Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The impact of data missingness mechanisms and degradation levels on rainfall agroclimatic data analyses

Tinetariro Wesley Chikati

tinechik@gmail.com

Meteorological Services Department of Zimbabwe https://orcid.org/0009-0006-8891-2195

Moven Manjowe

Meteorological Services Department of Zimbabwe

Tapiwa Michelle Masawi

Meteorological Services Department of Zimbabwe

Isaac Masawana

Meteorological Services Department of Zimbabwe

Research Article

Keywords: monte carlo simulation, missingness, agro-climatic, imputation, degradation

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10446862/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

The impact of data missingness mechanisms and degradation levels on rainfall agroclimatic data analyses

Tinetariro Wesley Chikati[1*], Moven Manjowe[1 2], Tapiwa Michelle Masawi[1], Isaac Masawana[1]

[1]Zimbabwe Meteorological Services Department, Cnr Bishop Gaul and Hudson Avenue, Belvedere, Harare, Zimbabwe

[2]University of Zimbabwe, Mount Pleasant, Harare, Zimbabwe

*Corresponding author. E-mail: tinechik@gmail.com;

Contributing authors: manjowem@gmail.com; masawim@weatherzw.org.zw; masawanaisaac@gmail.com

Abstract

This study investigates the effects of data missingness on five agro-climatic indices: Maximum Consecutive Dry Days (CDD), Maximum Consecutive Wet Days (CWD), Seasonal Totals, Season Onset, and Season Cessation. These impacts were assessed under three patterns of data missingness, Missing at Random (MAR), Missing Completely at Random (MCAR), and Missing Not at Random (MNAR), as well as at three missingness volumes: 5%, 10%, and 20%. To characterise the effects, three error metrics were used to compare indices calculated from a degraded time series against those from a complete one, using a Monte Carlo simulation that cycled through all possible combinations of missingness pattern, volume, and temporal position within the time series. The results showed that the MNAR mechanism caused the most severe bias in the indices, while MAR caused the least severe bias. The results also showed that uncertainty in the indices increased as the missingness volume increased. Additionally, drier areas were less affected by missingness than wetter regions. Among the indices, onset and cessation were the most resilient to data missingness, while CDD was the most sensitive.

Keywords: Monte Carlo, Missingness, Agroclimatic, Bias, Imputation, Degradation

1 Introduction

Weather data such as maximum temperature, minimum temperature, evaporation and precipitation constitute the starting point of drought monitoring and hydrological modelling [1]. It is therefore important that these datasets are reliable for operational scientific analyses used for decision making in different fields. For example, Zimbabwe has an agriculture-based economy which contributes approximately 11-14 percent to the nation's Gross Domestic Product (GDP) [2]. It is critical that all analyses, such as the determination of average season onset dates and rainfall trends, are representative of what is on the ground, to prevent the dissemination of potentially misleading information. The use of inaccurate information can lead to poor planning, resulting in reduced yields at the individual farmer level and, more broadly, across the whole country, negatively affecting people's livelihoods.

While satellite datasets are acknowledged as an alternative that is usually complete but degraded by errors associated with remote sensing [3], ground station data is often seen as error free, leading to its use in various bias correction methods [4]. However, in-situ data are not problem free: in many parts of the developing world, data collection is inconsistent, and observation networks are fragmented. Consequently, climate records contain spatial gaps and temporal discontinuities, leading to various levels of degradation. These gaps and discontinuities propagate into the products derived from analysing such data, thereby introducing uncertainty into climate decision support systems [5].

Station data degradation can arise from many sources, not limited to transmission failure for automatic weather stations (AWS), manual instrument failure, station closure or relocation, and human reading and recording errors. Such events are rarely uniform, and statistically, data missingness can occur depending on observed values, be linked to unobserved processes, or occur completely at random. These patterns of data degradation are grouped into three categories: Missing at Random (MAR), Missing Not at Random (MNAR), and Missing Completely at Random (MCAR), respectively [6]. Missingness influences the statistical properties of data as well as the accuracy of analysis. As an illustration, if gaps are clustered in specific critical periods of a rainfall season, for example, the end of October into November in the context of

Zimbabwe, they may have a disproportionately large impact on season onset dates compared to uniformly distributed gaps.

The clustering can occur due to factors such as extreme weather events or station maintenance cycles. For example, during Cyclone Idai in 2019, some stations failed because of unusually heavy rainfall that they were not engineered or calibrated to handle, resulting in data loss and degradation in the period around mid-March [7]. As a result, the effects of such limitations are significant. Despite this, many studies in the climate domain implicitly assume that missing data are randomly distributed in time. For example, [8] and [9] all assumed randomness, as indicated by their gap-filling methodologies, using techniques such as Linear Interpolation, Radial Basis Function, and Multiple Linear Regression. The mechanisms underlying these techniques rely on this assumption. While this simplifies analysis, it misrepresents reality, as missingness usually follows different patterns.

While the issue of missing data is acknowledged, with some studies even setting thresholds for data completeness, there remains a gap in the systematic evaluation of the impacts of different degradation structures and missingness mechanisms on climate analyses. A limited number of studies compare the effects of temporally structured and random data loss. Furthermore, even fewer studies evaluate how various imputation methods perform under such a variety of degradation and missingness structures. This lack of evaluation leaves uncertainty in the generalisation of research findings.

To address this gap, a controlled experiment was carried out in which a complete dataset was purposely degraded according to the three missingness structures, allowing the simulation of different degradation patterns. The aim of this study is to simulate and quantify the effect of different data degradation mechanisms on climate analysis products, and to evaluate the robustness of selected common analysis types under varying conditions of missingness. Specifically, the study compares the sensitivity of five seasonal metrics: Maximum Consecutive Wet Days (CWD), Maximum Consecutive Dry Days (CDD) [10], Season Onset, Season Cessation [11], and Seasonal Total Rainfall, to data missingness. These metrics highlight different properties of the rainfall season, such as rainfall distribution, intra-season structure, seasonal boundaries, and raw rainfall amount, respectively.

2 Materials and methods

2.1 Study area

Zimbabwe is situated in Southern Africa, as shown in Fig 1. The country has a sub-tropical climate, divided into seven Agro-Ecological Zones (AEZ), with yearly rainfall ranging from at least 1000 mm yr^{-1} in Region I to 300 mm yr^{-1} in Region Vb [12]. Zimbabwe has two distinct weather regimes, the wet and dry seasons. The wet season spans two contiguous years, running from October to April of the following year, while the dry season runs from May to September [13]. The country's weather and climate are monitored by the Zimbabwe Meteorological Services Department (ZMSD), which operates a network of 47 stations collecting key variables such as rainfall and temperature across the country's 10 provinces.

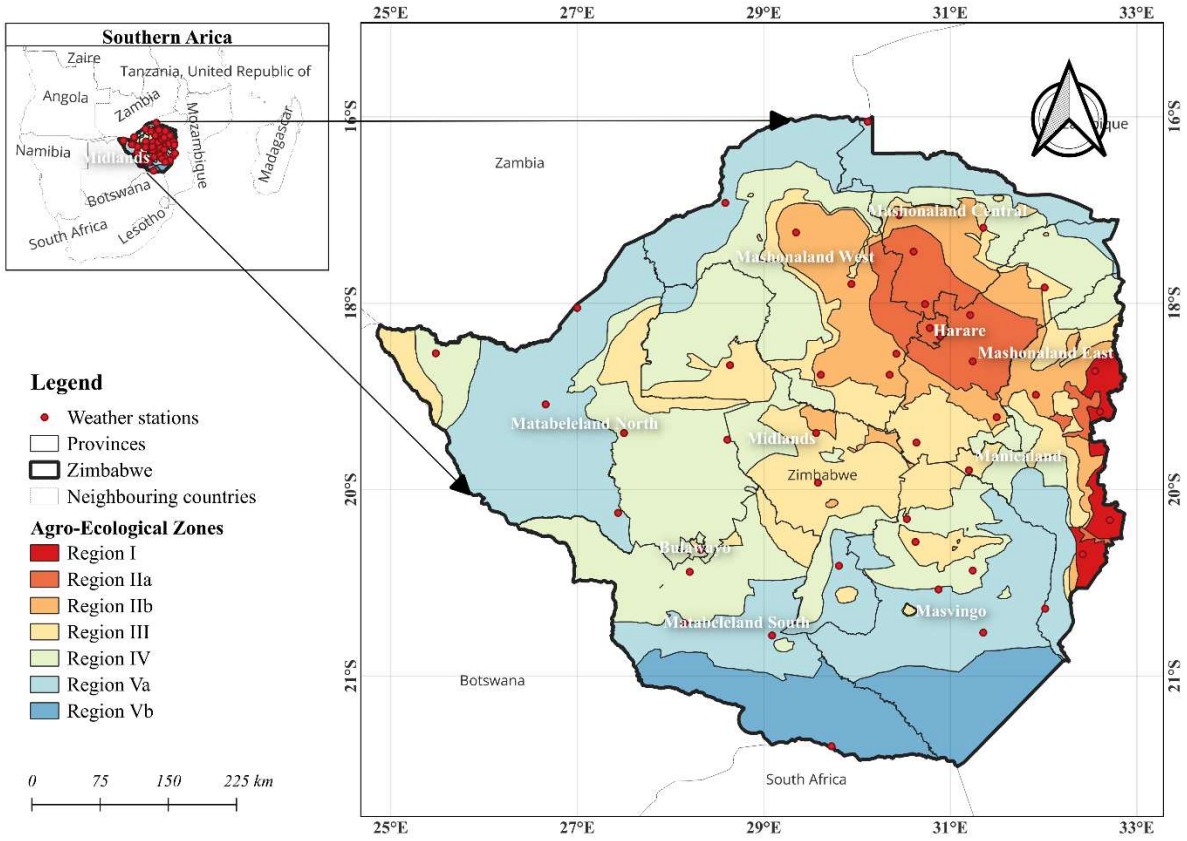


Fig 1. Map of the study area showing the agro-ecological zones and manned weather stations in Zimbabwe

