Analysis of small-scale urban air temperature patterns using crowd-sourced data

Preprint

M. Feichtinger[a,b], R. de Wit[a,*], G. Goldenits[a,c], T. Kolejka[a], B. Hollósi[a], M. Žuvela-Aloise[a] and J. Feigl[a]

[a]Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Hohe Warte 38, 1190 Vienna, Austria
[b]Karl-Franzens-Universität Graz, Department of Geography and Regional Sciences, Heinrichstraße 36, 8010 Graz, Austria
[c]Universität Wien, Department of Economics, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
[*]Corresponding author: rosmarie.dewit@zamg.ac.at

Preprint of Feichtinger et al. (2020): Case-study of neighborhood-scale summertime urban air temperature for the City of Vienna using crowd-sourced data (https://doi.org/10.1016/j.uclim.2020.100597)

Highlights:

- The urban heat island of Vienna was investigated using citizen weather stations
- A comparison to the operational network shows clear advantages for urban studies
- The implication of different spatial averaging methods is shown and discussed

Abstract

For the investigation of intra-urban air temperature patterns, high spatial resolution observations are needed. To reach such a dense monitoring network in a cost-efficient way, recently there have been efforts to utilize networks of citizen weather stations (CWS). Here, quality-controlled crowd-sourced CWS air temperature observations were used to study the temperature distribution in Vienna, Austrian, for an 11-day high-temperature period in August 2018. The statistical quality control was based on the procedure developed in a previous study, further optimized by incorporating two additional steps addressing the specific issue of radiative errors using solar radiation data from non-CWS reference stations. After application of the quality control, CWS air temperature data provide results comparable to stations of the national meteorological service network, and discrepancies between these data sets could be explained by differences in station location relative to the urban structure. In Vienna, 1357 unique stations were available during the study period, of which 1083 stations passed the full quality control procedure. Compared to the currently available operational network, this represents a major improvement in the spatial resolution of available temperature data. Hence, quality-controlled CWS data for Vienna seem a promising data source for further urban climate studies of the area.

Keywords: crowd-sourcing, urban heat island

1. Introduction

Cities are ever-growing arrangements of various urban structures and materials. These modifications in land use through urbanization and respective changes in the surface energy balance result in temperatures being amplified in cities, an effect referred to as the Urban Heat Island (UHI) effect (e.g. Landsberg, 1981; Oke, 1995). Even within city limits, microclimatic differences can be found, resulting in intra-urban heat islands (Yokobori and Ohta, 2009). To investigate such small-scale heat variations, station networks with high spatial resolution are needed for observational studies as well as the verification of urban modelling results. However, due to the high cost associated with meteorological measuring equipment, it is hardly common to find that many professional stations within a city. Even in Vienna, Austria, and its immediate surroundings, where a relatively dense observational network exists compared to other cities, only 15 stations operated by the Zentralanstalt für Meteorologie und Geodynamik (ZAMG, Austria's national meteorological service) and the City of Vienna Environmental Protection Department are available (GCOS, 2017). This network is depicted in Figure 1, clearly showing many areas not covered by observations. An additional issue related to performing urban studies with these observations is that, to follow World Meteorological Organisation (WMO) guidelines for climate and synoptic stations (WMO, 2008), most stations are located at similar land use classes, and measurements covering the full range of urban structure are not available.

Figure 1: Satellite image of Vienna and the surrounding areas. White markers indicate the location of the stations in the observational networks of the ZAMG and the City of Vienna.

To reach a dense monitoring network in a cost-efficient way, there has been a recent trend to utilize alternative networks based on citizen weather stations (CWS) (e.g. Chapman et al., 2017; Meier et al., 2017, Hammerberg et al, 2018). Many of these studies rely on data from the network of Netatmo stations. Netatmo is a French company specialized in "smart home" devices, and has a weather station module for hobby users among their products. If the user agrees, the collected meteorological data are uploaded to the Netatmo server automatically, and all station data can be downloaded for further study.

The term 'crowdsourcing' has first been coined in an economic context by Howe (2006) as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call". Muller et al. (2015) expanded this definition to include obtaining data from non-traditional sources such as public and smart phone sensors or CWS connected via the internet. While the crowd-sourced air temperature data have the advantage of low cost devices and a high spatial resolution network, the quality of measurements is a crucial subject when using such data (Bell et al., 2015; Chapman et al., 2017; Meier et al., 2017; Napoly et al., 2018). As stated by Muller et al. (2015), "crowdsourcing has the potential to overcome the spatial and temporal representativeness of standard data", but overcoming the issue of quality is the biggest challenge when using it in science.

The quality of received data can be biased due to calibration issues, design flaws (e.g. an unventilated and sealed temperature sensor, Chapman et al., 2017), communication and software errors, and metadata issues (Bell et al., 2015). In addition, an unsuitable installation location chosen by the user (e.g. the installation of the outside module inside a house) can result in large biases. However, Meier et al. (2017) noted that the main error sources are primarily related to the siting of the sensor rather than the sensor quality itself. To address these issues, a careful quality assurance or Quality Control (QC) is necessary in order to deal with several common error sources.

