



Modelling temperature extremes in the Limpopo province: bivariate time-varying threshold excess approach

Daniel Maposa¹ · Anna M. Seimela¹ · Caston Sigauke² · James J. Cochran³

Received: 28 July 2020 / Accepted: 27 January 2021 / Published online: 11 February 2021
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

A common problem that arises in extreme value theory when dealing with several variables (such as weather or meteorological) is to find an appropriate method to assess their joint or conditional multivariate extremal dependence behaviour. The method for choosing an appropriate threshold in peaks-over threshold approach is also another problem of endless debate. In this era of climate change and global warming, extreme temperatures accompanied by heat waves and cold waves pose serious economic and health challenges particularly in small economies or developing countries like South Africa. The present study attempts to address these problems, in particular, to deal with and capture dependencies in extreme values of two variables, by applying bivariate conditional extremes modelling with a time-varying threshold to Limpopo province's monthly maximum temperature series. Limpopo and North West provinces are the two hottest provinces in South Africa characterised by heat waves and the present study is carried out in the Limpopo province at Mara, Messina, Polokwane and Thabazimbi meteorological stations for the period 1994–2009. With the aim to model extremal dependence of maximum temperature at these four meteorological stations, two modelling approaches are applied: bivariate conditional extremes model and time-varying threshold. The latter approach was used to capture the climate change effects in the data. The main contribution of this paper is in combining these two approaches in bivariate extremal dependence modelling of maximum temperature extremes in the Limpopo province of South Africa. The findings of the study revealed both significant positive and negative extremal dependence in some pairs of meteorological stations. Among the major findings were the significant strong positive extremal dependence of Thabazimbi on high-temperature values at Mara and the strong negative extremal dependence of Polokwane on high-temperature values at Messina. The findings of this study play an important role in revealing information useful to meteorologists, climatologists, agriculturalists, and planners in the energy sector among others.

Keywords Bivariate extreme modelling · Climate change · Conditional extremes model · Laplace margins · Temperature extremes · Time-varying threshold

✉ Daniel Maposa
danmaposa@gmail.com

Extended author information available on the last page of the article

1 Introduction

In most statistical analysis applications, common focus is placed on the measures of central tendency of the data which may include the mean and/or median. In rare or extreme events, however, interest is in the tails of the underlying distribution of data. These rare or extreme events are usually outliers in a dataset which, in most cases, are discarded during data cleaning and analysis. Natural hazards, natural disasters, and most pandemic diseases such as the 1918 Spanish Influenza and the recent new coronavirus (Covid-19) are examples of rare events. Increase in the number, frequency and intensity of natural hazards has characterised the 21st century (Maposa et al. 2017). The increased number, frequency and intensity of natural hazards such as heat waves, cold waves, tornadoes, hurricanes, floods and droughts are generally attributed to climate change (Diriba and Debusho 2020; Diriba et al. 2015; Maposa et al. 2017). Extreme value theory (EVT) is the branch of statistics commonly used in analysing extreme events (Acero et al. 2014; Bhagwandin 2013; Coles 2001; Ferreira and de Haan 2015; Heffernan and Tawn 2004; Keef et al. 2013; Maposa et al. 2017; Nemukula et al. 2018).

There are two fundamental realisations in EVT modelling; the block maxima and the peaks-over threshold (POT) (Ferreira and de Haan 2015). The present study is based on the POT approach setting. In this study, the POT approach is applied to model extreme temperature in the Limpopo province of South Africa. The POT approach has several variations with regards to the selection or identification of the threshold. Some common approaches in threshold selection are the use of residual mean excess and stability plots in Keef et al. (2013) which usually depend on the subjective visual interpretation or assessment of the plots by the user, automated threshold selection in Thompson et al. (2009) which is based on the distribution of the difference of parameter estimates when the threshold is changed. More recent advanced approaches in threshold selection are in Nemukula et al. (2018) where a penalised cubic smoothing spline is used to perform a nonlinear detrending of the data prior to fitting bivariate threshold excess models to positive residuals above the threshold, while the other advanced approach in Sigauke and Bere (2017) involves using a time-varying threshold with generalised Pareto distribution (GPD) to capture the changing climatic effects in the data.

The present study will combine the latter two approaches to perform extreme value analysis of maximum temperatures in the Limpopo province of South Africa using a bivariate time-varying threshold approach. The literature of this nature, particularly with application to maximum temperature extremes, is scarce in the province and South Africa as a whole. In general, there is limited literature of this nature worldwide. Therefore, this approach will bring a novel EVT application approach to maximum temperature extremes in the province (Figure 1).

1.1 Background

Many areas of society throughout the world are susceptible to the effects of extreme values of temperature (Keelings and Waylen 2015; Nemukula 2018; Raghavendra et al. 2019). Temperature extremes such as heat waves and cold waves are deadly natural hazards although they occur more slowly and are more difficult to detect than a hurricane or a cyclone (DEA 2019; Henderson and Muller 1997; Sigauke and Nemukula 2018). Heat waves are reportedly occurring more frequently across much of the globe including South

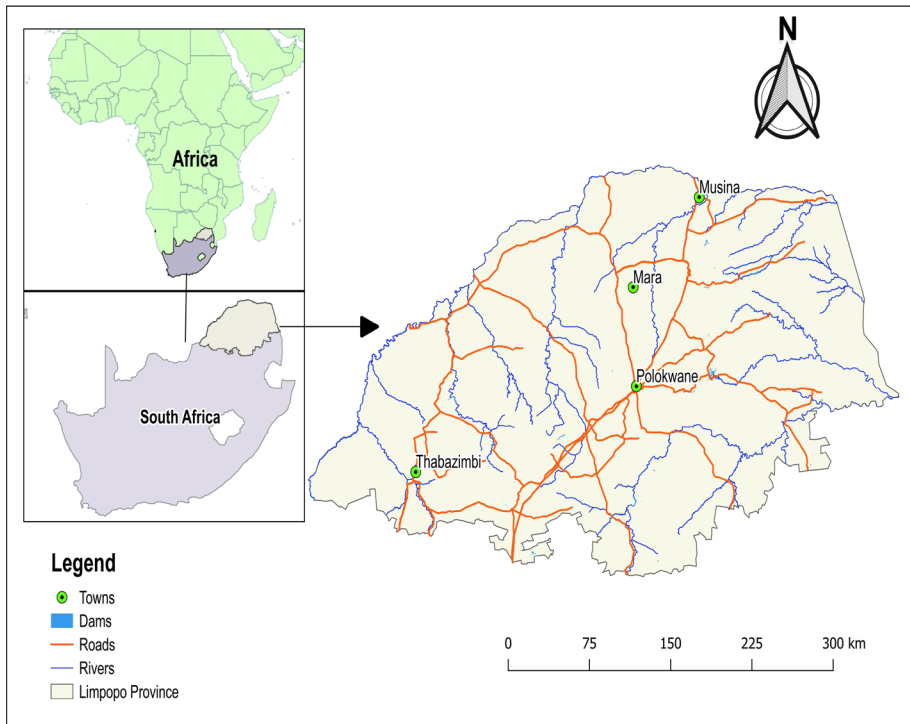


Fig. 1 Limpopo province map showing the four meteorological stations. Source: [Authors' own contribution]

Africa, and under a global warming climate they are expected to increase in frequency, intensity, and duration (Coumou and Robinson 2013; Coumou et al. 2013; Keelings and Waylen 2015; Sigauke and Nemukula 2018). Climate change is regarded as the most contributing factor to recent increases in global temperatures (Winter 2016). Worldwide, temperature extremes have a major impact on agricultural, economic, health and energy sectors (Raggad 2018; Reddy and Vincent 2017; Sigauke and Nemukula 2018). For instance, extremely high temperatures such as heat waves may result in loss of plant and animal species, losses in economic goods, high energy demand for air conditioning, death resulting from heart attacks, heat cramps, fainting, heat strokes and heat exhaustion (Makate et al. 2019; Sigauke and Nemukula 2018). Extreme low temperatures such as cold waves may result in water pipelines to freeze and burst, a rise in the demand for fuels and electricity, animals not able to graze and die of starvation, frostbites in humans and animals, and other serious medical ailments (Diriba et al. 2015; Henderson and Muller 1997).

