AgERA5 representation of seasonal mean and extreme temperatures in the Northern Cape, South Africa

AUTHORS:
Jacobus A. Kruger[1]
Sarah J. Roffe[1,2]
Adriaan J. van der Walt[1,3]

AFFILIATIONS:
[1]Department of Geography, University of the Free State, Bloemfontein, South Africa
[2]Agrometeorology Division, Agricultural Research Council – Natural Resources and Engineering, Pretoria, South Africa
[3]Afromontane Research Unit, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Adriaan van der Walt

EMAIL:
vanderwalta@ufs.ac.za

DATES:
Received: 26 Apr. 2023
Revised: 14 Sep. 2023
Accepted: 08 Nov. 2023
Published: 27 Mar. 2024

HOW TO CITE:
Kruger JA, Roffe SJ, van der Walt AJ. AgERA5 representation of seasonal mean and extreme temperatures in the Northern Cape, South Africa. S Afr J Sci. 2024;120(3/4), Art. #16043. https://doi.org/10.17159/sajs.2024/16043

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Jemma Finch

KEYWORDS:
heat waves, cold waves, AgERA5 reanalysis, weather stations, data set evaluation

FUNDING:
Society of South African Geographers; University of the Free State

Over regions with sparse observation networks, including South Africa's Northern Cape Province, gridded data sets represent valuable supplementary data sources enabling spatially detailed climate investigations. Their performance is, however, influenced by regional characteristics, thus a performance assessment should be a prerequisite for any regional application. Through a pairwise comparison with eight point-based temperature records, we evaluated the AgERA5 data sets representation of mean summer (November–March; Tms) and winter (May–September; Tmw) temperatures and respective seasonal heatwave and coldwave characteristics across the Northern Cape for 1980–2020. Correlations ranging from 0.48 to 0.92 for Tms and from 0.38 to 0.94 for Tmw reflect relatively strong, but varying, temporal correspondence between the AgERA5 data and stations. Low biases, averaging –0.08 (0.17) °C and ranging from -0.79 to 2.10 (-0.40 to 1.47) °C for Tms (Tmw) were evident. Biases for the heatwave (coldwave) magnitudes were low, averaging -0.38 (0.19) °C^2, and ranging from -1.55 to 1.47 (-2.05 to 2.91) °C^2. Biases for the heatwave (coldwave) frequency were also low, but typically overestimated, averaging 1.19 (0.73) days, and ranging from -1.33 to 5.60 (-1.61 to 3.39) days. Biases for the heatwave (coldwave) number were low and typically overestimated, averaging 0.27 (0.08) events, and ranging from -0.28 to 1.40 (-0.39 to 0.39) events. Despite some stations depicting consistently poor performance, the study results support further application of the AgERA5 product for spatiotemporal analyses of mean and extreme temperatures across the Northern Cape, provided limitations are adequately acknowledged. Further application of the fine-resolution AgERA5 product will greatly inform impact-based studies exploring mean and extreme temperature influences over the Northern Cape Province.

Significance:

- The AgERA5 product was assessed on its performance in representing average and extreme temperature characteristics over South Africa's Northern Cape Province.
- Good comparability between the AgERA5 product and point-based observations supports further application of the AgERA5 across the Northern Cape.
- The AgERA5 product offers a spatially detailed picture of mean and extreme temperatures across the Northern Cape, which is valuable for regions where weather stations are not available.
- The AgERA5 product is thus important for impact-based studies assessing, for instance, the impact of extreme temperatures on livestock and human health.

Introduction

Southern Africa is expected to experience above global-average warming, which will lead to drastic changes in regional extreme temperature event (ETE) characteristics.[1-3] Historical trends and future projections indicate that, compared to other South African provinces, the Northern Cape Province (Figure 1) has and will likely continue to experience among the largest increases in surface air temperature and hot ETE characteristics (e.g. heatwaves).[1,4] Conversely, historical trends and future projections over southern Africa typically show decreasing trends in the cold ETE characteristics (e.g. coldwaves).[5-7] During ETEs, prolonged exposure to thermal stress can have devastating impacts which can influence agricultural productivity, by reducing crop yields and potentially causing livestock mortalities, and human health, by exacerbating illnesses (e.g. headaches and asthma) and potentially leading to mortality.[7-9] These impacts are of concern, because in developing regions, such as the Northern Cape, associated implications are exacerbated due to a high reliance on weather and climate-sensitive activities (e.g. agriculture), and high levels of poverty and unemployment.[10,11]

Across the Northern Cape Province, interactions between tropical, temperate and subtropical weather systems, the regional topography, and the cold Benguela Current (and the Benguela Upwelling System) are known to drive the occurrences of cold and hot ETEs.[12] Through westerly troughs, cut-off lows and mid-latitude cyclone cold fronts, the mid-latitude westerlies and cold Benguela current (and the Benguela Upwelling System) contribute to the advection of cold air, from the southern Atlantic Ocean, over the Northern Cape and are known to be associated with cold snaps and coldwave events.[12-14] Typically, heatwaves are associated with mid-to-lower tropospheric high-pressure systems, limited cloud coverage, and enhanced incoming longwave radiation.[15,16] Troughs extending from the tropics transport warm air from the farther northern tropical regions and are also known to induce hot ETEs across South Africa and the Northern Cape.[17-19]

Weather station records have been the primary data source for investigations regarding ETEs across South Africa, yet large parts of South Africa, especially mountainous and remote regions, have sparse station network coverage.[20] Thus, station data alone cannot provide detailed spatial pictures required for climate studies. Station data sometimes have data quality issues and are not typically temporally complete due to technical issues and, in some cases, closure of stations.[20] There is thus an increasing need for an alternative, or supplementary data

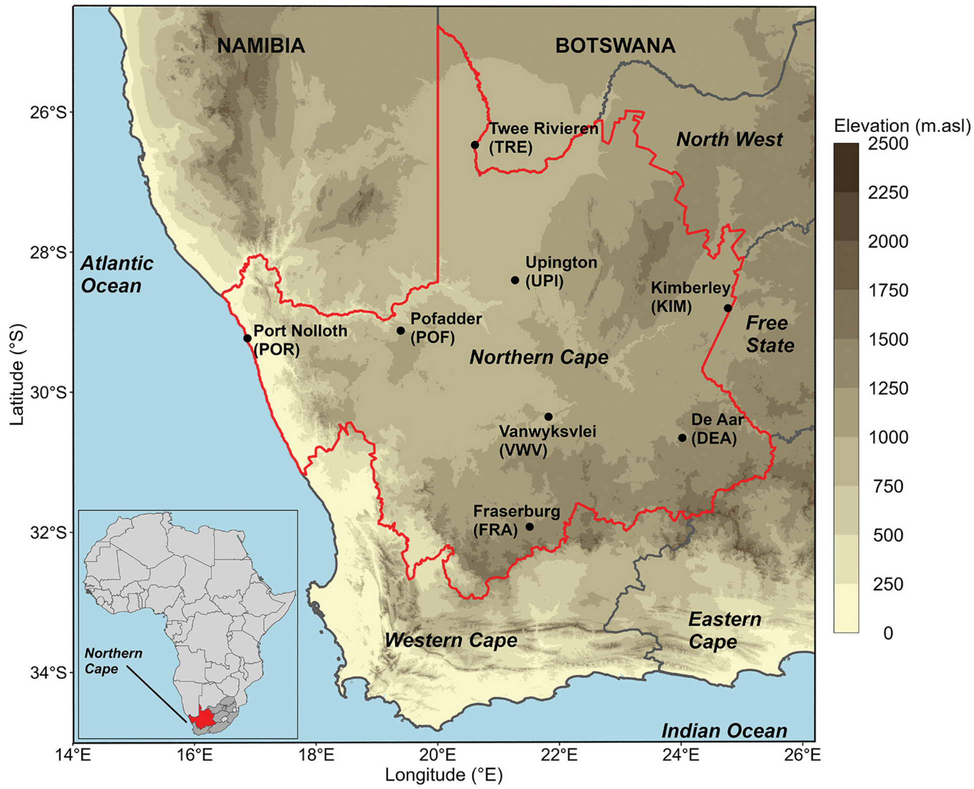

Figure 1: Study site map depicting elevation, neighbouring countries and provinces, bordering oceans, and locations of weather stations used across the Northern Cape Province. Details of the weather stations are presented in Supplementary table 1.

source, and gridded data sets offer such an alternative.[21,22] Various gridded temperature data sets exist which are generated using different methods, such as interpolating station data, analysing satellite imagery, and assimilating observations from stations and/or satellite imagery through simulations.[23-27] There is, however, uncertainty as to whether gridded data accurately represent temperature and ETEs, especially in regions adjacent to oceans, with steep near-surface temperature gradients, elevation gradients and/or precipitation gradients.[21,22] The Northern Cape represents a region with steep elevation gradients, where the high-lying interior plateau and low-lying coastal plain are separated by the mountainous region of the western Great Escarpment (Figure 1). The Northern Cape has a complex rainfall climatology, with a summer rainfall zone towards the east, a year-round rainfall zone from the central parts to the edge of the western Great Escarpment, and a winter rainfall zone across the coastal plain.[28] Therefore, over a region like the Northern Cape, a fine resolution data set should provide a more detailed representation of the region's climate, as studies demonstrate that finer resolution data sets often, but not always, provide improved climate representations compared to coarser products.[20,29]

