LCZ6, LCZ8) are used as model input. The resulting T_{est} values for these LCZs are compared to each other to study whether the urban fabric discernibly influences T_{est} . A daily average UHI per LCZ is calculated by subtracting the resulting temperatures from the daily averaged background temperature, taken from the WMO station. Note that this station is surrounded by a built-up area and cannot be considered as an ideal rural station, though T_{air} differences between LCZs will still be visible using this approach. From this analysis a daily mean UHI of $\sim 0.9\text{K}$ arises for LCZ8 and $\sim 0.3\text{K}$ for LCZ3 (Fig. 7a). Standard error in the mean for all LCZs is $\sim 0.09^\circ\text{C}$. Using LCZ-specific battery temperatures does not strongly affect the model output: that is, only the sign of the ME changes for LCZ8 (from -0.48° to 0.37°C ; Figs. 7b and 7c). Where the original model output underestimated the urban T_{air} , for LCZ8 the bias is positive, suggesting higher model temperatures as is indeed seen in the large positive UHI (Fig. 7a). For LCZ3 the bias as compared to the full set remains negative but decreases (to -0.27°C). Since the statistical distribution of the data is unknown, the significance of the UHI effect in these two LCZs is investigated using the nonparametric Kruskal–Wallis test for two independent samples. Test results (not shown here) confirm that the UHI magnitude between LCZs is significantly different. Hence, there is a discernable difference in T_{est} between these LCZs, which shows that the UHI can indeed be observed with this method.

c. Estimation of hourly air temperatures

Next, we explore whether the method can also correctly estimate hourly averaged temperatures, despite the significantly reduced number of measurements available (Fig. 3). The hourly T_{est} shows a relatively poor result (ρ^2 of 0.35) with a large spread (RMSE of 3.2°C) and an ME of roughly 0.9°C (Fig. 8a). It appears that the model results are delayed compared to the reference measurements (Fig. 8a); that is, the maximum T_{est} occurs several hours after the measured maximum temperature. Furthermore, the cooling rate in the evening is

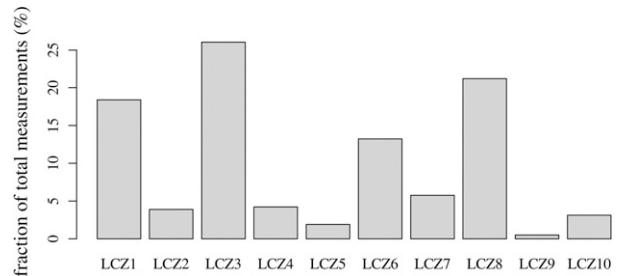


FIG. 6. Histogram of the distribution of battery temperature readings in 2013 over the urban LCZs in São Paulo. Fraction is calculated over the total of the 17 LCZs; the 7 rural LCZs (A–G) have not been plotted; and the LCZ corresponding to each number is explained in Fig. 2.

more rapid in the measured temperature, whereas T_{est} lags behind, cooling later and more slowly. This may be due to the heat capacity of the system (the phone itself, and the insulating layers between phone and air, and phone and body), causing a delay in response. To explore whether results might improve, a delay is introduced to Eq. (2):

$$T_{\text{est}}(t) = m_j [T_p(t + H) - T_0] + T_0. \quad (3)$$

Here $T(t)$ is the temperature ($^\circ\text{C}$) at hour t (hours UTC) and H is the delay in whole hours ($H = 1, 2, \dots, h$). Residual analysis of hourly T_{est} against the city average air temperature yields the best match between smartphone estimates and temperature measurements at $H = 4$ h. The ρ^2 doubles (from 0.36 to 0.72), and the large MAE and RMSE are reduced with over 1°C each, to 1.63° and 1.99°C , respectively (Fig. 8b). While the magnitude of the peaks (positive and negative) is still much larger than the measurements indicate, the timing of the estimated temperatures now corresponds much better to the observations. Analysis of the daily peaks in temperature reveals that on average the delay during the day is roughly 2 h between model and observations, whereas at night the delay can be longer, on average up to 3 or 4 h.

In search of a physical explanation for the delay, we formulate a simplified differential equation for the change in temperature of the phone as caused by the differences between the phone temperature, and the air and body temperatures:

$$\frac{\partial T_p}{\partial t} = \frac{k_e(T_{\text{air}} - T_p)}{mc} + \frac{k_b(T_b - T_p)}{mc}. \quad (4)$$

Here m is the mass of the phone, taken as 0.13 kg, and c is the specific heat of the phone, taken as $600\text{J kg}^{-1}\text{K}^{-1}$ (based on specific heat of glass and sand, for simplicity).