2.2 Datasets

Zimbabwe's revised AEZs [12] were obtained in Esri Shapefile format, and daily satellite rainfall estimates from 1981 to 2025 were acquired from CHIRPS version 2 (CHIRPS-V2). The dataset is available in gridded format, with a spatial resolution of 0.05° x 0.05° [14]. CHIRPS-V2 was used in this study as it has a proven record of being more accurate than other datasets, over Zimbabwe as shown by [15]. Forty-seven weather station coordinates were acquired from the ZMSD (Zimbabwe Meteorological Services Department, 2026), and seven stations were drawn at random, one from each unique AEZ. Only one station per zone was chosen to reduce simulation processing time. Table 1 shows the selected stations.

Table 1. Randomly selected stations, their coordinates, agro-ecological zones, and seasonal rainfall totals [12]

Station	X	Y	Agro-Ecological Zone	Annual Rainfall (mm)
Chipinge	32.62	-20.20	Region I	< 1000
Henderson	30.97	-17.58	Region IIa	750 - 1000
Kadoma	29.88	-18.32	Region IIb	750 - 1000
Mt Darwin	31.58	-16.78	Region III	650 - 800
Matopos	28.51	-20.38	Region IV	450 - 650
Binga	27.33	-17.62	Region Va	< 650
Beitbridge	29.99	-22.21	Region Vb	< 600

2.3 Methods

2.3.1 Monte Carlo simulation

To determine the effects of missingness, a Monte Carlo Simulation was implemented. The simulation was implemented programmatically using Python [16], supported by three main modules: NumPy [17], pandas [18], and SciPy [19]. Additionally, tqdm was used to show the progress of the simulation runs [20]. The rainfall data were artificially degraded according to the three data missingness mechanisms, MAR, MCAR, and MNAR, and for each of these structures the data were deleted at three different volumes, 5%, 10%, and 20%, to determine the impact of different missingness

volumes. The simulations were carried out on a computer equipped with a 12th Generation Intel® Core™ i5-1235U processor operating at 1.30 GHz, featuring 12 processing cores and 24 GB of RAM, running the Windows 11 Pro operating system.

A critical parameter in MCS is the number of individual repetitions (N). This is the number of iterations required to achieve a stable estimation of the target performance metrics without incurring a large computational overhead, while also reaching acceptable precision. A seed value of 42 was used, since it has been widely used in simulation studies. To ensure the missingness structures were reproducible and deterministic, a pseudo-random number generator (PRNG) was used to produce two random numbers, which could also be used as seeds for initialising the simulation, and these were then tested for convergence against the simulation results generated using 42. This was achieved by plotting the Spearman's correlation coefficients of the results generated from the simulation runs with the randomly generated seeds against the results produced from the run with seed 42, on a scatterplot. This was done to ensure that the observed differences in the performance metrics could be attributed only to missingness, rather than to random variation in the PRNG (Morris et al., 2019). N was determined using the formula:

$$\mathrm{N} = \frac{\mathrm{E(Coverage)} \times (1 - \mathrm{E(Coverage)})}{(\text{Monte Carlo SE})^2} \quad (1)$$

where N is the number of optimal simulations, E(Coverage) is the percentage of repetitions required to produce a true value, and SE is the standard error of estimation due to the use of a limited number of simulations. This was followed by the evaluation of the seasonal metrics: CDD, CWD, Season Onset, Season Cessation, and Seasonal Total Rainfall. The analysis was also carried out individually for the wet and dry seasons, as well as for each of the seven locations selected at random from Zimbabwe's seven AEZs.

2.4 Seasonal metrics calculation

2.4.1 Maximum consecutive dry days and consecutive wet days

A wet day is defined differently in different contexts. In this study, an agricultural context was adopted, defining a wet day as a day receiving greater than or equal to 3 mm of precipitation within 24 hours (from 0600z to 0559z the next day). Therefore,

CDD is the maximum number of consecutive days receiving precipitation of at least 3 mm (rainfall ≥ 3 mm d^{-1}). Consequently, CWD was defined as the maximum number of consecutive days receiving precipitation of less than 3 mm (rainfall < 3 mm d^{-1}) [21].

2.4.2 Season onset, cessation and seasonal total rainfall

Season Onset was defined as the first day after 1 October on which 20 mm is received within two consecutive days, given that a dry spell does not occur in the next 20 days of the wet spell [22,23]. Cessation was defined via a daily soil water balance model, developed by [24], as the first day after Season Onset when the cumulative water balance drops to 0 mm after subtracting a potential evapotranspiration loss of 5 mm d^{-1}. Finally, Seasonal Total Rainfall was defined as the summation of daily rainfall from 1 October to 30 April.

2.5 Determination of variance from the complete timeseries

To determine the statistical impact of missingness, the degraded time series were compared to the complete reference dataset. Rainfall is known to be non-normally distributed due to the presence of many 0 mm day^{-1} values; therefore, traditional error metrics such as Mean Bias and Standard Deviation were avoided, since they rely on the assumption that their residuals follow a normal distribution [25]. As a result, more robust error metrics were used. Median Bias was used to determine the systematic error since it is less affected by outliers and extensive data gaps [26,27]. To evaluate the dispersion of errors from the complete timeseries, the Normalised Median Absolute Deviation (NMAD) was used, as it requires at least 50% of the dataset to be missing to break down [28,29]. NMAD was then standardised against the baseline index from uncorrupted data to give a relative percentage, allowing comparison across different indices with different units of measurement. Finally, the impact of missingness on the temporal coherence of the time series was assessed using Spearman's correlation coefficient (ρ) [30]. The coefficient ranges from -1 to +1; values close to either extreme indicate strong correlation, while values close to zero indicate weak correlation.

3 Results

3.1 Seed sensitivity testing

The Pseudo-Random Number Generator generated 28,876,402 and 1,352,964,956 as the alternative seeds. Fig 2a and Fig 2b show the agreement between the Spearman's ρ from running the simulation with seed 42 versus 28,876,402, and seed 42 versus 1,352,964,956, respectively. The coefficient showed a strong positive correlation, with ρ = 0.9998 for both comparisons. As a result, 42 was used as the seed for the full run, considering its wide usage in the literature as well as for simplicity. Using Eq. (1) to determine the optimal number of runs showed that at least 1600 simulations were required, and 2000 was used instead, resulting in the simulation running for 309.1 minutes.

