Review of the paper
"Feasibility and limits of commercial microwave links for opportunistic air-temperature monitoring" , by P. Musil et al..
https://doi.org/10.5194/egusphere-2026-3888

This manuscript presents an analysis of temperature data collected by the internal sensors located at the endpoints of commercial microwave links (CMLs) and assesses whether these data can provide information about the air temperature. To this aim, the authors use physically-based models, regressions and machine learning methods. Overall, I think the results are of interest not only for meteorological purposes but also for a wide range of applications such as energy efficiency of buildings, optimization of solar panels deployment, and last but not least, CML-based studies of precipitation.

However, I believe that several aspects of the manuscript require further improvement before it can be considered suitable for publication. In my assessment, a major revision is needed for the reasons outlined below.

General comments:

- Paper organization. This is the major point. I think the paper is too long, has too many subsections and it is not well organized. Moreover, the same concepts are repeated a few times. It's a shame because the authors did a good job in analysing their dataset. My suggestions:
    - Sec. 2 has five subsections, Sec. 3 has seven: some of them are one-two paragraphs (3.1, 3,3, 3.4. 3,5), while Sec. 3.7 is 10 pages long ! This is a sign of bad paper organization.
    - Eqns. (1)-(3): are a special case of Eqn. (4). I would first introduce the general equation and eventually derive the simple expression that holds in the case of the in-lab experiment. I am not even sure that Eqns. (1)-(3) are really necessary. It would be enough to state that the experimental curves in Fig. 2 are well approximated by a first-order dynamic system (i.e. described by a first-order differential equation).
    - The in-lab experiment in Sec. 2.2 is a result. Can't see why it has been placed in Sec. 2 "Material and methods". Indeed it is basically described again in Sec. 3.1 of the Results (the "controlled experiment"). This experiment, in particular Fig. 2, justifies Eqn.(4) at least in the special case of the in-lab experiment, as the model curve fits well the experimental curves. I warmly suggest moving Sec. 2.2 in the "Results" section and merge it with Sec. 3.1
    - I suggest grouping Figures 6-9 into a four-panel figure for readability and to save space.
    - Sec. 3.3 does not contain any result. It can be removed. I suggest moving Figure 10 to Sec. 2. You can add it to Figure 5, obtaining a two-panel figure showing the CML deployment with respect to the weather station and the sources of heat. It would be a nice and complete overview of the investigated on-field scenario.
    - Section 3.4: this is another example of factors influencing the thermal behaviour of the ML unit. It makes sense to add it to Sec. 3.2, while deleting the short Sec. 3.4.

    - Section 3.5: another very short sub-section, reporting an example of the effect of maintenance and system resets on CMI temperature. It should be integrated into Sec. 3.2
    - Table 6 can be merged with Table 5 as they refer to the same set of data. Moreover, is it really necessary to show the three correlation coefficients in Table 6 ? The authors only comment on Pearson coefficient on lines 282-284.
- There are really many improper/inaccurate expressions throughout the paper. Examples: “cyclic diurnal and seasonal features”, “four aspects of feasibility”, “distributional similarity”, “blocked regression models”, “exploratory way”, “controlled experiment”, “central interpretation”, etc. Moreover, the word “proxy” and variants are compulsively used. Same for the prefix “per-”. Please carefully check the language.
- Sec. 3.7 deserves a discussion of its own, as I think it is a central part of the manuscript. This section shows the results of regression and ML algorithms. However, it is really hard to follow. It is written more like a technical report for ML experts than a scientific paper. I think this section has not been carefully revised by the authors before submission and contains some basic errors such as introducing figures in the wrong order, duplicated information and, in general, a poor organization of the text. It took me a lot of time to review it. Frankly, this stuff shouldn’t be the reviewer's work. The authors should address the following issues:
    - “The regression dataset is not intended to extend the single-site physical case study directly; rather, it provides an independent operational test whether the telemetry–reference relationship remains useful under blocked temporal and spatial validation”. Unclear statement. Please state clearly which is the scope of the regression analysis. The word “blocked” is confusing.
    - Lines: 291-295: the authors explain the format of the dataset they used as input to the regression analysis. It would be useful to explain better: 1) which is the format of the raw data logged by the MUs, and 2) which kind of pre-processing/data-reduction the authors did to get what they called “monthly export”
    - Line 300: “releasable feature matrix” is it the same as (or similar to) “monthly export” previously used ? If so, please use the same words for clarity.
    - Eqn, (13): can the authors write the full regression equation including all the features they used? Writing only one piece of information does not help at all.
    - Lines 306-311. “Numerical predictors were median-imputed and standardized using parameters estimated from the training part of each fold only. Categorical predictors were imputed from the training fold and one-hot encoded with unseen held-out categories ignored during prediction. This fold-wise preprocessing is important because both roof identity and device identity can otherwise leak information from the test block into the fitted model; in the leave-one-roof protocol, the held-out roof is therefore never available as a usable categorical level at prediction time.” This paragraph is really not intelligible. This is not a ML journal, it’s a journal focused on atmospheric measurements. A few points: a) “predictors” are, stated in a simple language, the inputs to the regression model; b) “Median-imputed” refers to an algorithm for replacing missing data, which is not obvious to somebody who is not so familiar with ML; c) “fold”, that is, “subset” for the non-expert reader, is used for the first time on line 307 without any explanation about the rationale used for dividing the dataset into subsets.

