Meteorologische Zeitschrift, Vol. **26**, No. 5, 525–547 (published online September 24, 2017)

Intra and inter 'local climate zone' variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany

Daniel Fenner[1*], Fred Meier[1], Benjamin Bechtel[2], Marco Otto[1] and Dieter Scherer[1]

[1]Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Berlin, Germany
[2]Institute of Geography, Department of Earth Sciences, Universität Hamburg, Hamburg, Germany

(Manuscript received March 16, 2017; in revised form June 2, 2017; accepted June 2, 2017)

Abstract

A one-year data set for the year 2015 of near-surface air temperature (T), crowdsourced from 'Netatmo' citizen weather stations (CWS) in Berlin, Germany, and surroundings was analysed. The CWS data set, which has been quality-checked and filtered in a previous study, consists of T measurements from several hundred CWS. It was investigated (1) how CWS are distributed among urban and rural environments, as represented by 'local climate zones' (LCZ), (2) how LCZ are characterised in T along the annual cycle and concerning intra-LCZ T variability, and (3) if significant T differences between LCZ (ΔT) can be detected with CWS data. Further, it was investigated how the results from CWS compare to reference data from standard meteorological measurement stations. It can be shown that all 'urban' LCZ are covered by CWS, but only few CWS are located in 'natural' LCZ (e.g. forests or urban parks). CWS data along the annual cycle show generally good agreement to reference data, though for some LCZ monthly means between both data sets differ up to 1 K. Intra-LCZ T variability is particularly large during night-time. Statistically significant ΔT can be detected with CWS data between various LCZ pairs, particularly for structurally dissimilar LCZ, and the results are in agreement with existing literature on LCZ or the urban heat island. Furthermore, annual mean ΔT in CWS data agree well with reference data, thus showing the potential of CWS data for long-term studies. Several challenges related to crowdsourced CWS data need further investigation, namely missing meta data, the non-standard measurement locations, the imbalanced availability in time and space, and potentials to combine CWS and reference data to benefit from the main advantages of both, i.e., the large number of stations and the high quality of data, respectively.

Keywords: urban climate, local climate zones, air temperature, crowdsourcing, citizen weather station, Berlin

1 Introduction

Due to high spatial and temporal variability of urban climates, dense sensor networks are needed to observe and study these atmospheric conditions. The trade-off between accuracy and (maintenance) costs of instruments, however, remains a major limitation in urban climatology and a discussion point until today (Chapman et al., 2015). Moreover, there is broad consensus that spatial heterogeneity of urban areas and their underlying structures (e.g. land cover, height of roughness elements, sky view factor, surface albedo) leads to a spatially non-uniform distribution of near-surface air temperature (T) (e.g. Oke, 1982; Eliasson and Svensson, 2003; Unger, 2004; Kim and Baik, 2005; Kolokotroni and Giridharan, 2008; Houet and Pigeon, 2011; Suomi and Käyhkö, 2012; Fenner et al., 2014; Schatz and Kucharik, 2014; Gál et al., 2016). Today, urban-rural air temperature differences are one of the most investigated features in studies of urban climatology (Arnfield, 2003; Stewart, 2011) and the term 'urban heat island' (UHI) expresses that urban areas are, in general, warmer than their natural surroundings. However, the term 'urban heat island' over-simplifies the temporal and spatial heterogeneity of urban climate. Besides, the terms 'urban' and 'rural', applied to distinguish measurement sites in UHI studies, do not accurately express the local settings of a measurement station (Stewart, 2011).

The need for a standardised approach to characterise measurement locations in terms of their local surroundings was addressed by Stewart and Oke (2012), introducing the Local Climate Zone (LCZ) concept (Fig. 1). The concept builds on existing climate-based classification schemes (e.g. Chandler, 1965; Eleffsen, 1991; Scherer et al., 1999; Oke, 2006; Loridan and Grimmond, 2012) but generalises them to make the scheme applicable to all urban regions. The LCZ concept classifies urban and natural environments into classes, which are distinguished by their surface parameters, e.g. building surface fraction, sky view factor, height of roughness elements, and anthropogenic heat flux. Each LCZ has a specific range of these parameters and the concept distinguishes between ten 'urban' or built-up, and seven 'natural' LCZ (Fig. 1). A number of studies could

*Corresponding author: Daniel Fenner, Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Rothenburgstraße 12, 12165 Berlin, Germany, e-mail:daniel.fenner@tu-berlin.de

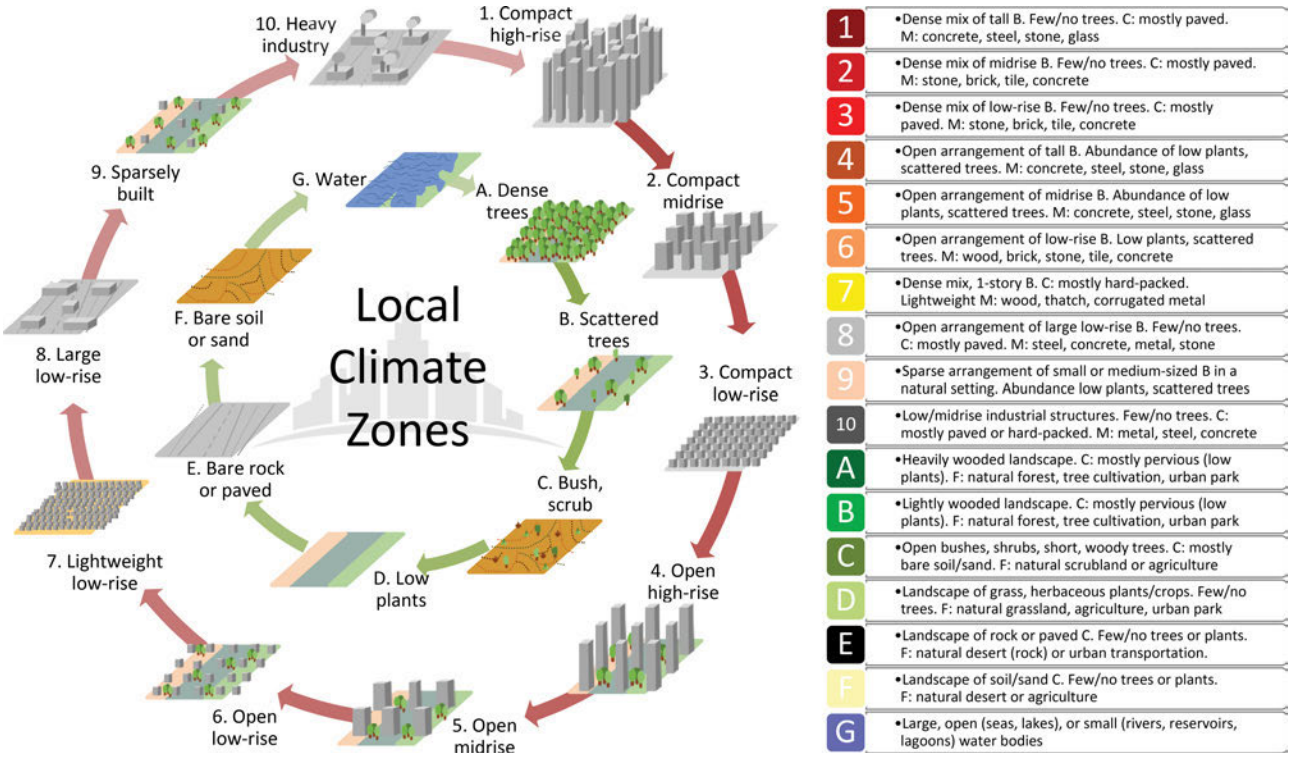

Figure 1: The local climate zone (LCZ) concept with 'urban' (LCZ 1–10) and 'natural' (A–G) LCZ, and their characteristics (adapted from Table 2 in Stewart and Oke (2012); text shortened, icons reworked). B: Buildings, C: cover, M: materials, F: function, tall: > 10 stories, mid-rise: 3–9 stories, low: 1–3 stories. Figure taken from Bechtel et al. (2017).

show that distinct T differences (ΔT) between stations, located in different LCZ, exist (e.g. Emmanuel and Krüger, 2012; Siu and Hart, 2013; Alexander and Mills, 2014; Fenner et al., 2014; Lehnert et al., 2014; Stewart et al., 2014; Arnds et al., 2017).

However, only few studies so far have investigated how T varies within one LCZ (i.e., intra-LCZ variability). Lehnert et al. (2014) detected non-zero ΔT between LCZ for the city of Olomouc, Czech Republic, but they also revealed distinct intra-LCZ T variability. Such features were confirmed by a recent study in Szeged, Hungary (Skarbit et al., 2017), showing that intra-LCZ variability is especially pronounced at night. Using data from mobile measurements for selected time periods, Stewart et al. (2014) and Leconte et al. (2015) also reported that remarkable spatial variability of T exists within the same LCZ, which they relate to microscale variability of surface characteristics such as building density or surface cover. Nonetheless, both studies (Stewart et al., 2014; Leconte et al., 2015) point out that spatially averaged T for the same LCZ yield representative values of local-scale T. However, these analyses were based on short-term observations, covering only few selected days, while long-term studies with stationary measurements addressing inter- and intra-LCZ T variability are still rare (e.g. Lehnert et al., 2014; Skarbit et al., 2017). Main reasons are the costs and difficulties of setting up and maintaining dense observation networks within urban areas for extensive time periods (Muller et al., 2013; Chapman et al., 2015).

The current state of urban atmospheric observation networks is discussed in a review by Muller et al. (2013), pointing out some high-density networks such as the Helsinki Testbed (Koskinen et al., 2011), the Birmingham Urban Climate Laboratory (Chapman et al., 2015; Warren et al., 2016), or the Oklahoma City Micronet (Basara et al., 2010; Hu et al., 2016). Nevertheless, the challenge remains to acquire observational data over longer time periods at high spatial resolution to adequately represent spatio-temporal heterogeneity of T as expected within urban areas (Grimmond, 2006; Muller et al., 2013). One approach to overcome these limitations is the use of data from meteorological stations maintained by citizens, also called 'citizen weather stations' (CWS). A number of studies showed the potential of using such data for scientific applications (Steeneveld et al., 2011; Wolters and Brandsma, 2012; Bell et al., 2013; Castell et al., 2015).

Recently, Chapman et al. (2016) and Meier et al. (2017) presented a new approach for investigating urban thermal climates using crowdsourcing techniques and obtaining freely-available CWS data from several hundred stations in London, United Kingdom (UK), and Berlin, Germany, respectively. They (Chapman et al., 2016; Meier et al., 2017) define crowdsourcing according to Muller et al. (2015) as the automated collection of data from privately owned sensors that are connected to the internet. The term 'crowdsourcing' is adopted in this study to express that the

CWS data stem from an undefined crowd using identical CWS devices. The data sets used by Chapman et al. (2016) and Meier et al. (2017) originate from 'Netatmo' weather stations (https://www.netatmo.com/product/weather/weatherstation). These CWS can be bought by interested citizens worldwide to monitor indoor and outdoor atmospheric conditions. Chapman et al. (2016) analysed night-time T and UHI distribution of London, UK, during a short summer period, showing that distinct features of the urban thermal climate can be investigated using these data. However, they, among others, also stress that quality assurance plays a key role for meaningful data analyses, and that this remains the biggest challenge when using crowdsourced data (Bell et al., 2013; Muller et al., 2015; Chapman et al., 2016).

Meier et al. (2017) presented CWS data from Netatmo weather stations in Berlin, Germany, and surroundings for an entire year, addressing the issue of quality assurance. They developed a systematic quality assessment procedure, taking the different sources of errors into account that are associated with T data from crowdsourced CWS. While Meier et al. (2017) showed that a number of challenges are linked to this novel data set for urban climate research, they also concluded that these challenges can be overcome, and that several benefits are linked to crowdsourcing T measurements. Main advantages of Netatmo CWS are the consistent use of the same type of sensors with the same technical specifications in all stations, high spatial density and coverage in many urban areas, and an application programming interface (API) provided by the company facilitating data acquisition.

In this study, the aim is to expand the analyses in Meier et al. (2017) to investigate whether crowdsourced data from Netatmo CWS are applicable for detecting and analysing spatial T differences of different urban environments. To classify these environments and the measurement sites, the LCZ concept is applied as an example for characterisation of urban morphology and surface cover. The objective is not to obtain spatially interpolated T from CWS data. The city of Berlin is chosen as test bed since it is a good example of a large city with a distinct urban climate without an influence of mountains or the sea (e.g. Hupfer and Chmielewski, 1990; Endlicher and Lanfer, 2003; Fenner et al., 2014). To make the overall approach transferable to other urban regions, CWS data are combined with the satellite-image-based LCZ classification method of Bechtel and Daneke (2012) and Bechtel et al. (2015). Specifically, the following research questions are addressed: (1) How is the spatial heterogeneity of Berlin's metropolitan region, as represented by LCZ, covered by CWS? (2) How are LCZ characterised in terms of crowdsourced T along the annual cycle, concerning intra-LCZ T variability, and in comparison with data from standard meteorological networks? (3) Can significant ΔT between different urban environments (i.e., inter-LCZ differences) be detected in CWS data, and can the data set thus be used for intra-urban differentiation of thermal climates?

2 Data and methods

2.1 Study area and period

The study focuses on the city of Berlin, Germany, and surrounding areas, located in eastern Germany (52.52 N, 13.40 E) with a population of approximately 3.5 million inhabitants in 2015. Overall, the topography in the region is relatively flat. The central parts of the city are at approximately 30 m above mean sea level (amsl) with only solitary peaks up to 120 m amsl high at the edges of the urban agglomeration. The river Spree runs through the centre of the city. The Müggelsee lake (7.4 km^2) is located in the south-eastern part of Berlin, while in the western part several lakes are connected via the river Havel (Fig. 2). Berlin's climate is characterised by a humid warm temperate climate (Cfb) according to Köppen's classification (Kottek et al., 2006).

The investigation period is the year 2015. Mean T in 2015 is 11.2 °C at the station of the German Weather Service (Deutscher Wetterdienst – DWD) in Berlin-Tempelhof (TEMP, Fig. 2; Table 1), which is 1.3 K above the long-term mean (1981–2010). Mean daily maximum (T_{max}) and minimum T (T_{min}) are also higher with 15.6 °C (+1.9 K) and 6.9 °C (+0.8 K), respectively. Especially the months January, August, November, and December show positive T anomalies (Fig. 3). Annual precipitation amounts to 507 mm a^{-1} in 2015, which is 69 mm below (−12 %) the long-term mean (1981–2010). While in January, October, and November more precipitation is recorded, the months February, May, August, and December are especially dry (Fig. 3).

2.2 Crowdsourced air temperature data

Crowdsourced T data from CWS of the private company 'Netatmo' (https://www.netatmo.com/) were used. Measurements of these stations are taken by two modules, i.e., an indoor and an outdoor module. The outdoor module measures T (specified accuracy by manufacturer: ±0.3 K in the range −40 to 64 °C) and relative humidity (RH), while the indoor module additionally measures air pressure, noise level, and carbon dioxide concentration. Measurements are taken at approximately five minute intervals (instantaneous values), which are automatically uploaded to the Netatmo server via private WiFi connection. If the user agrees, outdoor T and RH, and air pressure measurements are shared publicly (https://weathermap.netatmo.com/), and data can be acquired at no cost via an API.