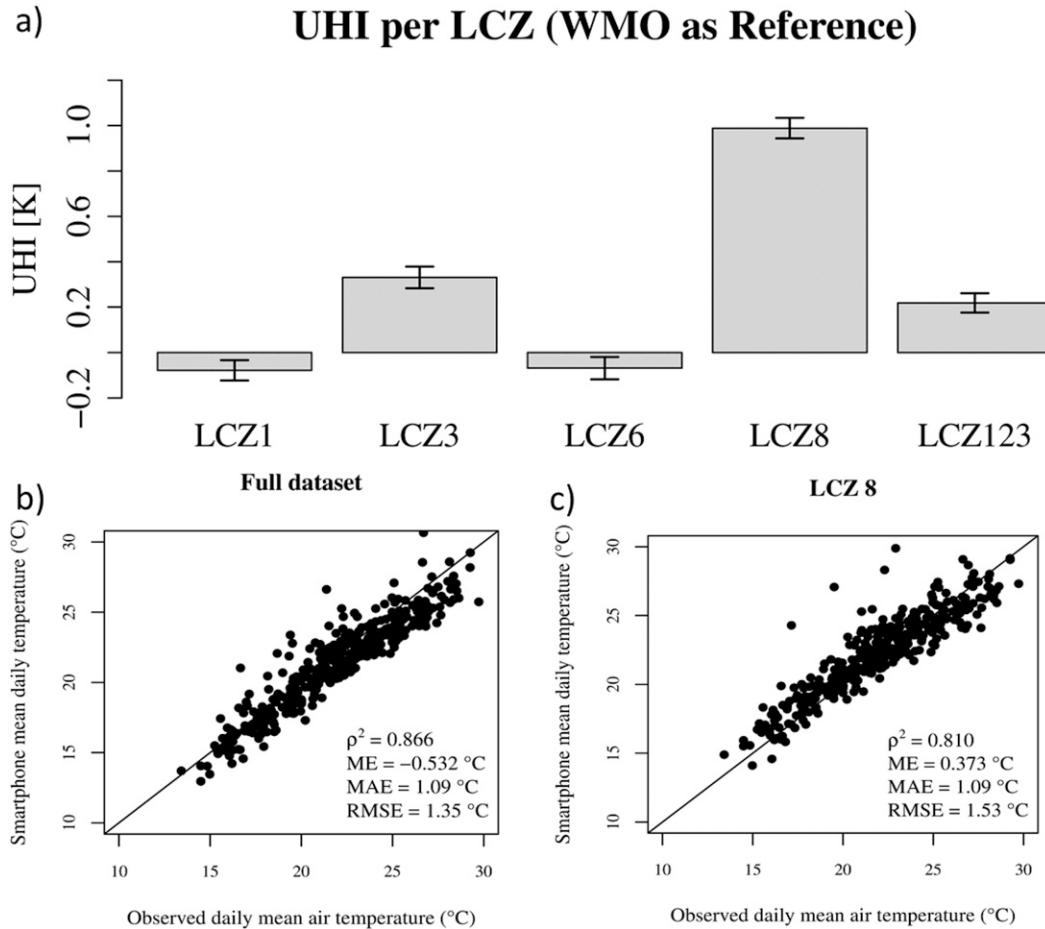


FIG. 7. (a) UHI (K) derived from T_{est} per LCZ, using Congonhas as reference background station. The mean standard errors (whiskers) are shown. Scatterplots of observations (horizontal axis) against T_{est} (vertical axis) of the full (b) dataset (Fig. 4) and (c) LCZ8. The 1:1 lines indicate where measurements would equal model results (straight black lines).

