Authors' response to RC2 - manuscript ESSD-2022-113

Responses to Reviewer comments: The text in blue represents the authors' reply. Text in italic represents text that was changed in the manuscript, with line numbers referring to those in the revised manuscript. We numbered the comments for easier referral.

General comments:

Small typographic errors (missing punctuation or erroneous capitalization) that proofreaders should catch.

Data accessible, clearly described, easy to work with. This reviewer prefers .csv to .tab but that seems a matter of individual preference.

I understand why authors might prefer CC-BY-NC but, due to proper ESSD cautions, have editors granted permission for -NC?

Overall: good well-organized description of QC levels and good summary of 'final' QC outcome, 0.15 + 0.56 down to 0.00 + 0.22 for temperature. In describing and summarizing these efforts and QC levels, the authors seem to have left out or passed over several issues.

Reply: We thank the reviewer for the thorough review. We have considered all your comments and revised the manuscript accordingly. We have proofread the manuscript to correct for additional typographic errors. We also like to mention that the data depository automatically transforms .csv files to .tab files. There is, however, a possibility to download the files as .csv. Regarding the CC-BY-NC license, this was not raised as a problem during the initial submission phase. If the editors would not grant permission for -NC, we can change this. We tried to rephrase and further explain the QC methodology in order to solve the issues mentioned below. Please see our point-by-point responses on the specific comments below.

Specific comments:

1. Daytime vs nighttime much used but little described. Based on daily hours sunlight, calculated by day, month or season? Or calculated based on measured surface radiation values where night = low light = < some minimum value of W/m2?

Reply: Thanks for the valid comment. We agree that the differentiation between night- and daytime was not clearly described. Currently it was only mentioned in the caption of figure 3. The distinction between day- and nighttime is based on the radiation measurements of each station; daytime is defined as $LC_RAD > 0$ W/m2, nighttime is defined as $LC_RAD = 0$ W/m2. This has been further clarified in the manuscript. We should further state that the differentiation between night and day is only for illustrative purposes. Only one RF model has been made for both day and night, the differentiation between day and night is however indirectly present in the model through the input variables "Hour", "Radiation" and "Radiation60".

'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 (radiation > 0 W/m2) and night (radiation = 0 W/m2), a positive mean temperature bias during daytime (0.32 °C) and a negative mean temperature bias during night time (-0.10 °C) is obtained.' (Lines 348-351) 'The random forest prediction of temperature bias showed the best results. By splitting up the results for day (radiation > 0 W/m2) and night (radiation = 0 W/m2) (Figure 6), a smaller standard deviation of the bias during night time (0.25) compared to daytime (0.31) is obtained. This differentiation between night and day is only for illustrative purposes, only one RF was built for both day and night. The statistical details of the random forest model are further summarized in Table 9.' (Lines 425-428)

2. Other parameters (RH, precip, wind) not corrected. Authors describe, correctly, the use of non-T parameters in various QC functions but otherwise give no hints about suitability, difficulty or desirability of addressing RH, precip, etc. Do readers await a future paper? Do we assume QC of T proved easier than other parameters? Authors give us no hint what to assume!

Reply: We understand the question of the reviewer. Before using the non-T parameters in various steps of the QC, we have made a qualitative assessment of the data quality by making scatterplots of the non-T parameters for the LC-R stations, compared to the AWS data. Overall, the measured parameters are within the accuracy given by the manufacturer. These scatterplots have been added as an Appendix (Appendix C) to the manuscript.

There are, however, deviations in the wind and radiation measurement that are attributable to the small location differences (in the order of meters) between the LC-R stations and the official sensors. For wind, we compared the LC-R data with professional 2 m wind measurements, but this height is in fact not the standard measurement height for wind since too many ground effects are still into play at this height. For radiation, we already mentioned the design flaw regarding the wind vane dropping a shadow, but also high nearby trees can influence the measurements for low solar elevations (in the case of LC-R01 and LC-R02).