To date, several studies have addressed the QC of temperature data from CWS. Chapman et al. (2017) studied nighttime temperature measurements by Netatmo weather stations for two months to quantify the urban heat island of London, United Kingdom. Their QC consisted of a simple statistical approach of removing measurements exceeding the threshold of three standard deviations from the mean of all stations. However, Chapman et al. (2017) mentioned, the QC step would have been more important in a study considering daytime values.

Meier et al. (2017) broadened the study period to one year and extended the QC to several levels addressing common error sources. They concluded that Netatmo, even being a non- traditional network, is suitable for observing the urban atmosphere. Nonetheless, calibrated and standardized sensor networks are essential to assure data quality. Hammerberg et al. (2018), using a modified QC based on the Meier et al. (2017) procedure, used Netatmo CWS data to obtain high-resolution temperature observations for Vienna. However, only a short time period was needed for the purpose of their study and only a fraction of stations was used compared to the current study.

Most recently, Napoly et al. (2018) developed an automated statistical QC based on data of the Netatmo stations themselves, without the need to use reference measurements from a professional network, thereby extending the possibility to use CWS data at locations without official monitoring stations. The developed QC was then applied to Berlin and Paris, to demonstrate the QCs transferability between cities (Napoly et al., 2018).

In the following, a QC method for CWS temperature observations, based on the procedure proposed by Napoly et al. (2018), is shown. In addition, the presented method specifically addresses radiative errors following Meier et al. (2017), thereby allowing the analysis of daytime data and diurnal temperature cycles. The QC methodology is applied to an example for the city of Vienna, Austria to study an 11-day high-temperature period in August 2018. A comparison between temperatures observed with weather stations from the Austrian national meteorological service (ZAMG) located in different urban areas and CWS in the vicinity is presented, and the results are discussed in light of differences in urban structure. Examples of the Vienna UHI are shown for two selected nights, and its temporal variability is addressed. Finally, the 2-D representation of CWS data is discussed.

2. Data and Methodology

Vienna, the Austrian capital, was chosen as study area. Vienna is the biggest city in Austria with a population of approximately 1.9 Million people (as of January, 2019) (Statistics Austria, 2019) and an urban structure typical of a Central European city. It consists of a densely built- up city center and a historic district. It is confined by the Vienna Woods to the West and by agricultural areas to the South, East and North, as can be seen in Figure 1. The city is transected by the Danube river, which separates the 21st (Floridsdorf) and 22nd (Donaustadt) district from the rest of the city. The highest point of Vienna is the Hermannskogel (542 m) in the North of the city. The lowest point Lobau (151 m) is located in the Southeast of Vienna.

To investigate the potential of Netatmo data for urban climate studies of Vienna, Austria, a study period from the 13th to the 23rd of August 2018 was chosen. This period was characterized by high temperatures in Vienna, where after the passing of a cold air front and cloudy weather conditions on August 14th, temperatures increased gradually as shown in Figure 2. Especially the last few days were characterized by temperatures well over 30°C, with a maximum of 34.8 °C recorded at the inner-city station Wien Innere Stadt on August 23rd at 14:00 UTC. The recorded maximum at the Hohe Warte station (shown in Figure 2) was 34.2°C on August 20th at 13:00 UTC. The prolonged heat period eventually ended on August 24th when a distinctive change in temperature, paired with rainfall, occurred.

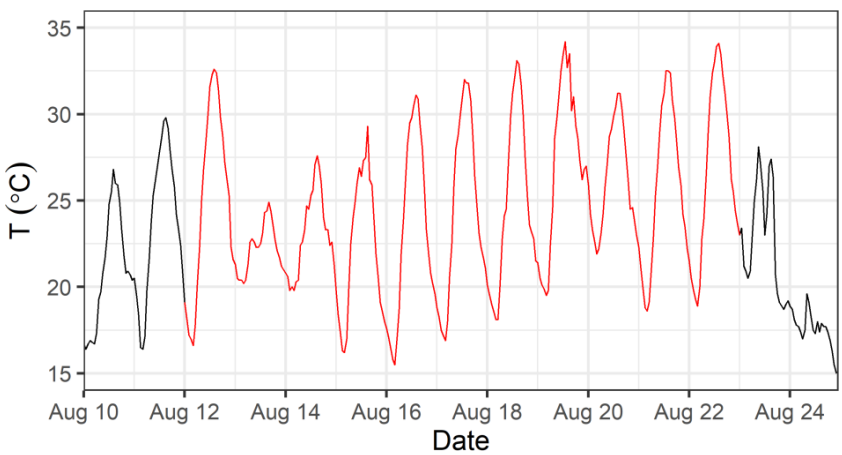

Figure 2: Temperature in Vienna (ZAMG station Wien Hohe Warte) from August 11th 2018 to August 25th 2018. The chosen period is shown in red. Dates are labelled at midnight (with time in UTC).

In this study, two kinds of data sources were used: crowd-sourced air temperature data, and 2-m air temperature observations from the ZAMG. The crowd-sourced dataset consisted of air temperature data from CWS from the company Netatmo. As the initial setup of the stations and data acquisition have been described in detail by Chapman et al. (2017), Meier et al. (2017), Fenner et al. (2017) and Napoly et al. (2018), only a brief summary is given below.