Among the available gridded temperature products, the ERA5, and specifically products downscaled from it (i.e. AgERA5), offer the highest resolution temperature data sets.[25,26] Although recent research has assessed the AgERA5 data set in representing mean and extreme temperatures across the Northern Cape, a gridded observation-based product was considered. Thus, there is still uncertainty as to whether the AgERA5 area-averaged grid cells accurately compare to point-based temperature records.[20] Therefore, using point-based weather station records from the South African Weather Service (SAWS), we aimed to apply comparative statistics to explore the performance of the AgERA5 data set in representing mean summer (Tms) and winter (Tmw) temperatures and respective seasonal heatwave and coldwave characteristics across the Northern Cape for 1980–2020. Considering the adverse implications associated with ETEs and the importance of reliable gridded data sets for spatial investigations of ETEs, the evaluation of such a data set is relevant as a prerequisite for further studies utilising the AgERA5 product over the Northern Cape.[29]

Data and methods

Data and pre-processing

The AgERA5 daily maximum and minimum temperature (Tx and Tn, respectively) outputs for 1980–2020 were utilised for this study; AgERA5 data are freely available for download from the Copernicus Climate Data Store.[26] The AgERA5 data set is a statistically downscaled and bias-corrected product that is available at a 0.1° resolution.[26] It is derived from the hourly ERA5 reanalysis which is available at a 0.25° resolution and combines numerical model, satellite and observation data using the European Centre for Medium-Range Weather Forecasts' Integrated Forecast System (ECMWF).[25,26] Before calculating the temperature indices using the AgERA5 data, several pre-processing steps were undertaken to prepare the data using the Climate Data Operators software.[30] This preparation included the temporal merging of daily NetCDF files, spatial clipping to the study domain extent (i.e. 25–33° S and 17–25°E), converting units from K to °C, and changing of variable names according to the Climpact user manual.[31]

To evaluate the performance of the AgERA5 data set across the Northern Cape, daily Tx and Tn spanning 1980–2020 were utilised from eight SAWS weather stations (Supplementary table 1), purposively selected following van der Walt and Fitchett[32], to produce an evenly distributed network of stations across the province (Figure 1). For inclusion, stations were required to have >90% data availability for 1980–2020.[33] Before statistical analyses, data quality was examined and cleaning was performed.[32,33] All dates were checked for duplication and gaps, and values were rounded to two decimal places for consistency.[33] Repetition and duplication of temperature values and where Tx $\leq$ Tn were among the errors identified, while outliers were identified using box plots and then verified through comparison with nearby stations.[28,32,33] All errors and outliers identified were deleted and recorded as missing values.[28,34,35] Missing values were filled with data from nearby stations if they were located within a 50 km radius and had a Spearman Correlation Coefficient (CC) of >0.70 between the existing records of both stations.[32,35-37] If there were still missing values for less than five consecutive days, a five-day running average was used to estimate these.[28,32]

Table 1: Tabulated results, per location mapped in Figure 1, of the comparison between the AgERA5 and SAWS data sets for average daily summer temperatures (Tms) and daily winter temperatures (Tmw). CC represents the Spearman Correlation Coefficient; RMSE represents the Root Mean Square Error; MD represents the Modified Index of Agreement. CC values denoted in bold represent statistically significant correlations at the 5% alpha level.

Index	Station	CC	MD	RMSE	Bias
Tms	DEA	**0.92**	0.79	0.34	-0.21
	FRA	**0.86**	0.46	0.87	-0.79
	KIM	**0.85**	0.72	0.51	0.08
	POF	**0.90**	0.70	0.38	-0.26
	POR	**0.48**	0.21	2.19	2.10
	TRE	**0.68**	0.57	0.76	-0.50
	UPI	**0.84**	0.58	0.52	-0.36
	VWV	**0.85**	0.45	0.79	-0.70
Tmw	DEA	**0.84**	0.64	0.52	-0.40
	FRA	**0.84**	0.71	0.41	-0.11
	KIM	**0.70**	0.56	0.70	0.49
	POF	**0.94**	0.72	0.57	-0.29
	POR	**0.38**	0.23	1.57	1.47
	TRE	**0.68**	0.51	0.74	0.52
	UPI	**0.78**	0.68	0.37	-0.16
	VWV	**0.84**	0.64	0.40	-0.27

Heatwave and coldwave indices

To calculate ETE indices we utilised the R Climpact package, developed by the World Meteorological Organisation (WMO) Expert Team on Sector-specific Climate Indices (ET-SCI).[31] Among the available heatwave and coldwave definitions, the Excess Heat Factor (EHF) and excess cold factor (ECF) were applied herein for the heatwave and coldwave calculations for the extended summer (November–March) and winter (May–September) seasons for 1980–2020, respectively.[38,39] The EHF/ECF incorporates two components based on average daily temperatures, representing acclimatisation and significance, which are combined into one factor: the EHF/ECF.[39] Heatwaves (coldwaves) are defined when the EHF (ECF) value is positive (negative) for at least three consecutive days.[31,39] More information about the EHF, ECF and respective heatwave and coldwave calculations are detailed in Herold and McComb[31] and Nairn and Fawcett.[39]

Duration, magnitude and frequency are heatwave and coldwave characteristics frequently used to describe such events.[4,6,38,40] Thus, similar heatwave and coldwave indices were computed as seasonally averaged outputs for the AgERA5 and SAWS data sets. These include: (1) heatwave (coldwave) frequency (HWF [CWF]) which represents the total number of days contributing to annual summer (winter) heatwave (coldwave) events, (2) heatwave (coldwave) magnitude (HWM [CWM]) represents the average temperature of summer (winter) heatwaves (coldwaves) per year, measured as $°C^2$ due to the EHF (ECF) definition and (3) heatwave (coldwave) number (HWN [CWN]) which represents the total number of summer (winter) heatwave (coldwave) events per year.[38,39]

Evaluation analysis

To evaluate the AgERA5 in representing Tms and Tmw, and respective seasonal ETE characteristics, we first explored the data sets' performance in characterising daily average Tms and Tmw, using the summer and winter periods for the heatwave and coldwave calculations. Tms and Tmw are fundamental as they can dictate underlying factors influencing the performance of seasonal ETE characteristics.[20] Thereafter, performance was explored for the respective ETE indices. In all instances, four pairwise statistical metrics were calculated to compare the AgERA5 and SAWS data sets; single grid cell values corresponding to the station coordinates were extracted to compute the evaluation metrics (Supplementary table 1). CC was used to measure the level of temporal consistency between the AgERA5 and SAWS data sets, where output values range between zero and one, and one is optimal.[28] Information on the standard metric errors, measuring deviation between the data sets, were calculated as Root Mean Square Error (RMSE) values, where outputs range between zero and one, with zero as the strongest score.[22] The bias metric was used to determine the tendency of the AgERA5 data set to overestimate or underestimate values compared to the SAWS data set, where positive (negative) bias values indicate an overestimation (underestimation) by the AgERA5 data set, while a value of zero is optimal.[41] The Modified Index of Agreement (MD) was used to identify both additive and proportional disparities in the SAWS and AgERA5 mean and variance ranges.[42,43] This provides a valuable skill score for the data; the resulting value ranges between zero and one, with one being the desired value.[42,43]

Results

AgERA5 representation of mean summer and winter temperatures

Box plots illustrating the distribution of Tms and Tmw across the weather stations and AgERA5 grid cells in the Northern Cape for 1980–2020 indicate that there is strong agreement between the data sets, with mean values deviating by <1 °C for most locations (excluding POR; Figure 2a–b). Excluding POR, the overlapping boxplot boxes for the stations and AgERA5 cells further indicate a high degree of agreement and strong correspondence between the two data sets (Figure 2a–b). For Tms, the box plots reveal that KIM and POR exhibit higher temperature values (i.e. warm bias) in the AgERA5 data set, whereas the remaining locations exhibit lower temperatures (i.e. cool bias; Figure 2a). This pattern is indicative of a predominant negative (cool) bias in the AgERA5 Tms proxies, which is apparent in the Supplementary figure 1 time series plots and the bias values (Table 1). Similarly, for Tmw, the box plots

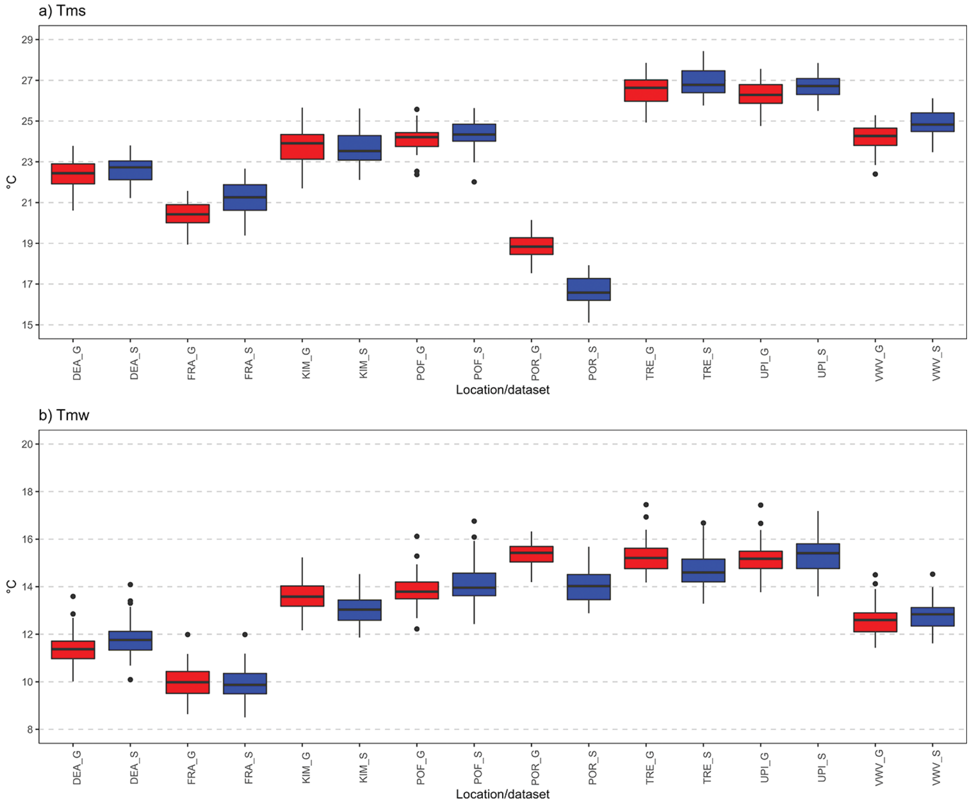

Figure 2: Box plots depicting the distribution of (a) average daily summer (Tms) and (b) winter (Tmw) temperatures for the AgERA5 (red) reanalysis and SAWS (blue) weather station data sets from 1980 to 2020 for locations mapped in Figure 1.

