https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

From Paper to Proof: Revealing Congo Basin Warming Through Rescued Climate Archives

Derrick Muheki[1], Koen Hufkens[2], Kim Jacobsen[3, 15], Hans Verbeeck[3], Pascal Boeckx[4], Dominique Kankonde Ntumba[5], Olivier Kapalay Moulasa[5], Bas Vercruysse[6], Julie M. Birkholz[6, 16], Christophe Verbruggen[6], Ed Hawkins[7], Seppe Lampe[1], Emmanuel Kasongo Yakusu[8, 9, 10], Fils Makanzu Imwangana[11,17], José Mbifo[12], Théophile Besango Likwela[12], Félicien Meunier[1,3,4], Olivier Dewitte[13], Peter Thorne[14], and Wim Thiery[1]

[1]Vrije Universiteit Brussel, Department of Water and Climate, 1050 Brussels, Belgium
[2]BlueGreen Labs (bv), 9120 Melsele, Belgium
[3]Ghent University, Department of Environment, 9000 Ghent, Belgium
[4]Ghent University, Isotope Bioscience Laboratory - ISOFYS, 9000 Ghent, Belgium
[5]Institut National pour l'Etude et la Recherche Agronomiques, Direction Générale, Kinshasa, Democratic Republic of the Congo
[6]Ghent University, Department of History, Ghent Centre for Digital Humanities, 9000 Ghent, Belgium
[7]University of Reading, National Centre for Atmospheric Science, Department of Meteorology, RG6 6ET Reading, United Kingdom
[8]Ghent University, Laboratory of Wood Technology, Department of Environment, 9000 Ghent, Belgium
[9]Royal Museum for Central Africa, Service of Wood Biology, 3080 Tervuren, Belgium
[10]Université de Kisangani, Faculté de Gestion des Ressources Naturelles Renouvelables, Kisangani, Democratic Republic of the Congo
[11]Université de Kinshasa, Faculté des Sciences, Kinshasa, Democratic Republic of the Congo
[12]Institut National pour l'Etude et la Recherche Agronomiques, Centre de Recherche de Yangambi, Yangambi, Democratic Republic of the Congo
[13]Royal Museum for Central Africa, Department of Earth Sciences, 3080 Tervuren, Belgium
[14]Maynooth University, ICARUS Climate Research Centre, Maynooth, Ireland
[15]Royal Museum for Central Africa, Department of Biology, 3080 Tervuren, Belgium
[16]Royal Library of Belgium, KBR Digital Research Lab, 1000 Brussels, Belgium
[17]Center for Geological and Mining Research, Department of Urban Geology and Environment, Kinshasa, Democratic Republic of Congo

Correspondence: Derrick Muheki (derrick.muheki@vub.be)

Abstract. The Congo Basin in Central Africa remains one of the few regions globally where the Intergovernmental Panel on Climate Change has neither assessed changes in hot extremes and extreme precipitation since the 1950s nor attributed such changes to anthropogenic influences, primarily due to the sparsity of in situ observational data. Although extensive daily weather records exist, spanning from the 1900s to the early 2000s and covering numerous stations across the basin, the majority of these remain archived on paper, limiting their accessibility for climate analysis. Here we present and analyse over 1 million temperature and precipitation observations from 37 weather stations in the Democratic Republic of the Congo, which have until now been unavailable to the research community. To produce this dataset, we first digitized (imaged) 9,885 paper sheets stored in local archives during two dedicated field campaigns. We subsequently apply an improved version of the MeteoSaver software

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

(version 1.1) to transcribe the imaged sheets, yielding daily time series of maximum, minimum, and average temperatures, precipitation, as well as dry bulb and wet bulb temperatures measured at three times per day (06:00, 15:00, and 18:00 local time). After quality control, analysis of multi-decadal temperature data across the basin reveals a consistent and accelerating warming signal since the 1960s, characterized by a warming shift in the distribution of daily maximum, minimum, and average temperatures with each successive decade. Median trends across 21 of the 37 stations with sufficient data availability are 0.22°C, 0.10°C, and 0.15°C per decade for daily maximum, minimum, and average temperatures, respectively, corresponding to approximately 0.7°C, 0.3°C, and 0.5°C of warming during the 1961–1990 period. We further find an increasing frequency of hot extremes and a decreasing frequency of cold extremes with each successive decade, with the most recent decade exhibiting nearly twice as many hot days per year and fewer cold days compared to the earliest decade. Analysis of precipitation at three stations with sufficient data indicates an increased frequency of heavy precipitation events at two stations and no substantial change at the third; however, limited spatial coverage due to data availability restricts broader conclusions for the DRC. Overall, this analysis of rescued weather data from Central Africa highlights the urgent need to close the knowledge gap on climate trends in the Congo Basin, one of the world's most data-sparse yet climatically significant regions.

1 Introduction

Recent studies have unequivocally identified anthropogenic climate forcing as the main driver for the observed increase in global mean air surface temperature over recent decades (Eyring et al., 2021; Bindoff et al., 2013). Similarly, the increase in frequency and intensity of hot extremes, such as heatwaves, in most regions of the world since the 1950s has also been attributed to human-induced climate change (Eyring et al., 2021; Seneviratne et al., 2021). However, unlike most regions, Central Africa remains one of the few areas where the Intergovernmental Panel on Climate Change (IPCC) has neither assessed changes in hot extremes and extreme precipitation since the 1950s nor attributed such changes to human influence, primarily due to the sparsity of in situ observational data (IPCC, 2021b; Eyring et al., 2021; Seneviratne et al., 2021).

The persistent data sparsity across Central Africa has been a significant impediment to studies assessing the regional effects of human-induced climate change on ecosystems, societies, and the water cycle (CSC, 2013), despite the region hosting the Congo Basin, which contains the world's second-largest river and tropical rainforest (Alsdorf et al., 2016). For example, while observed changes in the hydrological cycle and water resources have been robustly attributed to anthropogenic climate change in many regions (e.g., Gudmundsson et al., 2021; Grant et al., 2021; Marvel et al., 2019; Marvel and Bonfils, 2013; Zhang et al., 2007), the Congo Basin, remains a major observational blind spot.

Moreover, the Congo Basin is home to some of the world's most vulnerable populations and is experiencing rapid demographic growth in both urban and rural areas (Megevand, 2013; Sonwa et al., 2012; Yuh et al., 2024). These populations are increasingly exposed to climate-related hazards, such as landslides and gully erosion, which are further exacerbated by

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

human-induced land transformation associated with population growth (Depicker et al., 2021; Dille et al., 2022; Mawe et al., 2025).

Furthermore, satellite data suggest that the Congo Basin has experienced a drying trend in recent decades, characterized by reductions in precipitation, canopy water content, and forest photosynthetic capacity (Zhou et al., 2014; Hua et al., 2016; Wongchuig et al., 2025). Although Land Surface Models (LSMs) and Global Hydrological Models (GHMs) can be used to study hydroclimatic trends within the region, the quality of these models remains uncertain due to the limited availability of ground-based observations for validation (Zhou et al., 2014).

Substantial records of observed weather within the Congo Basin, stored in both local and international archives, present an opportunity to overcome these data limitations and thus facilitate the analysis of hydroclimatic trends in the basin (Alsdorf et al., 2016; Kasongo Yakusu et al., 2023). However, as in many parts of the world, millions of in-situ weather records from various stations remain preserved within archives as hard copies or in image form (e.g., Hawkins et al., 2022, 2023; Latapy et al., 2022; de Smeth et al., 2024). To date, daily temperature and precipitation records originally collected by the Institut National pour l'Etude Agronomique du Congo Belge (INEAC), with some stations dating back as early as 1908, though the majority begin in the 1930s, have been preserved in the State Archives of Belgium. These records, which extend up to 1960, were digitized under the Congo Basin Eco-climatological Data Recovery and Valorization (COBECORE) project, which involved photographing all available climatological and ecological records for 575 climate stations spread throughout the Congo Basin (Jacobsen et al., 2018; COBECORE, 2018). Additionally, post-1960 daily temperature and precipitation records for 37 weather stations in the Democratic Republic of the Congo (DRC) are hosted in the archives of the Institut National pour l'Etude et la Recherche Agronomiques (INERA), formerly INEAC, in Yangambi. These records span from 1960 to the 1990s for most stations, with a few extending into the early 2000s.

This paper describes the data rescue efforts within the INERA archives in Yangambi, for post-1960 handwritten daily temperature and precipitation records taken at 37 weather stations in the DRC. This process involves: (i) compiling a data inventory, (ii) digitizing (that is, imaging) the handwritten weather records from paper to digital format, and (iii) transcribing the digitized image into machine-readable format using MeteoSaver, an open-source software we developed based on machine learning (ML) algorithms (Muheki et al., 2026c). MeteoSaver takes the digitized images of tabular observation sheets as input and transcribes the data, in this case, handwritten daily weather data, into a standardized, machine-readable format known as the Station Exchange Format (SEF), as prescribed by the Copernicus Data Rescue Service (Muheki et al., 2026c). By automating transcription, MeteoSaver saves numerous man-years of manual data entry and demonstrates its potential for wider application in similar historical data rescue efforts.

In addition to this newly automatically transcribed temperature dataset, we incorporate three recently available, manually transcribed datasets for the region. First, Kasongo Yakusu et al. (2023) presents daily temperature and precipitation data for one station, Yangambi-Km.5, spanning 1960-2020. As one of the main findings, Kasongo Yakusu et al. (2023) reports a long-term increase of approximately 1°C in the annual average mean daily temperature over seven decades, and an increase in frequency of hot extremes, specifically warm days and nights, recorded at that station. Second, since 2000, INERA has also transcribed daily temperature and precipitation data for 6 of the 37 weather stations. Most of these records cover the period from the

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

1990s to the mid-2010s, with only two stations having data extending back to the 1960s, and one continuing up to the early 2020s. Third, we incorporate pre-1960 data from the COBECORE project for three stations that are also part of the 37-station post-1960 INERA dataset. We therefore include the Kasongo Yakusu et al. (2023), INERA and COBECORE datasets in our combined temperature and precipitation dataset, enabling the construction of long, continuous time series.

The resulting time series forms the largest in situ temperature and precipitation dataset available for the Congo Basin, and represents a major step toward advancing hydroclimatic and climate change research in this data-sparse region.

Using this consolidated temperature and precipitation dataset, we analyze temperature trends across the Congo Basin from the 1930s to the early 2000s. We examine long-term trends in daily maximum, minimum, and average temperatures; investigate shifts in their distributions across decades; and assess changes in the frequency of hot and cold extremes. We further investigate precipitation extremes through changes in daily precipitation return periods. Lastly, we compare the observed temperature trends with those derived from the state-of-the-art ERA5-Land reanalysis product developed by the European Centre for Medium-Range Weather Forecasts (ECMWF).

2 Data and Methods

2.1 Data

The *INERA-Section de Climatologie* archives in Yangambi (hereafter referred to as INERA archives) host daily meteorological observations from 37 weather stations throughout the DRC, including records of temperature, evaporation, and precipitation, along with various other weather data variables, with most spanning from the 1960s to the mid 1990s, and only a few through the early 2000s. Historically, these archives have served as the central repository for all weather data recorded at INERA stations throughout the country (Fig. 1). These records consist of paper bundles containing handwritten observations on tabular sheets, with each bundle representing weather data from a specific station for a given year, and each sheet representing one month's data from that station (Fig. 2).

Each sheet contains daily weather observations, their pentad totals and averages (5-day and/or 6-day), as well as monthly totals and averages. Specifically, for temperature records, each sheet provides daily maximum, minimum, and average values, along with the diurnal temperature range. There is also dry bulb and wet bulb temperatures measured at three times per day (06:00h, 15:00h, and 18:00h local time). Additionally, each sheet provides metadata, including the station name, year, month, observer's names, authorization signature, and occasionally extra comments at the bottom of the sheet (Fig. 2). No information on the location of the weather stations is given on these sheets, however, the data inventory we performed at the archives provided us with information on the station code, latitude, longitude, and altitude of each of the weather stations (see Appendix Fig. A1).

While the paper size, quality, colour, and maintenance condition vary across the bundles and stations, the format of the sheets and tables is generally consistent across these post-1960 records. However, it is important to note that in a small fraction of cases, the tables on the sheets were hand-drawn, possibly when the weather station ran out of standard paper, which presents

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

an additional challenge for non-manual transcription (see Appendix Fig. A2) In this dataset, 225 sheets were hand-drawn and were therefore excluded from automated transcription in this study.

The pre-1960 records digitized (imaged) during the COBECORE project follow a different table structure than the post-1960 sheets. Although they similarly provide daily maximum, minimum, and average temperatures for a given month and station, they do not include pentad totals or averages. Instead, these earlier sheets report only end-of-month totals and averages. They also lack the dry bulb and wet bulb temperatures recorded at three times per day as in the later records, and the table formats vary from decade to decade, and in a few cases between stations. Examples of these variations can be seen in Appendix Figs. B1, B2, and B3, corresponding to the 1930s, 1940s, and 1950s, respectively.

Figure 1. Two photographs showing the interior of the historical weather data archives at the INERA - Section de Climatologie site in Yangambi, Democratic Republic of the Congo. These shelves contain thousands of paper-based meteorological records spanning several decades, and form the foundation of our climate data rescue efforts. Photographs taken on 9 March 2023.

2.2 Data Rescue Missions

We undertook two data rescue missions: (i) a two-week campaign to construct a comprehensive data inventory (7-17 March 2023), and (ii) a two-month campaign for the digitization of the archived historical weather data sheets (August-September 2023). Throughout both the data inventory and digitization phases, we followed the guidelines for climate data rescue established by the World Meteorological Organization (WMO, 2016). In the following subsections, we describe our inventory and digitization campaigns in the INERA archives, DRC.

2.2.1 Data Inventory

The inventory process within the INERA archives involved two main steps: (i) collecting metadata, including information on all stations such as station names, codes, and location (latitude, longitude, and altitude), and (ii) cataloging the available data sheets and variables present for each station by year, along with the paper formats and maintenance condition.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

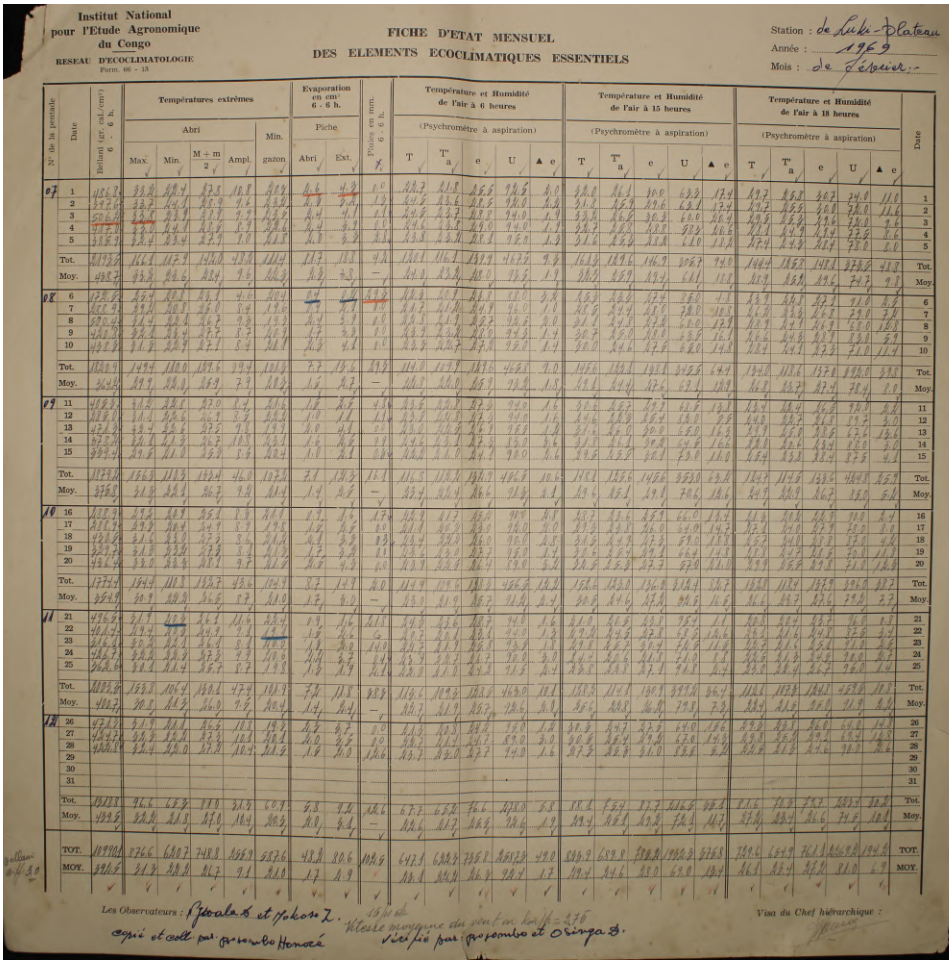


Figure 2. An example One Month observed weather data sheet for Station Luki-Plateau (S 5°37', E 13°06') in DRC available within the *INERA-Section de Climatologie* archives in Yangambi. The sheet contains the following information: (i) the station name ('Station'), year ('Année') and month ('Mois') in the top right corner, (ii) the data owner ('Institut National pour l'Étude Agronomique du Congo') in the top left corner, (iii) a large table (center) showing the pentad number ('N° de la pentade'), date ('Date'), a measure for total solar radiation ('Bellani'), extreme temperatures in degree Celsius ('Températures extrêmes') consisting of maximum, minimum and average temperatures, as well as the diurnal temperature range ('Ampl.'), Evaporation in cubic centimeters ('Evaporation'), Rainfall ('Pluies') in millimeters per day, and Dry- and Wet-bulb Temperatures and Humidity ('Température et Humidité') recorded at three times of the day, namely 06:00h, 15:00h, and 18:00h local time, and (iv) observer's names ('Les Observateurs') in the bottom left, authorization signature ('Visa du Chef hiérarchique') in the bottom right, and extra comments written in the bottom center of the sheet. Additionally, the table also shows 5-day (and/or 6-day) totals ('Tot.') and means ('Moy.'), as well as monthly totals and mean values. Note that the values in these sheets are all handwritten, and the handwriting styles and paper maintenance conditions vary throughout the dataset. In this case, the image represents a well-preserved example relative to the range of paper conditions present in the archive.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

The first step is important as it ensures that the locations of the recorded weather data are correctly linked to the accurate locations (latitude, longitude, and altitude) for each weather station (Fig. 3 and Appendix A1). This is crucial for any research, analysis and conclusions that would be drawn from the final transcribed data. The next step involved determining the total number of available weather data sheets in the archives for all 37 stations. A total of 9,885 data sheets (in a similar format to Fig. 2) were identified in the archives and checked for the presence of daily temperature and precipitation records (see Fig. 4). Additionally, the placement and order of the sheets within the bundles were verified to ensure they were correctly organized by month and year, and properly assigned to the corresponding station. These two steps are essential to prevent errors during the subsequent digitization and transcription processes, such as the loss of metadata or the incorrect assignment of data to the wrong station or date. The results from the inventory process were collected in a spreadsheet ('INERA_Stations_Data.xlsx') which is published alongside the database (see section Code and data availability).

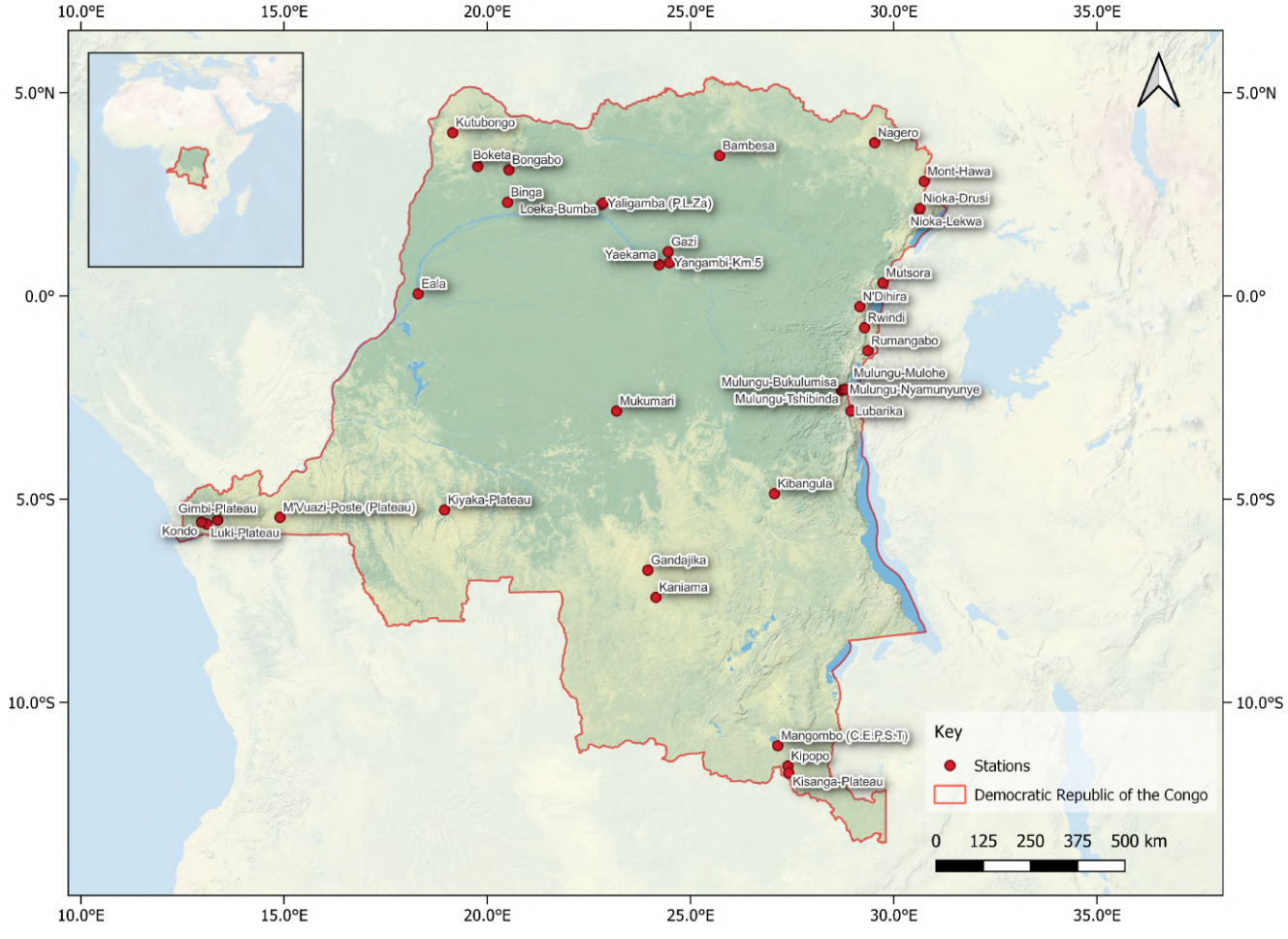


Figure 3. Map showing the location of the 37 weather stations in the DRC whose data is hosted by *INERA-Section de Climatologie* archives in Yangambi (Background map courtesy of ESRI National Geographic | Powered by Esri).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

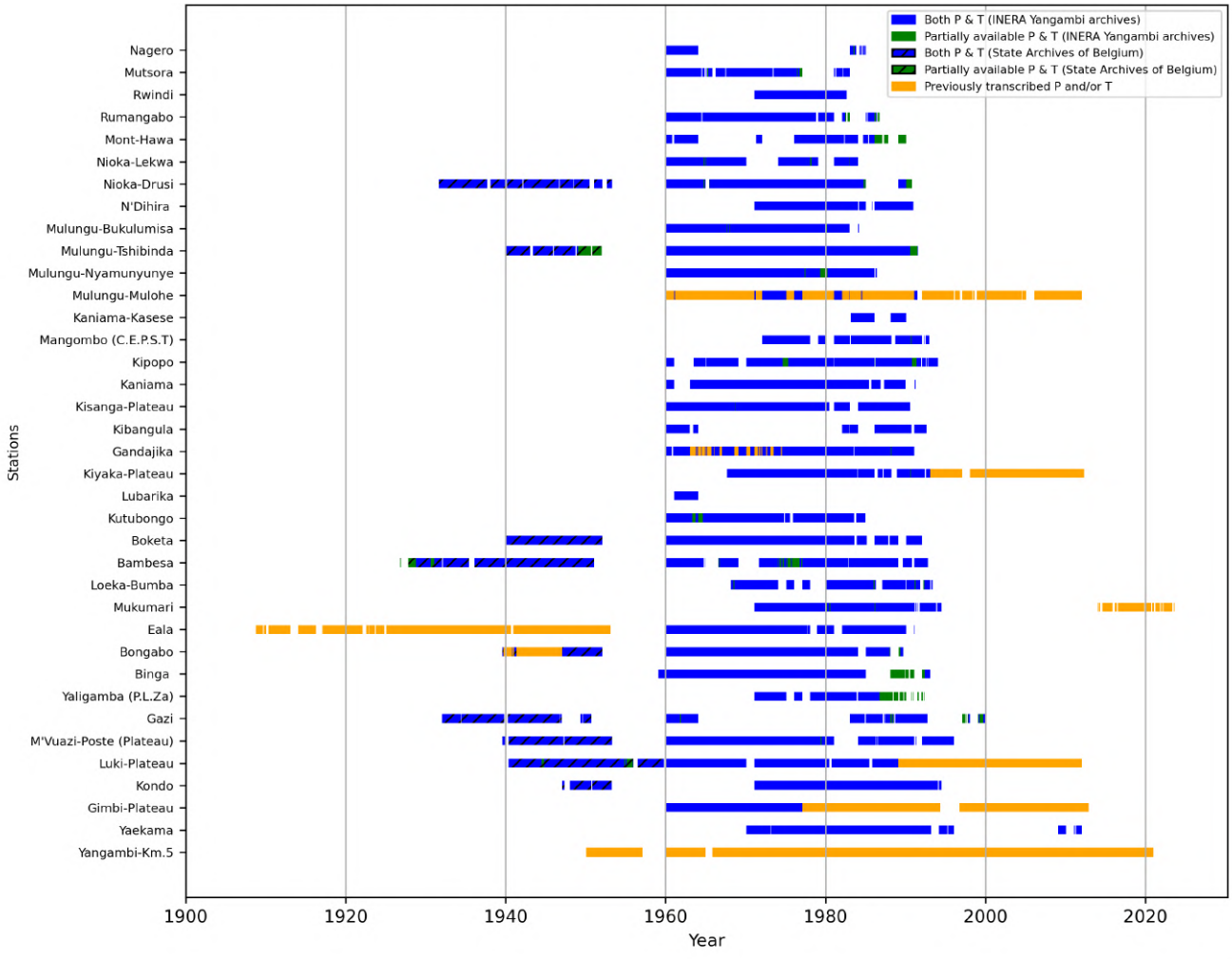


Figure 4. Inventory of available daily precipitation (P) and temperature (T) data for weather stations across the DRC, based on records from the *INERA-Section de Climatologie* archives in Yangambi and the State Archives of Belgium. Blue bars indicate months where both daily P and T data are available in hard copy at the *INERA-Section de Climatologie* archives in Yangambi; hatched blue bars indicate such availability at the State Archives of Belgium. Green markers denote periods where only one of the two variables (P or T) is available in hard copy. Orange bars represent months for which data have previously been transcribed, either by INERA, COBECORE, or as compiled in Kasongo Yakusu et al. (2023).

2.2.2 Digitization

To preserve and enable analysis of historical weather data stored at the INERA archives, we digitized all available data sheets using a high-resolution digital camera (Canon EOS 2000D with an EF-S 18-55mm IS II lens) mounted on a tripod, combined with a photo box (with a black background) equipped with uniform lighting (Fig. 5). This digitization step was essential not

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

only for the subsequent transcription steps, but also to prevent further data loss due to pests or fires, such as the recent February 2023 fires reported in one of the wings of the INERA Yangambi archives (ACP, 2023; Mamba, 2023). Given the lack of electricity at the archive site, the set-up was powered by two rechargeable batteries for the camera.

The resulting images were saved in JPG format, as it preserves high image quality while remaining an open and widely supported format. The images were organized into station-specific folders, and each image file was named using a standardized convention encoding key metadata: station code, year, month and page format. Thus, each file name follows the format `STN_YYYYMM_PF`, where `STN` is the station code, `YYYY` is the year, `MM` is the month, and `PF` is the page format (`SF` for standard format and `HD` for hand-drawn format). This naming convention ensures easy identification and retrieval of the 9,885 digital data copies.

Figure 5. Digitization campaign setup at the *INERA-Section de Climatologie* archives in Yangambi, including a digital camera mounted on a tripod and a photobox equipped with lighting fixtures to ensure high-quality digital copies of the historical weather records. Photograph taken on 15 August 2023.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

2.3 Transcription

To automatically transcribe all the digitized images of historical weather data sheets, we developed and applied a new version of MeteoSaver, version 1.1 (Muheki et al., 2026a). In the following subsections, we first briefly introduce the original software, and subsequently describe three main improvements implemented in the new release (v1.1).

2.3.1 Original MeteoSaver Description

MeteoSaver is an open-source machine learning (ML)-based software specifically developed to transcribe handwritten tabular data into machine-readable format ready for analysis (Muheki et al., 2026c). MeteoSaver takes images of tabular weather data as input and processes each image through six modules: (i) configuration, (ii) image pre-processing, (iii) table and cell detection, (iv) transcription, (v) quality assessment and quality control (QA/QC) and (vi) data formatting and upload (Muheki et al., 2026c).

In the configuration module, the software requires user-specific settings, including, but not limited to, the choice of computational environment (personal computer or High Performance Computing (HPC) infrastructure), directory paths for input and output files, and a description of the general table structure (e.g., number of rows and columns). These user settings enable the software to run smoothly across the subsequent modules. The image-preprocessing module enhances the quality of input images using adaptive threshold binarization thereby improving the performance of the subsequent table and text recognition steps. Following this, the table and cell detection module applies ML-based algorithms to detect only the tables from the images, correct for skew or tilted tables/images, and identify individual table cells. These detected cells are then subsequently passed to the transcription module, which uses a pre-trained Tesseract optical character recognition (OCR) model to recognise handwritten values. All transcribed values are then evaluated in the Quality Assessment and Quality Control (QA/QC) module, where each value is individually checked. These checks include maximum and minimum threshold tests, logical consistency checks, inter-variable relationship checks, checksum validation, and outlier identification. Finally, QA/QC 'confirmed' transcribed data are formatted per station into both the Station Exchange Format (SEF), as prescribed by the Copernicus Data Rescue Service, and the Comma-Separated Values (CSV) format, making the data ready for upload to open-access repositories. For a detailed description of the software, refer to Muheki et al. (2026c).

In the following subsections, we describe three main improvements implemented in the new release (v1.1).

2.3.2 Improved Table Detection

As an improvement to the Table and Cell Detection module in MeteoSaver v1.0, version 1.1 introduces more precise cell detection by using the pre-existing horizontal and vertical line detection based on structuring elements generated with the OpenCV (Open Source Computer Vision) library (Bradski and Kaehler, 2008).