The supporting information of O13 indicates that k is the conductivity multiplied by the surface area divided by the insulating material thickness. We take the typical dimensions of the phone as $10\text{ cm} \times 4\text{ cm}$, and the conductivity of the clothing layer between phone and body as $0.037\text{ K m}^2\text{ W}^{-1}$ (resembling a pair of pants; ASHRAE 2010). Furthermore, in our results $m_j = (1 + k_b/k_e) \approx 2$, so $k_b \approx k_e = 0.11\text{ WK}^{-1}$. Using these typical values, T_b at 37°C and a linear cooling of the atmosphere with $\approx 1\text{ K h}^{-1}$, we can simulate the phone’s cooling (heating) rate. This simple analysis indicates the phone arrives at a steady cooling rate after $\approx (1-2)\text{ h}$, depending on the exact initial values of $T_{air} - T_p$ and the specific heat and mass. The data seem to suggest a larger delay time (up to 4 h): in reality, the heat capacity of the phone will be larger than assumed, by including the heat capacity of the bag or clothes in which it is being carried. The inside T_{air} for those readings taken indoors will

influence the calibration: inside T_{air} reacts to outside T_{air} , with a lower amplitude and another delay factor, thus increasing the response time of the total smartphone system.

A second, implicit way to correct for the delay is by using 24 hourly m_j values, rather than a single fixed m_j , for the entire dataset. By determining one m_j value per hour, the variations in heat transfer efficiency over the day are taken into account, since m_j is the ratio of the thermal insulation k values [Eq. (2)]. Possible variations in human behavior (e.g., clothing) and the available measurements per hour can also be implicitly accounted for with this method. According to theory, the heat flow between phone and environment should decrease when the difference in temperature decreases. One would expect this to happen during the day when air temperature is relatively high and therefore closer to the smartphone battery temperature ($\approx 30^\circ\text{C}$). At night, the temperature difference is larger and the rate of heat

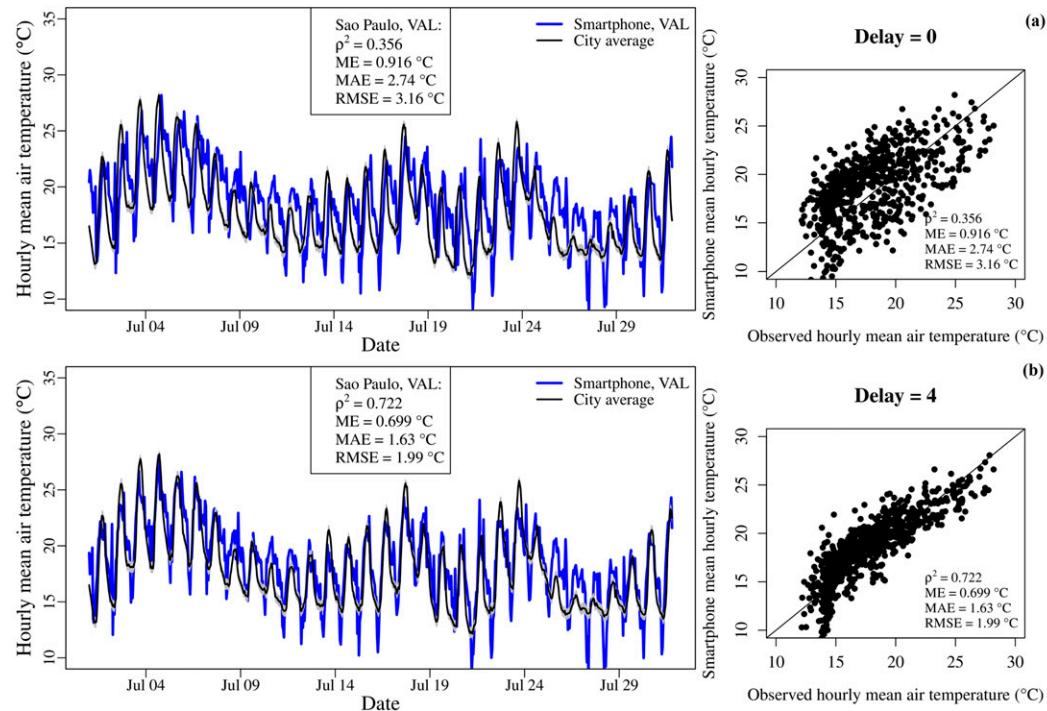


FIG. 8. (left) Average hourly air temperature of São Paulo during July 2014 estimated from smartphone battery temperature readings (blue line) validated against city average air temperature observations (black line), and (right) scatterplots of observations (horizontal axis) against T_{est} (vertical axis); the 1:1 lines indicate where measurements would equal model results (straight black lines). (top) Results of the validation without a delay correction in the smartphone data and (bottom) with a delay-corrected dataset with delay $H = 4$ h.

exchange would increase. When implementing an hourly variation of m_j , the results (not shown) are very similar to the results for the delay-corrected series (Fig. 8b). There are no appreciable differences between the two sets (i.e., hourly m_j vs delay corrected), indicating that m_j implicitly corrects for the delay in the battery response. Values for m_j vary between 1.4 and 2.2 throughout the day, with the higher values occurring during nighttime. A high value indicates k_e increases or that k_b decreases. A higher k_e indicates a larger temperature difference between phone and environment (O13), as does indeed occur during the night (if the reading takes place outdoors).

