https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium

Eva Beele[1], Maarten Reyniers[2], Raf Aerts[1,3,4], Ben Somers[1,5]

[1] Division Forest, Nature and Landscape, University of Leuven (KU Leuven), Celestijnenlaan 200E-2411, BE-3001 Leuven, Belgium
[2] Royal Meteorological Institute of Belgium, Ringlaan 3, BE-1180 Brussels, Belgium
[3] Risk and Health Impact Assessment, Sciensano (Belgian Institute of Health), Juliette Wytsmanstraat 14, BE-1050 Brussels, Belgium
[4] Division Ecology, Evolution and Biodiversity Conservation, University of Leuven (KU Leuven), Kasteelpark Arenberg 31-2435, BE-3001 Leuven, Belgium
[5] KU Leuven Urban Studies Institute, University of Leuven (KU Leuven), Parkstraat 45-3609, BE-3000 Leuven, Belgium

Correspondence to: Eva Beele (eva.beele@kuleuven.be)

Abstract.

The growing urbanization trend and increasingly frequent extreme weather events urge further monitoring and understanding of weather in cities. In order to gain information on these intra urban weather patterns, dense high quality atmospheric measurements are needed. Crowdsourced weather stations (CSW) could be a promising solution to reach such monitoring networks in a cost-efficient way. Because of their non-traditional measuring equipment and installation settings, the quality of these datasets remains however an issue of concern. This paper presents crowdsourced data from the Leuven.cool network, a citizen science network of around 100 low-cost weather stations (Fine Offset WH2600) distributed across Leuven, Belgium. The dataset is accompanied by a newly developed station specific temperature quality control (QC) and correction procedure. The procedure consists of three levels removing implausible measurements, while also correcting for inter (in between stations) and intra (station-specific) station temperature biases by means of a random-forest approach. The QC method is evaluated using data from four WH2600 stations installed next to official weather stations belonging to the Royal Meteorological Institute of Belgium (RMIB). A positive temperature bias with strong relation to the incoming solar radiation was found between the CSW data and official data. The QC method is able to reduce this bias from 0.15 ± 0.56°C to 0.00 ± 0.22°C. After evaluation, the QC method is applied to the data of the Leuven.cool network, making it a very suitable data set to study in detail local weather phenomena such as the urban heat island (UHI) effect.

1 Introduction

More than 50% of the world population currently lives in urban areas and this number is expected to grow to 70% by 2050 (UN, 2018). Keeping this growing urbanisation trend in mind and knowing that both the frequency and intensity of extreme weather events will increase (IPCC, 2021), it becomes clear that both our cities and its citizens are vulnerable for climate change. To plan efficient mitigation and adaptation measures, and hence mitigate future risks, information on intra-urban weather patterns is needed (Kousis et al., 2021). Dense high-quality atmospheric measurements are thus becoming increasingly important to investigate the heterogeneous urban climate. Due to their high installation and maintenance costs and strict siting instructions (WMO, 2018), official weather station networks are however sparse. As a results, most cities only have one or even no official station at all (Muller et al., 2015). Belgium only counts around 30 official weather stations distributed across a surface area of 30,689 km^2. 18 of them (Sotelino et al., 2018) are owned and operated by the Royal Meteorological Institute of Belgium (RMIB). These classical observation networks operate at a synoptic scale and are thus not suitable to observe city-specific or intra-urban weather phenomena such as the urban heat island (UHI) effect (Chapman et al., 2017).

The UHI can be measured by a number of methods. Fixed pair stations (e.g., Bassani et al., 2022; Oke, 1973) or mobile transect approaches (e.g., Kousis et al., 2021) have traditionally been used to quantify this phenomenon. Both methods are however not ideal as pair stations lack detailed spatial information while transects often miss a temporal component (Chapman et al., 2017; Heaviside et al., 2017). Other studies have quantified the UHI using remote sensing data derived from thermal sensors. Such methods can provide spatially continuous data over large geographical extents but are limited to land surface temperatures (LST) (Arnfield, 2003; Qian et al., 2018). As opposed to LST, canopy air temperatures (Tair) are however more closely related to human health and comfort (Arnfield, 2003). Finding the relationship between LST and Tair is known to be rather difficult and inconsistent (Yang et al., 2021). Numerical simulation models (e.g. UrbClim (De Ridder et al., 2015), SURFEX (Masson et al., 2013)) in which air temperatures are continuously modelled over space and time could be a possible solution. They do however still have some drawbacks. Due to computational power capacity, models only take into account a limited number of variables, making them less suitable for real-life applications (Rizwan et al., 2008). Additionally, they often lack observational data to train and validate their simulations (Heaviside et al., 2017).

The rise of crowdsourced data, especially in urban areas, could be a promising solution to bridge this knowledge gap (Muller et al., 2015). Such data are obtained through a large number of non-traditional sensors, mostly set up by citizens (cf. citizen science) (Bell et al., 2015; Muller et al., 2015). Crowdsourced datasets have already been successfully used for monitoring temperature (Chapman et al., 2017; Feichtinger et al., 2020; Fenner et al., 2017; Hammerberg et al., 2018; Meier et al., 2017; Napoly et al., 2018; de Vos et al., 2020), rainfall (de Vos et al., 2020, 2017, 2019), wind speed (Chen et al., 2021; de Vos et al., 2020) and air pollution (Castell et al., 2017; EEA, 2019) within complex urban settings. Because of their non-traditional measuring equipment and installation settings, the quality of these datasets remains however an issue of concern (Bell et al., 2015; Chapman et al., 2017; Cornes et al., 2020; Meier et al., 2017; Muller et al., 2015; Napoly et al., 2018; Nipen et al., 2020). Quality uncertainty arises due to several issues: (1) calibration issues in which the sensor could be biased either before the installation or drifts over time, (2) design flaws in which the design of the station makes it susceptible to inaccurate observations, (3) communication and software errors leading to incorrect or missing data, (4) incomplete metadata (Bell et al., 2015) and (5) unsuitable installation locations (Cornes et al., 2020; Feichtinger et al., 2020).

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Recent studies have therefore highlighted the importance of performing a data quality control in data processing applications (Båserud et al., 2020; Longman et al., 2018), especially before analysing crowdsourced temperature data (Bell et al., 2015; Chapman et al., 2017; Cornes et al., 2020; Feichtinger et al., 2020; Jenkins, 2014; Meier et al., 2017; Napoly et al., 2018; Nipen et al., 2020). Jenkins et al. (2014) and Bell et al. (2015) both conducted a field comparison in which multiple crowdsourced weather stations (CWS) were compared with official, and thus professional, observation networks. Both found a profound positive instrument temperature bias during daytime with strong relation to the incoming solar radiation. The use of crowdsourced data thus requires quality assurance and quality control (QA/QC) that both removes gross errors and corrects station-specific instrument biases (Bell et al., 2015). Using the findings of Bell et al. (2015) as a basis, Cornes et al. (2020) corrected crowdsourced air temperature data across the Netherlands using radiation from satellite imagery and background temperature data from official stations belonging to the Royal Netherlands Meteorological Institute (KMNI). To investigate the UHI in London, UK, Chapman et al. (2017) used Netatmo weather stations and removed crowdsourced observations that deviated more than three standard deviations from the mean of all stations. Meier et al. (2017) developed a detailed QC procedure for Netatmo stations using reference data from two official observation networks in Berlin, Germany. The QC consists of four steps, each identifying and removing suspicious temperature data. Their methods highlight the need for standard, calibrated and quality-checked sensors in order to assess the quality of crowdsourced data (Chapman et al., 2017; Cornes et al., 2020; Meier et al., 2017). Such official sensors are however not present in most cities, hindering the transferability of these QC methods. To this end, Napoly et al. (2018) developed a statistically based QC method for Netatmo stations independent of official networks (the R-package *CrowdQC*). The QC method was developed on data from Berlin, Germany and Toulouse, France and was later applied to Paris, France to demonstrate the transferability of this method. The procedure consists of four main and three optional QC levels, removing suspicious values, correcting for elevation differences and interpolating single missing values. Since the *CrowdQC* filtered dataset still contained some radiative errors, Feichtinger et al. (2020) combined the methods of Napoly et al. (2018) and Meier et al. (2017) to study a high temperature period in August 2018 in Vienna. Most recently, Fenner et al. (2021) presented the QC R-package *CrowdQC+*, which is a further development of the existing package *CrowdQC* developed by Napoly et al. (2018). The core enhancements deal with radiative errors and sensor response time issues (Fenner et al., 2021).

Current QC studies mostly identify and remove implausible measurements (Chapman et al., 2017; Meier et al., 2017; Napoly et al., 2018), instead of correcting for known temperature biases (Cornes et al., 2020). We do however know that both the siting and the design of CWS can introduce such a bias. By parameterising this bias, it can be learned and corrected for, hereby limiting the number of observations that is eliminated (Bell et al., 2015). Additionally, most QC procedures require data from official networks (Chapman et al., 2017; Cornes et al., 2020; Meier et al., 2017), while most cities do not have such measurements available (Muller et al., 2015). Lastly, previous research also noted that biases can be station specific, this because the design of a CWS is an important uncertainty source (Bell et al., 2015), indicating the need for station-specific

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

quality control methods. There is thus a need for station specific quality control and correction methods, independent of official weather station networks.

Here we report a statistically-based QC method for the crowdsourced air temperature data of the Leuven.cool network, a citizens science network of almost 100 weather stations distributed across Leuven, Belgium. The Leuven.cool network is a uniform network in the sense that only one weather station type (Fine Offset WH2600) is used for the entire network. To our knowledge, no quality control method has been developed for this sensor type. The stations were installed following a strict protocol, lots of metadata is available and both the dataflow and station siting are continuously controlled. This novel QC method removes implausible measurements, while also correcting for inter (in between stations) - and intra (station-specific) station temperature biases. The QC method only needs an official network during its development and evaluation stage, afterwards the method can be implemented independent of an official network.

The paper is organised as follows. Section 2 describes materials and methods, providing information on the study area, crowdsourced (Leuven.cool) dataset and official reference dataset. The development of the quality control method is explained in Section 3. In Section 4 the newly developed QC method is first tested on four crowdsourced stations installed next to three official stations from the Royal Meteorological Institute of Belgium (RMIB). This allows us to quantify the data quality improvement after every QC level. In Section 5 the QC method is applied to a network of CWS in Leuven, Belgium. Concluding remarks are summarized in Section 6. After applying this quality control and correction method, the crowdsourced Leuven.cool dataset becomes suitable to monitor local weather phenomena such as the urban heat island (UHI) effect.

2 Materials & methods

2.1 Study area

The QC method is developed for a citizens science weather station network “Leuven.cool”, based in Leuven, Belgium (50°52’39” N 4°42’16” E). The Leuven.cool project is a close collaboration between the KU Leuven, the city of Leuven and the RMIB aiming to measure the micro-climate in Leuven and gain knowledge on the mitigating effects of green and blue infrastructures (Leuven.cool, 2021). Leuven has a warm temperate climate with no dry season and a warm summer (Cfb) with no influence from mountains or seas and overall weak topography (Kottek et al., 2006). Leuven is the capital and largest city of the province of Flemish Brabant and is situated in the Flemish region of Belgium, 25 kilometres east of Brussels, the capital of Belgium. The city comprises the districts of Leuven, Heverlee, Kessel-Lo, Wilsele and Wijgmaal, covering an area of 56.63 km^2. The main characteristics of the study area are summarized in Table 1.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Table 1: Main characteristics of the study area

Climate		
Annual Min/Mean/Max daily temperature (°C)	6.9/11.2/15.5	Leuven, 1991-2020 (RMI, 2020)
Mean annual rainfall (mm y^{-1})	780.7	Leuven, 1991-2020 (RMI, 2020)
Köppen's classification	Cfb	(Kottek et al., 2006)
Demographics		
Size (km^2)	56.63	Figure 2
Population	101 315	(Bevolkingsregister Stad Leuven, 2021)

2.2 Leuven.cool dataset

Data from the citizens science network Leuven.cool are presented in this paper. The crowdsourced weather station network consists of 98 weather stations distributed across Leuven and surroundings. The meteorological variables are measured by low-cost consumer weather stations produced by the manufacturer Fine Offset: the WH2600 wireless digital weather station (Figure 1). The station's specifications, as defined by the manufacturer, are summarized in Appendix A.1. The weather station consists of an outdoor unit (sensor array) and a base station. The outdoor sensor array measures temperature, humidity, precipitation, wind speed, wind direction, solar radiation, and UV every 16 seconds. This outdoor sensor array transmits its measurements wirelessly, through the 868MHz radiofrequency, to the base station. This base station needs both power and internet - via a LAN connection - supply in order to send the data to a server. The data is forwarded to the Weather Observations Website (Kirk et al., 2020), a crowdsourcing platform initiated and managed by the UK Met Office. RMIB participates in this initiative and operates its own WOW portal (WOW-BE, 2021). The outdoor unit is powered by three rechargeable batteries which are recharged by a small built-in solar panel. A radiation shield protects both the temperature and humidity sensors against extreme weather conditions and the direct exposure of solar radiation.

Figure 1: The outdoor unit of the WH2600 wireless digital weather station at the Mathieu Layensplein in Leuven (LC-105) (a) and next to the official AWS equipment in Humain (LC-R05) (b). Pictures: Maarten Reyniers.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

From July 2019 onwards the weather stations were distributed along an urban gradient from green (private) gardens to public (semi-)grey locations following a stratified sampling design (Figure 2). The stratification was based on the concept of Local Climate Zones (LCZ) (Stewart and Oke, 2012). This LCZ scheme was originally developed as an objective tool for classifying urban-rural gradients, herby capturing important urban morphological characteristics (Verdonck et al., 2018). Stewart & Oke (2012) formally define these zones as "regions of uniform surface cover, structure, material, and human activity that span hundreds of meters to several kilometres in horizontal scale".

Stewart & Oke (2012) define 17 LCZ classes, divided into 10 urban LCZs (1-10) and 7 natural LCZs (A-G) . A LCZ map for Leuven was developed following a methodology proposed by Demuzere et al. (2021). Details on this LCZ map are available in Appendix B. Table 2 summarises the LCZ present in Leuven and the number of weather stations in each LCZ class.

Table 2: LCZs present in Leuven and the number of weather stations in each LCZ class.

LCZ ID	LCZ description	# Stations
LCZ 2	Compact midrise	15
LCZ 3	Compact low-rise	9
LCZ 5	Open midrise	16
LCZ 6	Open low-rise	29
LCZ 8	Large low-rise	12
LCZ 9	Sparsely built	12
LCZ A	Dense trees	0
LCZ B	Scattered trees	2
LCZ D	Low plants	3
LCZ G	Water	0

It can be noted that the weather stations are not evenly distributed across the different LCZ classes. Due to the complex urban settings in which the network is deployed, practical limitations apply to the eligible locations for installation. We rely on volunteering citizens, private companies and government institutions giving permission to install a weather station on their property. Further, the middle-sized city of Leuven does not contain all available LCZ classes. In the urban context high-rise, lightweight low-rise and heavy industry is missing. In the natural context, brush or shrub vegetation, bare rock or paved and bare soil or sand are not present in sufficiently large areas (Table 2). Lastly, the number of stations within more natural settings are limited due to the technical limitations of the weather stations; each outdoor unit needs a base station, with both power and LAN connection, within 50 to 100 meters in order to transmit its data.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.
Open Access Earth System Science Data Discussions

The weather stations were installed according to a strict protocol. In private gardens the weather stations were installed at 2 meters height using a steel pole with a length of 2.70 meters. Dry concrete was used to anchor the pole into the soil at a depth of 70 cm. Following the station's guidelines, stations were installed at an open location within the garden, at least 1 meter from interfering objects, such as nearby buildings and trees. In order to maximize the absorption of solar radiation by the solar panel and to assure correct measurements of wind direction and precipitation, the weather station was levelled horizontally and the solar panel of the weather station was directed towards the south Weather stations located on public impervious surfaces were installed on available light poles using specially designed L-structures to avoid direct effect contact with the pole. For security reasons an installation height between 3-4 meters was used.