Within the MeteoSaver version 1.0 framework, once the table boundaries were identified using OpenCV, individual table cells were detected through further image processing. This involved dilating the inverted binary image of the table (with horizontal or vertical lines removed), thereby turning the handwritten values into white blobs on a black background. Digits in

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

close proximity, as expected for multi-digit values, were merged into a single text blob, which was then treated as an individual cell (Muheki et al., 2026c). While this approach generally performs well in identifying individual cells, it occasionally misclassified text written near boarders, and close to text in adjacent cells, as a single blob.

To address this limitation, we introduce in version 1.1 an additional step in the module that re-erases the pre-identified vertical and horizontal lines after text blob identification. This serves as a secondary check to avoid merging of text blobs across different cells.

Additionally, we have optimized the overall table and cell detection process in our framework by using arrays to store the detected text components and noise features (e.g., dots), which reduces this module's processing time by more than half.

2.3.3 Improved Transcription

Building upon the transcription module in MeteoSaver v1.0, we further train the Tesseract optical character recognition (OCR) model on a broader range of handwriting styles to increase its accuracy in text recognition. This is achieved by feeding the tesstrain model with ground truth data in the form of images and their corresponding transcriptions as text files (Tesseract OCR, 2025a).

With every MeteoSaver v1.1 run, we generate an extensive training database from the INERA handwritten monthly data sheets, comprising more than 5 million clipped individual cell images in total and their corresponding confirmed transcriptions (post QA/QC checks), where available. By iteratively adding newly confirmed transcribed cell images, and their respective values, from preceding runs to this database, and retraining the model using tesstrain, the Tesseract OCR model continuously learns new handwriting styles across the dataset. This improvement makes use of Tesseract's character pattern recognition capabilities (Tesseract 3) and its Long Short-Term Memory (LSTM) neural network for line-based recognition (Tesseract 4) (Tesseract OCR, 2025b). For example, after five consecutive runs of MeteoSaver over the entire INERA database, the total number of QA/QC confirmed transcribed temperature and precipitation values increased from 321,456 to 1,019,153, corresponding to a 215% increase in confirmed values, highlighting the iterative learning capability of the software.

2.3.4 QA/QC extension to additional weather variables

Muheki et al. (2026c) demonstrates the use of the QA/QC module in MeteoSaver v1.0 to perform quality checks on transcribed daily maximum, minimum, and average temperatures (*Tx*, *Tn*, and *Tavg*). In version 1.1, we extended the QA/QC module to other recorded variables on the INERA sheets, namely: daily total precipitation (P, mm day^{-1}), and dry bulb temperature (T, °C), wet bulb temperature ($T'a$, °C), actual vapour pressure (e, hPa), relative humidity (U, %) and vapour pressure deficit (Δe, hPa) recorded at three times of the day (06:00h, 15:00h, and 18:00h local time) (Fig. 2). The following logic checks are now included in the QA/QC module.

Similar to the checks performed on the transcribed values for *Tx*, *Tn*, and *Tavg* in v1.0 (Sect. 2.3.1), the latest version applies user-defined (i.e., region-specific) maximum and minimum temperature thresholds to instantaneous T and $T'a$ as a first check (Muheki et al., 2026c). The multi-day (pentad) totals and averages are then calculated for transcribed T and $T'a$ values within these thresholds, as well as for P, and compared with their corresponding transcribed pentad totals and averages. If

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

the transcribed pentad total and/or average is equal to (or, for *T* and *T'a*, within the set uncertainty margin of 0.2°C of) the calculated pentad total and/or average, all the respective transcribed daily values for *T*, *T'a*, and *P*, leading up to that pentad total and/or average, are marked as confirmed. An uncertainty margin of 0.2°C is applied to temperature checks to account for potential human rounding errors in the original pentad averages or *Tavg* values. Such rounding may result in small discrepancies (typically < 0.2 °C) between the transcribed and QA/QC-calculated values.

In addition, logic checks are carried out on the transcribed values for *T*, *T'a*, *e*, *U*, and Δe for each day as follows:

(i) *T* must be greater than or equal to *T'a*. Values that do not meet this condition are flagged as inconsistent.

(ii) We calculate *U* (using Eq. 4), and compare it with the transcribed *U*. If the transcribed *U* is equal to (or within the set uncertainty margin of 0.2%) the calculated *U*, we mark the transcribed *U*, *e* and Δe values for that day and time as confirmed.

$$U = \frac{e}{e_s} * 100 \tag{1}$$

where, *e* is the actual vapour pressure and e_s is the saturation vapour pressure. Given that e_s values are not recorded in the DRC sheets, we derive them using *e* and Δe as follows:

$$\Delta e = e_s - e \tag{2}$$

Rearranging Eq. 2, e_s can be expressed in terms of Δe and *e* as:

$$e_s = \Delta e + e \tag{3}$$

Thus, by substituting terms in Eq. 1, we calculate *U* in terms of Δe and *e* as:

$$U = \frac{e}{\Delta e + e} * 100 \tag{4}$$

(iii) Following this, if the *U*, *e* and Δe values at a particular time are confirmed, 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.

$$e_s = 6.112 exp \frac{17.67T}{T + 243.5} \tag{5}$$

By changing the subject to *T* and substituting e_s,

$$T = \frac{243.5 * ln(\frac{\Delta e + e}{6.112})}{17.67 - ln(\frac{\Delta e + e}{6.112})} \tag{6}$$

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

(iv) If both the transcribed values of T and $T'a$ at a particular time and day are already marked as confirmed from their pentad totals and/or averages, we calculate the e using the psychometric equation from the Smithsonian meteorological tables (List, 1966) (Eq. 8) and subsequently the U and Δe values for that day and time are calculated (using Eqs. 1 - 5).

(v) However, if only one of the two transcribed values (T and $T'a$) at a particular time and day is marked as confirmed by either the pentad total or average, the confirmed value is used to check the relationship between the transcribed e, U and the other of the two (applying the Eqs. 1 - 9). If the calculated value of either T or $T'a$ matches the transcribed value (or within the uncertainty margin of 0.2°C), the respective transcribed value is marked as confirmed. This relationship check indirectly checks the transcribed U and Δe values and corrects the incorrectly transcribed one if the other is marked as confirmed.

$$e_s(T'a) = 6.112 \exp\left(\frac{17.67\,T'a}{T'a + 243.5}\right) \tag{7}$$

$$e = e_s(T'a) - \gamma\,(T - T'a) \tag{8}$$

where:

$$\gamma = 0.00066\,(1 + 0.00115\,T'a)\,p_0 \tag{9}$$

and p_0 is the standard atmospheric pressure at mean sea level (1013.25 hPa)

2.3.5 Application of MeteoSaver to the DRC data

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% for relative humidity values. 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).

Although most of the tables in our dataset share a similar structure, applying the software to such a large and diverse dataset, characterized by multiple handwriting styles, different paper sizes, varying paper quality and maintenance conditions, demonstrates its robustness and flexibility.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

The final output consists of QA/QC confirmed daily temperature and precipitation time series formatted in the Station Exchange Format (SEF), ready for subsequent analysis. This resulting dataset has been uploaded to the Copernicus Global Land and Marine Observations Database, where it will be made available to users via the Copernicus Climate Change Service Data Store, as well as on Zenodo (Muheki et al., 2026b).

2.4 Statistical Analysis

2.4.1 Validation

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.

We assess the accuracy of the automatic transcription by calculating the percentage of correctly transcribed daily temperature values (*Tx*, *Tn*, and *Tavg*) relative to their respective manually transcribed values. This comparison assumes that the manual transcriptions are accurate and therefore does not account for potential errors in the manually transcribed data itself. Similar to Muheki et al. (2026c), we apply an uncertainty margin of 0.2°C during this comparison. In cases where a value is incorrectly transcribed, we also examine the manually transcribed values for the preceding and following days to identify potential one-day shifts caused by row or cell misalignment during table and cell detection. We, therefore, present both the percentage of correctly transcribed values (hereafter referred to as the match rate) and the percentage of values correctly transcribed but shifted by ± 1 day (hereafter referred to as the one-day shift percentage).

Finally, we compute the Mean Absolute Error (MAE) of the automatically transcribed values, using the corresponding manually transcribed values on the same day as the reference.

2.4.2 Trend Analysis

We undertake trend analysis of *Tx*, *Tn*, *Tavg*, and *P* values. The analysis primarily covers the period from 1960 to the 1990s, with 11 stations with data pre-1960 and 6 stations extending into the 2000s.

It is important to note that some stations are excluded from this analysis due to substantial data gaps, incomplete years, or insufficient quality in the transcribed temperature records. Specifically, stations with fewer than 25 years of data and/or with significant missing periods (> 5 years) that would bias trend estimation are omitted. 21 of the 37 stations are considered for trend analysis in the DRC (Table 3).

To assess long-term changes in temperature, we apply the widely used Theil–Sen estimator (Wilcox, 2001) to determine multi-decadal trends in *Tx*, *Tn*, and *Tavg* for each station across its respective period of record. The Theil–Sen estimator was applied directly to the available time series, with missing values ignored in the trend estimation. Here, we express trends in degrees Celsius (°C) per decade. Additionally, we apply spatial interpolation of station-based trends using a Radial Basis Function (RBF) approach, a commonly used method for spatial interpolation of in situ meteorology data (Jaczewski et al.,

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

2025; Ryu et al., 2024), to visualize their spatial patterns across the DRC. The interpolation is performed using stations with a minimum record length of 25 years within the most overlapping observation window and is constrained to the DRC boundary.

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.

Furthermore, we investigate the changes in daily precipitation return periods at stations with records spanning over 30 years. For each station, we split the daily precipitation time series into an earlier and a later half by dividing the record at the midpoint of the valid daily observations, so that both halves contain a similar number of data points. For each half, we estimate empirical return-period curves using the Weibull plotting position, focusing on the upper tail of the distribution, to highlight changes in heavy precipitation events. We plot these curves on the same figure to visualise how the intensity of heavy precipitation events has changed between the earlier and later periods.

Lastly, we compare the observed temperature trends at each station with those derived from the ERA5-Land dataset. ERA5-Land provides hourly 2-meter air temperature data over land at a spatial resolution of 0.1° x 0.1° (approx. 9km), from 1950 to present (Muñoz Sabater, 2019). We derive daily *Tx*, *Tn*, and *Tavg* from the ERA5-Land hourly dataset. For each station, we extract ERA5-Land values from the grid cell whose center lies closest to the station coordinates, using the nearest-neighbor method. To ensure direct comparability, we compute the trends in ERA5-Land only over time instances where both the station observations and the corresponding ERA5-Land timeseries contain data.

3 Results

3.1 Database characteristics

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. 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. The dataset spans the period from 1908 to 2023. The longest station record is from Eala (1908–1991, with eight missing years), while the shortest record is from Lubarika (1961–1963). Most station records cover the 1960s through the 1990s (Tables 1 and 2, and Figs. 3 and 4). The percentage of QA/QC confirmed values out of the total transcribed values for the daily temperature records ranges from 29.1-97.5% across

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

the 37 stations, with a mean of 61%. This percentage is lower for daily precipitation records ranging from 10.3-89.4% with a mean of 30% across the stations. This difference reflects the more extensive QA/QC procedures applied to the transcribed daily temperature data. 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.

3.2 Validation

We compare the automatically transcribed daily temperature values (*Tx*, *Tn*, and *Tavg*) from five stations with their corresponding manually transcribed values produced by INERA. The validation results indicated that across these stations, the automatically transcribed daily temperature records had different accuracies across the stations ranging from 81.2% to 92.3%, with a median match rate of 87.4% and a median one-day shift percentage of 14.7% across all the five stations (Fig. 6).

Finally, we compute the mean absolute error (MAE) of the automatically transcribed values relative to the manually transcribed values on the same day. The MAE across the transcribed sheets of these five stations ranges from 0.18-0.51°C, with a median of 0.35°C (Fig. 6).

3.3 Trend Analysis

3.3.1 Warming Trends and Shifting Temperature Distributions

Of the 21 stations considered in the trend analysis (see Sect. 2.4.2), 16 exhibit an increasing trend in daily maximum temperature (*Tx*) over their respective observation periods (Table 3, Fig. 7, and Appendix Figs. D1-D5), with rates ranging from 0.14 to 0.51°C/decade (equivalent to 0.4–1.5°C over 30 years, with a median of 0.8°C warming over this period). Four stations show no change in *Tx*, while one record shows a decreasing trend of –0.20°C/decade (–0.6°C over 30 years). Across all 21 stations, the median *Tx* trend is 0.22°C/decade which corresponds to approximately 0.7°C of warming over a 30-year period.

For daily minimum temperature (*Tn*), 13 stations show an increasing trend of 0.03–0.33°C/decade (0.1–1°C over 30 years, with a median of 0.6°C), six show no change, and two record decreases of –0.21 and –0.44°C/decade (–0.6 and –1.3°C over 30 years, respectively). The median *Tn* trend across all the stations is 0.10°C/decade, equivalent to 0.3°C increase over 30 years (Table 3, Fig. 7, and Appendix Figs. D1-D5).

Average daily temperature (*Tavg*) increases at 18 stations, with trends of 0.05–0.41°C/decade (0.2–1.2°C over 30 years, with a median of 0.5°C). Two stations show no change in *Tavg*, and one shows a decline of –0.14°C/decade (–0.4°C over 30 years). Across these stations, the median Tavg trend is 0.15°C/decade, corresponding to approximately 0.5°C of warming across a 30-year period.

Overall, at 20 of the 21 assessed stations, at least one of *Tx*, *Tn*, or *Tavg* exhibits a positive trend, i.e., at each of these stations, at least one of the three variables shows an increase (Table 3, Fig. 7, and Appendix Figs. D1-D5). Spatial interpolation of

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Table 1. Summary of station data sheets transcribed using MeteoSaver v1.1, including also already manually transcribed data.

No.	Station name	Station No.	Latitude	Longitude	Altitude	Start Date	End Date	% Data Gap	% Confirmed values			
									Tx	Tn	Tavg	P
1	Yangambi-Km.5	001	N 0°46’	E 24°29’	485	Jan 1950	Nov 2020	5.3	96.6	97.5	97.0	89.4
2	Yaekama	004	N 0°46’	E 24°14’	388	Jan 1970	Dec 2011	36.5	54.8	55.1	55.6	10.6
3	Gimbi-Plateau	101	S 5°31’	E 13°22’	480	Jan 1960	Oct 2012	4.6	81.0	82.3	81.2	81.2
4	Kondo	103	S 5°34’	E 12°58’	230	Jan 1971	May 1994	2.8	56.0	55.2	56.6	19.5
5	Luki-Plateau	104	S 5°37’	E 13°06’	350	Jan 1960	Dec 2011	7.6	82.4	84.9	82.6	18.5
6	M’Vuazi-Poste (Plateau)	106	S 5°27’	E 14°54’	505	Jul 1939	Dec 1995	11.3	50.8	51.2	51.3	20.3
7	Gazi	201	N 1°05’	E 24°27’	456	Jan 1932	Nov 1999	62.9	33.0	34.3	34.4	10.3
8	Yaligamba (P.L.Za)	202	N 2°17’	E 22°51’	435	Jan 1971	Apr 1992	19.9	56.5	58.7	58.1	29.8
9	Binga	203	N 2°18’	E 20°30’	400	Jan 1959	Dec 1992	12.5	33.6	33.4	34.2	16.7
10	Bongabo	204	N 3°06’	E 20°32’	450	Jul 1939	Aug 1989	37.9	57.7	57.4	58.5	35.2
11	Eala	205	N 0°03’	E 18°18’	350	Sep 1908	Jan 1991	16.1	79.4	79.1	79.3	66.8
12	Mukumari	206	S 2°50’	E 23°11’	535	Jan 1971	Jul 2023	41.4	61.9	62.2	62.5	33.8
13	Loeka-Bumba	209	N 2°15’	E 22°49’	430	Feb 1968	Apr 1993	21.1	54.0	53.8	54.9	16.0
14	Bambesa	301	N 3°27’	E 25°43’	621	Jan 1960	Sep 1992	15.8	57.7	58.3	58.4	14.5
15	Boketa	302	N 3°11’	E 19°46’	475	Jan 1960	Dec 1991	6.8	47.9	48.4	49.1	24.4
16	Kutubongo	304	N 4°01’	E 19°09’	550	Jan 1960	Nov 1984	2.7	68.4	69.3	69.7	22.6
17	Lubarika	401	S 2°50’	E 28°57’	980	Jan 1961	Dec 1963	0.0	75.5	77.9	76.1	23.5
18	Kiyaka-Plateau	402	S 5°16’	E 18°57’	735	Aug 1967	Mar 2012	5.4	70.0	67.1	66.2	52.9

where: Tx = daily maximum temperature, Tn = daily minimum temperature, Tavg = daily average temperature, and P = precipitation. Altitude is given in meters above sea level (m a.s.l.). Confirmed values are calculated with an uncertainty margin of 0.2°C.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Table 2. (continued from Table. 1). Summary of station data sheets transcribed using MeteoSaver v1.1, including also already manually transcribed data.

No.	Station name	Station No.	Latitude	Longitude	Altitude	Start Date	End Date	% Data Gap	% Confirmed values			
									Tx	Tn	Tavg	P
19	Gandajika	405	S 6°45’	E 23°57’	780	Jan 1960	Dec 1990	0.8	51.7	50.8	50.7	30.4
20	Kibangula	406	S 4°52’	E 27°04’	685	Jan 1960	Jul 1992	63.9	72.2	73.9	73.7	37.3
21	Kisanga-Plateau	501	S 11°44’	E 27°25’	1187	Jan 1960	Jun 1990	5.2	51.4	44.0	51.7	25.0
22	Kaniama	503	S 7°25’	E 24°09’	949	Jan 1960	Feb 1991	12.6	29.1	29.2	29.2	22.4
23	Kipopo	504	S 11°34’	E 27°24’	1300	Jan 1960	Dec 1993	11.8	40.5	32.9	41.6	24.3
24	Mangombo (C.E.P.S.T)	506	S 11°04’	E 27°09’	1128	Jan 1972	Nov 1992	13.9	62.0	61.3	62.4	26.3
25	Kaniama-Kasese	508	-	-	-	Feb 1983	Dec 1989	30.1	59.8	62.0	59.7	33.8
26	Mulungu-Mulohe	601	S 2°18’	E 28°47’	1731	Jan 1960	Jan 2001	3.0	91.4	90.4	91.0	83.7
27	Mulungu-Nyamunyunye	602	S 2°18’	E 28°48’	1703	Jan 1960	Apr 1986	0.6	47.9	48.4	49.0	21.5
28	Mulungu-Tshibinda	603	S 2°19’	E 28°45’	2055	Jan 1960	Jun 1991	0.0	71.3	70.8	71.7	71.7
29	Mulungu-Bukulumisa	605	S 2°20’	E 28°43’	2378	Jan 1960	Jan 1984	4.5	73.1	73.1	72.6	22.2
30	N’Dihira	607	S 0°16’	E 29°10’	2190	Jan 1971	Nov 1990	5.0	65.5	64.5	65.5	24.2
31	Nioka-Drusi	701	N 2°09’	E 30°39’	1678	Jan 1960	Sep 1990	14.4	60.6	60.6	62.3	19.4
32	Nioka-Lekwa	702	N 2°07’	E 30°38’	1677	Jan 1960	Dec 1983	25.0	60.9	61.4	62.9	21.7
33	Mont-Hawa	703	N 2°49’	E 30°45’	1350	Jan 1960	Dec 1989	45.8	58.5	58.5	59.2	19.2
34	Rumangabo	903	S 1°21’	E 29°22’	1620	Jan 1960	Aug 1986	13.8	46.1	47.2	47.6	22.7
35	Rwindi	904	S 0°47’	E 29°17’	1040	Jan 1971	Jul 1982	0.0	69.2	69.4	70.0	27.9
36	Mutsora	905	N 0°19’	E 29°44’	1330	Jan 1960	Dec 1982	22.5	55.3	55.9	56.5	17.3
37	Nagero	906	N 3°46’	E 29°32’	740	Jan 1960	Dec 1984	78.7	57.1	57.9	57.8	18.6

where: Tx = daily maximum temperature, Tn = daily minimum temperature, Tavg = daily average temperature, and P = precipitation. Altitude is given in meters above sea level (m a.s.l.). Confirmed values are calculated with an uncertainty margin of 0.2°C.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

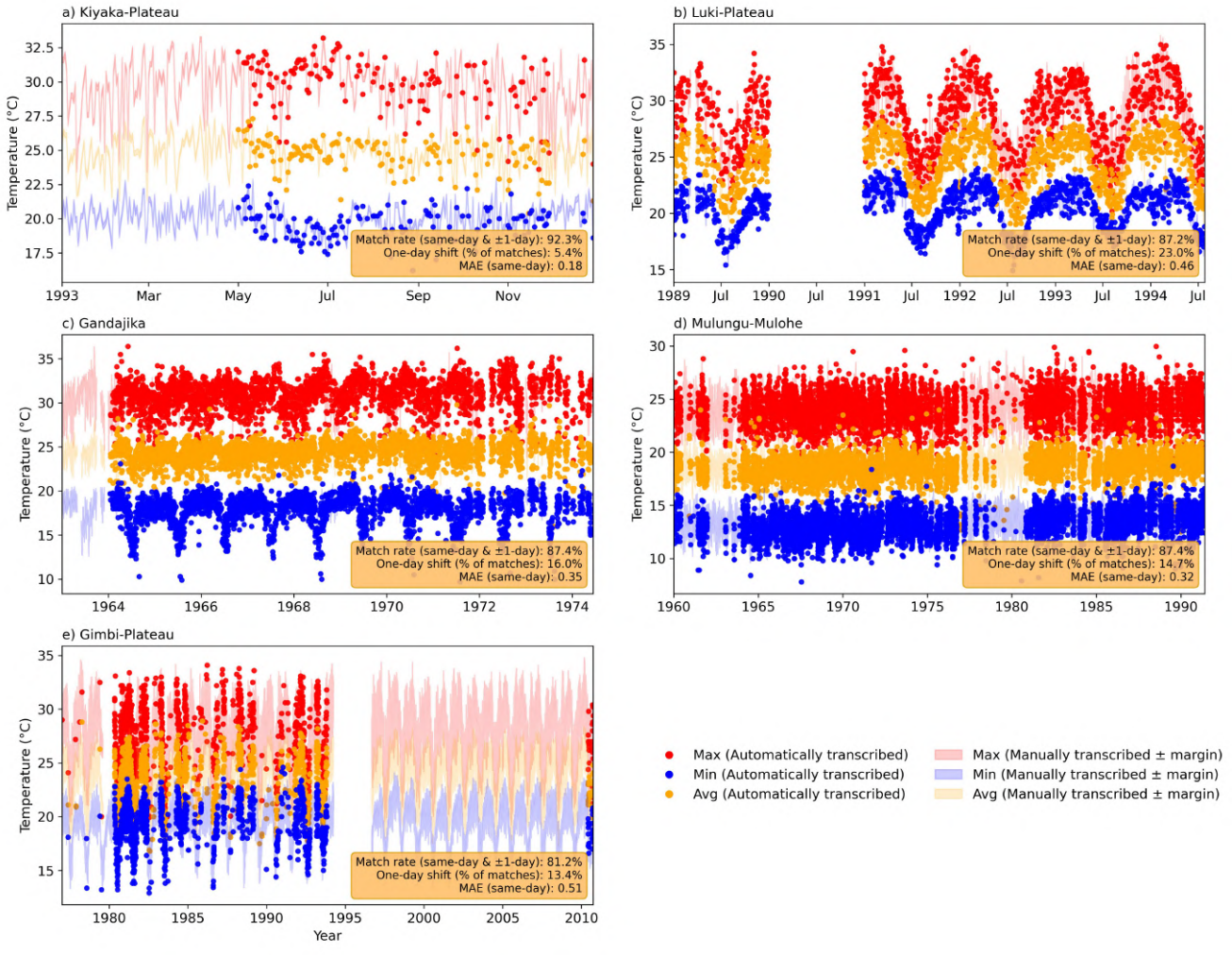


Figure 6. Time series plots of the daily maximum (red), average (orange), and minimum (blue) temperatures for stations: a) Kiyaka-Plateau (S 5°16', E 18°57'), b) Luki-Plateau (S 5°37', E 13°06'), c) Gandajika (S 6°45', E 23°57'), d) Mulungu-Mulohe (S 2°18', E 28°47'), and e) Gimbi-Plateau (S 5°31', E 13°22'). Automatically transcribed values are shown as solid markers, while manually transcribed values by INERA are illustrated as lighter time series bands, with a 0.2°C uncertainty margin applied during QA/QC checks (as done in Muheki et al. (2026c)). The match rate of the automatic transcription relative to the manual transcription, the one-day shift percentage (±1 day), and the mean absolute error (MAE) are indicated in the lower right corner of each panel.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

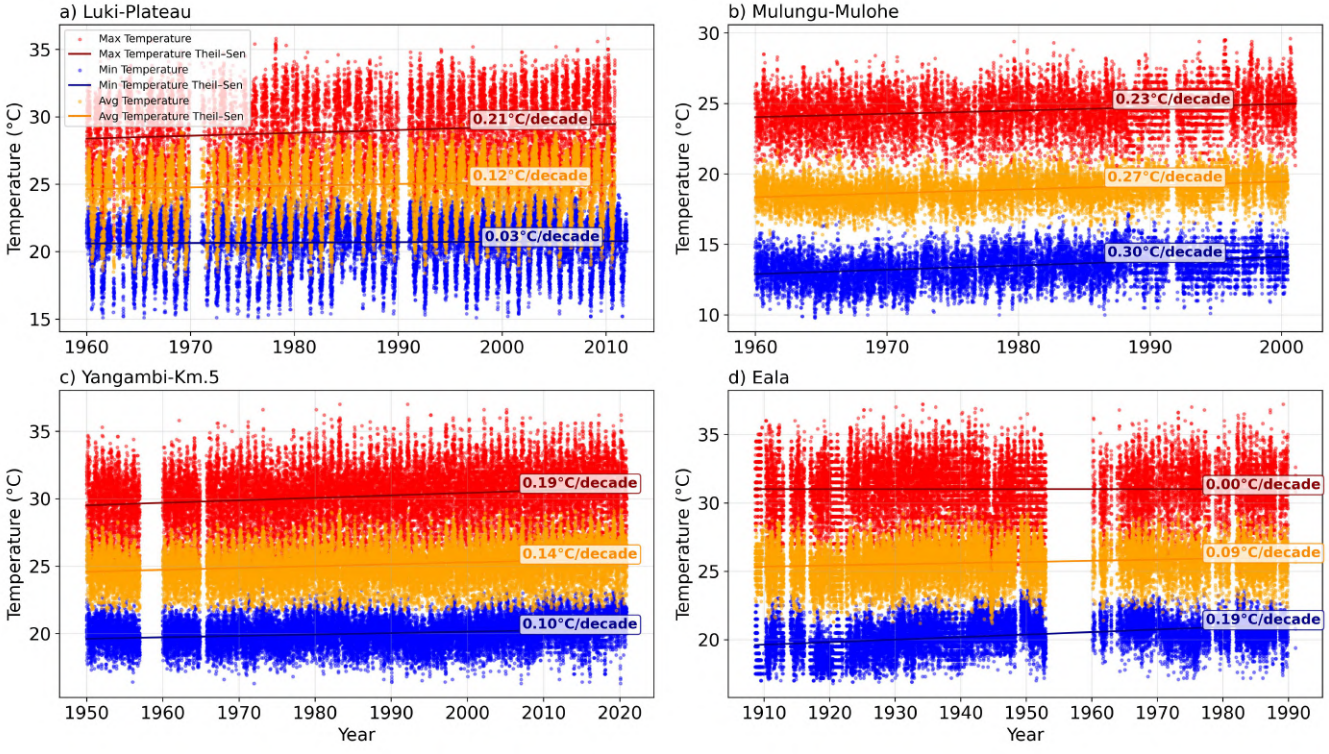


Figure 7. Daily temperature trends at stations: a) Luki-Plateau (S 5°37', E 13°06'), b) Mulungu-Mulohe (S 2°18′, E 28°47′), c) Yangambi-Km.5 (N 0°46', E 24°29'), and d) Eala (N 0°03', E 18°18') in the DRC. Colored dots represent daily values of maximum (*Tx*, red), average (*Tavg*, orange), and minimum (*Tn*, blue) temperatures. Solid lines indicate Theil–Sen trend estimates for each variable, with the corresponding slope (°C/decade) shown at the end of each line. This data includes both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. Note that these stations have different start and end years.

these temperature trends at the stations during the most overlapping observation window (1961–1990) illustrates a widespread warming signal throughout the network (Fig. 8).

We then examine the decadal temperature distributions both at individual stations and for the aggregated dataset across stations. For the latter, the analysis focuses on the most overlapping observation window (1961–1990) across all the stations.

In recent decades, the distributions of *Tx*, *Tn* and *Tavg* have shifted toward higher values, reflecting warmer modal temperatures and higher 95th percentile values. This pattern is observed at both individual stations (Fig. 9a-d and Appendix Figs. E1-E5) and in the aggregated dataset for the DRC (Fig. 10). For example, at Station Mulungu-Mulohe, the mode of *Tx* and *Tn* during the last two decades (1981-1990 and 1991-2000) is approximately 1°C higher than in the first decade (1961–1970) (Fig. 9b). Similarly, the 95th percentile values of *Tx* and *Tn* at this station are about 1°C higher in the last decade (1991-2000) compared to the first decade (1961-1970) (Fig. 9b).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Table 3. Daily temperature trends at 21 stations selected for the climate analysis in the DRC. These trends are estimated using the Theil–Sen estimator (Wilcox, 2001). The data used to estimate these trends includes both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA, and through the COBECORE project. Start and end years indicate the period of available observations for each station. Missing years represent the number of years without data within this period.

