Response to Reviewer 3's comments:

The manuscript titled "From Paper to Proof: Revealing Congo Basin Warming Through Rescued Climate Archives" by Derrick Muheki et al. presents a valuable community resource by rescuing and digitizing historical meteorological observations from the Democratic Republic of the Congo (DRC). The development of a QA/QC workflow and the creation of a long-term observational dataset with more than 1 million observations from 37 stations represent an important contribution to climate research in a data-sparse region.

The manuscript is well written, and the workflow and results are clearly presented. The comments below are primarily intended to improve methodological clarity and reproducibility.

1. The dataset combines observations collected over several decades across multiple weather stations. During this period, instrumentation, measurement precision, calibration procedures, and observational practices may have changed. It would be helpful if the authors could comment on whether such temporal and interstation inconsistencies were assessed or accounted for (e. g., through homogenization or other quality-control procedures), and discuss their potential influence on the derived long-term climate trends.

2. The QA/QC procedure uses a fixed uncertainty range of 0.2°C for temperature and 0.2% for relative humidity. Could the authors comment on how these thresholds were selected? In particular, were these based on empirical validation, rounding, or sensitivity analysis. Also, is it uncertainty in the range of the reported variable or an acceptance criterion?

3. In section 2.3.4, the authors mention that when U, e, and Δe are internally consistent but the reconstructed dry-bulb temperature, T, differs from the transcribed value, the algorithm replaces the transcribed T using equations (5) and (6). This appears to assume that the discrepancy is mainly attributed to T rather than to uncertainties in e and Δe. Since small uncertainty in e and Δe can also propagate into the reconstructed T, could the authors comment on why an asymmetric criterion is considered.

4. In section 2.4, the authors mention that the software obtained transcription for 37 stations was validated against manually transcribed records from five stations. It would be helpful to discuss how representative these stations are of the broader archive in terms of climatic regions, handwriting styles, and temporal coverage.

5. In section 3.1, the authors mention that 1, 011, 297 records achieved the "highest internal quality flag." Could the authors briefly describe the flagging scheme used here. Also, the confirmed fraction for precipitation is substantially lower (30%). It would be useful to clarify how this lower confirmation rate affects the reliability and representativeness of the precipitation analyses.

6. In the methodology section, the authors describe the construction of kernel density estimation (KDEs) for individual stations, but it is not clear how the aggregated temperature distributions (Fig. 10) are obtained from these station-level KDEs. In particular, whether the aggregated distributions are generated by pooling all daily observations or by combining station-level distributions, and whether any weighting is applied to account for differences in data availability among stations.

Overall, I appreciate the authors for their substantial effort involved in rescuing, digitizing, quality-controlling, and analyzing the historical meteorological dataset for the DRC. The dataset fills an important observational gap and will provide a useful resource for future climate and environmental studies. I believe the manuscript is suitable for publication after the authors have addressed the above mentioned minor points.

We thank the reviewer for their time and valuable suggestions to improve the manuscript. Below, we respond to the individual comments and illustrate modifications to the manuscript to accommodate the concerns raised. We believe that the manuscript will benefit from these modifications. The following convention is used in this document to illustrate the proposed text modifications in the original manuscript: modified text.

Individual comments:

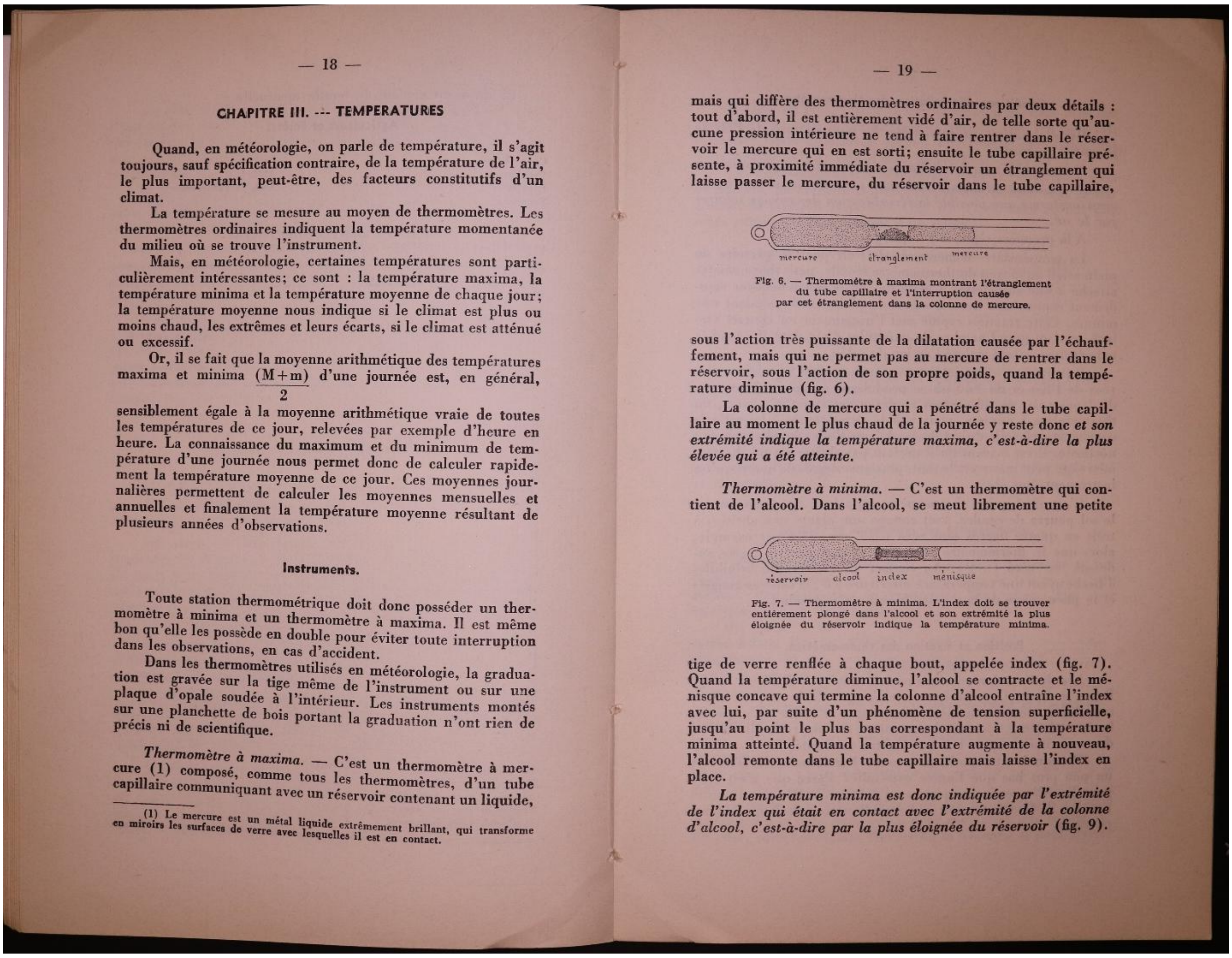
1. The dataset combines observations collected over several decades across multiple weather stations. During this period, instrumentation, measurement precision, calibration procedures, and observational practices may have changed. It would be helpful if the authors could comment on whether such temporal and interstation inconsistencies were assessed or accounted for (e. g., through homogenization or other quality-control procedures), and discuss their potential influence on the derived long-term climate trends.