An automatic work-flow to fetch these CWS data for Berlin and surroundings at one-hourly intervals was set up, storing the data in a local database. The details of

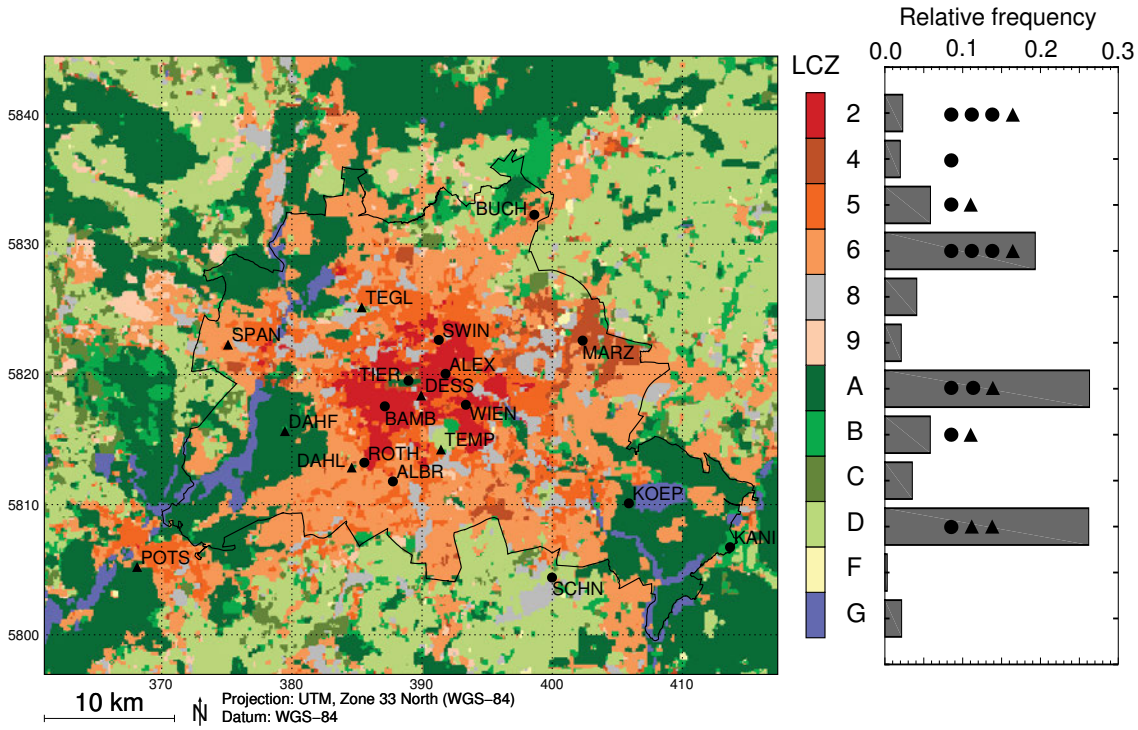

Figure 2: Local climate zone (LCZ) coverage in the study region Berlin, Germany, and locations of the reference measurement stations (black symbols). Dots mark stations that were used in the analyses; triangles mark stations that were filtered out when applying the kernel filter (cf. Section 2.4). The black line denotes the city border of Berlin. The relative frequency of each LCZ in the study region is given in the right panel, along with the number of reference sites per LCZ as symbols (dots and triangles).

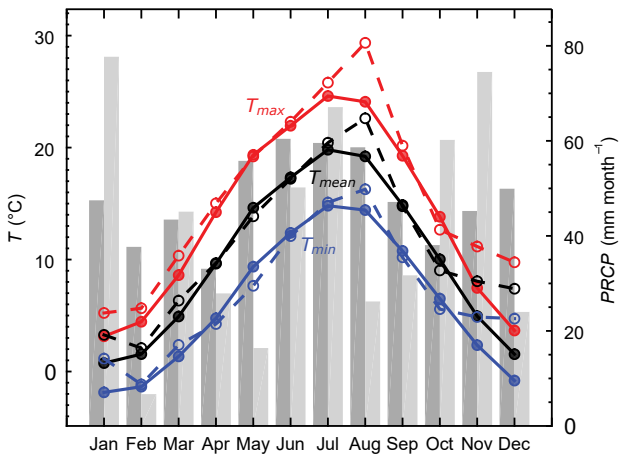

Figure 3: Climate diagram for the station Berlin-Tempelhof (TEMP) (13.4021° E, 52.4675° N, 48 m amsl) for the period 1981–2010 (solid lines, dark grey bars) and the year 2015 (dashed lines, light grey bars). Blue lines: mean daily minimum air temperature T_{min}, black lines: mean daily air temperature T_{mean}, red lines: mean daily maximum air temperature T_{max}, bars: monthly sum of precipitation *PRCP*.

the method are described in Meier et al. (2017). However, these raw data can hardly be used directly for scientific analyses. Common quality issues with these crowdsourced data are inconsistent meta-data, low data availability, outdoor devices probably installed inside buildings, and radiation errors due to exposure of sensors in sunlit places. Hence, Meier et al. (2017) developed a rigorous data quality assessment (QA) and filtering routine for T data. The QA procedure addresses the different error sources linked to crowdsourced data, i.e., soft- and hardware limitations, inconsistent meta-data, and installation deficiencies caused by the user. Meier et al. (2017) distinguish between different quality levels with level D being the highest level, which is the quality level of the crowdsourced data set used in this study.

Due to a server failure no data could be collected during nine days in May and eight days in August 2015. These gaps were filled with data acquired retrospectively; details are given in Appendix A. Afterwards, the complete data for the two months were quality-checked and filtered according to Meier et al. (2017).

The quality-checked data set used in the following consists of T measurements at one-hourly temporal resolution (instantaneous values for all times except the two periods in May and August with hourly mean values, see Appendix A) for up to several hundred measurement sites within Berlin and surrounding area in 2015. T data from CWS are denoted as ${}^{\mathrm{cws}}T$ in the following.

To account for different elevation heights of the measurement sites all T data were corrected with respect to a reference height of 48 m amsl using the dry adiabatic lapse rate (-9.8×10^{-3} K m^{-1}). For this, globally avail-

Table 1: Locations, land cover information, and local climate zones (LCZ) of reference stations. Ground sky view factor (SVF), excluding roofs, was calculated using SOLWEIG 2014a (Lindberg et al., 2008) including buildings and vegetation (trunk zone was set to a fraction of 0.25 of the vegetation height). Land cover information were obtained from EnvAtlBer (2014) and are given as mean values for a radius of 250 m. For sites POTS and SCHN no land cover information were available. Expert LCZ classification was based on the available land cover information, visual inspection of aerial photography, and local expert knowledge. Bold stations were used in the analyses, the others where filtered out when applying the kernel filter (cf. Section 2.4). Operators: DWD – Deutscher Wetterdienst, TUB – Technische Universität Berlin.

Site (Operator)	Longitude (degree)	Latitude (degree)	Building fraction (%)	Vegetation fraction (%)	Building height (m)	Vegetation height (m)	SVF (–)	LCZ (WUDAPT)	# LCZ-pixels[1]	LCZ (expert)[2]
ALBR (TUB)	**13.3486**	**52.4447**	**9.7**	**69.3**	**10.9**	**6.8**	**0.36**	**6 – Open low-rise**	**25**	**6 – Open low-rise**
ALEX (DWD)	**13.4054**	**52.5198**	**25.7**	**30.7**	**25.7**	**7.7**	**0.54**	**2 – Compact midrise**	**25**	**5 – Open midrise**
BAMB (TUB)	**13.3375**	**52.4964**	**34.5**	**33.5**	**20.8**	**9.5**	**0.24**	**2 – Compact midrise**	**25**	**2 – Compact midrise**
BUCH (DWD)	**13.5022**	**52.6309**	**13.1**	**34.4**	**10.0**	**7.5**	**0.55**	**6 – Open low-rise**	**25**	**6 – Open low-rise**
DAHF (TUB)	13.2252	52.4777	0.00	93.1	0.00	7.3	0.46	B – Scattered trees	10	–
DAHL (DWD)	13.3017	52.4537	5.7	82.5	8.7	6.7	0.39	6 – Open low-rise	19	–
DESS (TUB)	13.3783	52.5045	31.3	33.4	19.0	5.1	0.42	2 – Compact midrise	16	–
KANI (DWD)	**13.7309**	**52.4040**	**0.8**	**43.1**	**3.4**	**7.2**	**0.90**	**B – Scattered trees**	**25**	**B – Scattered trees**
KOEP (TUB)	**13.6157**	**52.4330**	**0.**	**99.0**	**0.**	**14.2**	**0.06**	**A – Dense trees**	**25**	**A – Dense trees**
MARZ (DWD)	**13.5598**	**52.5447**	**13.5**	**37.1**	**11.5**	**3.4**	**0.61**	**4 – Open high-rise**	**21**	**5 – Open midrise**
POTS (DWD)	13.0622	52.3813	–	–	–	–	–	A – Dense trees	18	–
ROTH (TUB)	**13.3158**	**52.4572**	**19.3**	**61.3**	**12.4**	**9.9**	**0.23**	**6 – Open low-rise**	**21**	**6 – Open low-rise**
SCHN (DWD)	**13.5306**	**52.3807**	–	–	–	–	–	**D – Low plants**	**25**	**D – Low plants**
SPAN (TUB)	13.1584	52.5364	12.6	37.7	4.2	2.0	0.74	5 – Open midrise	12	–
SWIN (TUB)	**13.3969**	**52.5431**	**24.5**	**40.3**	**16.4**	**6.9**	**0.34**	**5 – Open midrise**	**22**	**5 – Open midrise**
TEGL (DWD)	13.3088	52.5644	2.5	68.7	5.5	0.4	0.92	D – Low plants	12	–
TEMP (DWD)	13.4021	52.4675	0.1	81.2	3.0	0.2	0.96	D – Low plants	11	–
TIER (TUB)	**13.3636**	**52.5145**	**0.9**	**88.6**	**6.9**	**15.5**	**0.16**	**A – Dense trees**	**23**	**A – Dense trees**
WIEN (TUB)	**13.4291**	**52.4987**	**33.5**	**26.6**	**16.1**	**8.2**	**0.37**	**2 – Compact midrise**	**25**	**2 – Compact midrise**

[1]The number of pixels of the same WUDAPT-LCZ as the measurement site's pixel for a 5×5 kernel, centred on the pixel of the WUDAPT classification where the site is located. [2]LCZ classification only carried out for measurement sites with at least 20 pixels of the same WUDAPT-LCZ as the measurement site's pixel within the 5×5 kernel. These sites were used in the analyses.

able digital elevation data from the Shuttle Radar Topographic Mission (SRTM) (Farr et al., 2007) version 4 at 0.000833° (~ 90 m) spatial resolution was used, and the nearest pixel value was assigned to each site. Additionally, a uniform sensor height of 2 m above ground was assumed for each site.

2.3 Reference air temperature data

Measurement data from ten sites of the Urban Climate Observation Network (UCON), maintained by the Chair of Climatology at Technische Universität Berlin, were used as reference for crowdsourced data (Table 1, Fig. 2). Data from this network has previously been used for studies of long-term spatial and temporal characteristics of T (Fenner et al., 2014), and as observational data for model evaluation (Jänicke et al., 2017; Kuik et al., 2016). All UCON measurement sites are equipped with Campbell Scientific CS215 T and RH probes (accuracy for $T \pm 0.4$ K in the range 5 to 40 °C) in white radiation shields, actively ventilated during sunlit periods. Raw measurement data at one-minute resolution were quality-checked as described in Meier et al. (2017), and aggregated to hourly mean values for further analyses.

Additionally, T data from nine meteorological stations in and around Berlin maintained by DWD were used to complement UCON data (Table 1, Fig. 2). Measurement data (measurements taken with Eigenbrodt

LTS2000 T probe in white radiation shields, accuracy ±0.2 K) are available as quality-checked products at hourly resolution (DWD Climate Data Center 2016; Kaspar et al., 2013).

Both data sets combined form the reference data set of T, denoted as ^{ref}T, for comparison with crowdsourced data. Since the stations are located in a variety of local settings (cf. Table 1) they are highly suitable for this purpose. The stations are identical to the reference stations used in Meier et al. (2017) for the QA of CWS data but were supplemented by two more stations (KOEP and WIEN). All reference stations measure T in a height of 2 m above ground except at BAMB (2.5 m), DESS (3.5 m), and WIEN (33.6 m). Similar to CWS sites, data were corrected to a reference height of 48 m amsl according to the dry adiabatic lapse rate, taking into account the measurement height and using SRTM elevation data.

2.4 Local climate zone classification

Ideally, and following the recommendations by Stewart and Oke (2012), an LCZ classification process for a single measurement site should be carried out by (1) collecting the relevant metadata, (2) estimating the station's thermal source area, and (3) selecting the best suited LCZ for this site. Alternatively, a number of approaches have been made to derive LCZ maps for entire cities, using different types of data (e.g. Alexander and Mills, 2014; Lelovics et al., 2014; Unger et al., 2014; Geletic and Lehnert, 2016). One approach that offers the possibility to potentially classify LCZ for any urban region using remote-sensing data and free software was developed by Bechtel and Daneke (2012) and Bechtel et al. (2015). The method is used in an international effort to gather and distribute information on form and function of cities worldwide in a consistent manner, called the World Urban Database and Access Portal Tools (WUDAPT) (Bechtel et al., 2015; See et al., 2015). The WUDAPT-LCZ approach was successfully applied by, e.g., Brousse et al. (2016) for Madrid, Spain, and by Perera and Emmanuel (2016) for Colombo, Sri Lanka.

This LCZ classification is a supervised classification method, thus the user first creates training areas (TA) for each LCZ using Google Earth software (map data: Google, DigitalGlobe). The TA are used by SAGA GIS software (Conrad et al., 2015) to identify the underlying spectral and thermal properties of each LCZ from Landsat images, train a random forest classifier (Breiman, 2001), and subsequently apply the classifier to infer the most probable corresponding LCZ for each pixel. The whole classification process is iterative, i.e., after the first and following classifications the TA can be corrected by the user to optimise classification results.

The LCZ classification for Berlin was carried out using the available materials on the WUDAPT webpage (http://www.wudapt.org/). This includes Landsat scenes (LC81930232015084LGN00 and LC81930232015100 LGN00 from March and April 2015, respectively), and a predefined region of interest (ROI). Six iteration steps were carried out to improve the LCZ classification, and a post classification filter (majority filter with radius of two pixels) was applied to the last classification, avoiding a too granular classification image. The final classification was evaluated in terms of accuracy and robustness, showing an overall consistent and correct classification. Details on the evaluation are given in Appendix B.

The final LCZ classification (resolution 0.0011°; ~ $122 \times 75\,m^2$) was resampled to a regular $100 \times 100\,m^2$ horizontal resolution using nearest-neighbour interpolation for further analyses (Fig. 2). Afterwards, a LCZ was assigned to each CWS and to each reference station using a two-step procedure. First, the LCZ of the pixel containing the station was assigned. Second, all measurement sites located in regions of heterogeneous LCZ coverage were excluded to avoid obscuring the T signal with stations not representative for a specific LCZ. This was done by keeping a station if more than 20 pixels within a 5 × 5 kernel (i.e., 80 %), centred on the pixel assigned to the respective station (step 1), were of the same LCZ as the centre pixel. The kernel width corresponds to a distance of approximately 250 m around each site, which is consistent with the definition of 'local' by Stewart and Oke (2012), and which was also used by, e.g., Lelovics et al. (2014) as the minimal distance from the boundaries of a LCZ area for measurement sites in Szeged, Hungary, or for LCZ classification of stations by Skarbit et al. (2017).

This two-step LCZ assignment reduced the number of CWS and reference stations. From the original 19 reference stations seven did not pass the filter (Table 1). The remaining twelve stations are located in seven different LCZ: three in LCZ 6 (open low-rise), three in LCZ 2 (compact midrise), two in LCZ A (dense trees), and one each in LCZ 4 (open high-rise), 5 (open midrise), B (scattered trees), and D (low plants). The WUDAPT-LCZ of these twelve sites was also checked against a LCZ classification using detailed land cover information, aerial photography, and local expert knowledge (Table 1). Only for two sites, ALEX and MARZ, the two types of LCZ classifications do not match. Site ALEX was classified as LCZ 2 (dense midrise) and site MARZ as LCZ 4 (open high-rise) in the WUDAPT classification, while the classification with detailed land cover data locates both sites in LCZ 5 (open midrise). However, considering the values of surface properties for these LCZ as given by Stewart and Oke (2012), it becomes obvious that there is overlap between these LCZ, and hence a clear distinction cannot always be easily made. For a consistent methodology the WUDAPT-LCZ classification was used for all stations and sites ALEX and MARZ were included in the analyses.

Throughout the text the term 'LCZ' is used when referring to individual LCZ classes, i.e., all regions of one specific LCZ. This does not necessarily mean that these LCZ are one contiguous geographical area or zone.

Fig. 1 provides an overview of the different LCZ and their main characteristics. For a more detailed description of each LCZ the reader is kindly referred to Tables 2, 3, and 4, and supporting material in Stewart and Oke (2012).

2.5 Calculations and statistics

The thermal regime of LCZ and T differences between LCZ were analysed for each month of the year for different time periods during the day. Daily means (all 24 hours of a day) were computed as well as means during day- and night-time periods.

