TPM Vol. 32, No. S2, 2025
ISSN: 1972-6325
https://www.tpmap.org/

EXTREME PROBABILITY FORECASTING ALGORITHM: CORONAVIRUS SPREAD IN SOUTH AFRICA

DITEBOHO XABA[1], KATLEHO MAKATJANE[2] AND CLARIS SHOKO[3]
UNIVERSITY OF SOUTH AFRICA[1],PRETORIA, SOUTH AFRICA
UNIVERSITY OF BOTSWANA[2,3] ,GABORONE,BOTSWANA
xabald@unisa.ac.za; makatjanek@ub.ac.bw; shokoc@ub.ac.bw
CORRESPONDING AUTHOR: xabald@unisa.ac.za

Abstract- The spread of coronavirus in South Africa was characterised by extreme temporal dynamics, resulting in uncertainties in estimation and forecasting. A probabilistic approach can effectively address these uncertainties. We propose a combined forecasting model that integrates a Markov-switching autoregressive process with a truncated generalised extreme value (TGEV) distribution. Exploratory data analysis indicates that daily confirmed cases exhibit a fat-tailed, non-normal distribution characterised by notable regime shifts. The proposed MS-AR-TGEV model demonstrates superior predictive accuracy as indicated by performance metrics, including root mean square error (RMSE), mean absolute percentage error (MAPE), and continuous ranked probability score (CRPS); resulting in an effective model for forecasting COVID-19-related uncertainties in South Africa. The findings corroborate earlier research indicating that hybrid models demonstrate enhanced efficacy in forecasting time series that is characterised by intricate non-linear dynamics. The proposed model aids health sector personnel and government in planning and forecasting future epidemics that exhibit behaviour analogous to COVID-19.
Keywords*: Coronavirus, Ensemble Model Output Statistics, Forecasting, GeneralisedExtreme Value distribution, Markov switching model*

1.INTRODUCTION

The year 2020 will be remembered as a catastrophic period for humanity on Earth. In December 2019, a new type of coronavirus (2019-nCoV) was reported to be responsible for a pneumonia outbreak of unknown cause in Wuhan, Hubei province of China (Huang et al., 2020). The first death was reported on 10th January 2020, and it soon turned into a global pandemic (Sohrabi et al., 2020) affecting millions of people worldwide. The World Health Organization (WHO) confirmed that the virus belonged to the coronavirus family, which led countries to employ an array of measures to protect the health of their people; and these measures ranged from travel ban, quarantine, event cancellations and postponements, social distancing, mass testing, strict and moderate lockdowns (Acter et al., 2020). The economic and social repercussions of this virus were significantly more severe than the loss of life, particularly in developing and underdeveloped nations, and the potential impact of this virus on the African continent was alarming. The recommendations and practices regarding the use of public face masks during the ongoing coronavirus pandemic have varied significantly and were subject to rapid changes. The public's use of masks in public spaces has been a subject of controversy, particularly since April 3, 2020; and, this use of masks is significantly more common in various Asian countries, which have historically managed to limit the spread of the 2003 SARS epidemic within communities (Hung, 2003). Additionally, extensive mask usage is a key characteristic of the relatively effective response to the coronavirus (Young et al. 2020).

Nontheless, this study examines the extreme temporal dynamics of the coronavirus epidemic in South Africa (SA) in the time range of 05 March 2020 to 01 March 2023. An early look at basic day-lag maps shows that the way the epidemic spreads is similar in different places, suggesting that simple models can be useful for understanding how the epidemic spreads, particularly when it comes to the peak number of confirmed infections and when it happens. Therefore, we consider this modelling to be probabilistic, leading to uncertainty in forecasting, prediction, or estimation. When researchers create forecasts for an uncertain future, they must evaluate the associated weaknesses to help decision-makers understand these vulnerabilities.

The natural weaknesses in forecasts point to an ideal world where they should be probabilistic. That is to say, they should be probability distributions across future occurrences or amounts (Gneiting and others, 2014). Probabilistic predictions might seem like density forecasts measuring prediction uncertainties, prediction intervals, or quantiles. They represent a straightforward correction to the ideal dynamic as proposed by Gneiting and others (2014). But, for planning reasons, it is very crucial to assess vulnerabilities surrounding demand projections to prevent the construction of unneeded infrastructure and to guarantee that future demand is satisfied to help to subdue the coronavirus. Tay and Wallis (2000) described density forecasts for predicting a random variable's value in the future as assessments of the likelihood of different possible future values for that variable. It is crucial to evaluate the suitability of an alternative coronavirus source in a particular region before making investments; hence one of the objective of this study is to develop an effective forecasting model for newly reported coronavirus cases and to quantify extreme quantiles of coronavirus spread in South Africa. We assimilate a Markov-switching autoregressive (MS-AR) model with a truncated, generalised extreme value distribution (TGEV) and employ the ensemble method introduced by Baran et al. (2021) to accomplish this objective. This approach facilitates the acquisition of reliable estimates and addresses issues of uncertainty. The coronavirus disease has been designated a pandemic by the World Health Organisation, which represents a significant global public health crisis. The combination of a Markov switching model with a truncated, generalised extreme value distribution facilitates the identification of coronavirus trends as the regime shifts. We subsequently utilise the ensemble model output statistics (EMOS) to deliver a comprehensive predictive distribution of the analysed coronavirus trends by calibrating the ensemble forecasts of coronavirus spreads (waves). We therefore, fit a generalised extreme value distribution truncated at zero as the distribution of the predictions. This truncation resolves one of the crucial drawbacks of GEV-based EMOS models, which sometimes may lead to negative predictions for the COVID-19 spread (waves) and somewhat retains the attractive properties of the original distribution. But in a nutshesll, the spread of the virus is deemed positive hence truncation approach of the GEV to avoid misleading results. With this, we seek to contribute to the statistical and public health literature by applying a novel charter that combines EMOS with the Markov-switching autoregressive model, together with the truncated generalised extreme value distribution to model and predict the outbreak of coronavirus for South Africa probabilistically. The MS-AR model is important for detecting and describing regime-dependent phenomena for the coronavirus, as it can extract changes in different states. Since this model permits parameters to fluctuate among distinct hidden states, it is supposedly ideal for cases when the dynamics of a system undergo sudden changes (Makatjane and Xaba, 2016). Through this lens, it is possible, for instance, to differentiate between peak and trough transmission times of coronavirus. By adapting to these variations, the MS-AR model becomes more adaptable and accurate in predictions, unlike autoregressive (AR) models that exclusively rely on unchanging parameters. As a bonus, its probabilistic structure sheds light on the likelihood of regime transitions, which helps with public health crisis prediction and action timing. Finally, we can more accurately portray the heavy-tailed character of coronavirus case distributions by using the truncated generalised extreme value distribution, which improves the model's capacity to handle extreme values in the data. The truncated distribution is helpful because it can handle the unevenness and sharp peaks often found in high frequency data such as daiyl confirmed coronavirus spread cases, and by cutting off the distribution at zero, it makes sure that the predictions are always non-negative. Because it accounts for changing pandemic dynamics, the non-stationary TGEV version allows parameters to change over time, which enhances predicting performance even more. This enhances its usefulness in predicting the danger levels and return durations of future epidemic peaks. Finally, a full range of possible outcomes is created by fine-tuning the initial group predictions with the ensemble model output statistics, which is a powerful method for analysing data after it has been collected. Unlike set predictions, EMOS changes the ensemble outputs using statistical distributions like TGEV to account for uncertainty and accuracy in the predictions. It enables the model to provide trustworthy prediction intervals, which aid stakeholders in evaluating both the anticipated number of instances and the corresponding confidence levels. When used with predictions from multiple models, EMOS reduces errors, considers differences between models, and improves accuracy, all of which help make better forecasts. Forecasts are guaranteed to be accurate, well-calibrated, and probabilistically informative when EMOS is combined with MS-AR and TGEV.