reveal that the AgERA5 cells exhibit higher temperatures (i.e. warm bias) at KIM, POR and TRE, while lower temperatures (i.e. cool bias) are observed at the remaining locations (Figure 2b). This indicates a predominant negative (i.e. cool) bias in the AgERA5 Tmw data, which is also evident in the time series plots and tabulated bias values (Table 1; Supplementary figure 1). Compared to SAWS data, most locations depict lower AgERA5 box plot tail ends, which may translate to an underestimation of AgERA5 extremetemperatures (Figure 2). Among the stations, POR consistently depicts weaker correspondence as mean values deviate by >1 °C and the spread of the data does not overlap consistently for both Tms and Tmw (Figure 2).

Strong temporal consistency between the AgERA5 and station data sets is supported by the statistically significant CC values, demonstrating consistent interannual variability patterns, and the time series plots, which demonstrate consistent temporal tracking between the two data sets (Supplementary figures 1 and 2; Table 1). The CC values are moderate to strong in magnitude, further indicating strong consistency between the data sets (Table 1). CC values range between 0.48 (POR) and 0.92 (DEA) for Tms and from 0.38 (POR) to 0.94 (POF) for Tmw, with stronger temporal correspondence for Tms as reflected by higher correlation values for most stations (Table 1). Located along the western coast and farthest north, POR and TRE exhibit the lowest degree of temporal agreement for Tms (Tmw), with CC values of 0.48 (0.68) and 0.38 (0.68), respectively (Table 1). The MD values further indicate strong agreement between the data sets, with Tms values ranging between 0.21 (POR) and 0.79 (DEA) and Tmw values from 0.23 (POR) to 0.72 (POF; Table 1). The low, but varying, deviation in seasonal temperatures between SAWS and AgERA5 for Tms is characterised by RMSE values ranging from 0.34 °C (DEA) to 2.19 °C (POR), while for Tmw RMSE values range from 0.37 °C (UPI) to 1.57 °C (POR; Table 1). The AgERA5 Tms values are overestimated by 0.08 °C and 2.10 °C at KIM and POR, respectively, and Tmw values are overestimated by 0.49 °C, 1.47 °C and 0.51 °C at KIM, POR and TRE, respectively (Table 1). In general, for most stations, the AgERA5 data set underestimates Tms and Tmw, as evidenced by the mostly negative bias values and the time series plots per station (Table 1; Supplementary figures 1 and 2). Specifically, bias values range from -2.10 (POR) to 0.79 °C (FRA) for Tms and from -1.47 (POR) to 0.40 °C (DEA) for Tmw (Table 1). Despite these biases, their magnitudes are relatively small, which is consistent with the low RMSE values evident across the stations (excluding POR; Table 1).

AgERA5 representation of mean summer season heatwave characteristics

Box plots depicting the distribution of heatwave characteristics and time series plots illustrating temporal patterns across the stations and corresponding AgERA5 grid cells in the Northern Cape for 1980–2020 depict strong consistency between the data sets, as is evident from the overlapping boxes of the box plots, similar temporal variability patterns and consistent tracking for each time series (Figures 3–5; Supplementary figure 3). Specifically, the box plots indicate that most locations deviate, based on mean values, by <3 days for HWF, <1 °C^2 for HWM and <1 event for HWN (Supplementary figure 3). However, the box plots of heatwave characteristics at different locations show different degrees of overlap, less so for POR, suggesting that the data distributions for these variables differ across various locations and characteristics (Supplementary figure 3).

Moderate to strong statistically significant CC values (>0.5) for all heatwave characteristics supports that there is strong agreement between the AgERA5 and SAWS data sets (Table 2). For HWF, CC values range from 0.60 (TRE) to 0.87 (KIM), while for HWM, CC values range from 0.53 (POR) to 0.93 (VWV), and HWN CC ranges are from 0.53 (POR) to 0.85 (KIM; Table 2). CC values for POR, UPI and TRE are generally weaker for all heatwave characteristics (Table 1). This observation is consistent with relatively little overlap observed between the corresponding boxplot boxes and weaker MD values for POR, UPI and TRE (Supplementary figure 3). The degree of agreement between the SAWS and AgERA5 data sets, as measured by MD values, is moderate to strong and is generally higher in magnitude compared to Tms (Tables 1–2). MD values for HWF range between 0.47 (POR) and 0.80 (KIM), while for HWM the range is from 0.51 (KIM) to 0.77 (FRA), and for HWN the range is between 0.44 (POR)

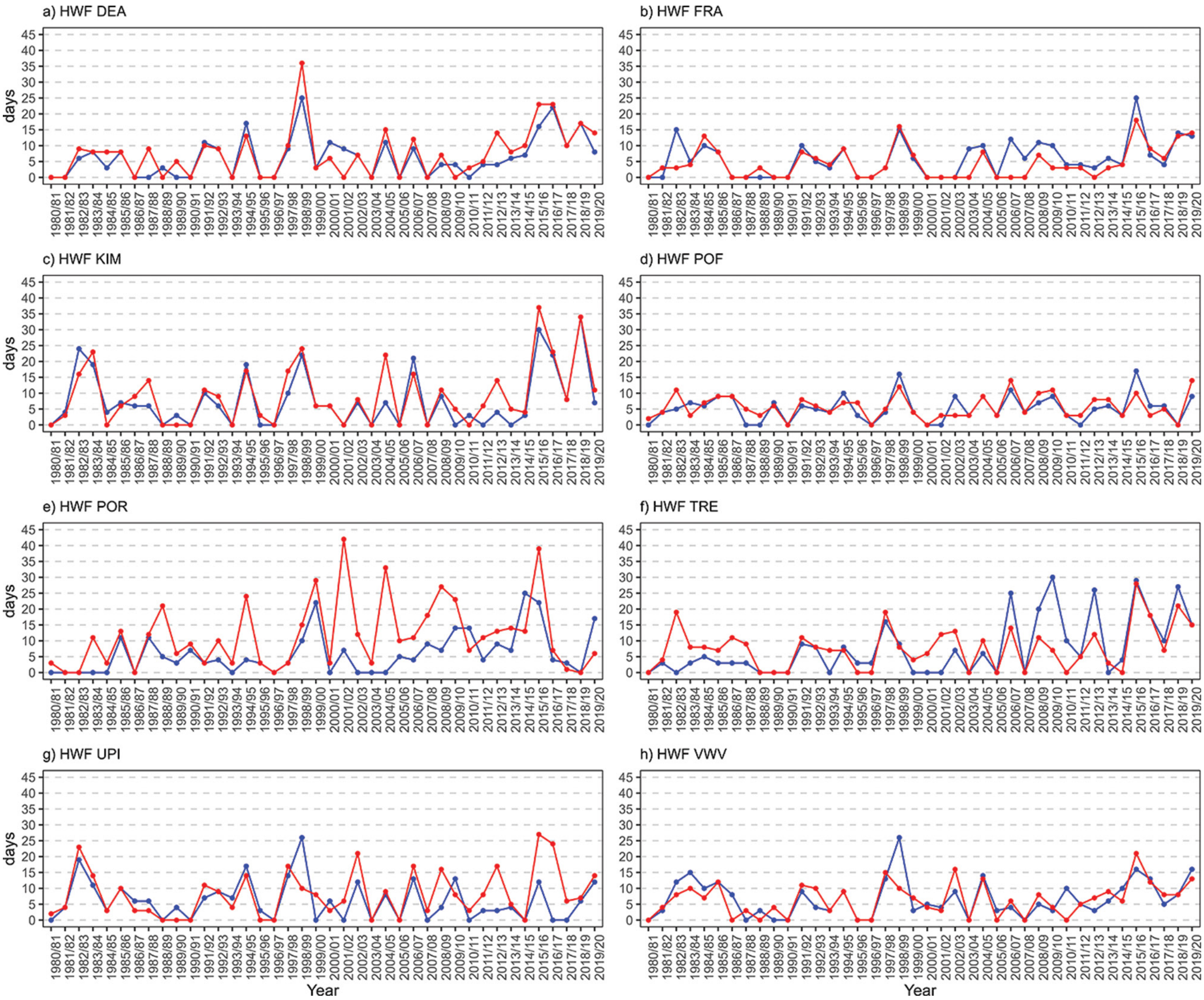

Figure 3: Time series plots depicting the total number of days contributing to annual summer heatwave events (HWF) for the AgERA5 (red) and SAWS (blue) data sets from 1980 to 2020 for locations mapped in Figure 1.