Response:

We agree that temporal and interstation inconsistencies related to instrumentation, measurement precision, calibration procedures, station location, and observing practices may affect long-term station records. Some ancillary documents *Instructions aux observateurs – INEAC* and *Guide pratique à l'usage des observateurs – Réseau Météorologique du Congo Belge* describe the instruments and observation procedures prescribed for use at the stations, including thermometer types, measurement techniques, and rain-gauge information. These documents are available through the COBECORE records (Jacobsen et al., 2018; Coppieters. G., 2013). In the revised manuscript, we will clarify that our QA/QC workflow was designed to identify transcription errors and physically inconsistent values, but does not constitute a homogenisation procedure. We will also clarify that some information on the types of instruments used and some measurement techniques is available in the COBECORE records, but that the available metadata are not sufficiently complete to reconstruct station-specific histories of instrument changes, calibration procedures, measurement precision, station relocations, or observing practices over the full observational period. We therefore propose the following text (below) to the Data description in **Section 2.1** on the available additional archive documents on instrumentation and measurement procedures available through the COBECORE records, and to the **Discussion Section 4.2** on how such non-climatic inhomogeneities may affect the magnitude of individual station trends and part of the observed distributional shifts. At the same time, we note that the widespread warming signal across multiple stations, together with its consistency with previous regional-scale assessments, supports the robustness of the broad regional warming pattern.

To this end, we propose adding this paragraph to the Data description in **Section 2.1**:

> In addition to the weather data sheets, additional archive documents available through the COBECORE records (Jacobsen et al., 2018; Coppieters. G., 2013), namely *Instructions aux observateurs – INEAC* and *Guide pratique à l'usage des observateurs – Réseau Météorologique du Congo Belge,* provide information on some of the instruments and observation procedures prescribed for use at the stations, including thermometer types, measurement techniques, and rain-gauge information. These documents demonstrate that standardised instrumentation and observation procedures were prescribed to ensure rigorous and consistent measurements across the INEAC/INERA observational network. Examples of these documents are shown in Appendix Figures B4-B5.



— 18 —

CHAPITRE III. --- TEMPERATURES

Quand, en météorologie, on parle de température, il s'agit toujours, sauf spécification contraire, de la température de l'air, le plus important, peut-être, des facteurs constitutifs d'un climat.

La température se mesure au moyen de thermomètres. Les thermomètres ordinaires indiquent la température momentanée du milieu où se trouve l'instrument.

Mais, en météorologie, certaines températures sont particulièrement intéressantes; ce sont : la température maxima, la température minima et la température moyenne de chaque jour; la température moyenne nous indique si le climat est plus ou moins chaud, les extrêmes et leurs écarts, si le climat est atténué ou excessif.

Or, il se fait que la moyenne arithmétique des températures maxima et minima $\frac{(M+m)}{2}$ d'une journée est, en général, sensiblement égale à la moyenne arithmétique vraie de toutes les températures de ce jour, relevées par exemple d'heure en heure. La connaissance du maximum et du minimum de température d'une journée nous permet donc de calculer rapidement la température moyenne de ce jour. Ces moyennes journalières permettent de calculer les moyennes mensuelles et annuelles et finalement la température moyenne résultant de plusieurs années d'observations.

Instruments.

