BMC Infectious Diseases

RESEARCH

Open Access

Endemic-epidemic modeling of the COVID-19 pandemic in South Africa

Rashider Aloni[1], Niel Hens[1†] and Tarylee Reddy[2*†]

Abstract

Background The COVID-19 pandemic has exhibited complex, multiwave dynamics with substantial spatial and temporal heterogeneity. In South Africa, repeated waves, driven by variant emergence, shifting public health policies, and uneven vaccine uptake, posed significant challenges to real-time surveillance and predictive modeling. There is a growing need for statistical frameworks that can capture these dynamics while offering interpretable insights for public health planning.

Methods We applied a spatio-temporal endemic-epidemic model to daily COVID-19 case counts across nine South African provinces from March 2020 to July 2022. The final model included fixed effects for time trends, seasonality, variant dominance, lagged vaccination coverage, government stringency, and weekend reporting patterns. Spatial transmission was modeled using power-law distance weights, and province-specific random intercepts were included in all components. Transmission was decomposed into endemic (background), autoregressive (within-province), and neighbourhood (interprovincial) contributions. Model validation involved 14-day internal forecasting, with predictive accuracy evaluated using 95% prediction intervals.

Results The final model provided the best fit based on AIC, mean log score, and dominant epidemic eigenvalue. Local transmission dominated overall spread, especially in provinces with sustained epidemic activity. The neighbourhood component highlighted Gauteng and Western Cape as key sources of spatial transmission. Omicron dominance significantly increased both background and interprovincial transmission, while higher vaccination coverage was associated with reduced spatial spread. The model achieved good forecasting performance, with most observed values falling within 95% prediction intervals. Divergence after Day 10 in forecasts suggested early signals of new wave onset.

Conclusion This study shows that the endemic–epidemic model offers a practical and interpretable way to monitor COVID-19 transmission across South Africa's provinces. By combining spatial structure, temporal patterns, and relevant covariates, the framework helps identify dominant transmission routes and detect emerging changes in epidemic pressure. These features make the model useful for near-real-time surveillance and for guiding locally targeted public-health responses, particularly in settings where resources and response capacity vary across regions.

Keywords COVID-19, South Africa, Endemic–epidemic model, Spatio-temporal modeling, Public-health surveillance

†Niel Hens and Tarylee Reddy contributed equally to this work.

*Correspondence:
Tarylee Reddy
tarylee.reddy@mrc.ac.za

[1]Data Science Institute, Hasselt University, Agoralaan, Building D, Diepenbeek, Limburg 3590, Belgium
[2]Biostatistics Research Unit, South African Medical Research Council, 491 Peter Mokaba Ridge Road, Overport, Durban, KwaZulu-Natal 4091, South Africa

Introduction

The COVID-19 pandemic has exposed critical limitations in traditional infectious disease modeling, particularly in the context of protracted, spatially heterogeneous outbreaks. Unlike typical epidemics that unfold as a single, self-limiting wave, COVID-19 has demonstrated a sustained, multiwave structure characterized by overlapping surges across different regions and time periods. These dynamics have been shaped by numerous factors, including the emergence of new variants, shifts in population behavior, variation in immunity, and evolving public health interventions [1, 2].

In South Africa, this complexity has been especially pronounced. The country has experienced several distinct epidemic waves, driven by successive variants of concern, Beta, Delta, and Omicron, each with differing transmissibility and immune escape properties. These waves have affected provinces at different times and intensities, reflecting variations in demographic structure, healthcare capability, vaccine uptake, and the timing and enforcement of non-pharmaceutical interventions such as lockdowns and mobility restrictions [3, 4]. Such spatio-temporal disparities underscore the need for analytical approaches that can capture both local persistence and interprovincial spread, while accounting for exogenous factors like vaccination and policy stringency.

Various modeling approaches have been applied in South Africa to guide decision-making during the pandemic. Among the most prominent was the National COVID-19 Epi Model (NCEM), produced by the South African COVID-19 Modelling Consortium (SACMC). This stochastic SEIR-type model incorporated age structure, disease severity, and treatment pathways, and was used extensively for scenario-based forecasting and healthcare resource planning. While effective for national-level projections, the NCEM required numerous fixed assumptions, such as age-specific contact matrices and healthcare utilization patterns, which can be difficult to update in real time, particularly in settings where granular surveillance data are limited.

Other efforts have adopted simpler approaches. Reddy et al. [5] employed parametric growth models, including Gompertz, Weibull, and logistic curves, to project cumulative cases and deaths, though these were largely restricted to the first wave and lacked consideration of spatial or immunological drivers. More recently, machine learning models have been explored to forecast short-term indicators like hospital admissions [6]. While promising for operational use, such models often function as black boxes and offer limited insight into the mechanisms underlying transmission.

Taken together, these approaches have offered valuable guidance during the pandemic, but many fall short in capturing the evolving nature of multiwave epidemics, especially where spatial variation and covariate effects are substantial. There remains a need for statistically robust, interpretable models capable of separating the contributions of local transmission, geographic spread, and external inputs, while integrating key public health covariates such as vaccination, variant prevalence, and policy measures.