No.	Station name	Start year	End year	Missing years	Trend (°C/decade)		
					Tx	Tn	Tavg
1	Yangambi-Km.5	1950	2020	4	0.19	0.10	0.14
2	Yaekama	1970	2011	1	0.32	0.13	0.22
3	Gimbi-Plateau	1960	2012	2	0.22	0.11	0.17
4	Luki-Plateau	1960	2011	2	0.21	0.03	0.12
5	M'Vuazi-Poste (Plateau)	1960	1995	4	0.25	0.00	0.15
6	Binga	1959	1992	8	-0.20	0.22	0.00
7	Eala	1908	1991	8	0.00	0.19	0.09
8	Bambesa	1960	1992	3	0.14	-0.44	-0.14
9	Boketa	1960	1991	2	0.31	0.00	0.17
10	Kutubongo	1960	1984	1	0.00	0.00	0.00
11	Kiyaka-Plateau	1967	2012	2	0.00	0.19	0.12
12	Gandajika	1960	1990	1	0.20	0.00	0.05
13	Kisanga-Plateau	1960	1990	2	0.28	0.29	0.39
14	Kaniama	1960	1991	3	0.51	0.00	0.19
15	Kipopo	1960	1993	3	0.38	-0.21	0.06
16	Mulungu-Mulohe	1960	2001	1	0.23	0.30	0.27
17	Mulungu-Nyamunyunye	1960	1986	2	0.00	0.28	0.16
18	Mulungu-Tshibinda	1960	1991	1	0.48	0.33	0.41
19	Mulungu-Bukulumisa	1960	1984	2	0.34	0.00	0.11
20	Nioka-Drusi	1960	1990	5	0.43	0.06	0.22
21	Rumangabo	1960	1986	3	0.19	0.26	0.26

Where: Tx = daily maximum temperature, Tn = daily minimum temperature, and Tavg = daily average temperature

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

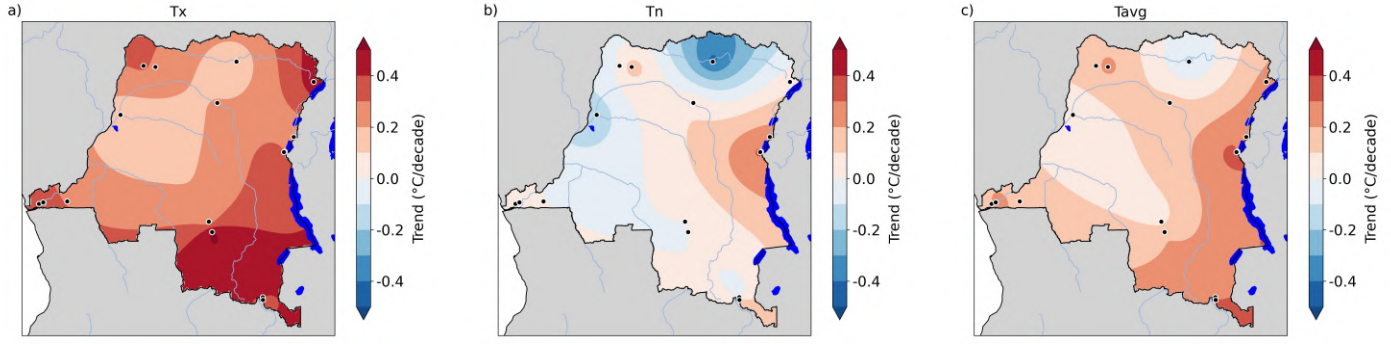


Figure 8. Spatial interpolation of temperature trends in daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures across the DRC, based on observations from 19 climate stations during the period 1961–1990. Trends are expressed in °C per decade and were derived using Theil–Sen slope estimates for each station. The interpolated fields were generated using the Radial Basis Function (RBF) method with a linear kernel for scattered data interpolation, and visualized using a continuous color scale. Station locations are indicated by black dots outlined with white circles. Lakes are visualised in dark blue. Note that this map serves as a visualization based on the limited number of available stations (n = 19) and may not fully represent the spatial variability across the region.

In the aggregated dataset, the distributions of *Tx*, *Tn*, and *Tavg* values over the overlapping window (1961-1990) shift progressively to warmer values with each successive decade (Fig. 10). The bimodal distribution of temperatures observed across all stations in the first decade (1961–1970) becomes less pronounced, evolving toward a more unimodal distribution in *Tx* and *Tavg*. This corresponds to an increase in the density of higher temperatures during the second decade (1971–1980) and an even greater increase in the last decade (1981–1990). An increase in the 95th percentile values for Tx and Tavg of just under 1 °C is also observed in the last decade compared to the first (Fig. 10).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

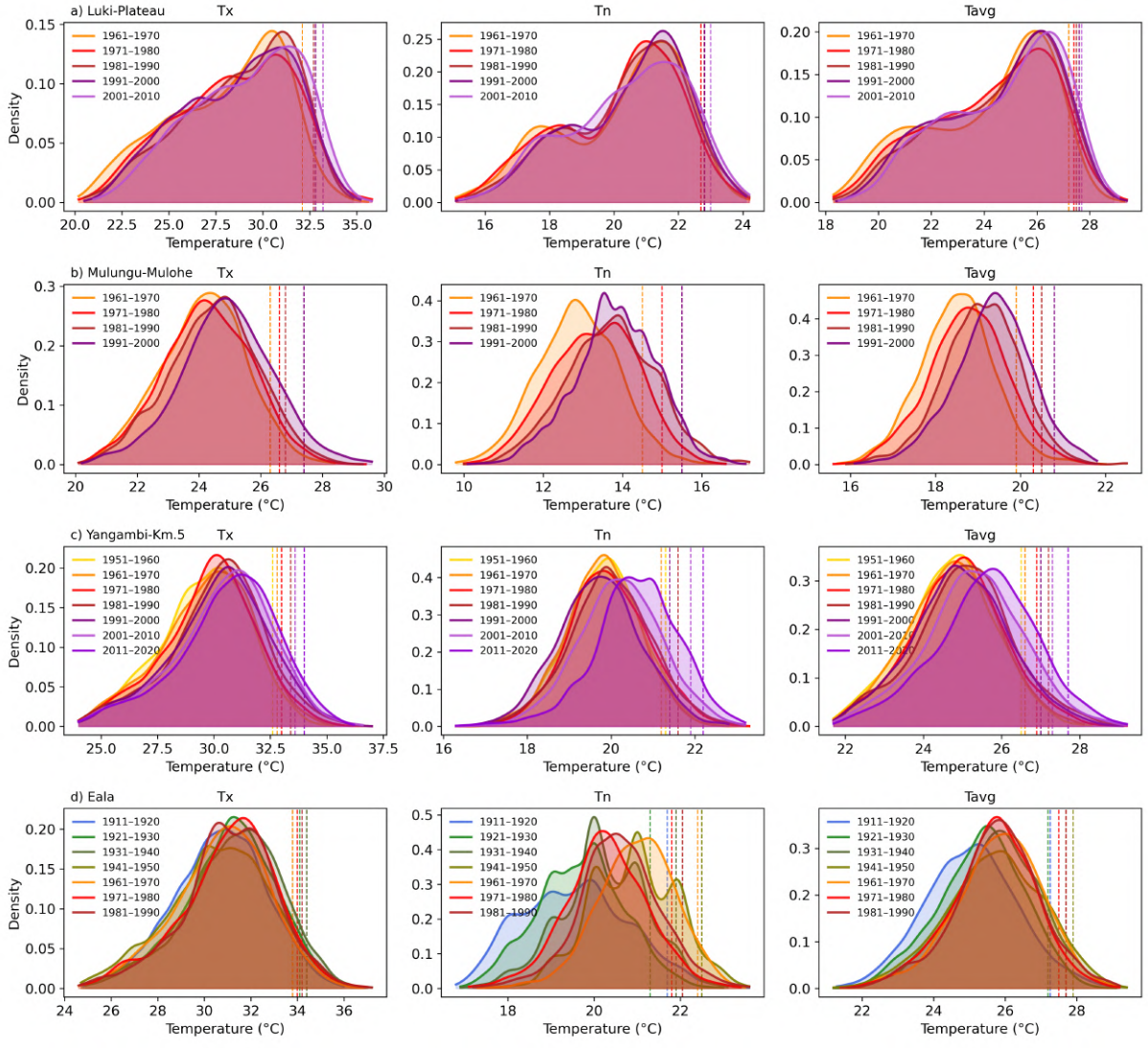


Figure 9. Decadal shifts in temperature distributions at stations: a) Luki-Plateau (S 5°37', E 13°06'), b) Mulungu-Mulohe (S 2°18′, E 28°47′), c) Yangambi-Km.5 (N 0°46', E 24°29'), and d) Eala (N 0°03', E 18°18') in the DRC. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). Vertical dashed lines indicate the 95th percentile values for each decade, with panels (a), (b) and (c) illustrating the progressive rightward shift towards higher temperatures and more extreme values over time.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

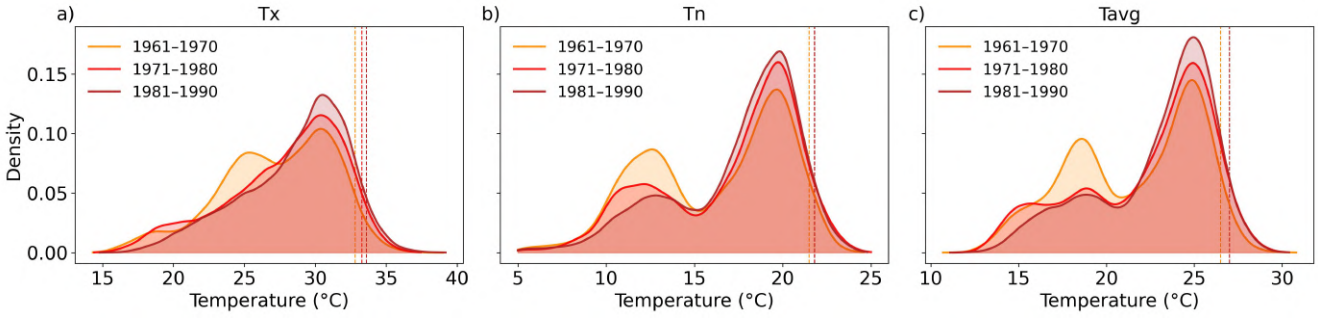


Figure 10. Decadal shifts in temperature distributions based on an aggregated dataset of 37 stations in the DRC for the period 1961 to 1990. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1961–1970 (yellow), 1971–1980 (red), and 1981–1990 (maroon). Vertical dashed lines indicate the 95th percentile values for each decade, illustrating the progressive rightward shift in temperature extremes over time.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

3.3.2 More Hot Days and Warm Nights, Fewer Cold Events

The average number of hot days per year, defined here as days with *Tx* exceeding the 95th percentile threshold, has more than tripled over the period 1961–2000 at several stations, including Luki-Plateau, Mulungu-Mulohe, and Yangambi-Km.5 (Fig. 11). Across all stations, during the most overlapping observation window (1961–1990), the number of hot days has on average increased from about 3 days per year with *Tx* above 33.3 °C (the 95th percentile threshold) in the first decade to about 5 days per year in the last decade (Fig. 12). Similarly, the average number of days with *Tavg* exceeding 26.9 °C (the 95th percentile threshold) increased from approximately 3 to 6 days per year, corresponding to a doubling over this period.

Regarding the 97.5th percentile, the most recent decade exhibits on average one additional hot day and warm night per year compared to the earliest decade. A similar increase of approximately one hot day is also observed at the 99th percentile for both *Tx* and *Tavg* (Fig. 12).

In contrast, the average number of cold days per year across all stations, defined as days with *Tx* below the 5th percentile, decreased from around 7 to 5 days per year between the first and last decades considered (Fig. 12). At more extreme thresholds (2.5th and 1st percentiles), cold days declined by approximately one day per year over the same period. Likewise, the average number of cold nights (*Tn* < 5th percentile) declined from about 5 to 4 days per year over the same period (Fig. 12).

These differences in decadal average number of hot days, warm nights and cold days are much more difficult to interpret on a station by station basis (as shown in Fig. 11 and Appendix Figs. F1-F5), as the extreme percentiles tend to be noisier, partially due to variations in time series completeness, internal climate variability, station data inhomogeneities (e.g., due to station location, instrument changes or observation practices), and possible local-scale effects (e.g., land-use and land-cover changes). However, the station-based analysis still shows a general tendency towards rising frequency of hot days/nights and declining cold events across decades, illustrating the shift in temperature distribution towards higher values. Moreover, when aggregated across all the stations in the DRC dataset (as shown in Fig. 12), we observe a clear trend of more hot days and fewer cold days across the DRC from 1960 to 1990.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

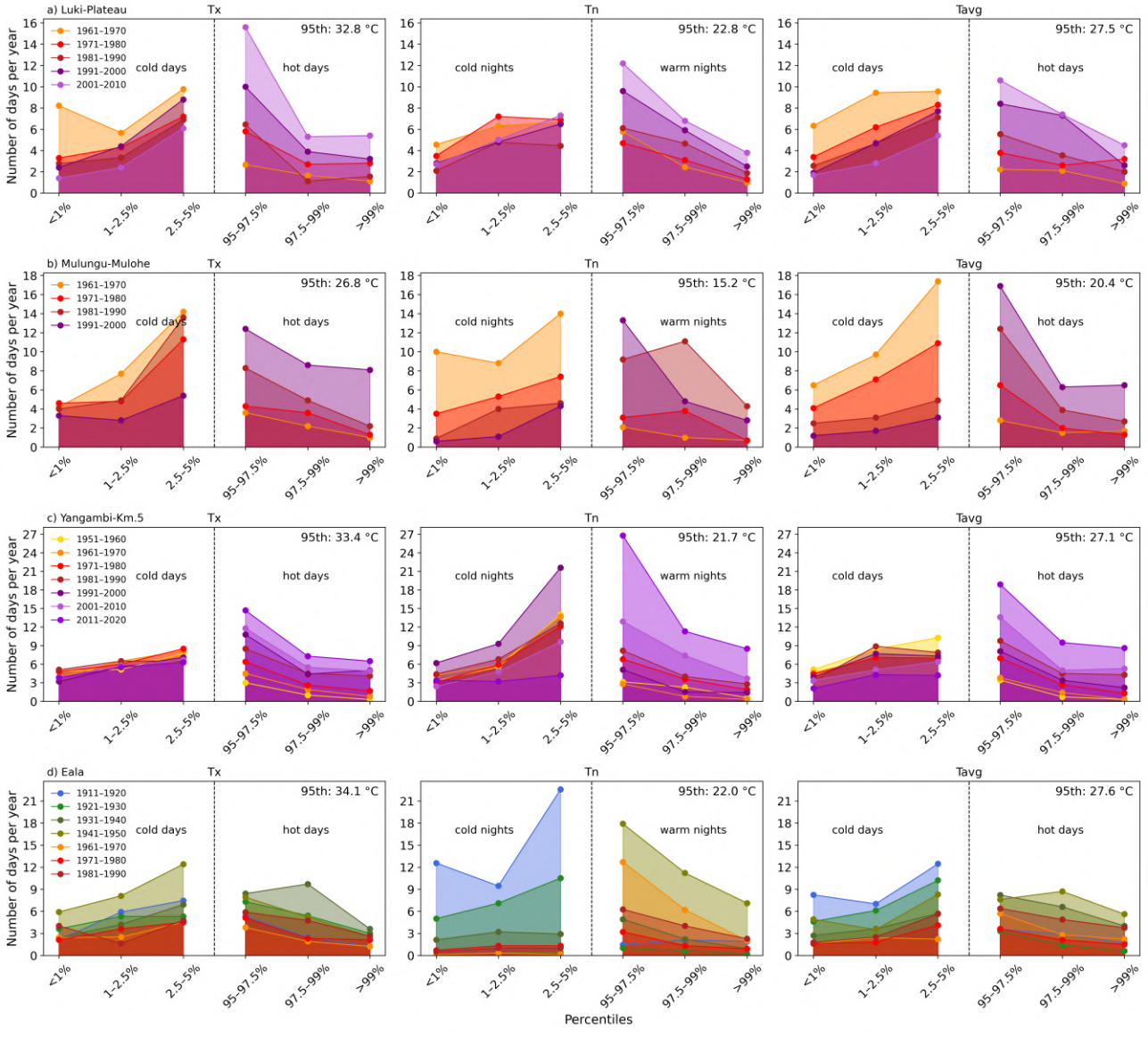


Figure 11. Frequency of hot and cold temperature extremes at stations: a) Luki-Plateau (S 5°37', E 13°06'), b) Mulungu-Mulohe (S 2°18′, E 28°47′), c) Yangambi-Km.5 (N 0°46', E 24°29'), and d) Eala (N 0°03', E 18°18') in the DRC. The plots show the average number of days per decade falling below the 5th percentile (cold events) and above the 95th percentile (hot events) for daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures, similar to Lorenz et al. (2019). Values are based on kernel density estimates calculated across the entire period. The number of days per year within each decade are shown: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). The middle 90% of the distribution (5th–95th percentile range) is omitted for clarity. The 95th percentile temperature for each variable is noted in the top-right corner of each plot.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

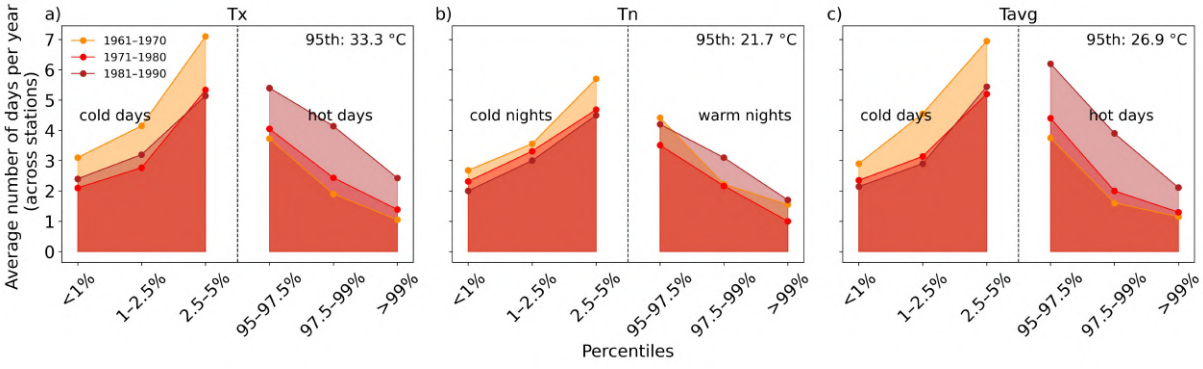


Figure 12. Frequency of hot and cold temperature extremes aggregated across all the 37 stations in the DRC from 1961 to 1990. The plots show the average number of days per decade falling below the 5th percentile (cold events) and above the 95th percentile (hot events) for daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures, similar to Lorenz et al. (2019). Values are based on kernel density estimates calculated across the entire period. The average number of days per year within each decade are shown: 1961–1970 (yellow), 1971–1980 (red), and 1981–1990 (maroon). The middle 90% of the distribution (5th–95th percentile range) is omitted for clarity. The 95th percentile temperature for each variable is noted in the top-right corner of each plot. The rising frequency of hot days/nights and declining cold events across decades illustrates the intensifying shift in temperature extremes.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

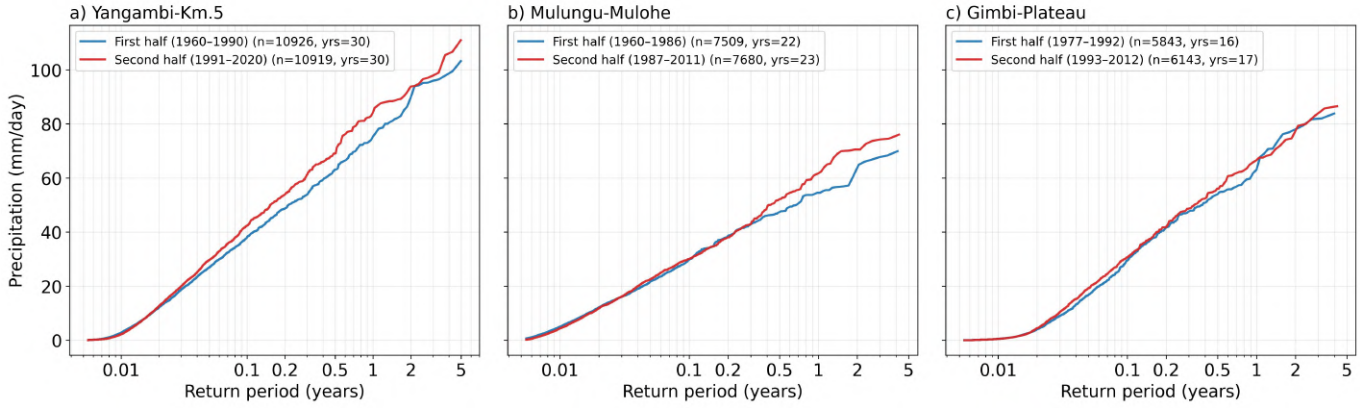


Figure 13. Changes in daily precipitation (P) return periods recorded at three stations: a) Yangambi-Km.5 (N 0°46', E 24°29'), b) Mulungu-Mulohe (S 2°18′, E 28°47′) and c) Gimbi-Plateau (S 5°31', E 13°22'). The blue lines represent the return periods of daily precipitation (in mm/day) during the first half of the available time series at each station, while the red lines represent those during the second half of the period. Note that while two of the stations have the same starting year (1960), one starts from 1977. All three stations differ in their end years.

3.3.3 Changes in Precipitation Return Periods

For precipitation we analyse three stations with more than 30 years of daily precipitation records, namely: Yangambi-Km.5, Mulungu-Mulohe, and Gimbi-Plateau. The former two stations show a clear increase in the frequency of heavy precipitation in the later periods compared to the earlier periods. For instance, at Mulungu-Mulohe, a daily precipitation amount of 63 mm/day has a return period of about 2 years in the earlier period (1960-1985), but about 1 year during in the later period (1986-2011) (Fig. 13b), indicating that such an event has become twice as frequent. Similarly, at Yangambi-Km.5, a daily precipitation amount of 83 mm/day is twice as frequent in the later period (1990-2020) than in the earlier period (1960-1989) (Fig. 13a). These increases are much more pronounced at Yangambi-Km.5 and Mulungu-Mulohe in comparison to Gimbi-Plateau where there are no substantial changes in the return periods between the two time periods (Fig. 13c). In general, precipitation events greater than 50mm/day are observed to occur more frequently in the later periods than in the earlier periods at Yangambi-Km.5 and Mulungu-Mulohe.

3.3.4 Comparison To Reanalysis

Comparison of the trends in transcribed observations of *Tx*, *Tn*, and *Tavg* with those in the ERA5-Land dataset reveals no systematic differences across the DRC (see Figs. 14-16). Although ERA5-Land generally captures the seasonal patterns well compared to the observations (as shown, for example in Fig. 14a), the trends in *Tx*, *Tn*, and *Tavg* are sometimes either underestimated or overestimated relative to the observed values, depending on the station and variable (see Fig. 14-16). These differences are typically of the same order of magnitude as the trends themselves, generally 10^{-1} °C/decade, with some dis-

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

crepancies on the order of 10^{-2} °C/decade. While most stations show the same sign of difference in trend as their nearby stations, a few exhibit opposite signs compared to their closest station, for example Kipopo and Kisanga-Plateau (see Fig. 15), further underscoring the non-systematic nature of differences between the observations and ERA5-Land, which may partly reflect non-climatic inhomogeneities in the station records.

We also observe that during the early 1950s, the ERA5-Land data displays a pronounced "jump" of approximately 4.5-5°C in temperature timeseries for *Tx*, *Tn*, and *Tavg* that is not captured by stations with observations during that period, such as Yangambi and Eala (see Figs. 14 c-d and Appendix Figs. H1 a-b). This suggests that the ERA5 reanalysis may be weakly constrained in this region during the early reanalysis period, highlighting the need for more historical observational data for this region to improve this product.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

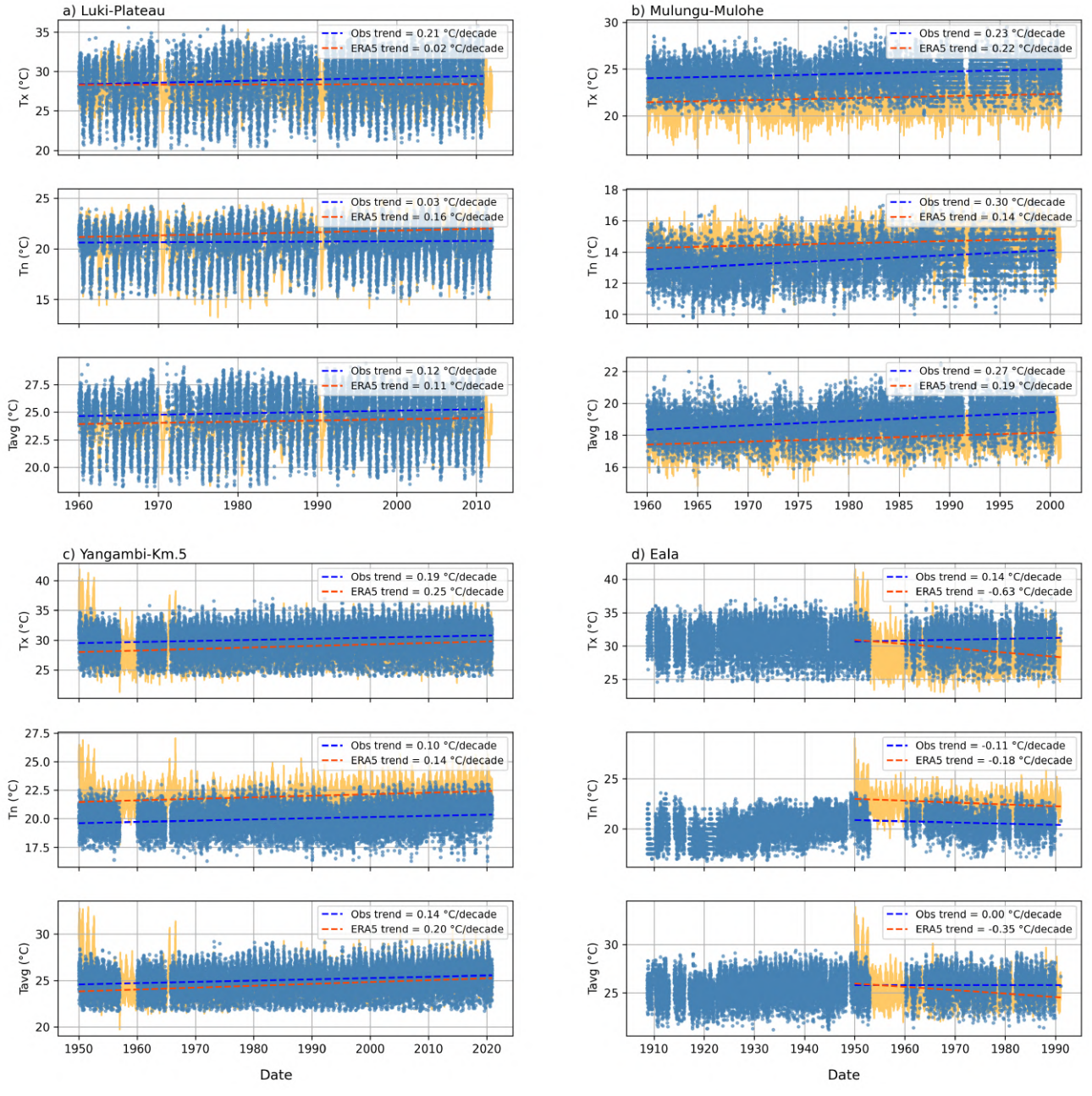


Figure 14. Comparison of trends in daily maximum, minimum and average temperatures at four example stations: a) Luki-Plateau (S 5°37', E 13°06'), b) Mulungu-Mulohe (S 2°18′, E 28°47′), c) Yangambi-Km.5 (N 0°46', E 24°29'), and Eala (N 0°03', E 18°18') in the DRC. Blue dots represent daily values of observed maximum, minimum and average temperatures, while solid yellow lines show the ERA5-Land timeseries. The dotted dark blue and orange lines indicate Theil–Sen trend estimates for each variable for observations (Obs) and ERA5-Land, respectively, with the corresponding slope (°C/decade) shown in the top left legend. The observations include both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. The trends are estimated using values where both the ERA5-Land data and transcribed observations are available. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

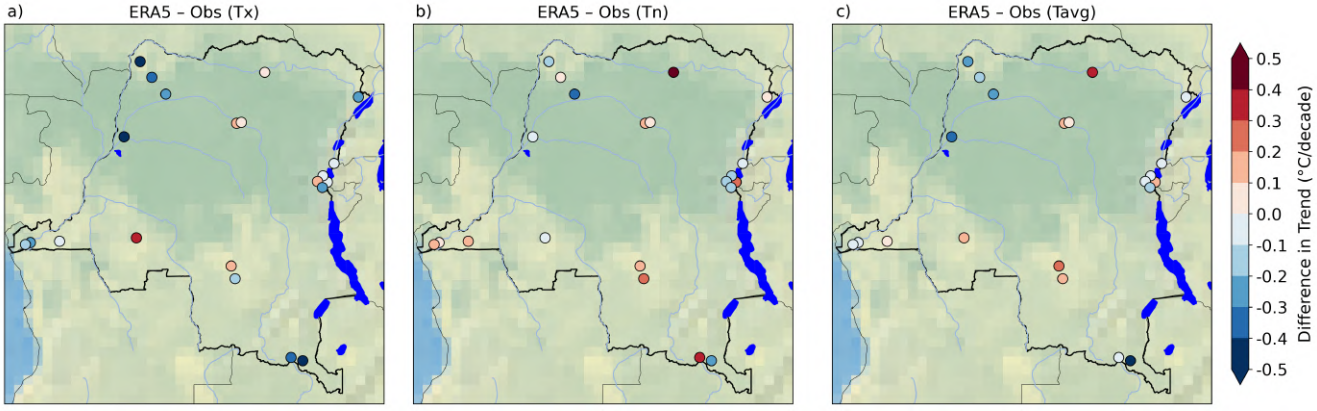


Figure 15. Map showing the differences between ERA5-Land and observed (Obs) trends in (a) daily maximum (Tx), (b) daily minimum (Tn), and (c)average (Tavg) temperatures across the DRC, based on observations from 21 climate stations and their corresponding ERA5-Land time series obtained using the nearest neighbor D. The green shading represents the Natural Earth shaded background used for geographic context (http://www.naturalearthdata.com/). Lakes are visualised in dark blue and occur primarily along the eastern border of DRC, while the Atlantic Ocean is shown in light blue to the west of the domain. Note that the stations have different start and end years.

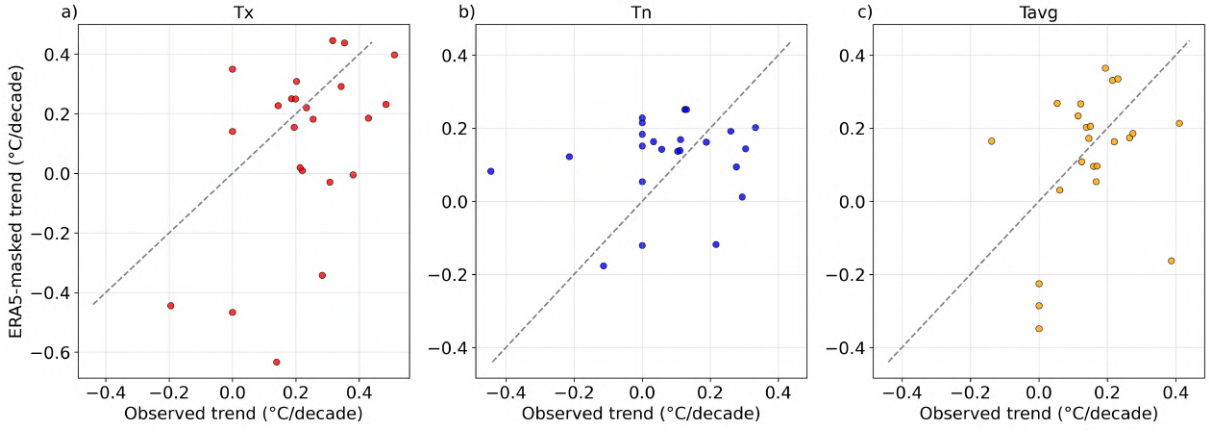


Figure 16. Graphs comparing the observed trends in (a) daily maximum (Tx), (b) daily minimum (Tn), and (c) daily average (Tavg) temperatures from 21 weather stations across the DRC with their corresponding ERA5-Land time series. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

4 Discussion

4.1 Transcription with MeteoSaver and its Limitations

In this study, we apply MeteoSaver v1.1 to transcribe 9,885 handwritten tabular historical weather data sheets from 36 INERA stations across the DRC. We obtain over one million QA/QC confirmed values for daily and sub-daily temperature and precipitation. Validation against previously manually transcribed data from five stations yields match rates ranging from 81.2% to 92.3%, with a median match rate of 87.4%, a median one-day shift percentage of 14.7%, and a median MAE of 0.35°C.