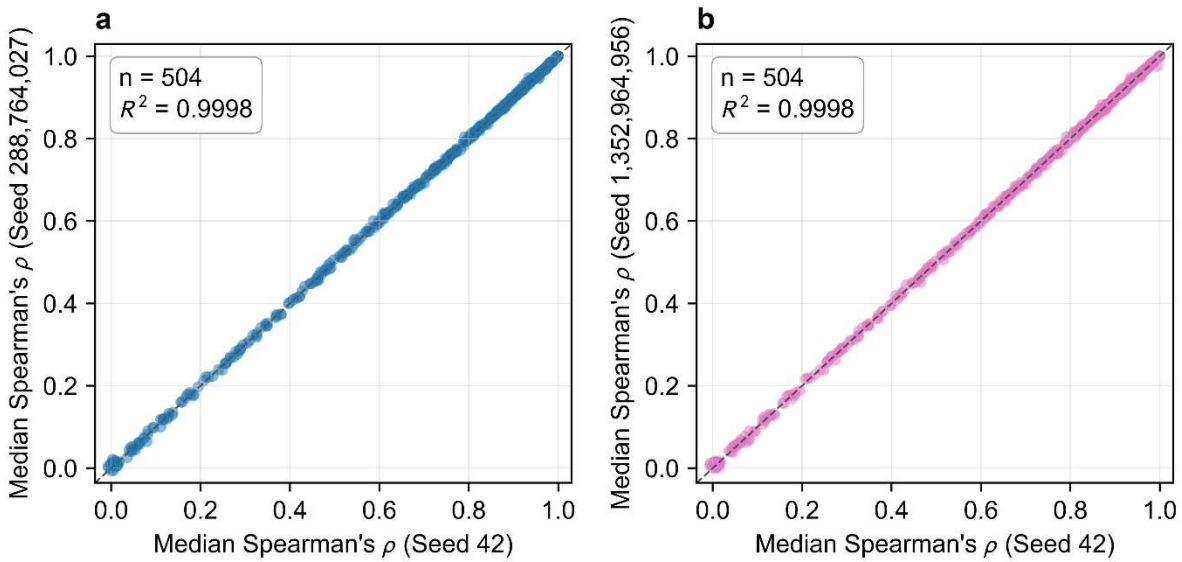


Fig 2. Scatterplots showing the correlation of randomly generated seeds with 42

3.2 Variance metrics

3.2.1 Normalised median absolute deviation

Fig 3a and Fig 3b show the Normalised Median Absolute Deviations (NMAD) for the wet and dry seasons, respectively. In the wet season, Season Onset and Cessation were the least affected by most missingness scenarios. Onset had NMAD=0% across all volumes, while Cessation peaked at 4% under the MNAR scenario with the maximum missingness volume of 20%. Under the MCAR mechanism, the NMAD for CDD scaled from 17% at 5% missingness volume to 56% at 20% missingness volume.

Seasonal Total Rainfall and CWD showed moderate vulnerability, with bias worsening under the MNAR mechanism, reaching 36% and 38%, respectively, at 20% missingness volume. In the dry season, CWD and Seasonal Total Rainfall had NMAD=0% for most scenarios, except for 20% MCAR and 20% MAR for Total Rainfall, as shown in Fig 3b. However, CDD variance became overwhelmingly high, ranging from 64% to a severe peak of 92% across all scenarios. This indicated that missing data during the dry season catastrophically destabilises the calculation of consecutive dry spells, regardless of whether the missingness is random (MCAR) or systematic (MNAR).

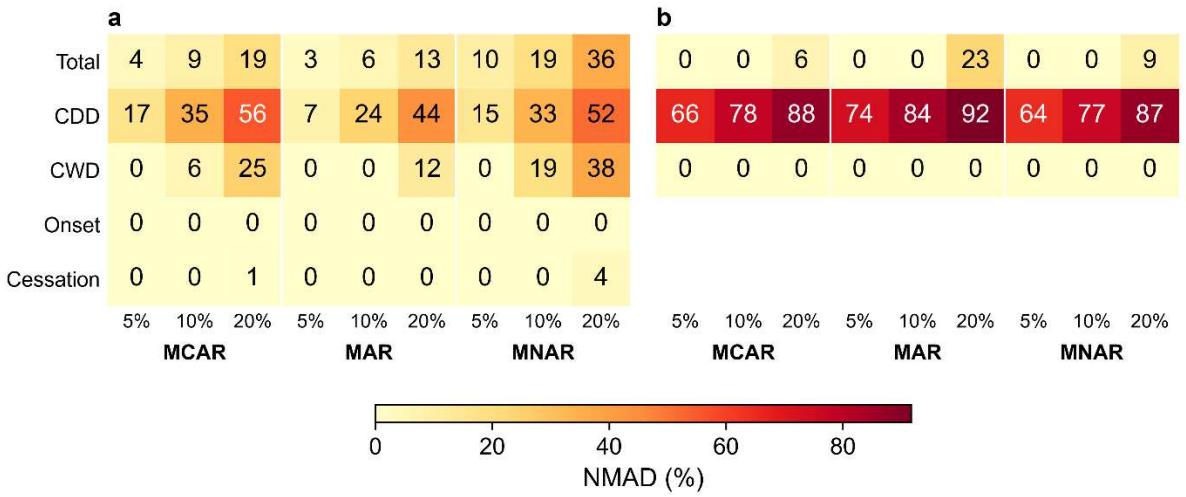


Fig 3. Effects of missingness mechanisms and volume on the Normalised Median Absolute Deviation (NMAD) of rainfall season indices in the (a) wet season and (b) in the dry season. Values are medians across all stations. Columns group missingness scenarios by mechanism (MCAR, MAR, MNAR) and volume (5%, 10%, 20%).

3.2.2 Normalised median absolute deviation

Fig 4a and Fig 4b show the systematic underestimation or overestimation of the agroclimatic indices under different structures and volumes of degradation. Seasonal Total Rainfall was the most biased metric during the wet season, as shown in Fig 4a. The Seasonal Total was increasingly underestimated, from -29.2 mm yr^{-1} at 5% MCAR up to -244.3 mm yr^{-1} under 20% MNAR. CDD showed the least bias (-15 mm yr^{-1} at 20% MCAR), while Season Onset and Cessation were not degraded. In the dry season, the bias profiles were inverted, as shown in Fig 4b. Total Rainfall and CWD exhibited 0 mm yr^{-1} bias across all missingness mechanisms and volumes. However,

CDD was underestimated, with values ranging from -98 mm yr^{-1} at 5% MNAR up to -138 mm yr^{-1} at 20% MAR.

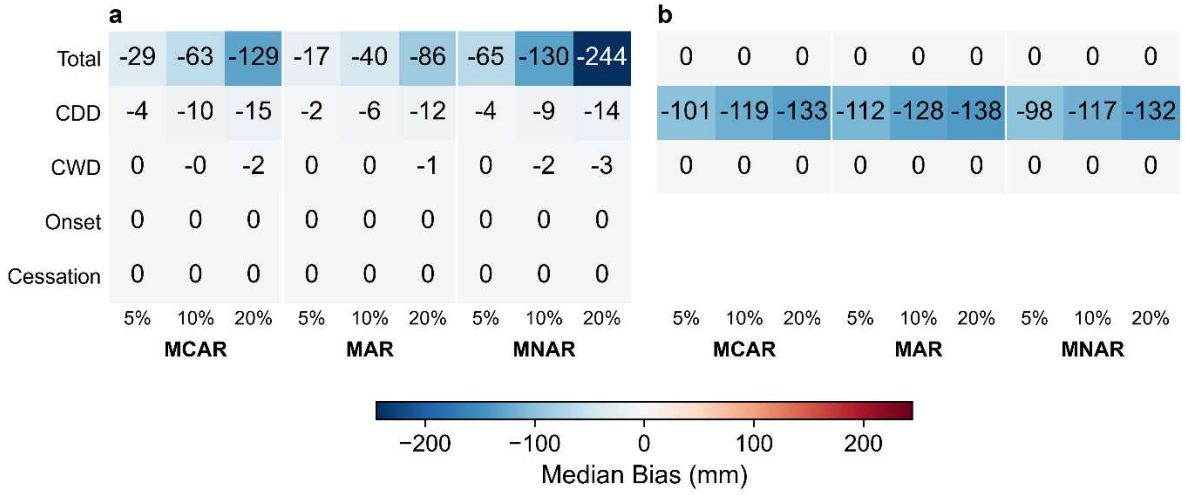


Fig 4. Effects of missingness mechanisms and volume on the Median Bias (MdB) of rainfall season indices in the (a) wet season and (b) in the dry season. Values are medians across all stations. Columns group missingness scenarios by mechanism (MCAR, MAR, MNAR) and volume (5%, 10%, 20%).

3.2.3 Spearman's correlation coefficient

Fig 5 shows whether the temporal distribution of the indices was preserved under data degradation. Despite the heavy bias in Seasonal Total Rainfall, the rank-order correlation remained consistently high across all mechanisms, ranging between 0.83 and 0.99 in the wet season, as shown in Fig 5a. Season Onset and Cessation also preserved their monotonic trend with the undegraded timeseries, with correlation remaining above 0.70, lowering slightly under 20% MCAR and MNAR. CDD lost almost all statistical reliability, as all mechanisms and volumes resulted in $\rho < 0.41$. In the dry season, all degradation scenarios maintained correlations above 0.86 and 0.81 for Seasonal Total Rainfall and CWD, respectively, as shown in Fig 5b. Conversely, the correlation between the undegraded timeseries and the degraded timeseries dropped to between 0.04 and 0.07.