The daytime period (1300–1600 UTC+1) corresponds to the hours of a day when daily maximum T most frequently occurs in Berlin and when the (urban) atmosphere is well mixed, resulting in typically small T differences between sites (Oke, 1982; Christen and Vogt, 2004; Chow and Roth, 2006; Erell and Williamson, 2007). The night-time period (4–7 hours after sunset UTC+1) covers the time when local-scale T differences between sites are well established, which typically occurs between three to five hours after sunset, and lasts until sunrise (Oke and Maxwell, 1975; Oke, 1982; Unger et al., 2001; Chow and Roth, 2006; Erell and Williamson, 2007). Moreover, cooling rates are similar among different urban environments during this period (Holmer et al., 2007; Holmer et al., 2013; Leconte et al., 2016), hence spatial T differences remain fairly constant (Holmer et al., 2007; Fenner et al., 2014; Hu et al., 2016). The end of the night-time period also takes into account the shortest night during the year, which is just over seven hours long in Berlin.

At first, mean T per time period for each day of the year was calculated for each site (CWS and reference stations) if at least 20 valid values for daily averages and three values for the four-hourly time periods were available. Then, if a station provided valid data for at least 80 % of the days of a month, monthly average values per time period were calculated for this station. Afterwards, all stations were grouped to their respective LCZ.

To test whether significant differences in ${}^{\mathrm{cws}}T$ between LCZ existed, a one-way analysis of variance (ANOVA) was carried out per month and each time period using the non-parametric Kruskal-Wallis test (Kruskal and Wallis, 1952). For significant results ($p < 0.05$) pairwise two-sided Wilcoxon-Mann-Whitney post-hoc tests (Wilcoxon, 1945; Mann and Whitney, 1947) were carried out for each LCZ pair to detect significant differences. Both tests do not require a normal distribution of the tested data, and were thus selected over parametric tests such as the t test (Gosset, 1908). The significance level of the post-hoc tests was adjusted according to the Holm-Bonferroni method (Holm, 1979) to account for multiple tests and to achieve an overall significance level of $p < 0.05$.

Differences in T between two LCZ x and y, i.e., $\Delta T_{\mathrm{LCZ}\ x\text{–}LCZy}$ were calculated as follows:

$$\Delta T_{\mathrm{LCZ}\ x\text{–}\mathrm{LCZ}\ y} = T_{\mathrm{LCZ}\ x} - T_{\mathrm{LCZ}\ y}$$

where $T_{\mathrm{LCZ}\ x}$ and $T_{\mathrm{LCZ}\ y}$ are the mean values across all sites in LCZ x and LCZ y, respectively. This follows the recommendations of Stewart et al. (2014) that spatial averages should be used to obtain locally representative values. Calculations of $\Delta T_{\mathrm{LCZ}\ x\text{–}\mathrm{LCZ}\ y}$ were carried out for CWS and reference data separately and not merging the two data sets, following the approach of Chapman et al. (2016).

Additionally, $\Delta T_{\mathrm{LCZ}\ x\text{–}\mathrm{LCZ}\ y}$ were analysed for 'ideal' weather situations with clear skies, calm winds, and no precipitation. These conditions favour the development of pronounced thermal differences among different environments (Magee et al., 1999; Morris et al., 2001; Erell and Williamson, 2007; Stewart and Oke, 2012; Arnds et al., 2017; van Hove et al., 2015). For this analysis, only days with mean wind speed $\leq 2\,\mathrm{m\,s^{-1}}$, mean cloud cover ≤ 2 octas, and precipitation $< 1\,\mathrm{mm\,d^{-1}}$ for the preceding day and the day itself at station TEGL were used. In total, seven of these 'ideal' days occurred in 2015 (10 April, 04 June, 29 August, 27 and 28 September, 01 October, 11 November), and mean T was calculated for each station and time period, if valid data for at least six of these seven days existed. Like for the monthly analysis, an ANOVA (Kruskal-Wallis test) and subsequent pairwise two-sided Wilcoxon-Mann-Whitney tests were carried out to detect significant differences between LCZ.

3 Results

3.1 LCZ distribution and coverage by CWS

Within the ROI LCZ A (dense trees) and D (low plants) exhibit highest spatial coverage. Together they cover more than 50 % of the area (Fig. 2). With approximately 20 % of the area, LCZ 6 (open low-rise) is the third most dominant LCZ, extending around Berlin's city core into the outskirts of the city and beyond the administrative border of Berlin. Within the city's core, LCZ 2 (compact midrise) dominates over LCZ 5 (open midrise), though the latter contributes to a higher overall spatial coverage. LCZ 1 (compact high-rise), 3 (compact low-rise), 7 (lightweight low-rise), 10 (heavy industry), and E (bare rock or paved) do not occur (Fig. 2).

The number of CWS with valid data within the ROI varies per month and time period (Fig. 4). In general, three effects can be noted: (1) more CWS with valid data are available for night-time than for daytime or daily mean values, (2) the number of stations increases over the course of the year 2015, and (3) the number of CWS is highly differing between LCZ. Overall, September has the highest number of available CWS ($n = 451$) for night-time, and December for daytime ($n = 318$) and daily mean values ($n = 375$). A distinctive drop in number of CWS in October can also be noted for daily mean and daytime values, while CWS availability during night-time is fairly constant throughout the year. Spatial distribution of CWS within individual LCZ is fairly uniform across the ROI (not shown).

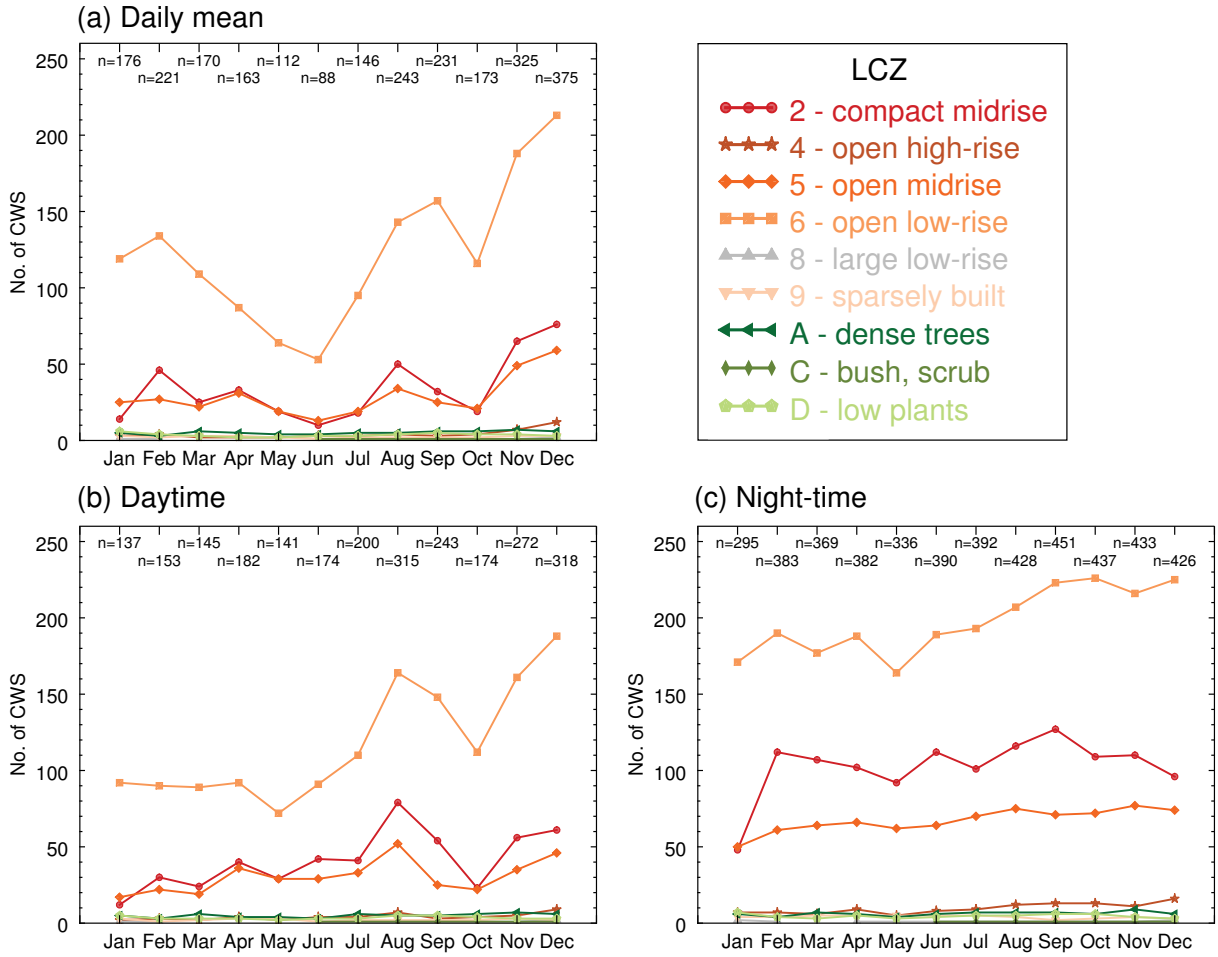

Figure 4: Number of Netatmo citizen weather stations (CWS) per local climate zone (LCZ) and month for (a) daily mean values, (b) daytime (1300–1600 UTC+1), and (c) night-time (4–7 hours after sunset UTC+1) in 2015. The total number of stations per month is written below the top horizontal axis.

At least one CWS with valid data is located in each existing LCZ except in LCZ B (scattered trees), F (bare soil or sand), and G (water). Much more stations are, however, found in 'urban' LCZ (1 to 10) than in 'natural' LCZ (A to G). The highest number of available CWS per LCZ is in LCZ 6, irrespective of month and time period (Fig. 4), with at least 48 % of all CWS in the study region (highest number: 226 CWS in LCZ 6 in October during night-time, i.e., 52 % of all CWS). LCZ 2 and LCZ 5 show similar numbers of CWS per month for daily mean and daytime values (up to 79 CWS for LCZ 2 in August), while for night-time nearly twice the amount of stations is located in LCZ 2 compared to LCZ 5 (Fig. 4). Between one and sixteen CWS are located in the remaining LCZ.

3.2 Monthly and annual air temperature characteristics per LCZ in 2015

Mean monthly T per LCZ is shown in Table 2 for CWS and reference stations, showing the annual cycle of T with highest values in August and lowest values in February for all LCZ. Highest mean monthly ${}^{\mathrm{cws}}T$ is recorded in LCZ 2 for all months except January and October. Comparison with mean monthly ${}^{\mathrm{ref}}T$ reveals a generally good agreement of $\leq \pm 0.3$ K between ${}^{\mathrm{cws}}T$ and ${}^{\mathrm{ref}}T$ for all LCZ in winter months (December, January, and February), March, April, and September. The months May to August show larger differences of up to +1.0 K (LCZ 4) between mean monthly ${}^{\mathrm{cws}}T$ and ${}^{\mathrm{ref}}T$. Taking the spatial standard deviations (SD) per LCZ and month into account, it can be noted that the ranges of mean $T \pm$ SD of CWS and reference data overlap (Table 2). In those cases, a clear distinction between mean monthly ${}^{\mathrm{cws}}T$ and ${}^{\mathrm{ref}}T$ cannot be made. Overall, LCZ D (low plants) and LCZ 2 show best agreements between mean monthly ${}^{\mathrm{cws}}T$ and ${}^{\mathrm{ref}}T$ (Table 2), while for LCZ 4 and LCZ 6 the deviations between CWS and reference data are $> \pm 0.3$ K in seven and six months, respectively.

Annual mean values of ${}^{\mathrm{cws}}T$ and ${}^{\mathrm{ref}}T$ per LCZ show overall small differences ($\leq \pm 0.2$ K) to one another, except for LCZ 4 and LCZ 6 (Table 2). LCZ 2 also exhibits highest ${}^{\mathrm{cws}}T$ and ${}^{\mathrm{ref}}T$ in the annual mean, while lowest T in 2015 is recorded in LCZ B (${}^{\mathrm{ref}}T$).

Table 2: Mean monthly and annual air temperature $T \pm$ standard deviation (SD) per local climate zone (LCZ) in Berlin and surroundings in 2015. First line per LCZ: Netatmo citizen weather stations, second line (italic): reference stations. Spatial SD was calculated across all sites per LCZ, no standard deviation is given if only one site per LCZ was available. Annual mean was calculated from mean monthly values if data were available for at least ten months in 2015.

	$T \pm$ SD (°C)												
LCZ	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	2015
2	3.4±0.4	3.0±0.6	6.8±0.5	10.6±0.7	15.2±0.5	18.2±0.5	20.9±0.6	23.4±0.5	15.5±0.4	9.5±0.5	8.8±0.7	7.9±0.5	11.9
	3.5±0.4	*2.6±0.4*	*6.9±0.4*	*10.5±0.4*	*14.6±0.3*	*17.7±0.3*	*20.9±0.4*	*23.1±0.6*	*15.4±0.3*	*9.6±0.2*	*8.6±0.2*	*7.8±0.2*	*11.8*
4	3.5±0.6	2.5±0.7	6.4±0.3	10.2±0.9	14.8±0.6	17.9±0.4	20.9±0.3	23.1±0.6	15.0±0.1	9.7±0.2	8.5±0.5	7.5±0.5	11.7
	3.2	*2.1*	*6.3*	*9.8*	*13.8*	*17.1*	*20.3*	*22.6*	*14.9*	*8.9*	*7.9*	*7.3*	*11.2*
5	3.5±0.4	2.6±0.5	6.7±0.6	10.2±0.7	14.6±0.6	18.1±0.5	20.9±0.6	23.0±0.6	15.2±0.6	9.3±0.6	8.4±0.7	7.7±0.5	11.7
	–	–	–	–	*13.9*	*17.3*	*20.4*	*22.6*	*14.9*	*9.1*	*8.3*	*7.7*	–
6	3.1±0.4	2.0±0.6	6.1±0.5	9.5±0.7	14.1±0.6	17.6±0.4	20.3±0.5	22.4±0.6	14.5±0.6	8.6±0.5	7.9±0.7	7.2±0.6	11.1
	2.9±0.2	*1.7±0.3*	*6.0±0.2*	*9.3±0.2*	*13.2±0.3*	*16.7±0.3*	*19.7±0.3*	*21.9±0.5*	*13.9±0.4*	*8.4±0.3*	*7.5±0.3*	*6.9±0.3*	10.7
8	2.7±0.1	1.5	–	–	–	–	–	–	–	–	8.3	7.7±0.1	–
	–	–	–	–	–	–	–	–	–	–	–	–	–
9	2.9±0.3	1.6±0.2	5.8±0.1	9.1±0.3	13.4±0.1	17.3	20.2±0.2	21.9±0.1	13.7±0.2	8.1±0.0	7.1±0.3	6.8±0.5	10.7
	–	–	–	–	–	–	–	–	–	–	–	–	–
A	3.1±0.8	1.8±0.6	6.1±0.4	9.4±0.5	13.8±0.5	17.1±0.4	19.8±0.4	21.5±0.3	14.2±0.4	8.8±0.2	7.6±0.5	7.0±0.4	10.8
	2.8±0.5	*1.5±0.9*	*5.8±1.0*	*9.2±0.8*	*13.3±0.6*	*16.7±0.5*	*19.7±0.4*	*21.9±0.5*	*13.9±0.8*	*8.4±0.9*	*7.4±0.7*	*6.7±0.7*	*10.6*
B	–	–	–	–	–	–	–	–	–	–	–	–	–
	2.6	*1.0*	*5.2*	*8.4*	*12.6*	*15.8*	*19.1*	*20.9*	*13.0*	*7.5*	*6.8*	*6.4*	*9.9*
C	–	–	–	–	–	17.1	19.8	21.3	13.7	8.3	7.5	7.1	–
	–	–	–	–	–	–	–	–	–	–	–	–	–
D	3.0±0.3	1.6±0.2	5.8±0.0	9.1±0.8	13.6±0.0	17.6±0.2	20.0±0.3	22.0±0.3	14.0±0.2	8.2±0.4	7.5±0.4	6.8±0.2	10.8
	2.8	*1.4*	*5.7*	*9.0*	*13.2*	*16.8*	*20.0*	*22.2*	*14.2*	*8.3*	*7.5*	*6.7*	*10.6*

3.3 Intra-LCZ variability

Boxplots per LCZ and month with mean T for each measurement site allow a detailed analysis of intra-LCZ T variability. Fig. 5 and Fig. 6 show the distribution of T among CWS and reference stations for the months February and August, respectively, and reveal a large variability for most LCZ. In general, summer months show higher intra-LCZ ${}^{\mathrm{cws}}T$ variability than winter months, and night-time hours larger variability than daytime or daily mean values (Fig. 5, Fig. 6). August is an exception to this with similar intra-LCZ variability during daytime and night-time, though the inter-quartile range is smaller for daytime than for night-time (Fig. 6). Also, with a higher number of CWS per month, higher intra-LCZ variability is found. A good example is LCZ 6 with most CWS and the largest range between minimum and maximum ${}^{\mathrm{cws}}T$ among all LCZ in most months. This range is as high as 6.6 K in August for daytime and 6.1 K in May for night-time (not shown). Likewise, the LCZ with second and third most CWS, i.e., LCZ 2 and LCZ 5, show considerable spread among monthly mean ${}^{\mathrm{cws}}T$ for all three time periods. It is also worth noting that the inter-quartile ranges in Fig. 5 and Fig. 6, as well as the standard deviation (spatially) in Table 2 for LCZ 2, 5, and 6 are broadly similar, showing that all three LCZ are much alike in terms of spatial T variability.