These results indicate that MeteoSaver performs robustly across diverse handwriting styles, variations in ink intensity of recorded values and table borders, paper maintenance conditions, and paper sizes, across this large dataset. An MAE of 0.35°C is of the same order of magnitude as typical observational uncertainties reported for historical thermometer measurements of 0.2°C (1σ) for stations within the latitude range of 20°S–20°N (Morice et al., 2012; Brohan et al., 2006; Folland et al., 2001). Moreover, the mismatches associated with one-day shifts (row shifts), rather than incorrect digit recognition, suggest row identification errors as one of sources of disagreement between the manually and automated transcription. Additionally, the validation checks we undertake against the manually transcribed data assume that those human transcriptions were performed accurately and therefore do not account for potential human keying errors, which have been shown to occur in manual transcription projects (Noone et al., 2024). Therefore, part of these differences observed in the validation here might be a result of the manually transcribed reference dataset. Overall, we consider the transcribed dataset sufficiently reliable for decadal-scale temperature trend analysis.

Nevertheless, we acknowledge that the following improvements need to be made to MeteoSaver to enhance its robustness and maximise the volume of transcribed data: (i) The transcription module can be improved through continued expansion of the OCR training dataset. This is because the iterative ML training strategy used in this study resulted in a 215% increase in QA/QC confirmed values after five consecutive runs over the INERA dataset. This substantial improvement illustrates the importance of feeding a large dataset of diverse handwriting images to the TesseractOCR *tesstrain* model. This approach would also improve the model for scalability, especially in the context of additional data rescue projects in data-sparse regions, through the use of multi-station datasets; (ii) Improvements to the row identification algorithm within the Table and Cell detection module would reduce the ± 1 day shift of correctly transcribed values and further improve alignment between detected values and calendar dates.

These software improvements would also increase its potential for further data rescue. For instance, while the output dataset presented here contains over one million QA/QC confirmed values for daily and sub-daily temperature and precipitation, approximately 300,000 additional temperature records and 200,000 precipitation records were transcribed but did not pass the current QA/QC checks. Engaging volunteers to review and verify such records could enhance data recovery while simultaneously generating additional training data to further refine the transcription module.

Similarly, the 225 hand-drawn sheets (as in Appendix Fig. A2), which we excluded from automatic transcription due to their structural complexity, could benefit from the improved software, as well as citizen science efforts, to expand both the training dataset and the observations dataset.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

In addition, the extensive pre-1960 climate records from 575 stations digitised (imaged) during the COBECORE project, as reported in Jacobsen et al. (2018), represent a major opportunity for further data rescue. Automatic transcription of these records could substantially expand the observational dataset, including additional meteorological variables (e.g., wind speed and direction, evaporation, and humidity). Although not all of these stations contain continuous long-term historical weather records, their transcription could potentially add millions of additional historical in-situ records for the Congo Basin.

Finally, this pipeline could also be extended to other national and international archives hosting historical climate data for the Congo Basin, such as the African Centre of Meteorological Applications for Development (ACMAD) database (Noone et al., 2024) and national meteorological and hydrological services in the region.

4.2 Warming Trends and Shifting Temperature Distributions

Twenty of the 21 stations considered in this study show an increasing trend in at least one of the *Tx*, *Tn*, or *Tavg* values. The median warming trends are 0.22°C, 0.10°C, and 0.15°C per decade for *Tx*, *Tn*, and *Tavg*, respectively, which correspond to approximately 0.7°C, 0.3°C, and 0.5°C of warming over 30 years. When scaled to multi-decadal periods, these station-based trends are of similar magnitude to the reported warming of 0.75-1.05°C since 1960 for Central Africa (Doherty et al., 2022), indicating consistency between in-situ observations and prior regional-scale assessments that did not have access to these data.

These warming trends are also reflected in the aggregated dataset across all stations. Each successive decade within the period 1961-1990 displays a shift toward higher temperature values in distributions of *Tx*, *Tn*, and *Tavg*, showing higher modal temperatures, and higher extremes, here represented by the 95th percentile values (Fig. 10). This pattern is also consistent at individual stations with longer time series extending beyond the common overlapping period i.e., stations with data before 1960 and/or continuing into the 2000s (see Fig. 9 a-d).

Together, these results show a clear and consistent warming signal across the DRC during 1961–1990, with evidence of continued warming at the few stations with records extending into the early 2000s. Given that the DRC constitutes a substantial portion of the Congo Basin (approx. 62% of the basin; (Group, 2021)), these findings are likely broadly representative of regional conditions, although they may not fully represent spatial variability across the entire Congo Basin.

However, it is important to note that, while QA/QC checks were applied to the transcribed data, homogenisation was not performed to detect and adjust for potential non-climatic inhomogeneities in the station records. Such inhomogeneities may arise from changes in station location, instruments, observation time or surrounding environment (e.g., land-use or land-cover change), and could introduce artificial (non-climatic) shifts in the temperature time series (Peterson et al., 1998). 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. Such changes can lead to localised warming effects (Zeppetello et al., 2020). Therefore, part of the distribution changes may reflect these non-climatic influences from one or more stations.

Moreover, most station records stop in the mid 1990s because of interruptions in observations during the First Congo War (1996-1997) and subsequent Second Congo War (1998-2003) that led to nationwide institutional breakdown (Weiss, 2000). This lack of sufficient data post 1990 is limiting given that recent studies report an increased global warming rate since the

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

1990s (Samset et al., 2023), suggesting that the full magnitude of recent warming in the region may not be captured within this available station data.

4.3 More Hot Days and Fewer Cold Events

Analysis of the aggregated dataset across all stations during the common overlapping period (1961-1990) reveals a clear increase in the frequency of hot days, represented here as the average number of days with *Tx* and/or *Tavg* values above the 95th percentile thresholds, with each successive decade (Fig. 12a,c). On average, the number of hot days per year nearly doubled between the earliest and most recent decades. These changes in hot extremes in the region have not been assessed in the IPCC Sixth Assessment Report (AR6), largely due to the sparsity of in situ observational data in Central Africa (IPCC, 2021a). Our results, therefore, contribute new observational evidence for changes in hot extremes in the region and underscore the importance of data rescue efforts in the region to enable robust detection and attribution studies.

In contrast, the frequency of cold days, here defined as days with *Tx* values below the 5th percentile, declined over the same period (Fig. 12 a). Taken together, the increase in frequency of hot days and the decrease in frequency of cold days provide evidence of a warming signal across the DRC during 1961–1990, consistent with the distributional changes discussed in Sect. 4.2. A similar pattern is also observed in more recent decades (2000s and 2010s) at stations with data extending beyond 1990 (Figs. 11 a-c).

Consistent with our findings, Aguilar et al. (2009) report increases in the frequency of warm extremes accompanied by decreases in cold extremes over the period 1955–2006 across Central Africa, based on observations from stations in Cameroon, the Central African Republic, São Tomé and Príncipe, the Republic of the Congo, Gabon, and eight additional stations in the DRC that are distinct from those analysed in this study. This indicates that these changes are not limited to the DRC but reflect broader regional trends across Central Africa.

4.4 Changes in Precipitation Return Periods

Unlike the transcribed daily temperature values, which were subjected to multiple QA/QC checks to confirm and correct transcription errors, the transcribed daily precipitation values were subjected to only one QA/QC check, namely the pentad total consistency check (Sect. 2.3.4). Consequently, the number of QA/QC-confirmed daily precipitation values were substantially lower than for temperature. For this reason, we analyse changes in precipitation return periods for only three stations with more than 80% QA/QC-confirmed values (Yangambi-Km.5, Mulungu-Mulohe, and Gimbi-Plateau), with the former two indicating an increase in the frequency of heavy precipitation events.

However, given that only three out of 37 stations in our dataset have more than 30 years of daily precipitation data, it is difficult to draw conclusions about the broader changes in frequency of heavy precipitation across the DRC. For this reason, our discussion mainly focuses on the changes in temperature across the region, for which we have a larger spatial dataset.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

4.5 Comparison To Reanalysis

In this study, we compare temperature trends derived from station observations with those from ERA5-Land to assess the consistency between in-situ data and the reanalysis. As this comparison involves station (point-based) observations and gridded ERA5-Land data (matched using the nearest-neighbor method), differences in absolute temperature values are expected. Therefore, we focus on comparing temporal trends over time periods where both datasets contain data.

Our results show that ERA5-Land generally captures the seasonal variability well relative to the station records and exhibits no systematic differences in trends across the DRC. However, ERA5-Land trends sometimes underestimate or overestimate the trends, depending on the station and variable. In some cases, even nearby stations show opposite signs in their differences in trend (Figs. 14 - 16). This may be partly due to non-climatic inhomogeneities in the station data (see Sect. 4.2), which were not explicitly accounted for in this analysis.

A notable discrepancy is observed in the early period of the record, where ERA5-Land exhibits a pronounced "jump" of approximately 4.5-5°C in temperature variations for *Tx*, *Tn*, and *Tavg* during the 1950s that is not evident in the station observations (Fig. 14 c and d). This likely results from the limited number of assimilated observations available to ERA5-Land during that period compared to more recent decades (Bell et al., 2021).

Overall, these results indicate that ERA5-Land captures large-scale variability and long-term warming trends in the region, particularly from the 1960s onward. However, caution should be taken when interpreting trends in the early-period (1950s), where the reanalysis may be weakly constrained. These findings underscore the importance of continued data-rescue efforts to recover and digitize archived historical observational records, which are essential for evaluating and constraining reanalysis products in data-sparse regions, especially in the early periods.

4.6 Potential of Our Data Rescue Approach and Opportunities

Throughout our study, we demonstrate our climate data rescue workflow, starting with data inventory of archived weather sheets, followed by the digitization (imaging) of these sheets, transcription using the open-source machine-learning software MeteoSaver, automated QA/QC and data formatting, and finally the upload of the rescued datasets to open-access repositories.

By documenting each step, we aim to provide a replicable workflow that can be applied to similar data rescue projects across Africa and other data-sparse regions where historical weather observations are still stored in local or international archives. These techniques can also be used by the broader hydro-climate research community to recover additional historical records such as river discharge, lake level, and coastal water level measurements (e.g., Latapy et al. (2022); de Smeth et al. (2024)).

The rescued dataset, in addition to daily maximum, minimum, and average temperatures and daily precipitation, also includes dry- and wet-bulb temperatures recorded three times per day (06:00, 15:00, and 18:00 local time). These sub-daily temperature measurements enable the derivation of additional variables, including actual vapour pressure, relative humidity, and vapour pressure deficit (Bolton, 1980; List, 1966) (see Fig. 2 and Sect. 2.3.4). This dataset therefore enables additional analyses beyond those presented here, including but not limited to extreme event attribution analyses, climate model evaluation, and constraining of regional climate models for the Congo Basin.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

5 Conclusions

In this study, we describe our climate data rescue mission in the Democratic Republic of Congo where we: (i) digitised (imaged) 9,885 historical weather sheets presently stored in the INERA archives in Yangambi, (ii) automated transcription of the data using an open-source machine-learning based software MeteoSaver version 1.1, resulting in over one million daily and sub-daily temperature and precipitation records from 37 stations across the DRC, (iii) uploaded the data to open-access repositories, and finally (iv) analysed the temperature trends at 21 stations and changes in precipitation return periods at three stations, as well as comparing the temperature trends in the station observations with ERA5 reanalysis.

This newly transcribed dataset represents one of the first large-scale digital historical climate datasets for the DRC and is made available to the wider research community. Our results show clear and consistent warming trends in daily maximum, minimum, and average temperatures across the DRC, with median increases of approximately 0.7°C, 0.3°C, and 0.5°C over a 30-year period, respectively. The warming shift in distributions of daily maximum, minimum and average temperatures with each successive decade during the period 1961-1990 further highlights the warming pattern in the region, with higher modal temperatures and more extreme values. This decadal shift in distributions is also evident in stations with data extending into the early 2000s and 2010s.

Consistent with these trends, we observe an increasing frequency of hot extremes across the region with each successive decade. During the period 1961-1990, the average number of hot days per year nearly doubled between the earliest and most recent decades. Simultaneously, fewer cold days were observed across all the aggregated station database over the same period. This pattern persists in more recent decades at stations with data extending beyond 1990. This further strengthens the evidence of an unequivocal warming trend in the region.

Comparison between station observations and ERA5-Land reanalysis shows that, while ERA5-Land captures large-scale variability and long-term warming trends, some discrepancies with the station observations remain, particularly during the early reanalysis period (1950s), when observational constraints were limited.

In conclusion, our results demonstrate consistent warming trends across the Congo Basin during the period 1961-1990 at all the considered stations, with continued warming during the 2000s at those with extended data, accompanied by increasing frequency of hot extremes. These findings provide new observational evidence of climate change in this data-sparse region and underline the critical need for more climate data rescue efforts to improve assessments of observed changes in climate, as well as to better constrain reanalysis products.

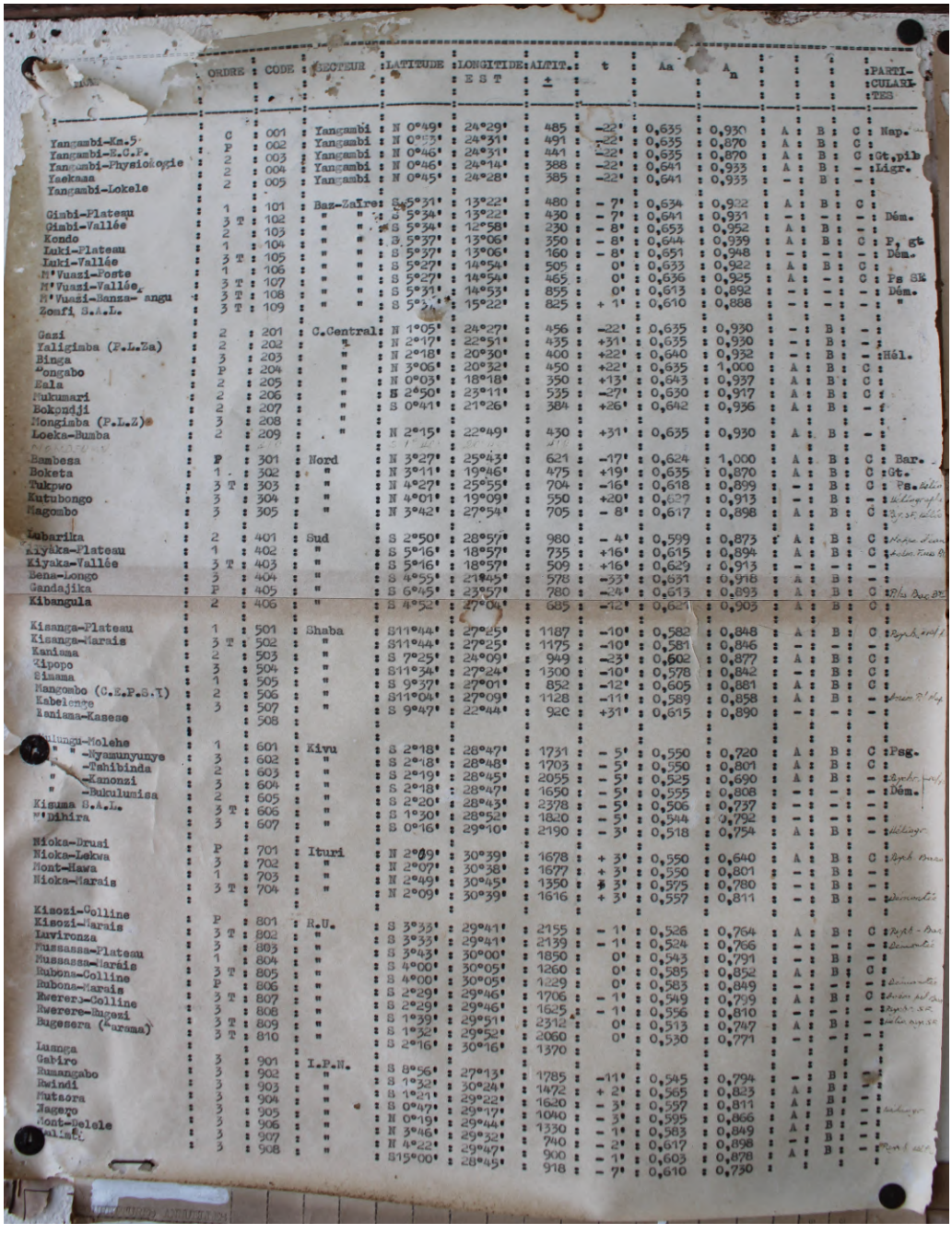
Code and data availability. MeteoSaver v1.1 is available on the GitHub repository of the Department of Water and Climate at the Vrije Universiteit Brussel (https://github.com/VUB-HYDR/MeteoSaver/tree/version-1.1_including_precip_and_dry_and_wet_bulb_temp). Additionally, all scripts used for the analysis of the transcribed data are available on the same repository (https://github.com/VUB-HYDR/2026_Muheki_et_al). The transcribed dataset is available on Zenodo (https://doi.org/10.5281/zenodo.18770063).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix A: Data Inventory and Digitization Campaign at INERA-Section de Climatologie archives in Yangambi, DRC

A1 INERA Station Metadata

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026



	ORDRE	CODE	SECTEUR	LATITUDE	LONGITUDE EST	ALTIT. ±	t	Aa	An				PARTICULARITES
Yangambi-Km.5	C	001	Yangambi	N 0°49'	24°29'	485	-22'	0,635	0,930	A	B	C	Nap.
Yangambi-E.C.P.	P	002	Yangambi	N 0°53'	24°31'	491	-22'	0,635	0,870	A	B	C	
Yangambi-Physiologie	2	003	Yangambi	N 0°46'	24°31'	441	-22'	0,635	0,870	A	B	C	Ct.pib
Yaekaaa	2	004	Yangambi	N 0°46'	24°14'	388	-22'	0,641	0,933	A	B	-	Ligr.
Yangambi-Lokele	2	005	Yangambi	N 0°45'	24°28'	385	-22'	0,641	0,933	-	B	-	
Gimbi-Plateau	1	101	Bas-Zaire	S 5°31'	13°22'	480	- 7'	0,634	0,922	A	B	C	
Gimbi-Vallée	3 T	102	"	S 5°34'	13°22'	430	- 7'	0,641	0,931	-	-	-	Dém.
Kondo	2	103	"	S 5°34'	12°58'	230	- 8'	0,653	0,952	A	B	-	
Luki-Plateau	1	104	"	S 5°37'	13°06'	350	- 8'	0,644	0,939	A	B	C	P. gt
Luki-Vallée	3 T	105	"	S 5°37'	13°06'	160	- 8'	0,651	0,948	-	-	-	Dém.
M'Vuazi-Poste	1	106	"	S 5°27'	14°54'	505	0'	0,633	0,922	A	B	C	
M'Vuazi-Vallée	3 T	107	"	S 5°27'	14°54'	465	0'	0,636	0,925	A	-	C	Ps SK
M'Vuazi-Banza-Mangu	3 T	108	"	S 5°31'	14°53'	855	0'	0,613	0,892	-	-	-	Dém.
Zomfi S.A.L.	3 T	109	"	S 5°3[illegible]	15°22'	825	+ 1'	0,610	0,888	-	-	-	"
Gazi	2	201	C.Central	N 1°05'	24°27'	456	-22'	0,635	0,930	-	B	-	
Yaligimba (P.L.Za)	2	202	"	N 2°17'	22°51'	435	+31'	0,635	0,930	-	B	-	
Binga	3	203	"	N 2°18'	20°30'	400	+22'	0,640	0,932	-	B	-	Hôl.
Mongabo	P	204	"	N 3°06'	20°32'	450	+22'	0,635	1,000	A	B	C	
Eala	2	205	"	N 0°03'	18°18'	350	+13'	0,643	0,937	A	B	C	
Fukumari	2	206	"	S 2°50'	23°11'	535	-27'	0,630	0,917	A	B	C	
Bokondji	2	207	"	S 0°41'	21°26'	384	+26'	0,642	0,936	A	B	-	f
Mongimba (P.L.Z)	3	208	"										
Loeka-Bumba	2	209	"	N 2°15'	22°49'	430	+31'	0,635	0,930	A	B	-	
Bambesa	P	301	Nord	N 3°27'	25°43'	621	-17'	0,624	1,000	A	B	C	Bar.
Boketa	1	302	"	N 3°11'	19°46'	475	+19'	0,635	0,870	A	B	C	Ct.
Tukpwo	3 T	303	"	N 4°27'	25°55'	704	-16'	0,618	0,899	-	B	C	Ps.
Kutubongo	3	304	"	N 4°01'	19°09'	550	+20'	0,627	0,913	-	B	-	
Nagombo	3	305	"	N 3°42'	27°54'	705	- 8'	0,617	0,898	A	B	C	
Lubarika	2	401	Sud	S 2°50'	28°57'	980	- 4'	0,599	0,873	A	B	C	
Kiyaka-Plateau	1	402	"	S 5°16'	18°57'	735	+16'	0,615	0,894	A	B	C	
Kiyaka-Vallée	3 T	403	"	S 5°16'	18°57'	509	+16'	0,629	0,913	-	-	-	
Bena-Longo	3	404	"	S 4°55'	21°45'	578	-53'	0,631	0,918	A	B	-	
Gandajika	P	405	"	S 6°45'	23°57'	780	-24'	0,613	0,893	A	B	C	
Kibangula	2	406	"	S 4°52'	27°04'	685	-12'	0,621	0,903	A	B	C	
Kisanga-Plateau	1	501	Shaba	S 11°44'	27°25'	1187	-10'	0,582	0,848	A	B	C	
Kisanga-Marais	3 T	502	"	S 11°44'	27°25'	1175	-10'	0,581	0,846	-	-	-	
Kaniama	2	503	"	S 7°25'	24°09'	949	-23'	0,602	0,877	A	B	C	
Kipopo	3	504	"	S 11°34'	27°24'	1300	-10'	0,578	0,842	-	B	C	
Sinama	1	505	"	S 9°37'	27°01'	852	-12'	0,605	0,881	A	B	C	
Mangombo (C.E.P.S.K)	2	506	"	S 11°04'	27°09'	1128	-11'	0,589	0,858	A	B	-	
Kabelenge	3	507	"	S 9°47'	22°44'	920	+31'	0,615	0,890	-	-	-	
Kanisma-Kasese		508	"										
Kilungu-Mokeke	1	601	Kivu	S 2°18'	28°47'	1731	- 5'	0,550	0,720	A	B	C	Psg.
" -Nyamunyunye	3	602	"	S 2°18'	28°48'	1703	- 5'	0,550	0,801	A	B	C	
" -Tshibinda	2	603	"	S 2°19'	28°45'	2055	- 5'	0,525	0,690	A	B	-	
" -Kanonzi	3	604	"	S 2°18'	28°47'	1650	- 5'	0,555	0,808	-	-	-	Dém.
" -Bukulumisa	2	605	"	S 2°20'	28°43'	2378	- 5'	0,506	0,737	-	-	-	
Kisuma S.A.L.	3 T	606	"	S 1°30'	28°52'	1820	- 5'	0,544	0,792	-	-	-	
M'Dihira	3	607	"	S 0°16'	29°10'	2190	- 3'	0,518	0,754	A	B	-	
Nioka-Drusi	P	701	Ituri	N 2°09'	30°39'	1678	+ 3'	0,550	0,640	A	B	C	
Nioka-Lokwa	3	702	"	N 2°07'	30°38'	1677	+ 3'	0,550	0,801	-	B	-	
Mont-Hawa	1	703	"	N 2°49'	30°45'	1350	+ 3'	0,575	0,780	-	B	-	
Nioka-Marais	3 T	704	"	N 2°09'	30°39'	1616	+ 3'	0,557	0,811	-	B	-	
Kisozi-Colline	P	801	R.U.	S 3°33'	29°41'	2155	- 1'	0,528	0,764	A	B	C	
Kisozi-Marais	3 T	802	"	S 3°33'	29°41'	2139	- 1'	0,524	0,766	-	-	-	
Luvironza	3	803	"	S 3°43'	30°00'	1850	0'	0,543	0,791	-	B	-	
Mussassa-Plateau	1	804	"	S 4°00'	30°05'	1260	0'	0,585	0,852	A	B	C	
Mussassa-Marais	3 T	805	"	S 4°00'	30°05'	1229	0'	0,583	0,849	-	-	-	
Rubona-Colline	P	806	"	S 2°29'	29°46'	1706	- 1'	0,549	0,799	A	B	C	
Rubona-Marais	3 T	807	"	S 2°29'	29°46'	1625	- 1'	0,556	0,810	-	-	-	
Rwerere-Colline	3	808	"	S 1°39'	29°51'	2312	0'	0,513	0,747	A	B	-	
Rwerere-Bugesi	3 T	809	"	S 1°32'	29°52'	2060	0'	0,530	0,771	-	-	-	
Bugesera (Kayama)	3 T	810	"	S 2°16'	30°16'	1370							
Lusanga													
Gabiro	3	901	I.P.N.	S 8°56'	27°13'	1785	-11'	0,545	0,794	-	B	-	
Rumangabo	3	902	"	S 1°32'	30°24'	1472	+ 2'	0,565	0,823	A	B	-	
Rwindi	3	903	"	S 1°21'	29°22'	1620	- 3'	0,597	0,811	A	B	-	
Mutsora	3	904	"	S 0°47'	29°17'	1040	- 3'	0,595	0,866	A	B	-	
Jagero	3	905	"	N 0°19'	29°44'	1330	- 1'	0,583	0,849	A	B	-	
Mont-Belele	3	906	"	N 3°46'	29°32'	740	- 2'	0,617	0,898	-	B	-	
Kalisi	3	907	"	N 4°22'	29°47'	900	- 1'	0,603	0,878	A	B	-	
	3	908	"	S 15°00'	28°45'	918	- 7'	0,610	0,730				

Figure A1. Station metadata document photographed at INERA-Section de Climatologie archives in Yangambi, DRC, during the data inventory and digitization campaign. The sheet contains the following information about all the INERA weather stations: (i) station name, (ii) classification order (ORDRE), (iii) station code (CODE), (iv) region (SECTEUR), (v) Latitude, (vi) Longitude (East) (LONGITUDE EST), (vii) altitude (ALTIT. ±), and (viii) other station specific information, classification and remarks (t, Aa, An and PARTICULARITÉS).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

A2 Example of a Hand-Drawn Post-1960 Observation Sheet

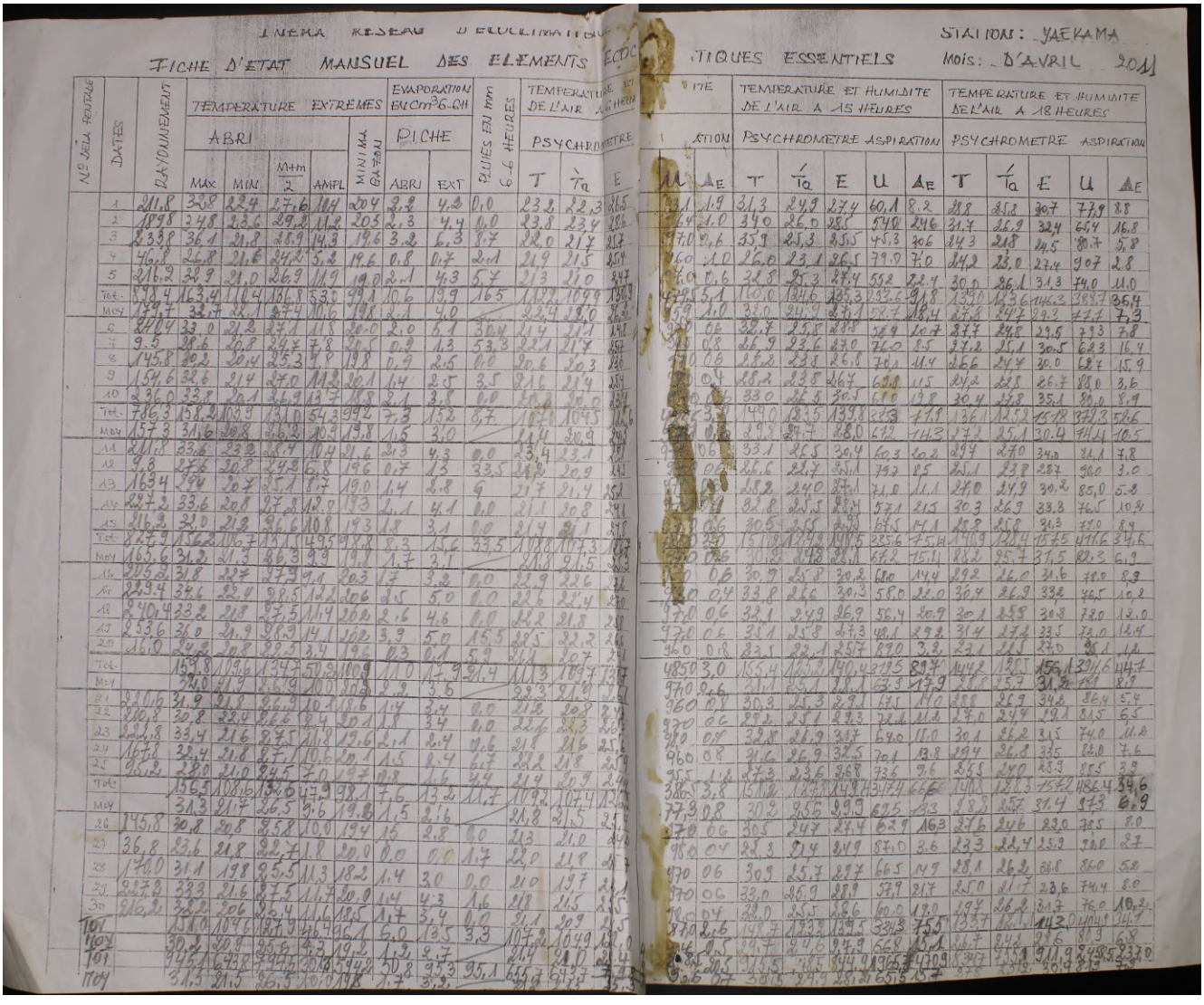
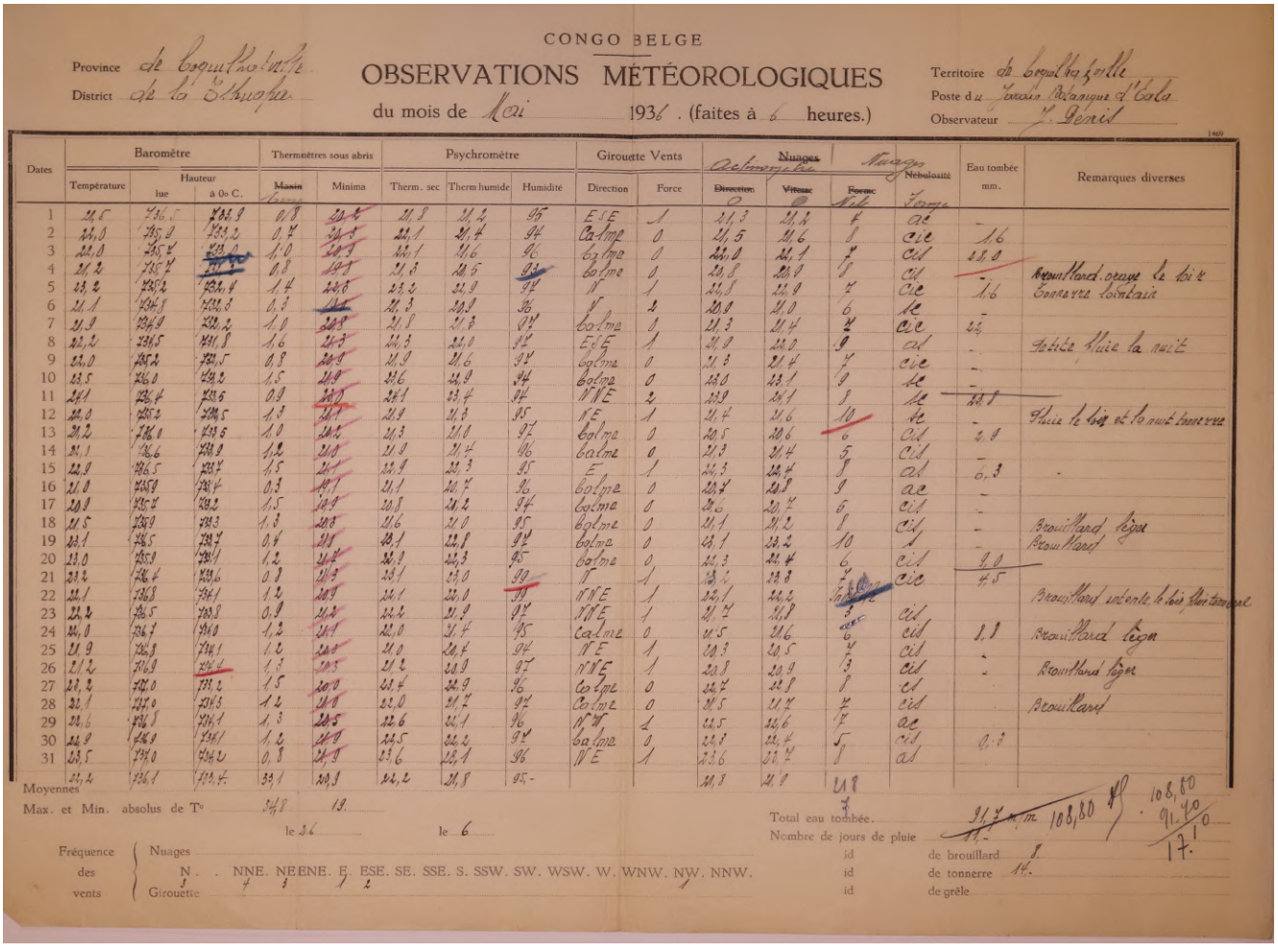