Stations by Netatmo consist of an indoor and an outdoor module enclosed in an aluminum shell. The indoor module measures air temperature (accuracy: ± 0.3°C), humidity (accuracy: ± 3%), barometric pressure (accuracy: ± 1 mbar), CO_2 level (accuracy: ± 50 ppm) and noise level. The outdoor module measures air temperature (accuracy: ± 0.3°C) and humidity (accuracy: ± 3%). The user can add further sensors to this standard weather station: additional indoor stations, a rain gauge (accuracy: 1 mm/h) and an anemometer for wind speed (accuracy: 0.5 m/s) and wind direction (accuracy: 5°) (Netatmo, 2019). Collected data are uploaded to the Netatmo server automatically, but shared publicly only if the user agrees. Public data can be downloaded via an application programming interface (API).

To evaluate the Netatmo measurements, a reference dataset was used consisting of 2-m air temperature observations obtained from the ZAMG operational weather station measurements taken at 10-minute intervals. For the purpose of comparison, hourly values were used. In addition, global radiation observations from the same network were incorporated in the QC. Except for the inner city station (Wien Innere Stadt), which is situated on a roof, the setup of the stations in the ZAMG network are World Meteorological Organization (WMO) synoptic station compliant (WMO, 2008) and measure air temperatures at 2-m level, over open areas with short grass.

In the reference period chosen for this study (August 13th to August 24th 2018), a total of 1357 single Netatmo stations sending air temperature data were available prior to application of the QC. According to Netatmo (2019), stations take measurements at a five-minute interval. However, timestamps differ as exact time of measurement depends on the initial setup of the station. To make the data comparable, values were assigned to the closest full hour in our analysis; e.g. 10:03 was assigned to 10:00 and 10:51 was assigned to 11:00. If more than one value was assigned to a full hour, the mean was calculated. On these data, the QC is applied.

In order to retain only the most trustworthy crowd-sourced air temperature data, a rigorous QC procedure was implemented (see Table 1). The approach followed in this study combines the work by Meier et al. (2017) and Napoly et al. (2018). Quality levels L1 to L5 are based on Napoly et al. (2018). Additionally, two further levels, L6 and L7, were introduced based on Meier et al. (2017). Napoly et al. (2018) introduced a statistic-based, automated procedure to filter crowd-sourced temperature data from typical errors, such as installation issues. The QC of Napoly et al. (2018) can be applied without the use of reference data from official meteorological networks, and is therefore easily implemented. To this end, an R-package ("CrowdQC", Grassmann et al., 2018; R Core Team, 2018) containing the necessary code was provided by the authors. The QC analysis consists of four main

levels and three optional levels of which all main and one optional level were used in the current study (Table 1). Please note that the QC is based on all data for the month of August, although results are presented for the reference period only.

Table 1: Quality levels and description of the shown method

Quality Level	**Description**	**Remaining data**	**Based on**
L1	Flag common location; Flag missing (NA) values	70.8 %	M1 in Napoly et al. (2018)
L2	Flag upper and lower part of the hourly distribution (1% and 95% quantile)	62.9 %	M2 in Napoly et al. (2018)
L3	Flag month if L2 flagged > 20% of the month	62.5 %	M3 in Napoly et al. (2018)
L4	Flag month if $R(T_{M3}, \text{median}(T_{M3}))_m < 0.9$	61.3 %	M4 in Napoly et al. (2018)
L5	Linear interpolation of hourly values (filling short data gaps)	65.1 %	O1 in Napoly et al. (2018)
L6	Systematic radiative error filter: positive and significant correlation between global radiation and air temperature difference $(T_{crowd} - T_{reference})$	57.8 %	C1 in Meier et al. (2017)
L7	Single value radiative error filter: flagging day-time values when air temperature difference $(T_{crowd} - T_{reference}) > 3 * SD$ in $T_{reference}$	55.3 %	C2 in Meier et al. (2017)

If upon station set up no location is specified by the Netatmo station user, an automatic location based on the IP address is determined, which often results in equal longitude and latitudes reported for multiple stations (Meier et al., 2017). Level L1 (M1 in Napoly et al., 2018) flags stations with equal longitude and latitude values, as this likely indicates an automatically assigned location not representing the true siting of the station. Additionally, Level L1 flags all missing (NA) values, which explains the high number of datasets not passing this particular level (Table 1).

In level L2 (M2 in Napoly et al., 2018) a statistical outlier detection is performed. To this end, prior to further analysis, altitude-corrected temperatures are calculated, to account for variations of temperature with height in order to adapt the methodology for cities with more complex topography. Owners of Netatmo stations are able to specify the height of their station. When inspecting the raw data, it is obvious that often faulty values are added (e.g. height above ground rather than altitude above mean sea level). To this end, altitude values of the digital elevation model from Geoland.at (2015) with a resolution of 10x10 meter were used. For the statistical outlier-analysis the quantiles used for flagging can be adapted; the defaults of 1% and 95% quantile were used in this study. Since Meier et al. (2017) concluded that most common error sources include increasing temperature (e.g. radiative errors) the upper outliers were treated more strictly by Napoly et al. (2018),

which was adopted in the current study as well.