The compensating effect of m_j on the delay in the battery temperature (Fig. 8b) is confirmed when calibrating hourly m_j values to the explicitly delay-corrected set [Eq. (3)]. When calibrating hourly m_j values to this explicitly corrected set, the range of m_j is halved (ranging between 1.7 and 2.1), though the diurnal pattern (lower m_j during the day) persists: no constant m_j is obtained for the delay-corrected set. Results do not notably improve: a 0.08°C reduction in the MAE, while ρ^2 and RMSE remain equal in comparison to the hourly m_j calibration on the uncorrected dataset. This means

that the observed delay in the smartphone battery estimates can be corrected for by either explicitly accounting for the delay [as in Eq. (3)] or taking 24 hourly m_j values rather than a single fixed value for the entire day.

5. Discussion

a. Relation to other studies

Our study extends O13 by employing a more extensive dataset for just one city, and using independent calibration and validation datasets. The São Paulo results of O13 show nearly the same ρ^2 based on two periods of 3 months (ρ^2 of 0.65 and 0.85 for winter and spring 2012, respectively), as our results provide a ρ^2 of 0.86. The mean absolute error (MAE) for São Paulo in O13 is only slightly higher than our values (1.2°C in O13 and $\approx 1.1^\circ\text{C}$ here). This indicates that even with a smaller dataset (O13 used on average 1383 measurements per day for São Paulo, whereas this study has roughly 10 times more), the daily averaged temperature on a citywide scale can be captured well. Though our study has only focused on one city, the O13 study was carried out for eight different cities in vastly different climate zones with different temperature seasonality and extremes. The

sound results of O13 indicate that the method is valid across a wide variation of climates, rather than only for São Paulo. The specific calibration constants of this work are optimized for São Paulo and are statistically not valid for any other city. However, this is not a fundamental limitation of the proposed method, since for other regions the model can be recalibrated using region-specific data.

Considering the data availability, we find that results deteriorate below a minimum number of battery measurements, even on the daily scale. As an illustration, Overeem et al. (2014) have applied the same method to Rotterdam and Amsterdam, the Netherlands; however, their model statistics are less satisfactory (ρ^2 of 0.77 and 0.67, respectively; MAE of 1.22° and 1.40°C, respectively). The daily data availability was much lower for these comparatively small cities (382 and 116 per day for Rotterdam and Amsterdam, respectively). Similarly, with 203 readings per day Muller et al. (2015) report an even lower ρ^2 of 0.52, and a higher MAE of 1.71°C for Birmingham, United Kingdom. Overeem et al. (2014) provide a relation between the data availability and model ρ^2 , showing that the results become inaccurate for <100 measurements per day. Also, >350 daily measurements are preferable for accurate results: the optimal number of measurements is 700+ (Fig. 5). O13 fulfills this requirement, but the data availability reported in Overeem et al. (2014) and Muller et al. (2015) lies below this threshold. Note that the ρ^2 found by Muller et al. (2015) is lower than the lowest ρ^2 in Fig. 5.

In our study 12×10^3 battery readings are available on average per day (section 3a), and the ρ^2 value is 0.87 for the daily analysis (section 4a; Fig. 4), which clearly illustrates the necessity of having enough data. Overall, our results tend to be of equal or better skill compared to earlier studies with the same heat transfer model. The high data availability provides possibilities for studying the method at an hourly scale or for making selections for separate city areas (LCZs).

b. Data quality and additional filtering

A notable issue is the uncertainty in the location of the smartphone, which is often in the order of tens of meters. This may introduce an uncertainty in coupling a battery reading to a location in the LCZ map (Fig. 2). However, this map is based on satellite imagery with a resolution of 120 m²; the uncertainty in smartphone location should fall within this range. In addition, the phone's GPS tracking is not always turned on (O13), so it remains difficult to discern between indoor and outdoor readings. However, the applied data selection (see O13; section 3a) aims to minimize the uncertainty. In addition, the calibration process will account for this effect as well if relevant. Importantly, the apparent time lag

between temperature changes in phone and environment (section 4c) suggests that readings taken inside may still have been affected by the outdoor temperature.

