Earth Syst. Sci. Data Discuss., referee comment RC2
https://doi.org/10.5194/essd-2022-221-RC2, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on essd-2022-221

Anonymous Referee #2

Referee comment on "OpenMRG: Open data from Microwave links, Radar, and Gauges for rainfall quantification in Gothenburg, Sweden" by Jafet C. M. Andersson et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-221-RC2, 2022

Review of the manuscript „OpenMRG: Open data from Microwave links, Radar, and Gauges for rainfall quantification in Gothenburg, Sweden" submitted to ESSD by Andersson et al.

Summary

This paper presents an unprecedented and important open dataset with data from CMLs, rain gauges and weather radar. The amount of data, its temporal resolution and the fact that the CML coordinates are real, make it very valuable for progressing the development of improved CML rainfall estimation and data fusion methods. The paper is very well written and structured. I have only some smaller comments regarding the manuscript which will require a minor revision. I also checked the data that is available at zenodo. The data is well organised and easy to parse. There are some minor things that could be optimised, though.

General comments

Naming convention of TX and RX: While I have seen the term TX and RX being used for CML data, using TSL and RSL (transmitted signal level) and (received signal level) is much more explicit. Why was TX and RX chosen and would it be possible to change to TSL and RSL?

Rainfall data from CMLs: I agree with reviewer #1 that it would be beneficial to add information on how to process the raw CML data into rain rates. I am, however, undecided whether or not the addition of processed CML rain rates is required or even a good idea. I see the presented dataset as the basis for the validation of existing and new CML processing techniques and not as a tool to use CML rainfall data in an hydrological

application, because (at least in my opinion) expert knowledge on CML processing is still required to interpret certain misbehaviour of CML rain rates caused by e.g. erratic behaviour of the CML data or by wet antenna attenuation. If the authors do not yet have an optimized processing for the presented CML data, it would be too much effort to produce it for adding it here. But, a small section with info regarding CML data processing methods and code, would be beneficial for sure.

Specific comments on dataset

- For some of the city gauges the metadata entry for „type" und „unit" is switched.

- The proj string that is given in the paper is not available as attribute in the NetCDF. To do the reprojection of the coordinates I had to copy-paste it from the paper. That is easy, but all information required to work with the data should be available from the data alone. There is a lot of info on the projection in the NetCDF in the variable „crs", but I could not find something that I can directly use in my code for doing the projection.

- I strongly suggest to add the longitude and latitude grid to the radar data NetCDF. I took me some time and some internet search to do the reprojection from the provided x and y data to lon and lat. CML and gauge data also come with lon-lat coordinates. Hence, for comparison with radar data, one has to generate the lon and lat grids. While this is doable, it is prone to errors and cumbersome (not the actual coding, but finding the solution, at least if you are not doing this on a daily basis). The code that I used in Python is below.

```
import pyproj
x_grid, y_grid = np.meshgrid(x_radar, y_radar)
transformer = pyproj.Transformer.from_crs('+proj=stere +lat_ts=60 +ellps=bessel
+lon_0=14 +lat_0=90', 'EPSG:4326', always_xy=True)
lon_grid, lat_grid = transformer.transform(xx=x_grid, yy=y_grid)
```

The lon_grid and lat_grid can be added to the NetCDF as variable along the (y, x) dimension. The variable with the pseudo-dBZ data should then have an attribute "coordinates": "longitudes latitudes", assuming that the lon and lat grid variable name is „longitudes" and „latitudes". I am not an expert on this, but I think it is also good to add an attribute „grid_mapping": „crs" to the pseudo-dBZ variable so that any NetCDF viewer will correctly plot the data on a map. As far as I can see the already existing variable „crs" has a lot of projection info. I do not know if this info is in line with the CF conventions (cfconventions.org), though.

- The usage of pseudo-dBZ is not required to have small NetCDF file size. NetCDF supports these types of conversion automatically when reading or writing. One can set the

encoding attributes "scale_factor", "add_offset" and specify the "datatype" of the variable (see http://cfconventions.org/cf-conventions/cf-conventions.html#_reduction_of_dataset_size). This is not an urgent issue, but if the radar NetCDF is redone because the lon and lat grids are added, this could also be improved.

