

Response to Reviewer Comments

OpenMesh: Wireless Signal Dataset for Opportunistic Urban Weather Sensing in New York City

ESSD manuscript ID: **essd-2025-238**

1 Summary (Review)

The authors have done a major update of the manuscript, the dataset and the accompanying source code. They have mostly followed my recommendations and have provided reasonable explanations when not doing so. The paper is now much more focused on presenting the data instead of doing additional analysis. There are, however, still some smaller issues which require a minor revision. The newly added text is not always of sufficient quality. The restructured and extended dataset is very new. But the accompanying source code on GitHub has some issues which I point out below and which require a minor revision.

Additional note on the documents provided for the revision: Unfortunately the revision is not easy to review because the track-changes document is very inconsistent. It seems, the authors did not use `latexdiff` or similar tools. One hint at that is that there is e.g. “`\blue{QC}`” occurring in the text which indicates that the authors have manually highlighted new text in blue. This is not only cumbersome, but also prone to errors. By comparing with the initial version of the manuscript I found several cases where changes are not highlighted. Removed text is not highlighted at all. In addition, the response letter is only a summary of the main and selected changes and not a point-by-point response. The authors did provide a point-by-point response as “author comment” during the initial discussion phase. But this “author comment” mostly propose changes for the revision. Only in the final response letter it is made clear what was actually done for the revision. I must say, though, that the whole procedure with “author comment” during the initial discussion and then another response letter when submitting the revision is probably a bit unclear to new authors of Copernicus journals. I also see the advantage of having a condensed response letter if large parts of the point-by-point response are obsolete due to restructuring. But, in the current case, with the inconsistent track-changes document and the response letter that only provides a summary I have to switch between several documents (authors comment, response letter, track-changes document, initially submitted document) to get the full picture. I have concentrated my review on the parts that seem new to me, hence, I might have missed some issues introduced elsewhere in the document.

Response: We once again thank the reviewer for the thorough and constructive evaluation, and for acknowledging the improvements made to the manuscript, dataset, and source code. For this revision, we addressed all comments below point-by-point, including text quality improvements, corrections to the manuscript, and a thorough cleanup of the GitHub repository ensuring all pipelines run correctly. As suggested, we used `latexdiff` to generate a proper track-changes document. We also attach separately the first major revision diff for convenience.

2 Comments on GitHub Source Code

The path in the following command in “Option A” in the README is wrong: `jupyter notebook src/fetch_data/download_and_read_openmesh.ipynb`

Response: Corrected.

If I run the notebook, which is at `src/fetch_data/OpenMesh/download_and_read_openmesh.ipynb`, it fails at the imports because `tqdm` is not in the requirements. After adding `tqdm` manually to the venv the notebook works correctly. The NetCDF files with the CML data that is downloaded looks good. Structure and naming conventions are correct.

Response: We added `tqdm` to the requirements. We appreciate the reviewer confirming that the CML data structure and naming conventions are correct.

`OpenMesh.zip` is only 13 MB but in the README on Github it says it is 330 MB. This should be updated.

Response: The README now correctly states both the compressed (13 MB) and extracted (330 MB) sizes.

The repo structure shown in the README is not up to date.

Response: Updated to reflect the current repository structure.

The code in the README for “Fetch Additional Weather Data” does not match the content of the repo. The directory is wrong and there is no `main.py` that can be run.

Response: Corrected. The README now points to the proper directory and `main.py` path.

There is additional info in the file `DATA_ORGANIZATION.md` but it not clear if, how and when the user should look at it. The repo structure in this file looks outdated. The info below the repo structure is unclear.

Response: We revised `DATA_ORGANIZATION.md` to reflect the current structure and clarified its content.

The notebook `asos_pipeline.ipynb` works correctly. But it is not mentioned in the README. I did not check other options regarding NOAA data.

Response: Added a reference to `asos_pipeline.ipynb` in the README.

The notebook `wu_pipeline.ipynb` fails in the first cell at the import of `from src.fetch_data` because it does not find such a module. It looks as if the code for adding the project root to the path did not work correctly. I did not debug this further since I do not have the API key for WU data and thus would not be able to run the full notebook anyway. But this should be fixed.

