

Review for manuscript ESSD-2022-113 by Beele et al.: “Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium”

### **General comments:**

The manuscript deals with a data set of urban meteorological observations made by nearly 100 low-cost weather stations in the city of Leuven, Belgium. Further, the authors present a dedicated quality-control (QC) procedure for the air-temperature data. Though the data set comprises several variables (e.g. air temperature, humidity, wind speed, global radiation, precipitation), the QC only addresses the air-temperature measurements. The development and evaluation of the QC is the focus of the manuscript.

The manuscript and data set are of high relevance, as few cities have dedicated urban measurement networks. The data set can thus be of value to many researchers for various applications. The manuscript is mostly well written and easily comprehensible. Quality of figures and tables, and their captions, should be improved (see details below). Number of figures and tables could potentially be reduced by combining them.

Overall, the quality of the research is good. My main remark is that the manuscript should highlight better the value of the data set itself for future research (which I think is high!), and that the developed and applied QC is relatively strictly focused on this type of data set and weather station. For example, one could not apply the QC to another crowdsourced data set of, e.g., Netatmo data (radiation data missing) or use it to quality control single/few stations (L1.3, L2).

My opinion is that the comprehensive data set and the fact that it is quality controlled to a high degree is the key selling point, not the developed QC itself. The fact that such a data set exists and that the network of stations will be maintained for some time (until when?) is great. I do not think that the QC is “easily transferable” to other cities, as you need a lot of measurement data for each station, which is not the case for most crowdsourced data sets. Thus I recommend to restructure the manuscript a bit to give more emphasis on the network (how set up, how maintained, by whom, ...) and the data (how collected, what is in the data set, ...), and a bit less on the QC. The data set itself on the repository should then be enhanced with additional material regarding meta data of the stations, e.g., Gubler et al. 2021 Supplementary Material (<https://doi.org/10.1016/j.uclim.2021.100817>).

I hope my comments are helpful to the authors.

Daniel Fenner

### **Specific comments:**

- Introduction: At some point when crowdsourced data is introduced, the authors should better highlight that this paper mainly deals with air-temperature data, especially regarding the QC.

- Introduction: In what way is the current data set a “crowdsourcing” data set? Please explain better and compare with the definition of crowdsourcing (e.g. Muller et al. 2015). Or is it more citizen science?

- l. 110-111: This is misleading in my opinion. It is true in your case, but a reader could get the impression that the QC could now (as it was developed and evaluated by you) be implemented in another city without other data. This is not true, as in L1.1 you also need data from an official station and since the thresholds were derived from your single network of stations and might need adjustment in other settings.

- l. 151: “stratified sampling”. Please reword, as the number of stations does not represent the percentages (coverage) of the LCZs within the study area. As far as I understand it, the stations were installed to cover a large range of LCZ. But does the number represent the spatial coverage of the LCZ or was it mainly a choice of where a station could be (administration, finding owners, etc.)?

- Table 6: Please make sure that the given parameter names are used in the same way as in the text.

- QC L1.2: Why are the values different for TOathresmax and TOathresmin? Please explain.
- QC L1.2: You give a threshold of 0.05 °C, yet the specified resolution of the sensor is only 0.1 °C. Of course you can obtain more digits due to the averaging, yet this does not reflect the sensor resolution. Additionally, in l. 427 you write 0.1 °C. I suggest to use 0.1 °C.
- QC L1.3: Table 6 and text l. 275: SOthresmin and SOthresmax should be dimensionless, no?
- QC L1.3: With such a low minimum number of stations (nstat = 1) one could falsely remove values that are actually of good quality, if you check each station one by one. Consider three stations, located in a line, each 2 km apart. If the middle station produces false values, it will get flagged out because it deviates too much from the other two (good!). But similarly, checking either one of the stations on the sides will flag that station too, as it deviates too much from the middle station (which actually produces false values) (bad!). Have you investigated using a larger number of neighbours or have you solved this issue somehow differently?
- QC L2: Why was the resampling done to 2 hours? Have you tried different lengths here and determined somehow the optimal duration (if so, how)? Further, are the episodes you find station specific or do they have to overlap across all stations? Please elaborate a bit further.
- Training & validation & RF model: How did you split the data set of LC\_R (e.g., randomly or by time)? Do you also train the RF on each LC\_X station later on (l. 552: “In QC level 3 the random forest model is applied to each station in order to obtain a site-specific prediction ...”)? The outcome of QC L3 for LC\_R is probably not surprising, since you train on the same data set (60 % of it) and then apply the RF model to the whole data set. Is the differentiation between day and night just for illustrative purposes or actually applied in the models? This section needs better description what has been done.
- Figure 4 and text: What is the authors explanation for the fact that largest biases are found in the middle of the radiation range and not at the top? Please elaborate a bit in the text.
- l. 573: You argue that your data set (since only one type of station) is free of certain effects that would require station-specific QC methods. Yet, in QC L3 you do a station-specific QC, no? Why couldn't you apply the QC to another type of weather station that measures the same variables? Please reword the sentence and/or explain better what you mean.
- l. 585: Who controls the stations and does the maintenance? Please explain in more detail.
- Figure 20: I suggest to include the months January – June in the top sub figure, simply to have the same x range as the other two. Please also ensure that the y range is the same in all sub figures.
- Figures and tables in general:
  - (a) Maps should contain coordinate axis.
  - (b) In all figures the subplot labels a, b, c, etc. are missing. This makes it quite difficult to understand some of them and needs to be resolved.
  - (c) In many figures where different sub-figures are present the axes ranges differ. I propose to have common axes ranges per figure. Alternatively, should should at least give a note in the caption that the ranges differ.
  - (d) In all figures with coloured background (e.g. Figure 4): Please explain the colour coding and ensure that this is the same in all sub-figures, at best in all of those figures for comparison. Further, how did you group the data (binsize, hourly averages?, etc.)?
  - (e) The captions need more information regarding the data that is displayed, mainly regarding time period and data sets (this also applies to all tables).
  - (f) Number of figures: Several figures could be combined, e.g. Figures 9, 13, 15; Figures 10, 14, 16. Please check all figures and check (a) if they are all relevant and (b) if they could be combined. The amount of figures (20) and tables (12) is too much in my sense and should be reduced. For figures that basically display