2.LITERATURE REVIEW

Literature related to the coronavirus pandemic and the consequences thereof is reviewed. The world fights against this crisis, and no one is certain about the future consequences and impact they will have. Lu et al. (2020) evaluated responses of Asian countries to identify key factors that played a role in their effective coronavirus management. The results highlighted various measures such as early detection, rapid testing, contact tracing, strict quarantine protocols, public health campaigns, and effective communication strategies. Kawohl and Nordt (2020) conducted statistical analyses to investigate the association between COVID-19, unemployment, and suicide. The results of these authors suggest that there is a complex and multifaceted relationship between the three associated factors.

A Bayesian structural time series models were also executed by Xie (2022) to capture the dynamics of the pandemic and incorporate various factors such as government interventions, population characteristics, and testing capacity. The results of the study show underlying patterns of coronavirus transmission and accurate forecasts of future case numbers. Rossouw et al. (2021), on the other hand, used a Markov switching dynamic regression (MSDR) model to analyse data from various sources, including surveys and economic indicators, to identify distinct regimes or states of happiness and investigate the transitions between these states. The results reveal significant shifts in happiness levels before and during the pandemic, indicating the influence of coronavirus on individuals' subjective experiences. de-Oliveira et al. (2021) employed GAM (herein referenced generalised additive model) functions to capture the nonlinear relationship between predictor variables and pandemic outcomes, allowing for a more accurate assessment of control measures. Furthermore, Markov-switching models were applied to identify distinct regimes or states of pandemic control and examine the transistion between these states over time. In addition, Al-Zoughool et al. (2022) used the stochastic continuous-time Markov chain model to simulate various lockdown scenarios and evaluate their impact on the spread of coronavirus and associated outcomes. The results of these authors suggest that the timing and duration of lockdown measures significantly influence the number of infections and the burden on the healthcare system. Lee et al. (2021) alos utilised the Sparse HP filter to contact rate data to identify abrupt changes that are not captured by conventional smoothing techniques. Somyanonthanakul et al. (2022) thereafter utilised time series modelling to capture the temporal dependencies and trends in coronavirus data, allowing for accurate forecasting. Their study provided factors that contribute to the spread of coronavirusand aid in developing effective forecasting models. Qu et al. (2022) also combined statistical modelling techniques with intelligent systems to develop a comprehensive structure for forecasting coronavirus outcomes. The results of this study provided valuable insights into the relationships between environmental factors and coronavirus outcomes, enabling better predictions of new cases and deaths. Douwes-Schultz et al. (2023) employed a Markov switching model to classify the outbreak into three distinct states, capturing different phases of the pandemic. The analysis is based on hospital admission data, which provides insights into the severity and spread of the virus. The results of these authors shed light on the transitions between these states and provide valuable information on the patterns of coronavirusoutbreaks and their impact on healthcare resources. Haimerl and Hartl (2023) employed regime switching model to capture the underlying factors driving changes in infection rates, such as policy interventions, behavioural changes, and the impact of new variants. The results provide insights into the different states of the pandemic and the effectiveness of control measures in containing the spread of the virus. Finally, He et al. (2020) utilised the Susceptible-Exposed-Infectious-Recovered (SEIR) model to analyse the progression and dynamics of COVID-19. The study analysed interactions among susceptible, exposed, infectious, and recovered individuals to elucidate the dynamics of the pandemic, encompassing infection rates, the efficacy of control measures, and the effects of interventions.

2.1 Research Highlights and Key Findings

This paper offers a unique forecasting system combining a truncated generalised extreme value distribution and EMOS with a Markov-switching Autoregressive model to probabilistically forecast the spread of coronavirus in South Africa. The model addresses key limitations in time series forecasting by including different behaviours during various phases of the outbreak, like low and high transmission periods, and by modelling the distribution of daily confirmed cases that have extreme values but cannot be negative. Importantly, the research measures uncertainty in both mild and severe outbreak situations; hence, it improves epidemic preparation and public health decision-making. The inclusion of the TGEV distribution lets one realistically estimate extreme values—e.g., epidemic peaks—without unreasonable negative projections. The EMOS, on the other hand, improves the calibration of predictive distributions, thereby strengthening and increasing the accuracy of the predictions. Table 1, below, summarises the highlights and results.

Table 1: Performance and Contributions of the MS-AR–TGEV–EMOS Model

Characteristic	Highlights	Findings	Contributions
Regime Shift Detection	Captures distinct epidemic phases via MS-AR	Identifies significant shifts in virus transmission	Enables early detection of transitions between low and high case periods
Tail Event Forecasting	TGEV models extremes while truncating negative values	Peaks in cases are accurately predicted	Improves accuracy and realism in extreme outbreak forecasting
Model Calibration	EMOS refines probabilistic forecasts	Enhances the sharpness and reliability of prediction intervals	Ensures well-calibrated forecasts for public health planning
Statistical Validation	High p-values in goodness-of-fit tests (CRPS, RMSE)	Demonstrates model robustness and predictive skill	Confirms the statistical reliability of forecasts
Asymmetry in Spread	Shape parameters show fat tails and left-skewness	Confirms severity of surges over declines	Provides risk insight into rapid outbreak escalation
Policy Implication Support	Scenario-based forecasts derived from the MS–AR-TGEV model	Helps plan interventions during volatile periods	Supports data-driven decisions in health and emergency response

2 Methods and Procedures

Our World in Data (https://www.ourworldindata/coronavirus-source-data) reports a daily number of confirmed coronavirus cases that are accessible to the public. We utilise this data from March 5, 2020, to October 18, 2023. We get data from the daily number of confirmed cases in South Africa.

2.1 Trend Test

We use the non-parametric Mann-Kendall (M-K) test statistic, Sen's slope estimator, and time series plots to analyse the long-term trend and variability of the daily spread of coronavirus in South Africa. According to Wi et al 2016), this test is commonly used, and the test statistic is defined as

$$S = \sum_{j=1}^{n-1} \sum_{j+1}^{n} \operatorname{sgn}\left(\delta_i - \delta_j\right), \tag{1.1}$$

where is the number of extreme values. If S is positive, then there is an increasing trend, but if S is negative, then there is a decreasing trend, and $\operatorname{sgn}\left(\delta_i - \delta_j\right)$ is a sign function given by

$$\operatorname{sgn}\left(\delta_i - \delta_j\right) = \begin{cases} 1, \delta_i - \delta_j > 0 \\ 0, \delta_i - \delta_j = 0 \\ -1, \delta_i - \delta_j < 0 \end{cases}. \tag{1.2}$$

Under the null hypothesis of no trend, the theoretical mean of S is 0, and its variance is given by

$$Var(S) = \left[n(n-1)(2n+5 - \sum_{p=1}^{g} t_p \left(t_p - 1\right)\left(2t_p + 5\right) \right] / 18 \tag{1.3}$$

Where g is the number of tied groups t_p and is the number of data points in the tied group.

2.2 Sen's Slope Estimator

The Sen's slope nonparametric estimator method is used to evaluate the trend of the time series data. The slope of data pairs can be initially estimated by using

$$\beta_i = Median\left[\frac{X_j - X_k}{j - k} \right] \forall k < j. \tag{1.4}$$

In Equation (1.4), X_j and X_k are the values of a time series at time j and k respectively.Whiletime j is after time $k\left(k - j\right)$. The median of values of β_i is the Sen's slope estimator test. A negative β_i value represents a decreasing trend, a positive β_i value represents anincreasing trend over time.