Moreover, our approach assumes the phone is carried in a pocket, which allows for an assumed equilibrium exchange of heat between body and phone. In practice this assumption may be violated, for example, because the phone is carried in a bag or elsewhere, and on an hourly scale the system may not be in equilibrium. However, calibrated m_j values appeared to be close to a priori estimated m_j values from clothing properties (O13), which supports confidence in the followed approach. Additionally, we assume m_j to be constant over time, whereas clothing thickness (insulation) will obviously undergo a diurnal and seasonal cycle. Possibly, using the light sensor that many smartphone brands now possess, a distinction can be made between indoor and outdoor measurements, if these data are available. A follow-up study that improves the heat transfer model by reducing the assumptions made could be very valuable for further research with these data. The weather can also influence human behavior. On very hot days or days with extreme precipitation, people are more likely to stay indoors, meaning that readings taken during those periods will not reflect the outside air temperature, but rather the indoor environment. For instance, in May 2014 several hail events occurred in São Paulo, during which the error between T_{est} and the observations was relatively high (up to 3°C on 19 May) compared to clear days. A sensitivity test where all days with rainfall higher than 0 mm were excluded (leaving 238 dry days for calibration and 254 for validation) did not significantly improve model results enough to justify losing several weeks of data completely in the rainy months. Less strict filters (1, 2, and 5 mm) made the results nearly identical to the results without any filtering for precipitation amount. Therefore, we decided not to pursue this sensitivity aspect any further. These issues are an inherent drawback of using smartphones for air temperature data, but by averaging a large amount of battery data in space (on the scale of a city or a LCZ) and in time (daily or hourly), the errors will be filtered out to a certain extent, as can be concluded from the results of this study. For a thorough analysis of the reaction of the smartphone battery to changes in air temperature over the course of the day, a controlled trial should be set up with a conventional temperature sensor and several smartphones logging battery temperature. Because of several limitations (such as an inability to do continuous battery logging), we could not perform such a trial, but we strongly recommend it for any future research.

Two additional analyses did not yield any improvement: an attempt was made to use only smartphones with similar hardware (Samsung GTI series), which in theory will have similar k coefficients, similar heat capacities, similar battery temperature sensors, and a similar thermal energy generated by the battery (P). However, this was no improvement over using all the smartphone data available. Additionally, we constructed an extra filter of the raw battery temperatures by Gaussian mixture modeling (Reynolds 2009). This statistical technique assumes that the dataset consists of several sub-distributions or data clusters, each with its own mean and standard deviation. A data cluster that is, for example, characterized by high temperatures could be influenced by the battery charging or extensive use of the phone. These faulty data could be filtered out to improve results. However, the resulting data clusters of the mixture modeling had mean temperature differences smaller than 1°C with standard deviations around 6°C , which hampers distinguishing of clusters.

c. T_0 calibration

Initially, results with the optimized T_0 (section 4a) were worse than the results using the reference T_0 from O13. The T_0 was found to be as high as 49°C , and the MAE of the validation results increased significantly (by $\approx 0.3^{\circ}\text{C}$ compared to the set using $T_0 = 39^{\circ}\text{C}$). This large T_0 value cannot be realistically interpreted as the approximate body temperature, which is ideally near 37°C , plus a constant [Eq. (S-9) in O13]. A more physically sound value for T_0 (39.8°C , used in the analyses) was obtained from repeating the optimization procedure for incrementally increasing random samples of battery temperature used for calibration. For an increasing number N of measurements, the T_0 and RMSE values decrease until $N = 3 \times 10^3$. Beyond this point T_0 remains constant at 39.8°C and RMSE does not appreciably decrease any more (value $\approx 1.49^{\circ}\text{C}$).