Level L3 (M3 in Napoly et al., 2018) flags all data of a month, if the station is flagged more than 20% of that month in L2. In this way, overly erroneous stations are removed (Napoly et al., 2018). Level L4 (M4 in Napoly et al., 2018) calculates the Pearson correlation coefficient between a single station and the median of all stations, which have not been flagged previously, on the basis of one month. All values of a station of said month are flagged if the correlation coefficient is lower than 0.9. Finally, in level L5 (O1 in Napoly et al., 2018) missing values of single time steps are interpolated linearly on a station basis. 65.1% of raw data remained after this first part of the QC.

While a visual inspection of the time series pre- and post-QC showed that this procedure (L1- L5) already greatly smoothed the data and reduced unexplainable peaks (not shown), some exceptionally hot areas remained. This is illustrated in Figure 3a, which shows the temperature distribution for a grid with 500 x 500 m cell size, where for each grid cell all stations within a 2 km radius are averaged following an inverse distance weighting (compare Napoly et al., 2018; their Figure 10). Here, the difference between each grid cell and the lowest temperature in the entire Vienna area for August 20th 2018 at 12:00 UTC is shown. Further analysis showed that these 'hot spots' only occurred during the daytime, hence it was hypothesized this behavior is an artefact caused by radiative errors rather than it being a physical feature of the urban climate. Therefore, a more stringent global radiation test was introduced based on the QC introduced by Meier et al. (2017).

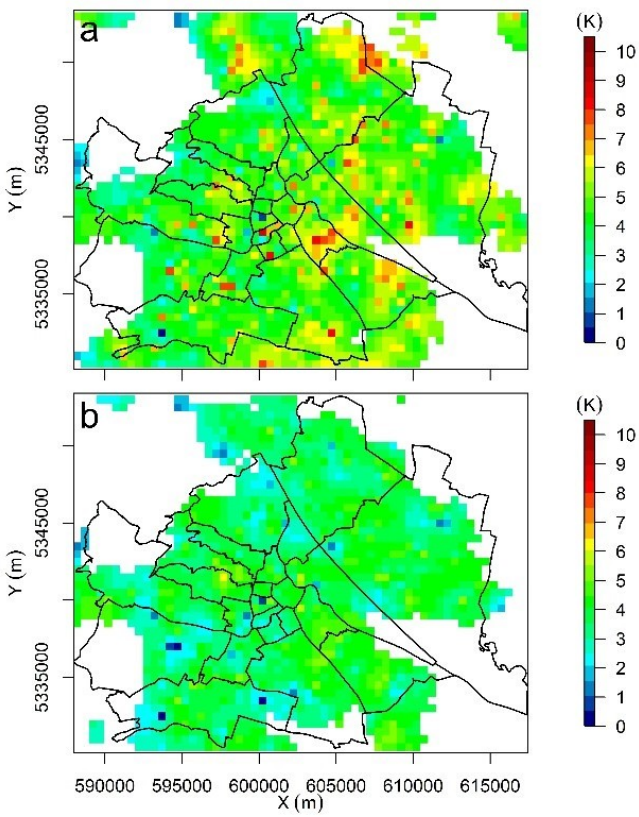

Figure 3: Temperature distribution (relative to the minimum value over the domain) based on Netatmo air temperature data for Vienna on August 20th 2018 at 12:00 UTC – Quality Level L5 (a) and L7 (b). Projection: WGS 84 / UTM zone 33N (EPSG: 32633), political boundaries: MA41 (2019).

Meier et al. (2017) developed four quality levels (A-D), including sub-levels, of which level C was used in the current study. This level handles faulty values due to radiative errors. Following level C1 from Meier et al. (2017), the Pearson correlation coefficient between the global radiation of the nearest ZAMG station and the hourly difference between the value of a single Netatmo and the mean of all reference stations in Vienna ($T_{crowd} - T_{reference}$) is calculated. If the correlation, which is calculated based on a full month of observations, is positive ($R > 0.5$) and significant ($p < 0.01$), it is assumed the observations are affected by solar radiation and the station is flagged for the specific month in level L6.

In level L7 (C2 in Meier et al. 2017), instead of filtering stations for a full month, single values with radiative errors are filtered. If the hourly difference between the value of Netatmo and the mean of all reference stations ($T_{crowd} - T_{reference}$) in Vienna was greater than three times the standard deviation of the reference stations of that hour, the single value was flagged.

As shown in Figure 3b, the unexplainable warm spots in Vienna disappeared after application of these additional QC steps, suggesting that indeed radiative errors were the cause. At the same time, cooler spots become apparent which are partially co-located with large green areas such as the Central Cemetery and the Schönbrunn palace park. Based on these results, the QC was extended with level L6-L7. Although this results in a further reduction of available data (55.3% of raw data passed this last quality level test), a total of 1083 stations are still available after the full QC procedure.