Fig. 5 and Fig. 6 also show overall agreement but also certain differences between CWS and reference data. While during daytime mean ${}^{\mathrm{ref}}T$ is slightly higher than mean ${}^{\mathrm{cws}}T$ for most LCZ during the warmer months of the year, for daily and night-time periods mean ${}^{\mathrm{ref}}T$ is mostly lower than mean ${}^{\mathrm{cws}}T$. All ${}^{\mathrm{ref}}T$ values are within the range of ${}^{\mathrm{cws}}T$ except for LCZ A and LCZ D (Fig. 5, Fig. 6).

3.4 Inter-LCZ differences

Significant differences in ${}^{\mathrm{cws}}T$ ($p < 0.05$) between at least two LCZ can be found for all months in 2015 for daily mean and night-time values. These differences are not only present between 'urban' LCZ and 'rural' surroundings, i.e., an UHI, but also between LCZ located within the city, i.e., inner-city T differences. For daytime values the ANOVA reveals only three months (September, November, and December) with significant differences.

Highest mean monthly $\Delta^{\mathrm{cws}}T_{\mathrm{LCZ}\,x-\mathrm{LCZ}\,y}$ are found for night-time, while daily mean and daytime values are lower. Fig. 7 shows $\Delta^{\mathrm{cws}}T_{\mathrm{LCZ}\,2-\mathrm{LCZ}\,y}$ along the annual cycle for the three time periods as LCZ 2 is the LCZ with highest mean monthly T and with the highest number of significant $\Delta^{\mathrm{cws}}T_{\mathrm{LCZ}\,x-\mathrm{LCZ}\,y}$. Additionally, the same analyses were conducted with all other possible LCZ pairs, which resulted in a lower number of significant differences. Especially high values of $\Delta^{\mathrm{cws}}T_{\mathrm{LCZ}\,2-\mathrm{LCZ}\,D}$ are found for the months May and August during night-time (> 3.5 K, Fig. 7c) and daily mean values. LCZ 2 shows significant ${}^{\mathrm{cws}}T$ differences to any other LCZ except LCZ 4 and LCZ 8 (containing only one or two CWS). For daytime, significant $\Delta^{\mathrm{cws}}T_{\mathrm{LCZ}\,2-\mathrm{LCZ}\,y}$ are only detected in November and December, though some months exhibit larger differences than values in these months (Fig. 7b).

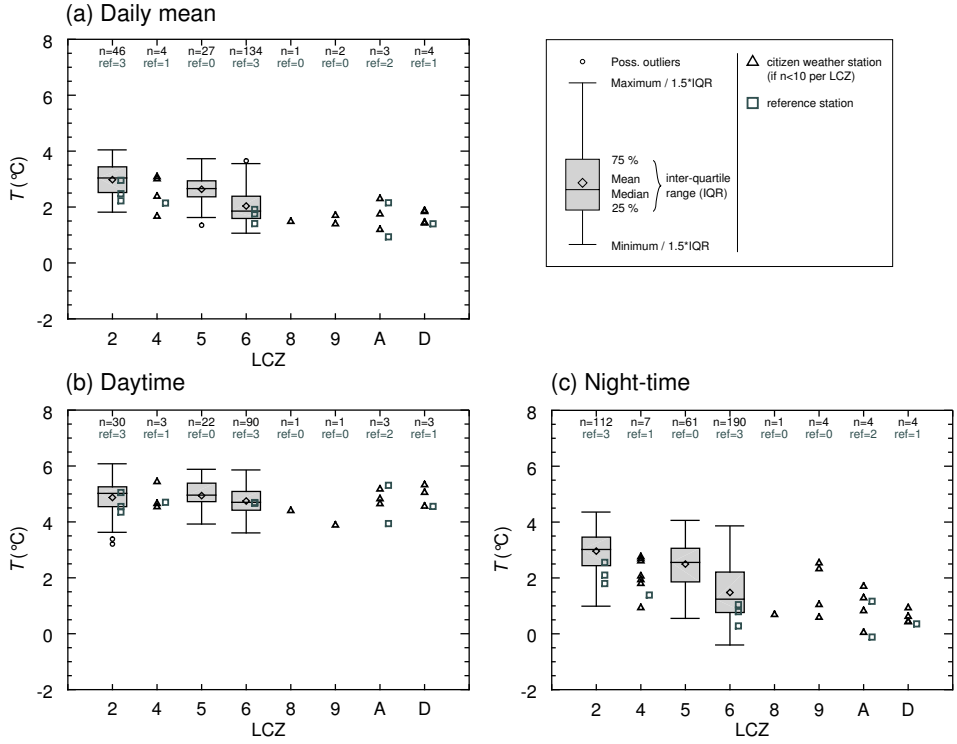

Figure 5: Air temperature T distribution for (a) daily mean values, (b) daytime (1300–1600 UTC+1) and (c) night-time (4–7 hours after sunset UTC+1) per local climate zone (LCZ) for Netatmo citizen weather stations (CWS) and reference stations in February 2015. Each distribution consists of mean values per station and time period in the respective LCZ. The number of sites per LCZ (CWS = n, reference = ref) is written below the top horizontal axis.

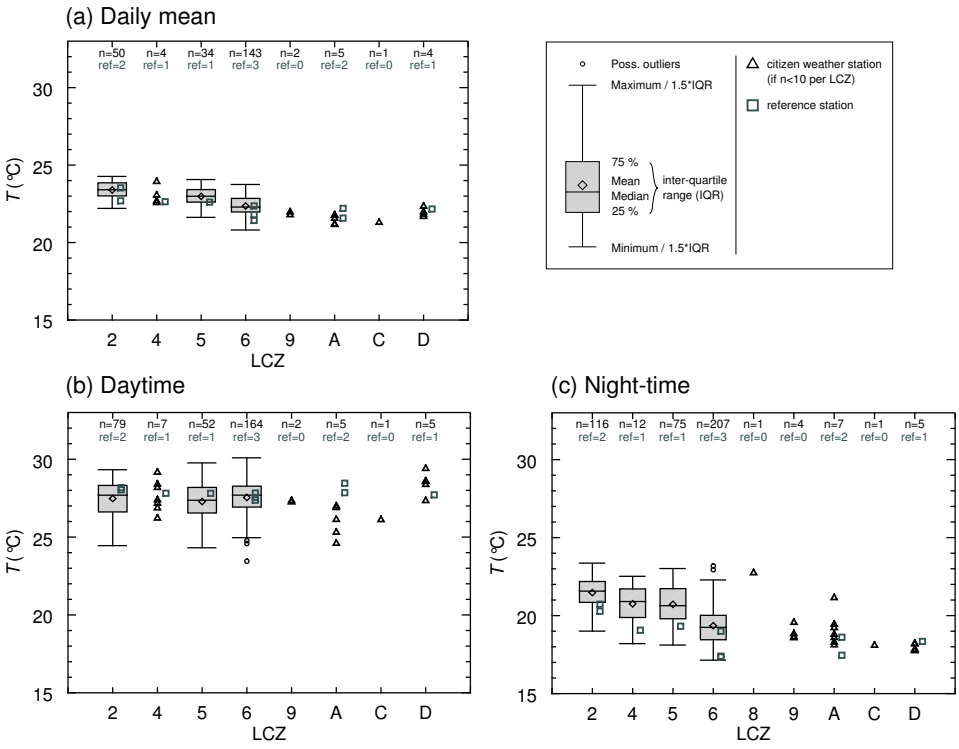

Figure 6: Same as Fig. 5 but for August 2015.

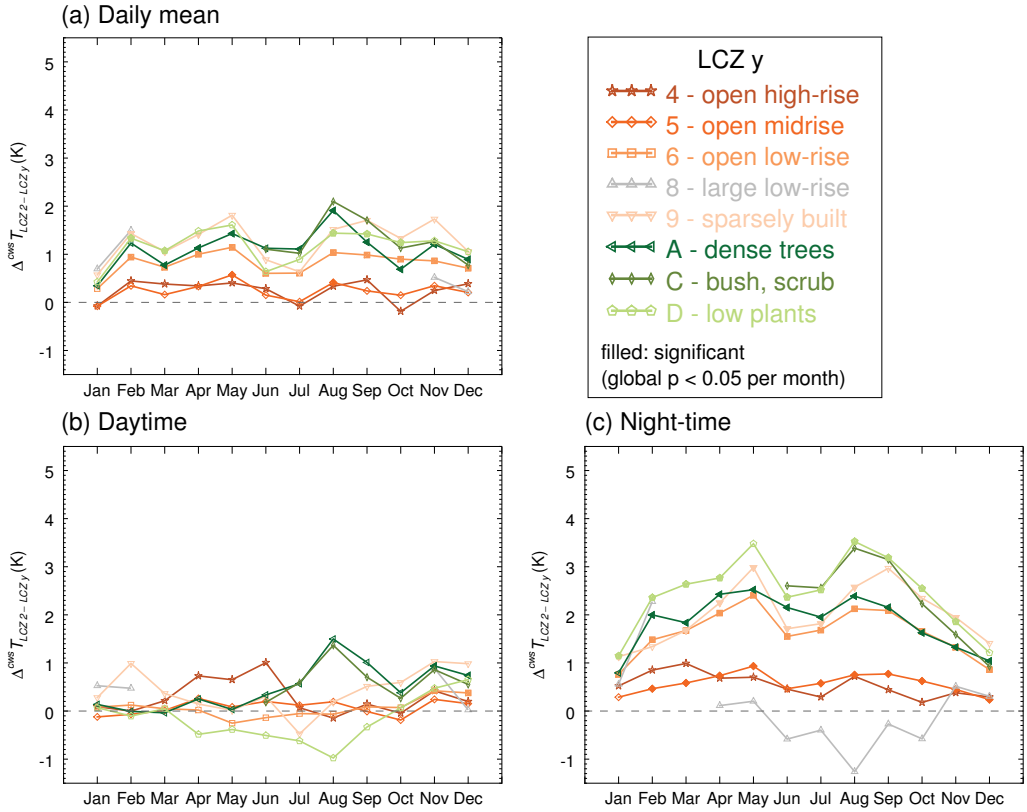

Figure 7: Annual cycle of air temperature differences ΔT for (a) daily mean values, (b) daytime (1300–1600 UTC+1) and (c) night-time (4–7 hours after sunset UTC+1) between local climate zone (LCZ) 2 (compact midrise) and different LCZ as observed by Netatmo citizen weather stations in and around Berlin in 2015. Values are mean values across all sites for the respective LCZ and month.

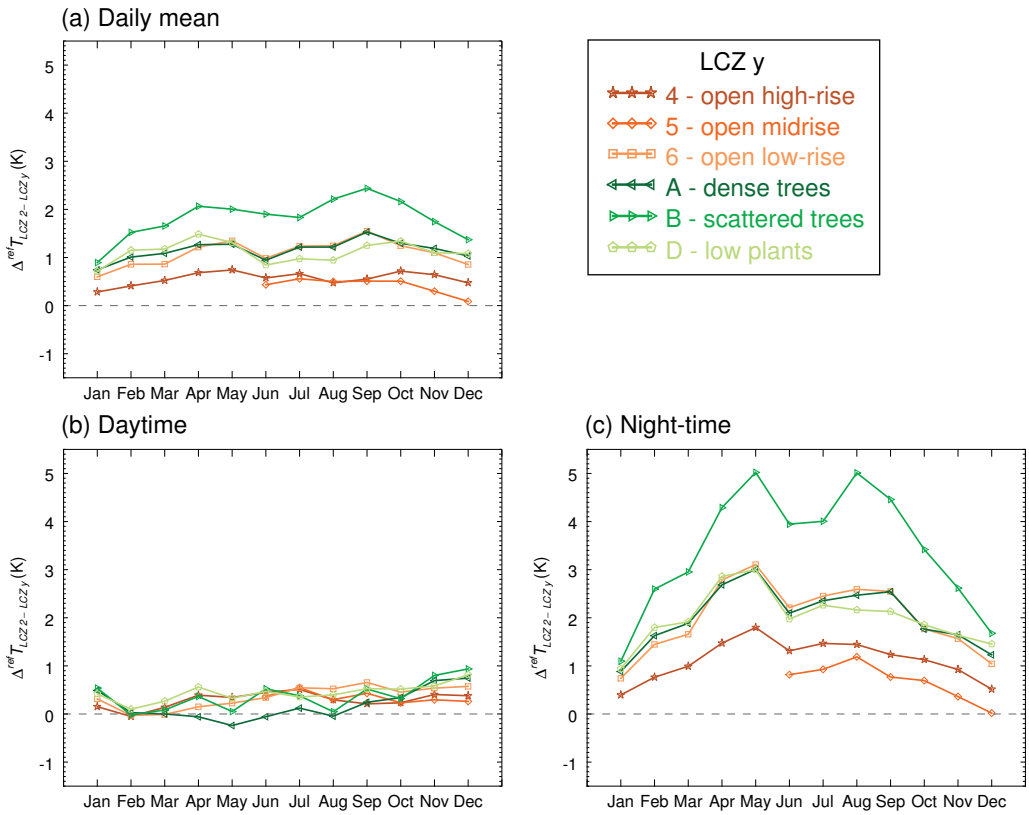

Figure 8: Same as Fig. 7 but for reference stations.

Table 3: Annual mean air temperature differences ΔT between local climate zones (LCZ) in Berlin and surroundings in 2015. Values were calculated as row minus column ($\Delta T_{\text{LCZ row–LCZ column}}$). First line per field: Netatmo citizen weather stations, second line (italic): reference stations. Standard deviation (SD) was calculated along the temporal dimension on the basis of daily values. Annual mean was calculated if data were available for at least ten months.

	Annual mean $\Delta T_{\text{LCZ row–LCZ column}}$ ±SD (K)						
LCZ	2	4	5	6	9	A	B	D
2		0.2±0.4	0.2±0.3	0.8±0.6	1.2±0.7	1.0±0.6	–	1.0±0.7
		0.6±0.4	–	*1.1±0.6*	–	*1.1±0.6*	*1.8±1.1*	*1.1±0.6*
4			0.0±0.3	0.6±0.5	0.9±0.6	0.8±0.5	–	0.8±0.6
			–	*0.5±0.4*	–	*0.6±0.4*	*1.2±0.9*	*0.5±0.4*
5				0.5±0.4	0.9±0.5	0.7±0.4	–	0.8±0.5
				–	–	–	–	–
6					0.4±0.3	0.2±0.3	–	0.3±0.3
					–	*0.1±0.2*	*0.7±0.6*	*0.0±0.4*
9						−0.2±0.2	–	−0.1±0.3
						–	–	–
A							–	0.1±0.5
							0.6±0.6	*0.0±0.4*
B								–
								−0.7±0.7
D								

The annual cycle of $\Delta^{\text{ref}}T_{\text{LCZ 2–LCZ }y}$ is given in Fig. 8, showing positive values during all months for daily mean (Fig. 8a) and night-time (Fig. 8c). Daytime $\Delta^{\text{ref}}T_{\text{LCZ 2–LCZ }y}$ are similar for most LCZ, always $< \pm 1$ K, and mainly positive (Fig. 8b). Largest differences are found for $\Delta^{\text{ref}}T_{\text{LCZ 2–LCZ }B}$ for daily mean and night-time values, especially in the months May and August for night-time (Fig. 8c).