2.3 Markov Switching Autoregressive Models

A special class of these Markov switching models (MSM) is the Markov Switching Autoregressive model. Given a time series $\{X_t : t = 1,2,3,...,n\}$, an MS-AR model provides an approximation to the system representation in the form (Hamilton, 2010)

$$X_t = \alpha_0^{(S_t)} + \alpha_1^{(S_t)} X_{t-1} + \cdots + \alpha_p^{(S_t)} X_{t-p} + \sigma^{(S_t)} \varepsilon_t \tag{1.5}$$

where, $\left(\alpha_0^S\right), \left(\alpha_1^S\right), \cdots, \left(\alpha_9^S\right), \left(\sigma^s\right) \in \Re^{p+1} \times (0,\infty)$ signifies the AR(p) model's unknown parameters that define how the observable process changes in the regime $S \in \{1, \cdots M\}$ while, ε_t is a series of independent and identically distributed Gaussian variables with a mean of zero and a variance of one that is not affected by the Markov chain S_t .To be more precise, we assume two states modelling and hence the underlying MS-AR (p) model is given by

$$X_t = \begin{cases} c_1 + \sum_{i=1}^{p} \phi_{1,i} X_{t-i} + \varepsilon_{1,t}, S_t = 1 \\ c_2 + \sum_{i=1}^{p} \phi_{2,i} X_{t-i} + \varepsilon_{2,t}, S_t = 2 \end{cases} \tag{1.6}$$

where the transition matrix isgiven by

$$\mathrm{P} = \begin{pmatrix} \mathrm{p}_{11} & \mathrm{p}_{21} \\ \mathrm{p}_{12} & \mathrm{p}_{22} \end{pmatrix}.$$

2.4 Proposed Non-stationary GEV-based EMOS Model

We are now looking at the ensemble model output statistics (EMOS) model of Lerch and Thorarinsdottir (2013), which is based on a GEV distribution, instead of the Truncated Normal EMOS method. The main change from the stationary GEV distribution is the inclusion of varying scale and shape parameters with time or maybe with tother factors Masingi and Maposa, 2021; Syafrina et al, 2019). So, we fit a non-stationary GEV distribution to the residuals from regime 1 of Equation (1.6). This distribution is defined as

$$GEVD\left(x;\mu(t),\sigma(t),\xi\right) = \exp -\left[1 + \xi \frac{x-\mu(t)}{\sigma(t)}\right]^{-\frac{1}{\xi}}, \xi \neq 0. \tag{1.7}$$

In the simplest case, the following regression structures could be examined for the location and scale parameters

$$\begin{aligned} \mu(t) &= \mu_0 + \mu_1 t + \mu_2 t^2 \\ \sigma(t) &= \exp\left(\sigma_0 + \sigma_1 t + \sigma_2 t^2\right), \\ \xi(t) &= \xi \end{aligned} \tag{1.8}$$

enabling the form parameter to stay the same and the time to change in a way that is quadratic (Panagoulia et al. (2014). For$x \geq 0$, the cumulative density function (CDF) for this truncated GEV (TGEV) distributionis given by

$$TGEV\left(\mu,\sigma,\xi\right) = \begin{cases} \dfrac{G\left(x \mid \mu,\sigma,\xi\right) - G\left(0 \mid \mu_0,\sigma_0,\xi\right)}{1 - G\left(0 \mid \mu_0,\sigma_0,\xi\right)}, G\left(0 \mid \mu_0,\sigma_0,\xi\right) < 1; \\ 1, G\left(0 \mid \mu_0,\sigma_0,\xi\right) = 1 \end{cases} \tag{1.9}$$

where the negative values are obviously excluded from the support set of the TGEV distribution. For $\xi < 1$ and $G\left(0 \mid \mu_0,\sigma_0,\xi\right) < 1$, the $TGEV\left(\mu,\sigma,\xi\right)$ distribution has a finite mean of

$$\begin{cases} \mu+\left(\Gamma\left(1-\xi\right)-1\right)^{\sigma/\xi}, \xi\neq 0, \xi\mu-\sigma>0; \\ \sigma-\sigma/\xi+\dfrac{\sigma\left(\Gamma_\ell\left(1-\xi,\left[1-\varsigma\mu/\sigma\right]^{-1/\xi}\right)\right)/\xi}{1-\exp\left(1\left[1-\xi\mu/\sigma\right]^{-1/\xi}\right)}, \xi\neq 0, \xi\mu-\sigma\leq 0; \\ \dfrac{\mu+\sigma\left(C-Ei\left(-\exp\left[\mu/\sigma\right]\right)\right)}{1-\exp\left(-\exp\left[\mu/\sigma\right]\right)}, \xi=0, \end{cases} \quad (1.10)$$

where Γ and Γ_ℓ denote the gamma and the lower incomplete gamma function defined by

$$\Gamma\left(\alpha\right)=\int_0^\infty t^{\alpha-2}e^{-t}dt \text{ and } \Gamma_\ell\left(\alpha,x\right)=\int_0^x t^{\alpha-2}e^{-t}dt,$$

and $E_i\left(x\right)$ is the exponential integral computed by $Ei\left(x\right)=\int_{-\infty}^{x}\frac{e^t}{t}d=C+\ln\left|x\right|+\sum_{k=1}^{\infty}\frac{x^k}{k!k}$ with C being the Euler–Mascheroni constant.

2.5 Training Data Selection and Verification Scores

This text examines the estimation of the score function as outlined by Gneiting and Raftery (2007). The optimal score principle entails the optimisation of an appropriate scoring method applied to a carefully chosen training dataset, which facilitates the estimation of unknown parameters in EMOS models. The standard EMOS modelling approach employs rolling training periods, whereby model parameters are computed for a designated date using ensemble predictions and corresponding validation observations from the prior calendar days. Considering the length of the training period, two classical methods exist for selecting the geographical distribution of the training data (Thorarinsdottir and Gneiting, 2010). The global (regional) method estimates a single set of parameters for the whole ensemble domain by using ensemble forecasts and observations from all health facilities in South Africa during the training period. The local estimate, using solely the training data from the designated station, yields varying parameter estimates for numerous centres (Lerch and Baran, 2017). The logarithmic score (logS) (Good, 1952) and the continuous ranked probability score (CRPS) (see, for instance, Wilks, 2011) represent the two most prevalent scoring systems.The first one is the negative logarithm of the predictive probability density function (PDF) calculated at the actual observation; the second one, for a predictive cumulative distribution function (CDF) F and a real value (actual observation) x, is defined as

$$CRPS\left(F,x\right)=\int_{-\infty}^{\infty}\left[F\left(y\right)-I_{\{y\geq x\}}\right]^2dy=E\left|X-x\right|-\frac{1}{2}E\left|X-x^{'}\right|, \quad (1.11)$$

where X and $X^{'}$ are independent random variables that follow F and have a finite initial moment, and I_H is the indicator function of set H .It is important to note that LogS and CRPS are negatively oriented scores; meaning that lower values indicate superior model performance. The CRPS can now be expressed in closed form, which enhances the efficiency of optimisation techniques. For TN, LN, and GEV laws, the studydirects the reader to the work of Friederichs and Thorarinsdottir (2012) for additional information. The Conditional Risk Premium Score (CRPS) of a Truncated Generalised Extreme Value (TGEV) distribution, which comes from a Generalised Extreme Value (GEV) cumulative distribution function (CDF), is the same as

$$CRPS\left(G_0,x\right)=\left(2G_0\left(x\right)-1\right)\left(x-\mu+\sigma/\xi\right)+\sigma/\xi\left(1-G\left(0\right)\right)^2\left[-2^\xi\Gamma_\ell\left(1-\xi,-2\ln G\left(0\right)\right)\right]$$
$$+2G\left(0\right)\Gamma_\ell\left(1-\xi,-\ln G\left(0\right)\right)+2\left(1-G\left(0\right)\right)\Gamma_\ell\left(1-\xi,\ln G\left(x\right)\right)$$

TPM Vol. 32, No. S2, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

for $\neq \xi 0$. To compare the predictive performance of the EMOS models for high coronavirus spread in South Africa, we consider a threshold-weighted continuous ranked probability score (twCRPS) of Gneiting (2011) which is given in Equation (1.12) as

$$twCRPS(F,x)=\int_{-\infty}^{\infty}\left[F(y)-I_{\{x\geq x\}}\right]^2\omega(y)dy \qquad (1.12)$$

where $(\omega y\geq)0$ is a weight function. Setting $(\omega y\equiv)1$ resulted in the CRPS in Equation (1.12).With the help of $\omega(y)=I_{\{y\geq r\}}$, one can address the coronavirus spread above a given threshold r; where the thresholds correspond approximately to 90th, 95th and 99th percentiles of the coronavirus spread observations of all considered health centres in South Africa.