3. Results

To evaluate the CWS observations, the quality-controlled data set was compared to observations obtained with the official ZAMG network. Figure 4 shows the hourly 2-m air temperatures at three different ZAMG stations (bold, black line) in Vienna during the study period. Additionally, it shows all Netatmo stations that passed the full QC (up to level L7) in a radius of 2 km from the respective ZAMG station (thin, colored lines). Data from the CWS generally follow the trend of the reference stations, and the amplitude and phase of the two data sets are similar. Even disturbances such as the passage of a cold front on August 14th 2018, are well represented in the crowd-sourced observations. Nonetheless, it can be seen that the temperatures derived from most of the CWS observations are slightly higher than the temperatures of ZAMG stations Hohe Warte and Donaufeld (Figure 4a, c). The same behavior (CWS being warmer) applies to the other stations in Vienna, which are not shown. Interestingly, this situation is different for the Inner City station (Figure 4b), which is situated on a rooftop in a densely built-up area. Here, the temperatures of the ZAMG station fall well within the range observed by the CWS. Implications of these findings are discussed in Section 5.

In Figure 5, the spatial temperature distribution derived using the CWS is shown for two nights towards the middle of the study period, August 16th 2018 at 1:00 UTC (panel a) and August 17th 2018 at 00:00 UTC (panel b) on a 500 x 500 m grid, following the method used to create Figure 3.

It is clear that not the entire domain shown in Figure 5 is covered by CWS observations. Within the city borders, no measurements are available in the national park Lobau in the Southeast as well as the hilly forest areas in the West and agricultural areas in the East of the city. Also around Vienna, sparsely populated areas are marked by missing observations. Nevertheless, for the populated, urban areas high-resolution temperature information could be derived.

Based on Figure 5, it can be seen that the highest temperatures are recorded in the inner parts of the city as well as parts of the 21st (Floridsdorf) and 22nd (Donaustadt) district in the western part of the city on both nights, and that the temperatures decrease with distance away from the center. The lowest temperatures are found at the elevated areas in the western part of the city (on the edge of the Vienna Woods area), as well as towards the Lobau floodplain in the Southeast. Furthermore, Vienna's topographical features are reflected in the temperature distribution, and show up as horizontal, alternating bands of higher (valleys) and lower (hills) temperatures in the West.

Albeit being similar, the temperature patterns for the two nights are not identical. Although the temperatures in the center of the city (1st district, Wien Innere Stadt) as well as those towards the city borders and the surrounding areas are very similar, higher temperatures are recorded for large parts of the city during the first night than during the second.

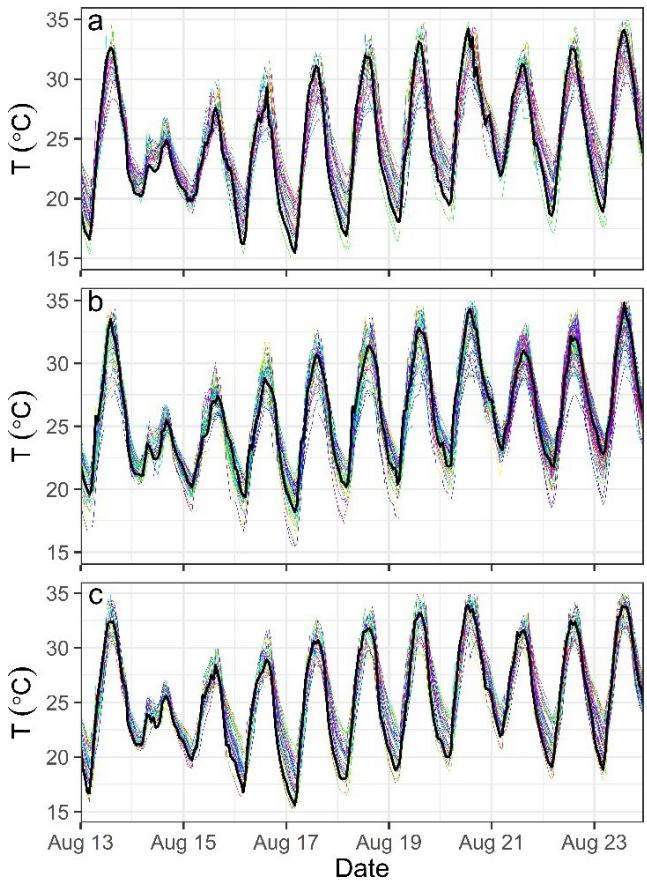

Figure 4: Air temperature near three different ZAMG stations in Vienna (a. Hohe Warte, b. Innere Stadt, c. Donaufeld) during the study period. Bold lines show air temperature at the ZAMG reference stations itself. Thin, colored lines show all quality-controlled Netatmo stations in a radius of 2 km of the reference station. Dates are labelled at midnight (with time in UTC).