The data is currently available from July 2019 (2019Q3) until December 2021 (2021Q4). The dataset can be downloaded in periods of three months and is thus available for each quarter. The raw 16 seconds measurements are aggregated to 10 minutes observations. This is done for three reasons: (1) an extremely high temporal resolution of 16 seconds is too high for most meteorological analyses, (2) the aggregation to 10 minutes is necessary to exclude the natural small-scale variability and noise on the observations and most importantly (3) the reference dataset of official measurement is only available in a 10 minute resolution. After resampling the data, some basic data manipulation steps are performed to obtain the correct units and resolution for every meteorological variable. The final dataset contains air temperature with the three quality level stages (see further), relative humidity, dewpoint temperature, solar radiation, rain intensity, daily rain sum, wind direction, and windspeed. We must stress that only the air temperature measurements undergo a quality check and correction procedure, further explained in the next sections. The variables other than temperature are, however, used in the correction procedure.

2.3 Reference dataset

Standard, calibrated and quality controlled reference measurements are used to develop the QC method and evaluate its performance. Since no official measurements are available in Leuven, we used data from three official RMIB stations in Uccle (6447 – 50.80°N 04.26°E, alt 100m), Diepenbeek (6477 - 50.92°N 5.45°E, alt 39m) and Humain (6472 - 50.19°N 5.26°E, alt 295m) (Figure 2).

The meteorological observation network of the RMIB consists of 18 automatic weather stations (AWS), ensuring continuous data collecting and limiting human errors. These weather stations report meteorological parameters such as air pressure, temperature, relative humidity, precipitation (quantity, duration), wind (speed, gust, direction), sunshine duration, shortwave solar radiation and infrared radiation every 10 minutes. The AWS network is set up according to the WMO guidelines (WMO, 2018).

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Since there is no AWS station available in the region of Leuven, four low-cost WH2600 weather stations were installed next to the official and more professional equipment of the RMIB in Uccle, Diepenbeek and Humain. Since these stations will serve as a reference, they were defined as LC-R01, LC-R02, LC-R04 and LC-R05 (Table 3). LC-R03 was installed for a short time in Diepenbeek, but has been removed due to communication problems and is not taken into account in our further analysis. Since January 2020, the oldest reference station LC-R01 is no longer active. This setup enables us to calculate the temperature difference or bias between the low cost reference stations and the official RMIB stations in Uccle, Diepenbeek and Humain.

Table 3: Specifics of Leuven.cool low-cost reference stations.

Station ID	Location	Installation date
Leuven.cool R01	Uccle	11/09/2018
Leuven.cool R02	Uccle	02/09/2019
Leuven.cool R04	Diepenbeek	06/11/2019
Leuven.cool R05	Humain	20/08/2020

In the rest of the paper, the terminology of Table 4 is used to refer to the different datasets and stations.

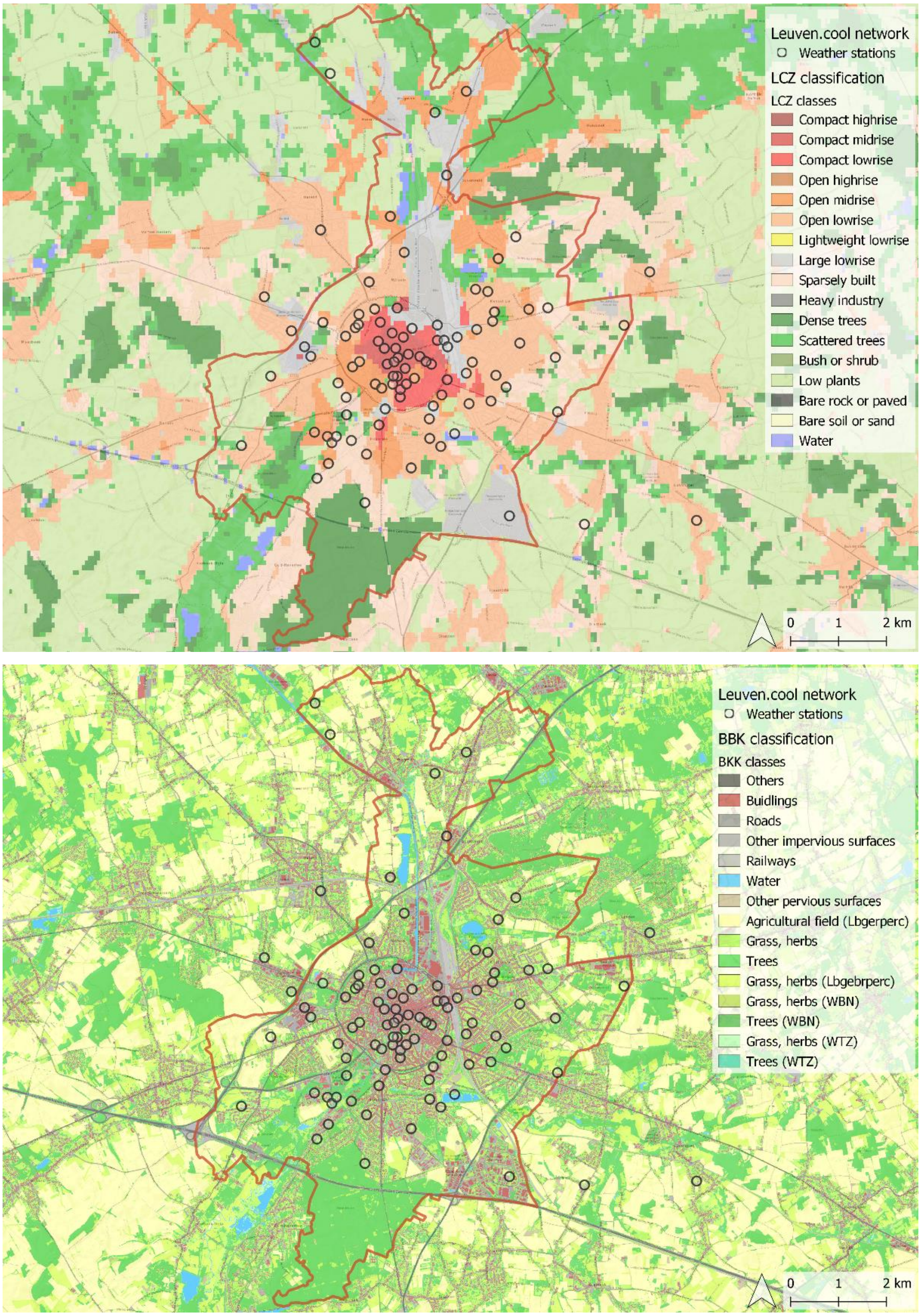
Table 4: Terminology of datasets and stations used in this paper

Terminology	Description
LC-X	The Leuven.cool (WH2600) stations installed in the study area (area of Leuven, Belgium)
LC-R	The Leuven.cool (WH-2600) stations installed next to the official weather stations operated by RMIB.
AWS	The Automatic Weather Stations owned and operated by RMIB. In our study, the AWS in Uccle, Diepenbeek and Humain are used.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

215



https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

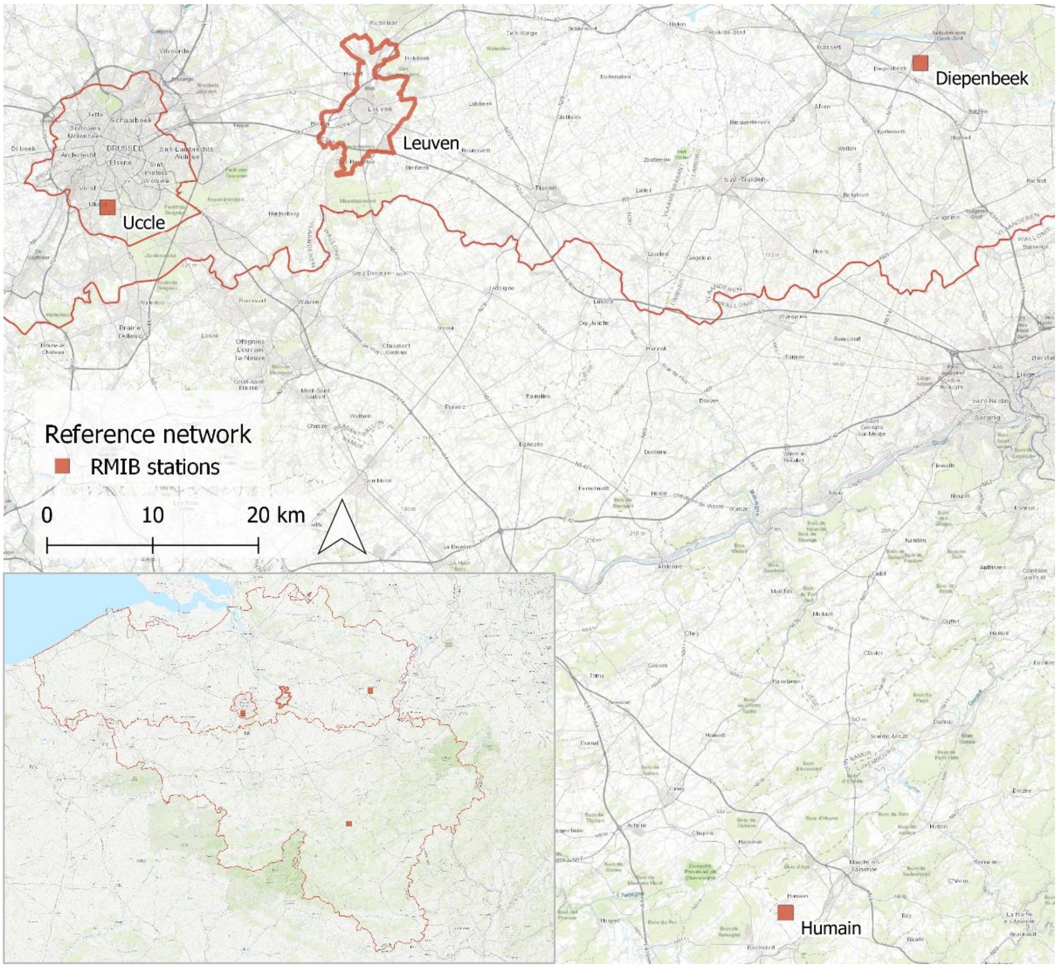


Figure 2: The Leuven.cool network (LC-X) with LCZ classification (a) and BBK (bodembedekkingskaart; land use map) classification (b) and Belgium delineated by the three official regions (Flanders, Wallonia and Brussels) with the location- of Leuven and the three RMIB stations (AWS) used in this study (c). Background map: Esri.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

3. Description of the quality control and correction method

The newly developed QC control method consists of three levels (Table 5), mostly focussing on eliminating calibration issues, design flaws and communication or software errors. Due to the strict installation protocol used for the Leuven.cool station network, some of the typical uncertainty sources are a priori discarded. Both the location and metadata of each station were controlled by experts, eliminating incomplete metadata or unsuitable installation locations. We further know that the low cost station used in this study has some design flaws (e.g. during clear sky with low sun conditions both the radiation and thermometer sensors experience shadow from the anemometer). Our correction method, however, is designed in such a way that these errors will be accounted for.

The first QC level removes implausible values mostly caused by software or communication errors. The second and third level correct for temperature biases. Both fixed inter (in between stations) biases due to sensor calibration uncertainties and variable intra (station-specific) biases due to the station's design and siting are parameterized and corrected for.

Table 5: Quality control levels, criteria for data filtering and potential error sources for crowdsourced air temperature measurements.

Quality control level	Description	Potential error sources
L1 Outlier detection		
L1.1 Range test	Range check against climatological extremes	Sensor malfunctioning
L1.2 Temporal outliers	Ensure realistic change in magnitude between consecutive observations of a specific station	Battery loss, server failure, connection issues, sensor malfunctioning
L1.3 Spatial outliers	Ensure realistic observation compared to neighbouring stations	Battery loss, server failure, sensor malfunctioning, outdoor sensor set up inside (not applicable in our setup due to the installation by team members)
L2 Inter station bias correction	Model the fixed in-between temperature bias	Sensor calibration issues
L3 Intra station bias correction	Model the variable station-specific temperature bias	Design flaws, outdoor sensor set up in sunlit conditions (no active ventilation)

3.1 Quality control level 1 – Outlier detection

The outlier detection algorithm uses a flag system in which every 10 min observation is assigned flag of either 0, 1 or -1 referring to "no outlier", "outlier" or "not enough information to determine whether observation is an outlier". The outlier detection method consists of three steps: a range test, a temporal outlier test and a spatial outlier test. The thresholds of the parameter settings used during each of these steps are explained in Table 6. We used an iterative procedure for threshold

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

optimization. Observations which received a flag of 1 and are thus defined as outliers are set to NA in the quality controlled (QC) level 1 dataset, hereby not considered during the following QC levels.

Table 6: Parameter settings for QC level 1 - Outlier detection.

Outlier parameter	Value (unit)	Description
Range outliers (RO)		
dev_reference	1 (°C)	Max allowed deviation between climatological min and max temperature of AWS stations in Uccle/Diepenbeek/Humain and LC-R in Uccle/Diepenbeek/Humain
dev	5 (°C)	Max allowed deviation between climatological min and max temperature of AWS station in Uccle and LC-X in Leuven
Temporal outliers (TO)		
TOathresmin	-3 (°C)	Min allowed difference between sequential 10 min observations
TOathresmax	2.5 (°C)	Max allowed difference between sequential 10 min observations
TObthresmin	0.05 (°C)	Min difference that should be noted in TObtimespan
TObtimespan	19 (-)	Number of consecutive 10 min observations in which temperature should change with TObthresmin
Spatial outliers (SO)		
range	2500 (m)	Range used to define neighbouring stations
SOthresmin	-3 (°C)	Min allowed Z-score
SOthresmax	3 (°C)	Max allowed Z-score
nstat	1 (-)	Minimum requirement of measurements in range

3.1.1 QC level 1.1 - Range outliers

During QC L1.1 a range test based on climatology is performed. Range outliers can occur when a station is malfunctioning or installed in a wrongful location. The latter has been largely eliminated by the installation protocol described in Section 2. Observations are flagged as 1 whenever they exceed the maxima or minima climate thresholds, plus/minus an allowed deviation (T_{max/min_AWS} ± dev). These thresholds are based on historical data from nearby official weather stations while the allowed deviations from the climate thresholds are based on local knowledge on environmental phenomena The thresholds are calculated as the maximum and minimum temperature from the official AWS station in Uccle within the 3 month period that currently undergoes the QC. Observations receive a flag equal to -1 when the no temperature observation is available.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

3.1.2 QC level 1.2 - Temporal outliers

In QC L1.2 temporal outliers are detected using both (a) a step test and (b) persistence test. Temporal outliers occur when an observation of a specific station is not in line with the surrounding observations of this station. The step test ensures that the change in magnitude between two consecutive observations lies within a certain interval; the test checks the rate of change and flags unrealistic jumps in consecutive values. Flags are set to 1 when observations increase more than 2.5°C ($TOa_{ThresMax}$) or decrease more than 3°C ($TOa_{ThresMin}$) in 10 minutes. Such steep increases or decreases in temperature are found when a station reconnects with its receiver after a period of hitches. Observations are assigned a flag equal to -1 when the difference between sequential observations cannot be calculated.