3 APPLICATION ON CORONAVIRUS DATA

This section presents an empirical analysis of real-world data, using daily confirmed coronavirus cases in South Africa from March 5, 2020, to March 1, 2023. This collection comprises 1,092 samples. We employed a Markov-switching autoregressive (MS-AR) model with a distinctive statistical distribution to analyse the changes in coronavirus cases and forecast potential epidemic surges in South Africa. Figure 1 presents a time series plot in the left panel, illustrating both upward and downward trends, as well as seasonal variations. The highest point occurred between December 2021 and January 2022. This result indicates that the series is non-stationary. The quantile-quantile (Q-Q) plot presented in Figure 1 further supports these findings. The Q-Q plot indicates that the distribution of newly confirmed coronavirus cases deviates from a normal distribution. The series adheres to a fat-tailed distribution. The kurtosis values presented in Table 1, all exceeding three, support the conclusions regarding the fat tail phenomenon; hence, the conclusion that the distribution of new confirmed cases in South Africa exhibits leptokurtic characteristics. Wong and Collins (2020) noted that the spread of the coronavirus exhibited a fat-tail distribution. These researchers aimed to determine if the virus propagation exhibited an exponential trend, and they employed three distinct methods to demonstrate that the tail behaves like a fat tail: 1) a zip plot, 2) a meplot, and 3) statistical estimators of the tail index. The Omicron variant is responsible for the increase in new confirmed coronavirus infections, alongside less stringent public health and social measures (WHO, 2022). As of early April, South Africa has reported 1,369 cases of the Omicron sub-variant BA.2, 703 cases of BA.4, and 222 cases of BA.5. BA.4 and BA.5 remain the primary concerns due to their significant mutations, which complicate the understanding of their impact on immunity.

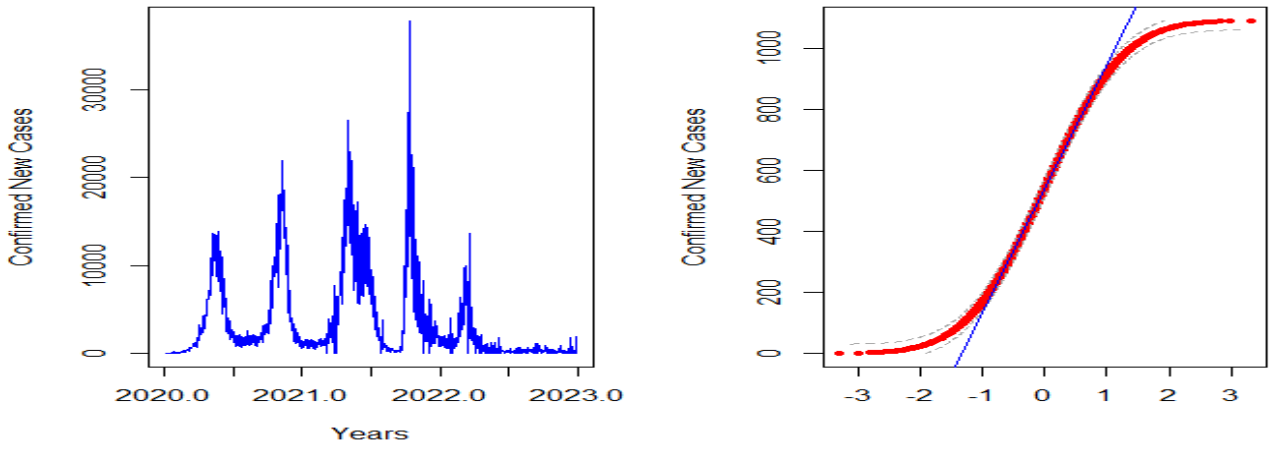


Figure 1: Time series plot for Confirmed Daily coronavirusCases

Table 1 provides a summary of the statistics related to confirmed cases of the coronavirus in South Africa. We analysed these statistics to further clarify the characteristics of the new coronavirus cases over a specified timeframe. The data indicates that the average number of new cases is on the rise, suggesting that the daily reported positive cases of the coronavirus in South Africa are increasing. The unconditional standard deviation stands at a notable 5099.562. This value indicates that the daily reported cases of the coronavirus exhibit considerable variation, resulting in a substantial number of cases. The reported kurtosis is notably high (7.579), and the

distribution exhibits a negative skewness of -1.8608, which confirms the presence of fat-tailed behaviour and asymmetry in the newly confirmed coronavirus cases in South Africa. Khan et al. (2021), in their investigation of extreme value theory and the coronavirus, identified fat-tailed behaviour in the NIFTY-50 that they analysed. Furthermore, Makatjane and Moroke (2022) revealed that high-frequency data, such as daily confirmed coronavirus cases, exhibit asymmetry and features fat tails.

Table 1: Descriptive statistics for Confirmed New Cases

	Central Tendency Measures				Normality Tests		
	Mean	Std dev	Skewness	Kurtosis	SW	KS	Jb
New Cases	3707.862	5099.562	-1.8608	7.579523	0.824(0.001)	0.984(0.001)	1584.4(0.001)

NB: values in () are probability values of JB, S-W and A-D

3.1 Results and Discussion

This section is divided into two parts: trend analysis and model fitting.

3.1.1 Trend Analysis Results

The Mann-Kendall test statistic and Sen's slope estimator are employed to examine the long-term trends of daily confirmed coronavirus cases in South Africa. Table 2 elucidates the results through the Mann-Kendall test statistic and Sen's slope. The results of the Mann-Kendall test indicated that the number of newly confirmed daily coronavirus cases exhibits a significant long-term monotonic decreasing trend, as evidenced by a negative value. The Sen's slope value indicates notable decreasing magnitudes of trends, aligning with the findings of the Mann-Kendall test. The decline is due to South Africa taking strong actions, such as setting up coordination systems at all levels, introducing control measures for key response areas, and enforcing public health and social rules. These measures encompassed movement restrictions, curfews, and the closure of businesses, educational institutions, and places of worship (WHO, 2020). Anne (2020) articulated three reasons that elucidate why the coronavirus has exhibited a lower mortality rate on the African continent compared to other regions.

These include: On 14 February, Egypt became the first country in Africa to confirm a case of coronavirus. Concerns arose that the emerging virus might rapidly strain the already vulnerable health systems across the continent. Consequently, from the outset, many African governments implemented significant measures to mitigate the virus's spread. Public health measures were implemented, including the avoidance of handshakes, frequent hand washing, social distancing, and the introduction of face mask usage. Certain countries, such as Lesotho, took action prior to the reporting of any cases. On 18 March 2020, Lesotho declared a state of emergency, subsequently closed schools, and initiated a three-week lockdown approximately ten days later, in alignment with several other southern African nations. Lesotho identified its initial confirmed cases shortly after the lockdown was lifted in early May (Anne, 2020). A survey conducted in August by PERC across 18 countries revealed high public support for safety measures, with 85% of respondents indicating they wore masks in the previous week. The reinforcement of strict public health and social measures allowed African Union (AU) member states to contain the virus between March and May. The report indicated that a "minor loosening of restrictions" in June and July correlated with a rise in reported cases throughout the continent. Since that time, a significant reduction in confirmed cases and fatalities has been observed in approximately half of the continent, potentially associated with the conclusion of the southern hemisphere winter. The youthful demographic in many African nations may have contributed to limiting the transmission of coronavirus. Globally, the majority of fatalities have occurred among individuals aged over 80, whereas Africa possesses the youngest population in the world, with a median age of 19 years. The pandemic predominantly affects younger populations, with approximately 91% of coronavirus infections in sub-Saharan Africa, including South Africa, occurring in individuals under 60 years of age, and over 80% of these cases being asymptomatic (WHO, 2022).

Table 2 Mann-Kendall test statistic and Sen's slope estimator

Variable	M-K Test Statistic	Kendall's Tau $\hat{\tau}$	p-Value	Sen's Slope
New Cases	-11.6255	-0.23496	0.001	-1.918

3.1.2 Markov-Switching Autoregressive-Non-Stationary TGEV

The study initially trains MS(k)-AR(p) using the ratio of 80% training and 20% validation sets to start our analysis. The goal here is to filter the coronavirus data to identify regime shifts and apply the upper regime, which is characterised by high variability, to our non-stationary TGEV distribution. We calculate the parameters for twelve AR(1) models subject to two regimes—MS(2)-AR(1) to MS(2)-AR(12)—using a method called expectation-maximisation. We use Final Prediction Error (FPE) and Predicted Residual Error Sum of Squares (PRESS) from Barron (2020), Bayesian Information Criterion (BIC) from Schwarz (1978), and Akaike Information Criterion

(AIC) from Akaike (1976) to select a simple lag length. Based on the outcomes of these criteria, we judge one's lag length to be parsimonious. Therefore, we proceed with a non-stationary MS(2)-AR(1) model, and Table 3 presents these findings.