Toute station thermométrique doit donc posséder un thermomètre à minima et un thermomètre à maxima. Il est même bon qu'elle les possède en double pour éviter toute interruption dans les observations, en cas d'accident.

Dans les thermomètres utilisés en météorologie, la graduation est gravée sur la tige même de l'instrument ou sur une plaque d'opale soudée à l'intérieur. Les instruments montés sur une planchette de bois portant la graduation n'ont rien de précis ni de scientifique.

Thermomètre à maxima. — C'est un thermomètre à mercure (1) composé, comme tous les thermomètres, d'un tube capillaire communiquant avec un réservoir contenant un liquide,

(1) Le mercure est un métal liquide extrêmement brillant, qui transforme en miroirs les surfaces de verre avec lesquelles il est en contact.

— 19 —

mais qui diffère des thermomètres ordinaires par deux détails : tout d'abord, il est entièrement vidé d'air, de telle sorte qu'aucune pression intérieure ne tend à faire rentrer dans le réservoir le mercure qui en est sorti; ensuite le tube capillaire présente, à proximité immédiate du réservoir un étranglement qui laisse passer le mercure, du réservoir dans le tube capillaire,

mercure étranglement mercure

Fig. 6. — Thermomètre à maxima montrant l'étranglement du tube capillaire et l'interruption causée par cet étranglement dans la colonne de mercure.

sous l'action très puissante de la dilatation causée par l'échauffement, mais qui ne permet pas au mercure de rentrer dans le réservoir, sous l'action de son propre poids, quand la température diminue (fig. 6).

La colonne de mercure qui a pénétré dans le tube capillaire au moment le plus chaud de la journée y reste donc *et son extrémité indique la température maxima, c'est-à-dire la plus élevée qui a été atteinte.*

Thermomètre à minima. — C'est un thermomètre qui contient de l'alcool. Dans l'alcool, se meut librement une petite

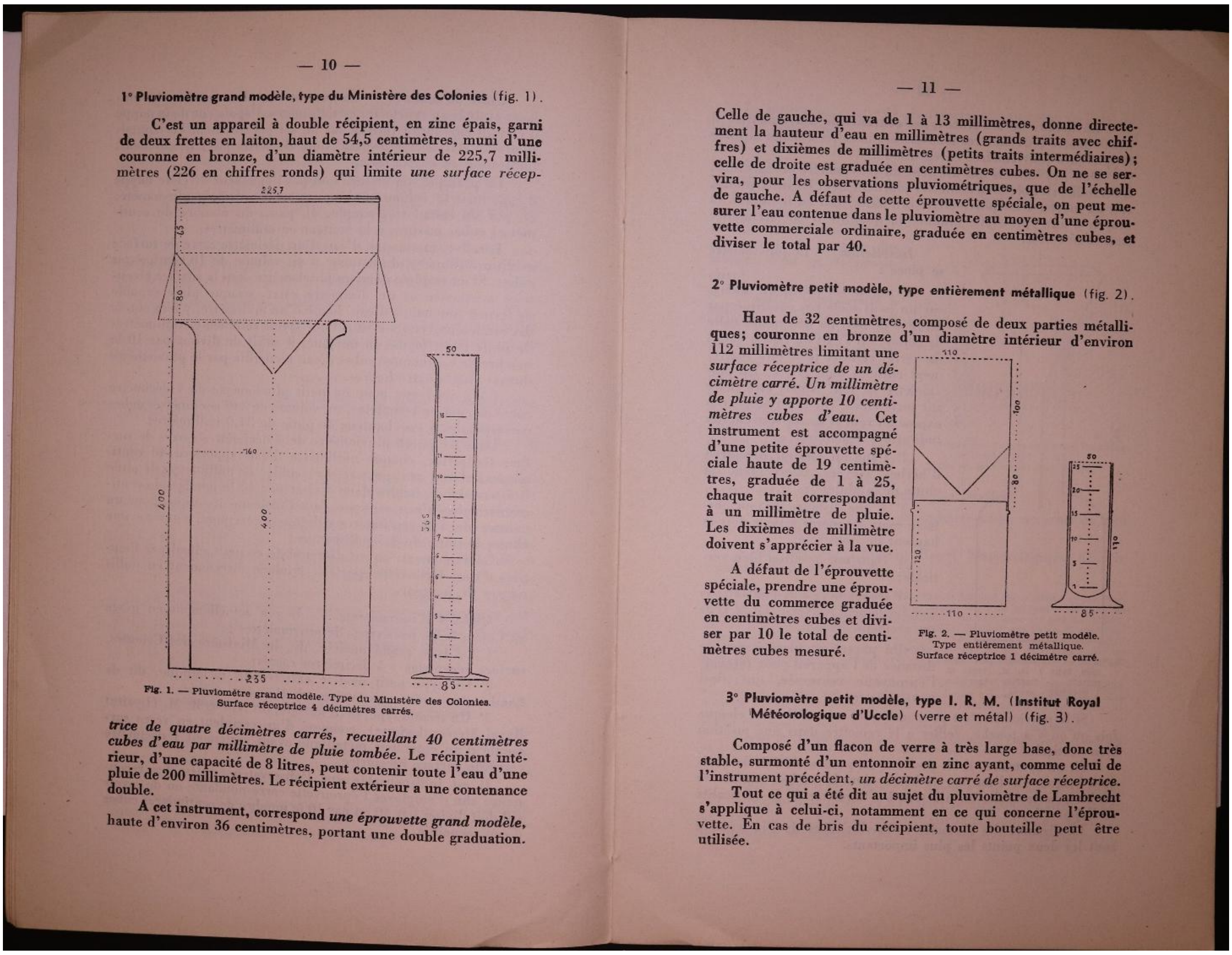
réservoir alcool index ménisque

Fig. 7. — Thermomètre à minima. L'index doit se trouver entièrement plongé dans l'alcool et son extrémité la plus éloignée du réservoir indique la température minima.