Specific comments on manuscript

P2 L15: I can somehow understand the reasoning behind the analogy of the limitation of weather radar and gauges regarding the tradeoff between coverage and resolution. However, the resolution at which weather radar provides rainfall information can be improved, and has already been improved, by more advanced scanning techniques that allow faster sweep times. For rain gauges the only option to increase resolution, in the sense of observability of spatial gradients of rainfall fields, is to place more gauges. „Resolution" is also a vague term here, since it is well defined for radar observations, but not for observations of rain gauge networks. Hence, the authors might want to reconsider this sentence. Please note, that this is more of a subjective comment.

P3 L5: The reference Bao et al 2017 does not fit here. Besides the reference to ITU the authors could cite work from the 70s where the k-R relation has been studied in detail empirically and theoretically

Atlas and Ulbrich, 1977: Path- and Area-Integrated Rainfall Measurement by Microwave Attenuation in the 1–3 cm Band, Journal of Applied Meteorology
Olsen et al., 1978: The aR^b relation in the calculation of rain attenuation, IEEE Transactions on Antennas and Propagation

P3 L20: It should be made clearer here that the data from the Netherlands was not aggregated by the researchers, but rather is provided like this by the network operators from the performance reports that network management systems typically produce.

P5 Fig 3: This is an important figure to define these terms which, up to know, do lack a clear citable definition and have been used inconsistently. Hence, I suggest to improve the clarity of the figure. Some suggestions: The dashed lines of the boxes combined with the choice of colours makes it hard to distinguish the individual relevant parts. Maybe having the tower and antennas drawn in black and then using easily distinguishable colours for the boxes can improve that. Shaded solid coloured boxes might be better than the boxes with dashed lines, or just solid lines of coloured boxes if towers and antennas are in black. The text for „sub-link" should maybe be on the top and not between the two arrows. It would also be good to add a second „link" to make it clearer in the figure why the term „hop" is needed. It is also not clear to me why the „node" and „hop" comprise also the towers.

P5 L14: Morais (2021) is an interesting reference that I was not aware of. But it should be made clearer here what specific part of this sentence it supports or provides more information on. Since this reference is a book, the effort to access the content is higher than for an online paper PDF. Hence, it would be good to be more specific here.

P6 L3: Is the information on antenna size for each CML also available? If yes, would it be possible to add it to the open dataset? This might be important information regarding the investigations of the wet antenna effect.

P7 L12: Is the sampling done at one fixed 10-second interval for all links, or are there individual sampling cycles or slight offsets for each CML?

P7 L14: I assume „synchronising time stamps" means that the raw time stamps from the DC are rounded to (the nearest?) 10-second time stamps. Update: Okay, further down you write that this is done at SHMI. But what is meant with „synchronising time stamps" then?

P13 L15: I would not say that -30 dBZ „should be considered as zero precipitation". For a standard Marshall-Palmer Z-R relation one would get R=0.04 mm/h for 0 dBZ, which I would considered to be the lower end of what we perceive as rainfall (also in the sense that the hydrometeors are falling and are not suspended in the air, which does not make a difference for the radar). One could debate if 5 dBZ or -5 dBZ is a better threshold, but for sure -30 dBZ is too low.

P13 Table 5: I had a quick look at the CML data. What I found is that all sublinks with a minimum RX value below -90 dBm had only garbage data. These sublinks have the IDs 208, 389, 391 and 603. The CMLs with min RX values between -85 dBm and -90 dBm (more than 100 CMLs) seem to look good (note that I only looked at a handful of time series). I am not sure how relevant this is and if this should be stated in the paper, since the dataset is what it is and it is also up to users to investigate all peculiarities. But you might want to consider to add more info on that issue.

P14 Fig 7: Would it be an option to use a log scale on the y-axis to better show the part of the RX data where rainfall occurs?

P19 L3: Looking at the radar images in Fig 13, I do not think that attenuation can cause the missing or too low radar rain rate at the gauge location since there is enough pixels with high rain rates around the gauge pixel. Since there is only one pixel offset between high radar rain rate and the gauge at 15:05, it looks a lot like spatial mismatch due to advection. But, if not yet done, the authors should check the grid definition of the radar grid, i.e. to which corner of the gird pixel (or the centroid) the coordinates refer to. When plotting a quadmesh (like non-equidistant lon-lat grids) the individual pixels are typically

defined by the lower left corner.

P19 L6: In addition to geographic position, also the spatial integration characteristics of the CML might lead to longer lasting but less peaked rain/attenuation events in the time series.

P19 L9: Since the radar data is not capturing the dynamics of the event correctly (assuming that the gauge and CML do) I would not say that the radar time series „is more realistic" than that of the gauge.

Technical corrections

P4 L1: Remove „with"

P22 L26: Remove „the" before „are provided"