and 0.80 (KIM; Table 2). KIM shows the highest MD values for HWF (0.80) and HWN (0.80), but the lowest value for HWM (0.51; Table 2). The consistency observed in the moderate to strong MD values reflects a high degree of temporal consistency between the data sets (Table 2). Relatively low deviation across the data sets is also evident from the RMSE values, which are relatively low for all heatwave aspects (Table 2); despite this, these RMSE values are higher than that for Tms (Tables 1–2). The RMSE values provide support for the consistency and general deviation observed in the box plots (Supplementary figure 3; Table 2). Specifically, the HWF values range from 2.80 (POF) to 10.92 (POR) days, while HWM ranges from 1.34 (TRE) to 6.45 (POR) °C^2, and HWN ranges from 0.71 (POF) to 2.40 (POR) events (Supplementary figure 3; Table 2). A larger degree of deviation exists between the stations and corresponding AgERA5 for HWF, and a smaller degree of deviation is evident for HWN (Table 2). Overall, the AgERA5 data set typically overestimates HWF and HWN, and consistent with the Tms biases, it typically underestimates HWM (Tables 1–2). For HWF, biases range between -1.33 (POR) and 5.60 days (FRA), while for HWM biases range from -1.55 (POR) to 1.47 °C^2 (KIM), and for HWN the range is between -0.28 (POR) and 1.40 events (FRA; Table 2). This pattern in the biases is also evident in the time series plots and box plots (Figures 3–5; Supplementary figure 3).

AgERA5 representation of mean winter coldwave characteristics

Box plots and time series plots for coldwave characteristics of the weather stations and AgERA5 grid cells in the Northern Cape from 1980 to 2020, demonstrate a high degree of agreement between the data sets, albeit to a lesser extent than for Tmw (Supplementary figure 4; Figures 2a, 3–5). Strong agreement for the coldwave aspects is evidenced by typically overlapping boxes in the box plots and comparable temporal variability in the time series (Supplementary figure 4; Figures 6–8). However, for some locations (e.g. POR), the box plots show less overlap and less temporal agreement in the time series plots (Supplementary figure 4; Figures 6–8). Specifically, the box plots indicate that most locations deviate, on average, by <2 days for CWF, <2 °C^2 for CWM and <0.5 events for CWN (Supplementary figure 4).

The mostly moderate to strong CC values (>0.5) for coldwave characteristics provide additional support for the conclusions drawn from the box plots and time series plots (Supplementary figure 4; Figures 6–8; Table 2). Lower CC values for CWM show that CWM was characterised by the weakest performance and CWM is the only index that has CC values that are statistically insignificant (Table 2). Specifically, for CWF, CC values range from 0.45 (POR) to 0.92 (DEA), for CWM, CC values range from 0.17 (POR) to 0.80 (UPI), and for CWN, the range is between 0.47 (POR) and 0.89 (DEA; Table 2). CC values generally suggest that the AgERA5 and SAWS data sets agree the most for CWF compared to CWM and CWN (Table 2). Additionally, the MD values demonstrate a similar pattern where the CWM aspect depicts the weakest performance (Table 2). For both the CC and MD values, weaker agreement for CWF and CWN is evident at KIM, POR, and TRE, whereas, for CWM, KIM has higher MD and CC values, and FRA has lower MD values compared to KIM for CWF and CWN (Table 2). More specifically, the MD values for CWF range from 0.40 (POR) to 0.80 (DEA), for CWM, values range from

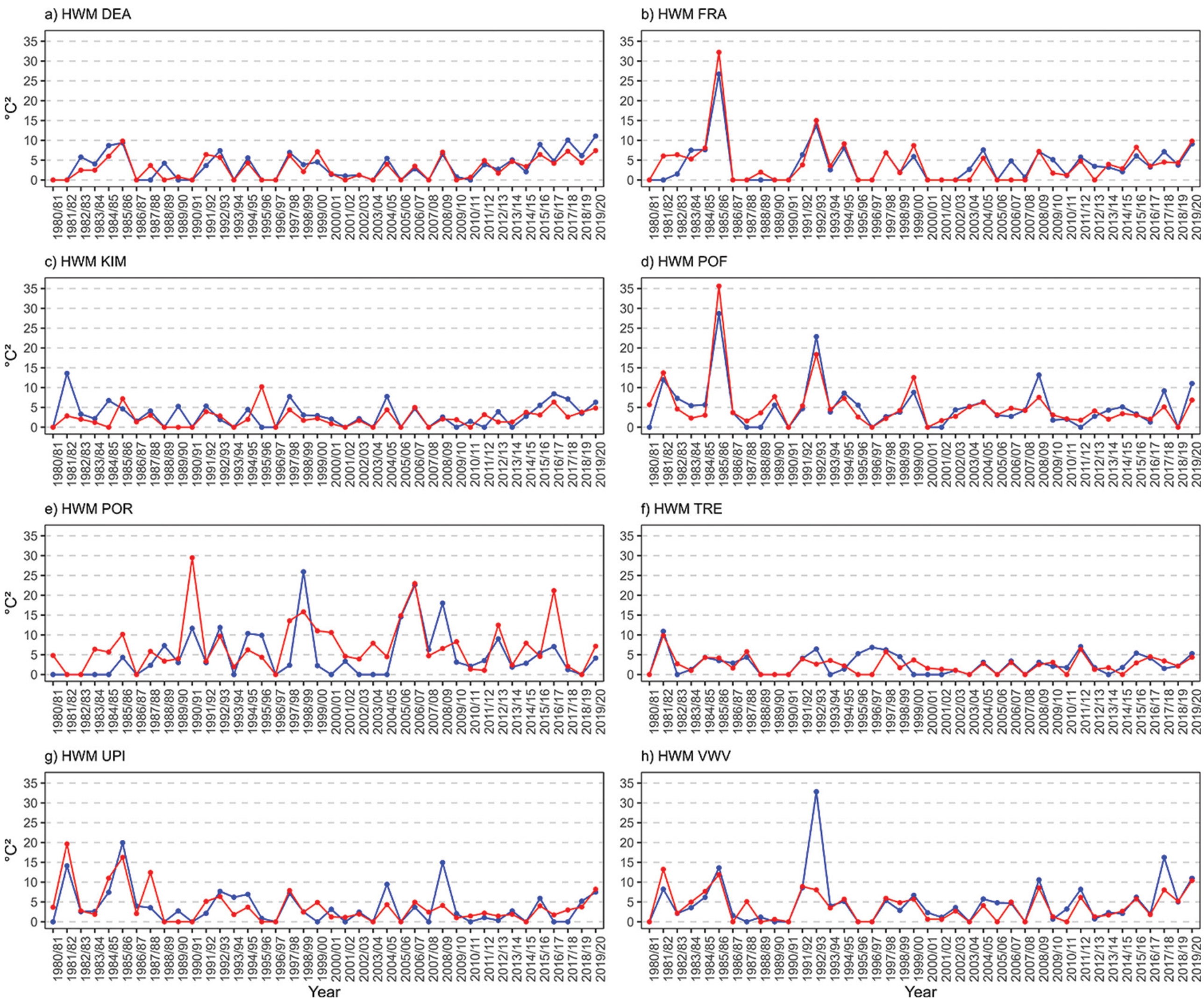

Figure 4: Time series plots depicting the average temperature of annual summer heatwave events (HWM) for the AgERA5 (red) and SAWS (blue) data sets from 1980 to 2020 for locations mapped in Figure 1.

0.35 (POR) to 0.68 (POR), and for CWN, values range from 0.48 (POR) to 0.84 (DEA; Table 2). Deviation between coldwave characteristics, as inferred from the AgERA5 and SAWS data sets' RMSE values, generally tends to be highest for CWF and lowest for CWN (Table 2). Regarding CWF, the RMSE values range from 3.25 (DEA) to 8.09 (POR) days, whereas for CWM, the range is from 2.32 (VWV) to 5.26 (KIM) °C², and for CWN, values range from 0.73 (DEA) to 1.52 (POR) events (Table 2). Box plots in Supplementary figure 4, patterns observed in the time series plots (Figures 6–8), and the positive bias values in Table 2 suggest that the AgERA5 data set overestimates coldwave characteristics to a greater extent than Tmw (Tables 1–2); positive biases for CWM are consistent with that for Tmw (Tables 1–2). Specifically, the biases for CWF range from -1.61 (FRA) to 3.39 days (POR), for CWM the range is from -2.05 (FRA) to 2.91 (KIM) °C², and for CWN the range is between -0.39 (FRA) and 0.71 (POR) events (Table 2). Bias values for POR and TRE are highest and lowest, respectively, for CWN and CWF, while the highest and lowest values for CWM are shown by POR and FRA, respectively (Table 2). This pattern is also apparent in the values calculated for CC and MD (Table 2).

Discussion and conclusions

Through a comparative analysis between weather stations and corresponding AgERA5 grid cells, we evaluated the accuracy of the AgERA5 representation of seasonal average and extreme temperature characteristics over the Northern Cape for 1980–2020. This study builds on research by Roffe and van der Walt[20] by evaluating the AgERA5 data set's ability to represent Tms and Tmw, and the respective seasonal heatwave and coldwave characteristics, against point-based SAWS station data, as opposed to such evaluations using a gridded observation data set. This approach is advantageous as fewer biases are introduced into the evaluation as the point-based weather station data used herein has not been interpolated.[20,44] We do, however, acknowledge that the station-based temperature data are not completely without biases due to, for instance, missing data and the methods used for estimation thereof, changes in station location and monitoring instrumentation, and land use/cover.[44] Despite the weather station data limitations, their use for evaluation provides a robust understanding of the AgERA5 performance over the Northern Cape. Thus, the insights gained from this comparative analysis highlight limitations and advantages of using the AgERA5 product over the Northern Cape; these are important to consider for future research applying this data set to the Northern Cape.