The main focus of the authors research is related to urban temperatures and the UHI effect. Currently no QC for the other parameters are planned.

'A qualitative assessment of the data quality of these variables in included in Appendix C.' (Line 204)

'Appendix C. Quality control other Leuven.cool variables

A qualitative assessment of the data quality was performed by making scatterplots of these variables for the LC-R stations compared to the AWS stations (Figure C1).

Overall, the measured parameters are within the accuracy given by the manufacturer. There are, however, deviations in the wind and radiation measurement that are attributable to the small location differences (in the order of meters) between the LC-R stations and the official sensors. For wind, we compared the LC-R data with professional 2m wind measurements, but this height is in fact not the standard measurement height for wind since too many ground effects are still into play at this height. For radiation, we already mentioned the design flaw regarding the wind vane dropping a shadow, but also high nearby trees can influence the measurements for low solar elevations (in the case of LC-R01 and LC-R02).







Figure C1: Scatterplots of LC-R versus AWS dew point temperature, humidity, radiation and wind speed for each reference station LC-R01, LC-R02, LC-R04, LC-R05. The identity line is shown in black. The colour scale indicates the density of observations, yellow indicating the highest, purple the lowest. The scatterplots include all measurements between the installation data of each LC-R and December 2021. For each variable the same ranges of x- and y-axis are used.'

(Lines 798-820)

3. Authors should, technically, express offsets and SD in K rather than C. Not often done, I accept.

Reply: Thanks for the comment. We have followed related literature on air temperature quality control and correction methods, which choose to express temperature offsets and SD in °C. A small clarification is however added to the manuscript.

'The outdoor sensor array measures temperature (°C, add 273.15 for K), humidity (%), precipitation (mm), wind speed (m/s), wind direction (°), solar radiation (W/m2), and UV (-) every 16 seconds.' (Lines 145-147)

4. Authors have lumped 2 m data with 3-4 m data? Early mention but then no subsequent treatment. One suspects both sensor elevation and surrounding (mostly impervious) surfaces would have a large effect but never mentioned? Shadowing by anemometer mentioned occasionally but that occurs independently of short vs high poles and regardless of underlying surface?

Reply: The installation height of the weather station is indeed not directly added as a parameter in the RF model (QC L3) in order to correct the T observations. This was not possible since all the LC-R stations were installed at a height of 2 m. We do however believe that the installation height is indirectly present in the RF model through the radiation and wind speed parameters. Furter, the altitude (above sea level) of the weather station will more clearly influence the measured temperature and is added as a parameter in the RF model.

We do acknowledge that nocturnal temperature inversions can appear under clear calm nights (under low radiation and low windspeed conditions). During such a nocturnal inversion we normally see an increase in temperature with height due to radiational cooling of the earth's surface (Ahrens, 2009). Such inversions however often occur at ground level, and thus below 2 m (Ahrens, 2009).

To quantify the temperature difference between 2 m and 3 m, a small test was performed using multiple AWS stations having T observations at both 2 m and 10 m. By interpolating the temperature at 2 m and

10 m, a temperature at 3 m is obtained. Next the average temperature difference and standard deviation between 3 m and 2 m was calculated for the whole of 2021 and for stable conditions $(T_{10 m} > T_{2 m})$ in which we expect the highest difference between 2 m and 3 m. For the whole of 2021 the mean difference equals 0 ± 0.09 °C, for the stable conditions we obtain a small difference of 0.08 ± 0.09 °C. We should thus state that stations installed at 3 m – 4 m can have an additional offset up to 0.08 ± 0.09 °C under stable conditions. This offset is much smaller than the night time temperature difference caused by the UHI effect (Chapman et al., 2017; Stewart, 2011; Venter et al., 2021).

We further stress that stations have only been installed at a height of 3 m to 4 m where it was needed due safety reasons.