Table 3 provides an overview of all available annual mean $\Delta T_{\text{LCZ }x\text{–LCZ }y}$ in 2015, displaying that largest differences between LCZ in CWS data amount to $\Delta^{\text{cws}}T_{\text{LCZ 2–LCZ }A/D} = 1.0$ K and for reference data to $\Delta^{\text{ref}}T_{\text{LCZ 2–LCZ }B} = 1.8$ K. Generally, higher $\Delta T_{\text{LCZ }x\text{–LCZ }y}$ are found for LCZ that are structurally dissimilar, e.g., $\Delta T_{\text{LCZ 2–LCZ }A}$. Table 3 also shows an overall good agreement between CWS and reference data with all differences between the two datasets $\leq \pm 0.4$ K.

3.5 Ideal weather situations

Under ideal weather conditions for the formation of local-scale ΔT, intra-LCZ variability is again highest for LCZ 6 (Fig. 9). LCZ 5, however, shows a similar range for night-time. The range for LCZ 2 is smaller and only slightly higher than for LCZ 4, though much more CWS are located in LCZ 2 ($n = 56$) than in LCZ 4 ($n = 4$). As for monthly mean values, intra-LCZ T variability is higher during night-time than during the day (not shown). Fig. 9 also shows, as seen for the monthly analyses in Fig. 5c and Fig. 6c, that T at reference stations is generally lower than at CWS during night-time.

All available $\Delta T_{\text{LCZ }x\text{–LCZ }y}$ during night-time under ideal weather conditions is presented in Table 4 for CWS and reference stations. Though values are higher than monthly or annual mean $\Delta T_{\text{LCZ }x\text{–LCZ }y}$, significant (overall $p < 0.05$) $\Delta^{\text{cws}}T_{\text{LCZ }x\text{–LCZ }y}$ can only be found for four LCZ pairs (Table 4). The highest significant $\Delta^{\text{cws}}T_{\text{LCZ }x\text{–LCZ }y}$ is for LCZ 2–6 (3.6 ± 1.1), two 'urban' LCZ. Comparing $\Delta^{\text{cws}}T_{\text{LCZ }x\text{–LCZ }y}$ and $\Delta^{\text{ref}}T_{\text{LCZ }x\text{–LCZ }y}$ shows that differences between the two data sets are small ($\leq \pm 0.2$ K) only for three pairs. For all other $\Delta T_{\text{LCZ }x\text{–LCZ }y}$ during night-time and under ideal weather conditions differences between CWS and reference data are at least ±0.7 K and up to 2.9 K (Table 4).

4 Discussion

4.1 LCZ coverage and measurement site locations

The CWS data set has some characteristics that are unusual as compared to classical meteorological measurement time series, especially changing number of measurement sites providing valid data throughout time (and space). Particularly remarkable are the large differences in the number of available CWS between daytime, night-time, and daily periods, and the strong variation across the months, particularly for daily and daytime values. Main reasons for these characteristics are the filter techniques applied by Meier et al. (2017). These filters include a daytime 'radiation' filtertofilter out unrealistically high $^{\text{cws}}T$, since Netatmo CWS are especially prone to radiation errors due to the compact construction with an unventilated aluminium cylinder as outer shell (Chapman et al., 2016; Meier et al., 2017). Such design flaws in view of obtaining accurate measurements are common among amateur weather stations (Bell et al., 2015), and must thus be considered in data quality assessments. This 'radiation' filter was applied by calcu-

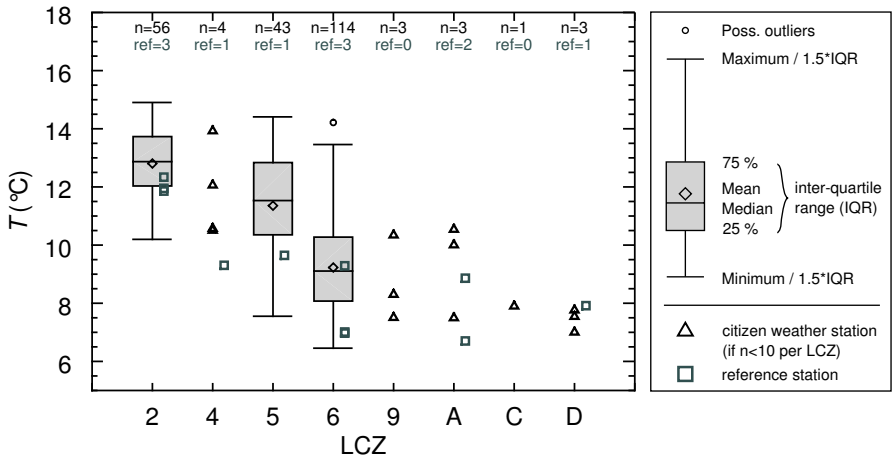

Figure 9: Air temperature T distribution during night-time (4–7 hours after sunset UTC+1) per local climate zone (LCZ) for Netatmo citizen weather stations (CWS) and reference stations during 'ideal' days in 2015 (seven days). Each distribution consists of mean values per station in the respective LCZ. The number of sites per LCZ (CWS = n, reference = ref) is written below the top horizontal axis. 'Ideal' days are days with mean wind speed $\leq 2\,\mathrm{m\,s^{-1}}$, mean cloud cover ≤ 2 octas, and precipitation $< 1\,\mathrm{mm\,d^{-1}}$ the preceding day and the day itself.

Table 4: Mean night-time air temperature differences ΔT between local climate zones (LCZ) in Berlin and surroundings during 'ideal' days in 2015 (seven days). Values were calculated as row minus column ($\Delta T_{\mathrm{LCZ\,row-LCZ\,column}}$). First line per field: Netatmo citizen weather stations, second line (italic): reference stations. Bold numbers mark significant differences (overall $p < 0.05$). Standard deviation (SD) was calculated along the temporal dimension, i.e., for the seven 'ideal' days. Grey-shaded fields mark the number of measurement sites for the respective LCZ. 'Ideal' days are days with mean wind speed $\leq 2\,\mathrm{m\,s^{-1}}$, mean cloud cover ≤ 2 octas, and precipitation $< 1\,\mathrm{mm\,d^{-1}}$ the preceding day and the day itself.

	Night-time $\Delta T_{\mathrm{LCZ\,row-LCZ\,column}} \pm$ SD (K), 'ideal' days								
LCZ	2	4	5	6	9	A	B	C	D
2	56	1.0±0.2	**1.5±0.5**	**3.6±1.1**	4.1±1.4	3.5±1.0	–	4.9±1.1	5.4±1.2
	3	*2.7±1.4*	*2.4±0.8*	*4.3±1.4*	–	*4.3±1.6*	*7.9±1.6*	–	*4.1±1.4*
4		4	0.4±0.6	2.5±1.1	3.0±1.4	2.4±1.1	–	3.9±1.2	4.3±1.3
		1	*−0.3±1.1*	*1.5±0.8*	–	*1.5±1.3*	*5.2±1.2*	–	*1.4±0.8*
5			43	**2.1±0.6**	2.6±0.8	2.0±0.6	–	3.5±0.6	**3.9±0.7**
			1	*1.9±0.8*	–	*1.9±1.0*	*5.5±1.1*	–	*1.7±1.1*
6				114	0.5±0.4	−0.1±0.4	–	1.3±0.2	1.8±0.3
				3	–	*0.0±0.6*	*3.6±0.6*	–	*−0.2±0.8*
9					3	−0.6±0.6	–	0.8±0.6	1.3±0.4
					–	–	–	–	–
A						3	–	1.4±0.4	1.9±0.7
						2	*3.6±0.6*	–	*−0.1±1.1*
B							–	–	–
							1	–	*−3.8±0.7*
C								1	0.5±0.4
								–	–
D									3
									1

lating the difference between hourly T at a specific CWS and mean T across the reference network at that hour (filter C2 in Meier et al. (2017)). If the positive difference was larger than three times the standard deviation in T across the entire reference network at that hour, the CWS measurement was filtered out. The analyses show that this filter markedly filters out CWS located in LCZ 2 and LCZ 5, explaining the large difference in the number of CWS in these LCZ between the daytime and night-time period (Fig. 4). In the absence of detailed meta data for CWS it cannot be said with certainty, but more stations within these LCZ could be poorly placed and exposed to radiation. It was checked whether CWS filtered during daytime show high T at night, which could indicate a poor siting, but no evidence for this was found. The fact that the number of CWS for daily and daytime values is much lower in October than during the months before and after is also likely due to the 'radiation' filter. Shedding of the leaves of trees usually happens in October in Berlin, while in September trees are still foliated, providing shadow and hence also shadowing CWS. In November, on the other hand, solar radiation and solar elevation rapidly drop. Thus, radiation errors of CWS are lower and occur less often.

As shown in Section 3.1, LCZ coverage in the study area is dominated by 'natural' LCZ, mainly LCZ A and D, followed by LCZ 6. Contrary to that, Netatmo CWS are located mainly in 'urban' LCZ, coming as no surprise as these are the areas were most users of CWS live. This gives proof of the fact that CWS can be of additional value for studies focusing on urban areas, since coverage in urban areas by traditional measurement networks is usually sparse (Muller et al., 2013). Moreover, when investigating intra-urban differentiation of climate conditions by, e.g., application of the LCZ concept, only a limited number of cities have, so far, implemented a network of stations that capture all urban environments, and mostly with only one station per LCZ. CWS in Berlin cover all classified 'urban' LCZ with at least one station per LCZ, while other LCZ are represented by more than 50 stations. It becomes clear that crowdsourced data from CWS can provide additional information to study, e.g., intra-LCZ variability of atmospheric conditions. Moreover, the large number of CWS enables the application of statistical tests in the detection of significant differences between LCZ pairs under consideration of spatial variability within individual LCZ. Reference networks, on the contrary, commonly have only one or very few sites per LCZ, which prevents statistical testing of inter-LCZ differences. However, many CWS in one LCZ are mainly in densely built-up environments (LCZ 2 or LCZ 5) or in populated regions with large spatial coverage (LCZ 6) in Berlin. Natural environments outside the urban area (e.g. LCZ A) or within the city, i.e., urban parks, as well as industrial areas (LCZ 8 or LCZ 10, the latter not being present in Berlin) are underrepresented or not represented at all in the CWS data set. Due to the low costs of these sensors this limitation could be overcome by researchers by installing additional sensors in LCZ with few CWS at the moment. Chapman et al. (2016) reported similar attributes of Netatmo data in London. Hence, this type of crowdsourced atmospheric data is far from being perfect in terms of representing all environments of an urban region. Reference data from professionally operated measurement networks are indispensable to cover such environments. While this imbalance is likely to remain, the increasing total number of CWS might help to obtain a more representative picture in less populated LCZ as well.

For existing standard measurement networks the question of how representative the given location of a station is for specific urban environments, such as represented by LCZ, arises (Lelovics et al., 2014). From the initial set of reference stations several were excluded from the analyses due to large spatial heterogeneity in the local surroundings in the WUDAPT-LCZ classification. It must be kept in mind, though, that this LCZ-classification methodology also has some uncertainties associated with it (cf. Appendix B). More importantly, however, the existing reference network grew throughout the past decades, some stations exist already more than 100 years (stations DAHL and POTS), and locations of measurement sites were not selected following the LCZ concept. This would be necessary to obtain clear LCZ signals (Lehnert et al., 2014). Nonetheless, it is apparent that a careful site location in homogeneous surroundings is essential for representative measurements, especially in urban regions (Oke, 2006).

4.2 Uncertainties in CWS measurements

Throughout the analysis of monthly T characteristics it was shown that mean ${}^{\mathrm{cws}}T$ is, in general, higher than ${}^{\mathrm{ref}}T$. While in most cases mean deviation is ≤ 0.3 K, it is notably higher in some months and for some LCZ, namely LCZ 4, 5, and 6. Due to the intra-LCZ T variability the deviation between single CWS and reference stations is several K. Chapman et al. (2016) described similar deviations between Netatmo CWS and standard measurement stations of up to several K, also depending on weather conditions. Specifically, weather conditions with low wind speeds and low cloud cover promoted deviations between CWS and reference data (Chapman et al., 2016). The analyses confirm this finding in such a way that while under the investigated 'ideal' days deviations in T and $\Delta T_{\mathrm{LCZ}\,x-\mathrm{LCZ}\,y}$ between CWS and reference data are stronger than in monthly or annual mean values, which include all weather situations.

One of the biggest challenges when dealing with crowdsourced CWS data is the issue of siting of these stations (Chapman et al., 2016). Presumably, most Netatmo CWS are not installed following standards for meteorological observations in cities (Oke, 2006), but situated at locations that are prone to microscale effects, such as courtyards, and close to building walls and on balconies. Hence, it can be assumed that most CWS do not measure locally representative T conditions in open environments or in regions of an urban street canyon where the air is well mixed. These regions are typically in the middle of the canyon and at heights of some metres above the ground. Close to building walls and to the ground, T might be notably higher; an effect that is particularly pronounced during daytime and smaller at night (Nakamura and Oke, 1988; Niachou et al., 2008). This would result in higher T in CWS data compared to reference data, which would explain the positive deviations in monthly mean ${}^{\mathrm{cws}}T$ in comparison to ${}^{\mathrm{ref}}T$. However, the analyses (Fig. 5, Fig. 6) revealed higher deviations between CWS and reference data during night-time, which could be due to more filtering ('radiation' filter, see above) during daytime. Detailed measurements in a variety of local environments with CWS and standard sensors in close proximity to building walls, as well as further away from buildings would enable a better understanding of the observed deviations.

A contrary effect on T data, i.e., lower values in CWS in comparison with reference data, is related to the measurement height of CWS. Heights can range from locations close to the ground up to several tens of metres above ground level. Higher measurement locations

within a street canyon would result in lower temperatures compared to locations closer to the ground for neutral or unstable conditions, which are typically prevailing in cities (Christen and Vogt, 2004). Since detailed meta data for CWS are not available, height and site conditions of CWS are unknown, and thus a height correction that takes into account the actual measurement height is not possible. It was only possible to correct for the differences in terrain elevation, assuming a uniform height of CWS of 2 m. As Chapman et al. (2016) noted, this issue of missing meta data "can only realistically be overcome by including a voluntary meta data section online for end users to document, with images, the siting and exposure of their instruments". Meta data of stations are thus a key challenge when using CWS data, and this needs particular attention in the future. It could be shown, however, that spatial mean T across all CWS in one LCZ is comparable to reference data, as well as spatial differences between LCZ. This supports previous studies, saying that spatial mean T should be used (Stewart et al., 2014; Leconte et al., 2015), and corresponds well to the notion that a larger atmospheric scale, i.e., the local scale, is a combination of smaller scale features, i.e., of micro-climatic effects (Oke, 2006). Hence, results from single or few CWS must be treated carefully for analyses of local-scale atmospheric conditions since crowdsourced CWS might measure microscale conditions that are unknown in absence of detailed meta data. The large number of stations, however, enables calculations of locally representative mean values, while at the same time allowing for a study of intra-LCZ T variability.

Further reasons for deviations between CWS and reference data could be more sensor-specific, such as sensor accuracy or the unventilated case of the Netatmo unit. However, Meier et al. (2017) showed with climate chamber experiments that the sensor accuracy of Netatmo CWS is within the specified accuracy range of ±0.3 K with only a slight warm bias (~ 0.5 K) around 0 °C, which is in the range of the deviations between CWS and reference data that were found. This being said, the unventilated case of Netatmo CWS is an important issue with these stations. The applied data filter techniques (Meier et al., 2017) address this issue, and it is assumed that the data presented here are mostly free of problems that could result from the compact construction of this type of CWS, namely radiation errors (Nakamura and Mahrt, 2005). In fact, the applied daytime 'radiation' filter applied in Meier et al. (2017) might be too strict, as discussed above.

Keeping in mind the discussed uncertainties, it was nonetheless found that most deviations of mean ${}^{\mathrm{cws}}T$ compared to ${}^{\mathrm{ref}}T$ in individual LCZ are small and that, considering spatial variability, both data sets cannot be distinguished from each other for many LCZ. Moreover, $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ showed good agreement between both data sets for monthly and annual means, whereas deviations where larger during selected 'ideal' days. While further investigations are needed to better understand the reasons behind these deviations and to assess the applicability of CWS data during such weather conditions, long-term studies can make use of the large amount of CWS in urban regions. If CWS and standard meteorological data are then to be used in combination with each other, it is crucial to document systematic differences between and the ranges of uncertainty of both data sets. By doing so, it is possible to benefit from the merits of both data sets, i.e., the high-quality data of reference measurements and their location in environments not covered by CWS, and the large number of CWS in areas where reference measurements are sparse.