The results presented herein reveal that the AgERA5 data set performs quite well over the Northern Cape, and is thus an invaluable data set to apply for future temperature-related research over the region. Over this data sparse region, the AgERA5 data set can be used, for instance, to analyse ETE impacts for areas with no existing weather station infrastructure.[45] This would be particularly valuable to understand the temperature patterns over Kakamas (a town without a weather station) for instance, where, in January 2023, anomalously hot temperatures caused seven fatalities.[46] This is not the only example in which ETEs have caused fatalities in South African regions where there are no weather

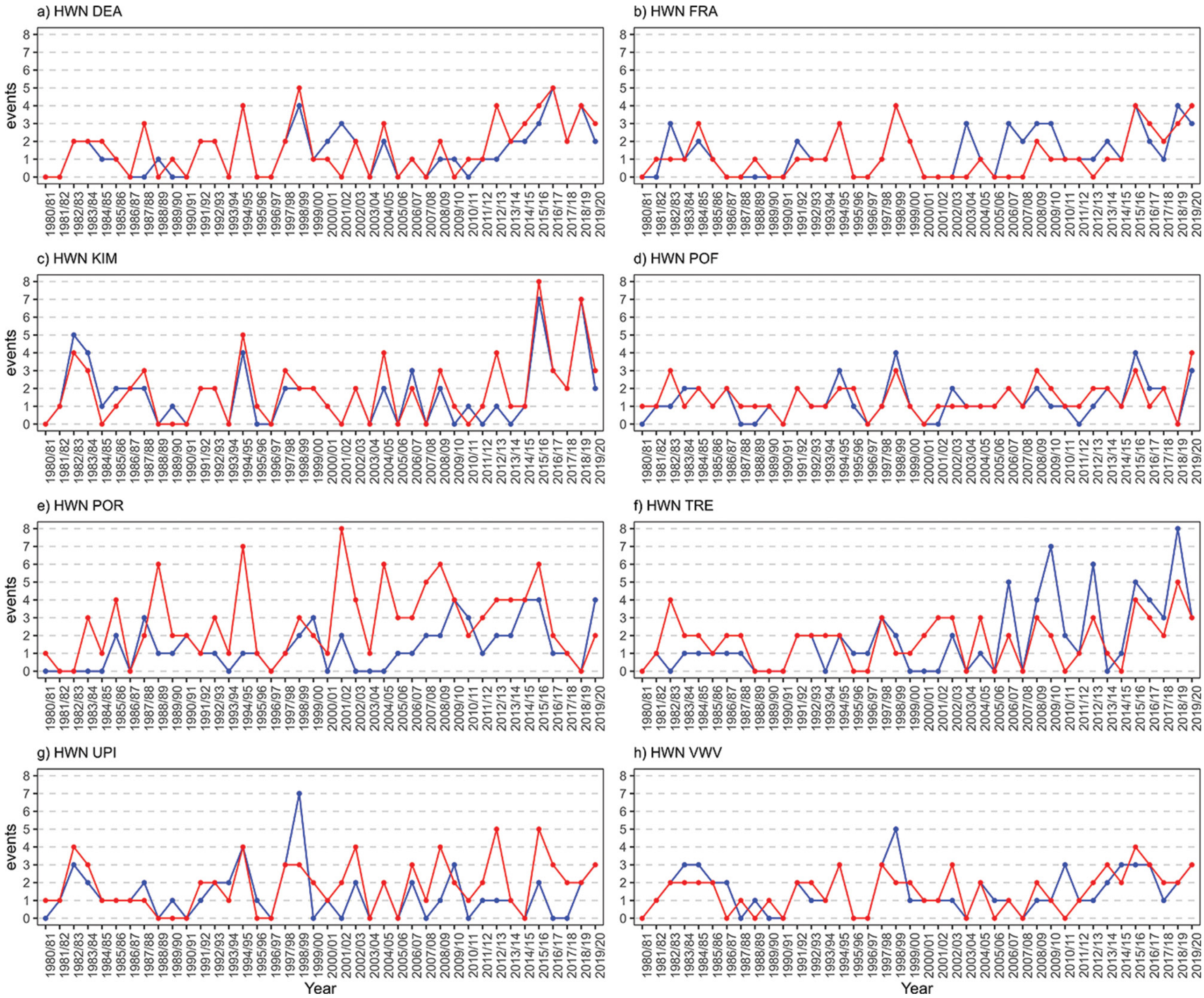

Figure 5: Time series plots depicting the total number of summer heatwave events (HWN) for the AgERA5 (red) and SAWS (blue) data sets from 1980 to 2020 for locations mapped in Figure 1.

Table 2: Tabulated results, per location mapped in Figure 1, of the comparison between the AgERA5 and SAWS data sets for the total number of days contributing to annual summer (winter) heatwave (coldwave) events (HWF [CWF]), the average temperature of annual summer (winter) heatwave (coldwave) events (HWM [CWM]), and the total number of summer (winter) heatwave (coldwave) events (HWN [CWN]). CC represents the Spearman Correlation Coefficient; RMSE represents the Root Mean Square Error; MD represents the Modified Index of Agreement. CC values denoted in bold represent statistically significant correlations at the 5% alpha level.

Index	Station	CC	MD	RMSE	Bias
HWF	DEA	**0.80**	0.76	4.04	1.28
	FRA	**0.72**	0.75	3.84	-1.33
	KIM	**0.87**	0.80	4.40	1.53
	POF	**0.78**	0.69	2.80	0.50
	POR	**0.66**	0.47	10.92	5.60
	TRE	**0.60**	0.64	6.96	0.10
	UPI	**0.65**	0.60	6.76	2.18
	VWV	**0.78**	0.69	4.26	-0.33

...Table 2 continues on next page

Table 2 continued...

Index	Station	CC	MD	RMSE	Bias
HWM	DEA	**0.74**	0.63	1.85	-0.74
	FRA	**0.79**	0.77	2.17	0.37
	KIM	**0.68**	0.51	2.90	-1.55
	POF	**0.73**	0.76	2.61	-0.51
	POR	**0.53**	0.54	6.54	1.47
	TRE	**0.69**	0.72	1.34	-0.41
	UPI	**0.66**	0.64	3.69	-0.51
	VWV	**0.93**	0.70	5.13	-1.19
HWN	DEA	**0.74**	0.78	0.99	0.23
	FRA	**0.68**	0.74	1.01	-0.28
	KIM	**0.85**	0.80	0.87	0.20
	POF	**0.74**	0.70	0.71	0.15
	POR	**0.53**	0.44	2.40	1.40
	TRE	**0.56**	0.59	1.64	-0.10
	UPI	**0.60**	0.58	1.48	0.50
	VWV	**0.68**	0.69	1.00	-0.05
CWF	DEA	**0.92**	0.80	3.25	0.83
	FRA	**0.73**	0.69	4.43	-1.61
	KIM	**0.71**	0.59	5.64	0.88
	POF	**0.78**	0.74	3.78	1.10
	POR	**0.45**	0.40	8.09	3.39
	TRE	**0.59**	0.62	6.59	-0.34
	UPI	**0.90**	0.77	3.45	1.41
	VWV	**0.82**	0.74	3.46	0.15
CWM	DEA	**0.52**	0.58	3.28	0.02
	FRA	0.25	0.36	3.67	-2.05
	KIM	**0.57**	0.56	5.26	2.91
	POF	**0.76**	0.68	2.05	0.23
	POR	0.17	0.35	3.08	0.74
	TRE	**0.52**	0.57	3.85	-0.59
	UPI	**0.80**	0.65	2.73	0.71
	VWV	**0.75**	0.64	2.32	-0.45
CWN	DEA	**0.89**	0.84	0.73	-0.10
	FRA	**0.70**	0.71	1.08	-0.39
	KIM	**0.62**	0.57	1.31	0.10
	POF	**0.78**	0.76	0.91	0.20
	POR	**0.47**	0.48	1.52	0.71
	TRE	**0.53**	0.63	1.47	-0.17
	UPI	**0.85**	0.76	0.81	0.17
	VWV	**0.66**	0.64	1.05	0.12