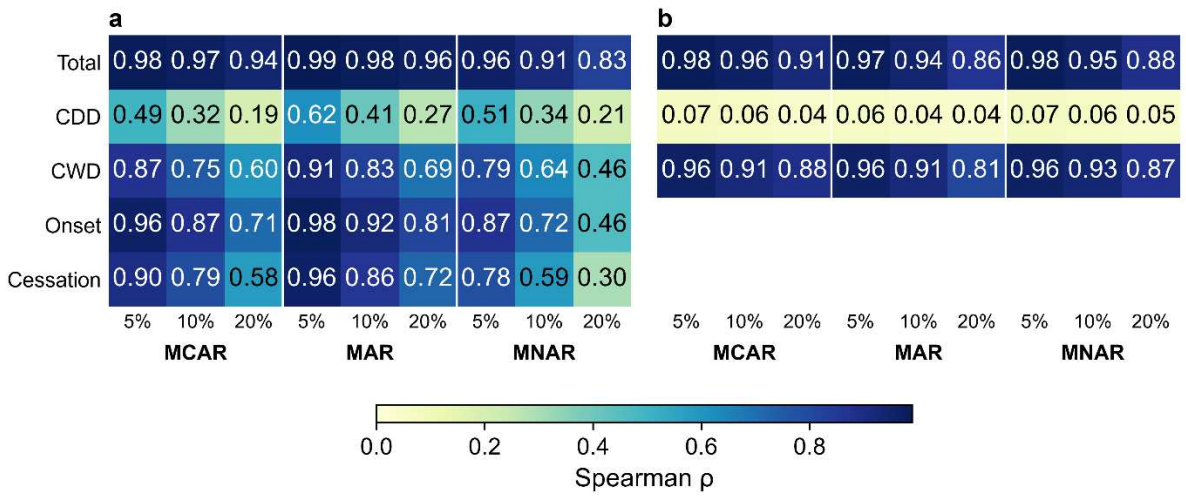


Fig 5. Effects of missingness mechanisms and volume on the Spearman rank correlation (ρ) of rainfall season indices in the (a) wet season and (b) in the dry season. Values are medians across all stations. Columns group missingness scenarios by mechanism (MCAR, MAR, MNAR) and volume (5%, 10%, 20%).

3.3 Spatial variability in data degradation impacts

The severity of bias in the indices was influenced by the station's baseline climate, with the highest impacts observed in higher Seasonal Total Rainfall regions. For Total Rainfall, Chipinge showed the most severe negative Median Bias across all three missingness mechanisms for both the dry and wet seasons, as shown in Fig 6a, Fig 6f, Fig 9a and Fig 9f. Beitbridge, a historically dry station, had the smallest bias for Seasonal Total Rainfall, as shown in Fig 6a and Fig 12a. While wetter regions had the highest volume loss, drier regions experienced significant degradation for continuity metrics like the Spearman's Rank Correlation Coefficient, particularly at high missingness volumes, as shown in Fig 8, Fig 11 and Fig 14. Errors in calculating CDD were most pronounced in dry stations such as Beitbridge, Binga and Matopos, evidenced by the dry season NMAD, which spiked at 90% under the 20% missingness volume, as shown in Fig 7g, Fig 10g, and Fig 13g. Additionally, the temporal integrity of the data collapsed significantly for CDD in drier stations even at 5% missingness volume, while Chipinge retained better rank correlation with the undegraded time series, as shown in Fig 8g, Fig 11g, and Fig 14g.

4 Discussion

The study investigated the impacts of data missingness on five agro-climatic indices: Maximum Consecutive Dry Days (CDD), Maximum Consecutive Wet Days (CWD), Seasonal Total Rainfall, Season Onset, and Season Cessation. These impacts were assessed under three patterns of missingness, Missing at Random (MAR), Missing Completely at Random (MCAR), and Missing Not at Random (MNAR), as well as at three different missingness volumes: 5%, 10%, and 20%. Overall, Missing Not at Random (MNAR) had the greatest impact on the agroclimatic indices, as shown by the -244.3 mm yr^{-1} Median Bias at 20% missingness volume in Fig 4a. This was followed by Missing Completely at Random (MCAR), while Missing at Random (MAR) had the least impact, as it had the lowest Median Bias for Seasonal Total Rainfall (Fig 4a) and the highest correlation with the undegraded timeseries (ρ = 0.99 at 5% missingness and 0.96 at 20% missingness, Fig 5a).

The study showed that bias in the seasonal metrics increases disproportionately as missingness volumes increase. This aligns with the literature, particularly studies on gap-filling climate datasets. [31] showed that the accuracy of climate analyses is mostly constrained by the percentage volume of missing data, and that the information given by the degraded data becomes severely biased once a particular threshold is surpassed. In this study, data missingness resulted in severe deviation from the undegraded time series at 20%, as shown, for example, by Fig 7 under the MCAR mechanism. This follows the same pattern observed by [32] in their study on data imputation, which saw biases inflating as missingness scaled up from 10% to 60%. This study also showed that the effects of data missingness were more pronounced in wetter regions such as Chipinge, while they were less pronounced in drier regions such as Beitbridge. Table 1 shows that Chipinge lies in Zimbabwe's Region I, while Beitbridge lies in Region Vb, which receive >1000 mm yr^{-1} and <600 mm yr^{-1}, respectively.

This implies that the magnitude of errors introduced by missingness is highly sensitive to the local climate baseline. It also means that even when various data imputation methods are employed, the errors are still more attributable to the missingness pattern of the data than to the imputation method itself. Navarro [33] found significantly lower bias in semi-arid regions than in humid ones, because drier regions already have low

annual rainfall totals, so when data is missing, not much information is lost. This is in line with the results of this study. Beitbridge, which lies in one of the driest regions in Zimbabwe, showed the smallest bias in Seasonal Total Rainfall, because the volume of rainfall at risk of being lost is small, resulting in flat or near flat missingness profiles, as presented in Fig 6a to Fig 6h under the MAR mechanism, and Fig 12a to Fig 12h under the MNAR mechanism. Overall, low baseline rainfall acts as a natural ceiling on how large the bias due to degradation can become, regardless of missingness mechanism.

Data missingness had the opposite effect on CDD. Low seasonal rainfall totals protected Total Rainfall from large biases; however, they degrade CDD estimates, because dry seasons consist of many contiguous dry days. As a result, when values went missing, the simulation had no way of determining whether a gap was wet or dry. This introduces a severe underestimation of true dry spell lengths. This is evident in Fig 13b, where the Normalised Median Absolute Deviation reached around 50% for all stations at 20% missingness volume in the wet season under the MNAR mechanism. Beyond bias in single-season analyses, these distortions carry consequences for longer-term climate monitoring. Data missingness can also prevent the effective employment of drought anticipatory action measures. For example, the Zimbabwe Meteorological Services Department, in collaboration with the World Food Programme (WFP), uses the two-month Standardised Precipitation Index (SPI) for tracking drought as part of their "Ready, Set & Go!" system [34].

If data is severely missing, the SPI figures can be misleading, resulting in potential agricultural losses. Missing data can also alter detected trends when such issues persist over multiple years [35], meaning trends could be reversed, or the rates at which they occur could be muted or enhanced. In response to the effects of data missingness, the Participatory Integrated Climate Services for Agriculture (PICSA) programme, developed by the University of Reading, set a limit for data missingness. For a station to qualify for PICSA, it should have at least 30 years of data available, and for those 30 years, the most recent 10 years should not have more than 10% of the data missing [36].

5 Conclusion

Finally, it is important that observation instruments such as manual and automatic weather stations are kept always running, to prevent data missingness. Maintenance checks should also be regular, to ensure that any anomalies are caught in time, before the volume of missing data crosses critical thresholds. An important limitation of the study lies in the use of the Monte Carlo approach. A finite number of simulations was used to simulate data degradation. While infinite repetitions would reveal the true effects of data missingness on agroclimatic analyses, this remains computationally impossible. Therefore, the results should be interpreted as illustrative rather than as a definitive characterisation of the full effects of different data missingness mechanisms on agroclimatic indices.

Appendix A Spatial variability in data degradation impacts

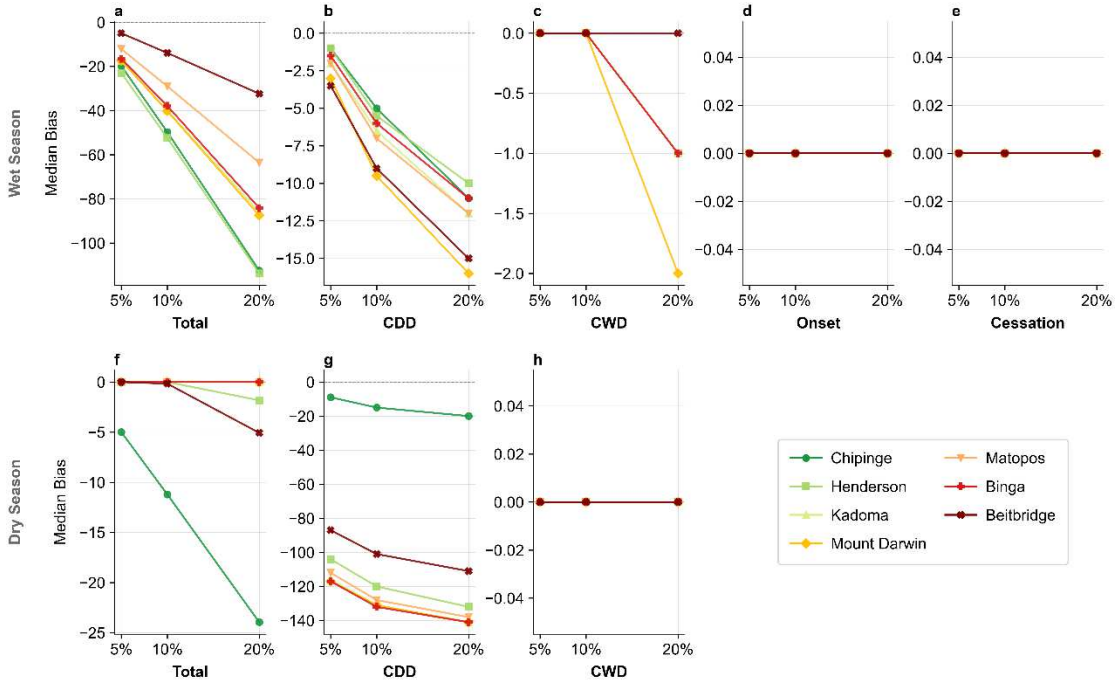


Fig 6. Characterisation of the effects of different volumes of Missingness at Random (MAR) on the rainfall season indices at 7 different stations using Median Bias (MdB)