Figure A2. Example of a hand-drawn monthly weather observation sheet for Station Yaekema (N 0°46', E 24°14') in the Democratic Republic of the Congo (DRC), as found in the archives of INERA, Yangambi.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix B: Pre-1960 weather data sheets at the State Archives of Belgium digitized by the COBECORE project

B1 Example of a Pre-1960 Sheet Format: 1930s


CONGO BELGE

OBSERVATIONS MÉTÉOROLOGIQUES

Province de Coquilhatville — Territoire de Coquilhatville
District de la Lulonga — Poste d'u Secteur Pédagogique d'Eala
du mois de Mai 1936 . (faites à 6 heures.) — Observateur J. Genist

Dates	Baromètre Température	Baromètre Hauteur lue	Baromètre Hauteur à 0° C.	Thermomètres sous abris Maxim.	Thermomètres sous abris Minima	Psychromètre Therm. sec	Psychromètre Therm humide	Psychromètre Humidité	Girouette Vents Direction	Force	Nuages Actinométrie Direction	Nuages Vitesse	Nuages Force	Nébulosité Forme	Eau tombée mm.	Remarques diverses
1	22,5	736,5	733,9	0,8	20,2	21,8	21,2	95	E.S.E	1	21,3	21,2	4	ac	-	
2	22,3	735,2	732,2	0,7	20,3	22,1	21,4	94	Calme	0	21,5	21,6	8	cic	1,6	
3	22,0	735,4	732,4	1,0	20,4	22,1	21,6	96	Calme	0	22,0	22,1	7	cid	12,0	
4	21,2	735,7	732,7	0,8	19,7	21,3	20,5	94	Calme	0	20,8	20,9	8	cid	-	Brouillard épais le soir
5	23,2	736,2	733,4	1,4	20,8	22,2	21,9	94	N	1	22,3	22,9	7	cic	1,6	Tonnerre lointain
6	21,1	734,3	731,3	0,3	19,2	21,3	20,9	96	E	2	20,8	21,0	6	ac	-	
7	21,3	734,0	731,2	1,0	20,8	21,8	21,2	97	Calme	0	21,3	21,4	7	cic	2,5	
8	22,2	736,5	733,8	1,6	20,3	22,3	22,0	97	E.S.E	1	21,9	22,0	9	ad	-	Petite pluie la nuit
9	22,0	736,2	733,5	0,8	20,9	21,9	21,6	97	Calme	0	21,3	21,4	7	cic	-	
10	23,5	736,0	733,2	1,5	21,0	22,6	22,3	94	Calme	0	23,0	23,1	9	ac	-	
11	24,1	736,4	733,6	0,8	20,0	24,1	23,4	94	N.N.E	2	23,9	24,1	8	ac	22,1	
12	22,0	736,2	733,5	1,3	20,7	21,9	21,3	95	N.E	1	21,4	21,6	10	ac	-	Pluie le soir et la nuit tonnerre
13	21,2	736,1	733,3	1,0	20,2	21,3	21,0	97	Calme	0	20,5	20,6	6	cid	2,1	
14	21,1	736,6	733,9	1,2	21,0	21,9	21,4	96	Calme	0	21,3	21,4	5	cid	-	
15	22,9	736,5	733,7	1,5	21,1	22,9	22,3	95	E	1	22,3	22,4	8	ad	6,3	
16	21,0	735,8	733,1	0,3	19,7	21,1	20,7	96	Calme	0	20,7	20,8	8	ac	-	
17	20,9	735,4	732,8	1,5	19,9	20,8	20,2	94	Calme	0	20,0	20,2	6	cid	-	
18	21,5	736,0	733,3	1,3	20,2	21,6	21,0	95	Calme	0	21,1	21,2	8	cid	-	Brouillard léger
19	22,1	736,5	733,7	0,8	21,1	22,1	21,8	94	Calme	0	22,1	22,2	10	1	-	Brouillard
20	23,0	736,9	734,1	1,2	21,4	23,2	22,3	95	Calme	0	22,3	22,4	6	cid	9,0	
21	23,2	736,4	733,6	0,8	21,3	23,1	23,0	99	N	1	23,2	23,3	7	cic	4,5	
22	22,1	736,8	734,1	1,2	20,3	22,1	22,0	99	N.N.E	1	22,1	22,2	7			Brouillard intense le soir éclair tonnerre
23	22,2	736,5	733,8	0,9	21,2	22,2	21,9	97	N.N.E	1	21,7	21,8	3	cid		
24	21,0	736,7	734,0	1,2	21,1	22,0	21,4	95	Calme	0	21,5	21,6	6	cid	1,1	Brouillard léger
25	21,9	736,8	734,1	1,2	20,9	22,0	21,4	94	N.E	1	20,3	20,5	4	cid	-	
26	21,2	736,9	734,1	1,3	19,5	21,2	20,9	97	N.N.E	1	20,3	20,3	3	cid	-	Brouillard léger
27	21,2	736,0	733,2	1,5	20,0	21,4	20,9	96	Calme	0	20,7	20,8	8	cl		
28	21,1	736,0	733,3	1,2	20,8	22,0	21,7	97	Calme	0	21,5	21,7	4	cid		Brouillard
29	22,6	736,8	734,1	1,3	20,5	22,6	22,1	96	N.N.E	1	22,5	22,6	7	ac		
30	22,9	736,8	734,1	1,2	20,9	22,5	22,2	97	Calme	0	22,3	22,4	5	cid	0,2	
31	23,5	736,0	733,2	0,8	21,9	23,6	23,1	96	N.E	1	23,6	23,7	8	ad		
Moyennes	22,4	736,1	733,4	33,1	20,3	22,2	21,8	95,-			22,3	22,4	6,8			

Max. et Min. absolus de T° 34,8 le 26 — 19. le 6

Total eau tombée 91,7 m/m 108,80 % 91,40 / 17.
Nombre de jours de pluie 11,- id de brouillard 8
id de tonnerre 14.
id de grêle

Fréquence des vents — Nuages — Girouette
N. 3 . NNE. 4 NE 3 ENE. 7 E. 2 ESE. SE. SSE. S. SSW. SW. WSW. W. WNW. NW. NNW.

Figure B1. Representative example of a hand-drawn monthly weather observation sheet from the 1930s, showing the typical format used during that period. This sheet corresponds to Station Eala (N 0°03', E 18°18') in the Democratic Republic of the Congo (DRC), and is preserved in the State Archives of Belgium. Digitized by the COBECORE project (COBECORE, 2018; Jacobsen et al., 2018).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

B2 Example of a Pre-1960 Sheet Format: 1940s

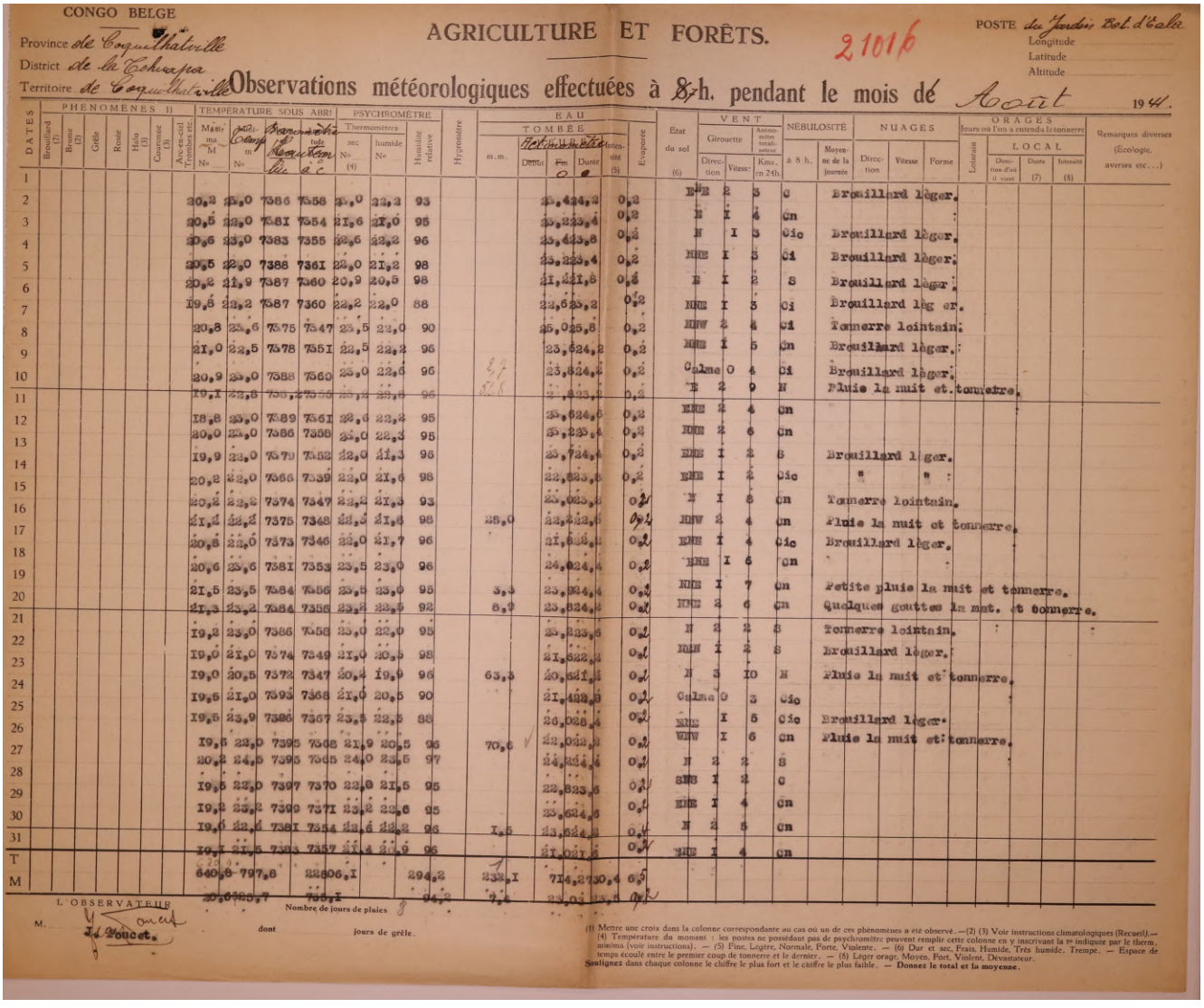
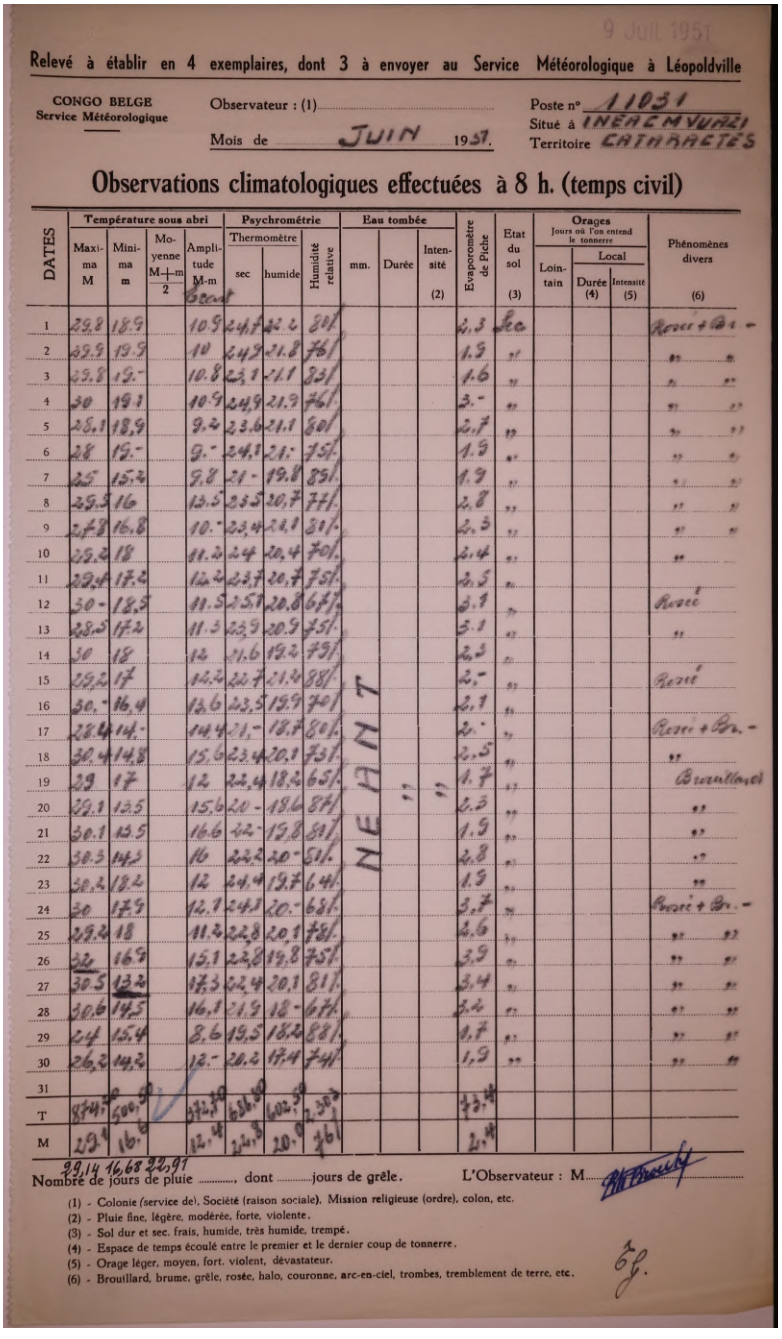


Figure B2. Representative example of a hand-drawn monthly weather observation sheet from the 1940s, showing the typical format used during that period. This sheet corresponds to Station Eala (N 0°03', E 18°18') in the Democratic Republic of the Congo (DRC), and is preserved in the State Archives of Belgium. Digitized by the COBECORE project (COBECORE, 2018; Jacobsen et al., 2018).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

B3 Example of a Pre-1960 Sheet Format: 1950s



Relevé à établir en 4 exemplaires, dont 3 à envoyer au Service Météorologique à Léopoldville

CONGO BELGE
Service Météorologique

Observateur : (1)

Mois de JUIN 1951.

Poste n° 11031
Situé à INEAC M'VUAZI
Territoire CATARACTES

Observations climatologiques effectuées à 8 h. (temps civil)

DATES	Température sous abri: Maxima M	Minima m	Moyenne M+m/2	Amplitude M-m	Psychrométrie: Thermomètre sec	Thermomètre humide	Humidité relative	Eau tombée: mm.	Durée	Intensité (2)	Evaporomètre de Piche	Etat du sol (3)	Orages: Jours où l'on entend le tonnerre: Lointain	Local: Durée (4)	Local: Intensité (5)	Phénomènes divers (6)
1	29.8	18.9		10.9	24.7	22.2	80				4.3	Sec				Rosée + Br. -
2	29.9	19.9		10	24.9	21.8	76				1.9	"				" "
3	29.8	19.-		10.8	23.7	21.1	83				1.6	"				" "
4	30	19.1		10.9	24.9	21.9	76				3.-	"				" "
5	28.1	18.9		9.2	23.6	21.1	80				2.7	"				" "
6	28	19.-		9.-	24.3	21.-	75				1.3	"				" "
7	25	15.2		9.8	21 -	19.8	85				1.9	"				" "
8	29.3	16		13.5	23.5	20.7	77				2.8	"				" "
9	28.8	16.8		10.-	23.4	21.1	81				2.3	"				" "
10	29.2	18		11.2	24	20.4	70				2.4	"				"
11	29.4	17.2		12.2	23.7	20.7	75				2.5	"				"
12	30.-	18.3		11.5	25.1	20.8	67				3.1	"				Rosée
13	28.5	17.2		11.3	23.5	20.9	75				3.1	"				"
14	30	18		12	21.6	19.2	73				2.3	"				"
15	29.2	17		14.2	22.7	21.4	88				2.-	"				Rosée
16	30.-	16.4		13.6	23.5	19.9	70				2.1	"				"
17	28.4	14.-		14.4	21.-	18.1	80				2.-	"				Rosée + Br. -
18	30.4	14.8		15.6	23.4	20.1	73				2.5	"				"
19	29	17		12	22.4	18.2	65	"	"		1.7	"				Brouillard
20	29.1	12.5		15.6	20 -	18.6	87				2.3	"				"
21	30.1	13.5		16.6	22-	19.8	81				1.9	"				"
22	30.3	14.3		16	22.2	20.-	81				2.8	"				"
23	30.2	18.2		12	24.4	19.7	64	NEANT			1.9	"				"
24	30	17.9		12.1	24.1	20.-	68				3.7	"				Rosée + Br. -
25	29.2	18		11.2	22.8	20.1	78				2.6	"				" "
26	32	16.9		15.1	22.8	19.8	75				3.9	"				" "
27	30.5	13.2		17.3	22.4	20.1	81				3.4	"				" "
28	30.6	14.5		16.1	21.9	18.-	67				3.4	"				" "
29	24	15.4		8.6	19.5	18.2	88				0.7	"				" "
30	26.2	14.2		12.-	20.2	17.4	74				1.9	"				" "
31																
T	874.-	500.8		332.20	688.80	602.-	2.305				73.4					
M	29.1	16.-		12.-	24.8	20.-	76				2.4					

Nombre de jours de pluie 29,14 16,68 22,91, dont ..... jours de grêle. L'Observateur : M. [signature]

(1) - Colonie (service de), Société (raison sociale), Mission religieuse (ordre), colon, etc.
(2) - Pluie fine, légère, modérée, forte, violente.
(3) - Sol dur et sec, frais, humide, très humide, trempé.
(4) - Espace de temps écoulé entre le premier et le dernier coup de tonnerre.
(5) - Orage léger, moyen, fort, violent, dévastateur.
(6) - Brouillard, brume, grêle, rosée, halo, couronne, arc-en-ciel, trombes, tremblement de terre, etc.

Figure B3. Representative example of a hand-drawn monthly weather observation sheet from the 1950s, showing the typical format used during that period. This sheet corresponds to Station M'Vuazi (S 5°27', E 14°54') in the Democratic Republic of the Congo (DRC), and is preserved in the State Archives of Belgium. Digitized by the COBECORE project (COBECORE, 2018; Jacobsen et al., 2018).

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix C: Configuration User Settings in MeteoSaver v1.1

Table C1. MeteoSaver configuration settings, descriptions and default values used in this project

Setting/Value	Description	Default Value
General		
run_mode	Defines the environment where the software will be run. Options: `local` for personal computers or `hpc` for High Performance Computing.	`hpc`
num_processors	Specifies the number of CPUs to use when running on an HPC infrastructure (in `hpc` mode).	18
Directories		
full_datadir	Directory containing folders of historical weather data sheet images, organized by station number.	`data/00_post1960...`
pre_QA_QC_transcribed_hydroclimate_data_dir	Directory for storing pre-QA/QC transcribed hydroclimate data.	`results/01_pre...`
post_QA_QC_transcribed_hydroclimate_data_dir	Directory for storing post-QA/QC transcribed hydroclimate data.	`results/02_post...`
validation_dir	Directory for validation results.	`results/03_valid...`
final_refined_daily_hydroclimate_data_dir	Directory for final refined hydroclimate data after all quality checks.	`results/04_final...`
transient_transcription_output_dir	Directory to store transient transcription outputs during processing.	`results/05_trans...`
manually_transcribed_data_dir	Directory for manually transcribed data, used for validation purposes.	`results/06_manual...`
result_maps_dir	Directory for results overlain on maps and combined plots.	`results/07_maps...`
metadata_file_path	Directory for all the stations metadata.	`data/01_meta...`
formatted_already_digitized_data_dir	Directory for previously transcribed data to add to our new transcribed database.	`data/02_form...`
region_shapefile_path	Shapefile of study region	`data/03_shap...`
era5_dir	Directory for ERA5-Land data.	`data/04_era5...`
Table and Cell Detection		
clip_up, clip_down, clip_left, clip_right	Clipping values (in pixels) to remove headers and row labels from detected tables.	430, 280, 200, 120
max_table_width, max_table_height	Maximum expected table width and height (in pixels) to ensure correct table detection.	3900, 3600
min_cell_width_threshold, max_cell_width_threshold, min_cell_height_threshold, max_cell_height_threshold	Minimum and maximum width and height (in pixels) allowed for detected table cells.	60, 140, 30, 50
no_of_rows, no_of_columns, no_of_rows_including_headers	Expected number of rows and columns in detected tables.	43, 24, 46
space_height_threshold, space_width_threshold, max_cell_height_per_box	Minimum height and width space (in pixels) between bounding boxes to detect missing cells, with height for newly added (missing) bounding boxes/cells.	50, 120, 40
Transcription		
ocr_model	Defines which OCR/HTR model to use (e.g., Tesseract-OCR, EasyOCR, PaddleOCR).	`Tesseract-OCR`
tesseract_path	Path to the Tesseract-OCR executable file, needed for Tesseract-OCR model.	`.../tesseract.exe`
system_tessdata_dir	Directory path to trained OCR/HTR language models for transcription (e.g., Tesseract tessdata).	`.../tessdata`

Note: Paths with '...' indicate truncated directory names that are relative to the repository root; see the Zenodo repository for full names.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Table C2. (Continued from Table C1) MeteoSaver configuration settings, descriptions and default values used in this project

Setting/Value	Description	Default Value
QA/QC		
daily_temperature_columns; daily_temperature_columns_and_diurnal_temperature_range	Columns to focus on for quality checks, here columns with Daily Maximum Temperature, Minimum Temperature,Average Temperature and the Diurnal Temperature Range.	D,E,F ; D,E,F,G
daily_precipitation_column	Column with daily precipitation.	K
dry_and_wet_bulb_temperature_columns	Columns with dry bulb and wet bulb temperature readings recorded at specific times of the day.	L,M,Q,R,V,W
max_temperature_threshold, min_temperature_threshold	Temperature thresholds (in °C) to flag invalid transcribed temperature readings.	40, 5
decimal_places	Number of decimal places in the data sheets.	1
uncertainty_margin	Defines the uncertainty margin to be used in temperature calculations and quality checks.	0.2
header_rows	Number of header rows in the sheet .	3
multi_day_totals, multi_day_averages	Flags to specify whether multi-day totals or averages (e.g., 5-day totals) are present in the sheets.	`True, True`
max_days_for_multi_day_total	Maximum number of days contained in multi-day totals (e.g., 5 or 6 days).	6
multi_day_totals_rows; final_totals_rows	Rows where multi-day totals or final totals are located (if applicable).	9,16,23,30,37,45 ; 45
excluded_rows; excluded_columns	Rows and columns to exclude during QA/QC checks involving multi-day totals (e.g., headers, date).	1,2,3,9,16,23,30,37,45 ; 1,2,3,15,20,25,26,27
additional_excluded_rows	Extra rows to exclude during QA/QC checks involving multi-day averages (if applicable).	10,17,24,31,38,46

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix D: Daily Temperature Trends at Individual Stations

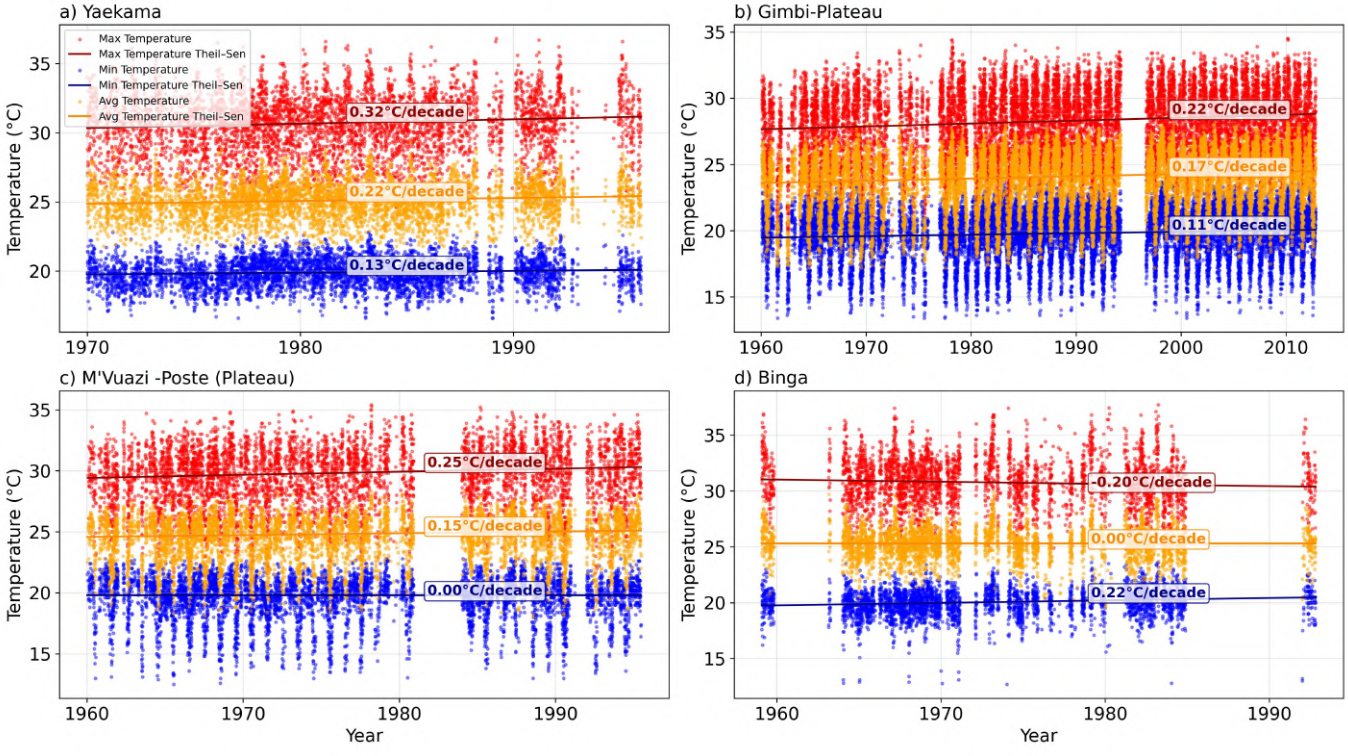


Figure D1. Daily temperature trends at stations: a) Yaekama (N 0°46' E 24°14'), b) Gimbi-Plateau (S 5°31' E 13°22'), c) M'Vuazi-Poste (Plateau) (S 5°27' E 14°54'), and d) Binga (N 2°18' E 20°30') in the DRC. Colored dots represent daily values of maximum (Tx, red), average (Tavg, orange), and minimum (Tn, blue) temperatures. Solid lines indicate Theil–Sen trend estimates for each variable, with the corresponding slope (°C/decade) shown at the end of each line. This data includes both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