In Africa, the impact of a changing climate varies by region (Sigauke and Nemukula 2018; Wright et al. 2014; Yamba et al. 2011). By the end of the century, Southern Africa is expected to experience an average temperature increase of about two degrees Celsius higher than the predicted average global increase (Wright et al. 2014). In the past four decades (1980–2015), Southern Africa experienced 491 climate disasters (meteorological, hydrological, and climatological) that resulted in 110,978 deaths, left 2.49 million people homeless and affected an estimated 140 million people (Reddy and Vincent 2017).

Changing weather conditions increase electricity demand due to the fact that in winter heating systems are used, while in summer air conditioning appliances are used (Sigauke and Nemukula 2018). This creates a big problem, particularly in South Africa where the national electricity supplier, ESKOM, is already battling with meeting the demands of the nation in energy supply (Chikobvu and Sigauke 2013; Sigauke and Nemukula 2018). ESKOM has experienced an increased demand for electricity supply over recent decades, consequently leading to rolling blackouts (Hohne et al. 2019). According to Yamba et al. (2011), energy demand is expected to change drastically in South Africa as a result of increasing temperatures and changing weather patterns, consequently affecting heating and cooling demands.

Extreme climate and weather events such as heat waves, cold waves and drought have negative impacts on the society, environment and resources management, particularly in developing countries like South Africa (Gebrechorkos et al. 2019; Sigauke and Nemukula 2018; Wolf et al. 2010). Climate change has resulted in rising temperature trends with associated changes in temperature extremes across the globe, which has the potential to impact on human health. It is generally anticipated that as the planet heats, climate variability will increase (Krugger and Sekele 2013; Reddy and Vincent 2017). Over the last five decades, South Africa has experienced a considerable increase in mean annual temperatures with hot and cold extremes increasing and decreasing in frequency across the country (DEA 2019; Diriba and Debusho 2020; Mbokodo 2017). Temperature is one of the main climatic elements that can indicate climate change (Toros et al. 2019; Worku et al. 2019; Wright et al. 2014). Global warming and its associated increase in temperature extremes pose a substantial challenge on natural systems. It is widely believed that the changing temperature due to global warming is permanently changing the earth's climate. That is, an increase in temperature is likely to lead to a global increase in drought condition, decrease in water supplies due to evapotranspiration and an increase in agricultural demand (Diriba et al. 2015; Nhamo et al. 2019; Ochanda 2016).

Limpopo province, where the present study is carried out, is one of the nine provinces of South Africa and is one of the hottest provinces in the country (Krugger and Shongwe 2004; Phophi et al. 2020). The province is among the lowest-ranked in terms of regional gross domestic product (GDP) per capita and it is the most vulnerable province to climate change impacts. Drought is one of the main problems in the province that affect the agricultural sector due to high temperatures and unreliable rainfall (Maponya and Mphandeli 2012b; Mphandeli and Maponya 2013; Tshiala and Olwoch 2010). The recent high temperatures in Limpopo province were experienced in the western bushveld and lowveld in October 2019 (Phophi et al. 2020). These extremely high temperatures can affect agricultural production in the province leading to scarce food and water resources, which is a big threat to a country like South Africa, where the population is rapidly growing (Makate et al. 2019; Maponya and Mphandeli 2012a; Tshiala and Olwoch 2010). South Africa is also concerned about public health around extreme hot events and how the impact of these events may change in the future (Mbokodo 2017; Wright et al. 2014). For instance, extremely high temperatures and prolonged heat waves can damage agricultural production, increase energy and water consumption and also badly affect human well-being and even cause loss of livestock, plant and human lives (Chikobvu and Sigauke 2013; Reddy and Vincent 2017). During the 21st century, the global surface temperature has increased by about 0.85 °C and many areas have experienced significant warming (Toros et al. 2019). Krugger and Shongwe (2004) found a considerable increase in temperature between 1960 and 2003 for the three stations Bela Bela, Polokwane and Messina (locally known as Musina) situated in the Limpopo province in north-eastern South Africa. The present study

is built against this background coupled with the challenges brought about by temperature extremes in the Limpopo province (Fig. 1).

1.2 Literature review on extremal dependence modelling

Several studies have modelled extremal dependence with application to various variables including temperature, wind speed, rainfall, air pollution, insurance claims, financial losses and many others. Southworth et al. (2020) give a detailed computational approach of multivariate extreme value data conditional modelling using *R* package called ‘texmex’. The authors cautioned that dependencies between variables in the body of the data do not necessarily imply dependencies in the extremes. Another issue that makes multivariate extreme value modelling more complicated than univariate is that for an observation to be considered multivariate extreme it has to be extreme in all components simultaneously. These authors, Southworth et al. (2020), explored and gave a detailed interpretation of pairwise extremal dependence and conditional multivariate extreme value modelling using the approach of Heffernan and Tawn (2004) which proceeds by first fitting the GPD models to the marginal variables before estimating the dependence structure. Similar to the GPD model for excesses over a given threshold, the modelling approach for Heffernan and Tawn (2004) also conditions on a variable exceeding a predetermined threshold. In the present study, we also follow closely the approach by Heffernan and Tawn (2004) and the *R* computational approach by Southworth et al. (2020). More details on the modelling framework of these approaches are given in the next section of models.

Another issue of importance in EVT methodology is the choice of a threshold when using the POT approach. Apart from the threshold selection approaches in Sigauke and Bere (2017) and Thompson et al. (2009) mentioned earlier in the introductory section, more recently Verster and Raubenheimer (2020) proposed a generalised model in the Bayesian approach that uses the properties of the posterior distribution to select an optimal threshold without a visual inspection. The Bayesian threshold approach by Verster and Raubenheimer (2020) is based on the Topp-Leone Pareto (TLPa) distribution and was shown to perform well. In another study on threshold choice, Minkah and de Wet (2014) investigated constant versus covariate dependent threshold in the POT approach. These authors, Minkah and de Wet (2014), proposed a covariate dependent threshold which is based on expectiles. They argued that although no threshold choice method is universally the best, strong arguments against the use of constant threshold is that an observation that may be considered extreme at some covariate level may not necessarily qualify as an extreme observation when considered at another covariate level. The newly proposed approach was compared with the constant and quantile regression thresholds in a simulation study based on exponential growth data for the estimation of the GPD tail index. The findings by Minkah and de Wet (2014) revealed that the covariate dependent threshold approach outperformed the other methods for smaller to medium values in the data, while for larger values of the response variable the constant threshold outperformed the other methods. Another threshold selection method slightly different from that of Minkah and de Wet (2014) was proposed by Thompson et al. (2009). The threshold selection approach by Thompson et al. (2009) is a pragmatic automated threshold selection method which is based on the distribution of the difference of parameter estimates when the threshold is changed. The similarity on methods by Thompson et al. (2009) and Minkah and de Wet (2014) is that the automated threshold selection can also be extended to depend on a covariate value such as the wave direction cosine. In a separate study, Sigauke and Bere (2017) used a GPD with time-varying covariates and thresholds to model daily peak electricity demand for South Africa.