The persistence test, on the other hand, makes sure that observations change minimally with time. Here we detect stations with connection issues, transmitting the same observation repeatedly. Observations changing less than 0.05°C ($TOb_{ThresMin}$) within 3 hours ($TOb_{Timespan}$) are flagged as 1. Whenever the difference between sequential observations cannot be calculated as observation get a flag equal to -1.

3.1.3 QC level 1.3 - Spatial outliers

QC L1.3 detects spatial outliers in the dataset. Spatial outliers occur when the observation of a specific station is too different compared to the observations from neighbouring stations. First neighbouring stations are defined as stations located within a 2.5 km radius (range). Next the Z-score or standard score is calculated for each observation following Eq. (1)

$$Z = \frac{x - \mu}{\sigma}, \qquad (1)$$

where x is the observed value, μ the mean value and σ the standard deviation across all neighbours. This standard score can be explained as the number of standard deviations by which the observed value is above or below the mean value of what is being observed. Whenever the Z-score is lower than -3°C ($SO_{ThresMin}$) or higher than 3°C ($SO_{ThresMax}$) the observation is seen as a spatial outlier and receives a flag equal to 1. When there are no neighbours available within the predefined range, or the Z-score cannot be calculated, each observation is flagged with -1.

3.2 Quality control level 2 – Inter station bias correction

The second quality control level corrects the data for the fixed offset or inter station temperature bias between the weather stations. This step is necessary since the temperature sensor are only calibrated by the manufacturer, and small calibration differences are expected for this consumer-grade weather sensor. Moreover, the Leuven.cool stations originate from different production batches, with possible hardware changes in the electronics. Calibration tests between multiple LC-X stations in the same controlled environment were both technically and logistically not feasible. Simultaneous measurements are only

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

available for two LC-R stations (LC-R01 and LC-R02 at the AWS of Uccle) for a period of four months, showing that sensor differences indeed exist and are non-negligible.

In order to quantify this inter station temperature bias, a rather pragmatic approach was followed to mimic a controlled environment: we selected episodes for which a similar temperature across the study area is expected. Such episodes occur under breezy cloudy conditions with no rainfall (Arnfield, 2003; Kidder and Essenwanger, 1995). In practice, the database is searched for suitable episodes every 6 months, currently ranging from 2019S2 to 2021S2. All 10 minute observations are resampled to 2 hour observations, hereby calculating the mean temperature, windspeed, radiation and rainfall across all weather stations. Next, suitable episodes are found by selecting episodes where the average rainfall intensity equals 0 mm/h and the average radiation lies below 100 W/m2. The selected episodes are ordered on average windspeed and limited to the top 10 results.

For these episodes, one can assume the temperature to be very uniform over the study area, and solely controlled by altitude (Lu Aigang et al., 2009). In practise, only episodes with a high correlation between temperature and altitude (> 0.7) are retained. By regressing temperature versus altitude for every episode and calculating the residuals e.g. the difference between the observed and predicted temperature, a fixed offset for each station and every episode is obtained. Finally, the median offset across all episodes is considered as the true offset for each station. These offsets are added to the QC level 1 temperature data in order to obtain the corrected QC level 2 temperature data.

3.3 Quality control level 3 – Intra station bias correction

During the third quality control level the QC level 2 temperature data is further corrected for the variable intra station temperature biases. This bias is present in the data since the measurements are made with non-standard equipment as compared to the AWS measurements (e.g., passive instead of active ventilation, dimension of the Stevenson screen). These biases change during day and night time, and according to their local environment (e.g., radiation and windspeed patterns). (Bell et al., 2015).

By identifying the climatic variables mostly correlated with the temperature bias between the low-cost reference stations (LC-R) and the official RMIB stations in Uccle, Diepenbeek and Humain (AWS), a predictor for temperature bias is created. To produce a robust model, data from all low-cost reference stations (LC-R01, LC-R02, LC-R04 and LC-R05) were used simultaneously to create a predictor for the intra station temperature bias.

For the construction of a predictor model, the dataset was split in training (0.60) and validation (0.40) data. The training data was used to train simple regression models, multiple regression models, random forest (RF) models and boosted regression trees (BRT). Since previous research (Bell et al., 2015; Cornes et al., 2020; Jenkins, 2014) has shown that both radiation and wind speed highly influence the temperature bias, the simple and multiple regression models are mostly based on these

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

variables. A previous study by Bell et al. (2015) also suggested that past radiation measurements, using an exponential weighting, are an even better prediction of the temperature bias, resulting in an advanced correction model (Bell et al., 2015). The potential predictor models are validated using the validation data, ensuring an fair evaluation of the model. The coefficient of determination (R^2) and root mean square error (RMSE) are calculated to identify the most optimal prediction model for the temperature bias. After validation, the prediction model can be applied to the weather station network in Leuven (LC-X), hereby providing a temperature bias for each observation of every station in function of its local climatic conditions. The predicted temperature bias is subtracted from the QC level 2 temperature data to obtain the QC level 3 corrected temperature dataset.

3.3.1 The intra station temperature bias

The overall temperature bias (i.e., all LC-R stations together) between the LC-R and the AWS data has a mean value of 0.10°C and a standard deviation of 0.55°C (Figure 3). By splitting up the temperature bias for day and night, a positive mean temperature bias during daytime (0.32°C) and a negative mean temperature bias during night time (-0.10°C) is obtained. Figure 3 further suggests a higher standard deviation during daytime (0.61 °C) compared to night time (0.37 °C), both with a remarkably skewed (and opposite) distribution.

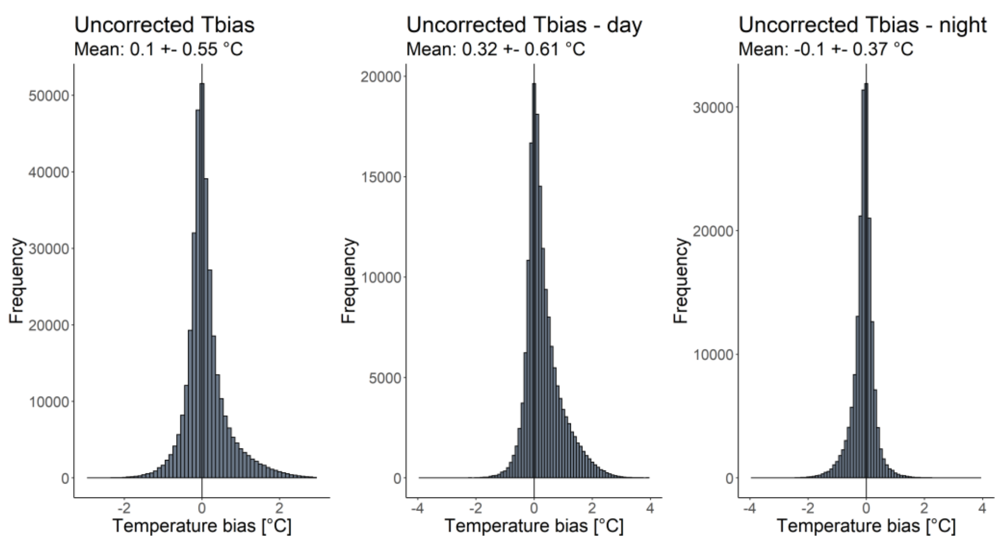


Figure 3: Histograms of the temperature bias between the low-cost reference stations (LC-R) and the official RMIB stations (AWS) for day and night (a), daytime defined by a radiation > 0 (b) and night time defined by a radiation = 0 (c). Mean biases and their standard deviations are given above the graphs.

To get a better understanding of the monthly and daily patterns of this temperature bias, Figure 4a shows the mean temperature bias in function of month and hour of the day. The stations show clear diurnal and seasonal patterns, confirming the positive

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

temperature bias during daytime and negative temperature bias during night time previously observed in Figure 3. In general, we see a positive bias that is high around midday and is more pronounced during the summer months, lasting for several hours during the day. In summer months, a shallow local minimum is seen around noon, which we expect to be an effect of the specific station design (shadow of anemometer). The night-time temperature bias is low for all months. A temperature bias of 0°C is reached for every month at a certain time of the day, the specific time at which this minimal temperature bias occurs, depends however on the season. Figure 4b shows the mean temperature bias in function of windspeed and radiation. As expected, a positive temperature bias is noticed for high solar radiation and low windspeed conditions. The rather strange high values at low radiation and high wind speed can be explained as outliers (Figure 4c). Figure 4c shows the sample size of each cell. After removing cells with sample size lower than 10, the final graph is obtained (Figure 4d).

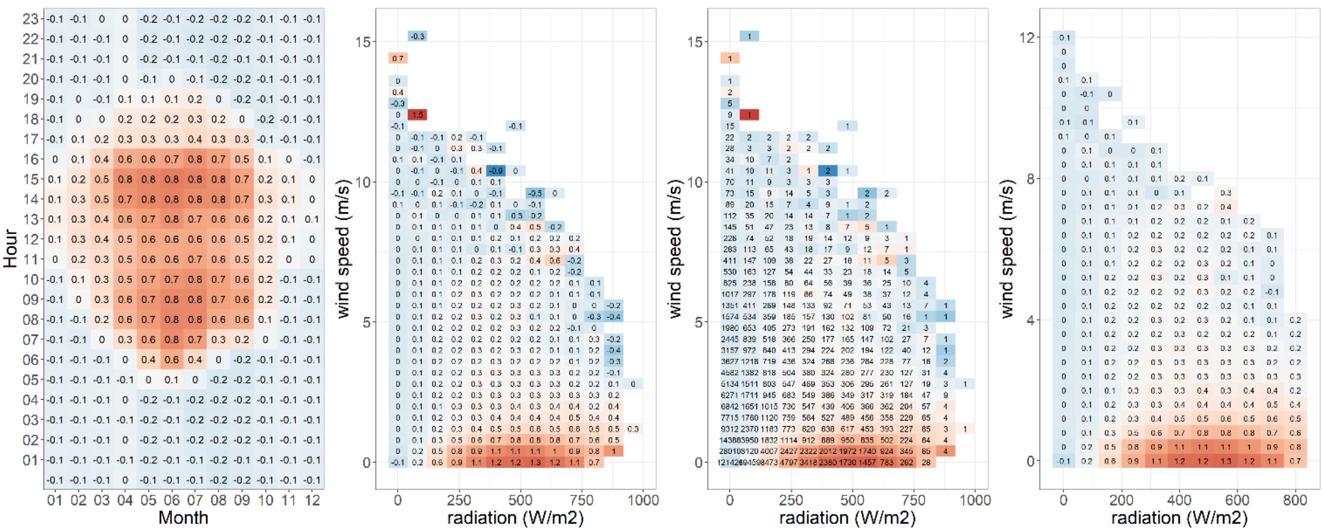


Figure 4: Temperature bias (°C) as a function of hour of the day and month of the year for all LC-R (a), temperature bias (°C) as a function of radiation and wind speed for all LC-R (b), temperature bias (°C) as a function of radiation and wind speed for all LC-R, the values written in each cell signify the sample size (c), temperature bias (°C) as a function of radiation and wind speed for all LC-R, cells with sample size lower than 10 are not shown in the graph (d).

3.3.2 Building a predictor for the intra station temperature bias

A correlation matrix between the temperature bias and other meteorological variables, measured by the low-cost weather station, is calculated (Table 7). The values indicate how the temperature bias will change for different meteorological conditions.

Table 7: Correlation matrix of temperature bias with other meteorological variables measured by the low-cost station.

Temperature (°C)	Dew point temperature (°C)	Humidity (%)	Radiation (W/m^2)	Radiation60 (W/m^2)	Wind speed (m/s)
0.41	0.18	-0.48	0.49	0.56	-0.01

The most correlated variable is radiation (0.49) directly followed by humidity (-0.48), temperature (0.41), dew point temperature (0.18) and wind speed (-0.01). As expected, taking the past radiation measurements into account further improves the correlation, reaching a maximum value when considering the last 60 minutes (0.56). An exponential weighting, giving higher importance to the radiation measurements closer to the temperature measurement, was used. This variable is further denoted as Radiation60 (Rad60).

The variables listed in Table 7 were used to build a predictor for the temperature bias. For this purpose, multiple models were calibrated in which the temperature bias is described as a function of only one or multiple meteorological variables. Below (Table 8 and Figure 5) only the models with the best performance are shown. Figure 5 shows the uncorrected temperature bias (a) as well as the corrected temperature biases after validation with six different models; a simple linear regression with the past radiation (b), a multiple linear regression with the past radiation and windspeed (c), with the past radiation and humidity (d), with the past radiation, windspeed and humidity (e), a random forest model (f) and a boosted regression trees model both including temperature, dew point temperature, humidity, radiation, radiation60, windspeed, altitude, month and hour.

Table 8: The coefficient of determination (R2) and root mean square error (RMSE) of the different models.

Model	R^2	RMSE
Radiation60	0.321	0.450
Radiation60 & windspeed	0.327	0.448
Radiation60 & humidity	0.336	0.445
Radiation60 & windspeed & humidity	0.342	0.443
Random Forest	0.741	0.279
Boosted regression trees	0.658	0.319

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.
Open Access Earth System Science Data Discussions

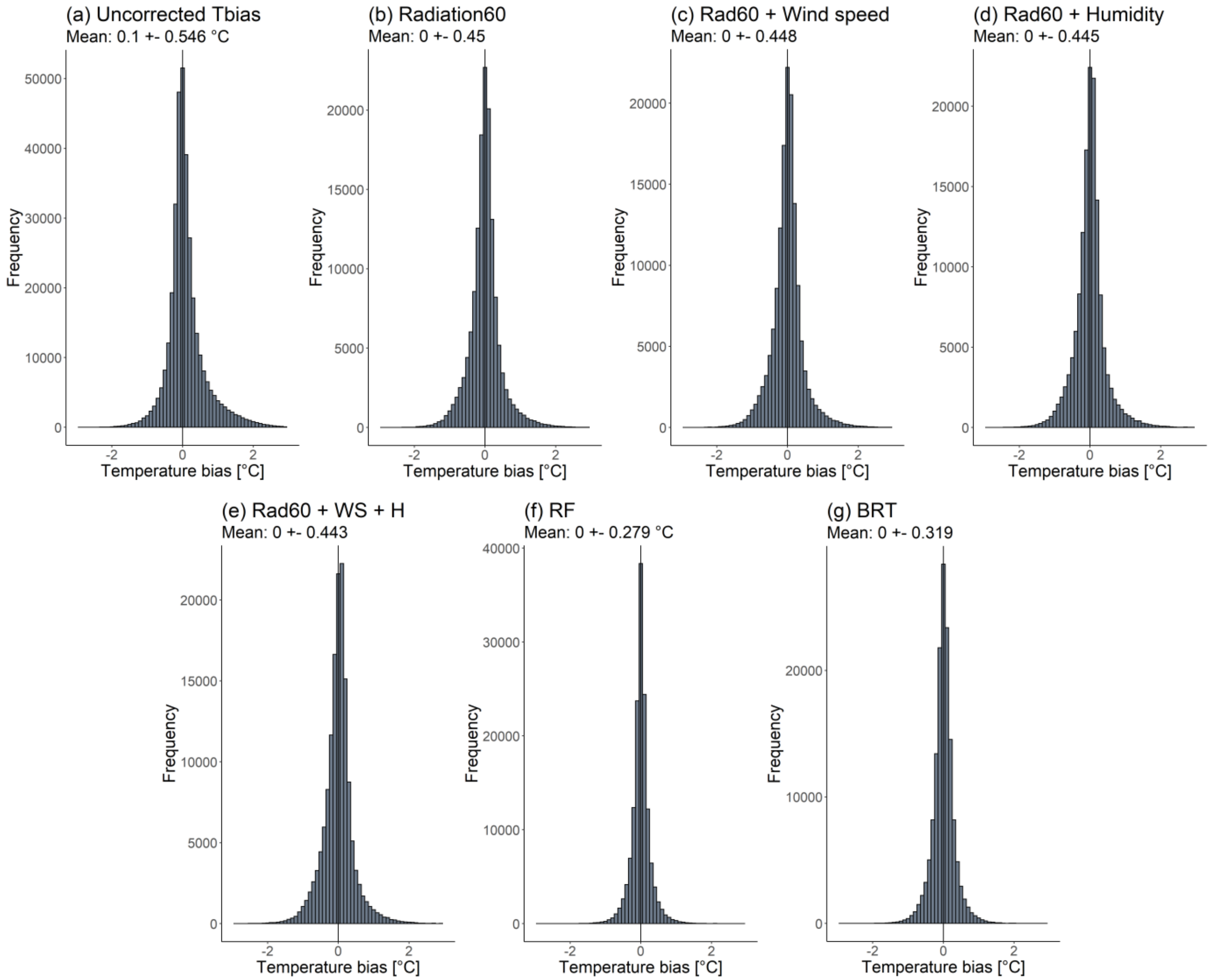


Figure 5: The uncorrected temperature bias (a) and the corrected temperature bias after validation with a simple linear regression with the past radiations (b), a multiple linear regression with the past radiations and windspeed (c), a multiple linear regression with the past radiations and humidity (d), a multiple linear regression with the past radiations, humidity and windspeed (e), a random forest model (f) and boosted regression trees model including temperature, dew point temperature, humidity, radiation, radiation60, windspeed, altitude, month and hour (g). Mean biases and their standard deviations are given above the graphs.