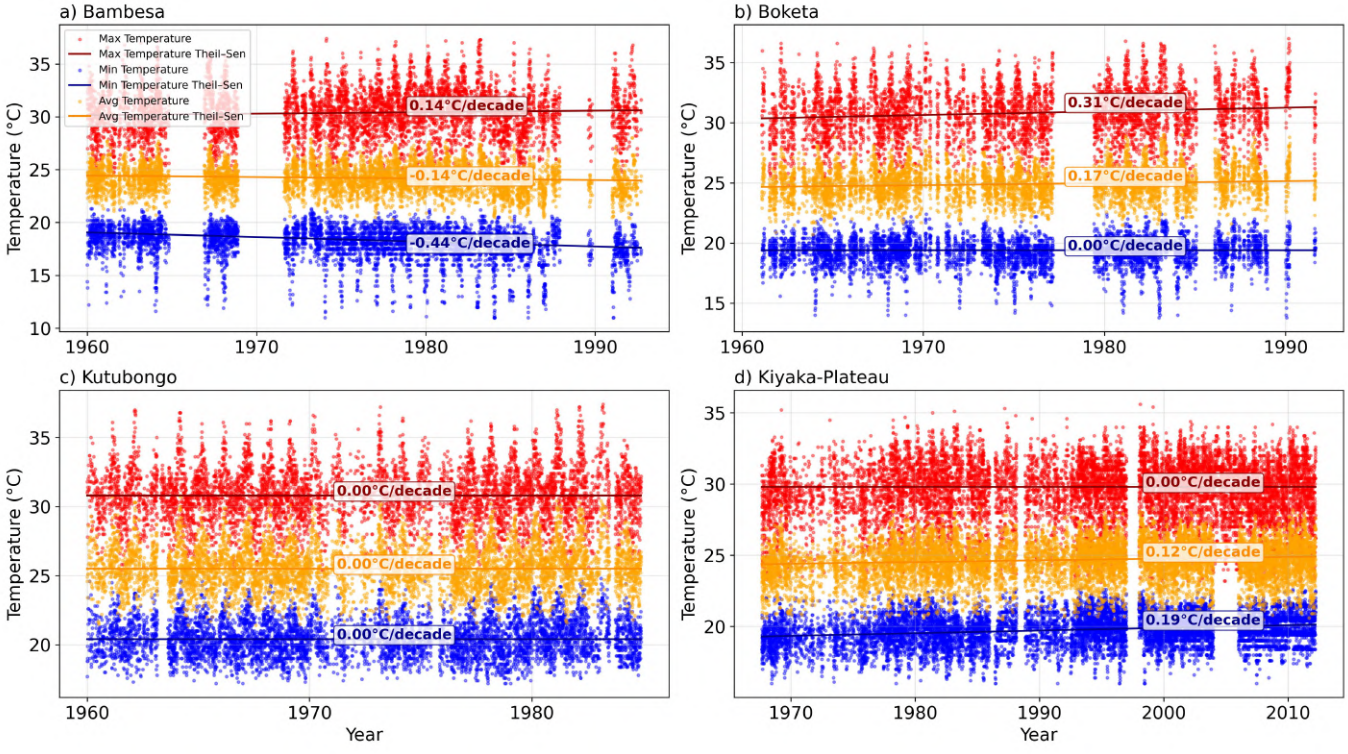


Figure D2. Daily temperature trends at stations: a) Bambesa (N 3°27' E 25°43'), b) Boketa (N 3°11' E 19°46'), c) Kutubongo (N 4°01' E 19°09'), and d) Kiyaka-Plateau (S 5°16' E 18°57') in the DRC. Colored dots represent daily values of maximum (Tx, red), average (Tavg, orange), and minimum (Tn, blue) temperatures. Solid lines indicate Theil–Sen trend estimates for each variable, with the corresponding slope (°C/decade) shown at the end of each line. This data includes both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

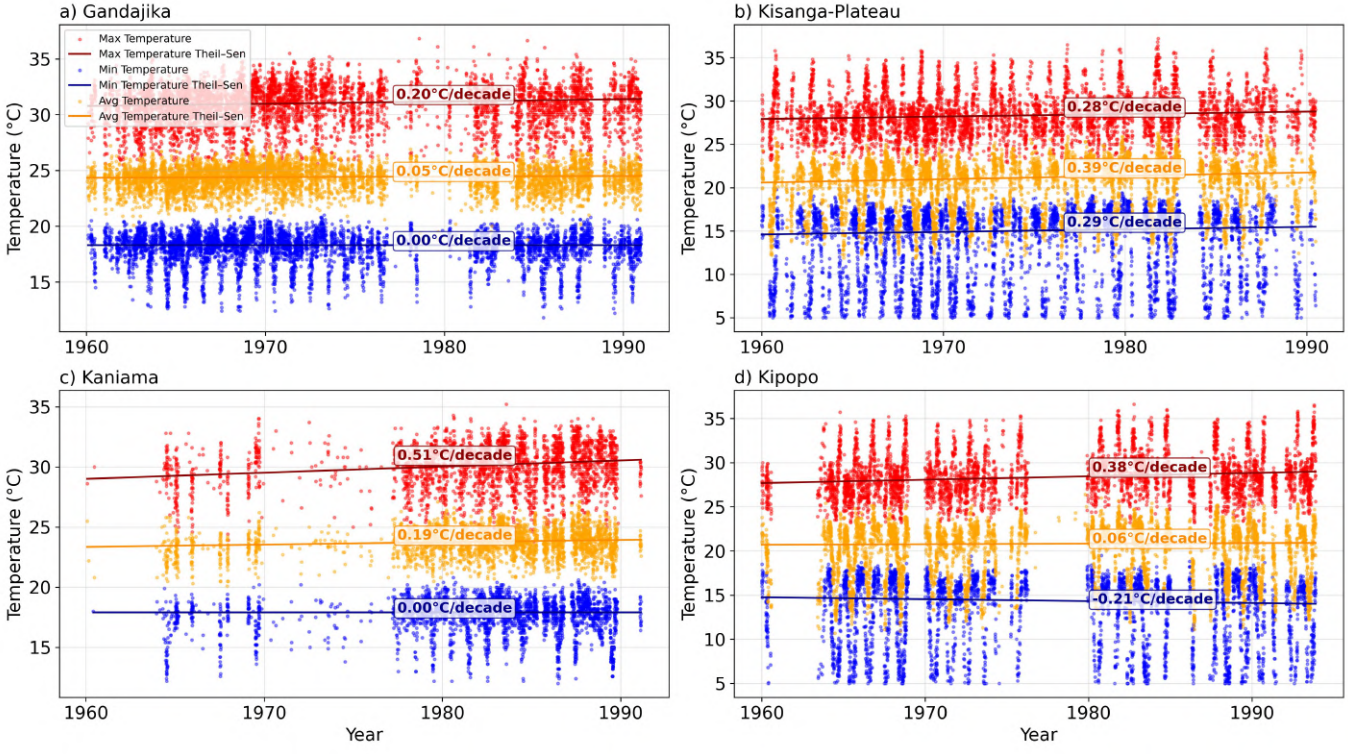


Figure D3. Daily temperature trends at stations: a) Gandajika (S 6°45' E 23°57'), b) Kisanga-Plateau (S 11°44' E 27°25'), c) Kaniama (S 7°25' E 24°09'), and d) Kipopo (S 11°34' E 27°24') in the DRC. Colored dots represent daily values of maximum (Tx, red), average (Tavg, orange), and minimum (Tn, blue) temperatures. Solid lines indicate Theil–Sen trend estimates for each variable, with the corresponding slope (°C/decade) shown at the end of each line. This data includes both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

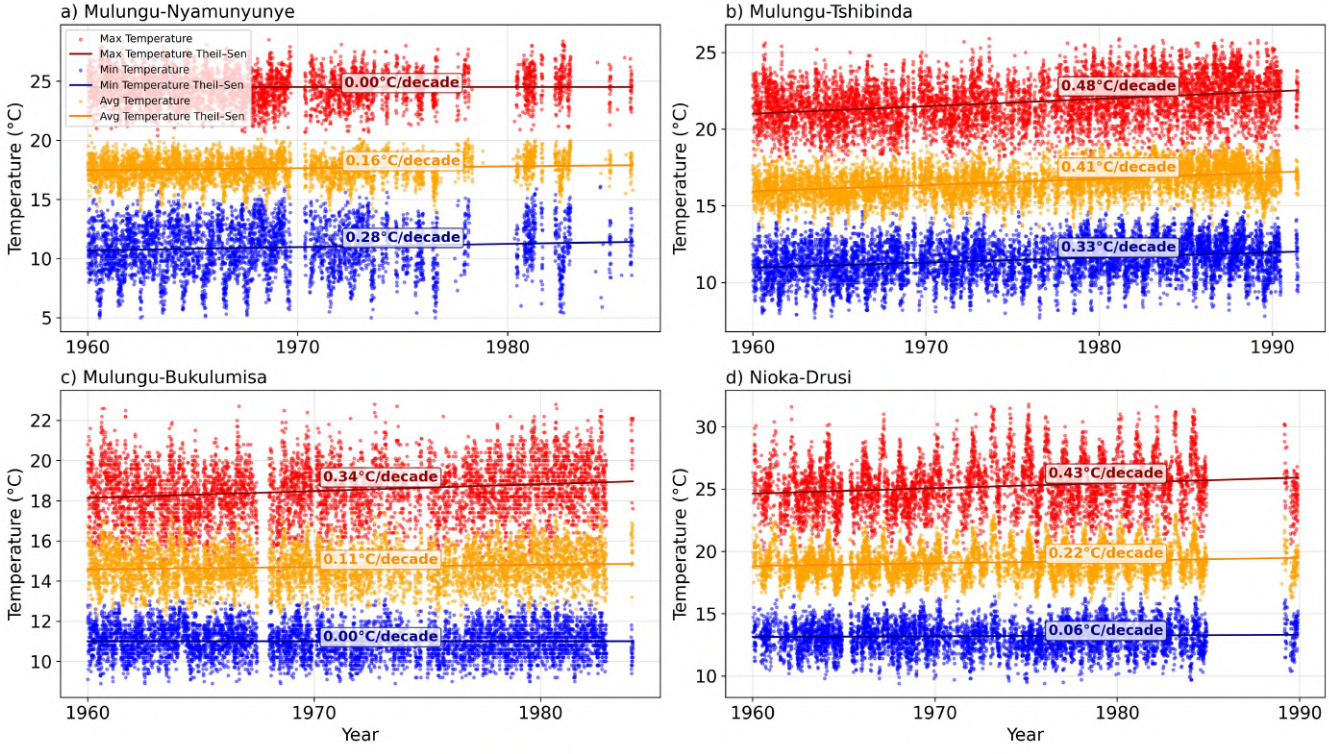


Figure D4. Daily temperature trends at stations: a) Mulungu-Nyamunyunye (S 2°18’ E 28°48’), b) Mulungu-Tshibinda (S 2°19’ E 28°45), c) Mulungu-Bukulumisa (S 2°20’ E 28°43’), and d) Nioka-Drusi (N 2°09’ E 30°39’) in the DRC. Colored dots represent daily values of maximum (Tx, red), average (Tavg, orange), and minimum (Tn, blue) temperatures. Solid lines indicate Theil–Sen trend estimates for each variable, with the corresponding slope (°C/decade) shown at the end of each line. This data includes both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

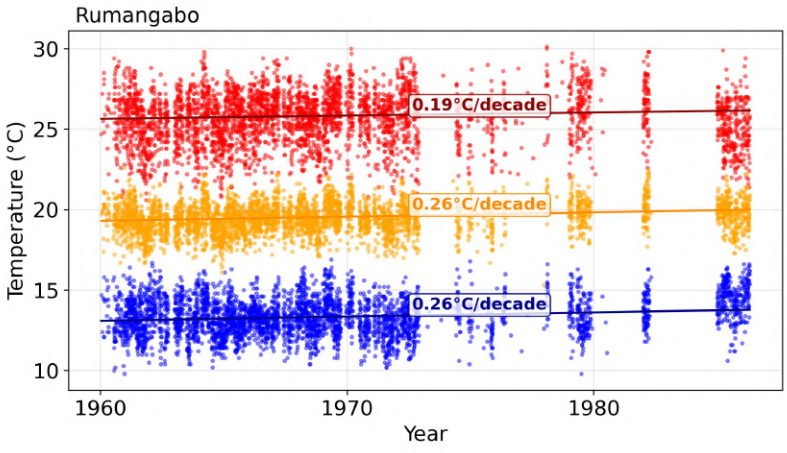


Figure D5. Daily temperature trends at station Rumangabo (S 1°21' E 29°22') in the DRC. Colored dots represent daily values of maximum (Tx, red), average (Tavg, orange), and minimum (Tn, blue) temperatures. Solid lines indicate Theil–Sen trend estimates for each variable, with the corresponding slope (°C/decade) shown at the end of each line. This data includes both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix E: Decadal shifts in temperature distribution at Individual Stations

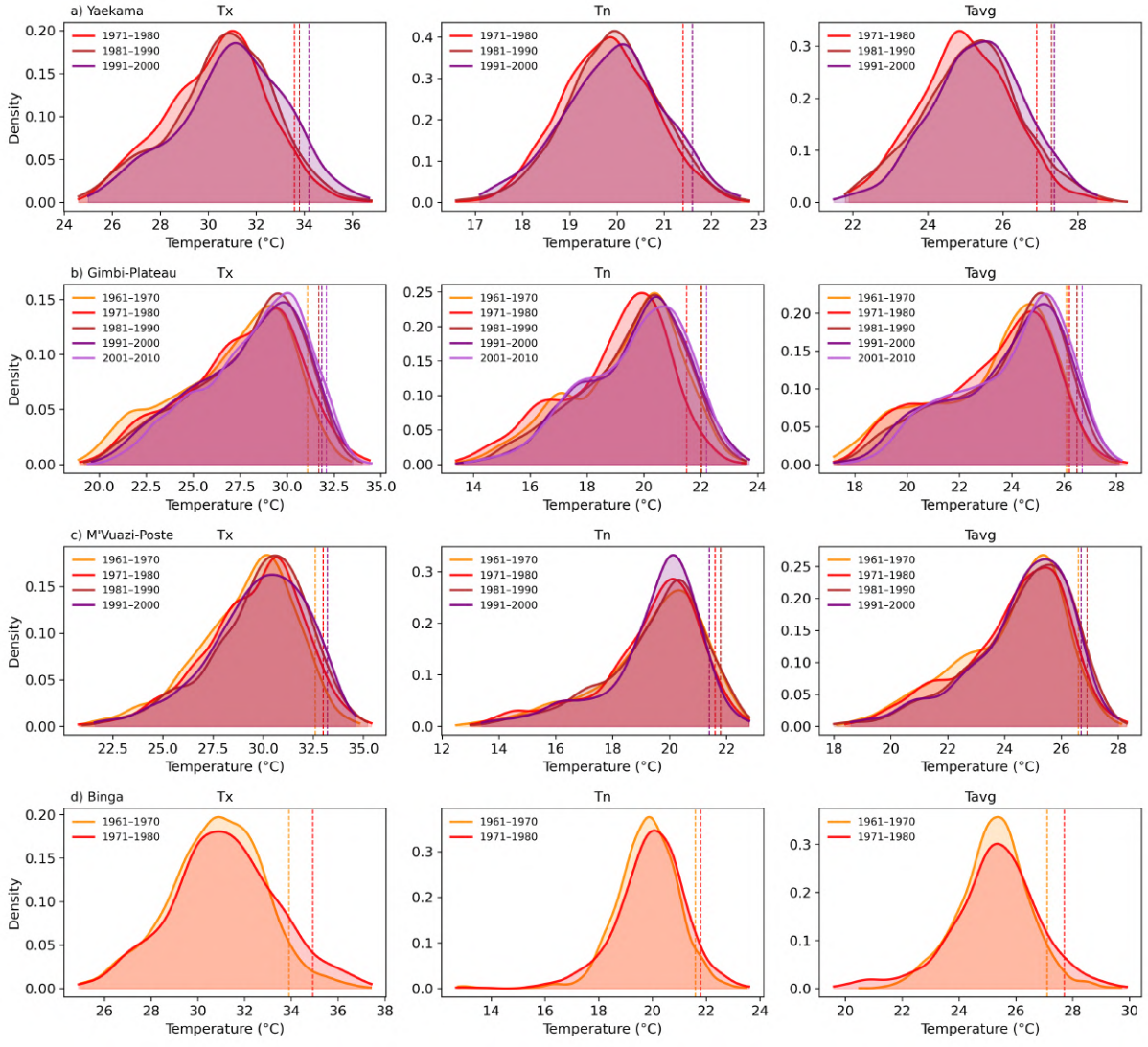


Figure E1. Decadal shifts in temperature distributions at stations: a) Yaekama (N 0°46' E 24°14'), b) Gimbi-Plateau (S 5°31' E 13°22'), c) M'Vuazi-Poste (Plateau) (S 5°27' E 14°54'), and d) Binga (N 2°18' E 20°30') in the DRC. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). Vertical dashed lines indicate the 95th percentile values for each decade, with panels (a), (b) and (c) illustrating the progressive rightward shift towards higher temperatures and more extreme values over time.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

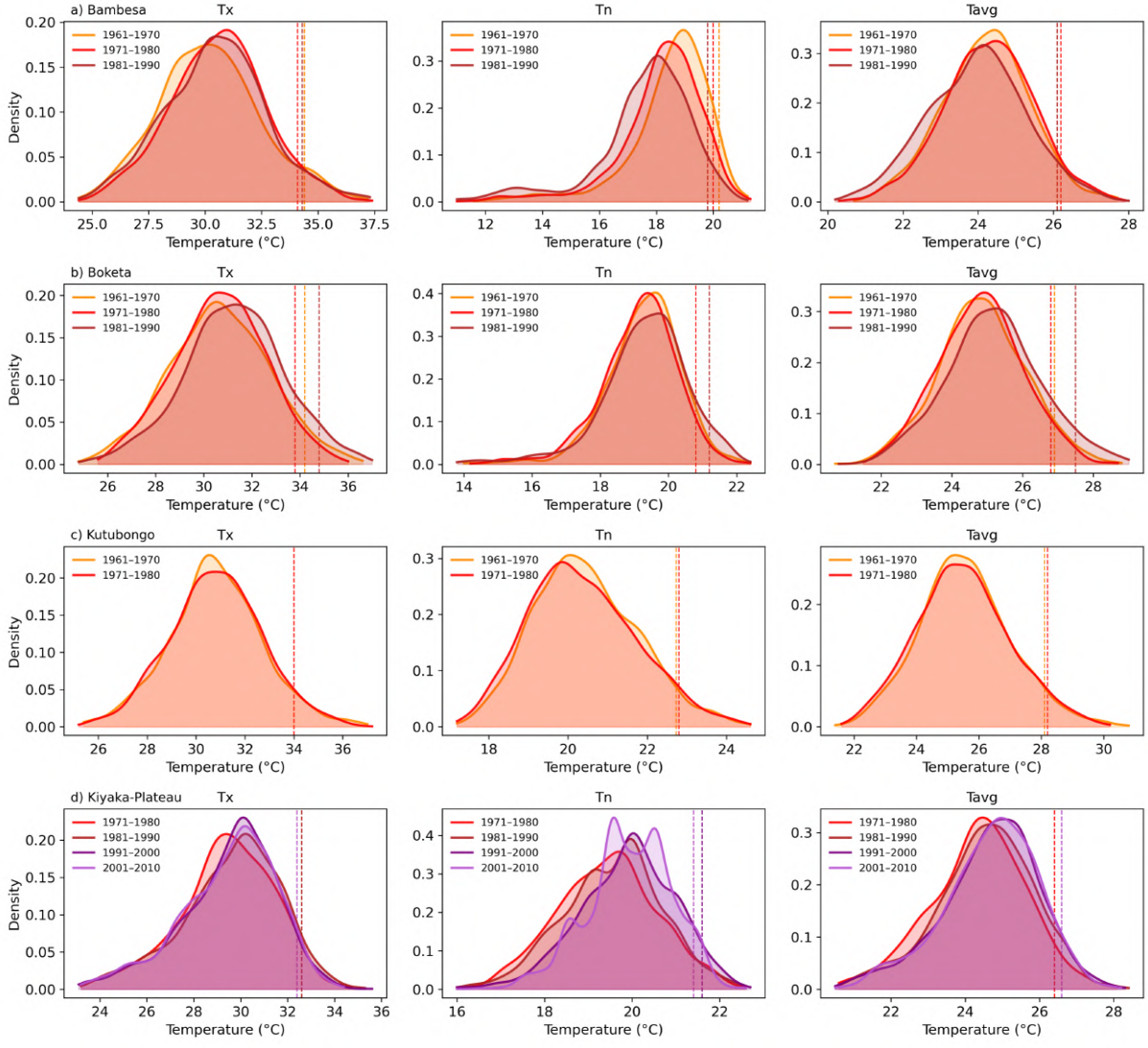


Figure E2. Decadal shifts in temperature distributions at stations: a) Bambesa (N 3°27' E 25°43'), b) Boketa (N 3°11' E 19°46'), c) Kutubongo (N 4°01' E 19°09'), and d) Kiyaka-Plateau (S 5°16' E 18°57') in the DRC. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). Vertical dashed lines indicate the 95th percentile values for each decade, with panels (a), (b) and (c) illustrating the progressive rightward shift towards higher temperatures and more extreme values over time.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

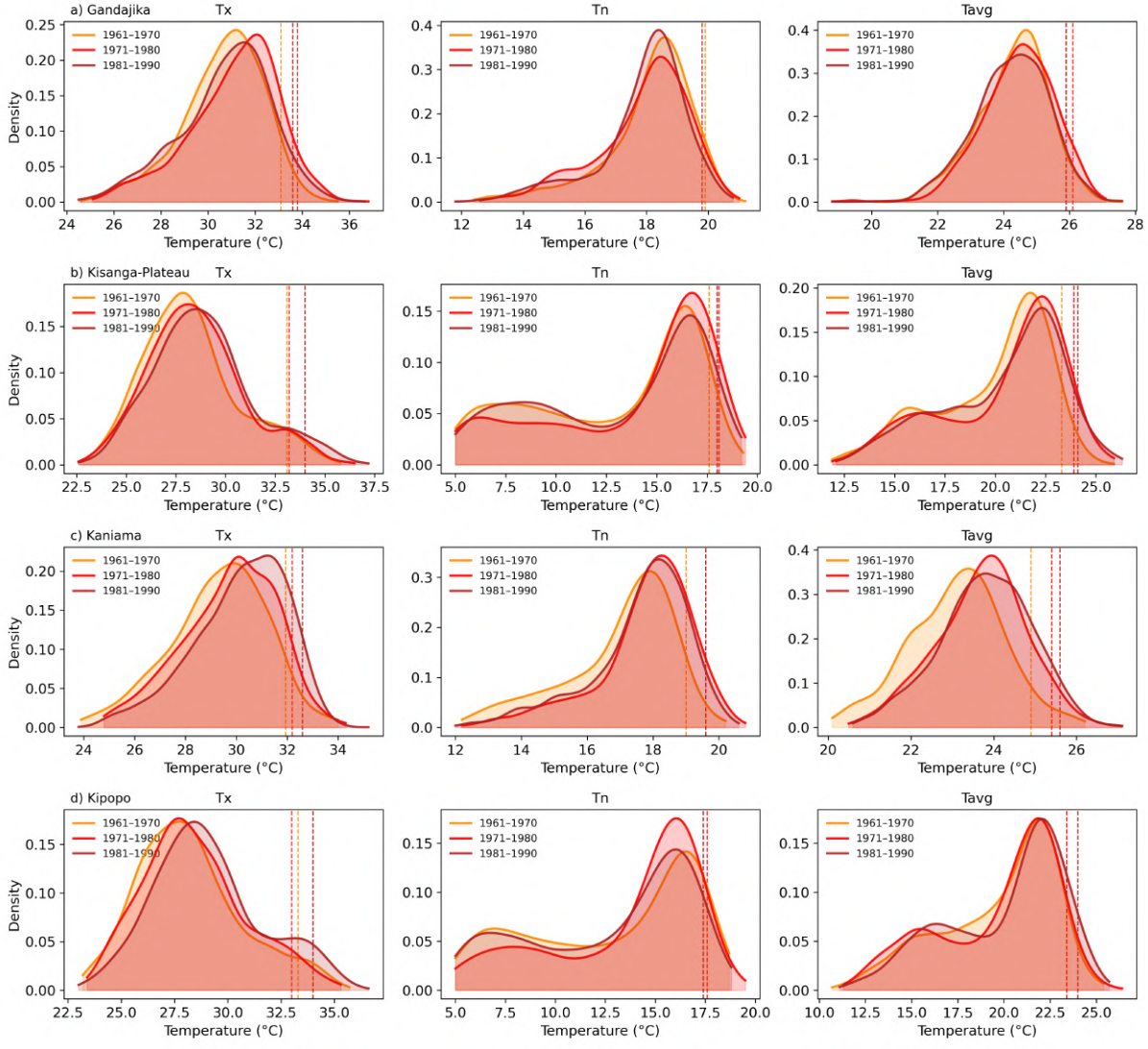


Figure E3. Decadal shifts in temperature distributions at stations: a) Gandajika (S 6°45' E 23°57'), b) Kisanga-Plateau (S 11°44' E 27°25'), c) Kaniama (S 7°25' E 24°09'), and d) Kipopo (S 11°34' E 27°24') in the DRC. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). Vertical dashed lines indicate the 95th percentile values for each decade, with panels (a), (b) and (c) illustrating the progressive rightward shift towards higher temperatures and more extreme values over time.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

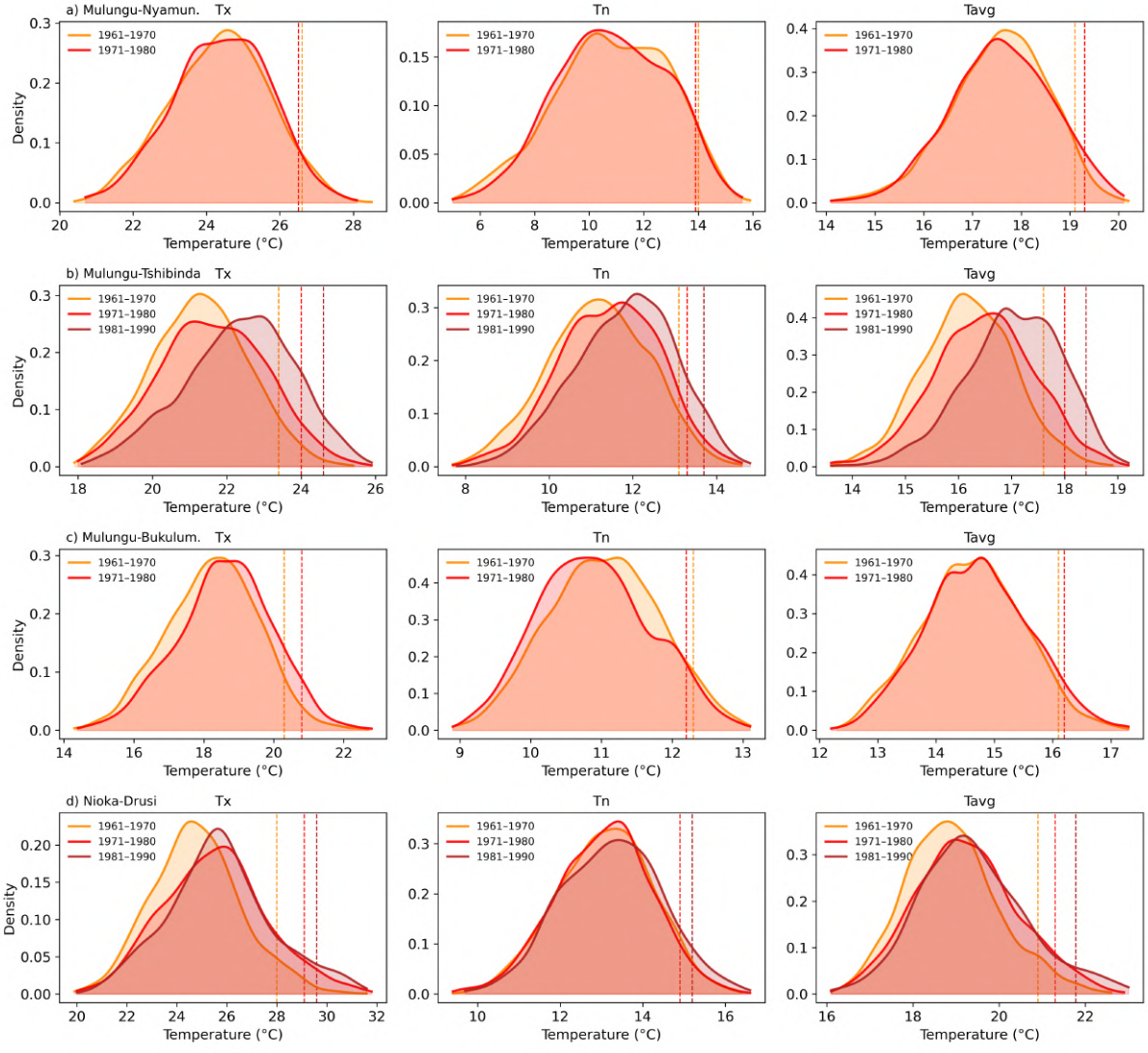


Figure E4. Decadal shifts in temperature distributions at stations: a) Mulungu-Nyamunyunye (S 2°18' E 28°48'), b) Mulungu-Tshibinda (S 2°19' E 28°45), c) Mulungu-Bukulumisa (S 2°20' E 28°43'), and d) Nioka-Drusi (N 2°09' E 30°39') in the DRC. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). Vertical dashed lines indicate the 95th percentile values for each decade, with panels (a), (b) and (c) illustrating the progressive rightward shift towards higher temperatures and more extreme values over time.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

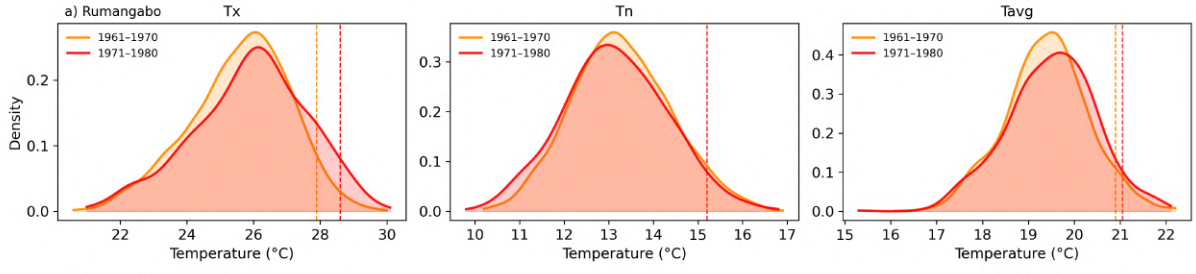


Figure E5. Decadal shifts in temperature distributions at station Rumangabo (S 1°21' E 29°22') in the DRC. Kernel density estimates of daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures are shown for each decade: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). Vertical dashed lines indicate the 95th percentile values for each decade, with panels (a), (b) and (c) illustrating the progressive rightward shift towards higher temperatures and more extreme values over time.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix F: Frequency of hot and cold temperature extremes at individual stations