4.3 Intra-LCZ variability

The analyses of monthly T per LCZ reveal intra-LCZ T variability of several K for some LCZ, especially for those with a large number of stations. Moreover, intra-LCZ T variability is more pronounced during night-time than during the day or in daily mean values, and in some cases as high as or even higher than differences between LCZ. This corresponds well to previous findings by Houet and Pigeon (2011), who also reported higher intra-class T variability in comparison to inter-class differences for some classes in Toulouse, France, applying the concept of Urban Climate Zones (UCZ; Oke, 2006), a predecessor of LCZ. They (Houet and Pigeon, 2011) also showed higher night-time (T_{min}) than daytime intra-class variability in winter and summer during dry days with low wind speeds and clear sky conditions. Analyses by Skarbit et al. (2017) for Szeged, Hungary, revealed intra-LCZ variability of up to 1.5 K for LCZ 6 at night-time during days with low wind speeds and low cloud cover, being larger than for LCZ 5 or LCZ 9, and larger than during the day. This corresponds well to our findings with higher nocturnal variability than during daytime. They (Skarbit et al., 2017) argue that this variability is likely due to micro-scale differences in exposure, surface cover, and anthropogenic heat sources at the measurement sites. Intra-LCZ T variability was also observed by Leconte et al. (2015) for different LCZ using mobile measurements during summer in Nancy, France. Intra-LCZ variability was especially pronounced in LCZ with heterogeneous urban fabric and was sometimes in the same range as ΔT between two LCZ (Leconte et al., 2015). Mobile measurements by Heusinkveld et al. (2014) also revealed considerable T variability within small distances, though no analyses concerning LCZ were carried out. In addition to that, Ellis et al. (2015) reported intra-neighbourhood (no LCZ or UCZ classification) T differences between measurement stations located either in open or shaded (by vegetation) surroundings of more than 1 K for daytime (T_{max}), which was higher than for night-time (T_{min}). Open stations measured significantly higher T_{max} than shaded ones (Ellis et al., 2015). Presumably, such processes contribute to the observed intra-LCZ T variability in CWS data, since Berlin's streets and courtyards are characterised by a high percentage of tree cover.

To further investigate whether the observed intra-LCZ variability is too large to differentiate between LCZ, i.e., groups of CWS, and whether these groups of CWS correspond to specific LCZ, a k-means cluster analysis (Steinhaus, 1957; Forgy, 1965; MacQueen, 1967; Lloyd, 1982) was carried out. Details of the analyses are given in Appendix C. It can be shown that the cluster algorithm can differentiate between groups of CWS, and that these groups of CWS correspond to different LCZ, despite the intra-LCZ variability. The LCZ concept is therefore applicable to characterise crowdsourced measurement stations according to their T characteristics. Also, since the results of the cluster analysis improve when applying the kernel filter for homogeneous LCZ coverage in the local surroundings (cf. Section 2.4, Appendix C), it becomes obvious that this filter is crucial and justified in order to obtain LCZ specific results, both concerning intra-LCZ T variability as well as inter-LCZ differences. However, the cluster analysis also shows that there is overlap between LCZ, which indicates that the LCZ concept does not fully explain the T variability observed with CWS. It must be kept in mind, though, that LCZ are also a discretisation of a continuum of urban structures. Each LCZ allows for a certain range of parameters such as building height, building surface fraction, and sky view factor, and therefore, intra-LCZ T variability is to be expected. A differentiation of LCZ into subclasses (e.g., LCZ 2_B as compact midrise buildings with an abundance of street trees) might lead to more distinguished results concerning T characteristics. In this respect, the WUDAPT approach to derive a LCZ classification has a limitation since it cannot differentiate between subclasses. Further research is thus needed to derive more detailed classifications, also concerning the challenges in applying the LCZ scheme to European cities, as pointed out by Wicki and Parlow (2017).

While some of the remaining observed intra-LCZ T variability is, on the one hand, probably linked to microscale features of the direct surroundings of a station, parts of it are, on the other hand, likely linked to the size of the urban area of Berlin and the therefore larger scale influence of the city onto the urban boundary layer itself. This 'meso-scale' effect of Berlin was shown to be at least 0.3 K when comparing T at stations located in the same LCZ, one inside and one outside of the city (Fenner et al., 2014). Since the applied methodology groups all stations into one LCZ, irrespective of their location within the study region, meso-scale effects might contribute to the intra-LCZ T variability observed with CWS data, but also present in the reference network. Skarbit et al. (2017) argue in a similar way to explain observed intra-LCZ T variability in Szeged, Hungary, also bringing forward that peripheral regions of the city might be subject to a country breeze of cool air and thus lower T measured at stations located at the city borders. Due to the large number of CWS in the study region, this effect could be studied in more detail, also similar to analyses by Bassett et al. (2016), but this goes beyond the scope of this study.

4.4 Inter-LCZ differences

The analyses with CWS data show statistically significant ΔT for a broad variety of LCZ pairs, both between intra-urban LCZ, i.e., pairs of 'urban' LCZ, as well as 'urban' and 'natural' LCZ. Further, different environments in Berlin differ in T especially during night-time, which is in line with existing literature on UHI and LCZ differences for mid-latitude cities (e.g. Yagüe et al., 1991; Erell and Williamson, 2007; Fortuniak et al., 2006; Houet and Pigeon, 2011; Lehnert et al., 2014, Leconte et al., 2015). Also, daytime $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ are small both in CWS and reference data, and few significant differences can be found in CWS data. This shows that the urban atmosphere is well mixed across the whole study region, and different local environments do not lead to distinct T differences.

Night-time ${}^{\mathrm{cws}}T$ differences are more pronounced during the warmer months of the year than during colder months, which is also typical for mid-latitude cities (e.g. Kłysik and Fortuniak, 1999; Morris et al., 2001, Skarbit et al., 2017), and which was previously reported for Berlin (Fenner et al., 2014). High night-time and daily mean inter-LCZ differences are especially present during May, August, and September 2015, which can be linked to the weather conditions during these months. The weather conditions were favourable for the development of local-scale T differences, i.e., low amounts of precipitation (May, August, and September) and high T (August) (Fig. 3). During August 2015, the region of Berlin was subject to heat wave conditions with twelve hot days at station TEGL, i.e., days with $T_{\mathrm{max}} \geq 30\,^{\circ}\mathrm{C}$. Such conditions intensify T differences between urban and rural environments compared to long-term mean values, as shown by Cheval et al. (2009) for Bucharest, Romania, and by Fenner et al. (2014) for Berlin. Results obtained for 'ideal' days further confirm the notion that dry, calm, and cloud free weather conditions promote local-scale T differences (Stewart and Oke, 2012) since $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ during these days are more pronounced compared to monthly and annual mean values. Overall, results obtained for the study year 2015 are in line with previous long-term investigations for Berlin (Fenner et al., 2014). Thus, it can be assumed that the results are not only due to special conditions during one single year but also representative for longer time frames.

Values of $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ were presented on a mean basis, e.g. monthly or annual mean values, and CWS data generally conformed to reference data. It is worth investigating, however, if crowdsourced CWS data can also capture the temporal variability along these time frames. For this, scatterplots showing absolute $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ and the corresponding temporal SD allow an easy-to-interpret possibility to (1) investigate if reference and CWS data show similar patterns, and to (2) detect if mean $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ values are reliable, i.e., if $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ is larger than SD. Fig. 10 displays such scatterplots for all available annual mean daily, daytime,

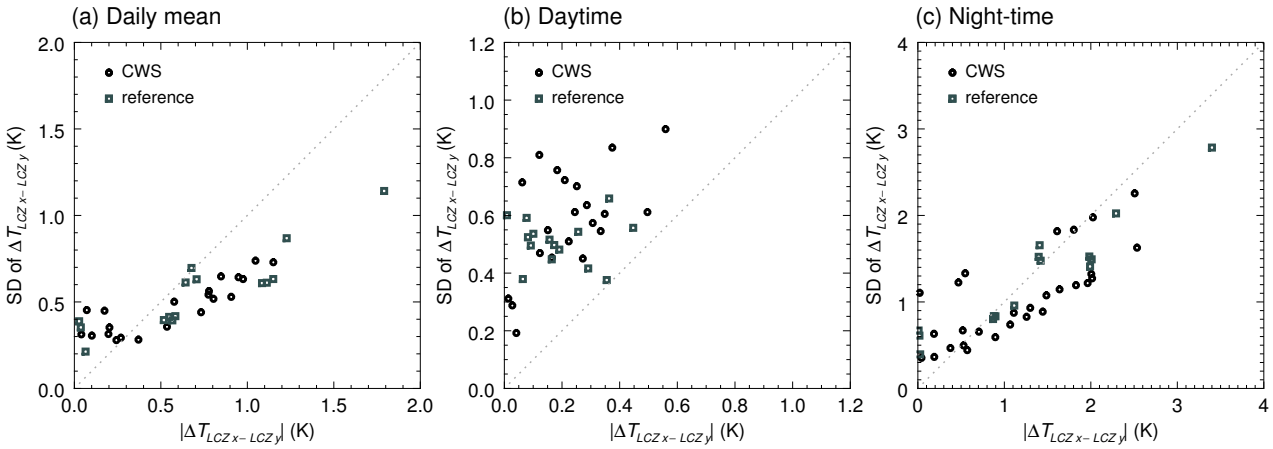

Figure 10: Absolute mean annual air temperature difference $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ between each pair of local climate zones (LCZ) and corresponding temporal standard deviation SD for (a) daily mean values, (b) daytime (1300–1600 UTC+1), and (c) night-time (4–7 hours after sunset UTC+1) as observed by Netatmo citizen weather stations (CWS, circles) and reference stations (squares) in and around Berlin in 2015. SD is calculated along the temporal dimension on the basis of daily values. Please note the different scaling of the axes in the three panels.

and night-time $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$. Similar figures for February and August 2015 can be found in Appendix D as examples for monthly analyses. Calculations of monthly mean $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ were carried out for each of the three time periods on a daily basis to also obtain a temporal SD. This approach is different to the calculations explained in Section 2.5, where $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ was calculated based on monthly mean T across all sites in LCZ x and y. The two methods result in slightly different values of $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$. However, the differences between the two methods are not larger than 0.3 K for the majority of months and LCZ pairs, being in the range of the measurement error of CWS and reference data. Some differences are larger and up to 1.3 K at night-time in April ($\Delta T_{\mathrm{LCZ}\,8\text{–}\mathrm{LCZ}\,D}$). These larger differences occur for LCZ with very few CWS and when data are not continuous over time, such as LCZ 8, with only one or two sites. It demonstrates that values of $\Delta^{\mathrm{cws}} T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ have to be interpreted carefully if only few CWS are available per LCZ, while for LCZ with multiple CWS the results do not depend on the applied method.

Fig. 10 shows that for daily mean and night-time values most LCZ pairs show reliable results for annual $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ since SD is smaller than $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ (Fig. 10a and Fig. 10c, respectively). Besides, Fig. 10 displays that CWS and reference data show a similar pattern, and that even for mean annual $\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ close to 0 K, a SD between 0.2 and 0.5 K is present. Hence, only $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ higher than that are reliable and interpretable. These statements concerning daily mean and night-time periods hold also true for individual months (Appendix D). For daytime values (Fig. 10b), though, results are different. For all LCZ pairs in CWS and reference data, annual mean $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ is lower than SD, and thus values are not interpretable since temporal variation is larger than the mean value. It shows that during daytime urban environments do not differ to one another systematically on an annual basis. On a monthly basis, however, some LCZ pairs in CWS data show higher $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ than SD (Appendix D), being different to reference data. These pairs with higher $|\Delta T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}|$ correspond to LCZ with few CWS and further demonstrate that a low number of sites can result in deviations to reference data that cannot be easily explained. We argue that if only few CWS are available per LCZ and for daytime periods, $\Delta^{\mathrm{cws}} T_{\mathrm{LCZ}\,x\text{–}\mathrm{LCZ}\,y}$ values have to be looked at carefully and interpreted with caution. While further analyses are required to understand these deviations for daytime periods, it can also be inferred from the analyses that CWS data are suitable for LCZ analyses for daily mean and night-time values.

5 Conclusions and outlook

This study of crowdsourced T data from CWS in Berlin, in combination with the LCZ concept, shows the applicability of this novel data set in urban climate research. The different (urban) environments are well represented by CWS with an unprecedented number of measurement stations for several LCZ. However, some classes are under-represented in CWS data, especially 'natural' LCZ. The annual cycle of T shows good agreement between CWS and reference data, though strong deviations between the two data sets occur for some LCZ and months. Such deviations are mainly due to the non-standard set-up of CWS, highlighting the need for further research to understand how measurements obtained from CWS can be combined with existing meteorological networks. In this respect, detailed meta data are crucial, which is a key issue of crowdsourced CWS data. This holds especially true for results obtained for daytime and with small numbers of CWS, when data analyses are more dependent on the individual qualities of a CWS in terms of measurement location and continuous data availability. For LCZ with a high number of CWS and if longer time frames are under consideration, extreme values are averaged out and mean values show good agreement with reference data.

The multiplicity of CWS in LCZ allow the investigation of intra-LCZ T variability, which is especially large during night-time. This intra-LCZ T variability is linked to intra-LCZ variation of urban structures, microscale heterogeneity of the surroundings of a measurement site, and likely a meso-scale influence across the entire urban region. Furthermore, the large number of CWS allows the application of statistical tests to detect significant ΔT between LCZ pairs, taking into account the spatial variability within individual LCZ. These inter-LCZ differences are especially pronounced at night-time and present between inner-city LCZ as well as 'urban' and 'natural' LCZ, showing that thermal climates of an urban region can be differentiated with CWS data. Further, annual mean ΔT at CWS are similar to values measured with reference stations, allowing us to conclude that crowdsourced data yield reliable information on differences of thermal characteristics between urban environments. In this respect, the LCZ concept proved to be applicable. However, T variation within individual LCZ is considerable and thus should not be neglected. Further research is needed to assess the impact of microscale heterogeneity and meso-scale influences, how these could be further considered when applying the LCZ scheme, but also how automated LCZ classification methods could be improved to derive LCZ subclasses. This could help to overcome some of the challenges when applying the LCZ concept on a large data set of measurement stations, and to exploit the full potential of the concept with respect to intra-urban distinction of local-scale environments.

With this study it was demonstrated that crowdsourced data from CWS are suitable as complementary data to existing measurement networks to study urban climate phenomena and to investigate features that are difficult to detect with a low number of measurement sites. Thus, crowdsourcing of atmospheric data is currently the only way to detect the large spatial heterogeneity of urban thermal climate with observations. Future research utilising crowdsourced atmospheric data could, e.g., focus on the question of how issues concerning missing meta data of CWS can be overcome. Further studies could make use of the dense distribution of CWS to analyse larger-scale advective UHI processes, investigate spatio-temporal characteristics of atmospheric humidity, which is also measured by Netatmo CWS, or target the question if data from CWS could provide dense information for studies focusing on spatially varying heat-related risks.

Acknowledgements

The authors would like to thank all Netatmo owners in Berlin and surroundings who shared their data. We are grateful to Tom Grassmann for helping with all issues concerning the collection of CWS data, as well as to Hartmut Küster and Ingo Suchland for maintaining the UCON weather stations, installed and operated from funds provided by the Technische Universität Berlin. Furthermore, we thank the German Weather Service for provision of atmospheric data. Lastly, the authors thank all colleagues for the fruitful discussions, as well as two anonymous reviewers for their valuable comments that helped to improve the manuscript. This research was funded by the Deutsche Forschungsgemeinschaft (DFG) as part of the research project 'Heat waves in Berlin, Germany – urban climate modifications' (Grant No. SCHE 750/15-1) and supported by the Cluster of Excellence 'CliSAP' (EXC177), University of Hamburg, funded by the DFG. The study was also part of the DFG Research Unit 1736 'Urban Climate and Heat Stress in Mid-Latitude Cities in View of Climate Change (UCaHS)' (Grant Nos. SCHE 750/8-1 and SCHE 750/9-1), and contributes to the research programme 'Urban Climate Under Change ($[UC]^2$)', funded by the German Ministry of Research and Education (FKZ 01LP1602A).