A simple linear regression based on the past radiation is already sufficient to suppress the mean temperature bias. Adding additional variables to the model, such as wind speed, humidity or both is statistically significant but only further decreases the RMSE by 0.003°C to 0.008°C. The RF and BRT models result in an RMSE of 0.279 and 0.319 and R^2 of 0.741 and 0.658 respectively, indicating a better precision and more robust models.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

The random forest prediction of temperature bias showed the best results. By splitting up the results for day and night (Figure 6), a smaller standard deviation of the bias during night time (0.25) compared to daytime (0.31) is obtained. The statistical details of the random forest model are summarized in Table 9.

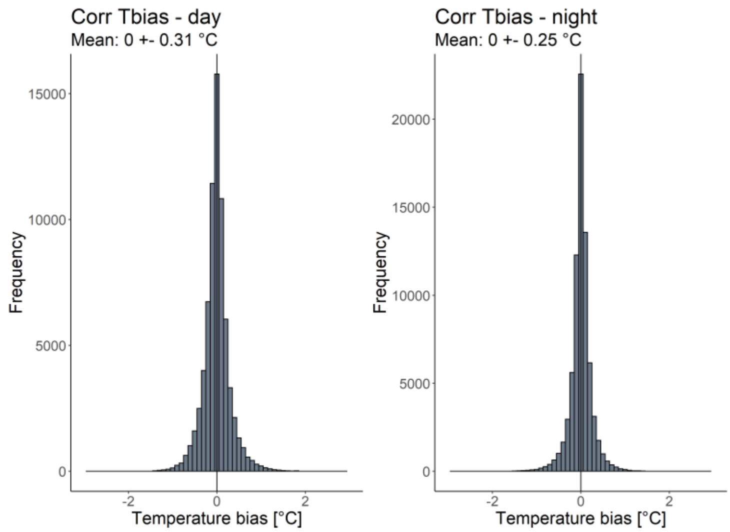


Figure 6: The corrected Tbias after validation with the random forest model for daytime (a) and for night time (b). Mean biases and their standard deviations are given above the graphs.

Table 9: Statistical details of the random forest temperature bias prediction model.

Formula	Tbias ~ QC L2 Temperature + Humidity + Dew point temperature + Radiation + Radiation60 + Windspeed + Altitude + Month + Hour
Number of trees	500
Number of variables tried at each split	3
Mean of squared residuals	0.091
% variance explained	69.5

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

4 Evaluation of the quality control and correction method

To evaluate the quality of the developed QC method, it is first applied to the four low cost WH2600 stations (LC-R) (Table 3) installed next to the official measuring equipment in Uccle, Diepenbeek and Humain (AWS). Comparing this LC-R dataset with AWS dataset, allows us to investigate the improvement or deterioration of the data quality after each QC level.

4.1 Quality control level 1 – Outlier detection

For QC L1.1 the range outliers are detected by comparing the temperature of each LC-R station with the climatic thresholds set by its nearby official AWS station (T_{max/min_AWS} ± dev_reference). As can be seen in Table 6, the allowed deviation from the climatic thresholds is smaller for the LC-R stations compared to the LC-X stations in the study area. This is due to the fact that the LC-R stations are installed next to the official AWS stations, no environmental factors should thus be taken into account. The temporal outliers are detected in QC L1.2 by comparing the rate of change between consecutive observations with the thresholds defined in Table 6. In QC L1.3 spatial outliers are detected using the Z-score. We should however stress that this analysis here is not ideal since every reference station only has 1 neighbour, the official AWS station. Only LC-R01 and LC-R02 located in Uccle have two neighbours during a period of 4 months, when both stations were active simultaneously.

The results show no spatial outliers for the LC-R stations. Some observations are however highlighted as range or temporal outliers. Table 10 summarises the number and percentage of observations flagged as 1. These observations are set to NA resulting in the temperature dataset with quality level 1. The temperature profiles of the LC-R stations versus the official AWS temperature (Figure 7; full coloured versus dashed grey line) highlight observations defined as range or temporal outliers as circles or squares respectively. Scatterplots in which the temperature of the LC-R stations is compared to the temperature of the AWS stations (Figure 8) use the same layout. The temperature difference between the LC-R stations and AWS stations ($\Delta T = T_{LC\text{-}R} - T_{AWS}$) is calculated as an effective quality measure.

Table 10: Number and percentage of observations flagged as outliers during QC level 1.

QC level	# flagged observations	% flagged observations
QC L1.1 Range test	21	0.006
QC L1.2 Temporal outliers	180	0.048
QC L1.3 Spatial outliers	0	0.000
Total	201	0.054

The results show only a few range outlier and even no spatial outliers for the LC-R reference stations. The procedure does however highlight 180 observations as temporal outliers (Table 10). These observations were highlighted during the persistence test, 180 observations change less than 0.1° within 2 hours. With only 0.054% of the observations flagged as

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

outliers we can conclude that the LC-R reference dataset does not contain a lot of outliers. Because of their importance in this QC method, especially in QC L3, these reference stations are indeed closely monitored hereby preventing and minimising the occurrence of outliers. Since only 0.054% of the data was set to NA, no difference in the ΔT statistics occurred. The histograms in Figure 9 thus represent the data with both QC level 0 and QC level 1. The mean temperature difference and standard deviation for all reference stations equals 0.15 ± 0.56°C.

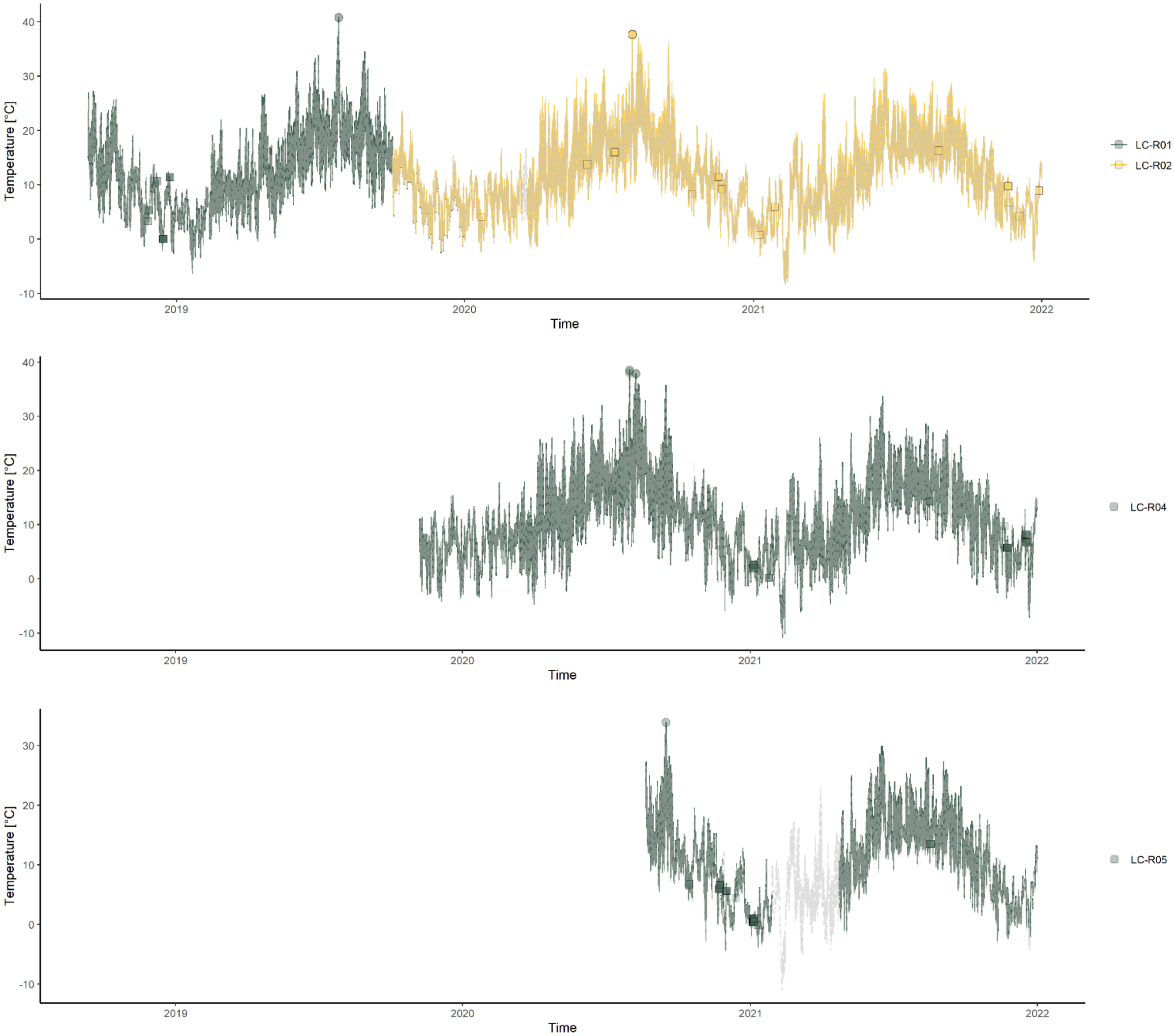


Figure 7: Temperature profile of LC-R stations in Uccle (LC-R01, LC-R02) (a), Diepenbeek (LC-R04) (b) and Humain (LC-R05) (c). The grey dashed line represents the official AWS temperature of a specific location. Observations defined as range outliers are symbolized by a circle, temporal outliers as a square.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

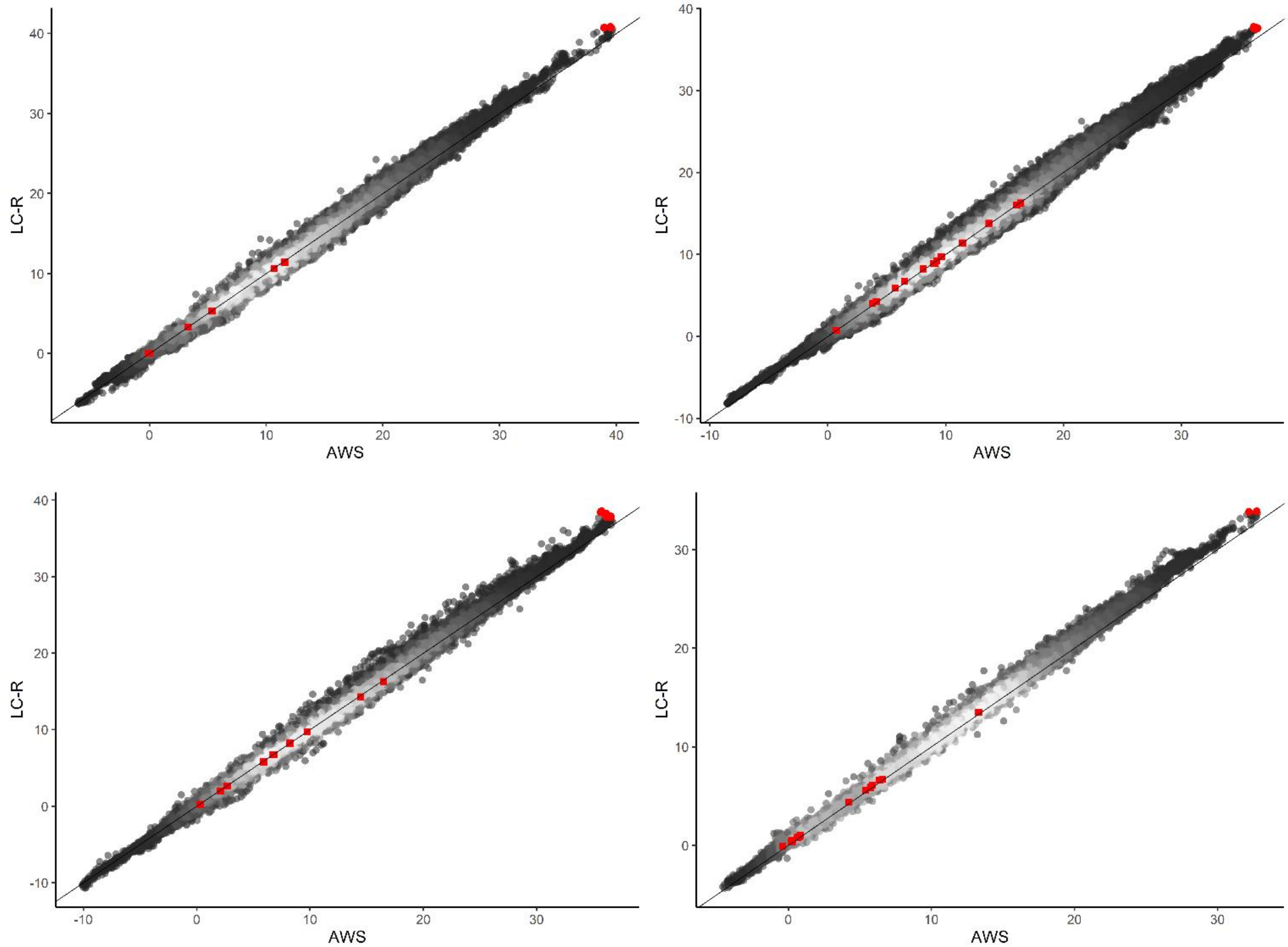


Figure 8: Scatterplots of LC-R versus AWS temperature for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 0. Observations defined as range outliers are symbolized by a red circle, temporal outliers by a red square.