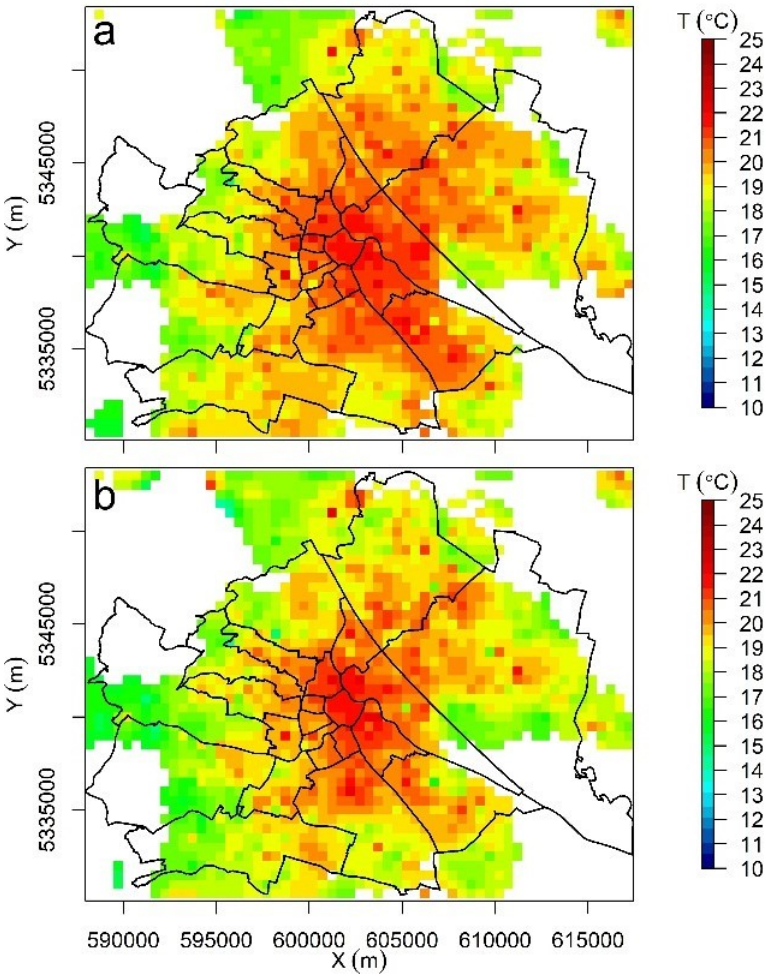

Figure 5: Spatial distribution of air temperature in Vienna, on August 16th 2018 at 01:00 UTC (panel a) and August 17th 2018 at 00:00 UTC (panel b). Projection: WGS 84 / UTM zone 33N (EPSG: 32633).

4. Discussion

The QC and code developed by Napoly et al. (2018) have two main advantages; (1) the code is freely available and easy to use, and (2) it relies solely on CWS data and no data from reference stations are needed, making it easily applicable in cities for which no reference data are available. However, it was found that the results for Vienna could be optimized by an additional QC step specifically designed to eliminate radiative errors. In this step, global radiation measurements from the ZAMG operational network were used, illustrating that in order to fully exploit the potential of crowd-sourced weather data reference stations might be beneficial in some cases.

Even after application of this optimized QC, a comparison of the CWS temperatures with the reference stations showed that CWS observed temperatures are generally higher than those observed by the reference stations (see Figure 4). While this may be interpreted as an overestimate of air temperatures by the CWS station due to the device design, a long-time comparison by Meier et al. (2017) showed good agreement between Netatmo and a Campbell CS215 reference station, with the Netatmo sensor only underestimating temperatures in the hours after sunrise and showing good agreement at all other times. Hence, it seems unlikely that these

results are solely due to measurement uncertainties and sensor inaccuracy.

Instead, these results may be explained by actual temperature differences between the reference location and the CWS sites. The CWS are mostly directly located in the built-up urban environment, whereas the reference stations are preferably located in open, green spaces. These locations have different energy balances, generally resulting in higher temperatures in the built-up environment in comparison to the green spaces. This interpretation is supported by the comparison between reference station Innere Stadt, which is located directly in the urban environment, and the CWS in a 2-km radius. Here, the reference temperatures are enveloped by the CWS observations, showing that temperatures are well represented when observing in similar environments. These findings increase our confidence that, after careful QC, CWS data can provide the valuable additional information related to urban microclimate, and holds great value for urban climate research.

Although the CWS provide valuable information about urban areas, Figure 5 clearly shows that not all areas of Vienna are covered by CWS. Especially for large nature areas such as the Lobau, and the edges of the Vienna Woods no CWS observations are available, limiting the possibility to capture information in uninhabited or sparsely populated areas. The topographical features of Vienna, on the other hand, are clearly reflected in the CWS-derived temperature distribution. This shows that the methodology of Napoly et al. (2018) can not only be used for cities with relatively flat topography (such as Berlin and Paris, with altitude differences of up to about 160 m), but also be used for cities with more complex terrain and higher altitude differences (approximately 400 m in case of Vienna).

In addition, it is important to keep in mind that the temperature distribution as shown in Figure 5 is based on interpolation of multiple stations. As a result, small-scale information may not be represented clearly or may be smoothed out using this method. This issue is illustrated in Figure 6, where three different representations of the temperature distribution on the 17th of August 2018 at 0:00 UTC are shown, where Figure 6b shows the method used to generate Figure 3 and Figure 5 with the location of the quality-controlled stations as black dots overlaid. In this method, stations in a radius of 2 km are averaged, applying an inverse distance weighting, to determine the temperature in each 500 x 500 m grid box. A clear advantage of this method is that the number of stations is increased when calculating the grid cell values, making the method less sensitive to possible outliers (Napoly et al., 2018) and resulting in a higher spatial coverage of the information. At the same time, however, it results in the mixed information and smoothing of small-scale features. This issue shows best, when looking at the positions of stations in relation to the grid boxes with temperature information. Indeed, it can be seen that grid box values are also available in areas where no stations are located; e.g. areas in and between Kurpark Oberlaa and the Vienna Central Cemetery in the south of Vienna. The same is true for the Danube River (indicated by the diagonal line crossing Vienna from the Northwest to the Southeast), a feature generally clearly visible in urban climate studies of Vienna (e.g. Žuvela-Aloise, 2016) yet not showing in the current

representation.