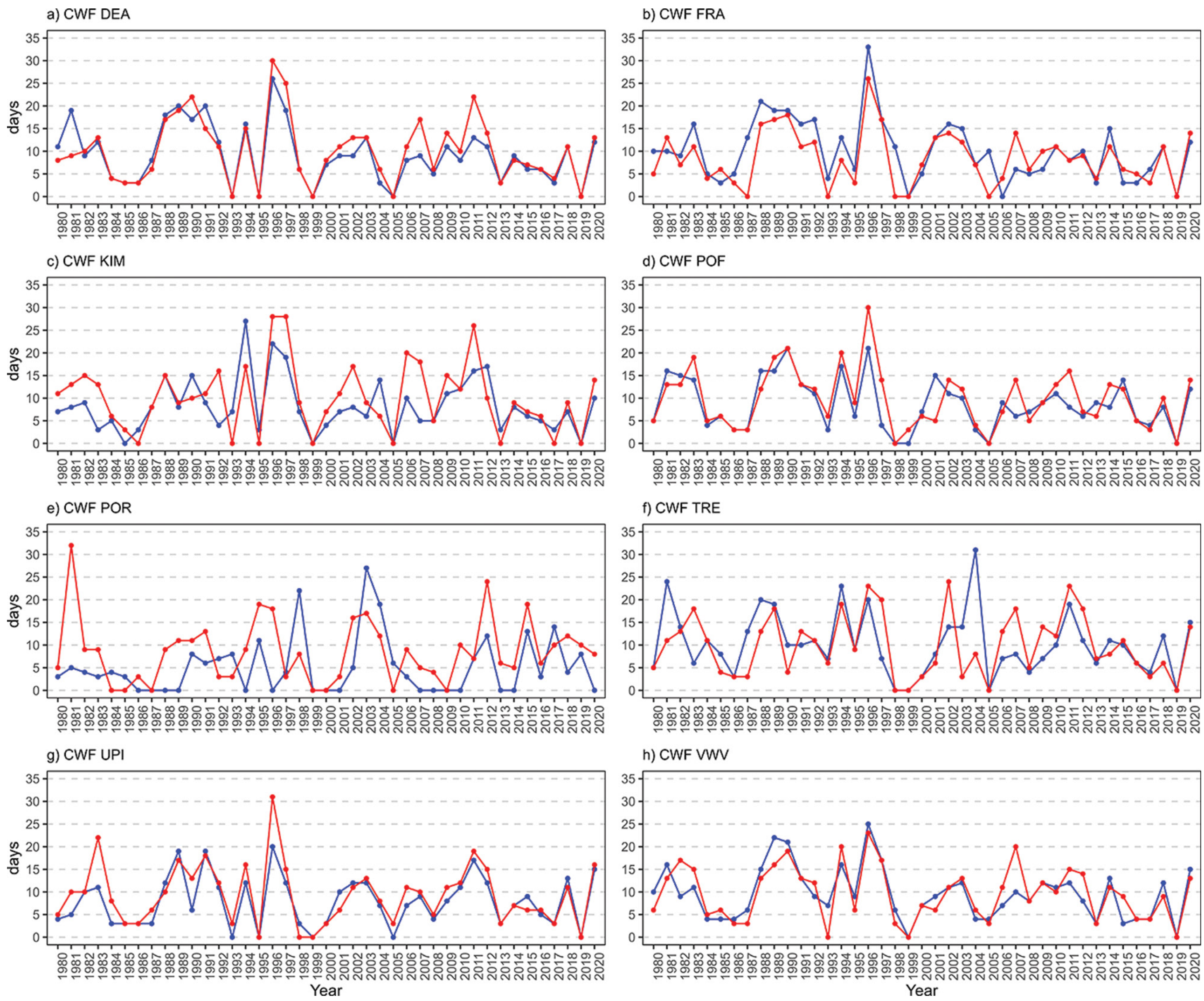

Figure 6: Time series plots depicting the total number of days contributing to annual winter coldwave events (CWF) for the AgERA5 (red) and SAWS (blue) data sets from 1980 to 2020 for locations mapped in Figure 1.

stations[47]; this highlights the value of a reliable gridded data set like the AgERA5. Moreover, by analysing ETEs with the AgERA5 data set, one can derive valuable information that can aid in developing efficient and prompt adaptive measures and strategies, which can aid in mitigating the negative impacts of ETEs, such as preventing heat stress, optimising crop management techniques, and enhancing infrastructure.[7-9,48]

In terms of the AgERA5 performance over the Northern Cape, general findings suggest strong, but varying, correspondence spatially, which is a result that is in agreement with the findings of similar studies.[20-21] Mean seasonal temperatures correspond and perform better than heatwave and coldwave characteristics from AgERA5 (Tables 1–2). This is likely due to differences at the temperature distribution extremes between the station and AgERA5 data sets (Figure 2). Furthermore, a common limitation in statistical comparisons, known as the double penalty problem, of fine-resolution data based on point-to-point analysis also contributes to the weaker performance detected for the ETE characteristics.[49] For instance, a temperature proxy may be penalised for a spurious observation due to imprecise station location, while an ETE proxy will be doubly penalised for issuing a false alarm for an ETE.[49] Based on daily to annual temperature averages, it is evident that Tms performs slightly better than Tmw (Table 1). Across the stations, Tms is characterised by higher CC and MD values, indicating better correlation and agreement between predicted and observed values (Table 1). The present analysis demonstrates a predominant underestimation of Tms and Tmw by the AgERA5 temperature proxies (Table 1), which is a finding that corresponds to results presented by Roffe and van der Walt[20] and Velikou et al.[21] Observed negative bias values between station and AgERA5 Tms and Tmw proxies can be considered a plausible explanation for the underestimation of HWM and CWM in the AgERA5 data set (Tables 1–2).[20] Among the heatwave and coldwave indices, HWM and CWM represent the ETE characteristics with the poorest performance (Table 2), and this is a finding that aligns with previous research.[20,21,29] While the AgERA5 data set performed well in representing heatwave and coldwave characteristics, these results highlight the need for caution when utilising the AgERA5 data set to examine HWM and CWM over the Northern Cape. The suboptimal performance of HWM and CWM therefore suggests that there is much room for improvement in the AgERA5 data set representation of temperatures at the tail ends of the temperature distribution. In contrast to HWM and CWM, HWF, HWN, CWF, and CWN exhibit positive biases relative to station data and demonstrate CC and MD values of greater magnitude (Table 2). While performance variations exist among different locations, and the difference in the performance of summer and winter ETE indices is relatively small, mean CC and MD values indicate superior performance for summer ETE indices compared to winter ETE indices (Tables 1–2). Stronger performance in summer ETE indices has also been observed in other studies.[20]

Although the overall performance was strong, some locations (i.e. POR, located along the western coast, and TRE, located in the Kalahari Desert) were consistently characterised by poorer performance (Tables 1–2), likely because these regions contain fewer stations compared to other parts of the Northern Cape.[20] Moreover, mean and extreme temperatures

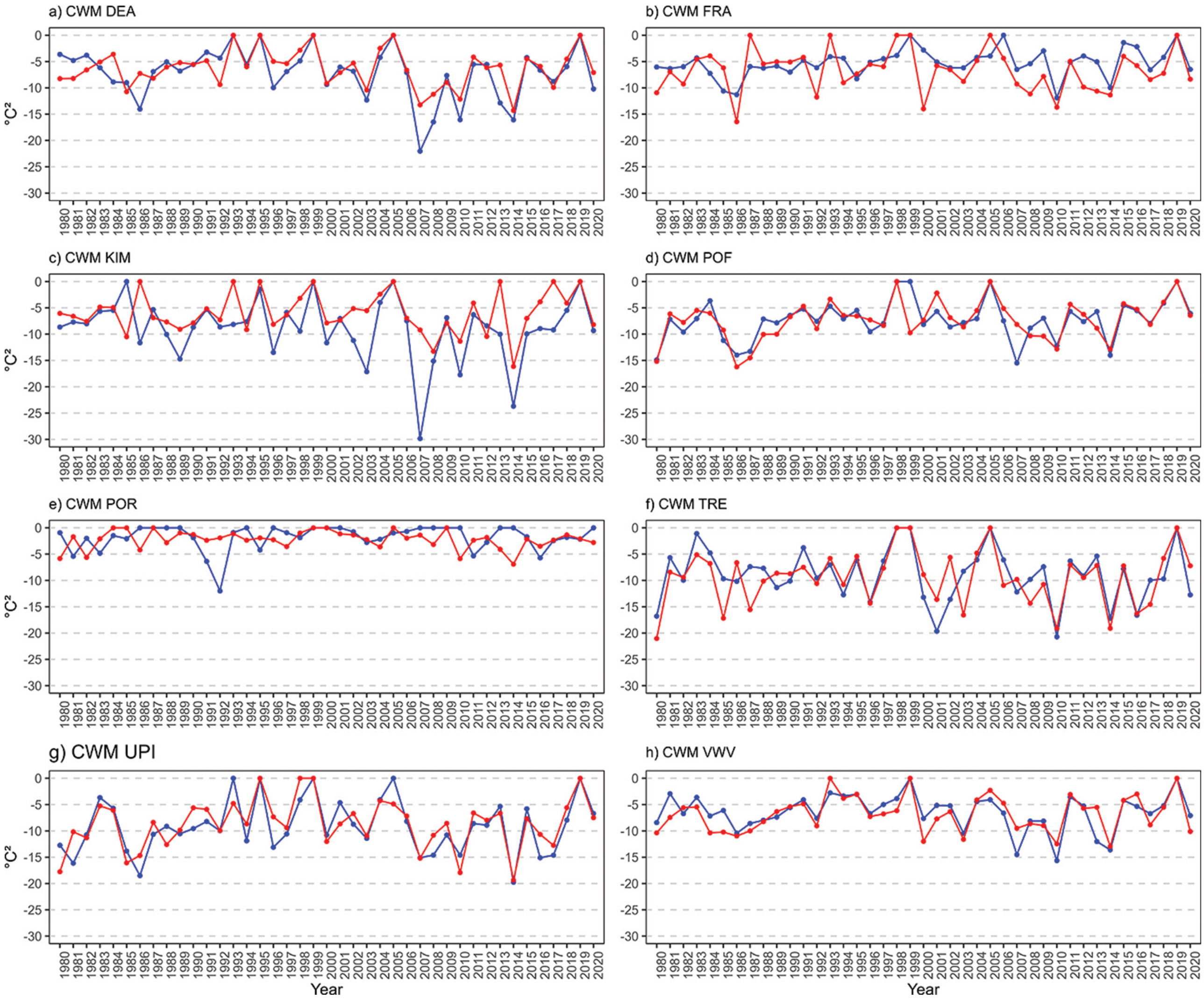

Figure 7: Time series plots depicting the average temperature of annual winter coldwave events (CWM) for the AgERA5 (red) and SAWS (blue) data sets from 1980 to 2020 for locations mapped in Figure 1.