tige de verre renflée à chaque bout, appelée index (fig. 7). Quand la température diminue, l'alcool se contracte et le ménisque concave qui termine la colonne d'alcool entraîne l'index avec lui, par suite d'un phénomène de tension superficielle, jusqu'au point le plus bas correspondant à la température minima atteinte. Quand la température augmente à nouveau, l'alcool remonte dans le tube capillaire mais laisse l'index en place.

La température minima est donc indiquée par l'extrémité de l'index qui était en contact avec l'extrémité de la colonne d'alcool, c'est-à-dire par la plus éloignée du réservoir (fig. 9).

Appendix Figure B4: Example of COBECORE archive documents describing thermometers used to measure daily maximum and minimum temperatures and observation procedures used at INEAC/INERA stations. These documents provide useful contextual metadata and demonstrate that standardised instrumentation and observation procedures were prescribed across the INEAC/INERA observational network. However, they do not constitute complete station-specific histories of instrument changes, calibration procedures, relocations, or observing practices over the full observational period (Source: *Guide pratique à l'usage des observateurs – Réseau Météorologique du Congo Belge*; available in the COBECORE records (Jacobsen et al., 2018; Coppieters. G., 2013).


— 10 —

1° Pluviomètre grand modèle, type du Ministère des Colonies (fig. 1).

C'est un appareil à double récipient, en zinc épais, garni de deux frettes en laiton, haut de 54,5 centimètres, muni d'une couronne en bronze, d'un diamètre intérieur de 225,7 millimètres (226 en chiffres ronds) qui limite *une surface réceptrice de quatre décimètres carrés, recueillant 40 centimètres cubes d'eau par millimètre de pluie tombée*. Le récipient intérieur, d'une capacité de 8 litres, peut contenir toute l'eau d'une pluie de 200 millimètres. Le récipient extérieur a une contenance double.

Fig. 1. — Pluviomètre grand modèle. Type du Ministère des Colonies. Surface réceptrice 4 décimètres carrés.

A cet instrument, correspond *une éprouvette grand modèle*, haute d'environ 36 centimètres, portant une double graduation.

— 11 —

Celle de gauche, qui va de 1 à 13 millimètres, donne directement la hauteur d'eau en millimètres (grands traits avec chiffres) et dixièmes de millimètres (petits traits intermédiaires); celle de droite est graduée en centimètres cubes. On ne se servira, pour les observations pluviométriques, que de l'échelle de gauche. A défaut de cette éprouvette spéciale, on peut mesurer l'eau contenue dans le pluviomètre au moyen d'une éprouvette commerciale ordinaire, graduée en centimètres cubes, et diviser le total par 40.

2° Pluviomètre petit modèle, type entièrement métallique (fig. 2).

Haut de 32 centimètres, composé de deux parties métalliques; couronne en bronze d'un diamètre intérieur d'environ 112 millimètres limitant une *surface réceptrice de un décimètre carré. Un millimètre de pluie y apporte 10 centimètres cubes d'eau.* Cet instrument est accompagné d'une petite éprouvette spéciale haute de 19 centimètres, graduée de 1 à 25, chaque trait correspondant à un millimètre de pluie. Les dixièmes de millimètre doivent s'apprécier à la vue.

A défaut de l'éprouvette spéciale, prendre une éprouvette du commerce graduée en centimètres cubes et diviser par 10 le total de centimètres cubes mesuré.

Fig. 2. — Pluviomètre petit modèle. Type entièrement métallique. Surface réceptrice 1 décimètre carré.

3° Pluviomètre petit modèle, type I. R. M. (Institut Royal Météorologique d'Uccle) (verre et métal) (fig. 3).

Composé d'un flacon de verre à très large base, donc très stable, surmonté d'un entonnoir en zinc ayant, comme celui de l'instrument précédent, *un décimètre carré de surface réceptrice.*

Tout ce qui a été dit au sujet du pluviomètre de Lambrecht s'applique à celui-ci, notamment en ce qui concerne l'éprouvette. En cas de bris du récipient, toute bouteille peut être utilisée.

Appendix Figure B5: Example of COBECORE archive documents describing rain gauges used to measure daily precipitation and observation procedures used at INEAC/INERA stations. These documents provide useful contextual metadata and demonstrate that standardised instrumentation and observation procedures were prescribed across the INEAC/INERA observational network. However, they do not constitute complete station-specific histories of instrument changes, calibration procedures, relocations, or observing practices over the full observational period (Source: *Guide pratique à l'usage des observateurs – Réseau Météorologique du Congo Belge*; available in the COBECORE records (Jacobsen et al., 2018; Coppieters. G., 2013).