Appendix A

To retrieve data retrospectively from the Netatmo server for a specific time period a different method of the API has to be used compared to the method to obtain near real-time data. The 'getmeasure' method was used for retrospective acquisition of data in May and August 2015. Data obtained from this method are hourly mean values compared to instantaneous values using the 'getpublicdata' method (as described in Meier et al., 2017). Comparison of data from the two methods for all days in August 2015 when both methods provided data reveals a hysteresis for the mean hourly values along the diurnal cycle (not shown). Mean hourly T from CWS are cooler (warmer) than instantaneous values of up to −0.3 K (+0.3 K) in the mean across all CWS during the morning hours after sunrise (during the evening hours around sunset). During night-time hours differences between CWS data from the two methods are below 0.1 K.

Appendix B

The evaluation of the final LCZ classification was carried out using two approaches: (1) a bootstrapping approach (Kaloustian and Bechtel, 2016) to test the consistency and robustness of the training areas (TA) and the resulting classification, and (2) with reference data from a second experienced expert to test the correctness of the TA.

The bootstrapping approach (1) conducts the LCZ classification based on the TA 25 times, each time using only 50 % of the TA (polygons) for training and using the other 50 % as evaluation areas. For each run, a number of accuracy measures are calculated based on a confusion matrix. The confusion matrix compares the LCZ assigned by the automated classifier in SAGA to the LCZ of the evaluation areas. The following accuracy measures were calculated: overall accuracy (OA^{all}), which is the ratio of correctly classified pixels to all pixels (within the evaluation set), OA^{urb} is the OA for the

'urban' LCZ only, and κ is a standard accuracy measure accounting for the different percentages of classes. Good results in the bootstrapping are a strong indication for consistent training data, but it is insensitive to consistently false labeling (e.g. if all dense tree areas were labeled as water and vice versa). Therefore, (2) the same accuracy measures were calculated based on the independent reference data and the final LCZ map.

Mean OA_{all} for the bootstrapping (case 1) is 0.68 and κ = 0.63, meaning that on average of the 25 runs 68 % of the classified pixels agree with the evaluation LCZ label. This is considered to be a good agreement since some variation can be expected, when using only half of the training data. For the 'urban' LCZ the accuracy is lower with $OA^{urb} = 0.57$, showing that for urban areas it is more difficult to obtain robust results. Both OA values are similar to evaluation results reported by Brousse et al. (2016) for their WUDAPT-LCZ classification for Madrid, Spain ($OA^{all} = 0.67$, $OA^{urb} = 0.59$).

Comparing the classification with reference LCZ data (case 2), OA^{all} is 0.95, OA^{urb} is 0.92, and κ is 0.94. This shows that the TA used in the final LCZ classification are of high quality and therefore are considered to be a good representation of the actual LCZ in Berlin.

Appendix C

A k-means cluster analysis (Steinhaus, 1957; Forgy, 1965; MacQueen, 1967; Lloyd, 1982) was carried out to (1) test whether groups of CWS can be distinguished by automated cluster algorithms, corresponding to LCZ, and (2) whether the approach of the applied kernel filter to filter out stations in regions of heterogeneous LCZ coverage is merited. The cluster analysis was applied on a monthly basis and carried out as follows. First, hourly T differences at each CWS were calculated relative to reference site ALEX (Table 1, located in LCZ 2). Second, mean monthly diurnal cycles of these differences were calculated per CWS, and all CWS that provided at least 80 % of values per hour and month were further considered. These mean values were then used as the dimensions in the k-means algorithm to cluster the stations. If all stations had no values during a specific hour (due to filtering or server outages) the number of dimensions was reduced by this hour. This approach using T differences was chosen to reduce the effect of weather on T at each CWS and to derive the characteristic thermal features along the diurnal cycle. The cluster method was firstly applied on all CWS (case a), and secondly only on those stations that passed the kernel filter for LCZ assignment (case b, cf. Section 2.4). In both cases, the number of unique LCZ among these CWS was chosen as the number of clusters in the cluster analysis. The cluster algorithm was run 100 times to check if the starting value of each cluster had an effect. However, in all 100 runs the same number of clusters was detected and the same clusters were assigned to the CWS.

Table A.1 and Table A.2 summarise the results for the month September 2015 as an example, showing the number of CWS per cluster and LCZ. For case a, it can be seen that some clusters correspond well with certain LCZ, but that there is also large overlap between clusters and different LCZ (Table A.1). LCZ 6, e.g., has one main cluster (9, >43 % of CWS in LCZ 6), but also all other clusters contain CWS in LCZ 6. After application of the kernel filter (case b) the clusters are much more clearly defined, resulting in a better automated differentiation of the CWS according to their T characteristics (Table A.2). Now, for LCZ 6, one cluster (4) contains the majority (>61 %) of CWS in this LCZ. Some confusion remains, however, and the LCZ do not correspond exactly to one cluster. Especially for LCZ 2 there is overlap with LCZ 5 and LCZ 6 for the two main clusters (1 and 8; Table A.2). Nonetheless, it can be derived from the results that, firstly, CWS can be grouped to clusters according to their thermal characteristics on the diurnal cycle and that these clusters correspond to certain LCZ. Secondly, the kernel filter to filter out stations in heterogeneous LCZ environments is justified, which highlights the importance that the thermal source area of the measurement station must be characteristic of the respective LCZ to obtain clear LCZ specific results.

Table A.1: Number of citizen weather stations (CWS) per local climate zone (LCZ) and cluster in September 2015 based on a k-means cluster analysis with eleven clusters. Cluster dimensions: mean monthly air temperature difference to reference site ALEX (Table 1) for each hour of the day with valid data for all sites (23 hours). All CWS were included in the analysis that provided ≥ 80 % valid data for each hour in the month.

LCZ/cluster	1	2	3	4	5	6	7	8	9	10	11	$\sum_{row}$
LCZ 2	17	18	30	3	4	45	5	5	4	3	11	145
LCZ 4	3	4	11	1	0	4	1	1	4	0	2	31
LCZ 5	11	12	38	9	6	26	3	16	12	4	1	138
LCZ 6	5	19	68	5	4	11	4	18	133	38	2	307
LCZ 8	2	0	0	1	0	4	1	0	5	1	1	15
LCZ 9	0	0	0	0	0	0	0	0	9	0	0	9
LCZ A	0	1	1	1	0	0	0	1	11	0	0	15
LCZ B	0	0	1	0	0	1	0	0	2	0	0	4
LCZ C	1	0	2	0	0	0	0	0	2	0	0	5
LCZ D	0	0	0	1	1	0	0	0	7	0	0	9
LCZ G	0	0	0	0	0	0	0	0	1	0	1	2
$\sum_{column}$	39	54	151	21	15	91	14	41	190	46	18	680

Table A.2: Same as Table A.1 but including only citizen weather stations that passed the kernel filter for homogeneous LCZ coverage (cf. Section 2.4).

LCZ/cluster	1	2	3	4	5	6	7	8	$\sum_{row}$
LCZ 2	25	2	9	13	1	10	4	48	112
LCZ 4	2	0	0	2	0	2	0	4	10
LCZ 5	9	5	7	12	2	0	2	20	57
LCZ 6	2	5	15	128	21	1	4	33	209
LCZ 8	1	0	0	0	0	0	0	0	1
LCZ 9	0	0	0	3	0	0	0	0	3
LCZ A	0	0	0	4	0	0	0	0	4
LCZ D	0	1	1	4	0	0	0	0	6
$\sum_{column}$	39	13	32	166	24	13	10	105	402

Appendix D

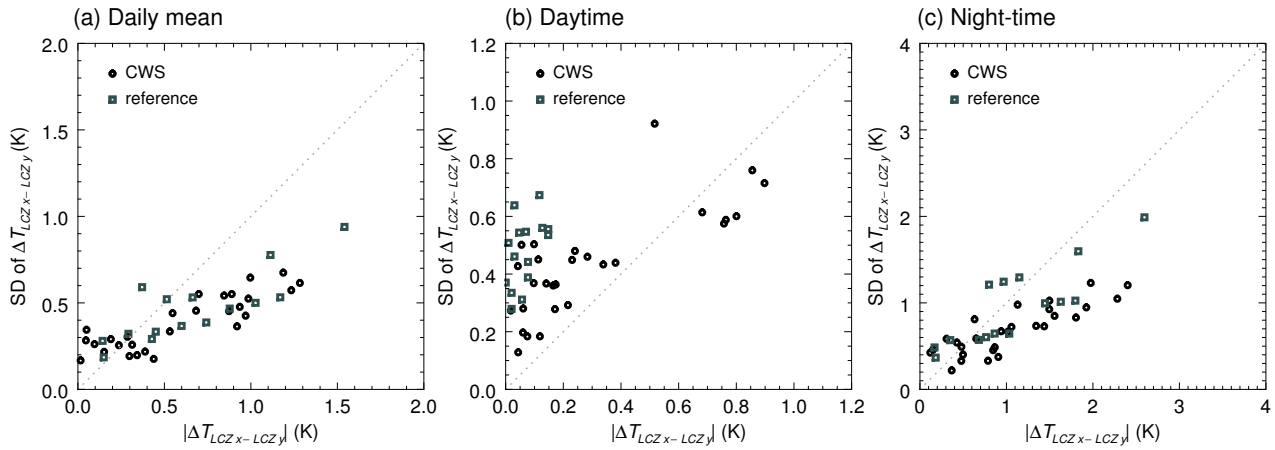

Figure A.1: Absolute mean air temperature difference $|\Delta T_{\mathrm{LCZ}\,x-\mathrm{LCZ}\,y}|$ between each pair of local climate zones (LCZ) and corresponding temporal standard deviation SD for (a) daily mean values, (b) daytime (1300–1600 UTC+1), and (c) night-time (4–7 hours after sunset UTC+1) as observed by Netatmo citizen weather stations (CWS, circles) and reference stations (squares) in and around Berlin in February 2015. SD is calculated along the temporal dimension on the basis of daily values. Please note the different scaling of the axes in the three panels.

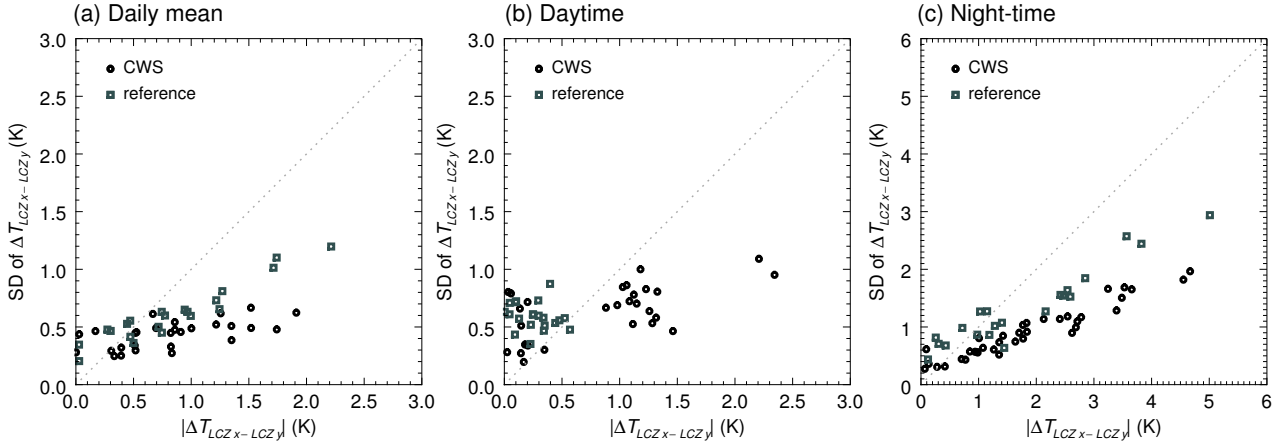

Figure A.2: Same as Fig. A.1 but for August 2015.

References

Alexander, P.J., G. Mills, 2014: Local Climate Classification and Dublin's Urban Heat Island. – Atmos. **5**, 755–774. DOI: 10.3390/atmos5040755.

Arnds, D., J. Böhner, B. Bechtel, 2017: Spatio-temporal variance and meteorological drivers of the urban heat island in a European city. – Theor. Appl. Climatol. **128**, 43–61. DOI: 10.1007/s00704-015-1687-4.

Arnfield, A.J., 2003: Two decades of urban climate research: a review of turbulence, exchange of energy and water, and the urban heat island. – Int. J. Climatol. **23**, 1–26. DOI: 10.1002/joc.859.

Basara, J.B., H.G. Basara, B.G. Illston, K.C. Crawford, 2010: The Impact of the Urban Heat Island during an Intense Heat Wave in Oklahoma City. – Adv. Meteor. **230365**. DOI: 10.1155/2010/230365.

Bassett, R., X. Cai, L. Chapman, C. Heaviside, J.E. Thornes, C.L. Muller, D.T. Young, E.L. Warren, 2016: Observations of urban heat island advection from a high-density monitoring network. – Quart. J. Roy. Meteor. Soc. **143**, 2434–2441. DOI:10.1002/qj.2836.

Bechtel, B., C. Daneke, 2012: Classification of Local Climate Zones Based on Multiple Earth Observation Data. – IEEE J. Sel. Top. Appl. **5**, 1191–1202. DOI:10.1109/JSTARS.2012.2189873.

Bechtel, B., P.J. Alexander, J. Böhner, J. Ching, O. Conrad, J. Feddema, G. Mills, L. See, I.D. Stewart, 2015: Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. – ISPRS Int. J. Geo-Inf. **4**, 199–219. DOI:10.3390/ijgi4010199.

Bechtel, B., M. Demuzere, P. Sismanidis, D. Fenner, O. Brousse, C. Beck, F. Van Coillie, O. Conrad, I. Keramitsoglou, A. Middel, G. Mills, D. Niyogi, M. Otto, L. See, M.-L. Verdonck, 2017: Quality of Crowdsourced Data on Urban Morphology – The Human Influence Experiment (HUMINEX) – Urban Sci. **1**, 15. DOI:10.3390/urbansci1020015.

Bell, S., D. Cornford, L. Bastin, 2013: The state of automated amateur weather observations. – Weather **68**, 36–41. DOI: 10.1002/wea.1980.

Bell, S., D. Cornford, L. Bastian, 2015: How good are citizen weather stations? Addressing a biased opinion. – Weather **70**, 75–84. DOI:10.1002/wea.2316.

Breiman, L., 2001: Random Forests. – Mach. Learn. **45**, 5–32. DOI:10.1023/A:1010933404324.

Brousse, O., A. Martilli, M. Foley, G. Mills, B. Bechtel, 2016: WUDAPT, an efficient land use producing data tool for mesoscale models? Integration of urban LCZ in WRF over Madrid. – Urban Climate **17**, 116–134. DOI:10.1016/j.uclim.2016.04.001.

Castell, N., M. Kobernus, H.-Y. Liu, P. Schneider, W. Lahoz, A.J. Berre, J. Noll, 2015: Mobile technologies and services for environmental monitoring: The Citi-Sense-MOB approach. – Urban Climate **14**, 370–382. DOI: 10.1016/j.uclim.2014.08.002.

Chandler, T.J., 1965: The Climate of London. – Hutchinson & Co, London, 292 pp.

Chapman, L., C.L. Muller, D.T. Young, E.L. Warren, C.S. B. Grimmond, X.-M. Cai, E.J. S. Ferranti, 2015: The Birmingham Urban Climate Laboratory: An Open Meteorological Test Bed and Challenges of the Smart City. – Bull. Amer. Meteor. Soc. **96**, 1545–1560. DOI:10.1175/BAMS-D-13-00193.1.

Chapman, L., C. Bell, S. Bell, 2016: Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. – Int. J. Climatol., published online December 2nd, 2016; DOI:10.1002/joc.4940.

Cheval, S., A. Dumitrescu, A. Bell, 2009: The urban heat island of Bucharest during the extreme high temperatures of July 2007. – Theor. Appl. Climatol. **97**, 391–401. DOI: 10.1007/s00704-008-0088-3.

Chow, W.T.L., M. Roth, 2006: Temporal dynamics of the urban heat island of Singapore. – Int. J. Climatol. **26**, 2243–2260. – DOI:10.1002/joc.1364.

Christen, A., R. Vogt, 2004: Energy and radiation balance of a central European city. – Int. J. Climatol. **24**, 1395–1421. DOI: 10.1002/joc.1074.

Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, J. Böhner, 2015: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. – Geosci. Model Dev. **8**, 1991–2007. DOI: 10.5194/gmd-8-1991-2015.

DWD Climate Data Center (CDC), 2016: Historical hourly station observations of 2 m air temperature and humidity. – Version v004.