at locations bordering the Atlantic Ocean are strongly influenced by the cold Benguela Current (and Upwelling System) which proves to be a phenomenon that is not accurately simulated by the ECMWF modelling system.[20,50] For TRE, in the Kalahari Desert region, the desert surface influences mean and extreme temperatures through cloud development mechanisms, which could also lead to parameterisation difficulties in the ERA5 reanalysis methods.[16,50] Grid cells bordering the cold Benguela current or falling within the Kalahari Desert regions should be treated with caution, and limitations and uncertainties should be appropriately acknowledged. A better understanding of these weaknesses is critical for improving future parameterisation methods for the underlying ECMWF model used for the ERA5 reanalysis; this will undoubtedly lead to better accuracy for calculated indices and long-term trends. Despite the weaknesses identified herein, this ERA5 reanalysis product has shown strong improvement based on its predecessor (ERA-Interim) and will likely improve still in the foreseeable future.[50] Furthermore, according to previous research comparing the AgERA5 data set to others, like AgCFSR and CPC, the AgERA5 outperforms these in terms of accuracy and reliability.[51,52]

The strengths and limitations of the AgERA5 presented herein support further application of the AgERA5 data set for characterising average and extreme temperatures over the Northern Cape. Findings suggest, based on moderate to strong CC values, that AgERA5 can effectively capture interannual variability patterns in the Northern Cape, and in turn will likely provide reliable trend results despite conflicting thoughts regarding the application of reanalysis for trend calculations.[20] This underscores the potential of AgERA5 as a valuable tool for studying long-term changes in the Northern Cape climate. Progressing forward, it would be important to utilise the ERA5 reanalysis to investigate synoptic-scale circulation patterns associated with the occurrence of coldwaves and heatwaves across the Northern Cape. Comprehending this using numerical weather prediction models can play an important role in the predictability of these events, ultimately informing better early warnings of ETEs over the Northern Cape.[53]

Acknowledgements

A.J.v.d.W. acknowledges the Emerging Scholar Research Programme of the University of the Free State and financial support from the Society of South African Geographers (SSAG) Centenary Award. A.J.v.d.W., J.A.K. and S.J.R. acknowledge financial support from the Faculty of the Natural and Agricultural Sciences and the Geography Department of the University of the Free State. We also acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF) for the publicly available AgERA5 data used in this study, which is available from the Copernicus Climate Change Service website (https://cds.climate.copernicus.eu/cdsapp#!/dataset/10.24381/cds.6c68c9bb?tab=overview).

Competing interests

We have no competing interests to declare.

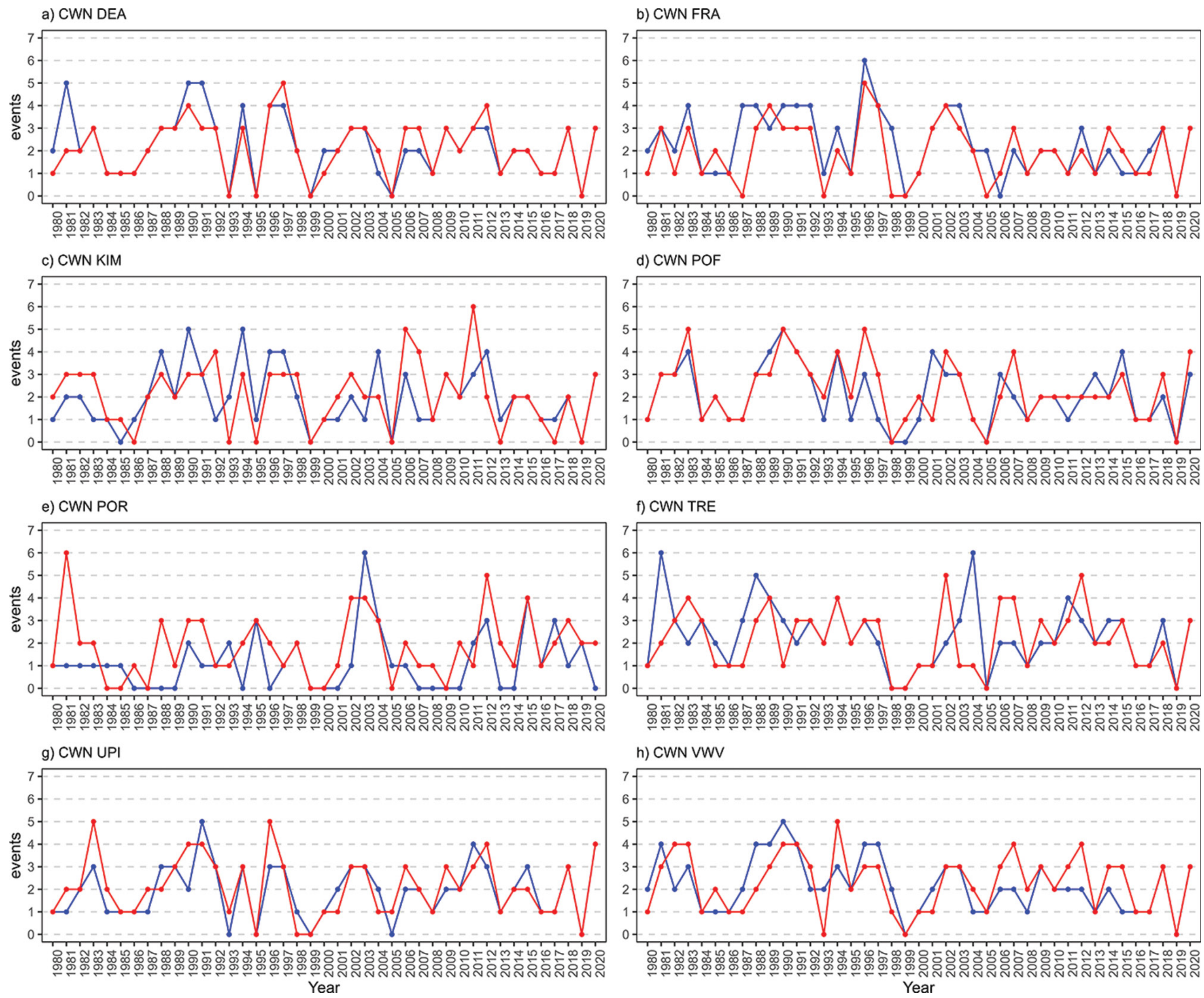

Figure 8: Time series plot representing the total number of winter coldwave events (CWN) for the AgERA5 (red) and SAWS (blue) data sets from 1980 to 2020 for all locations mapped in Figure 1.

Authors' contributions

J.A.K.: Conceptualisation, methodology, data collection, data analysis, data curation, student project management, validation, writing – initial draft. S.J.R.: Conceptualisation, methodology, data collection, data analysis, data curation, student project management, validation, writing – revisions, student supervision. A.J.v.d.W.: Conceptualisation, methodology, student project management, validation, writing – revisions, student supervision.

References

1. Mbokodo I, Bopape MJ, Chikoore H, Engelbrecht F, Nethengwe N. Heatwaves in the future warmer climate of South Africa. Atmosphere. 2020;11(7), Art. #712. https://doi.org/10.3390/atmos11070712

2. Engelbrecht F, Adegoke J, Bopape MJ, Naidoo M, Garl R, Thatcher M, et al. Projections of rapidly rising surface temperatures over Africa under low mitigation. Environ Res Lett. 2015;10(8), Art. #085004. https://doi.org/10.1088/1748-9326/10/8/085004

3. Engelbrecht FA, Monteiro P. The IPCC assessment report six working group 1 report and southern Africa: Reasons to take action. S Afr J Sci. 2021;117 (11–12), Art. #12679. https://doi.org/10.17159/sajs.2021/12679

4. Van der Walt AJ, Fitchett JM. Exploring extreme warm temperature trends in South Africa: 1960–2016. Theor Appl Climatol. 2021;143:1341–1360. https://doi.org/10.1007/s00704-020-03479-8

5. Iyakaremye V, Zeng G, Zhang G. Changes in extreme temperature events over Africa under 1.5 and 2.0 °C global warming scenarios. Int J Climatol. 2021;41(2):1506–1524. https://doi.org/10.1002/joc.6868

6. Van der Walt AJ, Fitchett JM. Trend analysis of cold extremes in South Africa: 1960–2016. Int J Climatol. 2021;41(3):2060–2081. https://doi.org/10.1002/joc.6947

7. Campbell S, Remenyi TA, White CJ, Johnston FH. Heatwave and health impact research: A global review. Health Place. 2018;53:210–218. https://doi.org/10.1016/j.healthplace.2018.08.017

8. Lobell DB, Bänziger M, Magorokosho C, Vivek B. Nonlinear heat effects on African maize as evidenced by historical yield trials. Nat Clim Change. 2011;1:42–45. https://doi.org/10.1038/nclimate1043

9. Nienaber JA, Hahn GL. Livestock production system management responses to thermal challenges. Int J Biometeorol. 2007;52:149–157. https://doi.org/10.1007/s00484-007-0103-x