We also propose the following modifications to the text in the **Discussion Section 4.2:**

However, it is important to note that, while QA/QC checks were applied to identify and correct transcription errors and physically inconsistent values, homogenisation was not performed to detect and adjust for potential non-climatic inhomogeneities in the station records. Such inhomogeneities, well-rehearsed throughout the literature, may arise from changes in station location, instruments, calibration procedures, measurement precision, observation times or practices, and could introduce artificial (non-climatic) shifts in the temperature time series (Williams et al. 2012; Peterson et al., 1998). While digitized records documenting the types of instruments used and some measurement procedures are available in the COBECORE records (Appendix Figures B4-B5 ; Jacobsen et al., 2018; Coppieters. G., 2013), the available metadata are not sufficiently complete to reconstruct station-specific histories of such changes throughout the full observational period.

Possible physical contributors may also have influenced differences among the station records. These include differences in elevation (Hock et al. 2019), changes in the immediate station environment, proximity to forests or water bodies, and local and remote land-use or land-cover change (Zeppetello et al., 2020). For instance, Mangaza et al. (2026) report a loss of approximately 9.7% of tropical forest in the Tshopo province, where three stations in our dataset are located (Yangambi-Km.5, Yaekama, and Gazi; the former two analysed), over the period 1900-2023 due to deforestation for agriculture and settlement. Within Yangambi alone, approximately 24 ha of primary forest was lost between 2002 and 2025, accounting for only about 2% of the total tree-cover loss in the area recorded over the same period (Global Nature Watch, 2026; Turubanova, S. et al. 2018; Hansen, M. C. et al. 2013). Between 2001 and 2025, permanent agriculture was associated with approximately 19% of the tree-cover loss in Yangambi (Global Nature Watch, 2026; Sims, M.J. et al. 2025; Hansen, M. C. et al. 2013). Such changes can lead to localised warming effects (Zeppetello et al., 2020).

More generally, differences in trend magnitude and sign among stations may reflect a combination of large-scale climate variability, local environmental conditions, and station-specific non-climatic effects. However, because this study did not perform a formal attribution analysis, we could not quantify the relative contributions of such possible drivers to the observed station-level trends. Therefore, part of the distribution changes may reflect these local physical and/or non-climatic influences from one or more stations, although the consistency with previous regional assessments suggests that the broad warming pattern is robust.

2. The QA/QC procedure uses a fixed uncertainty range of 0.2°C for temperature and 0.2% for relative humidity. Could the authors comment on how these thresholds were selected? In particular, were these based on empirical validation, rounding, or sensitivity analysis. Also, is it uncertainty in the range of the reported variable or an acceptance criterion?

Response:

We agree that the meaning of the term “uncertainty margin” in the MeteoSaver configuration module should be clarified in the manuscript. In the revised manuscript, we will clarify that the fixed margins of 0.2°C for temperature and 0.2 percentage points for relative humidity are used as acceptance criteria in the QA/QC consistency checks, rather than as estimates of the full observational uncertainty. These margins were selected to account for rounding in the handwritten records such as Tavg, and in derived quantities such as pentad averages. They therefore allow for small discrepancies arising from rounding when transcribed values are compared with recalculated values, averages, or other consistency relationships. We will also clarify that these fixed margins were based on reporting precision and rounding considerations, and not on a separate sensitivity analysis, in the proposed modified text in **Section 2.3.5**:

> We apply MeteoSaver v1.1 to 9,885 images of historical weather data sheets from 36 stations across the DRC (one of the 37 stations was previously transcribed by Kasongo Yakusu et al. (2023)), to automatically transcribe daily records of Tx, Tn, Tavg, P, T, and T’a. As a prerequisite for executing MeteoSaver, we input the user-defined settings in its configuration module, for example: (i) we set the computational environment to hpc, as we run the software on HPC infrastructure due to the large volume of images. Processing all images requires approximately 4.5 days on the VSC (Flemish Supercomputer Center) infrastructure using 18 parallel CPU cores. (ii) We specify all required input and output directories. (iii) We set region-specific maximum and minimum temperature thresholds to 40°C and 5°C, respectively (Muheki et al., 2026c; Alsdorf et al., 2016). (iv) We define the uncertainty margins for the QA/QC checks as 0.2°C for temperature values and 0.2

percentage points for relative humidity values. In the MeteoSaver configuration module, these margins are not estimates of the full observational uncertainty, but acceptance criteria used to account for rounding in the handwritten records and in derived quantities during internal consistency checks. Temperature values in the sheets are reported to one decimal place; therefore, a margin of 0.2°C allows small discrepancies arising from rounding when comparing transcribed values with recalculated values, pentad averages or consistency relationships such as that between Tx, Tn, and Tavg. Similarly, the 0.2 percentage-point margin for relative humidity accounts for rounding differences in the pentad averages and derived relative-humidity consistency checks. These margins were therefore selected based on reporting precision and rounding considerations, rather than through a separate sensitivity analysis. We detail the full list of our configuration settings in Appendix Tables C1 and C2, and in the configuration.ini file available in the Zenodo repository (Muheki et al., 2026b).

3. In section 2.3.4, the authors mention that when U, e, and Δe are internally consistent but the reconstructed dry-bulb temperature, T, differs from the transcribed value, the algorithm replaces the transcribed T using equations (5) and (6). This appears to assume that the discrepancy is mainly attributed to T rather than to uncertainties in e and Δe. Since small uncertainty in e and Δe can also propagate into the reconstructed T, could the authors comment on why an asymmetric criterion is considered.