Eliasson, I., M.K. Svensson, 2003: Spatial air temperature variations and urban land use – a statistical approach. – Meteorol. Appl. **10**, 135–149. DOI:10.1017/S1350482703002056.

Ellefsen, R., 1991: Mapping and measuring buildings in the canopy boundary layer in ten U.S. cities. – Energy Build. **16**, 1025–1049. DOI:10.1016/0378-7788(91)90097-m.

Ellis, K.N., J.M. Hathaway, L.R. Mason, D.A. Howe, T.H. Epps, V.M. Brown, 2015: Summer temperature variability across four urban neighborhoods in Knoxville, Tennessee, USA. – Theor. Appl. Climatol. **127**, 701–710. DOI: 10.1007/s00704-015-1659-8.

Emmanuel, R., E. Krüger, 2012: Urban heat island and its impact on climate change resilience in a shrinking city: The case of Glasgow, UK. – Build. Env. **53**, 137–149. DOI: 10.1016/j.buildenv.2012.01.020.

Endlicher, W., N. Lanfer, 2003: Meso- and Micro-Climatic Aspects of Berlin's Urban Climate. – Die Erde **134**, 277–293.

EnvAtlBer – Environmental Atlas Berlin, 2014: Building and Vegetation heights 2009/2010. – Provided by Berlin Senate Department for Urban Development and the Environment.

Erell, E., T. Williamson, 2007: Intra-urban differences in canopy layer air temperature at a mid-latitude city. – Int. J. Climatol. **27**, 1243–1255. DOI:10.1002/joc.1469.

Farr, T.G., P.A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, D. Alsdorf, 2007: The Shuttle Radar Topography Mission. – Rev. Geophys. **45**, DOI: 10.1029/2005RG000183.

Fenner, D., D. Scherer, F. Meier, A. Polze, 2014: Spatial and temporal air temperature variability in Berlin, Germany, during the years 2001–2010. – Urban Climate **10**, 308–331. DOI:10.1016/j.uclim.2014.02.004.

Forgy, E.W., 1965: Cluster analysis of multivariate data: efficiency versus interpretability of classifications. – Biometrics **21**, 768–769.

Fortuniak, K., K. Kłysik, J. Wibig, 2006: Urban-rural contrasts of meteorological parameters in Łódź. – Theor. Appl. Climatol. **84**, 91–101. DOI:10.1007/s00704-005-0147-y.

Gál, T., N. Skarbit, J. Unger, 2016: Urban heat island patterns and their dynamics based on an urban climate measurement network. – Hung. Geogr. Bull. **65**, 105–116. DOI: 10.15201/hungeobull.65.2.2.

Geletič, J., M. Lehnert, 2016: GIS-based delineation of local climate zones: The case of medium-sized Central European cities. – Moravian Geograph. Reports **24**, 2–12. DOI:10.1515/mgr-2016-0012.

Gosset, W.S., 1908: The Probable Error of a Mean. – Biometrika **6**, 1–25. DOI:10.2307/2331554.

Grimmond, C.S.B., 2006: Progress in measuring and observing the urban atmosphere. – Theor. Appl. Climatol. **84**, 3–22. DOI:10.1007/s00704-005-0140-5.

Heusinkveld, B.G., G.J. Steeneveld, L.W.A. van Hove, C.M.J. Jacobs, A.A.M. Holtslag, 2014: Spatial variability of the Rotterdam urban heat island as influenced by urban land use. – J. Geophys. Res. Atmos. **119**, 1–14. DOI: 10.1002/2012JD019399.

Holm, S., 1979: A Simple Sequentially Rejective Multiple Test Procedure. – Scand. J. Stat. **6**, 65–70.

Holmer, B., S. Thorsson, I. Eliasson, 2007: Cooling rates, sky view factors and the development of intra-urban air temperature differences. – Geografiska Annaler: Ser A, Phys. Geogr. **89**, 237–248. DOI:10.1111/j.1468-0459.2007.00323.x.

Holmer, B., S. Thorsson, J. Lindén, 2013: Evening evapotranspirative cooling in relation to vegetation and urban geometry in the city of Ouagadougou, Burkina Faso. – Int. J. Climatol. **33**, 3089–3105. DOI:10.1002/joc.3561.

Houet, T., G. Pigeon, 2011: Mapping urban climate zones and quantifying climate behaviors – An application on Toulouse urban area (France). – Env. Pollut. **159**, 2180–2192. DOI: 10.1016/j.envpol.2010.12.027.

Hu, X.-M., M. Xue, P.M. Klein, B.G. Illston, S. Chen, 2016: Analysis of Urban Effects in Oklahoma City using a Dense Surface Observation Network. – J. Appl. Meteorol. Clim. **55**, 723–741. DOI:10.1175/JAMC-D-15-0206.1.

Hupfer, P., F.-M. Chmielewski, 1990: Das Klima von Berlin. – Akademie Verlag, Berlin, 288 pp.

Jänicke, B., F. Meier, D. Fenner, U. Fehrenbach, A. Holtmann, D. Scherer, 2017: Urban-rural differences in near-surface air temperature as resolved by the Central European Refined analysis (CER): sensitivity to planetary boundary layer schemes and urban canopy models. – Int. J. Climatol. **37**, 2063–2079. DOI:10.1002/joc.4835.

Kaloustian, N., B. Bechtel, 2016: Local Climatic Zoning and Urban Heat Island in Beirut. – Procedia Eng. **169**, 216–223. DOI:10.1016/j.proeng.2016.10.026.

Kaspar, F., G. Müller-Westermeier, E. Penda, H. Mächel, K. Zimmermann, A. Kaiser-Weiss, T. Deutschländer, 2013: Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre. – Adv. Sci. Res. **10**, 99–106. DOI:10.5194/asr-10-99-2013.

Kim, Y.H., J.J. Baik, 2005: Spatial and temporal structure of the urban heat island in Seoul. – J. Appl. Meteorol. **44**, 591–605. DOI:10.1175/JAM2226.1.

Kłysik, K., K. Fortuniak, 1999: Temporal and spatial characteristics of the urban heat island of Łódź, Poland. – Atmos. Env. **33**, 3885–3895. DOI:10.1016/S1352-2310(99)00131-4.

Kolokotroni, M., R. Giridharan, 2008: Urban heat island intensity in London: An investigation of the impact of physical characteristics on changes in outdoor air temperature during summer. – Sol. Energy **82**, 986–998. DOI:10.1016/j.solener.2008.05.004.

Koskinen, J.T., J. Poutiainen, D.M. Schultz, S. Joffre, J. Koistinen, E. Saltikoff, E. Gregow, H. Turtiainen, W.F. Dabberdt, J. Damski, N. Eresmaa, S. Göke, O. Hyvärinen, L. Järvi, A. Karppinen, J. Kotro, T. Kuitunen, J. Kukkonen, M. Kulmala, D. Moisseev, P. Nurmi, H. Pohjola, P. Pylkkö, T. Vesala, Y. Viisanen, 2011: The Helsinki Testbed: A Mesoscale Measurement, Research, and Service Platform. – Bull. Amer. Meteor. Soc. **92**, 325–342. DOI:10.1175/2010bams2878.1.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, F. Rubel, 2006: World Map of the Köppen-Geiger climate classification updated. – Meteorol. Z. **15**, 259–263. DOI:10.1127/0941-2948/2006/0130.

Kruskal, W.H., W.A. Wallis, 1952: Use of Ranks in One-Criterion Variance Analysis. – J. Amer. Stat. Assoc. **47**, 583–621. DOI:10.1080/01621459.1952.10483441.

Kuik, F., A. Lauer, G. Churkina, H.A.C. Denier van der Gon, D. Fenner, K.A. Mar, T.M. Butler, 2016: Air quality modelling in the Berlin-Brandenburg region using WRF-Chem v3.7.1: sensitivity to resolution of model grid and input data. – Geosci. Model Dev. **9**, 4339–4363. DOI:10.5194/gmd-9-4339-2016.

Leconte, F., J. Bouyer, R. Claverie, M. Pétrissans, 2015: Using Local Climate Zone scheme for UHI assessment: Evaluation of the method using mobile measurements. – Build. Env. **83**, 39–49. DOI:10.1016/j.buildenv.2014.05.005.

Leconte, F., J. Bouyer, R. Claverie, M. Pétrissans, 2016: Analysis of nocturnal air temperature in districts using mobile measurements and a cooling indicator. – Theor. Appl. Climatol., published online August 10th, 2016; DOI:10.1007/s00704-016-1886-7.

Lehnert, M., J. Geletič, J. Husák, M. Vysoudil, 2014: Urban field classification by "local climate zones" in a medium-sized Central European city: the case of Olomouc (Czech Republic). – Theor. Appl. Climatol. **122**, 531–541. DOI:10.1007/s00704-014-1309-6.

Lelovics, E., J. Unger, T. Gál, C.V. Gál, 2014: Design of an urban monitoring network based on Local Climate Zone mapping and temperature pattern modelling. – Climate Res. **60**, 51–62. DOI:10.3354/cr01220.

Lindberg, F., B. Holmer, S. Thorsson, 2008: SOLWEIG 1.0 – Modelling spatial variations of 3D radiant fluxes and mean radiant temperature in complex urban settings. Int. J. Biometeorol. **52**, 697–713. DOI:10.1007/s00484-008-0162-7.

Lloyd, S., 1982: Least squares quantization in PCM. – IEEE Transactions on Information Theory **28**, 129–137. DOI:10.1109/tit.1982.1056489.

Loridan, T., C.S.B. Grimmond, 2012: Characterization of Energy Flux Partitioning in Urban Environments: Links with Surface Seasonal Properties. – J. Appl. Meteor. Climate **51**, 219–241. DOI:10.1175/jamc-d-11-038.1.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. – Proceedings of the fifth Berkeley symposium on mathematical statistics and probability **1**, 281–297.

Magee, N., J. Curtis, G. Wendler, 1999: The Urban Heat Island Effect at Fairbanks, Alaska. – Theor. Appl. Climatol. **64**, 39–47. DOI:10.1007/s007040050109.

Mann, H.B., D.R. Whitney, 1947: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. – Ann. Math. Stat. **18**, 50–60. DOI:10.1214/aoms/1177730491.

Meier, F., D. Fenner, T. Grassmann, M. Otto, 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.

Morris, C.J.G., I. Simmonds, N. Plummer, 2001: Quantification of the Influences of Wind and Cloud on the Nocturnal Urban Heat Island of a Large City. – J. Appl. Meteorol. **40**, 169–182. DOI:10.1175/1520-0450(2001)040<0169:QOTIOW>2.0.CO;2.

Muller, C.L., L. Chapman, C.S.B. Grimmond, D.T. Young, X. Cai, 2013: Sensors and the city: a review of urban meteorological networks. – Int. J. Climatol. **33**, 1585–1600. DOI:10.1002/joc.3678.

Muller, C.L., L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, 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.

Nakamura, R., L. Mahrt, 2005: Air Temperature Measurement Errors in Naturally Ventilated Radiation Shields. – J. Atmos. Oceanic Technol. **22**, 1046–1058. DOI:10.1175/JTECH1762.1.

Nakamura, Y., T.R. Oke, 1988: Wind, temperature and stability conditions in an east-west oriented urban canyon. – Atmos. Env. **22**, 2691–2700. DOI:10.1016/0004-6981(88)90437-4.

Niachou, K., I. Livada, M. Santamouris, 2008: Experimental study of temperature and airflow distribution inside an urban street canyon during hot summer weather conditions – Part I: Air and surface temperatures. – Build. Env. **43**, 1383–1392. DOI:10.1016/j.buildenv.2007.01.039.

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.

Oke, T.R., 2006: Initial guidance to obtain representative meteorological observations at urban sites. Instruments and Observing Methods Rep. **81**, WMO/TD-1250, WMO, Geneva, 47 pp.

Oke, T.R., G. Maxwell, 1975: Urban heat island dynamics in Montreal and Vancouver. – Atmos. Env. **9** (2), 191–200. DOI:10.1016/0004-6981(75)90067-0.

Perera, N.G.R., R. Emmanuel, 2016: A "Local Climate Zone" based approach to urban planning in Colombo, Sri Lanka. – Urban Climate, published online December 2nd, 2016; DOI:10.1016/j.uclim.2016.11.006.

Schatz, J., C.J. Kucharik, 2014: Seasonality of the Urban Heat Island Effect in Madison, Wisconsin. – J. Appl. Meteorol. Climate **53**, 2371–2386. DOI:10.1175/JAMC-D-14-0107.1.

Scherer, D., U. Fehrenbach, H.-D. Beha, E. Parlow, 1999: Improved concepts and methods in analysis and evaluation of the urban climate for optimizing urban planning processes. – Atmos. Env. **33**, 4185–4193. DOI:10.1016/s1352-2310(99)00161-2.

See, L., C. Perger, M. Duerauer, S. Fritz, B. Bechtel, J. Ching, P. Alexander, G. Mills, M. Foley, M. O'Connor, I. Stewart, J. Feddema, V. Masson, 2015: Developing a community-based worldwide urban morphology and materials database (WUDAPT) using remote sensing and crowdsourcing for improved urban climate modelling. – 2015 Joint Urban Remote Sensing Event (JURSE). DOI:10.1109/jurse.2015.7120501.

Siu, L.W., M.A. Hart, 2013: Quantifying urban heat island intensity in Hong Kong SAR, China. – Env. Monit. Assess. **185**, 4383–4398. DOI:10.1007/s10661-012-2876-6.

Skarbit, N., I.D. Stewart, J. Unger, T. Gál, 2017: Employing an urban meteorological network to monitor air temperature conditions in the 'local climate zones' of Szeged, Hungary. – Int. J. Climatol., published online March 9th, 2017; DOI: 10.1002/joc.5023.

Steeneveld, G.J., S. Koopmans, B.G. Heusinkveld, L.W.A. van Hove, 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. Atmos. **116**, published online, DOI:10.1029/2011JD015988.

Steinhaus, H., 1957: Sur la division des corps matériels en parties. – Bull. Acad. Polo. Sci. **12**, 801–804.

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.

Stewart, I.D., 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.

Stewart, I.D., T.R. Oke, E.S. Krayenhoff, 2014: Evaluation of the 'local climate zone' scheme using temperature observations and model simulations. – Int. J. Climatol. **34**, 1062–1080. DOI:10.1002/joc.3746.

Suomi, J., J. Käyhkö, 2012: The impact of environmental factors on urban temperature variability in the coastal city of Turku, SW Finland. – Int. J. Climatol. **32**, 451–463. DOI: 10.1002/joc.2277.

Unger, J., 2004: Intra-urban relationship between surface geometry and urban heat island: review and new approach. – Climate Res. **27**, 253–264. DOI:10.3354/cr027253.

Unger, J., Z. Sümeghy, J. Zoboki, 2001: Temperature cross-section features in an urban area. – Atmos. Res. **58**, 117–127. DOI:10.1016/S0169-8095(01)00087-4.

Unger, J., E. Lelovics, T. Gál, 2014: Local Climate Zone mapping using GIS methods in Szeged. – Hung. Geogr. Bull. **63**, 29–41. DOI:10.15201/hungeobull.63.1.3.

van Hove, L., C. Jacobs, B. Heusinkveld, J. Elbers, B. van Driel, A. Holtslag, 2015: Temporal and spatial variability of urban heat island and thermal comfort within the Rotterdam agglomeration. – Build. Environ. **83**, 91–103. DOI: 10.1016/j.buildenv.2014.08.029.

Warren, E.L., D.T. Young, L. Chapman, C.L. Muller, C.S.B. Grimmond, 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.

Wicki, A., E. Parlow, 2017: Attribution of local climate zones using a multitemporal land use/land cover classification scheme. – J. Appl. Remote Sens. **11**, 026001. DOI:10.1117/1.jrs.11.026001.

Wilcoxon, F., 1945: Individual Comparisons by Ranking Methods. – Biometrics Bull. **1**, 80–83. DOI:10.2307/3001968.

Wolters, D., T. Brandsma, 2012: Estimating the Urban Heat Island in Residential Areas in the Netherlands Using Observations by Weather Amateurs. – J. Appl. Meteor. Climatol. **51**, 711–721. DOI:10.1175/JAMC-D-11-0135.1.

Yagüe, C., E. Zurita, A. Martinez, 1991: Statistical analysis of the Madrid urban heat island. – Atmos. Env. **25**, 327–332. DOI:10.1016/0957-1272(91)90004-X.