Reply: The environment and surrounding surfaces of each sensor will indeed further influence the temperature measurements (Logan et al., 2020; J. Wang et al., 2022; Q. Wang et al., 2022; Ziter et al., 2019). The weather stations were however installed following a strict protocol, at least 1 m from interfering objects, making sure no direct effects of the environment incorrectly influence the T observations. On the other hand, the goal of this weather station network is to measure the micro-climate in Leuven, namely the difference between impervious and greener locations within and outside the city center. In other words, our stations are installed at both impervious and green locations within and outside the city center in order to measure (and in the future also explain) these temperature differences.

Reply: The shadowing of the anemometer is intrinsic to the station design and hence the same for every station, also the LC-R stations. As a consequence, the effect is included in our RF model.

5. UHI: This reader missed an overall assessment of data as QC'd here to address UHI. Do we need 0.1K? 1.0K? Have the authors come close with these corrections. In view of distance of reference AWS from Leuven and, for two RMIB stations at least, distance from true urban settings, do authors feel they have a network now suitable for addressing UHI. Not clear, perhaps needs/deserves further clarification. Given sensor height differences already mentioned plus apparent absence (or, avoidance) of the most urban land use categories, can users really trust this data for further UHI work? UHI mentioned frequently in introduction but not at all in conclusion paragraph. I agree with summary sentences but, if they could not in the end address UHI - which, in many indices, includes high net radiation, low RH and low wind - should they have given so much attention in introduction

Reply: We thank the reviewer for the valid comment. The resolution on which the UHI should be investigated is highly dependent on the application. For human thermal comfort studies, a resolution of 1° C would however be sufficient (Epstein & Moran, 2006; Georgi & Zafiriadis, 2006). Before the QC method, a mean temperature bias up to $0.15 \pm 0.56 \,^{\circ}$ C is noted. After correction the mean bias has been diminished to $0.00 \pm 0.28 \,^{\circ}$ C. Numerous studies have however showed that the UHI effect causes nighttime temperature differences up 6 to 9 $^{\circ}$ C during clear nights. These thresholds are much higher than mean bias obtained after correction, the quality controlled Leuven.cool dataset is thus suitable to study the UHI (Chapman et al., 2017; Stewart, 2011; Venter et al., 2021).

We further showed that the QC method can correct the temperature bias equally across different hours of the day and months of the year (Figure 16) as well as under different radiation and windspeed conditions (Figure 17). Most UHI indicators are based on either maximum daytime or minimum nighttime temperatures during extreme heat events (high radiation + low windspeed), while these time

periods show the highest temperature bias (Figure 10 and Figure 11). Before correction these indicators would have either a positive (daytime) or negative (nighttime) bias, the ideal meteorological conditions would result in a positive bias. This QC method ensures that day -and nighttime effects as well as effects due to certain meteorological conditions have been corrected for. As a result, we can trust UHI indicators calculated for specific day- and night time events.

'Numerous studies have shown that the UHI effect causes night time temperature differences up 6 to 9 °C during clear nights (Chapman et al., 2017; Venter et al., 2021; Stewart, 2011; Napoly et al., 2018; Feichtinger et al., 2020). These thresholds are much higher than mean bias obtained after correction; the dense quality controlled Leuven.cool dataset thus allows for microscale modelling of urban weather patterns, including the urban heat island (Chapman et al., 2017; de Vos et al., 2020; Napoly et al., 2018; Feichtinger et al., 2020).' (Lines 683-687)

Reply: Regarding the distance between the AWS and the LC-X in Leuven, we don't fully understand the question.

We agree that there is quite a difference between the AWS and the LC-X in Leuven, and that due to this difference the AWS stations cannot serve as a direct reference for the LC-X in Leuven. We should however state that the RF model is based on the T difference between the official AWS and the LC-R (installed next to the AWS), the RF model does not take into account the standalone measurements of the AWS itself. We further installed reference stations (LC-R) at three different locations (Uccle, Diepenbeek, Humain) to make the model more robust against spatial differences.

For addressing the UHI, the authors will not take AWS measurements into account but exclusively focus on the LC-X station in and outside Leuven.