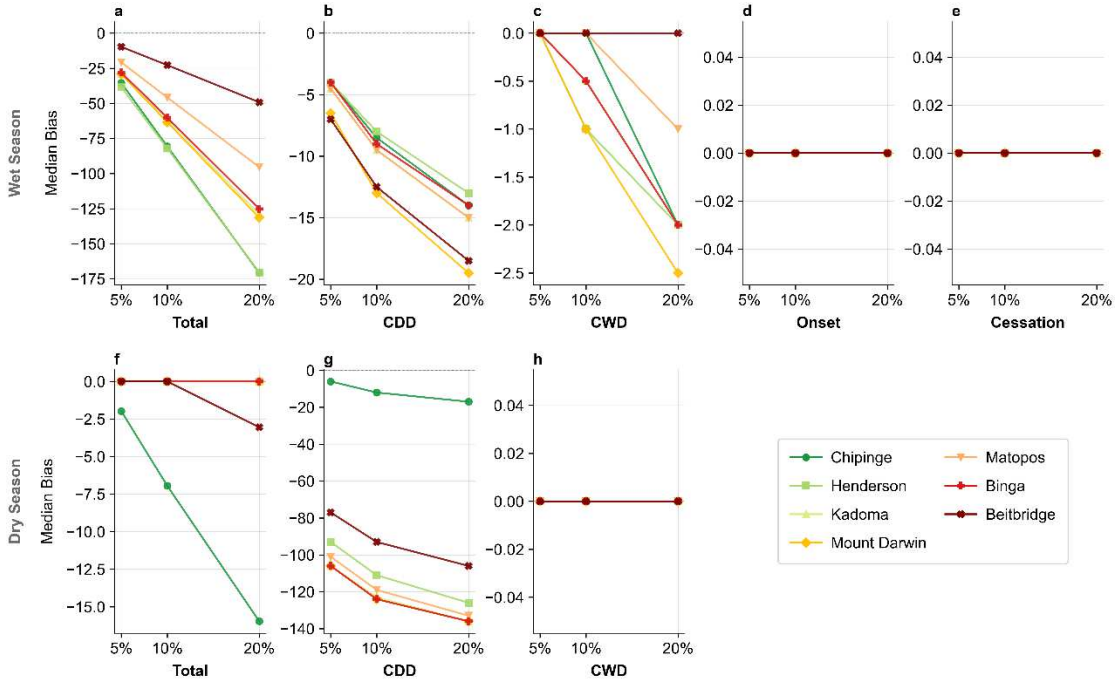


Fig 7. The effects of Missingness Completely at Random (MCAR) at different volumes on the rainfall season indices at 7 different stations quantified using Median Bias (MdB)

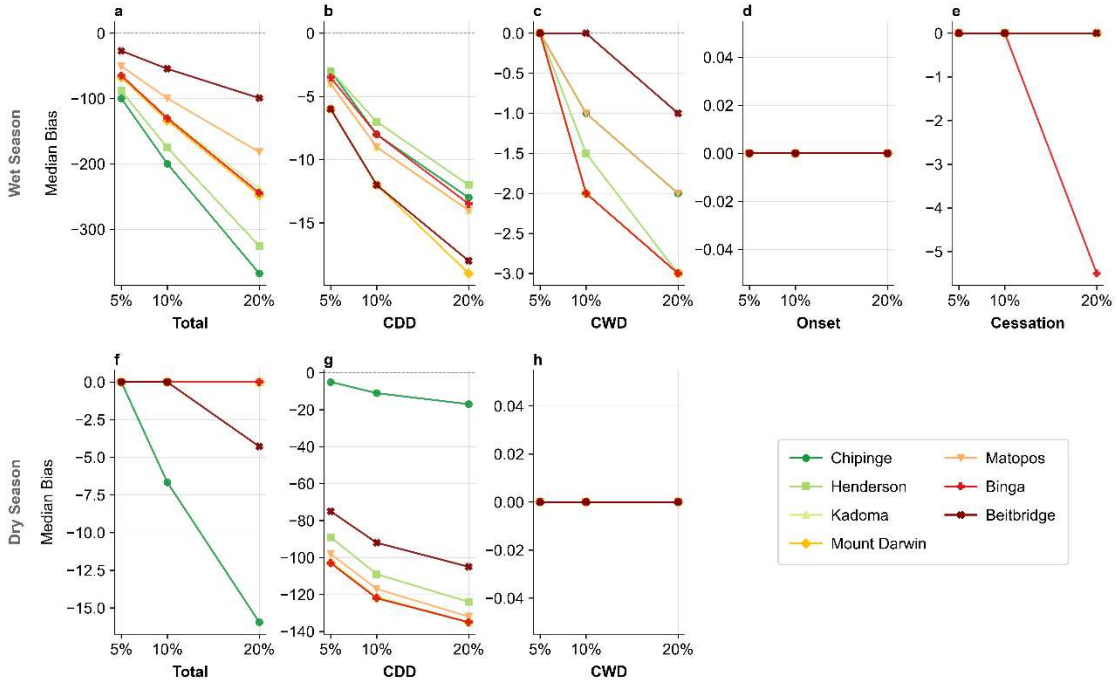


Fig 8. The effects of Missingness Not at Random (MNAR) at different volumes on the rainfall season indices at 7 different stations quantified using Median Bias (MdB)

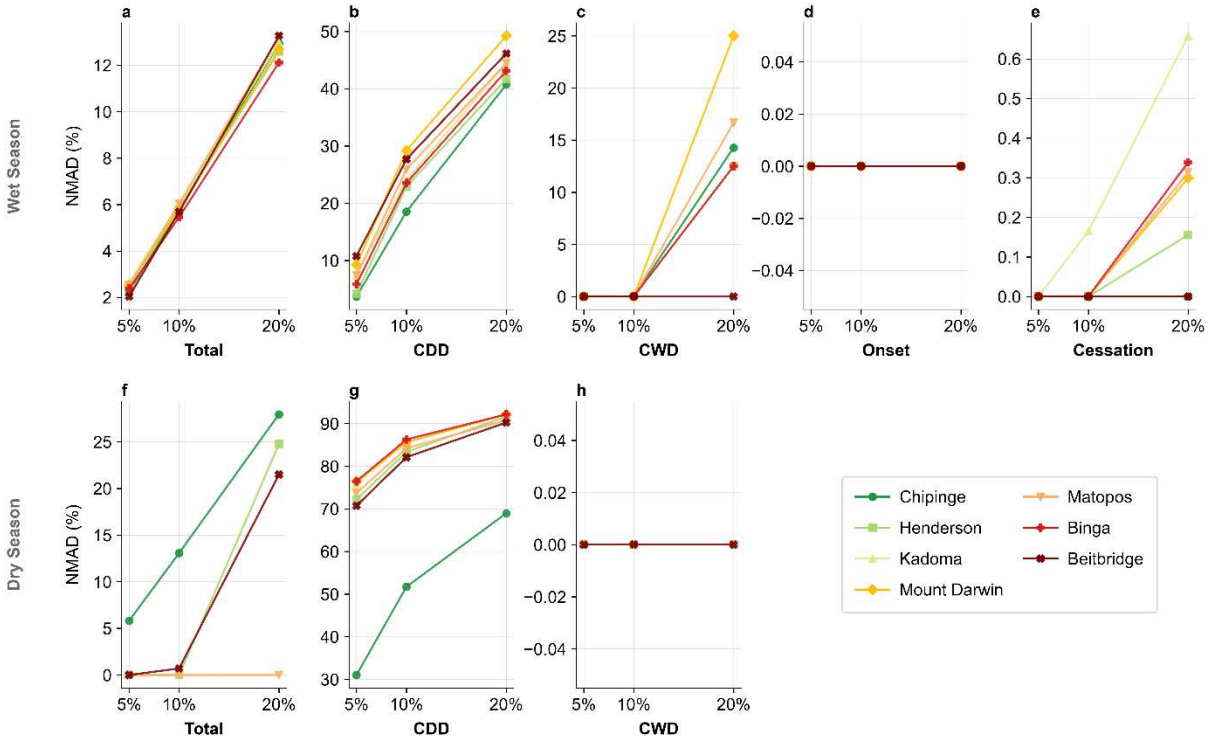


Fig 9. The effects of Missingness Not at Random (MAR) at different volumes on the rainfall season indices at 7 different stations quantified using NMAD