d. Weather station measurement data

With three different sources of measurement data, each with its own measurement accuracy, resolution, and location (footprint), it is nearly impossible to know which station represents “the truth.” What the “true” city temperature is remains a matter of definition. Since the city is heterogeneous, the temperature is increasingly influenced by local characteristics when moving from the boundary layer top to the surface layer (Barlow 2014). What is measured using the smartphones is the urban canyon temperature, influenced by the local microclimate. Using a city-averaged T_{air} constructed from all these measurements seems to be the most robust option, to represent the urban air temperature as

accurately as possible. However, this approach will not always yield the best model statistics. Particularly with the LCZ analysis, calibration and validation of the separate LCZs would ideally be performed with a station located in the same LCZ. Calibration of the model using specific LCZ air temperature data would make it better suited to detect differences in smartphone response between LCZs. The number of suitable hobby stations for use was scarce, however. For a city with more stations to choose from, a more thorough selection procedure (based on, e.g., measurement height, metadata, or neighborhood) could be performed according to the principles in Stewart (2011) and Bell et al. (2015). This might also reduce the high uncertainty at night, which can among others be caused by the high variability in measured minimum temperature (Brandsma and van der Meulen 2008), in combination with the low number of battery readings available during those hours. For estimating the absolute value of the UHI with smartphones, a robust rural background station is essential, but one was unavailable in this study.

e. Applicability

Though this article primarily functions as a proof of principle, smartphone-derived air temperatures can have various applications to complement conventional data. For instance, in developing countries, where weather stations are scarce but smartphone ownership is high, smartphones can add valuable information about the urban temperature. This knowledge can be vital during, for instance, heat waves, where knowledge of which neighborhoods are most prone to the UHI can potentially save lives. Additionally, whereas a traditional urban measurement network is very expensive to set up and maintain, and will be prone to vandalism, a smartphone network will not be hindered by these limitations, providing valuable data virtually for free. This will be particularly valuable for those cities for which funds for urban research are limited.

Alternatively, data assimilation in NWP models can be beneficial for NWP in the near future, since model resolution is steadily increasing to an extent that the influence of cities will be felt (ECMWF 2016). An urban scheme is often lacking within these NWP models, so data assimilation of the urban meteorological data will be crucial for reliable forecasts. Given the scarcity of urban data, even relatively coarse data such as the smartphone-derived temperatures could make a contribution to the forecasts. In broader terms, the developed methodology of this study may also be useful for algorithms that are being developed for application to other types of crowdsourced data. A preliminary test in which smartphone-based temperature data were

assimilated within the WRF modeling system for São Paulo revealed that maximum temperatures were forecasted by about 0.5–1 K better for the studied week (not shown).

6. Conclusions

This study utilizes a heat transfer model to translate smartphone battery temperature readings into citywide air temperatures, on both a daily and an hourly scale. This work extends the earlier research by [Overeem et al. \(2013b\)](#) by using an extensive dataset spanning 2 yr of over 10 million battery readings taken in São Paulo, Brazil. We use multiple measurement stations spread across the city for calibration, thereby better capturing the average urban air temperature than using a single WMO station. The extensive data availability allows for a division of the dataset per local climate zone (LCZ) to investigate spatial differences in temperature, as well as zooming in to the hourly temperature variations as captured by the smartphones. The consistent division into a separate calibration (the year 2013) and validation period (the year 2014) for both daily and hourly temperatures ensures that all results are statistically robust, and not subject to autocorrelation.

Estimated daily averaged air temperatures are good and can even be used to calculate temperatures of specific LCZs. A daily averaged UHI can be found in LCZ8 (large low-rise buildings) and LCZ3 (compact low-rise buildings): these LCZs have a significant difference in temperature in comparison to the official WMO airport station. However, insufficient battery temperature data are available to estimate hourly UHI. This would also need a proper rural background station: the airport is fully surrounded by a built-up area (LCZ3).

On the hourly scale, initial results for temperature were poor but were vastly improved after correcting for a seemingly delayed response of the battery temperatures to changes in air temperatures. An analogous improvement can be obtained by using 24 hourly calibration (m_i) constants rather than one average value for all hours. The incorrect magnitude of especially the nighttime lows remains an unsolved issue, possibly as a result of the low number of battery temperature readings taken at night. A larger set of battery temperatures, especially when taken at night, is required to reduce the nighttime underestimation. Making use of an urban test bed like Rotterdam ([Heusinkveld et al. 2014](#)) or Birmingham ([Muller et al. 2015](#); [Warren et al. 2016](#)), could aid with this issue.

From a large number of smartphone readings an accurate air temperature estimate for the daily and even hourly scale of a city can be obtained, which underlines the strength of crowdsourced data. With newer

smartphone models regularly carrying temperature, moisture, or pressure sensors, as well as applications such as mPing and WeatherSignal, there is no denying that measurements from smartphones may hold a lot of potential for future (urban) meteorological studies given their interconnectivity and everyday use in great numbers.