Reply: Regarding the apparent absence of most urban land use classes, we are a bit confused. Our weather station network mainly focusses on the urban LCZ classes, natural LCZ classes are not as well represented.

The selection procedure for suitable locations started as a stratified sampling based on the concept of LCZs (by looking at building height, building density and vegetation cover). The goal was to divide the stations across the LCZ classes so that they would represent the spatial coverage of each LCZ. We should state that the network was implemented with the intention of gaining knowledge on the mitigating effect of green and blue infrastructures within urban settings. The initial study area thus mainly included the city centre of Leuven. As a result, the network has a clear bias towards urban classes.

Figure 1 shows the distribution of the weather stations across the LCZ classes (left panel) and proportion of LCZ classes across study area (right panel). 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. Due to a technical limitation of the weather station, it cannot be installed in natural environments without a LAN connection within 50 to 100 meters of the weather station.



Figure 1: Distribution of weather stations across LCZ classes (left panel) and proportion of LCZ classes across study area (right panel).

Reply: We understand the reviewer's comment on the lacking of the UHI in the conclusion. We have added a section with additional applications of this dataset, including the investigation of the UHI effect. These applications were summarized in the conclusion.

'6. Application potential of the quality controlled and corrected Leuven.cool dataset

A validation of the proposed QC method showed that it can reduce the mean temperature difference and standard deviation from 0.15 ± 0.56 °C to 0.00 ± 0.28 °C. The QC method can correct the temperature difference equally across different hours of the day and months of the year (Figure 16) as well and under different radiation and windspeed conditions (Figure 17).

The quality-controlled Leuven.cool dataset enables a detailed comparison with other crowdsourced datasets for which less or even no metadata is available. As such, the Leuven.cool stations can serve as gatekeepers for other crowdsourced observations. In the past this role has been limited to standard weather station network which mostly only have a limited number of observations available (Chapman et al., 2017).

Numerous studies have shown that the UHI effect causes night time temperature differences up 6 to 9 °C during clear nights (Chapman et al., 2017; Feichtinger et al., 2020; Napoly et al., 2018; Stewart, 2011; Venter et al., 2021). These thresholds are much higher than mean bias obtained after correction. The dense quality controlled Leuven.cool dataset thus allows for microscale modelling of urban weather patterns, including the urban heat island (Chapman et al., 2017; de Vos et al., 2020; Feichtinger et al., 2020; Napoly et al., 2018). Since such high-quality datasets contain measurements with both high spatial and temporal resolution, they can easily be used to obtain spatially continuous temperature patterns across a region (e.g. Feichtinger et al., 2020; Napoly et al., 2018). Interpolation methods based on single pair stations or mobile transect methods are much less trustworthy (Napoly et al., 2018). Dense weather station networks can be used to investigate the inter- and intra LCZ variability within a city (Fenner et al., 2017; Verdonck et al., 2018). The dataset can further help investigate the relation between temperature and human and ecosystem health (e.g. Aerts et al., 2022; Troeyer et al., 2020) and their effect on evolutionary processes (e.g. Brans et al., 2022).

The dataset can also help refine existing weather forecast models which are currently mostly based on official rural observations (Sgoff et al., 2022). Nipen et al. (2020) showed that the inclusion of citizen observations improves the accuracy of short-term temperature forecasts in regions where official stations are sparse. Mandement & Caumont (2020) used crowdsourced weather stations to improve the observation and prediction of near convection. After quality control and correction, also the wind (Chen

et al., 2021) and precipitation measurements (de Vos et al., 2019) can be useful to improve detection and forecasting. Further, the Leuven.cool dataset could be a useful input in air pollution prediction models (e.g. IFD-model (Lefebvre et al., 2011)).' (Lines 673-701)

'The quality-controlled Leuven.cool dataset enables a detailed comparison with other crowdsourced datasets for which less or even no metadata is available. The dense dataset further allows for microscale modelling of urban weather patterns, such as the urban heat island, and can help identify the relation between temperature and human and ecosystem and their effect on evolutionary processes. Lastly the dataset could be used to refine existing forecast models which are currently mostly based on official rural observations.' (Lines 723-727)

Reply: We agree with the reviewer's comment that UHI effect is mostly investigated under high radiation, low relative humidity and low wind- speed conditions. As a consequence, we must make sure that the QC method is also valid under these meteorological conditions. If this is not the case, the dataset cannot be used to study the UHI with high accuracy.

