

“MeteoSaver v1.0: a machine-learning based software for the transcription of historical weather data” by Derrick et al.

<https://doi.org/10.5194/egusphere-2024-3779>

Preprint. Discussion started: 10 June 2025

General Assessment

This manuscript presents MeteoSaver v1.0, an open-source, machine-learning based pipeline for the transcription, quality control, and structuring of historical meteorological records. The work is technically strong, well motivated, and relevant for climate data rescue, particularly in data-scarce regions.

The system is a valuable contribution to the field of historical climate data rescue, and the open-source, modular design is commendable. The paper represents a valuable contribution at the interface of climate science, machine learning, and data engineering. To strengthen the manuscript, I recommend expanding or more clearly contextualizing the validation, clarifying accuracy requirements for climate applications, and addressing potential biases introduced by rule-based QC. Addressing these points will significantly enhance the practical usefulness of the software.

The manuscript reports several performance metrics (e.g., transcription match rates, MAE, quality flags). The reported median match rate of approximately 74% between MeteoSaver outputs and manual transcription is relatively low for climate data rescue applications, where accuracy requirements are often stringent. While the authors also report a median MAE of 0.3 °C for temperature, the relationship between these metrics and their implications for downstream climate analyses is not sufficiently discussed. The paper should sufficiently explain:

- What level of transcription error is acceptable for climate or meteorological analysis,
- How the reported errors might affect climatologies, trend analyses, or extreme-event detection,
- Whether the current performance is adequate for the intended use cases.

The authors should clearly link their validation results to the types of climate analyses for which MeteoSaver outputs are (or are not) suitable, and explicitly discuss limitations.

I recommend “minor revision”.

Below are few Minor Comments

- The resolution of figures 13 & 14 should be improved. Some validation figures would benefit from clearer guidance on how they should be interpreted by climate scientists.
- The discussion would benefit from a short paragraph situating MeteoSaver within the broader ecosystem of data rescue tools.
- The evaluation focuses primarily on temperature variables. Please clarify whether MeteoSaver currently supports other common variables (precipitation, humidity, pressure, radiation) and whether different QC logic would be required.
- The reported processing time of approximately 8 minutes per sheet on a laptop raises questions. Can you provide estimates for batch processing on HPC or

cloud infrastructure and discuss expected performance for thousands of sheets and potential bottlenecks.

- Either extend the validation to at least one additional archive (different country or table layout), or explicitly acknowledge and discuss the limitations of the current validation, clearly stating that the reported performance metrics may not generalize to other historical datasets.

p.14: “Following the transcription of the data, quality assessment and quality control (QA/QC) is carried out to ensure the final output data is highly accurate with reference to the original handwritten daily temperature records (see Fig. 9).”

>> The phrase “*highly accurate*” is not operationally defined. It would strengthen the methodology to clarify whether “accuracy” here refers to:

- internal consistency (logical constraints, totals, relationships),
- agreement with manually transcribed values,
- or conformity with physical bounds.

Clarifying this will help readers understand what the QA/QC module is designed to guarantee.

p.16:

“If this condition is not met, a specific adjustment, unique to our sheets, is applied: the first digit is removed from the value, and the cell is flagged to indicate this manipulation (see Fig. 11 a-b, with manipulated values in b shown in orange).”

>> This is a data transformation rule, not only a quality check. It would help to explicitly describe this as a *correction operation* and to specify its assumptions (e.g., why the first digit is assumed to be erroneous, and under what conditions this may fail).

“However, if the check is passed, the transcribed temperature values are then adjusted to match the required decimal places, set to one in this case (see Fig. 11 b–c).”

>> This step modifies the data but is not mathematically described. Please clarify:

- whether this is rounding, truncation, or scaling,
- and how uncertainty introduced by this step is handled.

“For the daily maximum temperature threshold, we use 40°C. For the daily minimum temperature threshold, we use 5°C.”

>> The manuscript would benefit from a brief discussion of how sensitive the results are to these fixed thresholds, and whether they are intended to be region-specific or globally applicable.

p.19:

“Only the confirmed (green) daily temperature values are passed to the next module, Data Formatting and Upload (sect. 3.6).”

>> This implies that a large portion of transcribed data may be excluded. Please indicate the proportion of discarded values and discuss potential impacts on time series completeness. Here the manuscript transitions from *checking* to *correcting*. Explicitly distinguishing these two roles would improve conceptual clarity.

“Two examples ... illustrate the sequence of QA/QC checks performed on the initial transcribed values, leading to the final confirmed values (flagged in green).”

>> Figure 11 shows the propagation of flags and value states, but the underlying equations and replacement rules are not visible in the figure. Consider annotating the panels with the rule names (threshold, digit removal, Eq. 1-4, etc.) to make the logic traceable.

p.20:

“At this stage, an additional check is performed, which was not included in the QA/QC module due to the availability of longer temperature series at this point.”

>> This introduces a new methodological step after the main QA/QC description. For structural clarity, it may be preferable to describe this earlier as an optional extension of Module 5.