10. Abatan AA, Abiodun BJ, Lawalc KA, Gutowski WJ. Trends in extreme temperature over Nigeria from percentile-based threshold indices. Int J Climatol. 2015;36(6):2527–2540. https://doi.org/10.1002/joc.4510

11. Harrington J, Otto F. Reconciling theory with the reality of African heatwaves. Nat Clim Chang. 2020;10:796–798. https://doi.org/10.1038/s41558-020-0851-8

12. Tyson PD, Preston-Whyte RA. The atmosphere and weather of southern Africa. 2nd ed. Cape Town: Oxford University Press; 2000. https://catalog.hathitrust.org/api/volumes/oclc/42309991.html
13. Jury MR, Parker B. Cold winters over the South African highveld: Pattern recognition and preliminary models. S Afr J Sci. 1999;94:473–481. https://hdl.handle.net/10520/AJA00382353_239
14. Levey KM. Interannual temperature variability and associated synoptic climatology at Cape Town. Int J Clim. 1996;16(3):293–306. https://doi.org/10.1002/(SICI)1097-0088(199603)16: 3%3C293::AID-JOC3%3E3.0.CO;2-3
15. Driver P, Reason CJC. Variability in the Botswana High and its relationships with rainfall and temperature characteristics over southern Africa. Int J Climatol. 2017;37(S1):570–581. https://doi.org/10.1002/joc.5022
16. Manatsa D, Reason C. ENSO–Kalahari Desert linkages on southern Africa summer surface air temperature variability. Int J Climatol. 2017;37(4):1728–1745. https://doi.org/10.1002/joc.4806
17. Mbokodo IL. Heat waves in South Africa: Observed variabilty, structure and trends [master's dissertation]. Thohoyandou: University of Venda; 2017.
18. Mkiva N. The Climatology of heat waves in the North West Province, South Africa [master's dissertation]. Potchefstroom: North-West University; 2020.
19. Van der Walt AJ, Fitchett JM. Extreme Temperature Events (ETEs) in South Africa: a review. S Afr Geogr J. 2021;104(1):70–88. https://doi.org/10.1080/03736245.2021.1907219
20. Roffe SJ, Van der Walt AJ. Representation and evaluation of southern Africa's seasonal mean and extreme temperatures in the ERA5-based reanalysis products. Atmos Res. 2023;284(106591):0169-8095. https://doi.org/10.1016/j.atmosres.2022.106591
21. Velikou K, Lazoglou G, Tolika K, Anagnostopoulou C. Reliability of the ERA5 in replicating mean and extreme temperatures across Europe. Water. 2022;14(4):543. https://doi.org/10.3390/w14040543
22. Schumacher V, Justino F, Fernández A, Meseguer-Ruiz O, Sarricolea P, Comin A. Comparison between observations and gridded data sets over complex terrain in the Chilean Andes: Precipitation and temperature. Int J Climatol. 2020;40(12):5266–5288. https://doi.org/10.1002/joc.6518
23. Harris I, Osborn TJ, Jones P, Lister D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci Data. 2020;7:109. https://doi.org/10.1038/s41597-020-0453-3
24. Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim. 2017:30(14):5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1
25. Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020;146(730):1999–2049. https://doi.org/10.1002/qj.3803
26. C3S. Copernicus climate change service climate data store [data set on the Internet]. c2020 [cited 2022 Mar 29]. https://doi.org/10.24381/cds.6c68c9bb
27. Oyler JW, Dobrowski SZ, Ballantyne AP, Klene AE, Running SW. Artificial amplification of warming trends across the mountains of the western United States. Geophys Res Lett. 2015;42(1):153–161. https://doi.org/10.1002/2014GL062803
28. Roffe SJ, Fitchett JM, Curtis CJ. Quantifying rainfall seasonality across South Africa on the basis of the relationship between rainfall and temperature. Clim Dyn. 2021;56:2431–2450. https://doi.org/10.1007/s00382-020-05597-5
29. Keller JD, Wahl S. Representation of climate in reanalyses: An intercomparison for Europe and North America. J Clim. 2021;34(5):1667–1684. https://doi.org/10.1175/JCLI-D-20-0609.1
30. Schulzweida U. CDO user guide. Hamburg: Max Planck Institute for Meteorology; 2021 [cited 2023 Apr 14]. Available from: https://code.mpimet.mpg.de/projects/cdo/embedded/index.html
31. Herold N, McComb J. Climpact [software on the Internet]. GitHub; 2021. [cited 2022 Sep 07]. Available from: https://github.com/ARCCSS-extremes/climpact
32. Van der Walt AJ, Fitchett JM. Statistical classification of South African seasonal divisions on the basis of daily temperature data. S Afr J Sci. 2020;116 (9–10), Art. #7614. https://dx.doi.org/10.17158/sajs.2020/7614
33. Kruger AC, Nxumalo M. Surface temperature trends from homogenized time series in South Africa: 1931–2015. Int J Climatol. 2017;37(5):2364–2377. https://doi.org/10.1002/joc.4851
34. Klein Tank AMG, Zwiers FW, Zhang X. Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. Geneva: World Meteorological Organization; 2009.
35. Durre I, Menne MJ, Gleason BE, Houston TG, Vose RS. Comprehensive automated quality assurance of daily surface observations. J Appl Meteorol Climatol. 2010;49(8):1615–1633. https://doi.org/10.1175/2010JAMC2375.1
36. Lakhraj-Govender R, Grab S, Ndebele NE. A homogenized long-term temperature record for the Western Cape Province in South Africa: 1916–2013. Int J Climatol. 2016;37(5):2337–2353. https://doi.org/10.1002/joc.4849
37. Wolski P, Conradie S, Jack C, Tadross M. Spatio-temporal patterns of rainfall trends and the 2015–2017 drought over the winter rainfall region of South Africa. Int J Climatol. 2021;41(S1):E1303–E1319. https://doi.org/10.1002/joc.6768
38. Perkins SE, Alexander L V. On the measurement of heat waves. J Clim. 2013;26(13):4500–4517. https://doi.org/10.1175/JCLI-D-12-00383.1
39. Nairn JR, Fawcett RJ. The excess heat factor: a metric for heatwave intensity and its use in classifying heatwave severity. Int J Environ Res Public Health. 2014;12(1):227–253. https://doi.org/10.3390/ijerph120100227
40. Perkins-Kirkpatrick SE, Gibson PB. Changes in regional heatwave characteristics as a function of increasing global temperature. Sci Rep. 2017;7, Art. #12256. https://doi.org/10.1038/s41598-017-12520-2
41. Wang C, Zhang L, Lee SK, Wu L, Mechoso CR. A global perspective on CMIP5 climate model biases. Nat Clim Change. 2014;4:201–205. https://doi.org/10.1038/nclimate2118
42. Li CZ, Zhang L, Wang H, Zhang YQ, Yu FL, Yan DH. The transferability of hydrological models under nonstationary climatic conditions. Hydrol Earth Syst Sci. 2012;16(4):1239–1254. https://doi.org/10.5194/hess-16-1239-2012
43. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res. 2005;30:79–82. https://doi.org/10.3354/cr030079
44. Sheridan SC, Lee CC, Smith ET. A comparison between station observations and reanalysis data in the identification of extreme temperature events. Geophys Res Lett. 2020;47(15), Art. #2020GL088120. https://doi.org/10.1029/2020GL088120
45. Van der Walt AJ, Kruger JA, Roffe SJ. Fine-resolution analysis of the spatiotemporal characteristics of heatwaves in the Maloti-Drakensberg region, southern Africa: 1979–2021. Trans R Soc South Africa. 2023;78(1/2):5–15. https://doi.org/10.1080/0035919X.2023.2175075
46. Khumalo Z. Heat wave claims lives of 5 people in Northern Cape. News24. 2023 January 20. Available from: https://www.sajs.co.za/vancouver-referencing-style
47. CRED. The international disaster database, EM-DAT query tool [database on the Internet]. c2023 [cited 2023 Sep 24]. Available from: https://www.emdat.be/
48. Stöllberger C, Lutz W, Finsterer J. Heat-related side-effects of neurological and non-neurological medication may increase heatwave fatalities. Eur J Neurol 2009;16(7):879–882. https://doi.org/10.1111/j.1468-1331.2009.02581.x
49. Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi, et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q J R Meteorol Soc. 2011;137(656):553–597. https://doi.org/10.1002/qj.828

50. Gleixner S, Demissie T, Diro GT. Did ERA5 improve temperature and precipitation reanalysis over East Africa? Atmosphere. 2020;11(9):996. https://doi.org/10.3390/atmos11090996

51. Araghi A, Martinez CJ, Olesen JE, Hoogenboom G. Assessment of nine gridded temperature data for modeling of wheat production systems. Comput Electron Agric. 2022;199, Art. #107189. https://doi.org/10.1016/j.compag.2022.107189

52. Lagos-Zúñiga M, Balmaceda-Huarte R, Regoto P, Torrez L, Olmo M, Lyra A, et al. Extreme indices of temperature and precipitation in South America: Trends and intercomparison of regional climate models. Clim Dyn. 2022. https://doi.org/10.1007/s00382-022-06598-2

53. Bobape M-JM, Sebego E, Ndarana T, Maseko B, Netshilema M, Gijben M, et al. Evaluating South African Weather Service information on Idai tropical cyclone and KwaZulu-Natal flood events. S Afr J Sci. 2021;117(3–4), Art. #7911. https://dx.doi.org/10.17159/sajs.2021/7911