We should however state that QC also works for high radiation and low windspeed conditions. The QC method can correct the temperature bias equally across different hours of the day and months of the year (Figure 16) as well as under different radiation and windspeed conditions (Figure 17). Most UHI indicators are based on either maximum daytime or minimum nighttime temperatures during extreme heat events (high radiation + low windspeed), while these time periods show the highest temperature bias (Figure 10 and Figure 11). Before correction these indicators could have either a positive (daytime) or negative (nighttime) bias, the ideal meteorological conditions would result in a positive bias. This QC method ensures that day -and nighttime effects as well as effects due to certain meteorological conditions have been corrected for.

References:

- Aerts, R., Vanlessen, N., Dujardin, S., Nemery, B., Van Nieuwenhuyse, A., Bauwelinck, M., Casas, L., Demoury, C., Plusquin, M., & Nawrot, T. S. (2022). Residential green space and mental health-related prescription medication sales: An ecological study in Belgium. *Environmental Research*, 211. https://doi.org/10.1016/j.envres.2022.113056
- Ahrens, C. D. (2009). *Meteorology Today: An Introduction to Weather, Climate, and the Environment* (Ninth Edit). Brooks/Cole.
- Brans, K. I., Tüzün, N., Sentis, A., De Meester, L., & Stoks, R. (2022). Cryptic eco-evolutionary feedback in the city: Urban evolution of prey dampens the effect of urban evolution of the predator. *Journal of Animal Ecology*, *91*(3), 514–526. https://doi.org/10.1111/1365-2656.13601
- Chapman, L., Bell, C., & 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. *International Journal of Climatology*, 37(9), 3597–3605. https://doi.org/10.1002/joc.4940
- Chen, Y., Weng, Q., Tang, L., Liu, Q., Zhang, X., & Bilal, M. (2021). Automatic mapping of urban green spaces using a geospatial neural network. *GIScience & Remote Sensing*, 00(00), 1–19. https://doi.org/10.1080/15481603.2021.1933367
- de Vos, Droste, A. M., Zander, M. J., Overeem, A., Leijnse, H., Heusinkveld, B. G., Steeneveld, G. J., & Uijlenhoet, R. (2020). Hydrometeorological monitoring using opportunistic sensing networks in the Amsterdam metropolitan area. *Bulletin of the American Meteorological Society*, 101(2), E167–E185. https://doi.org/10.1175/BAMS-D-19-0091.1
- de Vos, L., Leijnse, H., Overeem, A., & Uijlenhoet, R. (2019). Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring. *Geophysical Research Letters*, 46(15), 8820–8829. https://doi.org/10.1029/2019GL083731
- Epstein, Y., & Moran, D. S. (2006). Thermal comfort and the heat stress indices. In *Industrial Health*. https://doi.org/10.2486/indhealth.44.388
- Feichtinger, M., Wit, R. De, Goldenits, G., Kolejka, T., & Hollósi, B. (2020). Urban Climate Casestudy of neighborhood-scale summertime urban air temperature for the City of Vienna using crowd-sourced data. Urban Climate, 32(August 2019), 1–12. https://doi.org/10.1016/j.uclim.2020.100597
- 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. *Meteorologische Zeitschrift*, 26(5), 525–547. https://doi.org/10.1127/metz/2017/0861
- Georgi, N. J., & Zafiriadis, K. (2006). The impact of park trees on microclimate in urban areas. *Urban Ecosystems*, 9(3), 195–209. https://doi.org/10.1007/s11252-006-8590-9
- Lefebvre, W., Vercauteren, J., Schrooten, L., Janssen, S., Degraeuwe, B., Maenhaut, W., de Vlieger, I., Vankerkom, J., Cosemans, G., Mensink, C., Veldeman, N., Deutsch, F., Van Looy, S., Peelaerts, W., & Lefebre, F. (2011). Validation of the MIMOSA-AURORA-IFDM model chain for policy support: Modeling concentrations of elemental carbon in Flanders. *Atmospheric Environment*, 45(37), 6705–6713. https://doi.org/10.1016/j.atmosenv.2011.08.033
- Logan, T. M., Zaitchik, B., Guikema, S., & Nisbet, A. (2020). Night and day: The influence and relative importance of urban characteristics on remotely sensed land surface temperature. *Remote Sensing of Environment*, 247(June), 111861. https://doi.org/10.1016/j.rse.2020.111861
- Mandement, M., & Caumont, O. (2020). Contribution of personal weather stations to the observation of deep-convection features near the ground. *Natural Hazards and Earth System Sciences*, 20(1), 299–322. https://doi.org/10.5194/nhess-20-299-2020