The threshold selection approach by Sigauke and Bere (2017) makes use of a penalised cubic smoothing spline with a constant shift factor as a time-varying threshold. They used an intervals estimator method in declustering observations that exceed the threshold. They further included temperature as a covariate in the GPD parameters in order to explore its influence on electricity demand. The findings by Sigauke and Bere (2017) showed a better fit for the GPD model to the data when compared to the generalised extreme value (GEV) distribution. The present study will adopt the GPD time-varying threshold selection approach by Sigauke and Bere (2017) to cater for climate change effects in the maximum temperature data. Unlike the use of the method by Sigauke and Bere (2017) in univariate modelling, the present study extends its use to conditional multivariate extremal dependence modelling.

A study closely related to the present study in multivariate extreme value theory (MEVT) is that of Nemukula et al. (2018) who used bivariate threshold excess in modelling temperature extremes in the Limpopo province for three meteorological stations Lephalale, Polokwane and Thohoyandou. Similar to the present study, the approach by Nemukula et al. (2018) also used a penalised cubic smoothing spline to perform a nonlinear detrending of the temperature data before fitting bivariate threshold excess models to positive residuals above the threshold. The present study, however, extends the approach of Nemukula et al. (2018) by using a time-varying threshold instead of a constant threshold to capture the climate change effects in the monthly maximum temperature data series. Additionally, except for Polokwane meteorological station, three new stations Mara, Messina (also known as Musina in the local language) and Thabazimbi are used in the present study (Fig. 1).

Another literature of importance in this present study concerning MEVT is that of Tilloy et al. (2020) who evaluated the efficacy of bivariate extreme value modelling approaches in their estimation of risks generated by multi-hazard scenarios. These authors, Tilloy et al. (2020), fitted six distinct stochastic copula models to the synthetic datasets and concluded that there is no one shoe size fits all in bivariate extreme value modelling. They found that no one model was able to fit their synthetic data for all the parameters, instead, several models were appropriate to fit the data. Tilloy et al. (2020) limited their study to stochastic copulas and the bivariate case of multivariate models based on Heffernan and Tawn (2004). In their evaluation of bivariate extreme dependence, they discussed in detail issues on asymptotic dependence and asymptotic independence, as well as tail dependence measures. They argued that extremal dependence in practice tend to weaken at higher levels which may lead to dependence between variables being observed in the body of the joint distribution, while the multivariate distribution is in the maximum domain of attraction of independence. In their discussion of the bivariate models which include copulas and conditional extremes, Tilloy et al. (2020) advocated for the conditional extremes model by Heffernan and Tawn (2004) and Keef et al. (2013) which uses the Laplace margins. The conditional extremes model can accommodate both asymptotic dependence and asymptotic independence (see Heffernan and Tawn 2004; Keef et al. 2013, for more details). The present study also adopts the Heffernan and Tawn (2004) conditional extremes model with the addition of a time-varying threshold.

1.3 Research highlights

This paper addresses issues related to climate change, global warming and in particular maximum temperature extremes in the Limpopo province of South Africa. The study is based on combining two main approaches; bivariate conditional extremes model (Heffernan and Tawn 2004; Keef et al. 2013; Southworth et al. 2020) and time-varying threshold (Sigauke and Bere 2017). Conditional extremes modelling is crucial in

the study of the dependence structure among several variables. Conditioning on one variable helps to understand the significant positive (or negative) extremal dependence of the remaining variables on the large values of the conditioning variable. This paper presents an application of bivariate threshold excess modelling approach with a positive shift factor as a time-varying threshold to the monthly maximum temperature extremes in four meteorological stations of Limpopo province in South Africa namely Mara, Messina, Polokwane and Thabazimbi. Among the major findings were the significant strong positive extremal dependence of Thabazimbi on large temperature values of Mara and the strong negative extremal dependence of Polokwane on large temperature values of Messina.

The main contribution of this paper is in using a penalised cubic smoothing spline to perform a nonlinear detrending of the temperature data prior to fitting bivariate threshold excess models based on Laplace margins to positive residuals above the threshold and a positive shift factor as a time-varying threshold to capture the climate change or seasonality and/or cyclic effects in the data. The existing gap in the literature was in combining these two approaches using conditional extremes dependence modelling. The overall significance of this bivariate conditional extremes dependence modelling study lies on quantifying the dependence effects of maximum temperature extremes amongst the various meteorological stations in order to reveal some useful information needed for planning by climatologists, meteorologists, agriculturalists, decision-makers and planners in the energy sector.

The rest of the paper is organised as follows: Sect. 2 gives the theoretical framework of the statistical models considered in this study which include conditional multivariate extreme value modelling, threshold selection, bivariate threshold excess model, Laplace marginal transformation, as well as the data and variables. Sect. 3 presents empirical results and a comprehensive discussion of the results. The concluding remarks and areas for future research are presented in Sect. 4.

2 Models

2.1 Conditional multivariate extreme value modelling

Some joint tail models, copulas and other multivariate extreme methods assume that all the variables should be very large at the same time. This limitation is overcome by the use of conditional extremes model (CEM) (Heffernan and Tawn 2004; Keef et al. 2013; Tilloy et al. 2020). With the CEM, we seek to estimate among variables the dependence structure in which one variable is conditioned on being extreme and seeks to model the conditional distribution (Heffernan and Tawn 2004; Keef et al. 2013).

In this study, we adopt the Heffernan and Tawn (2004) methodology which uses a conditional multivariate approach. In this approach, the dependence structure is estimated after fitting the GPD first to the tails of the marginal distributions. The bivariate case is considered in this study, that is, given the threshold excesses of temperature, the conditional multivariate approach describes the temperature conditional distribution using a regression type model (Heffernan and Tawn 2004; Keef et al. 2013).

In the bivariate case, we focus on approximating a joint distribution $F(x_i, x_j)$ on $x_i > u_i$ and $x_j > u_j$ on regions of u_i and u_j that are sufficiently large.

2.1.1 Threshold selection

In this study, a time-varying threshold with a positive shift operator τ is used (for details of this modelling approach see Sigauke and Bere 2017). The time-varying threshold used is a penalised cubic smoothing spline given in Eq. (1),

$$\pi(t) = \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda \int (f''(t))^2 dt + \tau \quad (1)$$

where y_i denotes maximum temperature, λ is a smoothing parameter and $\tau \in \mathfrak{R}$ is a shift factor which should be large enough to allow asymptotic conditions to be satisfied when we fit the GPD. We extract observations above the time-varying threshold without the shift factor. The positive shift factor τ is then estimated using high quantiles in the ‘texmex’ R package.