This is explained later; d) “categorical-predictor”, “one-hot encoded”, ”held-out categories” again, is ML slang. Several readers won’t easily understand..

- Line 316: Would write “The two temporal protocols” for clarity.
- Line 317: “The (two) temporal protocols should therefore be interpreted as extrapolation in time for previously observed endpoints”. Strictly, it is correct only for “last-month hold out”, while “leave-one month out” also uses data in the future, unless one assumes that all years are the same, which is not true..
- Eqn. (14): not clearly written I would prefer something like $\hat{t}_{ws} = t_{MU} - C$ where $C = median(t_{MU} - t_{ws})$ with t explanation in the text. Also, is it really necessary to write Eqn (15), which is the same as (14)? The difference can be explained in the text.
- Line 326: sometimes the authors use the expression “pooled experiment”, while sometimes they use “pooled dataset” or “pooled multi-technology dataset”. Please use the same expression for clarity.
- Line 326: “In the pooled experiment this correction uses a single fold-specific scalar estimated from all training samples rather than a separate offset for each endpoint”. It is badly stated, while it becomes clear in Table 7. The authors calculated the scalar constant of the first baseline for each one of the three validation protocols (last-month hold out, leave-one month out, and leave-one-roof out). This is done for all the tested algorithms (Table 7), therefore a simple statement, on line 321 stating that all ML algorithms were tested over the above three protocols would help.
- Line 328: “clock-only ridge regression”, again machine learning slang. So it’s not clear at all how the second baseline has been obtained. Which are the inputs(predictors) ? I guess they are the ones of Eqn. (13). If this is the case, please explicitly mention this equation.
- Line 332: “scikit-learn implementation”, again a cryptic expression referring to a detail, which is not really essential in my view.
- Line 335-336: “To keep the experiment reproducible at paper scale, the fitted training set for each supervised model was capped by stratified sampling over month–roof groups”, is vaguely intelligible. Is it referring to the previous procedure of inner cross-validation or not? Please use simple words, for instance: “To reduce the computational effort, the training set of each supervised model was limited using stratified sampling, which preserves proportion …”
- line 340-341 “This baseline is intentionally not defined for”. This looks obvious. It can be removed.
- “Lined 342-343: if the authors reported the maximum number of training samples for this case, they should do the same on line 336.
- Table 7 and 8: these important tables are barely commented on in the text. Why? The short comment to Table 7 (lines 344-346) is obvious and already pointed out before. The same for the short comment to Table 8 (lines 348-349) Please comment and justify the difference in performance among methods. Then, going on reading the paper, I saw that Table 7 is commented on later (pp. 21-22). This is really confusing for the reader. Please comment on Table 7 when it is introduced (p. 19).