- Napoly, A., Grassmann, T., Meier, F., & Fenner, D. (2018). Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data. *Frontiers in Earth Science*, 6(August), 1–16. https://doi.org/10.3389/feart.2018.00118
- Nipen, T. N., Seierstad, I. A., Lussana, C., Kristiansen, J., & Hov, Ø. (2020). Adopting citizen observations in operational weather prediction. *Bulletin of the American Meteorological Society*, 101(1), E43–E57. https://doi.org/10.1175/BAMS-D-18-0237.1
- Sgoff, C., Acevedo, W., Paschalidi, Z., Ulbrich, S., Bauernschubert, E., Kratzsch, T., & Potthast, R. (2022). Assimilation of crowd-sourced surface observations over Germany in a regional weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 148(745), 1752– 1767. https://doi.org/10.1002/qj.4276
- Stewart, I. D. (2011). A systematic review and scientific critique of methodology in modern urban heat island literature. *International Journal of Climatology*, 31, 200–217. https://doi.org/10.1002/joc.2141
- Troeyer, K. De, Bauwelinck, M., Aerts, R., Profer, D., Berckmans, J., Delcloo, A., Hamdi, R., Schaeybroeck, B. Van, Hooyberghs, H., Lauwaet, D., Demoury, C., & Nieuwenhuyse, A. Van. (2020). Heat related mortality in the two largest Belgian urban areas : A time series analysis. *Environmental Research*, 188(February), 109848. https://doi.org/10.1016/j.envres.2020.109848
- Venter, Z. S., Chakraborty, T., & Lee, X. (2021). Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms. *Science Advances*, 7(22), 1–9. https://doi.org/https://dx.doi.org/10.1126/sciadv.abb9569
- Verdonck, M. L., Demuzere, M., Hooyberghs, H., Beck, C., Cyrys, J., Schneider, A., Dewulf, R., & Van Coillie, F. (2018). The potential of local climate zones maps as a heat stress assessment tool, supported by simulated air temperature data. *Landscape and Urban Planning*, 178(July 2017), 183–197. https://doi.org/10.1016/j.landurbplan.2018.06.004
- Wang, J., Zhou, W., & Jiao, M. (2022). Location matters: planting urban trees in the right places improves cooling. *Frontiers in Ecology and the Environment*, 20(3), 147–151. https://doi.org/10.1002/fee.2455
- Wang, Q., Wang, X., Zhou, Y., Liu, D., & Wang, H. (2022). The dominant factors and influence of urban characteristics on land surface temperature using random forest algorithm. *Sustainable Cities and Society*, 79(January), 103722. https://doi.org/10.1016/j.scs.2022.103722
- Ziter, C. D., Pedersen, E. J., Kucharik, C. J., & Turner, M. G. (2019). Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. *Proceedings of the National Academy of Sciences*, 116(15), 7575–7580. https://doi.org/10.1073/pnas.1817561116