Table 3. Two-Regime MS (2)-AR (1)

Regime 1				
Parameter	Coefficients	Std. error	t-value	p-value
$\hat{\mu}_1$	-179400	20353	-8.8144	0.001
$\hat{\phi}_1$	-1.5663	2.1663	-0.723	0.470
$\hat{\sigma}_1$	4271.4	27.098	157.6292	0.001
Regime2				
$\hat{\mu}_2$	-392820	14892	-26.378	0.001
$\hat{\phi}_2$	3.8722	1.244	3.1127	0.002
$\hat{\sigma}_2$	5433.6	26.071	208.4155	0.001
Transition Probabilities				
	$P_{11} = 0.994$			$P_{12} = 0.005$
	$P_{21} = 0.006$			$P_{22} = 0.995$

The two identified regimes exhibit differing interpretations in health and economic contexts. The variance in regime two surpasses that of regime one by 1162.2. The conditional distribution demonstrates considerable volatility and susceptibility to regime shifts, with an estimated daily count of 209 confirmed new coronavirus cases in South Africa. Upon transitioning the newly confirmed coronavirus to the second regime, the average daily confirmed cases declined to 392,820. This indicates that, during regime two, the likelihood of the virus transitioning to regime one is currently 0.006. The average duration of each regime supports this behaviour; where, regime one is expected to last for approximately 29 days and 4 hours, while regime two is projected to last for 57 days. We identified a significant regime shift in newly confirmed coronavirus cases in South Africa during the specified study period. Figure 2 displays the outcomes of filtered, smoothed, and predicted probabilities. Yang and Shaman (2022) attribute this regime shift to several coronavirus variants, namely Beta, Delta, and Omicron BA1. Despite South Africa's lower per capita case numbers relative to many other nations, the true extent of infections was likely much greater due to under-detection. In Gauteng, the estimated infection-detection rate during the initial pandemic wave was 4.59% (95% CI: 2.62–9.77%). The rate experienced a slight increase to 6.18% (95% CI: 3.29–11.11%) during the Beta wave and to 6.27% (95% CI: 3.44–12.39%) during the Delta wave. The estimates correspond with serological data. A sero-survey at the population level in Gauteng indicated a seropositivity rate of 68.4% among unvaccinated individuals following the Delta wave (Madhi et al., 2022). The number of reported cases at that time was about 6% of the population, and because some infections were missed in sero-surveys due to people losing antibodies and getting reinfected, it suggests that the total detection rate was below 10%.

TPM Vol. 32, No. S2, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

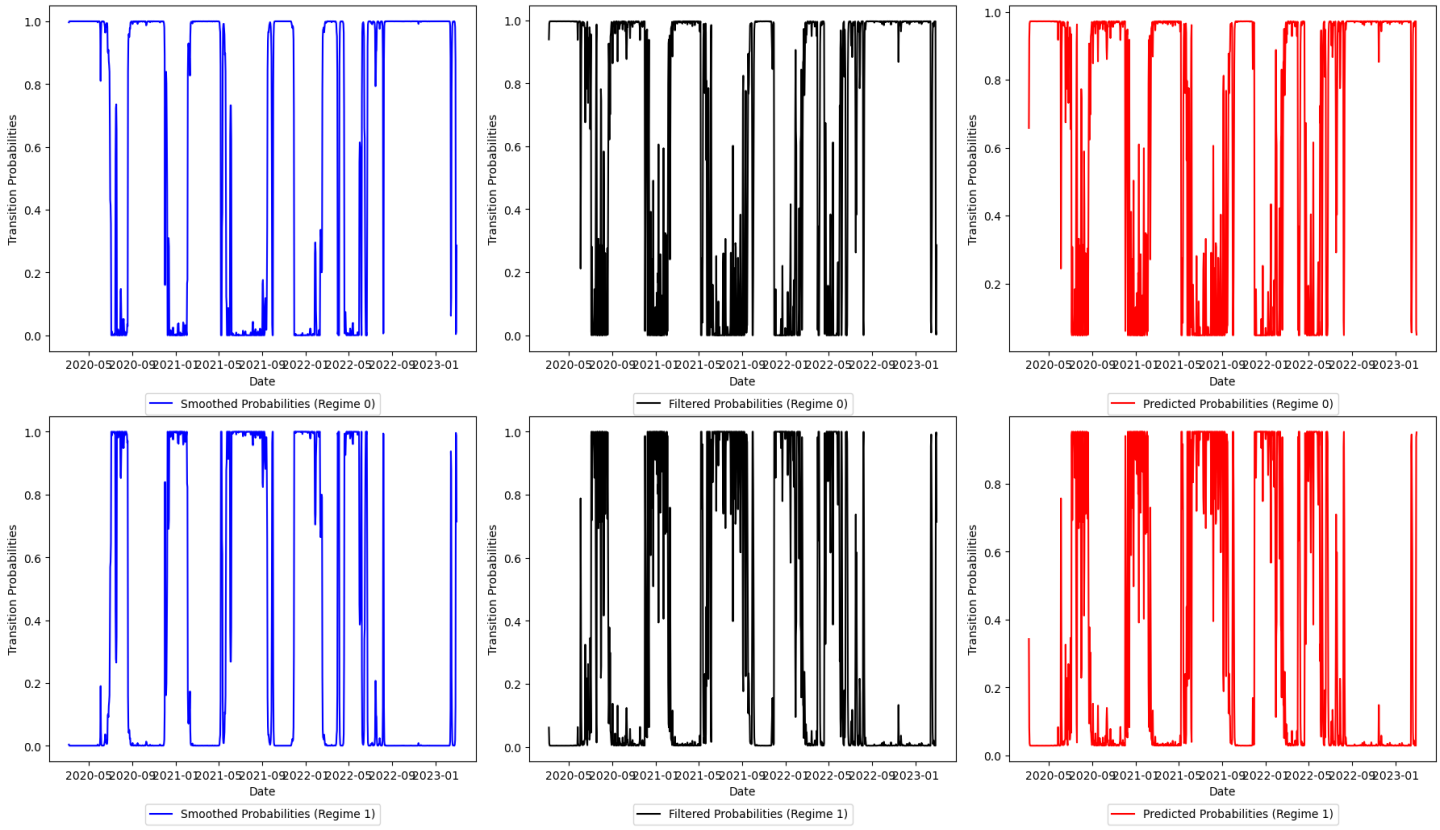


Figure 2: Filtered, smoothed and predicted probabilities

Assuming that the residuals from the upper regime of the fitted Ms(2)-AR(1) are of i.i.d sequence over a time span of n periods, a block maxima method (BMM) is applied to this sequence where the time span is ideally a calendar period.Figure 3 presents selected daily block maxima that are fitted to the proposed TGEV. The red dots in this figure are the extreme values of new confirmed corona virus cases.