2.1.2 Bivariate threshold excess model

For this current study, the multivariate modelling is limited to the pairwise combination of variables. A model used in EVT for exceedances above a threshold τ (i.e. $X > u$) is the GPD whose distribution function is given in Eq. (2),

$$G(x) = 1 - \eta \left[1 + \xi \left(\frac{x-u}{\sigma_u} \right) \right]^{-\frac{1}{\xi}} \quad (2)$$

where $\eta = \Pr(X > u)$, $\xi \neq 0$ and $\sigma > 0$ for a family defined on $\left\{ \left(1 + \xi \left(\frac{x-u}{\sigma_u} \right) \right) \right\} > 0$ and $x - u > 0$. Hence $F(x) \approx G(x)$ on $x > u$ for a sufficiently high threshold u with parameters η , ξ and σ_u (Coles 2001). The focus is to obtain a family that approximates a joint distribution $F(x, y)$ on $x > u_x, y > u_y$ regions for thresholds u_x and u_y that are sufficiently large.

2.1.3 Marginal transformation: Laplace margins

Before the dependence modelling, the regression type dependence model’s structure first transforms the margins to the standardised Laplace margins. The Laplace distribution is characterised by both symmetry and exponential tails. The use of the regression structure is simplified if Laplace margins are used compared to other transformations such as the Fréchet and Gumbel margins (Heffernan and Tawn 2004; Keef et al. 2013). This means that we only need a single model structure to describe both positive and negative dependence.

Let

$$Y_i = \begin{cases} \log \{2F_i(X_i)\}, & \text{for } X_i < F_i^{-1}(0.5) \\ -\log \{2[1 - F_i(X_i)]\}, & \text{for } X_i > F_i^{-1}(0.5) \end{cases} \quad (3)$$

for all $i \in D$. This shows that both tails of Y_i are exponentially distributed. In this study, we consider the bivariate case, that is $i = 1, 2$.

Let G denote the Laplace margins and assume $\mathbf{Y} = \{Y_1, \dots, Y_d\}$ are marginally Laplace distributed. If the \mathbf{Y}_j variable exceeds a sufficiently high threshold u , then the Heffernan and Tawn (2004) regression type model is given as:

$$\mathbf{Y}_{-j} = \alpha_{|j} \mathbf{Y}_j + (\mathbf{Y}_j)^{\beta_{|j}} \mathbf{R}_j, \quad (4)$$

where \mathbf{Y}_{-j} represents the vector \mathbf{Y} excluding the j th component, \mathbf{R}_j is a vector of residuals and $(d-1)$ dimensional parameter vectors $\alpha_{|j}$ and $\beta_{|j}$ which satisfy $(\alpha_{|j}, \beta_{|j}) \in [-1, 1]^{d-1} \times (-\infty, 1)^{d-1}$. For more details see Heffernan and Tawn (2004), Keef et al. (2013) and Southworth et al. (2020).

2.2 Estimation of parameters

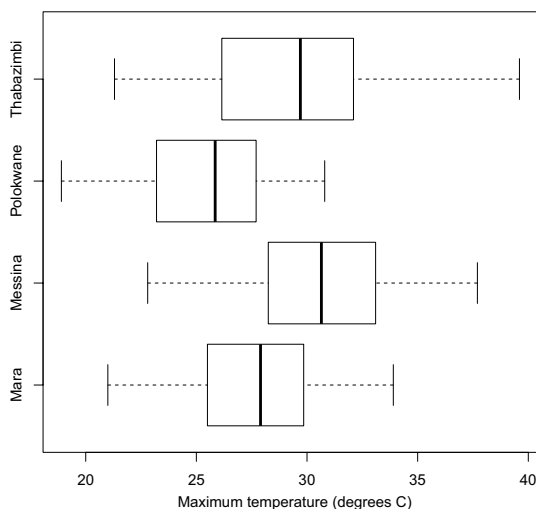
In this study, the parameters of the models are estimated by the maximum likelihood estimation (MLE) method (Heffernan and Tawn 2004; Keef et al. 2013; Southworth et al. 2020). Visual diagnostics are used to check whether the parameter estimates converge to the true population parameters. Within the R software package, we use ‘mexDependence’ to produce the profile-likelihood surface plots which are maximised to estimate the dependence model parameters (Southworth et al. 2020). Once we are satisfied with the model fit, we proceed to examine the estimated model parameters. It should be noted that parameters in the dependence structure are not easily interpretable, although values of a close to 1 (or -1) indicate strong positive (or negative) extremal dependence. According to Southworth et al. (2020), the dependence pairs of variables, say (X_1, X_2) , are described by a pair of parameters (a, b) and the associated empirical distribution of residuals $\mathbf{Z}_{|j}$.

2.3 Data and variables

Time series secondary data are used in this study. All the models considered in the present study have been tested using the monthly maximum temperature series measured in degrees Celsius ($^{\circ}\text{C}$) which were obtained from the South Africa Weather Service (SAWS) database. The data cover four major meteorological stations of the Limpopo province of South Africa, Mara (1949–2018), Messina [or Musina] (1934–2009), Polokwane (1932–2018) and Thabazimbi (1994–2018). The study area is presented in Fig. 1, and the distribution of maximum temperature for each meteorological station is shown in Fig. 2. The statistical software package used for the analysis of the data in this paper is R (see Southworth et al. 2020; Youngman 2020, among others).

As can be noted from the paragraph above, the longest series spans the period 1932–2018 while the shortest series spans the period 1994–2018. Also, one station has a series that terminates much quicker than the rest in 2009, meaning that the years 2010–2018 are missing and will be completely left out of the analysis for this particular station. Therefore, given the differences in the time spans from the four stations and that in the bivariate modelling approach the data variables must span the same time period, the present study considers only the monthly maximum temperature for the period 1994–2009 for all the stations. This means that all the exploratory results in the next section and the findings therein are based on the time period 1994–2009. While this maybe viewed as a weakness or limitation of the multivariate extreme value analysis approach used, the gains from the bivariate threshold approach used and insights from the expected findings on

Fig. 2 Box plots showing the distribution of maximum temperature for each meteorological station



dependence modelling with a time-varying threshold overshadow this limitation of data loss.

3 Empirical results and discussion

This section presents the empirical results of the monthly maximum temperature data for the four stations in the Limpopo province of South Africa: Mara, Messina, Polokwane and Thabazimbi for the period spanning from 1994 to 2009. The results are presented in form of tables and figures. In this section, comprehensive interpretation and discussion of the results are also given.

3.1 Exploratory data analysis

Figure 1 presents the map of the Limpopo province showing the four meteorological stations, while the distribution of maximum temperature for each meteorological station is shown in Fig. 2. The results in Fig. 2 reveal that the maximum temperature series for all the meteorological stations are positively skewed.

Table 1 presents the summary statistics of positive exceedances for monthly maximum temperature data, while Table 2 presents the dependence structure conditioning on each station. The estimates of the parameter a in Table 2 denote strong negative (or positive) extremal dependence for values close to -1 (or 1), respectively. Figures 3, 4, 5, 6 present the time series plots for Mara, Messina, Polokwane and Thabazimbi monthly maximum temperature, respectively, with a time-varying threshold. Figure 7 presents the pairwise scatter plot of the data, where the distribution of each station is shown on the diagonal. At the bottom of the diagonal are the bivariate scatter plots and at the top of the diagonal are the pairwise correlation values. Figure 8 presents the multivariate conditional Spearman's ρ correlation plots, where ρ denotes the Spearman's coefficient of correlation. The

Table 1 Summary statistics of positive exceedances

Statistic	ExcessMapos	ExcessMespos	ExcessPolpos	ExcessThapos
Minimum	0.03	0.02	0.05	0.03
First quartile	0.78	0.74	0.91	1.01
Median	1.54	1.68	1.61	1.97
Mean	1.71	1.85	1.70	2.25
Third quartile	2.53	2.76	2.32	3.18
Maximum	4.12	4.77	4.04	9.63