Response: Fixed the import path issue in `wu_pipeline.ipynb`.

The notebook `src/analysis/rain_analysis_example.ipynb` does not work as expected. The function `load_openmesh_cml()` expects the file `ds_openmesh.nc` in `dataset/links` but the notebook `download_and_read_openmesh.ipynb` which I did run to get the data, places the data in `src/src/data/openmesh/extracted/dataset/links`. If I manually move the data to the expected location the notebook works. The same is true for the PWS data. It seems that the download notebook has to be corrected. There were some other parts of the rain analysis notebook that might not have worked correctly, e.g. it said “PWS DataFrame not available”. But this did not throw an error and all the plots generated in the notebook seemed correct.

Response: We fixed the path inconsistency between the download and analysis notebooks so that data is saved and loaded from a consistent location. The PWS loading issue was also resolved.

In summary, the authors have to clean up their source code repository and make sure that the suggested pipelines do run correctly. I suggest to also simplify things. [...] As written above, I suggest to condense the info.

Response: We thoroughly cleaned up the repository following all suggestions above: simplified the directory structure, ensured all pipelines run correctly end-to-end, consolidated path information to avoid redundancy, and condensed the README files throughout.

3 Specific and editorial comments

Note that line numbers are the ones from the track-changes document.

L8: “dense” is unclear here. Maybe it would be better to state the number of PWS that are included. Since the domain size is already roughly known in the abstract that gives a good estimate of the station density.

Response: We specified the number and location of PWS. The text now reads: “1) 37 Weather Underground (WUnderground) personal weather stations (PWS)...”

L28: I do not understand what is meant with “Early OS forecasting applications include PWS networks providing hyperlocal observations...”. It sounds as if OS data is used for forecasting. But, I assume the OS data is assimilated into a NWP system which to improve the NWP-based forecast. This sounds misleading, or at least unclear, here. Also, “hyperlocal” is a term that does not make sense here. A point observation is of course “hyperlocal”. This sounds like a term from an advertisement. I suggest to remove it here and the other occurrences in the text.

Response: We revised the sentence to clarify that OS data is utilized in forecasting workflows rather than being an application itself, and removed “hyperlocal” throughout. The text now reads: “Early utilization of OS includes PWS networks providing dense observations...”

L31-33: “Support data assimilation” should maybe be “can be used for”. The major operational NWP system are not only “driven by homogeneous gridded products”. “and potentially inform future deep-learning methods” is vague.

Response: We revised this paragraph accordingly: adopted “can be used for,” corrected the NWP description to acknowledge diverse observational inputs, and specified that OS serves as a “*supplementary high-resolution input*” for deep learning.

L57, L60: “. . . and its relevance with background. . .” unclear what this part of the sentence should mean. Why is section 6 not mentioned here? Either add it here or remove the info on the very short and obvious sections here.

Response: We clarified the phrasing and added the missing sections as requested for the manuscript organization.

L85: “At higher bands [meaning V-Band and E-Band here], propagation is range-limited and blockage-sensitive”. Propagation of em-waves is always range-limited in some sense. But what you mean here, that range has a strong impact on the link budget is also true for lower frequencies of CMLs. Similarly for blockage which affects propagation already in the lowest GHz range.

Response: We refined the text to state that high-band propagation is “*significantly more sensitive to path loss and weather phenomena,*” favoring urban deployments that “*enable dense spatial coverage for environmental sensing.*”

L108: Write “increased sensitivity” instead of “direct sensitivity”.

L111: Maybe find a better formulation than “. . . are exploratory”.

Response: Updated to “increased sensitivity” (L108) and replaced “exploratory” with “emerging applications” (L111).

L146–147: Write “selected” and clarify what local services are “hosted by community members”.

Response: Updated to “selected” and replaced the general mention with specific examples of community-hosted infrastructure, including data repositories and network management systems.

L153-L155: These three types of nodes are not consistent with the naming conventions and the structure in Fig 3. I suggest to update the figure, which is not very clear anyway.