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

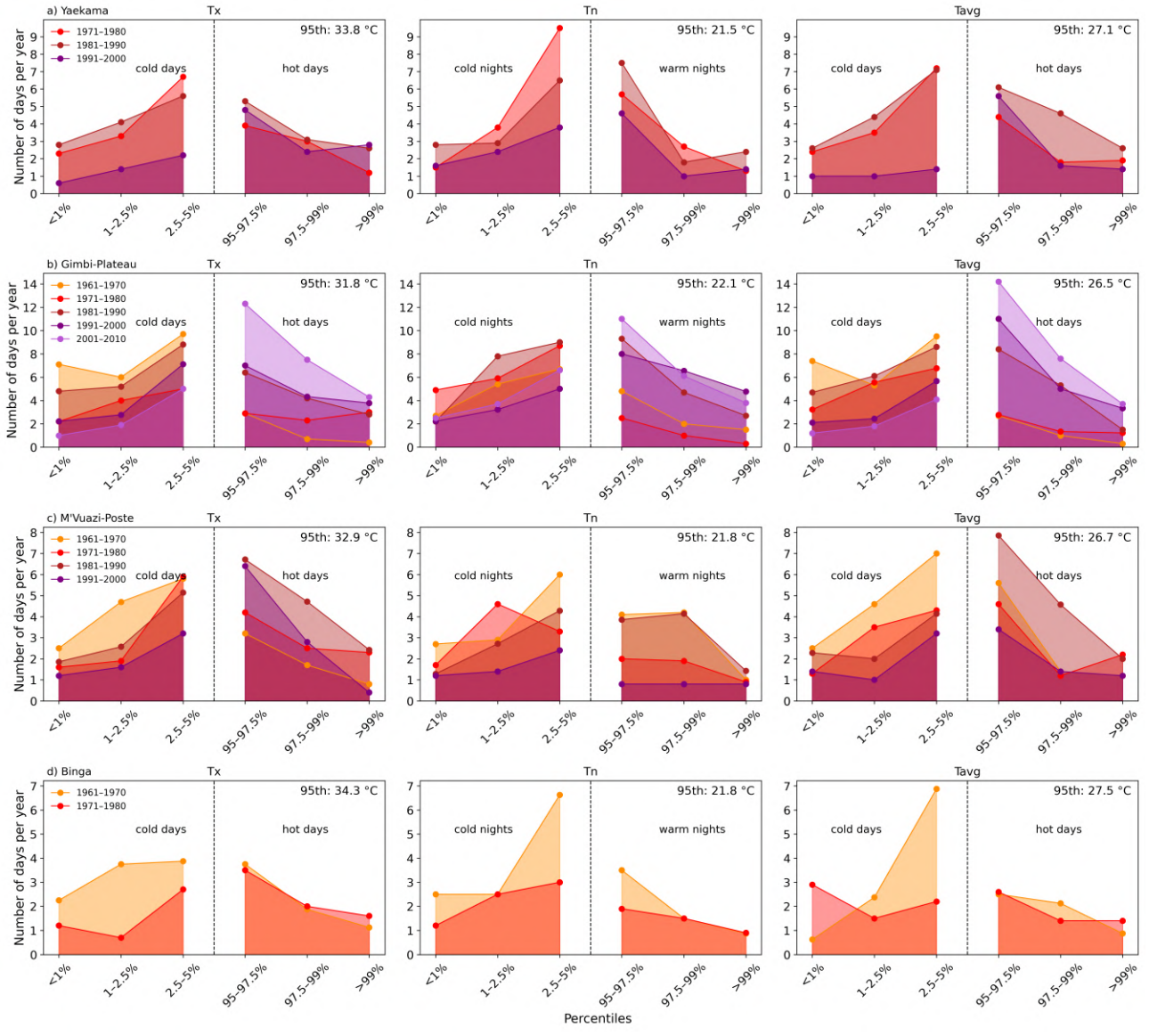


Figure F1. Frequency of hot and cold temperature extremes at stations: a) Yaekama (N 0°46' E 24°14'), b) Gimbi-Plateau (S 5°31' E 13°22'), c) M'Vuazi-Poste (Plateau) (S 5°27' E 14°54'), and d) Binga (N 2°18' E 20°30') in the DRC. The plots show the average number of days per decade falling below the 5th percentile (cold events) and above the 95th percentile (hot events) for daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures, similar to Lorenz et al. (2019). Values are based on kernel density estimates calculated across the entire period. The number of days per year within each decade are shown: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). The middle 90% of the distribution (5th–95th percentile range) is omitted for clarity. The 95th percentile temperature for each variable is noted in the top-right corner of each plot.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

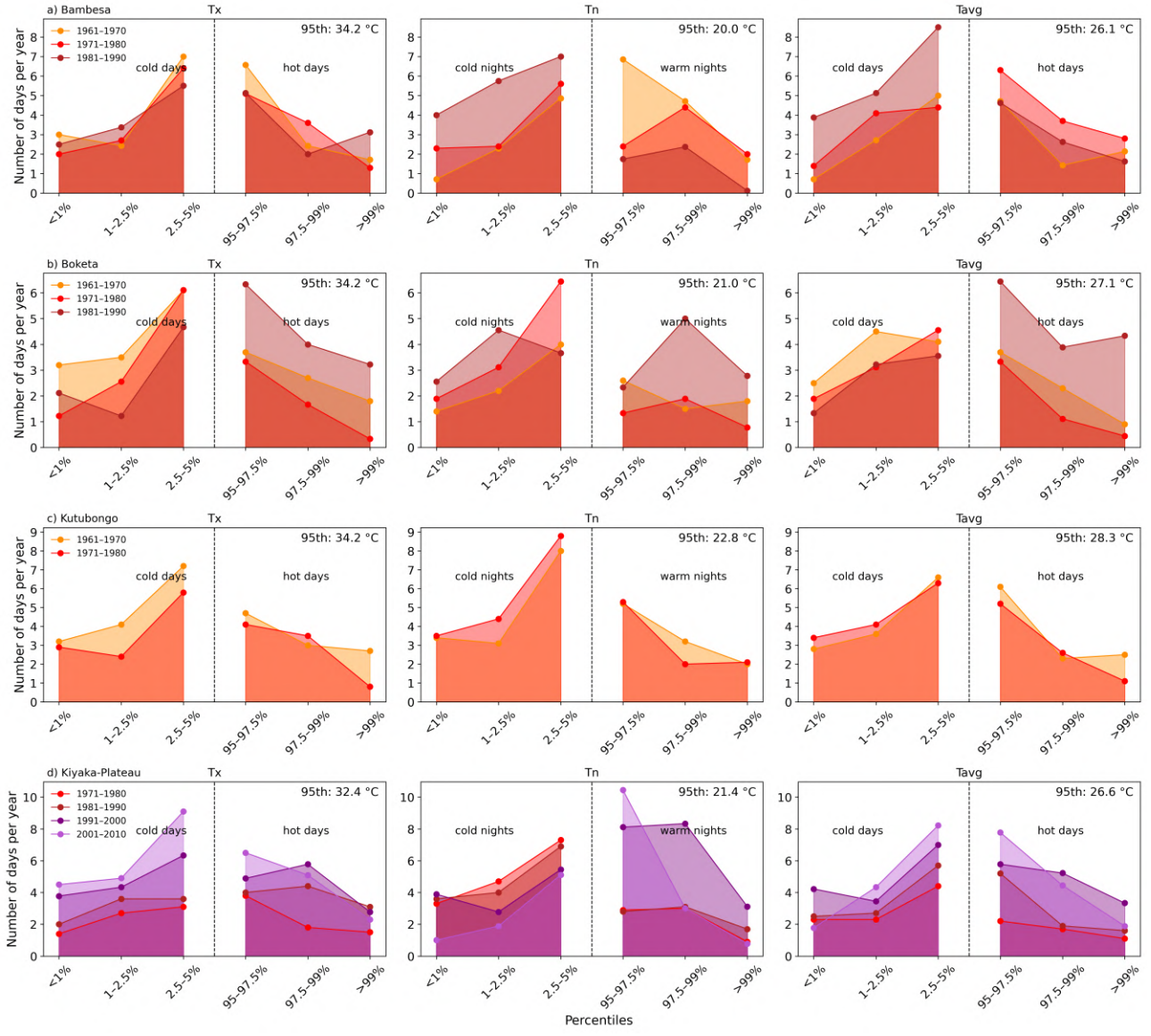


Figure F2. Frequency of hot and cold temperature extremes at stations: a) Bambesa (N 3°27' E 25°43'), b) Boketa (N 3°11' E 19°46'), c) Kutubongo (N 4°01' E 19°09'), and d) Kiyaka-Plateau (S 5°16' E 18°57') in the DRC. The plots show the average number of days per decade falling below the 5th percentile (cold events) and above the 95th percentile (hot events) for daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures, similar to Lorenz et al. (2019). Values are based on kernel density estimates calculated across the entire period. The number of days per year within each decade are shown: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). The middle 90% of the distribution (5th–95th percentile range) is omitted for clarity. The 95th percentile temperature for each variable is noted in the top-right corner of each plot.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

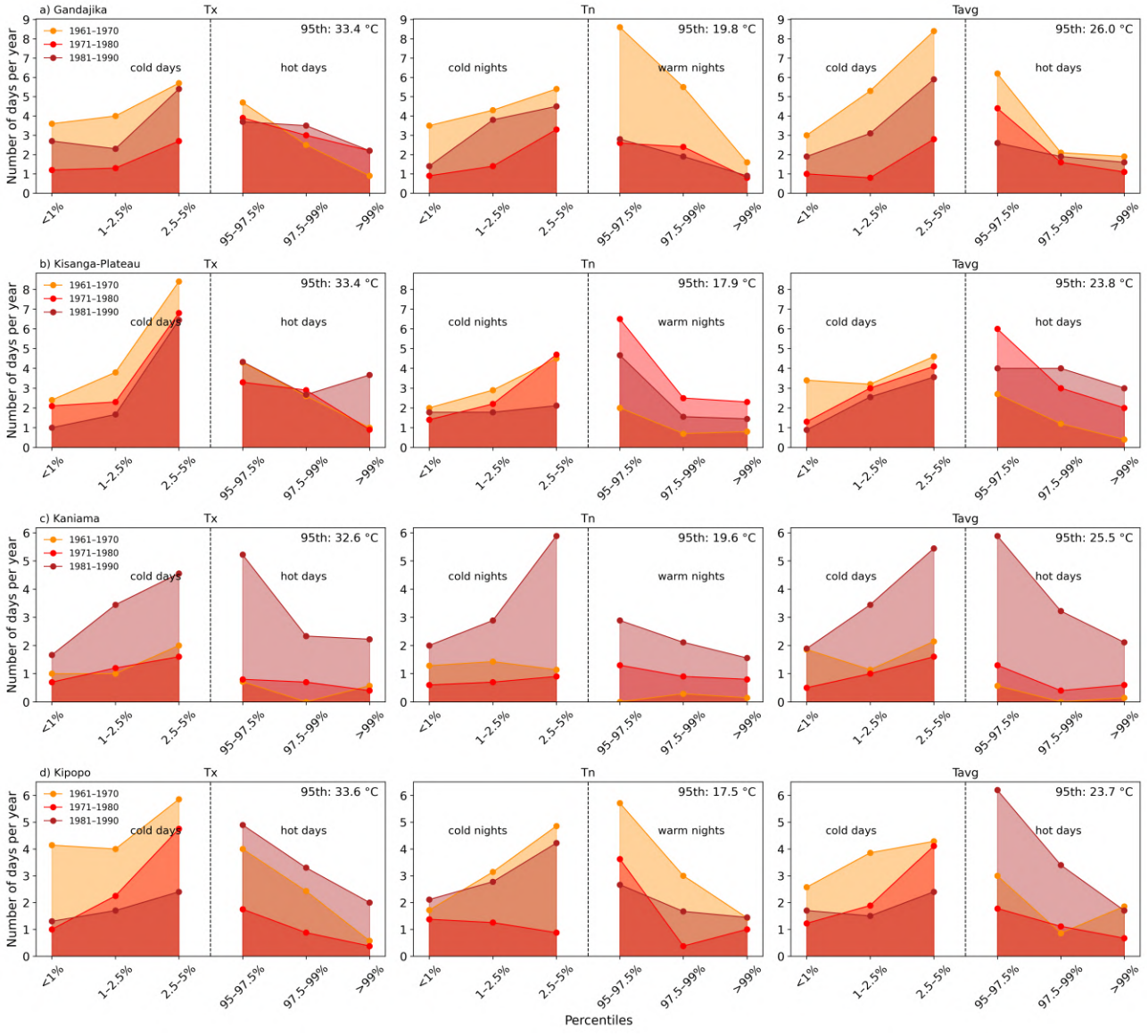


Figure F3. Frequency of hot and cold temperature extremes at stations: a) Gandajika (S 6°45' E 23°57'), b) Kisanga-Plateau (S 11°44' E 27°25'), c) Kaniama (S 7°25' E 24°09'), and d) Kipopo (S 11°34' E 27°24') in the DRC. The plots show the average number of days per decade falling below the 5th percentile (cold events) and above the 95th percentile (hot events) for daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures, similar to Lorenz et al. (2019). Values are based on kernel density estimates calculated across the entire period. The number of days per year within each decade are shown: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). The middle 90% of the distribution (5th–95th percentile range) is omitted for clarity. The 95th percentile temperature for each variable is noted in the top-right corner of each plot.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

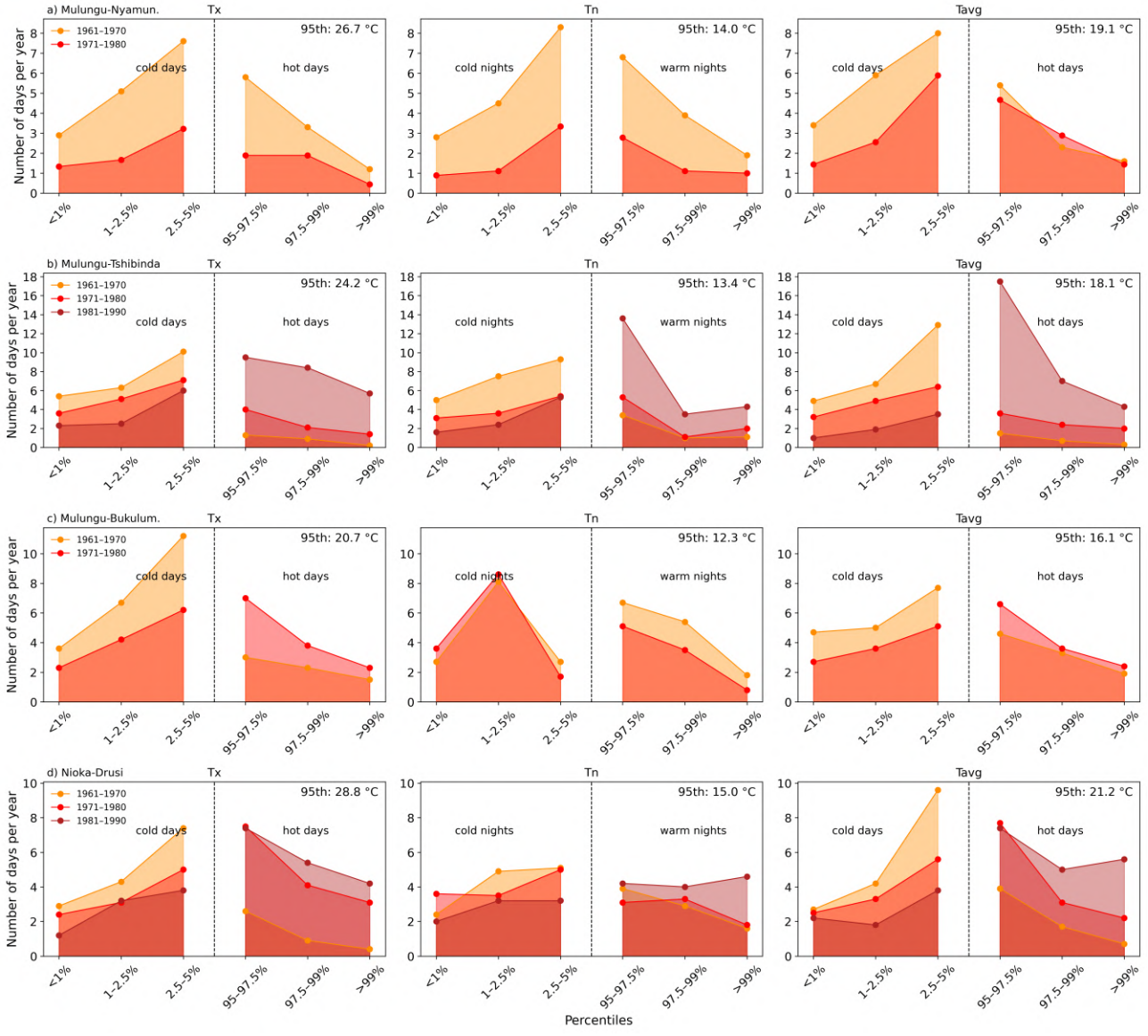


Figure F4. Frequency of hot and cold temperature extremes at stations: a) Mulungu-Nyamunyunye (S 2°18' E 28°48'), b) Mulungu-Tshibinda (S 2°19' E 28°45), c) Mulungu-Bukulumisa (S 2°20' E 28°43'), and d) Nioka-Drusi (N 2°09' E 30°39') in the DRC. The plots show the average number of days per decade falling below the 5th percentile (cold events) and above the 95th percentile (hot events) for daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures, similar to Lorenz et al. (2019). Values are based on kernel density estimates calculated across the entire period. The number of days per year within each decade are shown: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). The middle 90% of the distribution (5th–95th percentile range) is omitted for clarity. The 95th percentile temperature for each variable is noted in the top-right corner of each plot.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

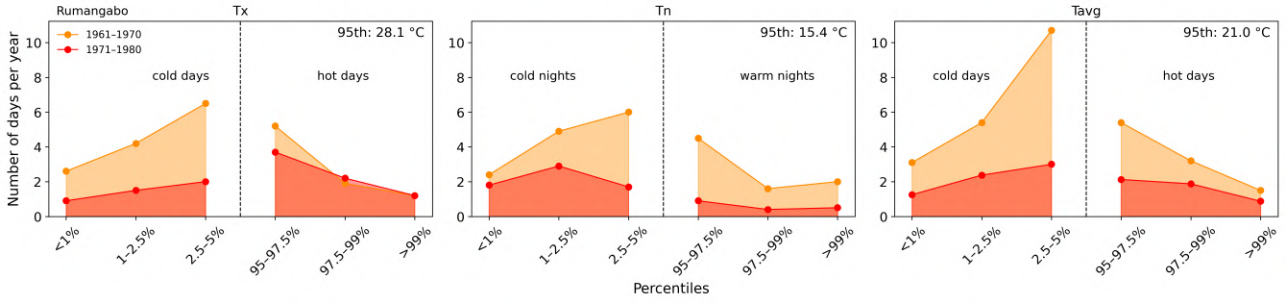


Figure F5. Frequency of hot and cold temperature extremes at station Rumangabo (S 1°21' E 29°22') in the DRC. The plots show the average number of days per decade falling below the 5th percentile (cold events) and above the 95th percentile (hot events) for daily maximum (Tx), minimum (Tn), and average (Tavg) temperatures, similar to Lorenz et al. (2019). Values are based on kernel density estimates calculated across the entire period. The number of days per year within each decade are shown: 1911-1920 (blue), 1921-1930 (light green), 1931-1940 (dark olive green), 1941-1950 (olive green), 1951-1960 (gold), 1961–1970 (yellow), 1971–1980 (red), 1981–1990 (maroon), 1991–2000 (indigo), 2001-2010 (light purple) and 2011-2020 (dark violet). The middle 90% of the distribution (5th–95th percentile range) is omitted for clarity. The 95th percentile temperature for each variable is noted in the top-right corner of each plot.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix G: Comparison of trends in daily maximum, minimum and average temperature at Individual Stations with ERA5-Land

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

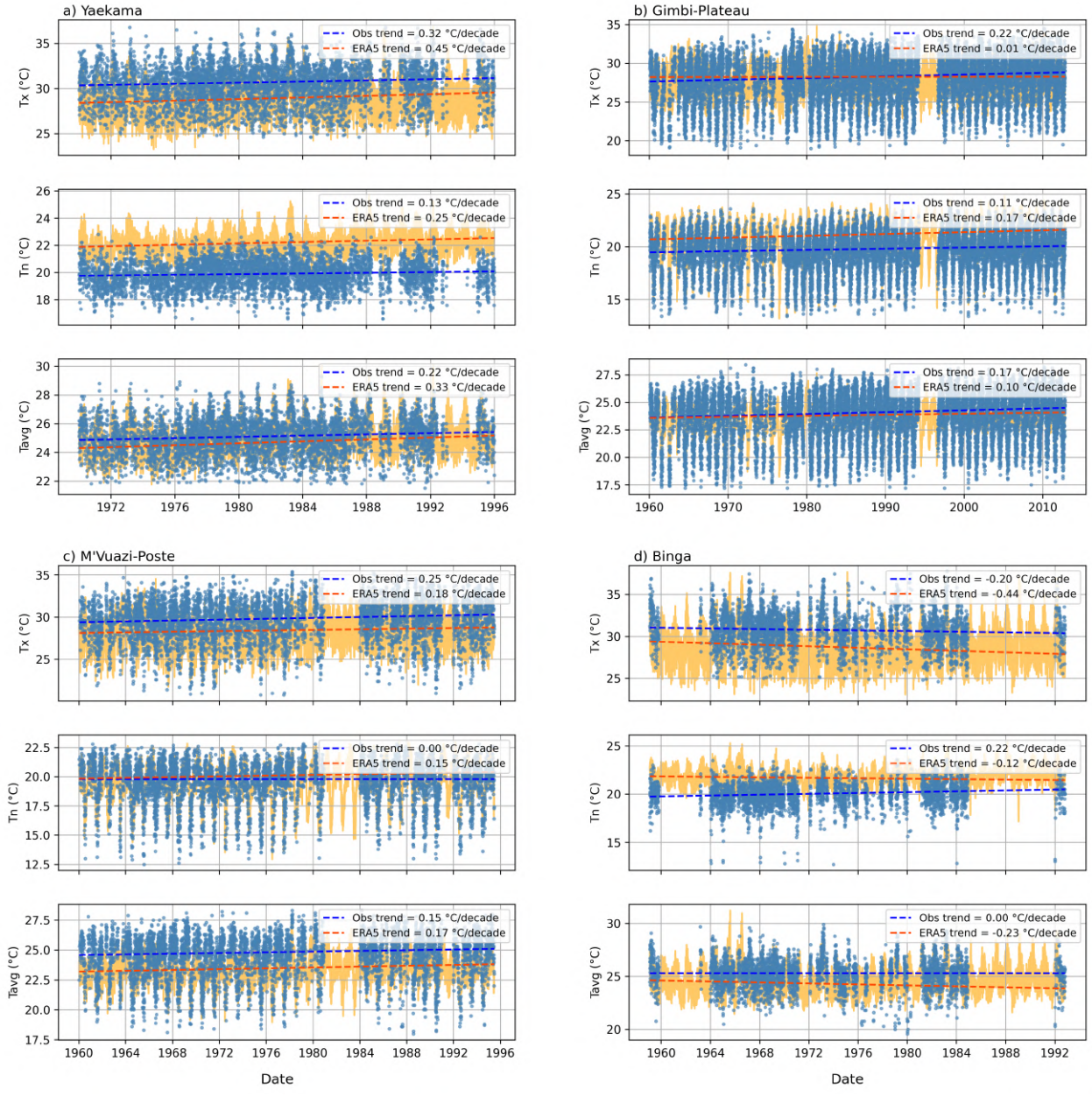


Figure G1. Comparison of trends in daily maximum, minimum and average temperatures at stations: a) Yaekama (N 0°46' E 24°14'), b) Gimbi-Plateau (S 5°31' E 13°22'), c) M'Vuazi-Poste (Plateau) (S 5°27' E 14°54'), and d) Binga (N 2°18' E 20°30') in the DRC. Blue dots represent daily values of observed maximum, minimum and average temperatures, while solid yellow lines show the ERA5-Land timeseries. The dotted dark blue and orange lines indicate Theil–Sen trend estimates for each variable for observations (Obs) and ERA5-Land, respectively, with the corresponding slope (°C/decade) shown in the top left legend. The observations include both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. The trends are estimated using values where both the ERA5-Land data and transcribed observations are available. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

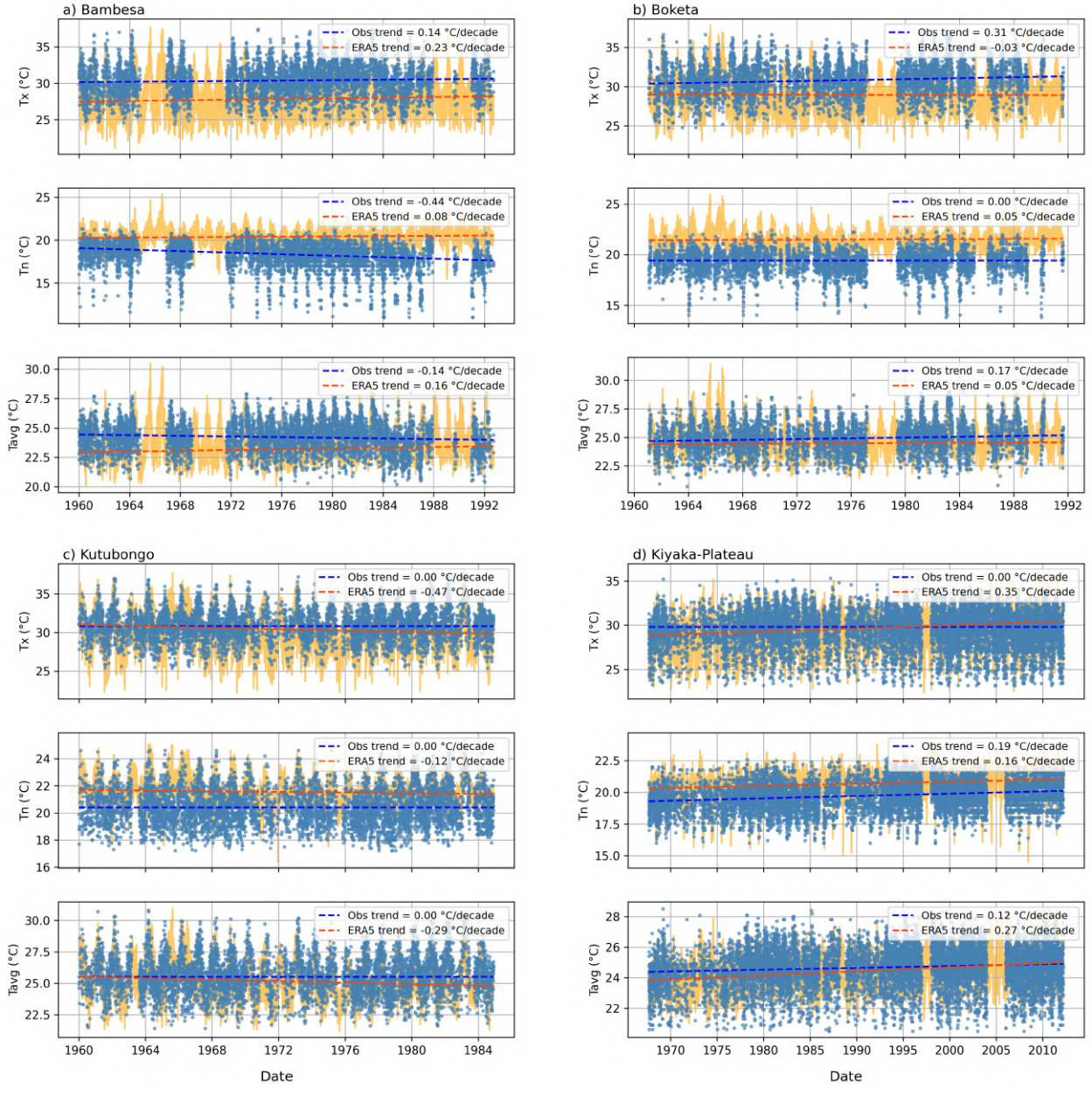


Figure G2. Comparison of trends in daily maximum, minimum and average temperatures at stations: a) Bambesa (N 3°27' E 25°43'), b) Boketa (N 3°11' E 19°46'), c) Kutubongo (N 4°01' E 19°09'), and d) Kiyaka-Plateau (S 5°16' E 18°57') in the DRC. Blue dots represent daily values of observed maximum, minimum and average temperatures, while solid yellow lines show the ERA5-Land timeseries. The dotted dark blue and orange lines indicate Theil–Sen trend estimates for each variable for observations (Obs) and ERA5-Land, respectively, with the corresponding slope (°C/decade) shown in the top left legend. The observations include both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. The trends are estimated using values where both the ERA5-Land data and transcribed observations are available. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

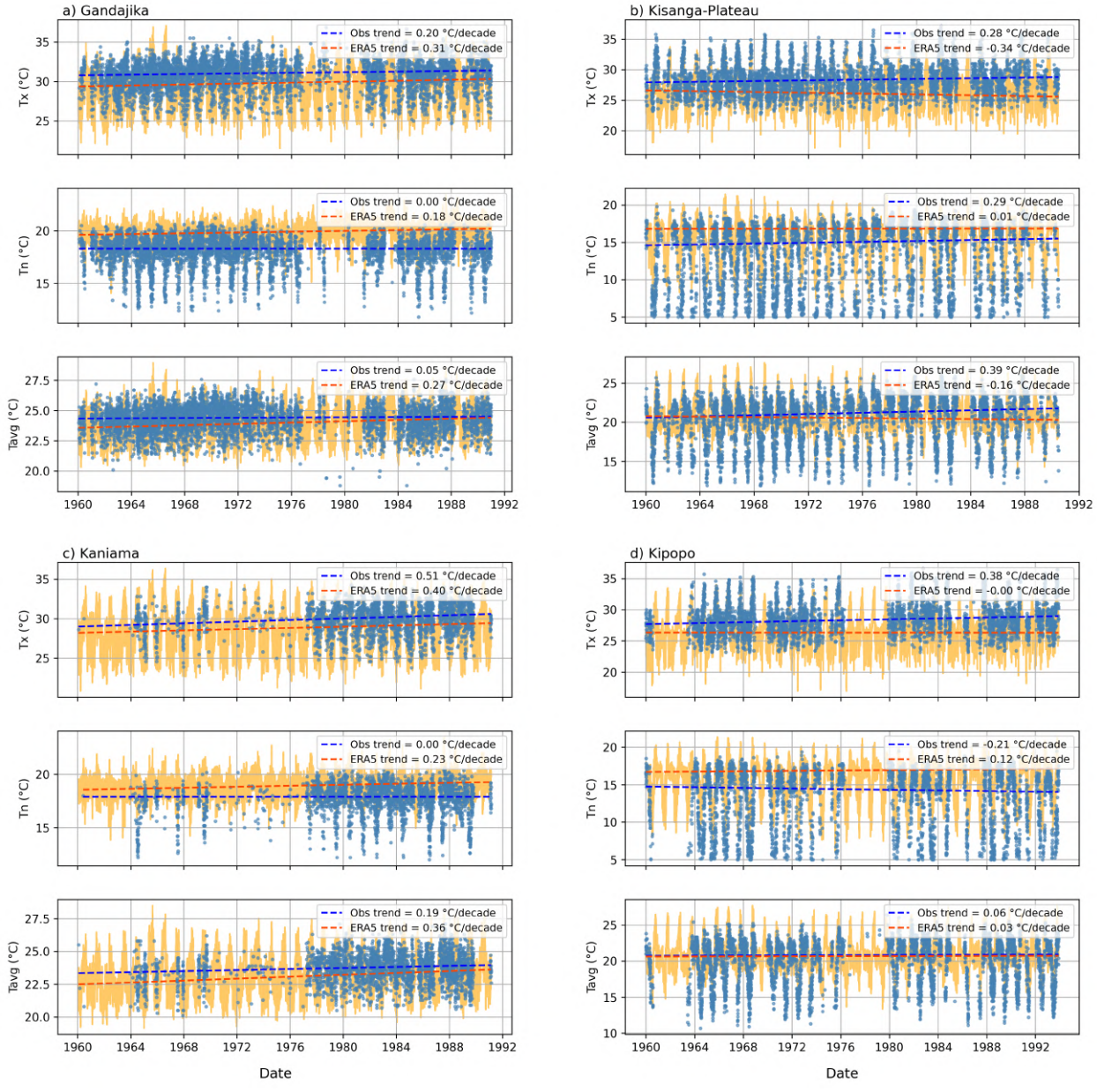


Figure G3. Comparison of trends in daily maximum, minimum and average temperatures at stations: a) Gandajika (S 6°45' E 23°57'), b) Kisanga-Plateau (S 11°44' E 27°25'), c) Kaniama (S 7°25' E 24°09'), and d) Kipopo (S 11°34' E 27°24') in the DRC. Blue dots represent daily values of observed maximum, minimum and average temperatures, while solid yellow lines show the ERA5-Land timeseries. The dotted dark blue and orange lines indicate Theil–Sen trend estimates for each variable for observations (Obs) and ERA5-Land, respectively, with the corresponding slope (°C/decade) shown in the top left legend. The observations include both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. The trends are estimated using values where both the ERA5-Land data and transcribed observations are available. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