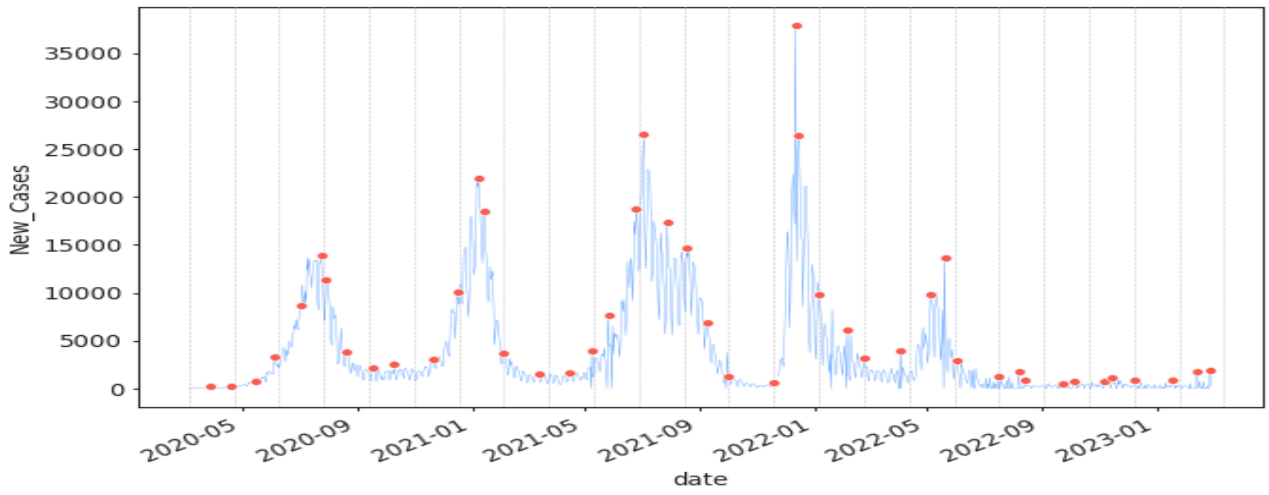


Figure 3: Block Maxima Source: Researcher's Own Computation

Table 4 displays the parameter estimates along with their corresponding standard errors for the non-stationary truncated Generalised Extreme Value (TGEV) distribution applied to the block maxima of confirmed COVID-19 cases in South Africa. The estimated shape parameter $(\xi = 0.378991)$ is positive, suggesting that the fitted distribution is classified within the Fréchet class (Type II extreme value distribution). This classification is appropriate for modelling heavy-tailed phenomena, consistent with Chan's (2016) findings of a similar positive

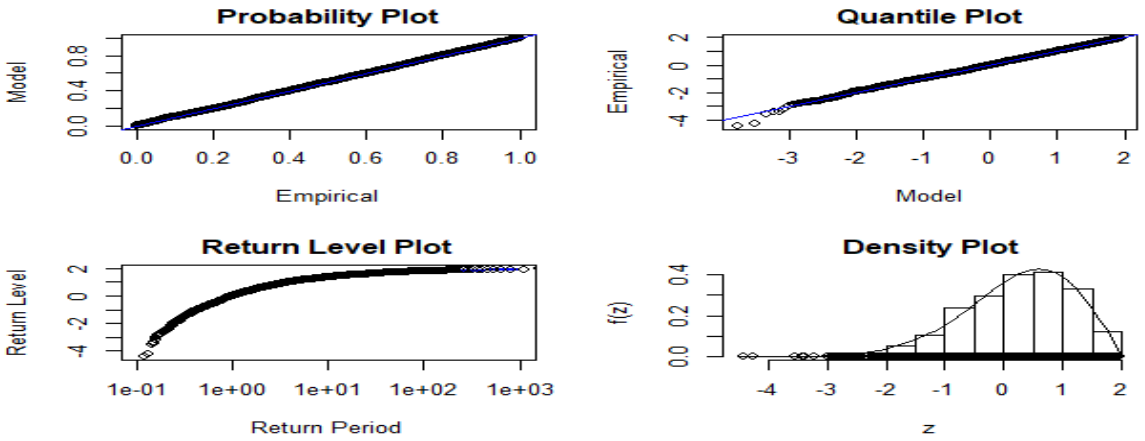
shape parameter in the context of statistically modelling extreme values for dependent variables. This result indicates a significant probability of extreme values, highlighting the potential for severe virus outbreaks. The location parameter is modelled as a linear function of time $\mu(t) = \mu_0 + \mu_1(t)$, with the intercept $(\mu_0 = 277190.9)$ indicating the baseline of extreme values and the slope $(\mu_1 = 5037.35)$ reflecting temporal trends. The estimated scale parameter $(\sigma = 3.7022)$ indicates the variability in block maxima.The asymptotic variance-covariance matrix helps clarify the estimates' reliability. The variances of the location parameters, especially the slope, are notably high, raising concerns regarding the stability and significance of the trend component. The relationships between parameters, like the negative link between the location intercept and slope, hint at possible problems in clearly identifying the model. A significant positive covariance between the shape and scale parameters indicates a potential interdependence in estimating tail heaviness and spread. The findings indicate that, despite some estimation challenges, a non-stationary TGEV distribution is appropriate for modelling extreme COVID-19 cases in South Africa. The model effectively shows the extreme values in the data, and its ability to adjust to changes over time highlights its importance in changing disease situations.

Table 4Truncated GEV distribution Estimates

Block	Maxima	$\hat{\xi}$	$\text{se}(\hat{\xi})$	$\hat{\sigma}$	$\text{se}(\hat{\sigma})$	$\hat{\mu}_0$	$\text{se}(\hat{\mu}_0)$	$\hat{\mu}_1$	$\text{se}(\hat{\mu}_1)$
12	156	0.378991	0.1073862	277190.9	2420.4853	5037.3515	190656.3	3.7022	5.2803
		Asymptotic Variance Covariance							
		0.0115318		259.9267		-437.2455		218895	
		259.926746		5858749		-18651610		-704.7218	
		-437.245452		-18651610		25374910		-2481.263	

Even though strong COVID-19 spikes are uncommon, the positive shape parameter and heavy-tailed distribution suggest that South Africa is vulnerable to them. This tail risk emphasises health sector preparedness. Critical care infrastructure, personnel, and emergency resources must be scalable and robust to case spikes in hospitals and health systems. Unanticipated extremes might overburden health resources, compromising patient treatment and increasing mortality. The likely (albeit statistically uncertain) rising trend in the location parameter implies that future waves may be more severe, especially if caused by novel mutations or public health failings. Moreover, extreme COVID-19 and other outbreaks threaten South Africa's vulnerable economy. Renewing lockdowns, mobility restrictions, or public health demands may disrupt labour markets, lower productivity, and hurt tourism, retail, and other industries. Policy planning and fiscal forecasting methods may have underestimated future shocks due to a non-stationary risk profile. We therefore need to invest in contingency finances, flexible social assistance, and health infrastructure as a risk mitigation strategy. The findings illustrate the connection between epidemic extremes and socioeconomic resilience in South Africa.

For model diagnosis, TGEV residual values are exponentially distributed. Most of them exhibit straight lines, indicating that the TGEV is a suitable model for new South African coronavirus infections. Chinhamu et al. (2015) observed similar findings. More block sizes improve GEV distribution data fit. Therefore, the Fisher theorem is applicable only when the block size is defined (Fisher and Tippett, 1928). Since the shape parameter is estimated positively, the return level curve asymptotes to a limited level; therefore, the estimated curve is almost quadratic, while the estimate is near zero. After accounting for sample variability, the curve accurately represents empirical estimates. Finally, the density estimate of the data histogram looks consistent. Thus, the four diagnostic graphs support the fitted TGEV distribution.



3.2 Model Performance Analysis

This section aims to identify the model that most accurately replicates the data while generating fewer forecasts. To do this, the studyused MAE, MAPE, RMSE, CRPS, and Bias, which are recognised methods for measuring errors in statistics. Confirmed new coronavirus cases in nine provinces of South Africa are non-exchangeable. For post-processing, the studybenchmarked with truncated normal and GEV, ultimately employing EMOS to combine the forecasts of MS-AR and TGEV. The studycalibrated the ensemble forecasts for the calendar year 2020 using a 30-day training period, which we derived from a comprehensive preliminary analysis (see Baran and Lerch, 2015). Table 5 shows a summary of the verification metrics, coverage, and the average width of the 77.78% (i.e., one standard uncertainty) central prediction intervals for different EMOS models, the raw ensemble, and the MS-AR-TGEV model. Although the MS-AR-TGEV model is significantly outperformed by the raw ensemble in terms of average CRPS, MAE, and RMSE values, it perfectly predicts extreme values as demonstrated by low mean twCRPS values. The calibration of the raw ensemble forecasts is inadequate—weighted ensemble forecasts are too tight, which results in a lack of uncertainty coverage and under-dispersed prediction intervals. By contrast, the MS-AR-TGEV model gives much wider intervals that account for the better coverage. The EMOS post-processing leads to a clear improvement in the calibration and forecast quality of the raw ensemble. All EMOS-based predictions consistently perform significantly better than both the raw ensemble and MS-AR-TGEV, particularly in terms of the mean twCRPS for the extreme coronavirus spread.