Reference station LC-R01 only has a small positive mean ΔT, while the mean ΔT of LC-R02 and especially LC-R05 is remarkably higher. The mean ΔT of LC-R04 equals zero. The standard deviations of LC-R04 and L-R05 are noticeably smaller than those of LC-R01 and LC-R02 (Figure 9). As expected, the temperature difference between the LC-R and AWS stations is not constant and is correlated with other variables. A higher difference is obtained during a summer day with low cloud and low windspeed conditions (Figure 10 and Figure 11).

Open Access Earth System Science Data Discussions

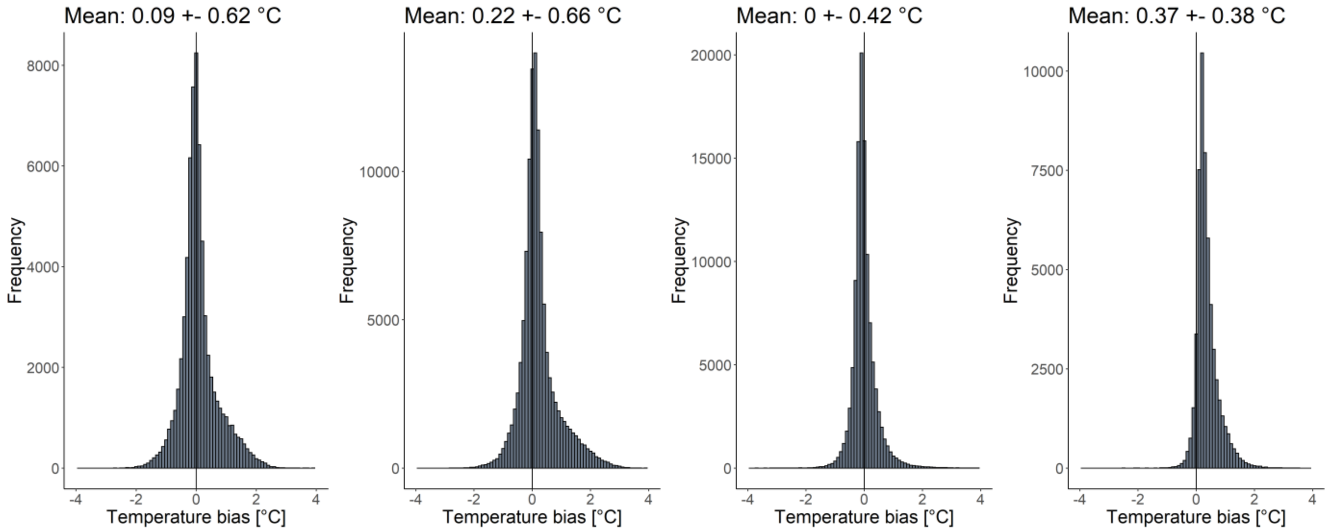


Figure 9: Histograms of temperature difference (ΔT = $T_{LC\text{-}R}$ - T_{AWS}) for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 0 and QC level 1. Mean differences and their standard deviations are given above the graphs.

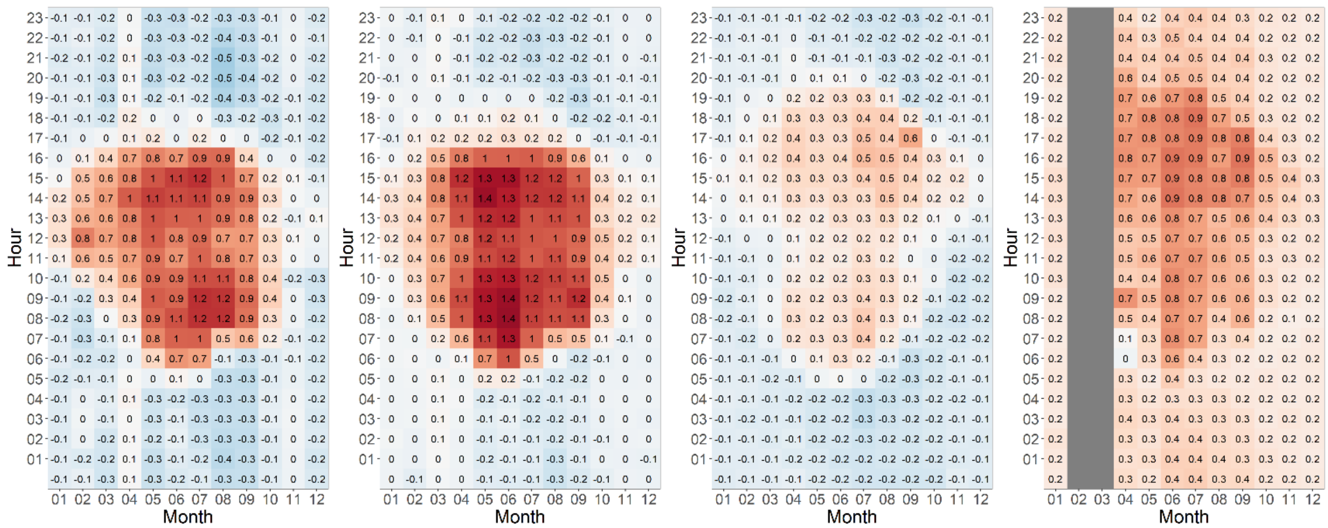


Figure 10: Temperature difference (ΔT = $T_{LC\text{-}R}$ - T_{AWS}) as a function of hour of the day and month of the year for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 1.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

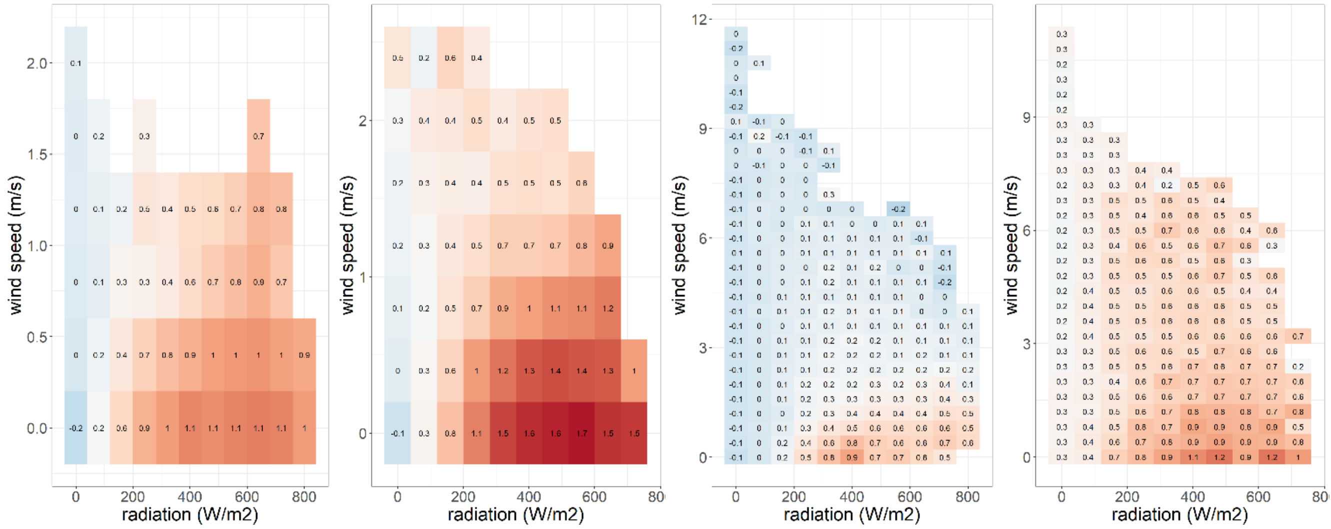


Figure 11: Temperature difference (ΔT = $T_{LC\text{-}R}$ - T_{AWS}) as a function of radiation and wind speed for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 1. Cells with sample size lower than 10 are not shown in the graph.

4.2 Quality control level 2 – Inter station bias correction

During QC level 2 temperatures are corrected for the fixed offset between stations or inter station bias, due to intrinsic sensor differences at the level of the electronics. The proposed methodology of searching for episodes with a very uniform temperature field over the study area (see Section 3), cannot be applied here, due to the large distance between the three locations with LC-R stations.

Here we selected episodes for which we expect a similar temperature between the LC-R and AWS stations. This occurs again under breezy cloudy conditions with no rainfall (Figure 11). For each reference station, all 10 minute observations are resampled to 2 hour observations, hereby calculating the mean temperature, windspeed, radiation and rainfall. Next, suitable episodes are found by selecting episodes where the average rainfall intensity equals 0 mm/h and the average radiation lies below 100 W/m2. The selected episodes are ordered on average windspeed and limited to the top 10 results. The mean LC-R and AWS temperature is calculated for each episode, next an offset between both is calculated. Finally the median offset across all episodes is considered as the true offset for each station. These offsets are subtracted from the QC L1 temperature data in order to obtain a corrected temperature; the QC level 2 dataset.

Reference stations LC-R01 and LC-R04 have a small negative offset, equal to -0.029°C and -0.072°C respectively. Stations LC-R02 and LC-R05 have positive and notably larger offsets equal to 0.113°C and 0.243°C (Figure 12).

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

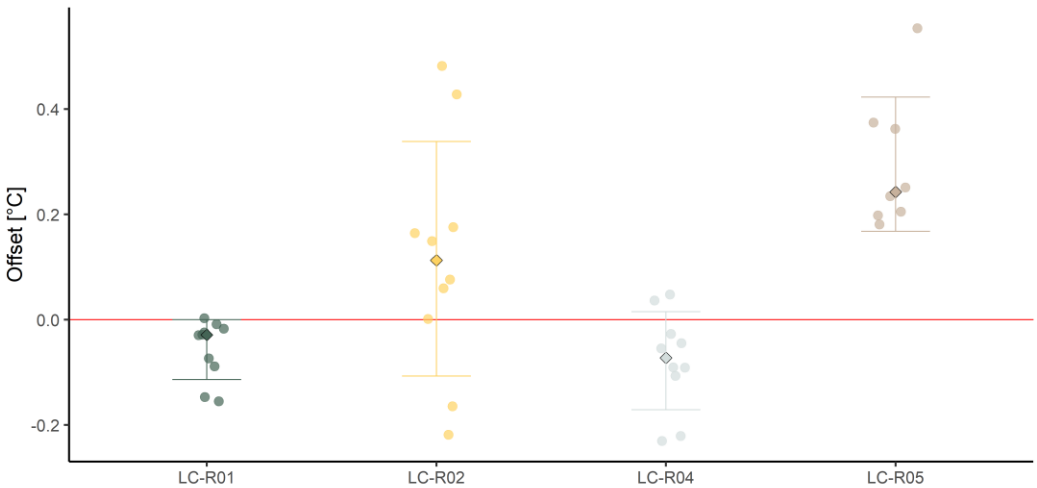


Figure 12: Offsets during the selected episodes (dots), the median offset (diamond) and its error bar (mean ± standard deviation) for each reference station (LC-R). For reference the zero-line is plotted in red.

To check the quality improvement of QC level 2, the temperature difference ($\Delta T = T_{LC\text{-}R} - T_{AWS}$) between the LC-R and AWS stations is again calculated for every station (Figure 13). Reference stations LC-R01 and LC-R04 show a small increase in their mean ΔT, the mean ΔT of LC-R02 and LC-R05 however decreases. As a result, the ΔT of all stations becomes more equal. Since QC level 2 only added a fixed temperature offset, the standard deviation of all ΔT remain the same. The inter station bias correction further highlights the seasonal and daily pattern of the ΔT, especially for reference station LC-R05 (Figure 14).

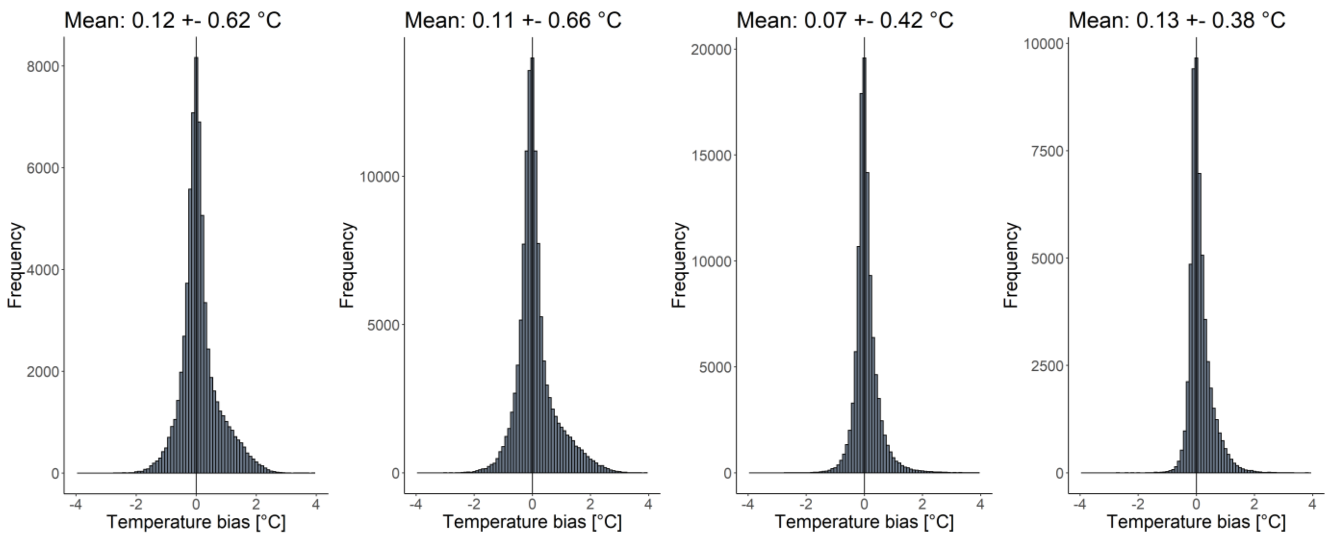


Figure 13: Histograms of temperature difference ($\Delta T = T_{LC\text{-}R} - T_{AWS}$) for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 2. Mean biases and their standard deviations are given above the graphs.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

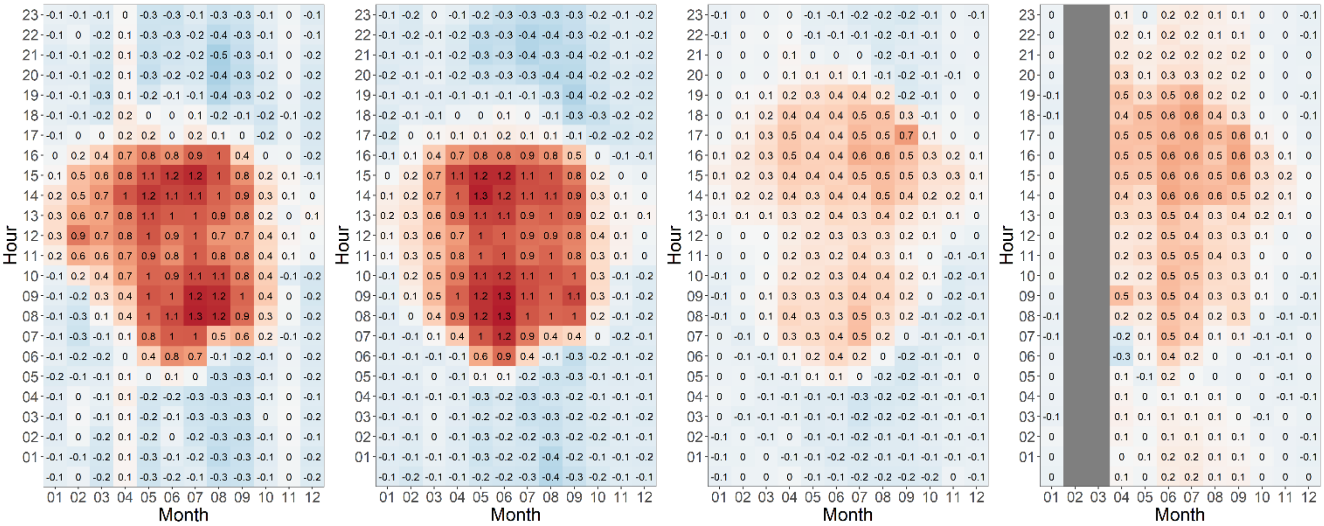


Figure 14: Temperature difference (ΔT = $T_{LC\text{-}R}$ - T_{AWS}) as a function of hour of the day and month of the year for each reference station LC-R01 (a), LC-R02 (b), -LC-R04 (c), LC-R05 (d) at QC level 2.