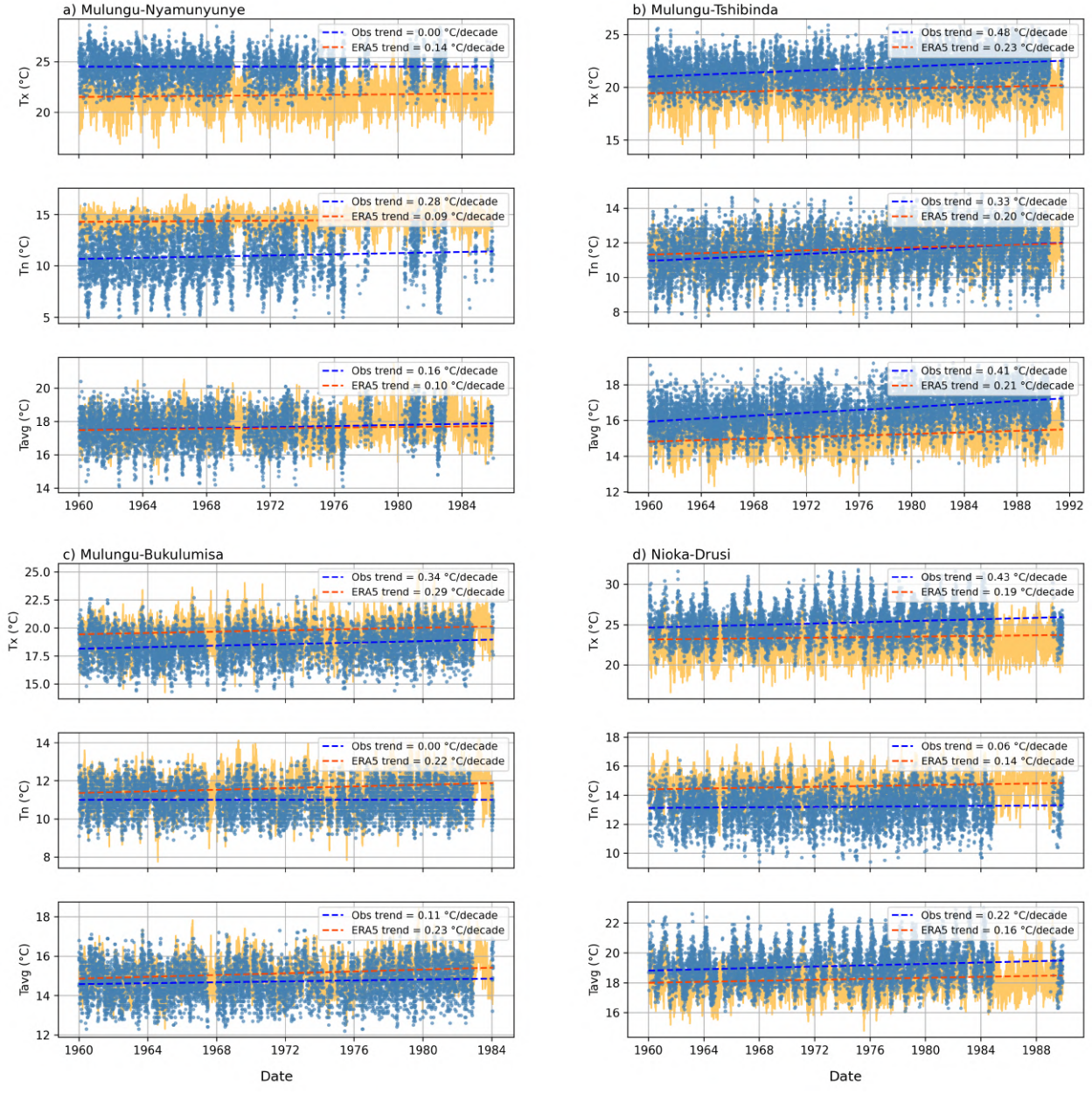


Figure G4. Comparison of trends in daily maximum, minimum and average temperatures at stations: a) Mulungu-Nyamunyunye (S 2°18' E 28°48'), b) Mulungu-Tshibinda (S 2°19' E 28°45), c) Mulungu-Bukulumisa (S 2°20' E 28°43'), and d) Nioka-Drusi (N 2°09' E 30°39') in the DRC. Blue dots represent daily values of observed maximum, minimum and average temperatures, while solid yellow lines show the ERA5-Land timeseries. The dotted dark blue and orange lines indicate Theil–Sen trend estimates for each variable for observations (Obs) and ERA5-Land, respectively, with the corresponding slope (°C/decade) shown in the top left legend. The observations include both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. The trends are estimated using values where both the ERA5-Land data and transcribed observations are available. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

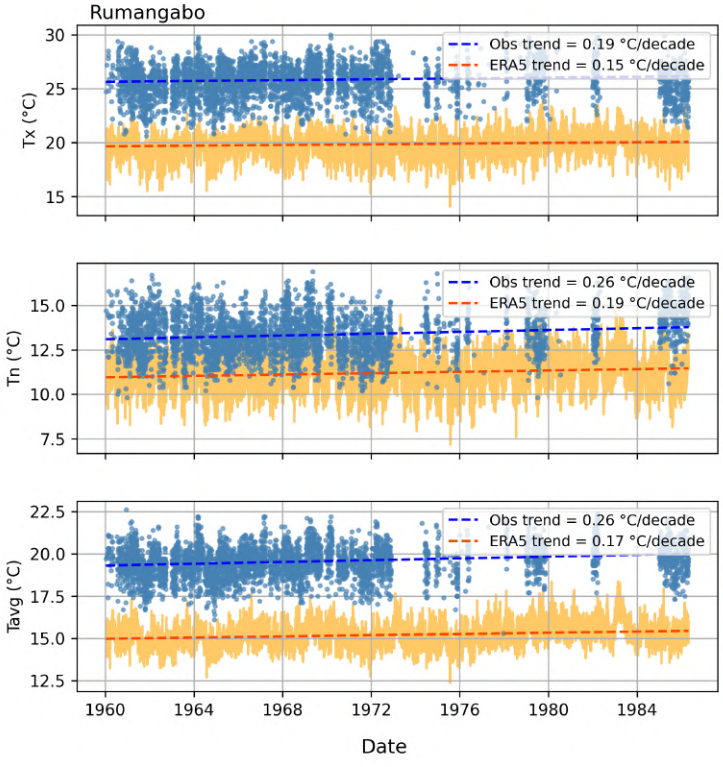


Figure G5. Comparison of trends in daily maximum, minimum and average temperatures at station Rumangabo (S 1°21' E 29°22') in the DRC. Blue dots represent daily values of observed maximum, minimum and average temperatures, while solid yellow lines show the ERA5-Land timeseries. The dotted dark blue and orange lines indicate Theil–Sen trend estimates for each variable for observations (Obs) and ERA5-Land, respectively, with the corresponding slope (°C/decade) shown in the top left legend. The observations include both automatically transcribed records using MeteoSaver and manually transcribed records previously available at INERA. The trends are estimated using values where both the ERA5-Land data and transcribed observations are available. Note that these stations have different start and end years.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Appendix H: ERA5-Land timeseries for daily maximum, minimum and average temperatures at stations: a) M'Vuazi-Poste (Plateau) (S 5°27' E 14°54'), and d) Binga (N 2°18' E 20°30') in the DRC.

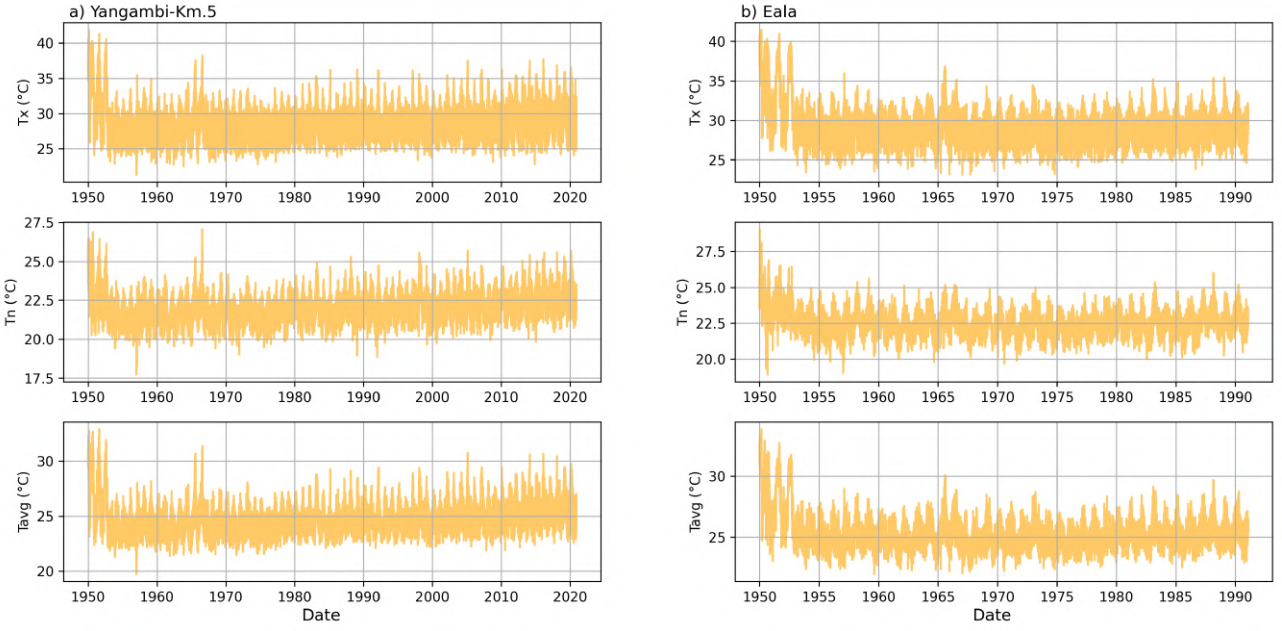


Figure H1. ERA5-Land timeseries for daily maximum, minimum and average temperatures at stations: a) Yangambi-Km.5 (N 0°46', E 24°29'), and b) Eala (N 0°03', E 18°18') in the DRC. The solid yellow lines show the ERA5-Land timeseries highlighting a pronounced jump in the 1950s for both stations that is not captured by stations with observations during that period (see Fig. 14 c-d). Note that these stations have different end years dependent on the transcribed in situ observations in the INERA dataset (as in Fig. 14 c-d)

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Author contributions. DM and WT designed the study. DM undertook the digitization of the post-1960 weather data sheets at the INERA archives. DKN, OKM, JM and TBL facilitated access to the INERA dataset (both hardcopy and manually-digitized data), while K. Hufkens, KJ, and HV provided access to the COBECORE dataset. DM, WT, K. Hufkens, and SL developed the scripts and conducted the development of MeteoSaver with assistance from BV, CV, EH, HV, JMB, KKTC, PB, and PT. EH, PT, K. Healion, SN, K. Hufkens, and KJ contributed expertise in data rescue throughout the project. All authors provided feedback during the data analysis and the writing of this paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. We would like to thank the Institut National pour l'Etude et la Recherche Agronomiques (INERA) situated in the Democratic Republic of the Congo (DRC) for granting us access to the extensive historical weather database available in the archives in Yangambi, DRC. We also thank the COBECORE project for providing access to pre-1960 DRC data, including manually transcribed records for the Eala, Bongabo, and Yangambi-Km.5 stations by Innocent Banzi Ngulu-kulu. DM is a research fellow at the Research Foundation - Flanders (11M8823N). WT received funding from the European Research Council (ERC) under the European Union's Horizon Framework research and innovation programme (grant agreement No 101076909; 'LACRIMA' project), the Interuniversity BOF projects (iBOF) (grant reference: iBOF/25/081; Cong'eau project), and the European Union's Horizon 2020 (grant agreement No 101081369; 'SPARCCLE'). Compute and storage resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation – Flanders (FWO) and the Flemish Government.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

References

ACP: La toiture de l'atelier mécanique de l'INERA/Yangambi détruite par le feu, Agence Congolaise De Presse, https://acp.cd/communication/environnement/la-toiture-de-latelier-mecanique-de-linera-yangambi-detruite-par-le-feu/, (last access: 17 July 2024), 2023.

Aguilar, E., Aziz Barry, A., Brunet, M., Ekang, L., Fernandes, A., Massoukina, M., Mbah, J., Mhanda, A., do Nascimento, D. J., Peterson, T. C., Thamba Umba, O., Tomou, M., and Zhang, X.: Changes in temperature and precipitation extremes in western central Africa, Guinea Conakry, and Zimbabwe, 1955–2006, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011010, 2009.

Alsdorf, D., Beighley, E., Laraque, A., Lee, H., Tshimanga, R., O'Loughlin, F., Mahé, G., Dinga, B., Moukandi, G., and Spencer, R. G. M.: Opportunities for hydrologic research in the Congo Basin, Reviews of Geophysics, 54, 378–409, https://doi.org/10.1002/2016RG000517, 2016.

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Soci, C., Villaume, S., Bidlot, J.-R., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.-N.: The ERA5 global reanalysis: Preliminary extension to 1950, Quarterly Journal of the Royal Meteorological Society, 147, 4186–4227, https://doi.org/https://doi.org/10.1002/qj.4174, 2021.

Bindoff, N., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X.: Detection and Attribution of Climate Change: from Global to Regional, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., book section 10, p. 867–952, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, ISBN ISBN 978-1-107-66182-0, https://doi.org/10.1017/CBO9781107415324.022, 2013.

Bolton, D.: The Computation of Equivalent Potential Temperature, Monthly Weather Review, 108, 1046 – 1053, https://doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2, 1980.

Bradski, G. and Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library, Computer Programming/Robotics, O'Reilly Media, Incorporated, ISBN 9780596516130, https://books.google.be/books?id=J0GbAgAAQBAJ, (last access: 09 February 2026), 2008.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D.: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/https://doi.org/10.1029/2005JD006548, 2006.

COBECORE: Congo basin eco-climatological data recovery and valorisation, https://cobecore.org/index.html, (last access: 25 June 2025), 2018.

CSC: Climate Service Centre Report No. 11, in: Climate Change Scenarios for the Congo Basin, edited by [Haensler, A., Jacob, D., Kabat, P., and F., L., Hamburg, Germany, https://www.climate-service-center.de/imperia/md/content/csc/csc-report11_optimized.pdf, (last access: 7 April 2026), 2013.

de Smeth, K., Comer, J., and Murphy, C.: Hydrometric data rescue and extension of river flow records: Method development and application to catchments modified by arterial drainage, Geoscience Data Journal, 11, 176–196, https://doi.org/10.1002/gdj3.206, 2024.

Depicker, A., Jacobs, L., Mboga, N., Smets, B., Van Rompaey, A., Lennert, M., Wolff, E., Kervyn, F., Michellier, C., Dewitte, O., and Govers, G.: Historical dynamics of landslide risk from population and forest-cover changes in the Kivu Rift, Nature Sustainability, 4, https://doi.org/10.1038/s41893-021-00757-9, 2021.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Dille, A., Dewitte, O., Handwerger, A. L., d'Oreye, N., Derauw, D., Ganza Bamulezi, G., Ilombe Mawe, G., Michellier, C., Moeyersons, J., Monsieurs, E., Mugaruka Bibentyo, T., Samsonov, S., Smets, B., Kervyn, M., and Kervyn, F.: Acceleration of a large deep-seated tropical landslide due to urbanization feedbacks, Nature Geoscience, 15, https://doi.org/10.1038/s41561-022-01073-3, 2022.

Doherty, A., Pearce, M., Calow, R., Daoust, G., Higazi, A., Burgin, L., and Osborne, R.: Climate risk report for the Central Africa region, Met Office, https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/services/government/central-africa-climate-risk-report-final.pdf, (last access: 18 February 2026), 2022.

Eyring, V., Gillett, N., Achuta Rao, K., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., book section 3, pp. 423–551, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009157896.005, 2021.

Folland, C. K., Rayner, N. A., Brown, S. J., Smith, T. M., Shen, S. S. P., Parker, D. E., Macadam, I., Jones, P. D., Jones, R. N., Nicholls, N., and Sexton, D. M. H.: Global temperature change and its uncertainties since 1861, Geophysical Research Letters, 28, 2621–2624, https://doi.org/https://doi.org/10.1029/2001GL012877, 2001.

Grant, L., Vanderkelen, I., Gudmundsson, L., Tan, Z., Perroud, M., Stepanenko, V. M., Debolskiy, A. V., Droppers, B., Janssen, A. B. G., Woolway, R. I., Choulga, M., Balsamo, G., Kirillin, G., Schewe, J., Zhao, F., del Valle, I. V., Golub, M., Pierson, D., Marcé, R., Seneviratne, S. I., and Thiery, W.: Attribution of global lake systems change to anthropogenic forcing, Nature Geoscience, 14, 849–854, https://doi.org/10.1038/s41561-021-00833-x, 2021.

Group, T. W. B.: Climate Risk Profile: Congo, Democratic Republic, World Bank Publications, https://climateknowledgeportal.worldbank.org/sites/default/files/2021-06/15883-WB_Congo%2C%20Democratic%20Republic%20Country%20Profile-WEB.pdf, (last access: 11 April 2026), 2021.

Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., Leonard, M., Liu, J., Schmied, H. M., Papadimitriou, L., Pokhrel, Y., Seneviratne, S. I., Satoh, Y., Thiery, W., Westra, S., Zhang, X., and Zhao, F.: Globally observed trends in mean and extreme river flow attributed to climate change, Science, 371, 1159–1162, https://doi.org/10.1126/science.aba3996, 2021.

Hawkins, E., Burt, S., McCarthy, M., Murphy, C., Ross, C., Baldock, M., Brazier, J., Hersee, G., Huntley, J., Meats, R., O'Grady, J., Scrimgeour, I., and Silk, T.: Millions of historical monthly rainfall observations taken in the UK and Ireland rescued by citizen scientists, Geoscience Data Journal, 10, 246–261, https://doi.org/10.1002/gdj3.157, 2022.

Hawkins, E., Brohan, P., Burgess, S. N., Burt, S., Compo, G. P., Gray, S. L., Haigh, I. D., Hersbach, H., Kuijjer, K., Martínez-Alvarado, O., McColl, C., Schurer, A. P., Slivinski, L., and Williams, J.: Rescuing historical weather observations improves quantification of severe windstorm risks, Nat. Hazards Earth Syst. Sci., 23, 1465–1482, https://doi.org/10.5194/nhess-23-1465-2023, 2023.

Hua, W., Zhou, L., Chen, H., Nicholson, S. E., Raghavendra, A., and Jiang, Y.: Possible causes of the Central Equatorial African long-term drought, Environmental Research Letters, 11, 1–13, https://doi.org/10.1088/1748-9326/11/12/124002, 2016.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009157896, (last access: 17 September 2024), 2021a.

IPCC: Summary for Policymakers, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors,

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 1–31, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009157896.001, 2021b.

Jacobsen, K., Van Hirtum, L., Amara, M., Beeckman, H., Meeus, S., Vandelook, F., Van den Bulcke, J., Stoffelen, P., Verbeeck, H., and Hufkens, K.: Climate data rescue from the Belgian colonial archives : helping to close the data-gap over Central Africa, in: Early Instrumental Meteorological Series, Conference abstracts, https://biblio.ugent.be/publication/8587593, (last access: 25 June 2025), 2018.

Jaczewski, A., Marosz, M., and Miętus, M.: PL1GD-T: a high-resolution gridded daily air temperature dataset for Poland, Earth System Science Data, 17, 3857–3871, https://doi.org/10.5194/essd-17-3857-2025, 2025.

Kasongo Yakusu, E., Van Acker, J., Van de Vyver, H., Bourland, N., Mbifo Ndiapo, J., Besango Likwela, T., Lokonda Wa Kipifo, M., Mbuya Kankolongo, A., Van den Bulcke, J., Beeckman, H., Bauters, M., Boeckx, P., Verbeeck, H., Jacobsen, K., Demarée, G., Gellens-Meulenberghs, F., and Hubau, W.: Ground-based climate data show evidence of warming and intensification of the seasonal rainfall cycle during the 1960–2020 period in Yangambi, central Congo Basin, CLIMATIC CHANGE, 176, 28, http://doi.org/10.1007/s10584-023-03606-0, 2023.

Latapy, A., Ferret, Y., Testut, L., Talke, S., Aarup, T., Pons, F., Jan, G., Bradshaw, E., and Pouvreau, N.: Data rescue process in the context of sea level reconstructions: An overview of the methodology, lessons learned, up-to-date best practices and recommendations, Geoscience Data Journal, 10, 396–425, https://doi.org/10.1002/gdj3.179, 2022.

List, R.: Smithsonian meteorological tables, Smithsonian miscellaneous collections, Smithsonian Inst., https://books.google.be/books?id=wws50AEACAAJ, (last access: 02 February 2026), 1966.

Lorenz, R., Stalhandske, Z., and Fischer, E. M.: Detection of a Climate Change Signal in Extreme Heat, Heat Stress, and Cold in Europe From Observations, Geophysical Research Letters, 46, 8363–8374, https://doi.org/10.1029/2019GL082062, 2019.

Mamba, F.: Tshopo: une partie de la toiture de la station climatologique de l'Inera/Yangambi réduite en cendres, mediacongo.net, https://www.mediacongo.net/article-actualite-117493_tshopo_une_partie_de_la_toiture_de_la_station_climatologique_de_l_inera_yangambi_reduite_en_cendres.html, (last access: 17 July 2024), 2023.

Mangaza, L., Batsi, G., Peroches, A., Masson, C., Sonwa, D. J., Lhoest, S., Makana, J.-R., Hubau, W., Lejeune, P., and Fayolle, A.: Dynamics and determinants of forest cover changes in the inner Congo basin, Trees, Forests and People, 23, 101 126, https://doi.org/10.1016/j.tfp.2025.101126, 2026.

Marvel, K. and Bonfils, C.: Identifying external influences on global precipitation, Proceedings of the National Academy of Sciences, 110, 19 301–19 306, https://doi.org/10.1073/pnas.1314382110, 2013.

Marvel, K., Cook, B. I., Bonfils, C. J. W., Durack, P. J., Smerdon, J. E., and Williams, A. P.: Twentieth-century hydroclimate changes consistent with human influence, Nature, 569, 59–65, https://doi.org/10.1038/s41586-019-1149-8, 2019.

Mawe, G. I., Landu, E. L., Dujardin, E., Imwangana, F. M., Bielders, C., Hubert, A., Michellier, C., Nzolang, C., Poesen, J., Dewitte, O., and Vanmaercke, M.: Mapping urban gullies in the Democratic Republic of the Congo, Nature, 644, https://doi.org/10.1038/s41586-025-09371-7, 2025.

Megevand, C.: Deforestation Trends in the Congo Basin: Reconciling Economic Growth and Forest Protection, World Bank, https://doi.org/10.1596/978-0-8213-9742-8, 2013.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/https://doi.org/10.1029/2011JD017187, 2012.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Muheki, D., Hufkens, K., Jacobsen, K., Verbeeck, H., Boeckx, P., Kankonde Ntumba, D., Kapalay Moulasa, O., Vercruysse, B., Birkholz, J. M., Verbruggen, C., Hawkins, E., Lampe, S., Kasongo Yakusu, E., Makanzu Imwangana, F., Mbifo, J., Besango Likwela, T., Meunier, F., Dewitte, O., Thorne, P., and Thiery, W.: MeteoSaver v1.1, https://github.com/VUB-HYDR/MeteoSaver/tree/version-1.1_including_precip_and_dry_and_wet_bulb_temp, (Last access: 27th Feb 2026), 2026a.

Muheki, D., Hufkens, K., Jacobsen, K., Verbeeck, H., Boeckx, P., Kankonde Ntumba, D., Kapalay Moulasa, O., Vercruysse, B., Birkholz, J. M., Verbruggen, C., Hawkins, E., Lampe, S., Kasongo Yakusu, E., Makanzu Imwangana, F., Mbifo, J., Besango Likwela, T., Meunier, F., Dewitte, O., Thorne, P., and Thiery, W.: Daily and Sub-daily Temperature and Daily Precipitation Records from 37 INERA stations across the Democratic Republic of the Congo (1.0), Zenodo, https://doi.org/10.5281/zenodo.18770063, [Data set], 2026b.

Muheki, D., Vercruysse, B., Chandrasekar, K. K. T., Verbruggen, C., Birkholz, J. M., Hufkens, K., Verbeeck, H., Boeckx, P., Lampe, S., Hawkins, E., Thorne, P., Ntumba, D. K., Moulasa, O. K., and Thiery, W.: MeteoSaver v1.0: a machine-learning based software for the transcription of historical weather data, EGUsphere, https://doi.org/10.5194/egusphere-2024-3779, preprint. Under review in Geoscientific Model Development (GMD), 2026c.

Muñoz Sabater, J.: ERA5-Land hourly data from 1950 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.e2161bac, (Accessed on 25-JUL-2025), 2019.

Noone, S., D'Arcy, C., Donegan, S., Durkan, W., Essel, B., Healion, K., Hersbach, H., Madden, S., Marshall, J., McConnell, L., Mensah, I., Scroxton, N., Thiesen, S., and Thorne, P.: Investigating the potential for students to contribute to climate data rescue: Introducing the Climate Data Rescue Africa project (CliDaR-Africa), Geoscience Data Journal, pp. 1–17, https://doi.org/10.1002/gdj3.248, 2024.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., and Parker, D.: Homogeneity adjustments of in situ atmospheric climate data: a review, International Journal of Climatology, 18, 1493–1517, https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T, 1998.

Ryu, S., Song, J. J., and Lee, G.: Interpolation of Temperature in a Mountainous Region Using Heterogeneous Observation Networks, Atmosphere, 15, https://doi.org/10.3390/atmos15081018, 2024.

Samset, B. H., Zhou, C., Fuglestvedt, J. S., Lund, M. T., Marotzke, J., and Zelinka, M. D.: Steady global surface warming from 1973 to 2022 but increased warming rate after 1990, Commun Earth Environ, 4, https://doi.org/10.1038/s43247-023-01061-4, 2023.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., book section 11, pp. 1513–1765, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009157896.013, 2021.

Sonwa, D. J., Nkem, J. N., Idinoba, M. E., Bele, M. Y., and Jum, C.: Building regional priorities in forests for development and adaptation to climate change in the Congo Basin, Mitigation and Adaptation Strategies for Global Change, 17, https://doi.org/10.1007/s11027-011-9335-5, 2012.

Tesseract OCR: tesstrain, https://github.com/tesseract-ocr/tesstrain, (last access: 7 October 2025), 2025a.

Tesseract OCR: tesseract, https://github.com/tesseract-ocr/tesseract, (last access: 7 October 2025), 2025b.

https://doi.org/10.5194/egusphere-2026-2107
Preprint. Discussion started: 4 May 2026

Weiss, H.: War and Peace in the Democratic Republic of the Congo, Current African issues, Nordiska Afrikaninstitutet, ISBN 9789171064585, https://books.google.be/books?id=4ZG_ALnn-70C, 2000.

Wilcox, R. R.: Theil-Sen estimator, in: Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy, pp. 194–197, Springer Nature, ISBN 978-1-4419-5524-1, https://doi.org/10.1007/978-1-4419-5525-8, 2001.

WMO: Guidelines on Best Practices for Climate Data Rescue, World Meteorological Organisation, https://library.wmo.int/idurl/4/55395, (last access: 30 September 2024), 2016.

Wongchuig, S., Papa, F., Fleischmann, A. S., Sierra, J. P., Boucharel, J., Espinoza, J. C., Kitambo, B., Oliveira, R. J., Paris, A., Paiva, R., Casas, P., and Tshimanga, R.: Recent significant drying in Central Congo Basin linked to weakened Walker circulation and warmer Atlantic, npj Climate and Atmospheric Science, 8, https://doi.org/10.1038/s41612-025-01225-3, 2025.

Węglarczyk, S.: Kernel density estimation and its application, ITM Web of Conferences: XLVIII Seminar of Applied Mathematics, 23, 8, https://doi.org/10.1051/itmconf/20182300037, 2018.

Yuh, Y. G., N'Goran, K. P., Kross, A., Heurich, M., Matthews, H. D., and Turner, S. E.: Monitoring forest cover and land use change in the Congo Basin under IPCC climate change scenarios, PLoS ONE, 19, https://doi.org/10.1371/journal.pone.0311816, 2024.

Zeppetello, L. R. V., Parsons, L. A., Spector, J. T., Naylor, R. L., Battisti, D. S., Masuda, Y. J., and Wolff, N. H.: Large scale tropical deforestation drives extreme warming, Environ. Res. Lett., https://doi.org/10.1088/1748-9326/ab96d2, 2020.

Zhang, X., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., Stott, P. A., and Nozawa, T.: Detection of human influence on twentieth-century precipitation trends, Nature, 448, 461–465, https://doi.org/10.1038/nature06025, 2007.

Zhou, L., Tian, Y., Myneni, R. B., Ciais, P., Saatchi, S., Liu, Y. Y., Piao, S., Chen, H., Vermote, E. F., Song, C., and Hwang, T.: Widespread decline of Congo rainforest greenness in the past decade, Nature, 508, 86–90, https://doi.org/10.1038/nature13265, 2014.