Table 5. Model Performance

Forecast	CRPS	MAE	MAPE	BIAS	RMSE	twCRPS ($r = 9$)	Cover
Training Set							
MS-AR	0.738	0.799	1.247	0.376	1.357	0.150	83.59
GEVD	0.741	0.913	1.002	0.436	1.362	0.149	80.44
MS-AR-GEVD	0.737	0.802	1.037	0.537	1.355	0.145	81.21
Truncated normal	0.736	0.710	1.038	0.497	1.356	0.145	82.13
Truncated GEV	0.901	0.622	1.041	0.468	1.373	0.175	68.22
Ensemble	1.046	0.613	1.037	0.517	1.822	0.173	82.54
MS-AR-Truncated GEV	0.803	0.599	0.979	0.315	1.352	0.163	84.87
Test Data							
MS-AR	0.102	0.437	1.001	0.376	0.978	0.010	92.19
GEVD	0.102	0.513	1.079	0.436	1.362	0.010	93.16
MS-AR-GEVD	0.112	0.441	1.067	0.537	1.001	0.010	94.84
Truncated normal	0.127	0.428	1.986	0.497	0.991	0.010	92.89
Truncated GEV	0.931	0.522	1.732	0.468	0.976	0.011	95.84
Ensemble	0.099	0.424	1.007	0.517	1.876	0.010	48.16
MS-AR-Truncated GEV	0.098	0.425	1.008	0.315	1.000	0.012	97.38

4 CONCLUSION

The uncertainties regarding the spread of coronavirus present considerable global challenges, especially for public health systems and policymakers who must rapidly address changing outbreak dynamics. The accurate prediction and forecasting of daily COVID-19 cases are essential for timely interventions and the efficient allocation of healthcare resources. This study enhances the forecasting of coronavirus spread in South Africa through the development of a hybrid probabilistic model that combines a Markov-Switching Autoregressive (MS(k)-AR(p)) process with a non-stationary Truncated generalised Extreme Value (TGEV) distribution.This combined MS-AR–TGEV system successfully identifies changes in transmission patterns and the extreme spikes in cases. The results show that the MS(2)-AR(1)-TGEV model predicts better, especially in uncertain and changing situations, making it a more reliable choice than traditional models. The findings present a practical solution for epidemic monitoring in South Africa and offer a generalisable methodology for forecasting future outbreaks characterised by complex, non-linear, and extreme behaviours, akin to those observed during the COVID-19 pandemic.
This study presents several recommendations aimed at enhancing epidemic forecasting and preparedness. Public health authorities should use probabilistic models, like the MS(2)-AR(1)-TGEV framework, in their regular monitoring systems because they are better at measuring uncertainty and predicting severe outbreak situations. The model's regime-switching and tail detection capabilities render it appropriate for early warning systems and strategic resource allocation in times of increased risk. The model's adaptability renders it applicable to other infectious diseases that demonstrate comparable non-linear and extreme behaviours. Regular retraining of the model with real-time data is advisable to maintain accuracy, particularly during times of variant emergence or policy changes. Furthermore, enhancing analytical capacity via focused training in ensemble forecasting and extreme value theory will allow health agencies to more effectively use these models. Policymakers should use scenario-based forecasts from this model to inform proactive interventions, including the implementation or relaxation of lockdown measures, the deployment of medical supplies, and the management of public health communications.

Acknowledgements
The authors would like to acknowledge the University of South Africa and the University of Botswana for giving us time to do this study.
Conflict of Interest
The authors do not have any conflicting interests concerning the publishing of this work.

REFERENCES

1. Acter T, Uddin N, Das J, Akhter A, Choudhury TR, Kim S. Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency. Science of the Total Environment. Aug 2020. doi:https://doi.org/10.1016/j.scitotenv.2020.138996
2. Akaike H. An information criterion (AIC). Math Sci. ;14:5-7. 1976.
3. Al-Zoughool M, Oraby T, Vainio H, Gasana J, Longenecker J, Al Ali W, AlSeaidan M, Elsaadany S, Tyshenko MG. Using a stochastic continuous-time Markov chain model to examine alternative timing and duration of the COVID-19 lockdown in Kuwait: what can be done now?. Archives of Public Health. 8;80(1):22, Jan 2022. doi:https://doi.org/10.1186/s13690-021-00778-y
4. Baran S, Lerch S. Log-normal distribution based Ensemble Model Output Statistics models for probabilistic wind-speed forecasting. Quarterly Journal of the Royal Meteorological Society.141(691):2289-99. Jul 2015.https://doi.org/10.1002/qj.2521
5. Baran S, Szokol P, Szabó M. Truncated generalized extreme value distribution-based ensemble model output statistics model for calibration of wind speed ensemble forecasts. Environmetrics. 32(6):e2678. Sep 2021 doi:https://doi.org/10.1002/env.2678.
6. Barron AR. Predicted squared error: A criterion for automatic model selection. InSelf-organizing methods in modeling (pp. 87-103). CRC Press. Nov 2020.
7. Chan, Ka Shing. "Statistical modeling of extreme values for dependent variables." PhD diss., 2016.
8. Chinhamu K, Huang CK, Huang CS, Chikobvu D. Extreme risk, value-at-risk and expected shortfall in the gold market. The International Business & Economics Research Journal (Online).14(1):107; 2015. https://doi.org/10.19030/iber.v14i1.9035