4.3 Quality control level 3 – Intra station bias correction

From Section 3, we recall that the random forest prediction of temperature bias showed the best results After applying this prediction model on the complete reference dataset (LC-R), a level 3 corrected temperature is obtained for each LC-R station.

To evaluate the quality improvement of QC level 3, the temperature difference (ΔT = $T_{LC\text{-}R}$ - T_{AWS}) between the LC-R and AWS stations is again calculated for every station (Figure 15). Histograms of the temperature show a mean ΔT of almost 0°C for each reference station, the standard deviation clearly decreased compared to QC level 2 (Figure 13). When the ΔT is plotted in function of each month and hour of the day, one can notice that the diurnal and seasonal pattern is completely corrected for (Figure 16). Also effects of wind speed and radiation are effectively eliminated (Figure 17). The mean temperature difference and standard deviation for all LC-R stations equals to 0.00 ± 0.22°C.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

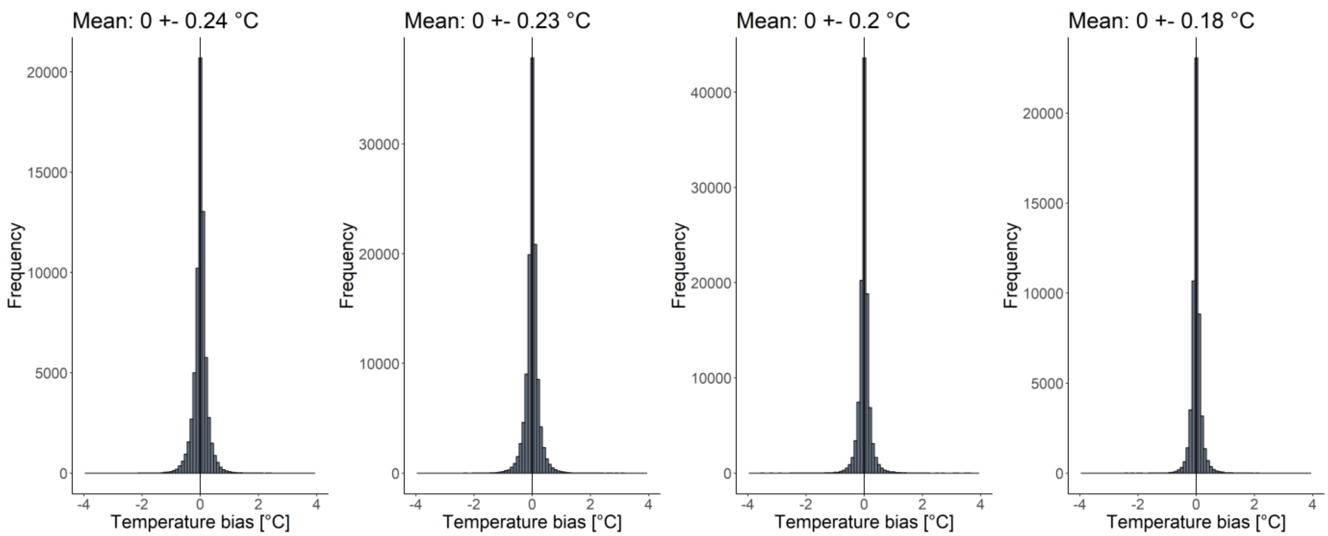


Figure 15: Histograms of temperature difference ($\Delta T = T_{LC-R} - T_{AWS}$) for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 3. Mean biases and their standard deviations are given above the graphs.

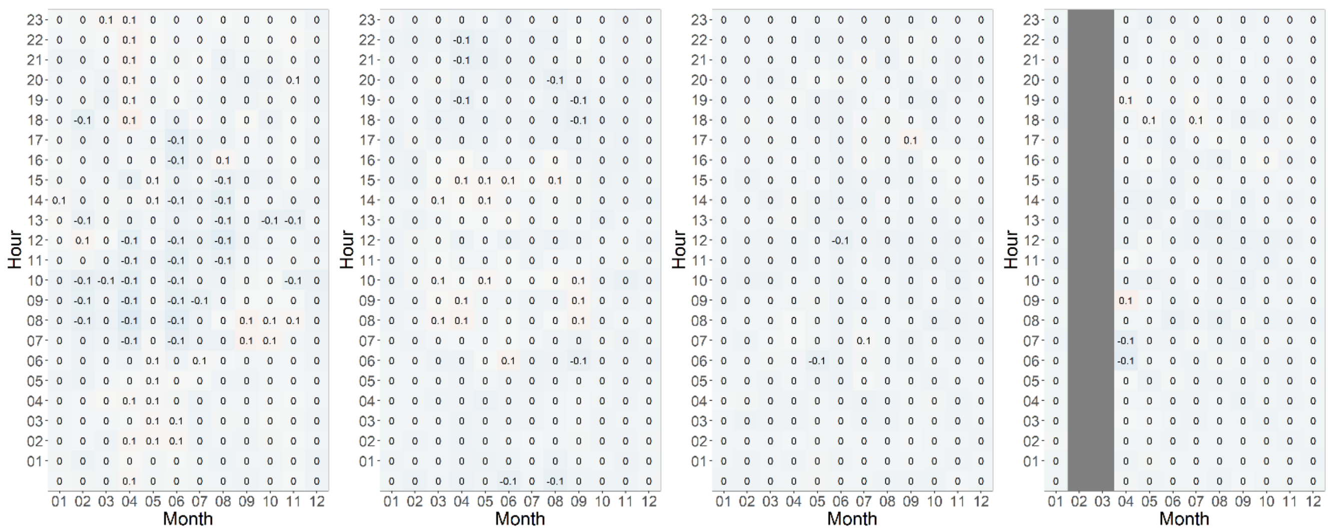


Figure 16: Temperature difference ($\Delta T = T_{LC-R} - T_{AWS}$) as a function of hour of the day and month of the year for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 3.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

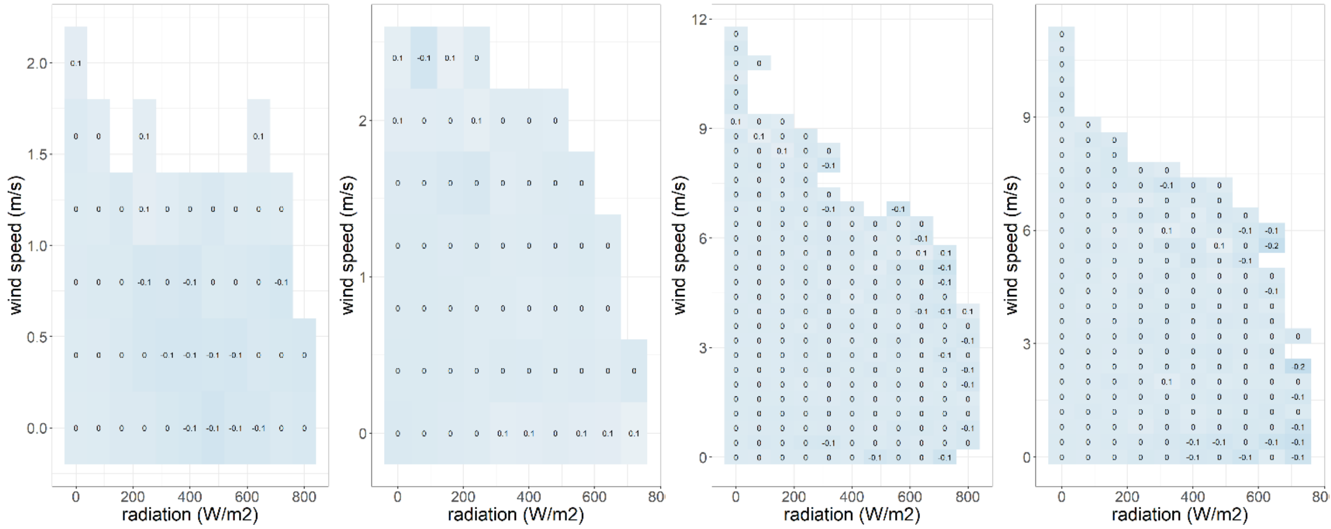


Figure 17: Temperature difference ($\Delta T = T_{LC-R} - T_{AWS}$) as a function of radiation and wind speed for each reference station LC-R01 (a), LC-R02 (b), LC-R04 (c), LC-R05 (d) at QC level 3. Cells with sample size lower than 10 are not shown in the graph.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

5 Application of the QC method to the stations in the study area

In this section the newly developed QC method is applied to the low-cost stations of the Leuven.cool network (LC-X). The Leuven.cool dataset currently ranges from July 2019 (2019Q3) until December 2021 (2021Q4). The QC method is performed 4 times a year, each time for a period of three months.

5.1 Quality control level 1 – Outlier detection

During QC level 1 range, temporal and spatial outliers are removed using climatological thresholds from Uccle, neighbouring observations and neighbouring stations respectively. Table 11 summarizes the number of observations flagged as outliers in each step. For each year, only between 0.5% and 1% of the data is defined as outliers and thus eliminated, indicating that the raw data quality is rather good compared to other citizens science networks.

Table 11: Number and percentage of observations flagged as outliers during QC level 1 for 2019, 2020 and 2021.

QC level	# 2019	% 2019	# 2020	% 2020	# 2021	% 2021
QC L1.1 Range test	0	0.000	9	0.000	26	0.000
QC L1.2 Temporal outliers: step	22	0.001	276	0.006	169	0.003
QC L1.2 Temporal outliers: persistence	796	0.039	1216	0.025	4796	0.092
QC L1.3 Spatial outliers	11769	0.581	31846	0.658	26186	0.503
Total	12587	0.621	33317	0.689	31177	0.599

The simple spatial outliers test performed by Chapman et al. (2017) yielded comparable results: 1.5% of the data was omitted in this study. Other studies reported a much higher fraction of eliminated data. Meier et al. (2017) only kept 47% of the raw data after conducting their 4 step QC analysis. Napoly et al (2018) and Feichtinger et al. (2020) kept 58% and 55% of the data respectively. CrowdQC+ as a further development of CrowdQC, results in an even lower data availability, only 30% of the raw data remains after the QC (Fenner et al., 2021). These high numbers of omitted data can however be explained by (1) a great number of CWS installed indoors (not applicable in our setup), thereby lacking the typical diurnal temperature patterns, and (2) radiative errors due to solar radiation exposure of poorly designed devices, resulting in very high temperature observations (Fenner et al., 2021; Napoly et al., 2018). In the QC method presented in this paper, calibration and radiative errors are, however, rather than omitted, corrected for during QC level 2 and 3.

5.2 Quality control level 2 – Inter station bias correction

In QC level 2 a fixed offset for each weather station is obtained. These offsets, induced by the intrinsic differences on the level of the sensors' electronics, are subtracted from the station's temperature, thereby accounting for calibration errors. The

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.
Open Access Earth System Science Data Discussions

obtained offsets, median offset and error bar for each station are plotted in Figure 18. The mean offset of all stations equals 0.010°C. Station LC-102 has the highest offset equal to 0.349°C, station LC-074 has the lowest offset equal to -0.220°C. It can be noticed that the error bars of most stations are rather small, which reinforces our confidence of a valid determination of the fixed calibration offset.

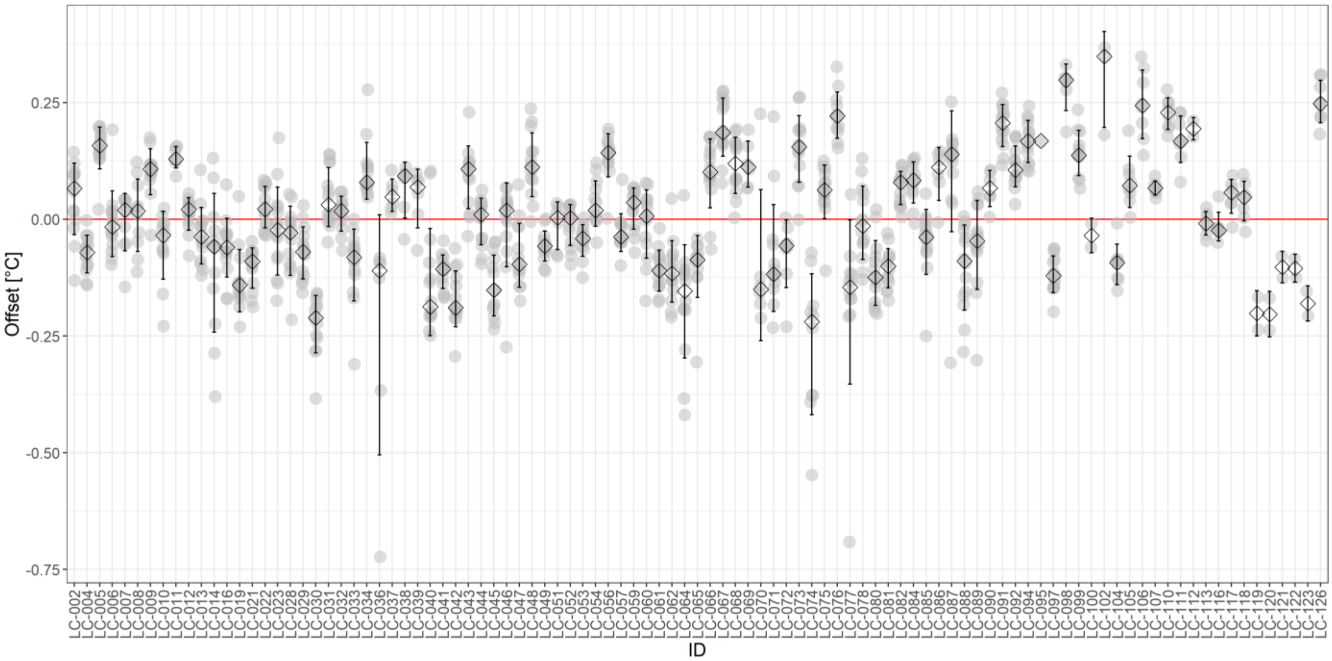


Figure 18: Offsets during the selected episodes (grey dots), the median offset (black diamond) and its error bar (mean ± standard deviation) for each station active during at least one episode. For reference the zero-line is plotted in red.

For stations that were not active during one of the selected timeframes, we were not able to determine their calibration offset (stations LC-003, LC-096, LC-108, LC-109, LC-114, LC-124, LC-125, LC-127 and LC-128). As a consequence, no corrected temperature could be calculated, meaning that those stations are not considered during the following QC level (QC level 3). The search for episodes should be extended with upcoming periods of 6 months in order to resolve this problem.