KEY (The following notation applies throughout this paper):

ExcessMapos represents Mara positive exceedances,

ExcessMespos represents Messina positive exceedances,

ExcessPolpos represents Polokwane positive exceedances,

ExcessThapos represents Thabazimbi positive exceedances

Table 2 Estimates of dependence models

	Dependence parameters	Mara	Messina	Polokwane
Conditioning on Thabazimbi	a	0.2405*	−0.001963	0.2575*
	b	0.2938	−0.128200	0.3147
	Dependence parameters	Mara	Messina	Thabazimbi
Conditioning on Polokwane	a	−0.2902*	−0.3007*	−0.02911
	b	0.0711	0.4061	−0.21600
	Dependence parameters	Mara	Polokwane	Thabazimbi
Conditioning on Messina	a	−0.1475*	−0.59110*	−0.4077*
	b	0.6061	−0.06039	−0.1408
	Dependence parameters	Messina	Polokwane	Thabazimbi
Conditioning on Mara	a	0.4060*	−0.4624*	0.5155*
	b	−0.3634	−0.5239	0.1690

*Denotes significant positive or negative values of a at 5% significance level

results in Fig. 8 suggest a weak positive association in Messina and Polokwane, Messina and Thabazimbi, and Polokwane and Thabazimbi, respectively.

To reduce the amount of diagnostic plots output produced, in this paper, we only show the diagnostic plots for Thabazimbi. Figure 9 presents the marginal diagnostic plots for Thabazimbi. Figure 10 presents the diagnostic plots for conditioning Thabazimbi on Mara, Messina and Polokwane. The plots from top to bottom show; dependence model residuals across the range of the extreme conditioning variable, the absolute

Fig. 3 Mara's monthly maximum temperature with a time-varying threshold

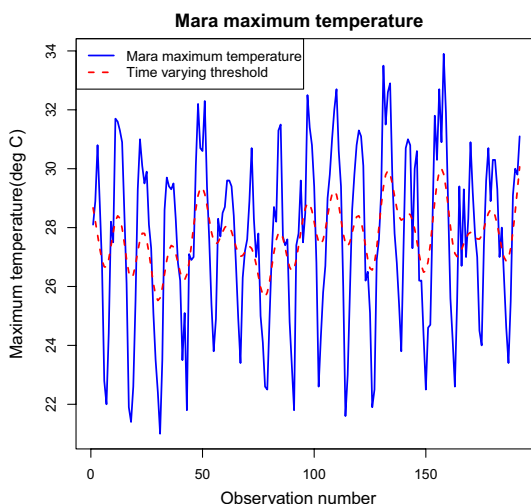
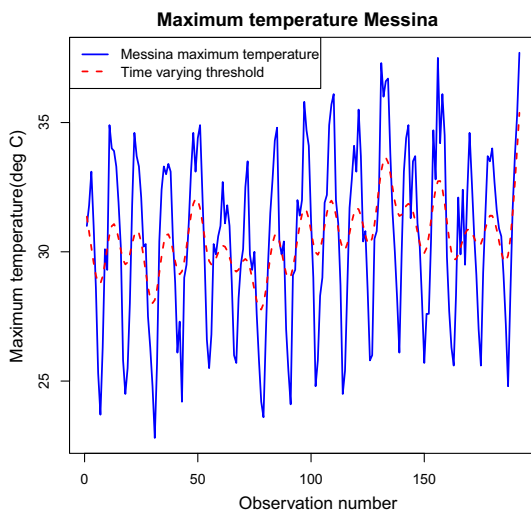


Fig. 4 Messina's monthly maximum temperature with a time-varying threshold



values of the centred and scaled values of the residuals across the range of the extreme conditioning variable and the original untransformed data with contours showing quantiles of the fitted conditional model. The results in Fig. 10 revealed that the parameter estimates are stable at the 75th percentile since the horizontal lines are smoothest.

3.1.1 Discussion of results

The pairwise correlations in Fig. 7 show the extremal dependence amongst the four stations. The results in Fig. 7 reveal a strong positive correlation between Messina and Mara (0.48), a very weak positive correlation between Polokwane and Mara (0.00115) and a

Fig. 5 Polokwane's monthly maximum temperature with a time-varying threshold

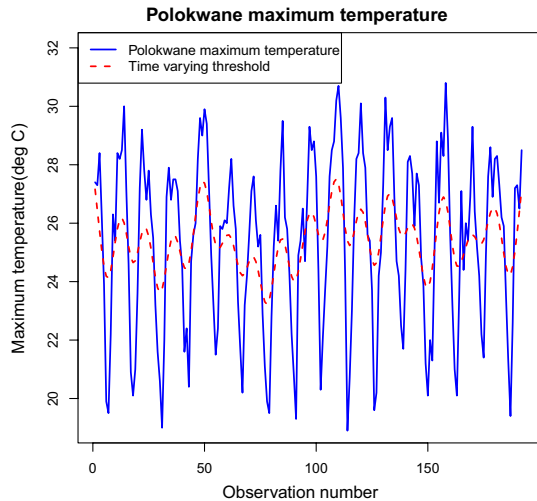
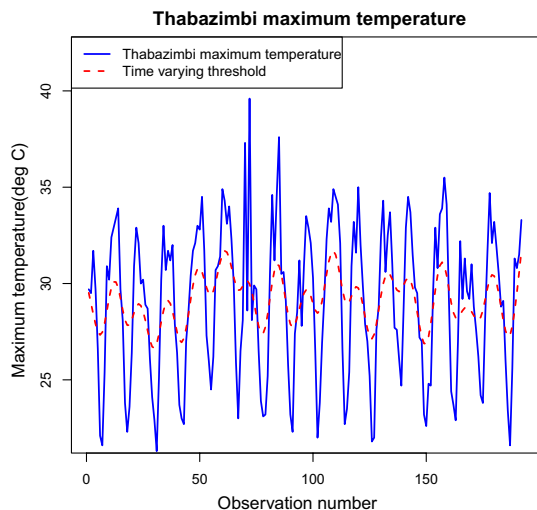


Fig. 6 Thabazimbi's monthly maximum temperature with a time-varying threshold



weak positive correlation between Thabazimbi and Mara (0.234), between Polokwane and Messina (0.144) and between Thabazimbi and Polokwane (0.299). The multivariate conditional Spearman's correlation plots in Fig. 8 support the results of the pairwise correlation in Fig. 7. The marginal diagnostic plots in Fig. 9 for Thabazimbi show that the points are linear, suggesting that the model fit the data well.