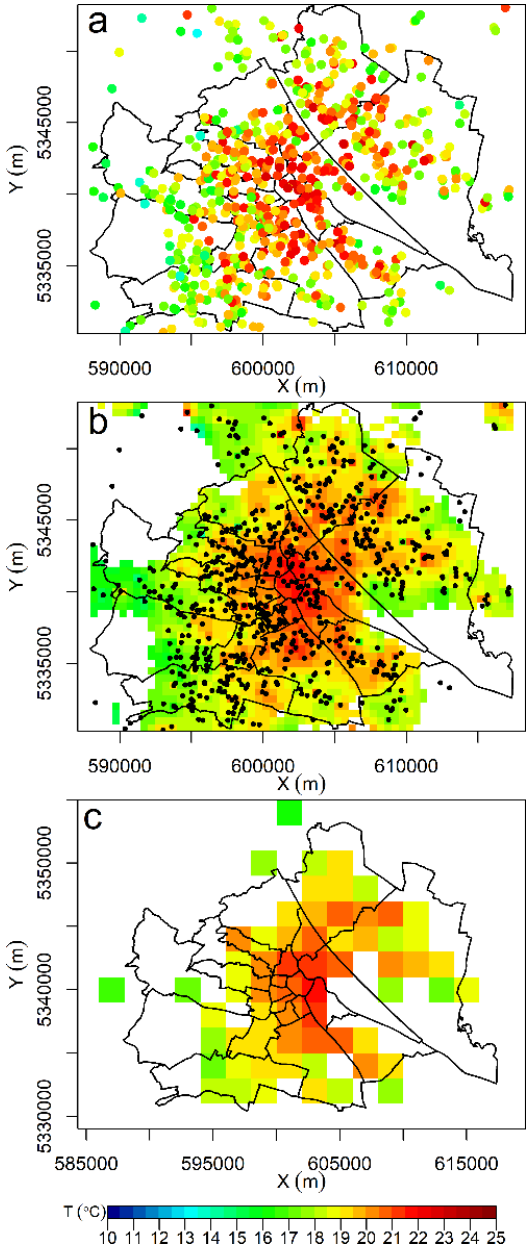

Figure 6: Spatial temperature distribution on August 17th 2018 at 00:00 UTC using Discrete Points (panel a), inverse distance weighting interpolation (panel b, only averaging when a minimum of 5 values in a radius of 2 km are available), and the mean of stations in grid box (panel c, only averaging when a minimum of 5 stations is available in the 2 x 2 km grid box).

To solve the issue of mixing of small-scale information, there are two possibilities. On the one hand, the temperatures from single stations can be shown as discrete points, such as illustrated in Figure 6a. Here, only the local information is presented, and no averaging is applied, potentially decreasing the trustworthiness of the result (Napoly et al., 2018). Based on this view, the visual recognition of spatial patterns is more complicated than based on Figure 6b, however, for a comparison to e.g. urban model results these data are of use. Additionally, the general pattern of temperature (decreasing towards the outer bounds of the city) can clearly be observed also when relying on single CWS. Another possibility is enlarging the cell size and only average over the stations in that specific grid box, shown in Figure 6c. As only stations in the cell are averaged, no mixing of information occurs. However,

large cell sizes are needed to balance the number of stations over which is averaged (at least five in this case) to reduce the effect of outliers. While all versions are correct, it is important to consider artefacts may arise resulting from the chosen gridding when interpreting and using these data.

5. Conclusion and outlook

In this study, quality-controlled crowd-sourced CWS air temperature observations provided by Netatmo were used to study the high-resolution air temperature distribution in the Austrian capital, Vienna, for an 11-day heat wave period in August 2018. The QC was mainly based on the procedure developed by Napoly et al. (2018), which considers common error sources in crowd-sourced temperature data such as installation errors, statistical outliers and inconsistency in metadata. Even though Vienna's topography is more complex than the cities studied in Napoly et al. (2018), the analysis has shown that the method produces reliable results. However, it was found that the results could further be optimized by incorporating two additional steps based on Meier et al. (2017) addressing the specific issue of solar radiative errors.

Through various steps of analysis, it was found that after a rigorous QC, CWS air temperature data provide results comparable to operational stations, and that differences between these data sets could be explained by differences in urban structure. Indeed, Vienna's land cover structure (see also Figure 1) can clearly be recognized in the temperature distribution maps, with highest temperatures found in the built-up areas in the city center, and lowest temperatures in the green areas at the city borders.

In Vienna, 1357 unique stations were available during the study period, of which 1083 stations passed the full QC procedure. Compared to the 15 available stations of the combined ZAMG and City of Vienna observational network, this represents a major improvement in the spatial resolution of available temperature data. Considering the high spatial resolution of usable data points, CWS hold great potential for the analysis of the Vienna UHI as well as intra-urban heat islands, which is only expected to increase further as the number of CWS increases. In this case, the possibility to analyze small-scale features will increase. Additionally, the problem of missing data as well as averaging and mixing of information by taking into account a relatively large radius of influence would be reduced, thereby increasing the quality of and resolution of the data further.