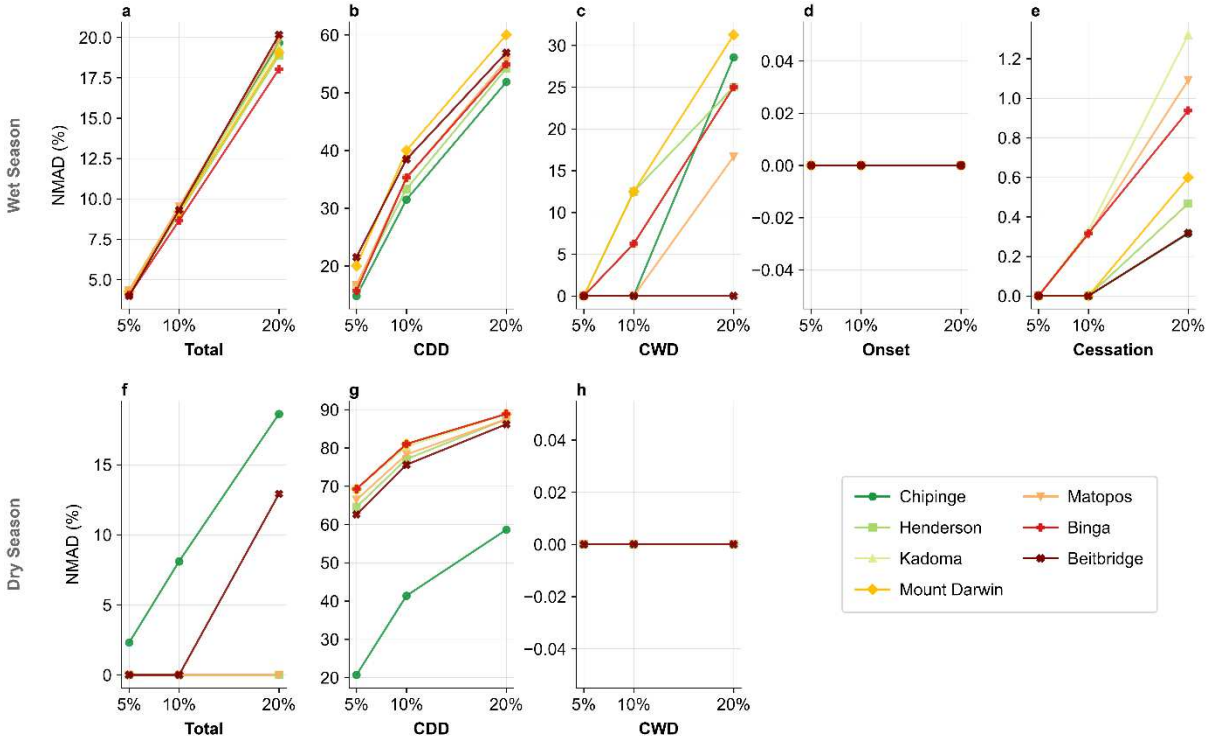


Fig 10. The effects of Missingness Completely at Random (MCAR) at different volumes on the rainfall season indices at 7 different stations quantified using NMAD

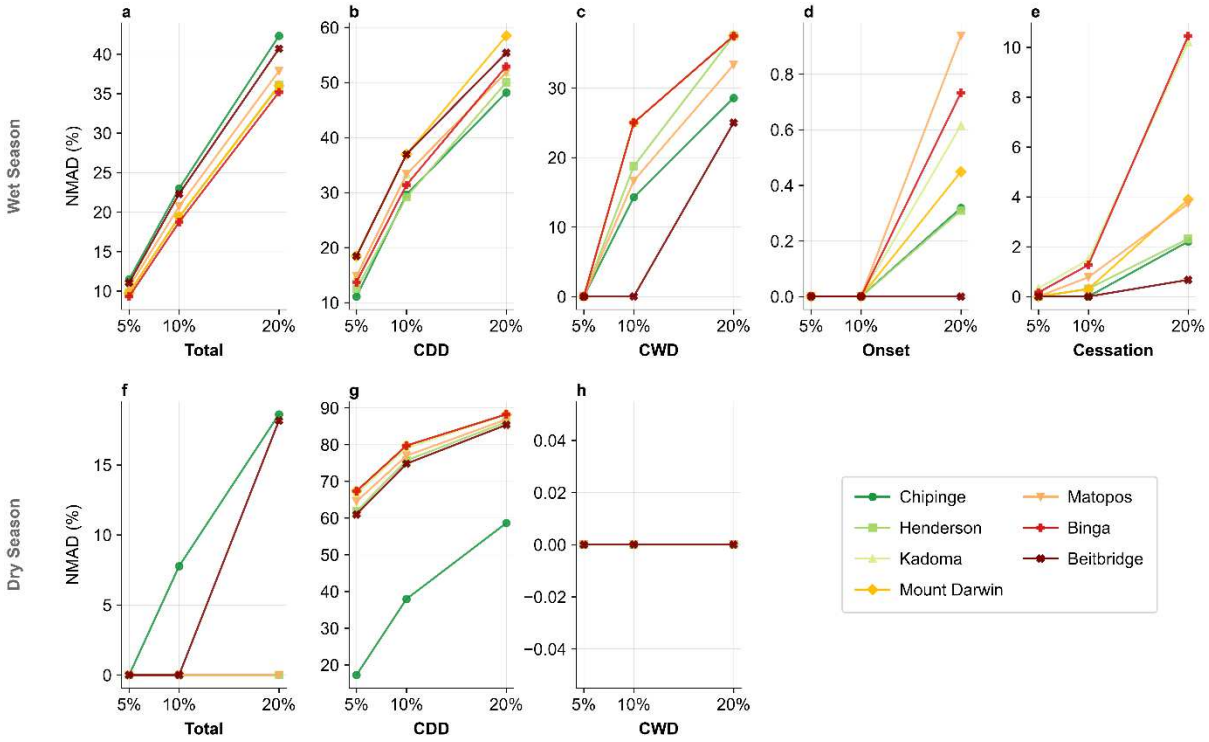


Fig 11. The effects of Missingness Not at Random (MNAR) at different volumes on the rainfall season indices at 7 different stations quantified using NMAD

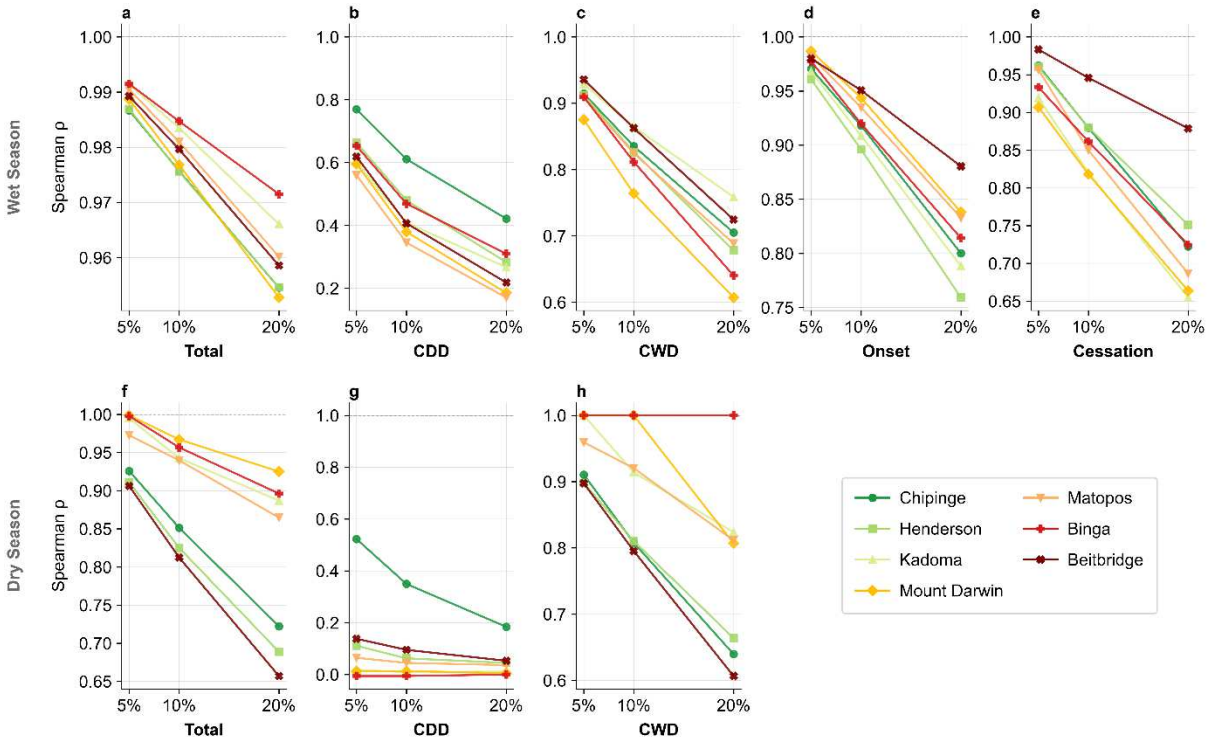


Fig 12. The effects of Missingness at Random (MAR) at different volumes on the rainfall season indices at 7 different stations quantified using Spearman's ρ