5.3 Quality control level 3 – Intra station bias correction

In QC level 3 the random forest model is applied to each station in order to obtain a site-specific prediction for its temperature bias. As expected, this prediction shows the same pattern as seen for the LC-R stations: generally, we see a positive bias that peaks around midday and is more pronounced during both summer months and low cloud and low windspeed conditions (Figure 19). Note that the actual bias calculation in QC level 3 is performed for every timestamp and every LC-X station separately, using the other weather variables measured by the station, as input for the RF model. After subtracting this temperature bias from the observed temperature, corrected temperature for each station is obtained.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

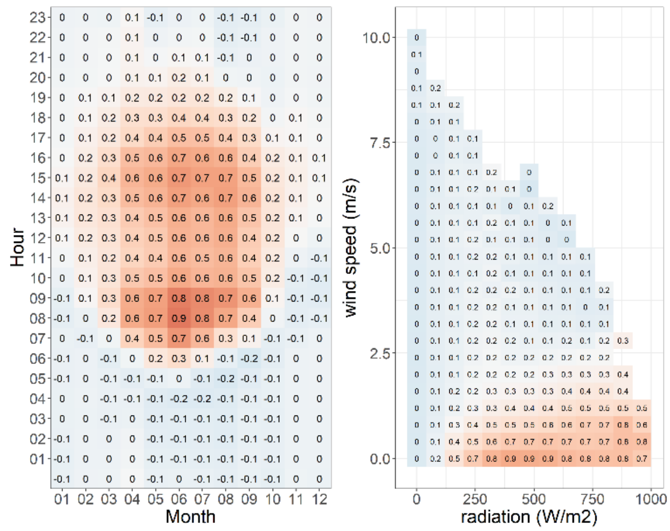


Figure 19: Prediction of the temperature bias (°C) as a function of hour of the day and month of the year for all LC-X stations (a) and prediction of the temperature bias (°C) as a function of radiation and wind speed for all LC-X stations, cells with sample size lower than 10 are not shown in the graph (b).

These results are in line with the findings of Jenkins et al. (2014) and Bell et al. (2015):both found a significant positive instrument temperature bias during daytime with strong relation to the incoming solar radiation for multiple types of crowdsourced weather stations. To our knowledge, previous to our study, only Cornes et al. (2020) has used the findings of Bell et al. (2015) to actually correct crowdsourced air temperature data. Radiation from satellite imagery and background temperature data from official stations was used to parameterize the short-wave radiation bias, as a consequence no correction was performed for night time. The data correction has reduced the error from ± 0.2 - 0.8°C to ± 0.2 - 0.4°C. These results are comparable with our results, although here slightly smaller errors of only 0.18 - 0.24°C are obtained (Figure 15). Cornes et al. (2020) do suggest to incorporate wind speed as an additional co-variate in order to incorporate the effect of passive ventilation. The random forest model described in this study does include additional co-variates, including wind speed, but most importantly only needs data from the weather station itself. No satellite imagery or official stations are needed once the random forest model is built. Cornes et al. (2020) further highlights the need for station specific quality controls in order to remove the confounding effect of different instrument types. The use of a unique station type in this study assures that no such effects are present in our dataset.

5.5 Overall impact of the QC method on the dataset

To assess the impact of the different QC stages on the global dataset, several violin boxplots were created both on monthly and yearly base. Figure 20 illustrates the monthly violin plots for 2019, 2020 and 2021 at each quality control level. Table 12 summarises the mean monthly temperature and its standard deviation for each quality control level.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

The violin plots, and accompanying mean temperature and standard deviation, do not change much over the different QC levels. Since QC level 1 removes outliers from the dataset we would expect a lower standard deviation for QC level 1 compared to QC level 0. Figure 20 does indicate the removal of some outliers, but Table 12 does not confirm the expected decrease in standard deviation. This can be explained by the low percentage of observations defined as outliers, per year only 0.5 to 1 % of the data was defined as outlier. Due to the strict installation protocol, most errors were already eliminated upfront. If errors do occur as a result of station malfunctioning, they are quickly resolved since the dataflow and station siting are continuously controlled.

During QC level 2 each station is corrected for its inter station temperature bias. Since both positive and negative biases, ranging from 0.349°C to -0.220°C, are possible no clear change in the mean temperature is expected between QC level 1 and QC level 2. Because this QC level corrects each station with a fixed offset the standard deviation should stay the same. Both of these assumptions are confirmed by Figure 20 and Table 12.

During the third QC level we do see a clear change in mean temperature and standard deviation. For the summer months a reduction in both the mean temperature and standard deviation up to respectively -0.40°C and -0.36°C is noted. The change in standard deviation shows a monthly pattern with a higher reduction during the summer months and almost no change during winter. The change in mean temperature is not as consistent and seems dependent on the observed temperatures. A higher reduction in mean temperature is noted for the hot summers of 2019 and 2020 compared to the rather cold summer of 2021. These results can easily be explained by the daily and seasonally patterns of the predicted temperature bias (Figure 19).

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.
Open Access Earth System Science Data Discussions

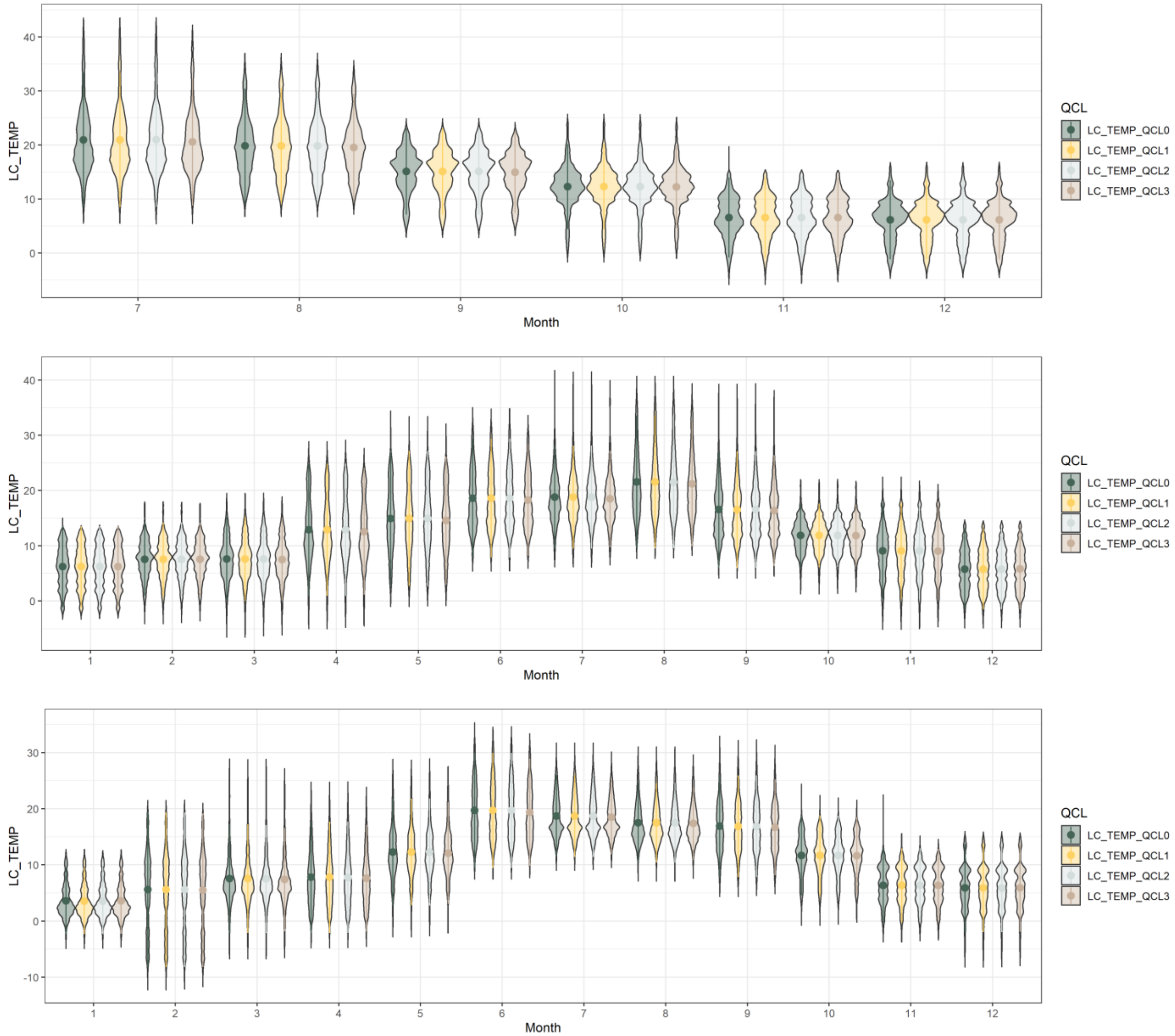


Figure 20: Monthly violin plots of temperature data (°C) at quality control level 0 (raw data), level 1 (outliers removed), level 2 (inter station bias correction) and level 3 (intra station bias correction) for 2019 Q3-4 (a), 2020 Q1-2-3-4 (b) and 2021 Q1-2-3-4 (c).

Table 12: The mean monthly temperature (°C) and its standard deviation at quality control level 0 (raw data), level 1 (outliers removed), level 2 (inter station bias correction) and level 3 (intra station bias correction).

	QC level 0		QC level 1		QC level 2		QC level 3		ΔL0-L3	
	Mean T	Std T	Mean T	Std T	Mean T	Std T	Mean T	Std T	Mean T	Std T
2019										

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

7	20.95	6.21	20.93	6.23	20.96	6.23	20.54	5.89	-0.40	-0.32
8	19.82	5.32	19.81	5.33	19.84	5.34	19.53	4.96	-0.29	-0.36
9	15.08	3.99	15.08	3.99	15.08	3.99	14.91	3.76	-0.17	-0.23
10	12.26	3.93	12.25	3.93	12.26	3.93	12.22	3.83	-0.04	-0.10
11	6.54	3.70	6.53	3.70	6.54	3.70	6.54	3.68	0.00	-0.03
12	6.12	3.65	6.12	3.65	6.12	3.65	6.15	3.65	0.03	0.00
2020										
1	6.21	3.68	6.20	3.68	6.21	3.68	6.21	3.66	0.00	-0.02
2	7.57	3.33	7.57	3.33	7.57	3.33	7.56	3.29	-0.02	-0.05
3	7.62	3.76	7.62	3.77	7.62	3.76	7.49	3.69	-0.13	-0.07
4	12.89	5.97	12.88	5.97	12.88	5.97	12.55	5.71	-0.33	-0.26
5	14.95	6.10	14.94	6.11	14.93	6.10	14.58	5.79	-0.37	-0.32
6	18.60	5.38	18.60	5.38	18.61	5.39	18.26	5.05	-0.35	-0.33
7	18.79	4.66	18.79	4.67	18.80	4.67	18.47	4.36	-0.33	-0.30
8	21.55	6.00	21.54	6.00	21.52	6.00	21.18	5.63	-0.37	-0.37
9	16.54	5.28	16.53	5.28	16.54	5.28	16.35	5.01	-0.19	-0.27
10	11.88	2.86	11.88	2.86	11.87	2.85	11.84	2.79	-0.05	-0.08
11	9.08	4.34	9.07	4.34	9.07	4.34	9.05	4.31	-0.03	-0.03
12	5.79	3.63	5.78	3.63	5.79	3.62	5.80	3.62	0.01	0.00
2021										
1	3.58	2.95	3.57	2.96	3.57	2.95	3.58	2.94	0.01	-0.02
2	5.59	6.88	5.59	6.89	5.59	6.88	5.50	6.84	-0.09	-0.04
3	7.55	4.84	7.55	4.84	7.51	4.80	7.40	4.61	-0.16	-0.23
4	7.80	4.95	7.80	4.95	7.78	4.96	7.60	4.76	-0.21	-0.20
5	12.30	4.75	12.30	4.75	12.29	4.76	12.10	4.52	-0.20	-0.23
6	19.70	5.11	19.70	5.11	19.69	5.12	19.37	4.77	-0.33	-0.34
7	18.72	3.64	18.71	3.64	18.71	3.65	18.48	3.29	-0.23	-0.34
8	17.51	3.53	17.51	3.54	17.51	3.54	17.38	3.22	-0.13	-0.31
9	16.88	4.52	16.87	4.52	16.86	4.52	16.68	4.26	-0.20	-0.26
10	11.68	3.55	11.67	3.55	11.66	3.56	11.63	3.44	-0.05	-0.11
11	6.38	3.29	6.37	3.29	6.36	3.29	6.37	3.26	-0.01	-0.03
12	5.90	3.92	5.89	3.92	5.88	3.93	5.90	3.92	0.00	0.00

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

6 Conclusion

This study presents the data from the citizen science weather station network Leuven.cool, which consists of around 100 weather stations in the city of Leuven, Belgium. The crowdsourced weather stations (Fine Offset WH2600) are distributed across Leuven and surroundings, measuring the local climate since July 2019. The dataset is accompanied by a newly developed station specific temperature quality control procedure. The quality control method consists of three levels, removing implausible measurements, while also correcting for inter (in between stations) - and intra (station-specific) station temperature biases. This QC method combines suggestions of previous developed methods but improves them by correcting aberrant temperature observations rather than removing them. As a result, more data can be retained allowing researchers to study the highly heterogeneous urban climate in all its detail. Moreover, the QC method uses information from the crowdsourced data itself and only requires reference data from official stations during its development and evaluation stage. As a consequence, the method is easily transferable to other urban regions not having an official weather station. A validation of the proposed QC method was carried out on four Leuven.cool stations installed next to official equipment, and showed that it is able to reduce the mean temperature difference and standard deviation from 0.15 ± 0.56°C to 0.00 ± 0.22°C. Knowing that both the frequency and intensity of heat waves will only increase during the upcoming years, dense high-quality datasets such as the Leuven.cool datasets become highly valuable for studying local climate phenomena, planning efficient mitigation and adaptation measures and hence mitigating future risks.

Data and code availability

All data described in this paper and scripts used to design, evaluate and apply the QC method are stored in RDR, KU Leuven's Research Data Repository and accessible through the following DOI: https://doi.org/10.48804/SSRN3F (Beele et al., 2022).

Author contributions

All authors contributed to the design of the weather station network. EB and MR practically implemented the design and collected the CWS data, MR assembled AWS data and EB processed all data used. All authors contributed to the design of the QC method, EB did the programming of the QC. EB carried out all analyses and mainly wrote the manuscript. All authors discussed the results and contributed to the writing of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Acknowledgements

We thank the many people who have contributed to the establishment and maintenance of the Leuven.cool project. In particular we thank all citizens and private companies of/within Leuven who voluntary made their garden/terrain available for our research. We also thank Tim Guily (city of Leuven) and Hanne Wouters (Leuven2030) for their logistic support and Margot Verhulst, Jordan Rodriguez Milis, Remi Chevalier and Jingli Yan (KU Leuven) for their technical support during the realisation of the network. The protocol of this study was approved by the Social and Societal Ethics Committee of the KU Leuven (G-2019 06 1674). The weather stations used in this study were sponsored by both the city of Leuven and the KU Leuven. E.B. holds a SB-doctoral fellowship of the Research Foundation Flanders (FWO, 1SE0621N).

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Appendices

645 **Appendix A. Specifications of the WH2600 digital weather station**

The technical specifications of the Fine Offset WH2600 weather station are given in Table A1.

Table A1: Technical specifications of the Fine Offset WH2600 weather station as given by the manufacturer.