The conditional bivariate dependence structure model results are presented in Table 2. Conditioning on Thabazimbi station, the estimates of the dependence parameters for Mara, Messina and Polokwane stations are $a = 0.2405$, $a = -0.001963$ and $a = 0.2575$, respectively. This implies that Mara and Polokwane stations have a positive extremal dependence on high-temperature values at Thabazimbi, while Messina has a negative extremal dependence on high-temperature values at Thabazimbi. Conditioning on Polokwane, the estimates of the dependence parameters for Mara, Messina and Thabazimbi are $a = -0.2902$, $a =$

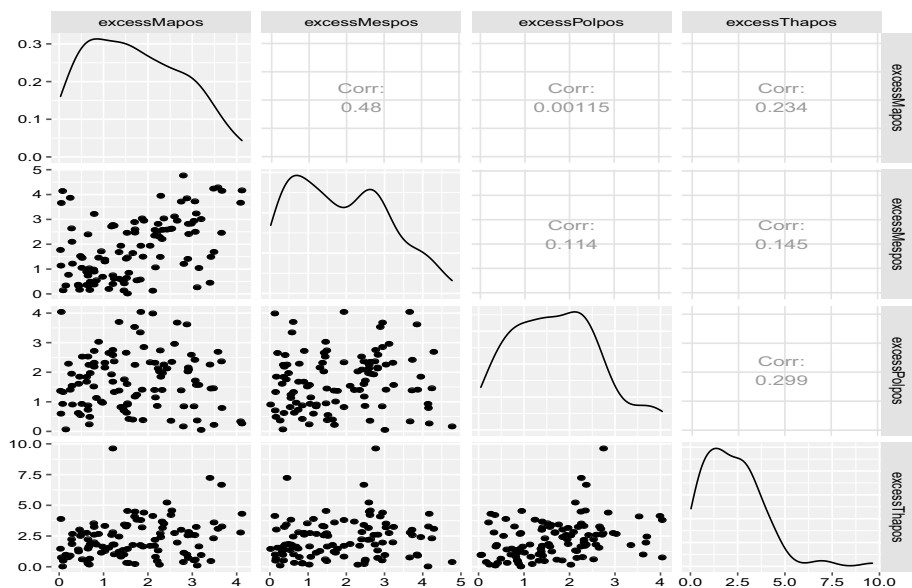


Fig. 7 Pairwise scatterplot and correlation of the data

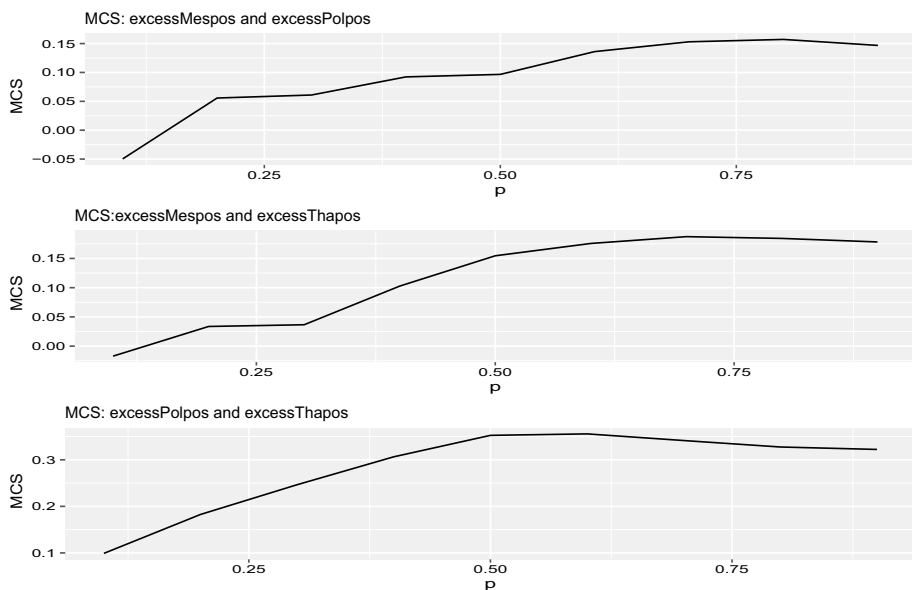


Fig. 8 Multivariate conditional Spearman's correlation

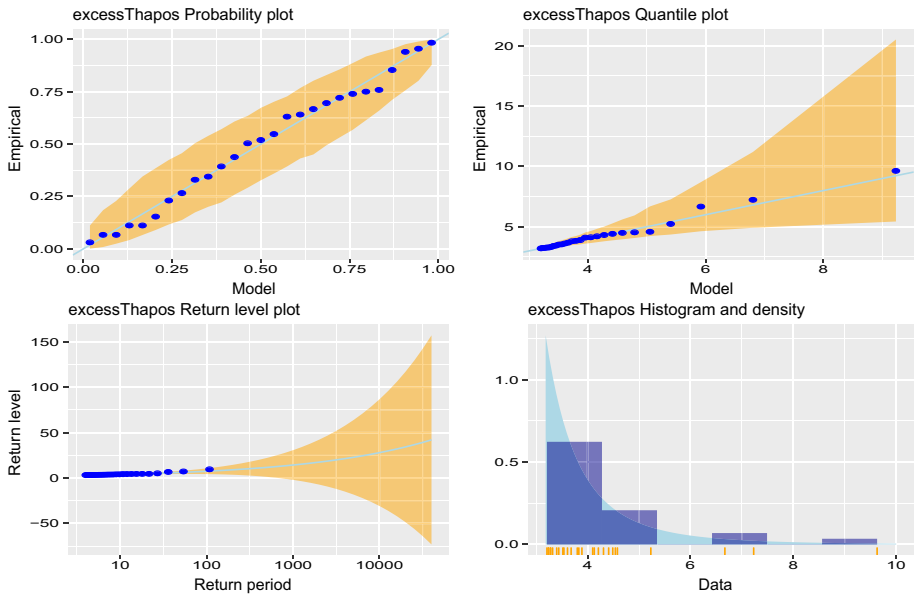


Fig. 9 Marginal diagnostic plots for Thabazimbi

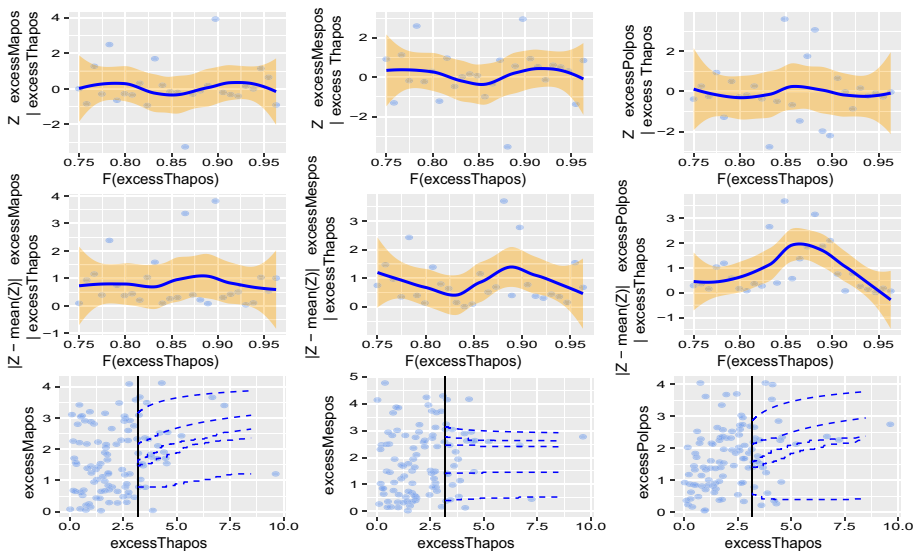


Fig. 10 Diagnostic plots for conditioning on Thabazimbi

-0.3007 and $a = -0.02911$, respectively. This implies that Mara, Messina and Thabazimbi have a significant negative extremal dependence on high-temperature values at Polokwane.