Response:

We agree that the original text did not sufficiently explain the rationale behind this asymmetric QA/QC criterion. In the revised manuscript, we will clarify that this correction is only applied after *U, e,* and *Δe* have passed the preceding internal consistency checks described in step (ii) of Section 2.3.4. In that step, U is recalculated from *e* and *Δe*, and the three values are considered consistent if the recalculated U agrees with the transcribed U within the set margin of 0.2 percentage points. At the subsequent step, if the reconstructed dry-bulb temperature *T* is not equal to the transcribed *T* (or within the set uncertainty margin of 0.2°C), then *T* is the only variable inconsistent with the reconstructed relationship. We therefore consider a single transcription error in *T* more likely than compensating errors in *U, e,* and *Δe* that would still preserve their mutual consistency. We will also clarify that this does not assume that *e* and *Δe* are free of uncertainty. Small discrepancies in the recalculated T and the transcribed *T* arising from rounding or derived-variable precision are accounted for using the 0.2°C acceptance margin. The revised text will explicitly state that this is a rule-based QA/QC consistency correction, rather than as a quantitative estimate of the propagated uncertainty in the calculated *T*. We therefore propose to modify the text in **Section 2.3.4** as follows:

iii) Following this, if the U, e and Δe values at a particular time have passed the internal consistency check in step (ii), we subsequently calculate the T using a modified Magnus-Tetens formula (Bolton, 1980) (Eqs. 5 - 6). If the calculated T is equal to (or within the set uncertainty margin of 0.2°C of) the transcribed value, we mark the transcribed value as confirmed; otherwise, it is replaced with the calculated value. In the case of replacement, this asymmetric correction is applied because, at this stage, U, e and Δe have been shown to be mutually consistent, and T, in that case, is the only variable inconsistent with the reconstructed relationship. We therefore consider a single transcription error in T more likely than compensating errors in , U, e, and/or Δe that nevertheless preserve their mutual consistency. However, this step does not assume that e and Δe are free of uncertainty. Small discrepancies arising from rounding or derived-variable precision of T are accounted for through the 0.2°C margin. The correction should therefore be interpreted as a rule-based QA/QC consistency correction, rather than as a quantitative estimate of the propagated uncertainty in the calculated T.

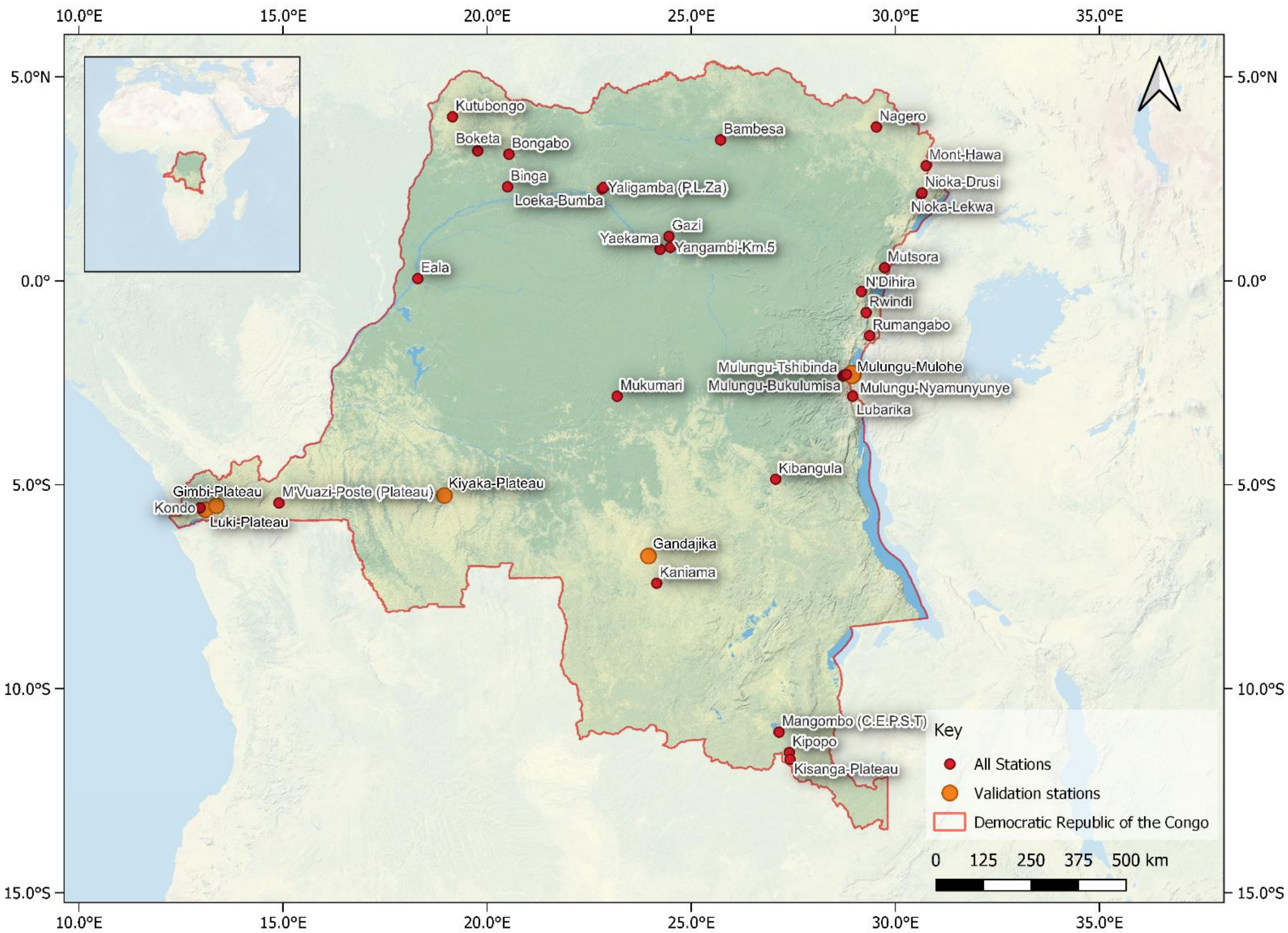
4. In section 2.4, the authors mention that the software obtained transcription for 37 stations was validated against manually transcribed records from five stations. It would be helpful to discuss how representative these stations are of the broader archive in terms of climatic regions, handwriting styles, and temporal coverage.