9. de Oliveira AM, Binner JM, Mandal A, Kelly L, Power GJ. Using GAM functions and Markov-Switching models in an evaluation framework to assess countries' performance in controlling the COVID-19 pandemic. BMC Public Health.1-4; Dec 2021: doi:https://doi.org/10.1186/s12889-021-11891-6.
10. Douwes-Schultz D, Schmidt AM, Shen Y, Buckeridge D. A three-state coupled Markov switching model for COVID-19 outbreaks across Quebec based on hospital admissions. arXiv preprint arXiv:2302.02488. Feb 2023.doi:
11. https://doi.org/10.48550/arXiv.2302.02488
12. Fisher RA, Tippett LH. Limiting forms of the frequency distribution of the largest or smallest member of a sample. InMathematical proceedings of the Cambridge philosophical society (Vol. 24, No. 2, pp. 180-190). Cambridge University Press. Apr 1928.
13. Friederichs P, Thorarinsdottir TL. Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. Environmetrics.23(7):579-94. Nov 2012 doi: https://doi.org/10.1002/env.2176
14. Gneiting T. Making and evaluating point forecasts. Journal of the American Statistical Association. Jun 2011;106(494):746-62.. doi:https://doi.org/10.1198/jasa.2011.r10138
15. Gneiting T, Katzfuss M. Probabilistic forecasting. Annual Review of Statistics and Its Application.1(1):125-51; Jan 2014. doi:https://doi.org/10.1146/annurev-statistics-062713-085831.
16. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association;102(477):359-78.Mar 2007 doi:https://doi.org/10.1198/016214506000001437.
17. Good IJ. Rational decisions. Journal of the Royal Statistical Society: Series B (Methodological).14(1):107-14.Jan 1952.doi:https://doi.org/10.1111/j.2517-6161.1952.tb00104.x.
18. Haimerl P, Hartl T. Modeling COVID-19 Infection Rates by Regime-Switching Unobserved Components Models. Econometrics. 3;11(2):10.Apr 2023. doi:https://doi.org/10.3390/econometrics11020010.
19. Hamilton JD. Regime switching models. InMacroeconometrics and time series analysis (pp. 202-209). London: Palgrave Macmillan UK,2010.
20. He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. Nonlinear dynamics. 101:1667-80. Aug 2020.doi:https://doi.org/10.1007/s11071-020-05743-y
21. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet. 395(10223):497-506. Feb 2020.doi:https://doi.org/10.1016/S0140-6736(20)30183-5.
22. Hung LS. The SARS epidemic in Hong Kong: what lessons have we learned?. Journal of the Royal Society of Medicine.96(8):374-8Aug 2003.doi:https://doi.org/10.1177/014107680309600803.
23. Justus CG, Hargraves WR, Mikhail A, Graber D. Methods for estimating wind speed frequency distributions. Journal of Applied Meteorology (1962-1982). 1:350-3. Mar 1978.
24. Justus CG, Hargraves WR, Mikhail A, Graber D. Methods for estimating wind speed frequency distributions. Journal of Applied Meteorology (1962-1982). 1:350-3. Mar 1978 doi:https://www.jstor.org/stable/26178009.
25. Kawohl W, Nordt C. COVID-19, unemployment, and suicide. The Lancet Psychiatry. 7(5):389-90. May 2020 doi:https://doi.org/10.1016/S2215-0366(20)30141-3.
26. Khan M, Aslam F, Ferreira P. Extreme value theory and COVID-19 pandemic: evidence from India. Economic Research Guardian.11(1):2-10. 2021.
27. Lee S, Liao Y, Seo MH, Shin Y. Sparse HP filter: Finding kinks in the COVID-19 contact rate. Journal of econometrics. 1;220(1):158-80. Jan 2021 doi:https://doi.org/10.1016/j.jeconom.2020.08.008.
28. Lerch S, Baran S. Similarity-based semilocal estimation of post-processing models. Journal of the Royal Statistical Society Series C: Applied Statistics. 66(1):29-51. Jan 2017. doi:https://www.jstor.org/stable/44681861.
29. Lerch S, Thorarinsdottir TL. Comparison of non-homogeneous regression models for probabilistic wind speed forecasting. Tellus A: Dynamic Meteorology and Oceanography.1;65(1):21206.Dec 2013. doi:https://doi.org/10.3402/tellusa.v65i0.21206.
30. Lu N, Cheng KW, Qamar N, Huang KC, Johnson JA. Weathering COVID-19 storm: Successful control measures of five Asian countries. American journal of infection control. 48(7):851. Jul 2020. doi:https://doi.org/10.1016/j.ajic.2020.04.021.
31. Madhi SA, Kwatra G, Myers JE, Jassat W, Dhar N, Mukendi CK, Nana AJ, Blumberg L, Welch R, Ngorima-Mabhena N, Mutevedzi PC. Population immunity and Covid-19 severity with Omicron variant in South Africa. New England Journal of Medicine. 7;386(14):1314-26. Apr 2022. doi:https://doi.org/10.1056/NEJMoa2119658.

32. Makatjane K, Moroke N. Examining stylized facts and trends of FTSE/JSE TOP40: a parametric and Non-Parametric approach. Data Science in Finance and Economics. 2(3):294-320.2022. doi:https://doi.org/10.3934/dsfe.2022015.
33. Makatjane, K., and Xaba, D. (2016). An early warning system for inflation using Markov-Switching and logistic models approach. Risk Governance & Control: Financial Markets & Institutions, 6(4).
34. Masingi VN, Maposa D. Modelling long-term monthly rainfall variability in selected provinces of South Africa: Trend and extreme value analysis approaches. Hydrology. 23;8(2):70. Apr 2021 doi:https://doi.org/10.3390/hydrology8020070
35. Panagoulia D, Economou P, Caroni C. Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. Environmetrics. 25(1):29-43. Feb 2014. doi: https://doi.org/10.1002/env.2252
36. Qu Z, Sha Y, Xu Q, Li Y. Forecasting New COVID-19 Cases and Deaths Based on an Intelligent Point and Interval System Coupled With Environmental Variables. Frontiers in Ecology and Evolution. 2;10:875000. May 2022. doi:https://doi.org/10.3389/fevo.2022.875000.
37. Rossouw S, Greyling T, Adhikari T. The evolution of happiness pre and peri-COVID-19: A Markov Switching Dynamic Regression Model. Plos one. 10;16(12):e0259579.Dec 2021. doi:https://doi.org/10.1371/journal.pone.0259579.
38. Schwarz G. Estimating the dimension of a model. The annals of statistics. 1:461-4. Mar 1978. doi:https://www.jstor.org/stable/2958889.
39. Sen PK. Estimates of the regression coefficient based on Kendall's tau. Journal of the American statistical association. 1;63(324):1379-89. Dec 1968. doi:https://doi.org/10.1080/01621459.1968.10480934.
40. Shinde V, Bhikha S, Hoosain Z, Archary M, Bhorat Q, Fairlie L, Lalloo U, Masilela MS, Moodley D, Hanley S, Fouche L. Efficacy of NVX-CoV2373 Covid-19 vaccine against the B. 1.351 variant. New England Journal of Medicine. 20;384(20):1899-909. May 2021. doi:https://doi.org/10.1056/NEJMoa2103055.
41. Sohrabi C, Alsafi Z, O'neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International journal of surgery. 1;76:71-6. Apr 2020 .doi:https://doi.org/10.1016/j.ijsu.2020.02.034.
42. Somyanonthanakul R, Warin K, Amasiri W, Mairiang K, Mingmalairak C, Panichkitkosolkul W, Silanun K, Theeramunkong T, Nitikraipot S, Suebnukarn S. Forecasting COVID-19 cases using time series modeling and association rule mining. BMC medical research methodology. 1;22(1):281. Nov 2022. doi:https://doi.org/10.1186/s12874-022-01755-x.
43. Syafrina AH, Norzaida A, Ain JJ. Stationary and nonstationary generalized extreme value models for monthly maximum rainfall in Sabah. InJournal of Physics: Conference Series 1 (Vol. 1366, No. 1, p. 012106). IOP Publishing. Nov 2019.
44. Tay AS, Wallis KF. Density forecasting: a survey. Journal of forecasting. 19(4):235-54. Jul 2000. doi:https://doi.org/10.1002/1099-131X(200007)19:4.
45. Thorarinsdottir TL, Gneiting T. Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. Journal of the Royal Statistical Society Series A: Statistics in Society. 173(2):371-88. Apr 2010. doi: https://doi.org/10.1111/j.1467-985X.2009.00616.x.
46. World Health Organization. WHO encouraged by South Africa's declining COVID-19 trend [Internet].
47. World Health Organization (WHO). Southern Africa faces uptick in COVID-19 cases [Internet]. 2022 [cited 2023 Dec 19]. Available from: https://www.afro.who.int/news/southern-africa-faces-uptick-covid-19-cases.
48. Wi S, Valdés JB, Steinschneider S, Kim TW. Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima. Stochastic environmental research and risk assessment. 30:583-606. Feb 2016. doi:https://doi.org/10.1007/s00477-015-1180-8.
49. Wilks DS. Statistical methods in the atmospheric sciences. Academic press; Jul 2011.
50. Wong F, Collins JJ. Evidence that coronavirus superspreading is fat-tailed. Proceedings of the National Academy of Sciences. 24;117(47):29416-8. Nov 2020. doi:https://doi.org/10.1073/pnas.2018490117.
51. Xie L. The analysis and forecasting COVID-19 cases in the United States using Bayesian structural time series models. Biostatistics & Epidemiology. 2;6(1):1-5. Jan 2022. doi:https://doi.org/10.1080/24709360.2021.1948380.
52. Yang W, Shaman JL. COVID-19 pandemic dynamics in South Africa and epidemiological characteristics of three variants of concern (Beta, Delta, and Omicron). Elife. 11;2022. doi:https://doi.org/10.7554/eLife.78933.

53. Young BE, Ong SW, Kalimuddin S, Low JG, Tan SY, Loh J, Ng OT, Marimuthu K, Ang LW, Mak TM, Lau SK. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. Jama. 21;323(15):1488-94. Apr 2020. doi:https://doi.org/10.1001/jama.2020.3204
54. Yuen RA, Baran S, Fraley C, Gneiting T, Lerch S, Scheuerer M, Thorarinsdottir TL. R package ensembleMOS, Version 0.8. 2: Ensemble Model Output Statistics. Elérheto: https://cran. r-project. org/package= ensembleMOS [Letöltve: 2019.06. 16]. 2018.