the same but only differ in the QC level, you could also write in the caption: “Same as figure X but for QC level Y”.

- The authors should somewhere better describe the actual data set on the repository (e.g. which variables, which resolution, what kind of meta data, etc.). This would be extremely helpful for other researchers to make use of the data. This section could, e.g., be put before the Conclusions.

- Conclusions: The authors could highlight more clearly the high value of the quality-controlled data set. The large amount of stations in a relatively small city could, e.g., enable detailed comparison with other crowdsourced data where no meta data are available and where the station setup is unknown (e.g. Netatmo). As such, the stations could act as “gatekeepers” for other stations (Chapman et al. 2017). The data set could also be of use to modellers (micro-scale modelling).

- Conclusions: The authors conclude that the method is “easily transferable to other urban regions not having an official weather station”. I do not understand how the authors come to this conclusion when such information is needed in QC L1.1. Also, “easily transferable” only in that sense, that quite a lot of other data are needed for the QC of air-temperature data. That is certainly not the case for the vast amount of citizen weather stations that exist, from which data could be crowdsourced. The authors should state more clearly what is needed for the QC to be transferred to another regions and reword the section.

- Meta data in the actual data set: I noticed that latitude and longitude values are given with only three digits. I highly recommend to give at least four digits for better precision and to subsequently allow more detailed spatial analyses on the micro scale.

- Do the authors have any idea about the quality of the other variables that are measured and used in the quality control? It could make sense to have relatively simple range and persistence tests (as you developed) implemented for all variables before the detailed QC of air temperature. Please comment on that in the text somewhere.

### **Technical corrections:**

- Compound adjective (e.g. “air temperature” but “air-temperature data”). Please check the whole manuscript carefully

- l. 32: consider “... that both cities and their citizens are ...”

- l. 50: “... in which air temperature is continuously modelled ...”

- l. 101: consider “Thus, there is a need for ...”

- l. 104: “Here we report on a ...”

- l. 169: “... is limited due to the ...”

- l. 178: “direct effect contact” ? Please reword.

- l. 182: “aggregated” – Do you mean temporally averaged?

- l. 194 (and other instances in the text, Table A1): There should be a space between the number and the unit, except for geographical degrees; cf. chapter 5.4.3 in <https://www.bipm.org/documents/20126/41483022/SI-Brochure-9.pdf/fcf090b2-04e6-88cc-1149-c3e029ad8232>).

- l. 266-267: Please check sentence and reword.

- l. 285: “non-negligible” – How did you determine that? What is the basis for the statement?

- l. 298: “i.e.” is instead of “e.g.”

- l. 300: Could be helpful to say “... across all ten selected episodes is considered ...”

- Table 7: What is actually shown here? What kind of correlation coefficient? Per definition, a correlation coefficient has no units. Please remove them and specify more clearly in the caption what is shown.

- Figure 8: What is the black line? Please explain in the caption.

- l. 447 & 448: Where is the information regarding the summer day shown? Please specify or reword.

**References:**

- Chapman, L., Bell, C. and Bell, S. (2017): Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. *Int. J. Climatol.* 37 (9): 3597-3605. DOI: 10.1002/joc.4940
- Gubler, M., Christen, A., Remund, J. and Brönnimann, S. (2021): Evaluation and application of a low-cost measurement network to study intra-urban temperature differences during summer 2018 in Bern, Switzerland. *Urban Clim.* 37: 100817. DOI: 10.1016/j.uclim.2021.100817
- Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. R. (2015): Crowdsourcing for climate and atmospheric sciences: current status and future potential. *Int. J. Climatol.* 35 (11): 3185-3203. DOI: 10.1002/joc.4210