- Table 7 caption reports a number of inessential information- The caption should only describe the content of the table, which are KPIs for several regression and ML methods and three different validation protocols. Moreover, units should be reported after the numerical values for readability (this holds also for the following tables): the reader should be able to interpret the numbers without reading the caption. Finally, the column “Folds” is not really essential in my view (same for Table 8).
- Table 8 the first part (Scope = ‘all’) just C&P the results of table 7, therefore it is useless. I suggest adding the original part of Table 8 in the end of Table 7, adding “All technologies” and "Technology A” as subtitles of the two parts.
- Line 352: “fully identified thermal model” is awkward and does not add any information. It can be removed, adding a full stop after “structure”.
- Line 352 I guess “reduced-order model” is the same as “regression structure”. Again, please use the same words to indicate the same thing..
- Line 353: the “inter-technology spread” is not evident at this point. it is explained in Table 9, which is introduced later.
- Line 353 “the quasi-steady floor” is the same as the “self-heating floor” on Line 276? If so, please use the same words.
- Table 9: please write the units (°C) after the numerical values.
- Line 358: “in the 2025 regression matrix”, refers to regression over Dataset 2. Please use “Dataset 2” instead of “2025” for clarity. Otherwise the reader needs to go back to Table 2 and check.
- Lines 365-367: unclear in several parts: a) “We do not fit a dedicated per-technology baseline here”. So the “median offset” reported in Table 9 is not the parameter of the constant offset regression model for technology A? So what is it?; b) “this reading identifies" which reading are the authors referring to?; c) the “global offset, which is clearly insufficient (Table 7)”, do the authors refer to the performance of the ”constant offset” regression?. I think the overall statement should be written with different words and made more clear.
- Line 368: “assess” instead of “expose”
- Line 368: “temporal validation” is awkward
- Line 370 “Offset” or “MAE?”
- Table 10: please report the same name of the algorithms as in Tables 7 and 8 for clarity.
- Table 10: Are these results meaningful? It is expected that the leave-one month out performance should bring rather homogeneous validation performance throughout the months in a continental climate with well defined and regular seasonal cycles. For instance, leaving out one winter month does not significantly change the proportion of summer and winter months in the training set as summer and winter are basically 50% of time each. Please evaluate whether it is the case to remove Table 10.
- Line 377-378 “the telemetry and auxiliary predictors contain information beyond a static per-device correction, *especially for within-month variability*” How did the authors come to this conclusion from Table 7?
- Figs 15-17 are introduced before Fig. 14.
- Line 402: Please specify that Figure 15 shows density scatterplots, while Figure 16 shows a standard scatterplot.

    - Figure 15: please indicate the three models below or above each subfigure (according to the journal template) as: a) Constant-offset, b) linear SVR, and c) XGBoost.
    - Figs 15-17 are not commented at all. Either the authors comment on it or it's better to remove them. I think they should be retained and commented on. For instance, Fig. 15 shows some interesting differences among the 3 methods. For instance, why is the cloud on the rightmost subplot rotated with respect to the 45° line, while the central one is not?
    - Figure 18 and Figure 19 report the same numbers as in Table 7 and Table 8, right? If my understanding is correct, they do not add any information and can be easily removed.
    - Figure 14 should be moved ahead on p. 21 when the authors mention the reduce-order physically based model.
    - Lines 417-421: suddenly the authors step back to Table 7 commenting on the results of "leave-one-roof". This has already been done before (lines from 382 on). Indeed the comments are the same! This is definitely annoying.
    - Line 421-428: these comments have been partially addressed when commenting on Table 7. Indeed part of the content of Table 11 is partially already in previous Table 7. I suggest removing Table 11 and just mention the differences in performance of individual roofs.
    - Lines 435-437: conclusions are unclear. The authors did a regression analysis using different validation protocols and methods. The purpose of regression is clear. However, the authors state that their analysis, which in my view has been rigorous and well done (left aside poor reporting), should be treated as "a supporting feasibility result". They found that generalization to unseen devices of unseen roofs is a problem, but that was expected given the multiple intervening factors. I think that the author's conclusions are not the best way to give value to their work.. I would delete the last statement.
- Local temperature measurements at dense spatial and temporal resolution have many applications. However, as 1) the authors propose temperature measurements based on CML thermal sensors, 2) CMLs have been extensively investigated as non-conventional rainfall sensors, and 3) several authors reported that the received signal level, which is the input to rainfall intensity estimators, is significantly correlated with the temperature, which may result in false positives (i.e. rain during dry weather), then, these measurements can be very useful to assess whether signal variations are due to the temperature of CML devices. Please add this important aspect/application somewhere with references. The authors just wrote a very short statement about this in the Discussion (lines 449-450).
- Given the many different reasons for the offset between MU temperature and air temperature, it would make sense to provide guidelines about tests/preliminary on-site checks that should be performed to discard or flag potentially problematic MUs in advance, to help the regression/ML algorithms work better. This piece of information may complete the Discussions section.