Response:

We thank the reviewer for this comment. The five validation stations were selected because they were the only INERA stations for which overlapping manually transcribed records were available. We agree that this should be clarified and that these stations should not be presented as a fully representative sample of all climatic regions, handwriting styles, and temporal periods in the broader archive. The validation assesses MeteoSaver performance across the available manually transcribed subset, rather than providing a complete evaluation of all possible archive conditions. In the revised manuscript, we will clarify this in section 2.4. We will also note that the validation subset is nevertheless useful because the five stations are geographically distributed across the DRC (see proposed Appendix Fig. A3 below), originate from different station records and observers, and span different periods from the 1960s to the 2010s, including both short and multi-decadal records. As a result, the validation includes variation in handwriting styles, table completion practices, and paper conditions. Below is our proposed modified text in **Section 2.4**:

> Following the transcription and quality control of the daily temperature and precipitation data from 37 weather stations across the DRC using MeteoSaver, we validate the automatically transcribed data against manually transcribed records by INERA for the five stations where such data were available and overlapped with our automatically transcribed time series, namely: Kiyaka-Plateau, Luki-Plateau, Gandajika, Mulungu-Mulohe, and Gimbi-Plateau. While these five stations may not constitute a fully representative sample of all climatic regions in the DRC, handwriting styles, and temporal periods in the broader archive, they provide a useful validation subset because they are geographically distributed across the DRC (Appendix Fig. A3), originate from different stations and observers, and span different periods from the 1960s to the 2010s, including both short and multi-decadal records. They therefore capture differences in handwriting styles, table completion practices, and paper maintenance conditions. The validation should therefore be interpreted as an assessment of MeteoSaver performance across the available manually transcribed subset, rather than as a complete evaluation of all possible archive conditions.

Following this, we propose to add the figure below (referred to in the proposed text above) to the Appendix, showing the geographical distribution of the validation dataset across the DRC:



Appendix Figure A3: Locations of weather stations in the DRC whose records are archived at the INERA-Section de Climatologie in Yangambi. Red circles show the full station network, while orange circles highlight the subset of stations for which manually transcribed reference data were available and used for validation in Sect. 2.4.1. (Background map courtesy of ESRI National Geographic | Powered by Esri).

5. In section 3.1, the authors mention that 1, 011, 297 records achieved the "highest internal quality flag." Could the authors briefly describe the flagging scheme used here. Also, the confirmed fraction for precipitation is substantially lower (30%). It would be useful to clarify how this lower confirmation rate affects the reliability and representativeness of the precipitation analyses.

Response:

We agree that the meaning of the "highest internal quality flag" should be clarified. In the revised manuscript, we will further explain that this flag refers to values that passed all applicable automated QA/QC checks described in Sect. 2.3, including variable-specific physical threshold checks and internal consistency checks. These values are referred to throughout the manuscript as QA/QC confirmed values. We propose the following modifications to the text in Section 3.1 as follows:

> In total, 1,011,297 daily and sub-daily temperature records, as well as daily precipitation observations, across the 37 weather stations achieved the highest internal quality flag, meaning that they passed all applicable QA/QC checks described in Section 2.3. These

records are hereafter referred to as QA/QC confirmed values. Together with manually transcribed daily temperature and precipitation records from previous efforts, namely Kasongo Yakusu et al. (2023) (95,395 records) and the COBECORE project (246,815 records), they form the foundation of the first large-scale, digital historical climate dataset for the Democratic Republic of the Congo.

In this revision, we will also discuss the lower confirmation rate for precipitation more clearly. We will clarify that this lower fraction does not necessarily imply that all unconfirmed precipitation values are erroneous, but mainly reflects the more limited set of automated QA/QC checks available for precipitation. Unlike temperature, for which several physical and internal consistency checks could be applied, precipitation was checked only using pentad total consistency. Therefore, the precipitation analyses are restricted to stations with sufficiently high fractions of QA/QC confirmed daily precipitation values and should be interpreted as station-specific examples rather than as a spatially representative assessment of precipitation extremes across the DRC. Below are our proposed modifications to the text in Sect. 3.1:

> This percentage is lower for daily precipitation records ranging from 10.3-89.4% with a mean of 30% across the stations. This lower confirmation rate does not necessarily imply that all unconfirmed precipitation values are erroneous but rather reflects the more limited set of automated QA/QC checks available for precipitation, in comparison to temperature. The QA/QC module includes only pentad total consistency checks for precipitation, whereas daily temperature values are subjected to additional consistency checks, including the physical relationship between Tx, Tn, and Tavg, their consistency with the diurnal temperature range (Ampl.), and maximum and minimum threshold checks (see Figure 2) (Muheki et al., 2026c). These additional checks increase the likelihood of detecting and correcting transcription errors, thereby leading to a higher percentage of confirmed temperature values. Consequently, the precipitation analyses in this study are restricted to station records for which more than 30 years of data are available and at least 80% of the transcribed daily precipitation values were confirmed through the QA/QC procedure, and the results should therefore be interpreted as station-specific examples rather than as a spatially representative assessment of precipitation extremes across the DRC.