Response: We revised the text and updated Figure 3 to ensure consistent terminology and reflect the NYC Mesh tiered topology.

L182: What does the special font type of “updated” mean?

Response: It referred to a JSON field name. We removed the special formatting and clarified this in the text.

L196: Why are there unidirectional hops? What purpose do they serve in this communication network? Or are they just a results of removing the sublink in the opposite direction because its data was too messy?

Response: Physical network connections are bidirectional. However, the dataset contains unidirectional hops for two reasons: (1) API access for client-side logging can be restricted for privacy and therefore not stored in our database, and (2) as the reviewer suggests, opposing sublinks are sometimes filtered out due to insufficient data quality.

L198: "...results in 4 hops..." unclear what this part of the sentence means.

Response: We clarified the sentence. Some hops combine bidirectional and dual-band configurations operating under the same link ID, resulting in more than two sublinks. We removed the specific count (4) to improve readability, with full topology details are documented in the repository.

L205: What does "feeds" mean here?

Response: We replaced "feeds" with "data streams" for clarity.

L207: "We also leveraged NYC Mesh's unique features such as multi-band environments, two frequencies directional pairs (dual-band), and diversity in lengths and spectrum". Not sure what a "multi-band environment" is. Also, dual-band links are a common technology nowadays. But, yes, this is the first open dataset with such dual-band links. Also, "diversity of lengths and spectrum" is given in other open CML datasets.

Response: We removed "multi-band environment" and revised the text to focus on our selection criteria, prioritizing V-band frequencies and directional dual-band pairs. We now emphasize that these configurations provide an additional data layer complementing existing open CML datasets, rather than claiming uniqueness in diversity of lengths and spectrum.

L229: "...even well-maintained... ...provide only limited spatial coverage" this relation between well-maintained and the coverage does not make sense, please rephrase.

Response: Rephrased to focus on the sparse spatial distribution of official stations.

L230: "However, official stations have a sparse spatial distribution,..." This is an example for text that I could not find in the old manuscript but it is not marked in the track-changes document. I actually looked a bit more in detail here because it kind of doubles with the sentence before and I wanted to check what the history of this section was. I have to assume there are more of these non-marked changes...

Response: We removed the repetition. All modifications in this revision are properly tracked using `latexdiff`.

L288: Sentence two times has “data where obtained...”

Response: Fixed.

L334: Something seems wrong with this part of the sentence “that was selected sensors”.

Response: Removed this sentence as it was unnecessary for the data description.

L413: Add point after “freezing”.

L441: I suggest to remove “controlled” because this is not a lab-setup or a dedicated field experiment.

Response: Both adopted.

L449: “NEXRAD KOKX and MRMS” needs more explanation, terms have not been introduced.

Response: We expanded the acronyms and added brief descriptions.

L467: Write “such as smart cities” instead of “such smart cities”.

Response: Fixed.

L467: This is, by the way, another example where text was updated/extended/changed but not highlighted in the track-changes document. This makes reviewing really unpleasant.

Response: Please see our response to the main comment regarding tracked changes.

4 Summary of Changes

4.1 Code Repository

The repository was cleaned up and all pipelines verified to run correctly end-to-end. The README and documentation were condensed and updated to reflect the current structure. Beyond the manuscript dataset, the repository supports fetching raw data from ASOS and Weather Underground for custom time intervals. To support further research, we uploaded the full PWS observation period in NetCDF format to a separate Zenodo record, approved by Weather Underground for non-commercial use, accessible directly through the repository. While the WUnderground API provides the most flexible access to hourly data (a guide for obtaining an API key is included), the Zenodo archive offers a convenient ready-to-use alternative with original, non-processed records. Note that the repository is under active development, with ongoing

work to integrate shared community libraries and optimize the codebase.

4.2 Manuscript

We revised the text throughout to correct typos, improve clarity, and ensure consistency, addressing all reviewer comments point-by-point. A `latexdiff` track-changes document is provided for this revision. The first major revision diff, which contains substantial changes from the original submission, is attached separately.