Conditioning on Messina, the estimates of the dependence parameters for Mara, Polokwane and Thabazimbi are $a = -0.1475$, $a = -0.59110$ and $a = -0.4077$, respectively. This

implies that Mara, Polokwane and Thabazimbi have significant negative extremal dependence on high-temperature values at Messina, the strongest negative being that of Polokwane on high-temperature values at Messina. Conditioning on Mara, the estimates of the dependence parameters for Messina and Thabazimbi are $a = 0.4060$ and $a = 0.5155$, respectively, which shows significant positive extremal dependence, while the estimate of the dependence parameter for Polokwane ($a = -0.4624$) shows significant negative extremal dependence on high-temperature values at Mara. These results reveal that Thabazimbi exhibits the strongest positive extremal dependence on high-temperature values at Mara.

Modelling the extremal dependence of extreme temperature in the Limpopo province of South Africa, especially in this era of climate change and global warming, may help reduce the deleterious impacts of maximum temperature extremes in the province. The impacts of maximum temperature extremes in the Limpopo province include heat waves which often result in drought in the agricultural sector. Also, heat waves cause deaths and other health-related problems in the health sector. In addition, heat waves lead to excessive use of air conditioning systems which poses serious challenges to both the economic and energy sectors due to the high demand for electricity. As discussed in the previous section of the background, Limpopo is among the lowest-ranked province in terms of regional GDP per capita in the country, meaning people living in this province cannot afford to have this extra burden resulting from maximum temperature extremes. More details on the deleterious impacts of maximum temperature extremes can be found in Henderson and Muller (1997), Maposa et al. (2017), Nemukula (2018) and Nemukula et al. (2018).

4 Conclusion

Climate change and global warming with its associated extreme temperatures characterised by heat and cold waves pose serious economic and health challenges particularly in developing countries like South Africa where the national electricity supplier, ESKOM, is already battling with meeting the demands of the nation in energy supply leading to occasional rolling blackouts. Increasing temperatures and changing weather conditions affect the heating and cooling systems consequently leading to increased electricity demand for air conditioning in summer and heating systems in winter. Extremely high temperatures also cause drought in the agricultural sector resulting in economic hardships and loss of human lives and livestock. Associated with global warming and climate change are heat wave challenges in the province. The present study has attempted to address these challenges by applying bivariate conditional extremes modelling with a time-varying threshold to Limpopo province monthly maximum temperature series.

Conditional extremes modelling is crucial in studying the dependence structure among several variables. Conditioning on one variable helps to understand the significant positive (or negative) extremal dependence of the remaining variables on the large values of the conditioning variable. This paper presented an application of bivariate threshold excess modelling approach with a positive shift factor as a time-varying threshold to the monthly maximum temperature extremes in four meteorological stations of Limpopo province, South Africa. The study considered Mara, Messina, Polokwane and Thabazimbi meteorological stations in Limpopo province. The findings of the pairwise correlation matrix revealed that Messina, Polokwane and Thabazimbi all have positive correlation on Mara, the strongest being that of Messina and the weakest being that of Polokwane. Other weak

pairwise correlations were also revealed for the pairs Polokwane and Messina, Thabazimbi and Messina, and Thabazimbi and Polokwane.

In addition to pairwise correlation, digression conditioning was performed in the study. Firstly, conditioning on Thabazimbi station, the values of the estimated dependence parameters show that the maximum temperatures of Thabazimbi have a positive extremal dependence on Mara and Polokwane, and a negative extremal dependence on Messina. Secondly, conditioning on Polokwane station, the values of the estimated dependence parameters show that Mara, Messina and Thabazimbi all have negative extremal dependence on Polokwane, the strongest being that of Messina on Polokwane. Thirdly, conditioning on Messina station, the values of the estimated dependence parameters show that Mara, Polokwane and Thabazimbi all have negative extremal dependence on high-temperature values of Messina, the strongest being that of Polokwane on Messina. Lastly, conditioning on Mara station, the values of the estimated dependence parameters show that the maximum temperatures of Mara have positive extremal dependence on Messina and Thabazimbi, and negative extremal dependence on Polokwane. The strongest positive conditional extremal dependence was revealed to be that of Thabazimbi on Mara. This suggests that the occurrence of extreme high temperature in Mara would imply some extreme high temperature at Thabazimbi. This explanation is similar for all meteorological stations with conditional significant positive extremal dependence, while the opposite is true for meteorological stations with significant negative extremal dependence.

The major contribution of this paper is in using a penalised cubic smoothing spline to perform a nonlinear detrending of the temperature data before fitting bivariate threshold excess models based on Laplace margins to positive residuals above the threshold and a positive shift factor as a time-varying threshold to capture the climate change effects in the data. The existing gap in the literature was in combining these two approaches in conditional extremes dependence modelling. The overall significance of this bivariate conditional extremes dependence modelling study lies in quantifying the dependence effects of maximum temperature extremes amongst the various meteorological stations in order to reveal some useful information needed for planning by climatologists, meteorologists, agriculturalists, decision-makers, and planners in the energy sector among others.

Future studies in extremal dependence modelling of temperature extremes may consider exploring the use of extreme value copulas, although this approach demand paying special attention to asymptotic independence. With the increased availability of meteorological stations, another important consideration in the future will be an investigation of spatio-temporal dependence between temperature extremes using the conditional extremes model (Dutfoy et al. 2014; Heffernan and Tawn 2004).

Acknowledgements The authors would like to acknowledge South African Weather Service (SAWS) for providing the data that was used for the analysis in this paper. We are also greatly indebted to the National Research Foundation (NRF) Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS) who provided funds for postgraduate studies. Lastly, we thank the University of Limpopo, University of Venda and University of Alabama for their support in research.

Author contributions DM, CS and AMS conceptualised the research. AMS and DM obtained the data and cleaned it. DM, CS, AMS and JJC analysed the data and drafted the original manuscript. All authors reviewed the manuscript. All authors read and approved the final manuscript. All authors approved the final manuscript and agreed to submit it for publication.

Funding Not applicable. The research was not financially funded by any organisation or institution.

Data availability The data are readily available upon request from South African Weather Service (SAWS).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.



Code availability Codes that are not provided in the Appendices can be provided upon request from the corresponding author.