6. In the methodology section, the authors describe the construction of kernel density estimation (KDEs) for individual stations, but it is not clear how the aggregated temperature distributions (Fig. 10) are obtained from these station-level KDEs. In particular, whether the aggregated distributions are generated by pooling all daily observations or by combining station-level distributions, and whether any weighting is applied to account for differences in data availability among stations.

Response:

We thank the reviewer for this comment. We agree that the construction of the aggregated temperature distributions in Fig. 10 was not sufficiently clear. In the revised manuscript, we will clarify that these aggregated KDEs were not obtained by averaging station-level KDEs. Instead, all QA/QC-confirmed daily observations from the stations with data during the common analysis period 1961--1990 were pooled into a single regional dataset. Decade-specific KDEs were then calculated from the pooled daily values Tx, Tn, and Tavg. No additional station-level weighting was applied. We have clarified that Fig. 10 should therefore be interpreted as the distribution of available QA/QC-confirmed station-day observations across the DRC during 1961--1990, rather than as an equal-weighted average of station-level distributions. We propose to add this clarification to the Methods section and briefly note it in the caption of Fig. 10. Below are the proposed modifications to the manuscript:

In the Methods Section 2.4.2:

In parallel, we apply the kernel density estimation (KDE) method (Węglarczyk, 2018) to construct probability density functions of Tx, Tn, and Tavg, grouped by decade per station. This approach enables us to (i) examine shifts in the overall temperature distributions over time, assessed through the changes in the mode and the 95th percentile of each variable's distribution per decade, and (ii) evaluate changes in the temperature extremes per station, based on the frequency of days falling outside the middle 90% of the distribution (i.e., below the 5th percentile and above the 95th percentile). Similar to Lorenz et al. (2019), we define these extremes for each variable using KDE-derived percentiles calculated across the entire station record and then count how many days per decade fall below (cold days/nights) or above (hot days/ warm nights) these thresholds. This approach enables a robust assessment of both the central shift in temperature and the changing frequency of temperature extremes over time at each station. For the aggregated temperature distributions across stations, we pool all QA/QC-confirmed daily observations from the stations into a single regional dataset and restrict the analysis to the most overlapping observation window, 1961--1990 (Fig. 4). Decade-specific KDEs are then calculated from the pooled daily values of Tx, Tn, and Tavg. The aggregated KDEs should therefore be interpreted as distributions of available QA/QC-confirmed station-day observations across the DRC during 1961--1990, rather than as equal-weighted averages of station-level distributions.

In the Results Section 3.3.1:

We then examine the decadal temperature distributions both at individual stations and for the aggregated dataset across stations. For the latter, all available QA/QC-confirmed station-day observations are pooled across stations, and the analysis focuses on the most overlapping observation window (1961--1990).

Proposed modification to the caption of Fig 10:

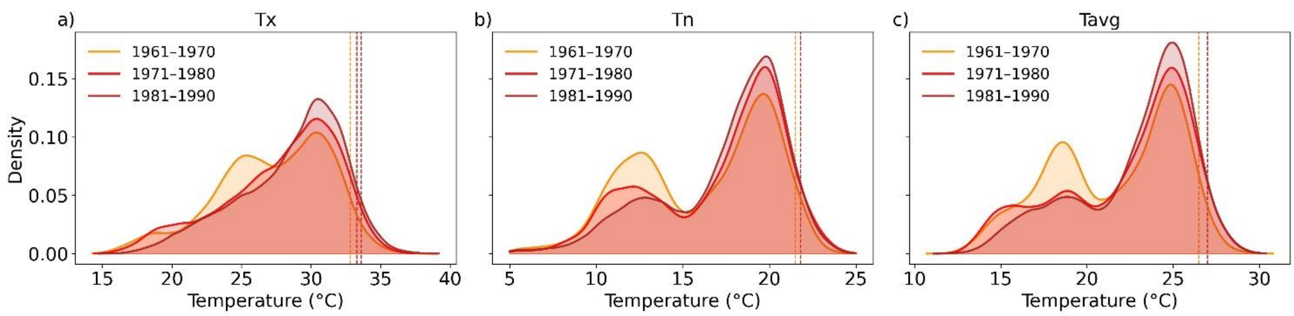


Figure 10. Decadal shifts in temperature distributions based on an aggregated dataset of 37 stations in the DRC for the period 1961 to 1990. The aggregated distributions were obtained by pooling all available QA/QC-confirmed station-day observations across stations during 1961--1990; no additional station-level weighting was applied. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1961–1970 (yellow), 1971–1980 (red), and 1981–1990 (maroon). Vertical dashed lines indicate the 95th percentile values for each decade, illustrating the progressive rightward shift in temperature extremes over time.