Comments on specific sections:

Section 1: Introduction:

- Line 22: “data transport and backhaul”. As backhauling links transport data, what’s the difference?
- Line 25: “and related metadata” instead of “or related metadata”. This is not a detail but a fundamental need. Rainfall intensity can be extracted by CML raw data only if the corresponding metadata are provided.
- Line 27: the authors mention “E-band links” assuming the reader is familiar with CML operational frequencies, which I don’t think is always the case. Please mention the typical range of CML frequencies useful for rainfall estimation, moving lines 77-78 here. Moreover, specify the E-band frequency range here, that is, where the term E-band is introduced for the first time.
- Line 29: the authors cite a paper on Satellite Microwave Links (Gianoglio et al., 2023). Indeed there are several SML papers in the list of references. However, this paper is based on the characteristics of CML equipment. I think this is misleading. If the authors think that references to SML papers are necessary, it’s ok. However, they should mention SMLs and their similarities and differences with CMLs..
- Line 29: write “than the free air-temperature”
- Paragraph starting from line 49: a well-known drawback of CML data is ownership. These data are generally not publicly available. As far as I know, the sample of temperature data provided by the authors on the zenodo repository is a unicum. This important aspect must be mentioned.

Section 2: Materials and methods

- Do the authors have info about the accuracy of temperature measurements performed by the sensors within microwave units?
- Line: 75 Add an introductory statement explaining how this section is organized.
- Line 77-78: better to move in the intro as are general information about CML
- Table 1, Eqn. (1)-(3) and all pertinent points in the paper: please use different symbols for time and temperature (e.g. t and T, respectively)
- Eqn. (1): please mention units for the quantities involved
- Line 188: “environment” looks better than “surroundings”
- Lines 118.-119: already stated a few lines above (line 106)
- Lines 123-124: unclear statement and unnecessary in my view.
- Figure 2: There is no evidence of saturation, especially for the red curve, probably the experiment should have been conductor for a longer time. Please: 1) change “time” into something “Time since power-on” or equivalent, and 2) replace “Approx” with “1st order model” in the legend, 3) in the caption, “first-order fits” looks coarse, would state “fits with a first-order dynamic model”.
- Figure 3, caption: the last statement is unimportant. It should be deleted.
- Line 131: “CML distributor”. Do you mean vendor?
- Line 132 better to specify ”The main dataset used in this study”
- Line 137: “reflections” is unclear. If I got it, it would be better to state “exposed to direct or indirect solar radiation”
- Lines 140-142, it’s badly stated. I suggest changing into; “In addition to the five-unit rooftop dataset, we consider a comparison between opposite ends of the same link, denoted as A and B. The A-units are installed on a roof near heat exchangers, whereas the B-units are not exposed to these exchangers”.
- Line 143-145: not here. This is a general description of the datasets and should be written before detailing the datasets.

- Line 150: 327 m instead of 327 km I guess.
- Section 2,.4: the first statement is a repetition of already stated concepts.The first statement of a section/subsection should describe objectives and/or content of the section/subsection, when deemed appropriate.
- Eqn. (4): the authors assume that the effect of all the sources of heat appearing in the right member has the same impact on temperature variations of the MU (through the thermal capacitance), despite these sources being of very different type. I guess local heating generated by the HW (PTH) is concentrated in certain areas of the unit while solar radiation affects the entire unit. Moreover, defining the thermal capacitance of a complex device is not trivial. I understand that Eqn. (4) is a simplified model. However, the authors should better clarify the assumptions behind. it. In addition, please provide units. The authors define PHS as “heat” and qsw as “radiative forcing”. I think all these terms represent powers.
- Figure 5, caption: the figure shows the “main local factors” affecting the CML rooftop temperature. Then the authors enumerate such local factors, including “solar loading”, which I think is not correct.
- Line 170: “variations” instead of “tendency
- Lines 186-89: unclear statement. The authors anticipate the operation of supervised models. They never talk about it. Please just state that this equation will provide a physical base for the ML learning models in a later section (if I got it). Also “cyclic diurnal and seasonal features” is awkward, better “features capturing diurnal and seasonal cycles” . What do you mean by “daylight indicator” ?
- Section 2.5 should be improved in my view. Here the authors explain the analysis conducted on the datasets and enumerate the KPIs used.
    - The four types of investigations listed on lines 191-195 should be better explained. Please state which is the objective of each type of analysis using an appropriate scientific language. Expressions such as “distributional similarity” or “temporal agreement” (do you mean analyze whether the two sources show similar temporal patterns or variations?) or “blocked regression models” are unclear. The statement “situations that reveal physical controls on the offset” is awkward and harshly understandable,
    - Line 191: What are “four aspects of feasibility” ? Simply state that four types of analysis are conducted and enumerate them.
    - Lines 200-207: most of the paragraph sounds like a repetition of what has been stated in commenting Eqn. (4). Moreover, if ML analysis is secondary, why add it to the paper? The paper should only contain key results. Additional or non-essential stuff should be included in the supplementary material. And the last statement of the paragraph does really not help in understanding the authors’ intentions.Finally, the ML stuff fills 10 pages (Sec. 3.7) ! So it’s not really secondary.
    - The KPIs in Eqns. (11-12) are different from the ones listed in the previous page (Pearson, Kendall, and Spearman correlations). Please list out only the KPIs used in the analysis.
- Line 194: “seasonal effects” rather than “different seasons”
- Eqn. (11)-(12) the KPIs defined here do not .