References

- Acero FJ, Garcia AG, Gellego MC, Parey S, Dacunha-Castelle D (2014) Trends in summer extreme temperatures over the Iberian Peninsula using nonurban station data. *J Geophys Res: atmos* 119:1–15. <https://doi.org/10.1002/2013JD020590>
- Bhagwandin L (2013) Multivariate extreme value theory with an application to climate data in the Western Cape province. MSc dissertation, University of Cape Town (2017)
- Chikobvu D, Sigauke C (2013) Modelling influence of temperature on daily peak electricity demand in South Africa. *J Energy South Africa* 24(4):63–70
- Coles S (2001) An introduction to statistical modelling of extreme values. Springer-Verlag, London
- Coumou D, Robinson A (2013) Historic and future increase in the global land area affected by monthly heat extremes. *Environ Res Lett* 8(3):034018. <https://doi.org/10.1088/1748-9326/8/3/034018>
- Coumou D, Robinson A, Rahmstorf S (2013) Global increase in record breaking monthly-mean temperatures. *Clim Change* 118(3–4):771–782
- DEA (2019) National climate change adaptation strategy. Department of Environmental Affairs (DEA) Report, Pretoria
- Diriba TA, Debusho LK (2020) Modelling dependency effect to extreme value distributions with application to extreme wind speed at Port Elizabeth, a frequentist and Bayesian approaches. *Comput Stat* 35:1449–1479. <https://doi.org/10.1007/s00180-019-00947-2>
- Diriba TA, Debusho LK, Botai J (2015) Modeling extreme daily temperature using generalized Pareto distribution at Port Elizabeth. *Annu Proc South African Stat Assoc (SASA)* 1:41–48
- Dutfoy A, Parey S, Roche R (2014) Multivariate extreme value theory - a tutorial with applications to hydrology and meteorology. *Depend Model* 2:30–48
- Ferreira A, de Haan L (2015) On the block maxima method in extreme value theory: PWM estimators. *Annal Stat* 43(1):276–298. <https://doi.org/10.1214/14-AOS1280>
- Gebrechorkos SH, Hulsman S, Bernhofer C (2019) Changes in temperature and precipitation extremes in Ethiopia, Kenya and Tanzania. *Int J Climatol* 39(1):18–30
- Heffernan JE, Tawn JA (2004) A conditional approach for multivariate extreme values. *J R Stat Soc, B* 66(3):497–546
- Henderson KG, Muller RA (1997) Extreme temperature days in South Central United States. *Clim Res* 8(2):151–162
- Hohne PA, Kusakana K, Numbi BP (2019) A review of water heating technologies: an application to the South African context. *Energy Rep* 5:1–19
- Keef C, Papastathopoulos I, Tawn JA (2013) Estimation of the conditional distribution of a multivariate variable given that one of its components is large. additional constraints for the Heffernan and Tawn model. *J Multivar Anal* 115:396–404
- Keelings D, Waylen P (2015) Investigating teleconnection drivers of bivariate heat waves in Florida using extreme value analysis. *Clim Dyn* 44(11–12):3383–3391
- Kruger AC, Sekele SS (2013) Trends in extreme temperature indices in South Africa: 1962–2009. *Int J Clim* 33(3):661–676
- Kruger AC, Shongwe S (2004) Temperature trends in South Africa: 1960–2003. *Int J Climatol: A J R Meteorol Soc* 24(15):1929–1945
- Makate C, Makate M, Mango N, Siziba S (2019) Increasing resilience of smallholder farmers to climate change through multiple adoption of proven climate-smart agriculture innovations: Lessons from Southern Africa. *J Environ Manag* 231:858–868
- Maponya P, Mphandeli S (2012) Climate change and agriculture production in South Africa: impacts and adaptation options. *J Agric Sci* 4(10):48
- Maponya P, Mphandeli S (2012) Impact of drought on food scarcity in Limpopo province, South Africa. *Afr J Agric Res* 7(37):5270–5277
- Maposa D, Cochran JJ, Lesaona M (2017) Modelling extreme flood heights in the lower Limpopo River basin of Mozambique using a time-heterogeneous generalised Pareto distribution. *Stat Interface* 10:131–144
- Mbokodo IL (2017) Heat waves in South Africa: Observed variability structure and trends. MSc dissertation, University of Venda

- Minkah R, de Wet T (2014) Constant versus covariate dependent threshold in the peaks-over threshold method. *J Appl Probab Stat* 9(1):64
- Mpandeli NS, Maponya PI (2013) Coping with climate variability in Limpopo province, South Africa. *Peak J Agric Sci* 4(1):54–64
- Nemukula MM (2018) Modelling temperature in South Africa using extreme value theory. MSc dissertation, University of Witwatersrand
- Nemukula MM, Sigauke C and Maposa D (2018) Bivariate threshold excess models with application to extreme high temperatures in Limpopo province of South Africa. *Proceedings of the 60th Annual Conference of SASA, South Africa*, 33–40
- Nhamo L, Matchaya G, Mabhaudhi T, Nhlengethwa S, Nhemachena C, Mpandeli S (2019) Cereal production trends under climate change: Impacts and adaptation strategies in Southern Africa. *Agriculture* 9(2):30
- Ochanda OO (2016) Time series analysis and forecasting of monthly air temperature changes in Nairobi, Kenya. MSc dissertation, University of Nairobi
- Pereira L (2017) Climate change impacts on agriculture across Africa, Oxford research encyclopedia of environmental science. Oxford University Press, Oxford
- Phophi MM, Mafongonya P, Lottering S (2020) Perceptions of climate change and drivers of insect pest outbreaks in vegetable crops in Limpopo province of South Africa. *Climate* 8(2):27
- Raggad B (2018) Statistical assessment of changes in extreme maximum temperatures over Saudi Arabia, 1985–2014. *Theor Appl Climatol* 132(3–4):1217–1235
- Raghavendra A, Dai A, Milrad SM, Cloutier-Bisbee SR (2019) Floridian heatwaves and extreme precipitation: future climate projection. *Clim Dyn* 52(1–2):495–508
- Reddy CL, Vincent K (2017) Climate risk and vulnerability. a handbook for Southern Africa, 2nd edn. CSIR, South Africa
- Sigauke C, Bere A (2017) Modelling non-stationary time series using peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand. *Energy* 119:152–166
- Sigauke C, Nemukula MM (2018) Modelling extreme peak electricity demand during a heatwave period: A case study. *Energy Syst*, pp 1–23
- Southworth H, Heffernan JE, Metcalfe PD (2020) Texmex: Statistical modelling of extreme values. R package version 2.4.4
- Thompson C, Cai Y, Reeve D, Stander J (2009) Automated threshold selection methods for extreme wave analysis. *Coast Eng* 56:1013–1021
- Tilloy A, Mulamud BD, Winter H, Joly-Laugel A (2020) Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios. *Nat Haz Earth Syst Sci* 20:2091–2117. <https://doi.org/10.5194/nhess-20-2091-2020>
- Toros H, Mokari M and Abbasnia M (2019) Regional variability of temperature extremes in the maritime climate of Turkey: A case study to develop agricultural adaptation strategies under climate change. *Modeling Earth Systems and Environment*, 1–9
- Tshiala M, Olwoch JM (2010) Impact of climate variability on tomato production in Limpopo province, South Africa. *J Agric Res* 5(21):2945–2951
- Verster A, Raubenheimer L (2020) A different approach for choosing a threshold in peaks over threshold. *arXiv:2006.05748*, [stat.ME]
- Winter H (2016) Extreme value modelling of heatwaves. PhD Thesis, Lancaster University
- Wolf J, Adger WN, Lorenzoni I (2010) Heat waves and cold spells: an analysis of policy response and perception of vulnerable population in the UK. *Environ Plann* 42(11):2721–2734
- Worku G, Tereri E, Bantider A, Dile YT (2019) Observed changes in extremes of daily rainfall and temperature in Jemma Sub-Basin, Upper Blue Nile Basin Ethiopia. *Theor Appl Climatol* 135(3–4):839–854
- Wright CY, Garland RM, Norval M, Vogel C (2014) Human health impacts in a changing South African climate. *South Afr Med J* 104(8):579–582
- Yamba FD, Walimwipi H, Jain S, Zhou P, Cuamba B, Mzezewa C (2011) Climate change/variability implications on hydroelectricity generation in the Zambezi River Basin. *mitigation and adaptation. Mitiga Adapt Strateg Global Change* 16(6):617–628
- Youngman BD (2020) Evgam: An R package for generalized additive extreme value models. *Journal of Statistical Software*, VV(II): *arXiv:2003.04067* (stat). <https://doi.org/10.18637/jss.v000.i00>

Authors and Affiliations

Daniel Maposa¹  · Anna M. Seimela¹ · Caston Sigauke²  · James J. Cochran³

¹ Department of Statistics and Operations Research, University of Limpopo, Private Bag X1106, Sovenga Polokwane, South Africa

² Department of Statistics, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa

³ Department of Information Systems, Statistics and Management Science, University of Alabama, Tuscaloosa, Alabama, USA