Outdoor sensor array	
Transmission distance in open field	100m
Temperature range	-40°C - 60°C
Temperature accuracy	+/- 1°C
Temperature resolution	0.1°C
Relative humidity range	1% - 99%
Relative humidity accuracy	+/- 5%
Rain volume range	0 – 9999mm
Rain volume accuracy	+/- 10%
Rain volume resolution	0.3mm (if rain volume < 1000mm)
	1mm (if rain volume > 1000mm)
Wind speed range	0 - 50m/s
Wind speed accuracy	+/- 1m/s (wind speed < 5m/s)
	+/- 10% (wind speed > 5m/s)
Light range	0 - 400k Lux
Light accuracy	+/- 15%
Measuring interval	16 sec

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Appendix B. Specifications LCZ map

The LCZ map was created with the LCZ generator developed within the WUDAPT project (Demuzere et al., 2021). A grid area of 15 by 15 km was drawn around the city centre of Leuven. Within this grid area training polygons for 10 LCZ types were drawn using the Training Area Template kml file proposed by the WUDAPT project. Based on these training polygons the LCZ generator provides the user with a final LCZ map and corresponding accuracy metrics. The LCZ map was reprojected from ESPG:4326 – WGS 84 to ESPG: 31370 – Belge Lambert 72 using the *projectRaster* function and nearest neighbour method in R.

The resulting LCZ map has an overall accuracy of 0.73, the accuracy for each LCZ can be found in Figure B1. The confusion matrix is presented in Table B1.

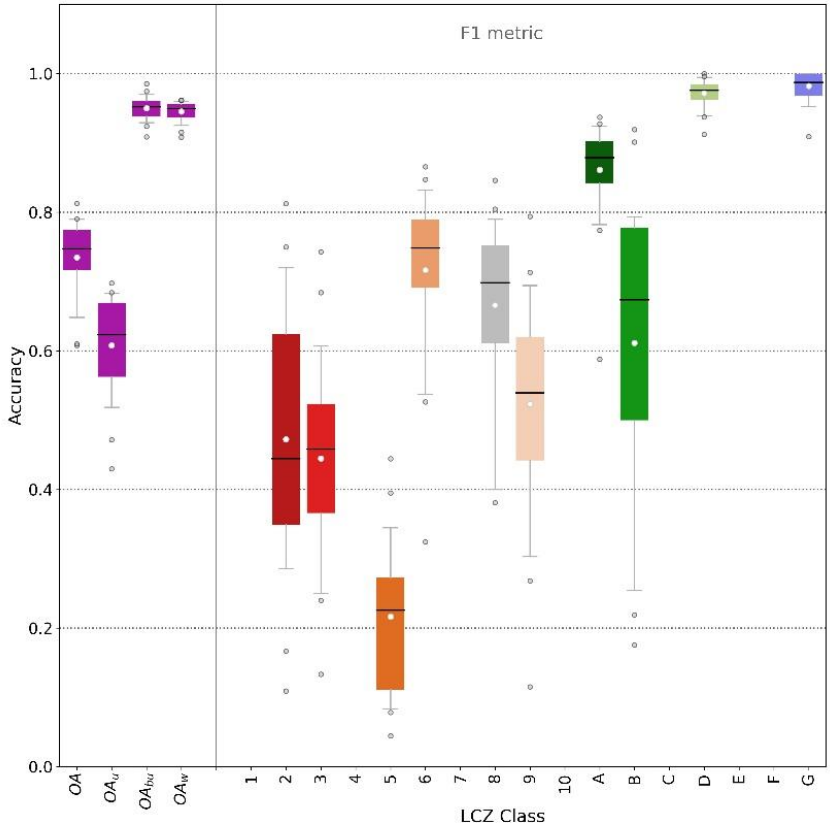


Figure B1: Accuracy metrics LCZ map. A boxplot of the overall accuracy (OA), the overall accuracy of urban classes (1-10) (OA_U), the overall accuracy of built zones versus natural zones (OA_{BU}), the weighted accuracy (OA_W) and the F1 metric of each LCZ class are given are given.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Table B1: Confusion matrix LCZ map. The matric summarises the number of cells wrongly or correctly classified for each LCZ class. The user accuracy (UA) and producer accuracy (PA) for each LCZ class and the overall accuracy are included as well.

	LCZ 2	LCZ 3	LCZ 5	LCZ 6	LCZ 8	LCZ 9	LCZ 11	LCZ 12	LCZ 14	LCZ 17	Total	UA (%)
LCZ 2	8.0	7.0	0.0	0.0	3.0	0.0	0.0	1.0	0.0	0.0	19.0	42.1
LCZ 3	6.0	2.0	1.0	1.0	3.0	0.0	0.0	0.0	0.0	0.0	13.0	15.4
LCZ 5	2.0	2.0	20.0	9.0	15.0	2.0	0.0	0.0	1.0	0.0	51.0	39.2
LCZ 6	0.0	4.0	7.0	75.0	3.0	7.0	0.0	0.0	0.0	0.0	96.0	78.1
LCZ 8	1.0	1.0	9.0	10.0	37.0	1.0	0.0	1.0	0.0	0.0	60.0	61.7
LCZ 9	0.0	1.0	2.0	7.0	0.0	34.0	2.0	18.0	1.0	1.0	66.0	51.5
LCZ 11	0.0	0.0	0.0	0.0	0.0	1.0	160.0	35.0	0.0	0.0	196.0	81.6
LCZ 12	0.0	0.0	0.0	1.0	0.0	1.0	0.0	8.0	0.0	0.0	10.0	80.0
LCZ 14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	95.0	0.0	95.0	100.0
LCZ 17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0	15.0	100.0
Total	17.0	17.0	39.0	103.0	61.0	46.0	162.0	63.0	97.0	16.0	621.0	
PA (%)	47.1	11.8	51.3	72.8	60.7	73.9	98.8	12.7	97.9	93.8		73.1

670

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

References

Arnfield, A. J.: Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island, Int. J. Climatol., 23(1), 1–26, doi:10.1002/joc.859, 2003.

Båserud, L., Lussana, C., Nipen, T. N., Seierstad, I. A., Oram, L. and Aspelien, T.: TITAN automatic spatial quality control of meteorological in-situ observations, Adv. Sci. Res., 17, 153–163, doi:10.5194/asr-17-153-2020, 2020.

Bassani, F., Garbero, V., Poggi, D., Ridolfi, L., Hardenberg, J. Von and Milelli, M.: Urban Climate An innovative approach to select urban-rural sites for Urban Heat Island analysis : the case of Turin (Italy), Urban Clim., 42(February), 101099, doi:10.1016/j.uclim.2022.101099, 2022.

Beele, E., Reyniers, M., Aerts, R. and Somers, B.: Replication Data for: Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium, , doi:https://doi.org/10.48804/SSRN3F, 2022.

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

Bevolkingsregister Stad Leuven: Demografie, Leuven in cijfers [online] Available from: https://leuven.incijfers.be/dashboard/dashboard/demografie (Accessed 15 December 2021), 2021.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D. and Bartonova, A.: Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates ?, Environ. Int., 99, 293–302, doi:10.1016/j.envint.2016.12.007, 2017.

Chapman, L., Bell, C. and Bell, S.: 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., 37(9), 3597–3605, doi:10.1002/joc.4940, 2017.

Chen, J., Saunders, K. and Whan, K.: Quality control and bias adjustment of crowdsourced wind speed observations, Q. J. R. Meteorol. Soc., 147(740), 3647–3664, doi:10.1002/qj.4146, 2021.

Cornes, R. C., Dirksen, M. and Sluiter, R.: Correcting citizen-science air temperature measurements across the Netherlands for short wave radiation bias, Meteorol. Appl., 27(1), 1–16, doi:10.1002/met.1814, 2020.

Demuzere, M., Kittner, J. and Bechtel, B.: LCZ Generator: A Web Application to Create Local Climate Zone Maps, Front. Environ. Sci., 9, 1–29, doi:10.3389/fenvs.2021.637455, 2021.

EEA: Assessing air quality through citizen science, Copenkagen., 2019.

Feichtinger, M., Wit, R. De, Goldenits, G., Kolejka, T. and Hollósi, B.: Urban Climate Case-study of neighborhood-scale summertime urban air temperature for the City of Vienna using crowd-sourced data, Urban Clim., 32(August 2019), 1–12, doi:10.1016/j.uclim.2020.100597, 2020.

Fenner, D., Meier, F., Bechtel, B., Otto, M. and Scherer, D.: Intra and inter "local climate zone" variability of air temperature

© Author(s) 2022. CC BY 4.0 License.

as observed by crowdsourced citizen weather stations in Berlin, Germany, Meteorol. Zeitschrift, 26(5), 525–547, doi:10.1127/metz/2017/0861, 2017.

Fenner, D., Bechtel, B., Demuzere, M., Kittner, J. and Meier, F.: CrowdQC + — A Quality-Control for Crowdsourced Air-Temperature Observations Enabling World-Wide Urban Climate Applications, Front. Environ. Sci., 9(December), 1–21, doi:10.3389/fenvs.2021.720747, 2021.

Hammerberg, K., Brousse, O., Martilli, A. and Mahdavi, A.: Implications of employing detailed urban canopy parameters for mesoscale climate modelling: a comparison between WUDAPT and GIS databases over Vienna, Austria, Int. J. Climatol., 38(February), e1241–e1257, doi:10.1002/joc.5447, 2018.

Heaviside, C., Macintyre, H. and Vardoulakis, S.: The Urban Heat Island: Implications for Health in a Changing Environment, Curr. Environ. Heal. reports, 4(3), 296–305, doi:10.1007/s40572-017-0150-3, 2017.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press., 2021.

Jenkins, G.: A comparison between two types of widely used weather stations, Weather, 69(4), 105–110, doi:10.1002/wea.2158, 2014.

Kidder, S. Q. and Essenwanger, O. M.: The Effect of Clouds and Wind on the Difference in Nocturnal Cooling Rates between Urban and Rural Areas, J. Appl. Meteorol., 34(11), 2440–2448, doi:10.1175/1520-0450(1995)034<2440:TEOCAW>2.0.CO;2, 1995.

Kirk, P. J., Clark, M. R. and Creed, E.: Weather Observations Website, Weather, 76, 47–49, doi:https://doi.org/10.1002/wea.3856, 2020.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. and Rubel, F.: World map of the Köppen-Geiger climate classification updated, Meteorol. Zeitschrift, 15(3), 259–263, doi:10.1127/0941-2948/2006/0130, 2006.

Kousis, I., Pigliautile, I. and Pisello, A. L.: Intra - urban microclimate investigation in urban heat island through a novel mobile monitoring system, Sci. Rep., 11(9732), 1–17, doi:10.1038/s41598-021-88344-y, 2021.

Leuven.cool: Leuven.cool, Leuven 2030 [online] Available from: https://www.leuven.cool/ (Accessed 15 December 2020), 2021.

Longman, R. J., Giambelluca, T. W., Nullet, M. A., Frazier, A. G., Kodama, K., Crausbay, S. D., Krushelnycky, P. D., Cordell, S., Clark, M. P., Newman, A. J. and Arnold, J. R.: Compilation of climate data from heterogeneous networks across the Hawaiian Islands, Sci. Data, 5(1), 180012, doi:10.1038/sdata.2018.12, 2018.

Lu Aigang, Wang Tianming, Kang Shichang and Pang Deqian: On the Relationship between Latitude and Altitude Temperature Effects, in 2009 International Conference on Environmental Science and Information Application Technology, vol. 2, pp. 55–58, IEEE., 2009.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J. C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A. L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J. F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V. and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6(4), 929–960, doi:10.5194/gmd-6-929-2013, 2013.

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

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. R.: Crowdsourcing for climate and atmospheric sciences: Current status and future potential, Int. J. Climatol., 35(11), 3185–3203, doi:10.1002/joc.4210, 2015.

Napoly, A., Grassmann, T., Meier, F. and Fenner, D.: Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data, Front. Earth Sci., 6(August), 1–16, doi:10.3389/feart.2018.00118, 2018.

Nipen, T. N., Seierstad, I. A., Lussana, C., Kristiansen, J. and Hov, Ø.: Adopting citizen observations in operational weather prediction, Bull. Am. Meteorol. Soc., 101(1), E43–E57, doi:10.1175/BAMS-D-18-0237.1, 2020.

Oke, T. R.: City size and the urban heat island, Atmos. Environ. Pergamon Pres, 7, 769–779, 1973.

Qian, Y., Zhou, W., Hu, X. and Fu, F.: The Heterogeneity of Air Temperature in Urban Residential Neighborhoods and Its Relationship with the Surrounding Greenspace, Remote Sens., 10(6), 965, doi:10.3390/rs10060965, 2018.

De Ridder, K., Lauwaet, D. and Maiheu, B.: UrbClim - A fast urban boundary layer climate model, Urban Clim., 12, 21–48, doi:10.1016/j.uclim.2015.01.001, 2015.

Rizwan, A. M., Dennis, L. Y. C. and Lia, C.: A review on the generation, determination and mitigation of Urban Heat Island, J. Environ. Sci., 20(1), 120–128, doi:10.1016/S1001-0742(08)60019-4, 2008.

RMI: Klimaatstatistieken van de Belgische gemeenten: Leuven, Leuven., 2020.

Sotelino, L. G., De Coster, N., Beirinckx, P. and Peeters, P.: Intercomparison of Shelters in the RMI AWS Network., 2018.

Stewart, I. D. and Oke, T. R.: Local climate zones for urban temperature studies, Bull. Am. Meteorol. Soc., 93(12), 1879–1900, doi:10.1175/BAMS-D-11-00019.1, 2012.

UN: World Urbanization Prospects: The 2018 Revision, New York., 2018.

Verdonck, M. L., Demuzere, M., Hooyberghs, H., Beck, C., Cyrys, J., Schneider, A., Dewulf, R. and Van Coillie, F.: The potential of local climate zones maps as a heat stress assessment tool, supported by simulated air temperature data, Landsc. Urban Plan., 178(July 2017), 183–197, doi:10.1016/j.landurbplan.2018.06.004, 2018.

de Vos, Droste, A. M., Zander, M. J., Overeem, A., Leijnse, H., Heusinkveld, B. G., Steeneveld, G. J. and Uijlenhoet, R.: Hydrometeorological monitoring using opportunistic sensing networks in the Amsterdam metropolitan area, Bull. Am. Meteorol. Soc., 101(2), E167–E185, doi:10.1175/BAMS-D-19-0091.1, 2020.

de Vos, L., Leijnse, H., Overeem, A. and Uijlenhoet, R.: The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam, Hydrol. Earth Syst. Sci., 21(2), 765–777, doi:10.5194/hess-21-765-2017, 2017.

de Vos, L., Leijnse, H., Overeem, A. and Uijlenhoet, R.: Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring, Geophys. Res. Lett., 46(15), 8820–8829, doi:10.1029/2019GL083731, 2019.

https://doi.org/10.5194/essd-2022-113
Preprint. Discussion started: 2 May 2022
© Author(s) 2022. CC BY 4.0 License.

WMO: Guide to meteorological instruments and methods of observation. Volume I – Measurement of Meteorological Variables, Geneva., 2018.

WOW-BE: Weather Observations Website - Belgium, KMI - MetOffice [online] Available from: https://wow.meteo.be/nl/, 2021.

Yang, Q., Huang, X., Yang, J. and Liu, Y.: The relationship between land surface temperature and artificial impervious surface fraction in 682 global cities: Spatiotemporal variations and drivers, Environ. Res. Lett., 16(2), doi:10.1088/1748-9326/abdaed, 2021.