Acknowledgments. The work of this paper was largely carried out at the Royal Netherlands Meteorological Institute. We thank Prof. Amauri Oliveira (University of São Paulo) for providing USP's FluxNet station data; Dr. Gerald Mills (University College Dublin), Michael Foley (UCD), and Maria de Fatima Andrade (USP) for creating and supplying the LCZ WUDAPT data for São Paulo; and James Robinson and the OpenSignal company for supplying the battery temperature dataset. We extend our thanks to the hobby meteorologists for maintaining and uploading the data from their private weather stations. We thank Prof. Berthold Horn (Massachusetts Institute of Technology) for sharing the idea of Gaussian mixture modeling. Gert-Jan Steeneveld and Arjan Droste acknowledge funding from the Netherlands Organization for Scientific Research (NWO) VIDI Grant "The Windy City" (File 864.14.007) and NWO eScience project "ERA-URBAN" (File 027.014.203).

REFERENCES

- Arnfield, A. J., 2003: Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.*, **23**, 1–26, doi:[10.1002/joc.859](#).
- ASHRAE, 2010: Thermal environmental conditions for human occupancy. ANSI/ASHRAE Standard 55-2010, 39 pp.
- Barlow, J. F., 2014: Progress in observing and modelling the urban boundary layer. *Urban Climate*, **10**, 216–240, doi:[10.1016/j.uclim.2014.03.011](#).
- Bell, S., D. Cornford, and L. Bastin, 2013: The state of automated amateur weather observations. *Weather*, **68**, 36–41, doi:[10.1002/wea.1980](#).
- , —, and —, 2015: How good are citizen weather stations? Addressing a biased opinion. *Weather*, **70**, 75–84, doi:[10.1002/wea.2316](#).
- Brandsma, T., and J. P. van der Meulen, 2008: Thermometer screen intercomparison in De Bilt (the Netherlands)—Part II: Description and modeling of mean temperature differences and extremes. *Int. J. Climatol.*, **28**, 389–400, doi:[10.1002/joc.1524](#).
- ECMWF, 2016: ECMWF strategy 2016–2025: The strength of a common goal. 27 pp. [Available online at http://www.ecmwf.int/sites/default/files/ECMWF_Strategy_2016-2025.pdf.]
- Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfusz, 2014: mPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342, doi:[10.1175/BAMS-D-13-00014.1](#).
- Hajat, S., and T. Kosatky, 2010: Heat-related mortality: A review and exploration of heterogeneity. *J. Epidemiol. Community Health*, **64**, 753–760, doi:[10.1136/jech.2009.087999](#).