Section 3: Results

- Line 213: “small hardware revision difference”, is unclear. There are two different HW versions or what?
- “Line 222: “temporal pattern” rather than “large-scale temperature evolution”
- Line 223: “the CML values remain shifted” is understandable but, once again, awkward.
- Line 223. “short deviations”. If you mean “short-time fluctuations”, I see them in both time series of Figure 6. Please explain.
- Line 224: “oriented in different directions” rather than “facing different directions”.
- Figure 6: the weather station temperature peaks at the same time as MU 2. Any explanation?
- Line 225: “becomes even more evident” rather than “is even clearer"
- Line 231: “during summer and winter days.” is not accurate as it suggests that many days are considered. Would rather say, “during a typical summer day and a typical winter day.”
- Line 233: “makes the CML temperature more nonlinear with respect to the reference” is poorly stated. The two time series show a weak correlation.
- Figure 9: The units for rain are wrong: if it is (as I guess) rainfall intensity, units are mm/h. Otherwise, please explain.
- Figure 11, caption “over one-year” better than “year.-long” and “at the opposite ends” rather than of “opposite ends”
- Sec. 3.5: the authors should provide an estimate of the fraction of time when CML temperature is not available or not stable. I guess it’s a very low percentage, which represents a plus, when compared to the outage time of an operational weather station.
- As far as I got all figures commented in Sec. 3.1-3.6, refer to Dataset 1 of table 2, that is, to the MUs deployed over the same rooftop. Better to remind the reader about it from time to time or when they show the first example.
- Line 270: the statement starting with “We report…” is useless and in my view it confuses the reader.
- The title of Sec. 3.7 should be improved including, for instance, the term “Machine-learning”. The current title “Temperature estimation” is too generic
- Section 3.7: see above in my “General comments”.

Section 4: Discussion

- The organization of this section can be improved. I suggest the authors summarize the advantages and drawbacks of the CML-based approach in two separate paragraphs Now, they are scattered throughout the section. In particular, the drawbacks depend partly on the measurement devices themselves and partly on the limitation of the experiment they carried out as well as on the available dataset. Please highlight this difference.
- One important drawback of the CML-based approach not mentioned by the authors is data ownership. These data are generated by networks owned by mobile operators, that is private companies. Hence, these data would be publicly available only upon agreement with the owners.
- Line 451-460: the authors state that the precision of their method is not enough for measuring the temperature field in the city environment. That’s true, but this is just a possible application. The method may be more useful for other applications. Please elaborate further.

- Line 463: “A/B comparison” is unclear
- Line 491: Is “They” referring to caveats? it’s not really clear
- Line 496: “main contributions” to what? Please add the missing part. It’s written in the caption of Table 12. However, it should be written also in the text for clarity.
- Table 12: would prefer “(vertical and horizontal separation)” in the fourth line below the header instead of numbers given for the horizontal separation only.
- Line 503: would just write “offset”
- Line 504 “In the present data this remaining gain is about 0.27 ◦C for leave-one-month validation”. Unclear. I guess the authors refer to the statement on lines 432-434. Please mention that it is MAE gain, that it is obtained by Linear SVR and add reference to Table 7.
- Lines 517-523: I would simply state that device resets and restarts produce temperature transients of a duration estimated by the authors through the in-lab experiment and add it to the drawbacks of the CML-based approach. The impact of this shortcoming depends on the frequency of such events, which, I guess, is rare, as these devices should operate with a five-nine availability or similar.
- Line 524: “Finally, the scope of the evidence should be stated clearly”
- Lines 524-529: partially stated a few lines above (511-516). Please avoid repetitions.
- Lines 530-544: again, the same concepts are repeated a few times

Section 5: Conclusions

- They are well written. No comments here.

References

- No comments