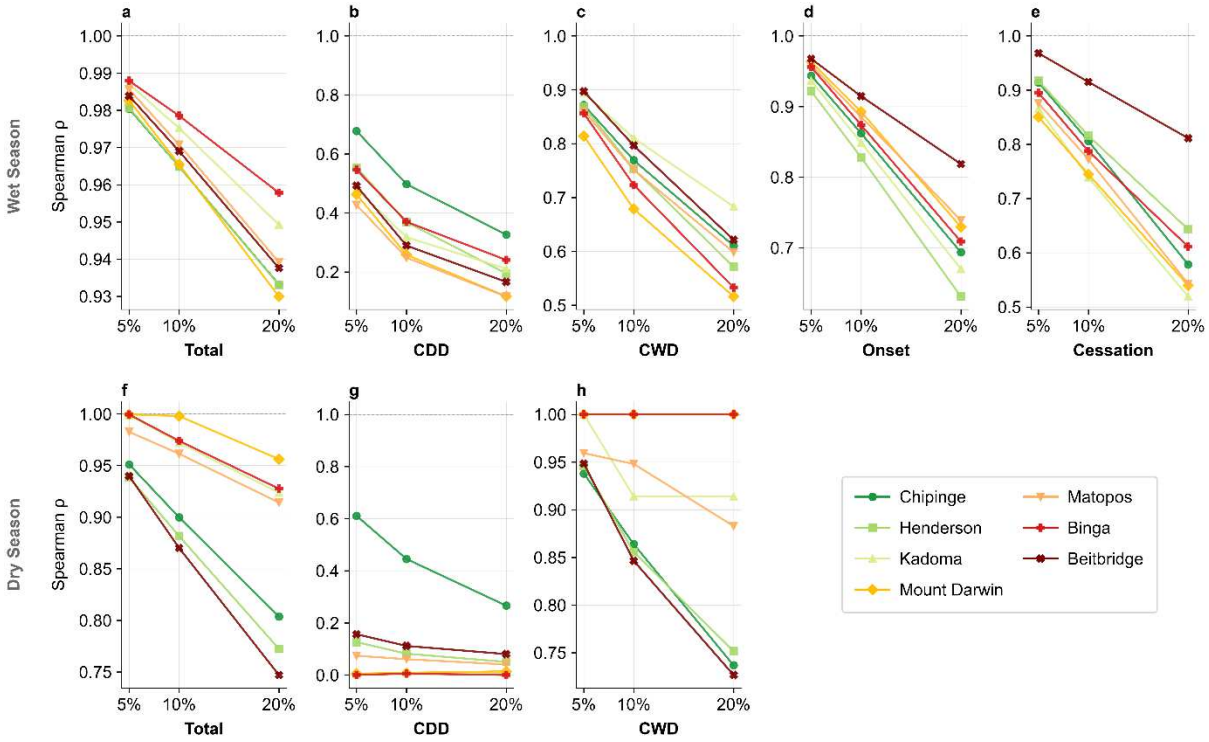


Fig 13. The effects of Missingness Completely at Random (MCAR) at different volumes on the rainfall season indices at 7 different stations quantified using Spearman's ρ

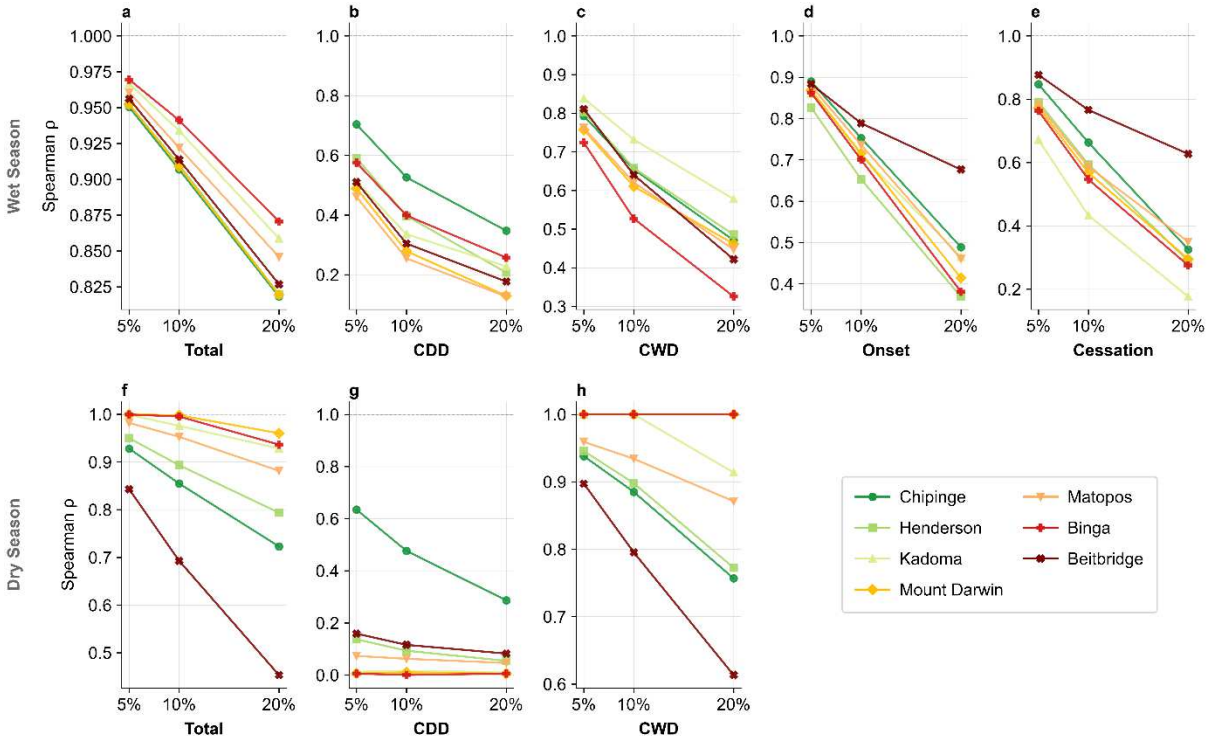


Fig 14. The effects of Missingness Not at Random (MNAR) at different volumes on the rainfall season indices at 7 different stations quantified using Spearman's ρ

References

1. Fentahun T, Gashaw T. Analysis of rainfall and temperature data to determine climate change in Dilla Zuria District, Southern Ethiopia. 2014.

2. ZIMFA. Ministry of Foreign Affairs & International Trade. Zimbabwe in Brief. Agriculture [gov] [Internet]. Government of Zimbabwe; 2024 [cited 2026 Apr 22]. Available from: https://www.zimfa.gov.zw/index.php/about-us/zimbabwe-in-brief/agriculture

3. de By RA, Knipers RA, Sun Y, Ellis MC, Weir MJC, Kraak MJ, Georgiadou Y, Radwan MM, van Westen CJ, Kainz W, Sides EJ. Principles of remote sensing and geographic information systems. 2. ed. Enschede: International Institute for Aerospace Survey and Earth Sciences; 2001. 490 p. (ITC educational textbook series; no. 1/2).