Based on the current analysis, quality-controlled Netatmo data for Vienna seem a promising data source for further urban climate studies of the Vienna area. As a next step, high-resolution urban model results will be validated and compared to the Netatmo data, to determine how well the intra-urban temperature distributions match between the two methods. Moreover, to further strengthen the confidence in the usability of CWS data, a Netatmo station was set up in the immediate vicinity of the ZAMG Hohe Warte station in January 2019, making it possible to obtain more information on biases similar to Meier et al. (2017). These results will be reported in a future study.

Acknowledgements

The authors would like to thank the owners of Netatmo stations as well as Netatmo for making their temperature data freely available online. This paper has been made possible by the availability of the temperature data on the Netatmo website. Furthermore, we would like to thank D. Fenner for interesting discussions and the support related to the data acquisition.

Funding: MF was supported by the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT) through a FEMTech grant.

References

Bell S., Cornford D. and Bastin L. (2015): How good are citizen weather stations? Addressing a biased opinion. In: Weather 70, 3, 75-84. doi: 10.1002/wea.2316

Chapman L., Bell C. and Bell S. (2017): 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. In: International Journal of Climatology 37, 3597-3605. doi: 10.1002/joc.4940

Fenner D., Meier F., Bechtel B., Otto M., Scherer D. (2017): Intra and inter 'local climate zone' variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. In: Meteorologische Zeitschrift 26, 5, 525-547. doi: 10.1127/metz/2017/0861

Geoland.at (2015): Digital terrain model of Austria, based on airborne laserscanning. https://www.data.gv.at/katalog/dataset/dgm (last accessed 15.01.2019)

GCOS (2017): Global Climate Observing System: Austrian Inventory Report 2017, Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMWFW), Zentralanstalt für Meteorologie und Geodynamik (ZAMG). Eds.: Silke Adler and Hermine Fürst, ZAMG, pp 113. available: https://www.zamg.ac.at/cms/de/dokumente/topmenu/gcos/gcos-report-2017

Grassmann T., Napoly A., Fenner D. and Meier F. (2018): CrowdQC. Quality control for crowdsourced data from CWS. R package - version 1.2.0.

Hammerberg K., Brousse O., Martilli A. and Mahdavi A. (2018): Implications of employing detailed urban canopy parameters for mesoscale climate modelling. A comparison between WUDAPT and GIS database over Vienna, Austria. In: International Journal of Climatology 38, Suppl. 1, e1241-e1257. doi: 10.1002/joc.5447

Howe J. (2006): Crowdsourcing. A Definition. https://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html (last accessed 21.02.2019)

Landsberg H.E. (1981): The Urban Climate, Academic Press, New York.

MA41 (2019): Political boundary of the Federal Province of Vienna, Magistrat Wien - Magistratsabteilung 41 – Stadtvermessung, UUID: 86c29af5-c52c-423d 8764-82c1bb037936 https://www.data.gv.at/katalog/dataset/86c29af5-c52c-423d 8764-82c1bb037936.

Meier F., Fenner D., Grassmann T., Otto M. and Scherer D. (2017): Crowdsourcing air temperature from citizen weather stations for urban climate research. In: Urban Climate 19, 170-191. doi: 10.1016/j.uclim.2017.01.006.

Muller C. L., Chapman L., Johnston S., Kidd C., Illingworth S., Foody G., Overeem A. and Leigh R. R. (2015): Crowdsourcing for climate and atmospheric sciences: current status and future potential. In: International Journal of Climatology 35: 3185-3203. doi: 10.1002/joc.4210

Napoly A., Grassmann T., Meier F. and Fenner D. (2018): Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data. In: Frontiers in Earth Science 6, Article 118. doi: 10.3389/feart.2018.00118

Netatmo (2019): Smart Home Weather Station's technicals specifications. https://www.netatmo.com/en-us/weather/weatherstation/specifications (last accessed 11.01.2019)

Oke T. (1995): The heat island of the urban boundary layer: Characteristics, causes and effects. In Cermak, J. E. et al. (eds.), Wind Climate in Cities, Kluwer Academic Publishers, 81- 107.

R Core Team (2018): R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (last accessed 21.02.2019)

Statistics Austria (2019): Population at the beginning of the year since 1981 by Länder. http://www.statistik.at/web_en/statistics/PeopleSociety/population/population_stock_and_population_change/total_population_at_the_beginning_of_the_year_quarter/106933.html (last accessed 30.07.2019)

WMO (2008): Guide to Meteorological Instruments and Methods of Observation. ISBN 978- 92-63-10008-5. https://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf (last accessed 07.02.2019)

Yokobori T. and Ohta S. (2009): Effect of land cover on air temperatures involved in the development of an intra-urban heat island. Clim Res 39: 61–73. doi: 10.3354/cr00800

Žuvela-Aloise M., Koch R., Buchholz S., Früh B. (2016) Modelling the potential of green and blue infrastructure to reduce urban heat load in the city of Vienna. Clim. Chang. doi:10.1007/s10584-016-1596-2