- Heusinkveld, B. G., G. J. Steeneveld, L. W. A. van Hove, C. M. J. Jacobs, and A. A. M. Holtslag, 2014: Spatial variability of the Rotterdam urban heat island as influenced by urban land use. *J. Geophys. Res. Atmos.*, **119**, 677–692, doi:10.1002/2012JD019399.
- Kotthaus, S., H. C. Ward, C. S. B. Grimmond, and J. G. Evans, 2012: Flux measurements in Southern England by KCL and CEH. *FluxLetter*, Vol. 5, No. 1, FLUXNET Office, University of California, Berkeley, Berkeley, CA, 25–31. [Available online at <http://www.fluxdata.org/Shared%20Documents/FluxLetter/FluxLetter%20Vol5%20no1.pdf>.]
- Mass, C. F., and L. E. Madaus, 2014: Surface pressure observations from smartphones: A potential revolution for high-resolution weather prediction? *Bull. Amer. Meteor. Soc.*, **95**, 1343–1349, doi:10.1175/BAMS-D-13-00188.1.
- Meier, F., D. Fenner, T. Grassmann, M. Otto, and D. Scherer, 2017: Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, **19**, 170–191, doi:10.1016/j.uclim.2017.01.006.
- MetOffice, 2011: Weather observations website. Accessed 12 May 2017. [Available online at <http://wow.metoffice.gov.uk/>.]
- Mills, G., B. Bechtel, J. Ching, L. See, J. Feddema, M. Foley, P. Alexander, and M. O'Connor, 2015: An introduction to the WUDAPT project. *Ninth Int. Conf. on the Urban Climate/12th Symp. on the Urban Environment*, Toulouse, France, Météo-France, 6 pp. [Available online at http://www.meteo.fr/cic/meetings/2015/ICUC9/LongAbstracts/gd2-1-6521222_a.pdf.]
- Miralles, D. G., A. J. Teuling, C. C. van Heerwaarden, and J. Vilà-Guerau de Arellano, 2014: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat. Geosci.*, **7**, 345–349, doi:10.1038/ngeo2141.
- Molenaar, R. E., B. G. Heusinkveld, and G. J. Steeneveld, 2016: Projection of rural and urban thermal comfort in The Netherlands in 2050. *Int. J. Climatol.*, **36**, 1708–1723, doi:10.1002/joc.4453.
- Muller, C. L., L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R. R. Leigh, 2015: Crowdsourcing for climate and atmospheric sciences: Current status and future potential. *Int. J. Climatol.*, **35**, 3185–3203, doi:10.1002/joc.4210.
- Oke, T. R., 1982: The energetic basis of the urban heat island. *Quart. J. Roy. Meteor. Soc.*, **108**, 1–24, doi:10.1002/qj.49710845502.
- Oleson, K. W., A. Monaghan, O. Wilhelmi, M. Barlage, N. Brunzell, J. Feddema, L. Hu, and D. F. Steinhoff, 2015: Interactions between urbanization, heat stress, and climate change. *Climatic Change*, **129**, 525–541, doi:10.1007/s10584-013-0936-8.
- Overeem, A., H. Leijnse, and R. Uijlenhoet, 2013a: Country-wide rainfall maps from cellular communication networks. *Proc. Natl. Acad. Sci. USA*, **110**, 2741–2745, doi:10.1073/pnas.1217961110.
- , J. C. R. Robinson, H. Leijnse, G. J. Steeneveld, B. K. P. Horn, and R. Uijlenhoet, 2013b: Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophys. Res. Lett.*, **40**, 4081–4085, doi:10.1002/grl.50786.
- , —, —, —, —, and —, 2014: Stadstemperatuur in beeld met smartphones. *Meteorologica*, **23**, 16–18.
- Reid, C. E., M. S. O'Neill, C. J. Gronlund, S. J. Brines, D. G. Brown, A. V. Diez-Roux, and J. Schwartz, 2009: Mapping community determinants of heat vulnerability. *Environ. Health Perspect.*, **117**, 1730–1736, doi:10.1289/ehp.0900683.
- Reynolds, D., 2009: Gaussian mixture models. *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds., Springer, 659–663.
- Ronda, R. J., G. J. Steeneveld, B. G. Heusinkveld, J. J. Attema, and A. A. M. Holtslag, 2017: Urban finescale forecasting reveals weather conditions with unprecedented detail. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-16-0297.1, in press.
- Snik, F., and Coauthors, 2014: Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters. *Geophys. Res. Lett.*, **41**, 7351–7358, doi:10.1002/2014GL061462.
- Sosko, S., and S. Dalyot, 2015: Towards the use of crowdsourced volunteered meteorological data for forest fire monitoring. *GEOProcessing 2015: The Seventh International Conference on Advanced Geographic Information Systems, Applications, and Services*, C.-P. Rückermann et al., Eds., 127–132, IARIA. [Available online at http://www.thinkmind.org/index.php?view=article&articleid=geoprocessing_2015_6_50_30173.]
- Steeneveld, G. J., S. Koopmans, B. G. Heusinkveld, L. W. A. van Hove, and A. A. M. Holtslag, 2011: Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands. *J. Geophys. Res.*, **116**, D20129, doi:10.1029/2011JD015988.
- Stewart, I. D., 2011: A systematic review and scientific critique of methodology in modern urban heat island literature. *Int. J. Climatol.*, **31**, 200–217, doi:10.1002/joc.2141.
- , and T. R. Oke, 2012: Local climate zones for urban temperature studies. *Bull. Amer. Meteor. Soc.*, **93**, 1879–1900, doi:10.1175/BAMS-D-11-00019.1.
- United Nations, 2012: World urbanization prospects: The 2011 revision: Highlights. Rep. ST/ESA/SER.A/322, Population Division, Dept. of Economic and Social Affairs, 302 pp. [Available online at http://www.un.org/en/development/desa/population/publications/pdf/urbanization/WUP2011_Report.pdf.]
- Warren, E. L., D. T. Young, L. Chapman, C. Muller, C. S. B. Grimmond, and X. M. Cai, 2016: The Birmingham Urban Climate Laboratory—A high density, urban meteorological dataset, from 2012–2014. *Sci. Data*, **3**, 160038, doi:10.1038/sdata.2016.38.