4. Gumindoga W, Rientjes THM, Haile AT, Makurira H, Reggiani P. Performance of bias-correction schemes for CMORPH rainfall estimates in the Zambezi River basin. Hydrol Earth Syst Sci. 2019 Jul 12;23(7):2915–38. doi:10.5194/hess-23-2915-2019

5. Addi M, Gyasi-Agyei Y, Obuobie E, Amekudzi LK. Evaluation of imputation techniques for infilling missing daily rainfall records on river basins in Ghana. Hydrological Sciences Journal. 2022 Mar 12;67(4):613–27. doi:10.1080/02626667.2022.2030868

6. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association. 1988 Dec;83(404):1198–202. doi:10.1080/01621459.1988.10478722

7. Munsaka E, Mudavanhu C, Sakala L, Manjeru P, Matsvange D. When Disaster Risk Management Systems Fail: The Case of Cyclone Idai in Chimanimani District, Zimbabwe. Int J Disaster Risk Sci. 2021 Oct;12(5):689–99. doi:10.1007/s13753-021-00370-6

8. Duarte LV, Formiga KTM, Costa VAF. Comparison of Methods for Filling Daily and Monthly Rainfall Missing Data: Statistical Models or Imputation of Satellite Retrievals? Water. 2022 Oct 6;14(19):3144. doi:10.3390/w14193144

9. Nkuna TR, Odiyo JO. Filling of missing rainfall data in Luvuvhu River Catchment using artificial neural networks. Physics and Chemistry of the Earth, Parts A/B/C. 2011;36(14–15):830–5. doi:10.1016/j.pce.2011.07.041

10. Libanda B, Nkolola NB. An ensemble empirical mode decomposition of consecutive dry days in the Zambezi Riparian Region: Implications for water management. Physics and Chemistry of the Earth, Parts A/B/C. 2022 Jun;126:103147. doi:10.1016/j.pce.2022.103147

11. Dunning CM, Black ECL, Allan RP. The onset and cessation of seasonal rainfall over Africa. Journal of Geophysical Research: Atmospheres. 2016;121(19):11,405-11,424. doi:10.1002/2016JD025428

12. Manatsa D, Mushore D Terence, Gwitira I, Chemura A, Shekede M Davis, Mugandani R, Sakala C Lucy, Ali H Leo, Masukwedza I Gibbon, Mupuro M Jonh, Muzira M Nyaradzo. Revision Of Zimbabwe's Agro-Ecological Zones [Revision Report] [Internet]. Zimbabwe: Zimbabwe National Geospatial and Space Agency (ZINGSA); 2020 Sep [cited 2026 May 23]. p. 113. Revision Report no.: 1. Available from: https://www.researchgate.net/profile/Desmond-Manatsa/publication/347966377_Report_on_Revised_Agroecological_Zones_of_Zimbabwe_in_press/links/5feab23c299bf140885658b2/Report-on-Revised-Agroecological-Zones-of-Zimbabwe-in-press.pdf?origin=publication_detail&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uRG93bmxvYWQiLCJwcmV2aW91c1BhZ2UiOiJwdWJsaWNhdGlvbiJ9fQ

13. Bailey M, Heinrich D, Kruczkiewicz A. Climate Profiles of Countries in Southern Africa: Zimbabwe [Profile] [Internet]. The Hague: Red Cross Red Crescent Climate Centre; 2021 [cited 2026 Jul 2]. Available from: https://www.climatecentre.org/wp-content/uploads/Climate-Profiles-of-Countries-in-Southern-Africa-Zimbabwe.pdf

14. Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci Data. 2015 Dec 8;2(1):150066. doi:10.1038/sdata.2015.66

15. Dinku T, Funk C, Peterson P, Maidment R, Tadesse T, Gadain H, Ceccato P. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. Quart J Royal Meteoro Soc. 2018 Nov;144(S1):292–312. doi:10.1002/qj.3244

16. Python.org. Python Release Python 3.13.11 [Python] [Windows 11]. Wilmington, Delaware, United States: Python Software Foundation; 2025 [cited 2026 Jul 13]. Available from: https://www.python.org/downloads/release/python-31311/

17. Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, Van Kerkwijk MH, Brett M, Haldane A, Del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nature. 2020 Sep 17;585(7825):357–62. doi:10.1038/s41586-020-2649-2

18. The pandas development team. pandas-dev/pandas: Pandas [Internet]. Zenodo; 2026 [cited 2026 Jul 10]. Available from: https://zenodo.org/doi/10.5281/zenodo.21003741 doi:10.5281/ZENODO.21003741

19. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, Van Der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, Van Mulbregt P, SciPy 1.0 Contributors, Vijaykumar A, Bardelli AP, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN,

Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik DV, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, De Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020 Mar 2;17(3):261–72. doi:10.1038/s41592-019-0686-2

20. Casper da Costa-Luis, Stephen Karl Larroque, Kyle Altendorf, Hadrien Mary, richardsheridan, Mikhail Korobov, Noam Raphael, Ivan Ivanov, Marcel Bargull, Nishant Rodrigues, Shawn, Mikhail Dektiarev, Michał Górny, mjstevens777, Matthew D. Pagel, Martin Zugnoni, CrazyPython, Charles Newey, Antony Lee, pgajdos, Todd, Staffan Malmgren, redbug312, Orivej Desh, Nikolai Nechaev, Mike Boyle, Max Nordlund, MapleCCC, Jack McCracken, Hugo van Kemenade. tqdm: A fast, Extensible Progress Bar for Python and CLI [Internet]. Zenodo; 2026 [cited 2026 Jul 10]. Available from: https://zenodo.org/doi/10.5281/zenodo.21237029 doi:10.5281/ZENODO.21237029

21. Zimbabwe Meteorological Services Department. Meteorological Services Department of Zimbabwe. Your weather and climate services for a resilient future [Organisation] [Internet]. Harare, Zimbabwe; 2026 [cited 2026 Apr 28]. Available from: https://www.weatherzw.org.zw/

22. Sivakumar MVK. Predicting rainy season potential from the onset of rains in Southern Sahelian and Sudanian climatic zones of West Africa. Agricultural and Forest Meteorology. 1988. doi:10.1016/0168-1923(88)90039-1

23. Tadross MA, Hewitson BC, Usman MT. The Interannual Variability of the Onset of the Maize Growing Season over South Africa and Zimbabwe. Journal of Climate. 2005 Aug 15;18(16):3356–72. doi:10.1175/JCLI3423.1

24. Mugalavai EM, Kipkorir EC, Raes D, Rao MS. Analysis of rainfall onset, cessation and length of growing season for western Kenya. Agricultural and Forest Meteorology. 2008 Jun;148(6–7):1123–35. doi:10.1016/j.agrformet.2008.02.013

25. Osatohanmwen P. A New Two-Component Hybrid Model for Highly Right-Skewed Data: Estimation Algorithm and Application to Rainfall Data from South Tyrol, Italy. Mathematics. 2025 Sep 16;13(18):2987. doi:10.3390/math13182987

26. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27–38. doi:10.1093/biomet/80.1.27

27. Kyriakou S, Kosmidis I, Sartori N. Median bias reduction in random-effects meta-analysis and meta-regression. Stat Methods Med Res. 2019 Jun;28(6):1622–36. doi:10.1177/0962280218771717

28. Höhle J, Höhle M. Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS Journal of Photogrammetry and Remote Sensing. 2009 Jul;64(4):398–406. doi:10.1016/j.isprsjprs.2009.02.003

29. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology. 2013 Jul;49(4):764–6. doi:10.1016/j.jesp.2013.03.013

30. King AP, Eckersley RJ. Descriptive Statistics II: Bivariate and Multivariate Statistics. In: Statistics for Biomedical Engineers and Scientists [Internet]. Elsevier; 2019 [cited 2026 Jul 2]. p. 23–56. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780081029398000116 doi:10.1016/B978-0-08-102939-8.00011-6

31. Aieb A, Madani K, Scarpa M, Bonaccorso B, Lefsih K. A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. Heliyon. 2019 Feb;5(2):e01247. doi:10.1016/j.heliyon.2019.e01247

32. Rodrigues DT, Gonçalves WA, Silva CMSE, Spyrides MHC, Lúcio PS. Imputation of precipitation data in northeast Brazil. An Acad Bras Ciênc. 2023;95(2):e20210737. doi:10.1590/0001-3765202320210737

33. Navarro Céspedes JM, Hernández JH, Alcántara Concepción PC, Morales Martínez JL, Carreño Aguilera G, Padilla Benítez F. A comparison of missing values imputation methods applied to precipitation of two semi-arid and humid regions of México. ATM. 2022 Aug 2. doi:10.20937/ATM.53095

34. Vhumisai G, Chaibva N, Gororo C, Chikukwa K Kelvin, Nobre G Guimarães, Ferrari I, Viljonen E. Zimbabwe 2024-25 Rainy Season Analysis and Performance of the Drought Anticipatory Action System [Internet]. Zenodo; 2025 May [cited 2026 Jul 20]. Available from: https://zenodo.org/doi/10.5281/zenodo.15592113 doi:10.5281/ZENODO.15592113

35. Keim BD, Wilson AM, Wake CP, Huntington TG. Are there spurious temperature trends in the United States Climate Division database? Geophysical Research Letters. 2003;30(7). doi:10.1029/2002GL016295

36. Clarkson G, Dorward P, Poskitt S, Stern RD, Nyirongo D, Fara K, Gathenya JM, Staub CG, Trotman A, Nsengiyumva G, Torgbor F, Giraldo D. Stimulating small-scale farmer innovation and adaptation with Participatory Integrated Climate Services for Agriculture (PICSA): Lessons from successful implementation in Africa, Latin America, the Caribbean and South Asia. Climate Services. 2022 Apr;26:100298. doi:10.1016/j.cliser.2022.100298