The endemic–epidemic (EE) modeling framework [7, 8] addresses many of these challenges. By decomposing disease incidence into endemic (background), autoregressive (local), and neighborhood (spatial) components, the EE approach provides a transparent structure for assessing transmission dynamics. Extensions to the framework have incorporated flexible spatial interaction functions [9], random effects [8], and time-varying covariates such as mobility, interventions, and susceptibility [10, 11]. Although widely applied in high-income settings such as Belgium, Switzerland, and Italy, the framework has seen limited application in sub-Saharan Africa. One notable example from Rwanda focused on early pandemic dynamics [12], but lacked full pandemic coverage, spatial decay modeling, and the incorporation of variant and vaccination effects.

This study extends the endemic–epidemic framework to daily COVID-19 data from South Africa's nine provinces over a 29-month period. Our model integrates power-law spatial weights, province-specific random intercepts, and covariates capturing variant dominance, lagged vaccination uptake, policy stringency, seasonality, and weekend effects. By quantifying the relative influence of each transmission component, the model offers new insights into the drivers of COVID-19 spread across the country and provides an interpretable foundation for future response planning in resource-constrained settings.

Methods

Study setting and time frame

This study focused on the nine provinces of South Africa over a period spanning from March 5, 2020, to July 25, 2022. The analysis was performed on daily confirmed COVID-19 case counts, resulting in a total of 862 time points per province. Two days were excluded due to substantial reporting irregularities affecting multiple provinces. This dataset captures multiple epidemic waves, including those associated with Beta, Delta, and Omicron variants.

Data sources and preprocessing

Daily cumulative confirmed case counts were obtained from publicly released epidemiological reports and provincial dashboards published by the National Institute for Communicable Diseases (NICD) [13]. Vaccination data, capturing the number of individuals who had received

at least one dose, were sourced from the South African Department of Health [14]. Policy measures were quantified using the Oxford COVID-19 Government Response Tracker stringency index [15]. Circulating variant prevalence over time was derived from NICD weekly genomic surveillance reports [16] and supplemented using GISAID's public SARS-CoV-2 sequence data [17].

All datasets were aligned to a common daily time scale. Dates with known anomalies (e.g., retrospective case backlogs) were checked for consistency. Case data were transformed from cumulative to daily incident counts by differencing. Negative or implausible values, typically caused by retrospective corrections, were removed. Province-level population sizes were taken from Statistics South Africa mid-year estimates for 2020 [18], and used to define population-based offsets in the model.

Spatio-temporal model structure and covariates

We applied the endemic–epidemic (EE) modeling framework for multivariate time series of infectious disease counts, decomposing incidence into three components: endemic (baseline risk), autoregressive (within-province transmission), and neighbourhood (interprovincial spread). Let Y_{it} represent the number of newly confirmed cases of COVID-19 in province i ($i = 1, \ldots, 9$) on day t ($t = 1, \ldots, T$), where T is the total number of days under study. The covariate vector associated with province i at time t is denoted by $\mathbf{x}_{it}$. It is assumed that, conditional on previous incidence values and relevant covariates, the number of new infections at time t follows a Negative Binomial distribution:

$$Y_{it} \mid Y_{i,t-1}, \mathbf{x}_{it} \sim \mathrm{NegBin}(\mu_{it}, \psi_i),$$

where μ_{it} is the expected number of cases and $\psi_i \geq 0$ is a dispersion parameter that varies by province. Under this formulation, the variance of Y_{it} is given by $\mu_{it}(1 + \psi_i \mu_{it})$, which accommodates overdispersion typically observed in infectious disease data. When $\psi_i = 0$, the model simplifies to a Poisson distribution with equal mean and variance.

The conditional mean μ_{it} is further decomposed into three distinct transmission sources:

$$\mu_{it} = \underbrace{\eta_{it}}_{\text{Endemic}} + \underbrace{\lambda_{it} Y_{i,t-1}}_{\text{Autoregressive}} + \underbrace{\theta_{it} \sum_{j \neq i} w_{ji} Y_{j,t-1}}_{\text{Neighborhood}}, \quad (1)$$

where w_{ji} are spatial weights defined using a power-law decay based on interprovincial distances. Each model component was specified on the log scale with covariates and random effects as follows:

Endemic component (η_{it}):

$$\begin{aligned} \log(\eta_{it}) = \log(N_i^*) &+ \beta_0^{(end)} + \beta_1^{(end)} t \\ &+ \sum_{k=1}^{3} \beta_k^{(var)} \cdot \mathrm{Variant}_{k,t} + \beta^{(str)} \cdot \mathrm{Stringency}_t \\ &+ \beta^{(w)} \cdot \mathrm{Weekend}_t + \beta^{(v)} \cdot \mathrm{Vaccination}_{t-7} \\ &+ \sum_{s=1}^{3} \left[\gamma_s \sin\left(\frac{2\pi s t}{365}\right) + \delta_s \cos\left(\frac{2\pi s t}{365}\right) \right] \\ &+ u_i^{(end)} \end{aligned} \quad (2)$$

Autoregressive component (λ_{it}):

$$\begin{aligned} \log(\lambda_{it}) = \; &\beta_0^{(ar)} + \beta_1^{(ar)} t + \sum_{k=1}^{3} \beta_k^{(ar)} \cdot \mathrm{Variant}_{k,t} \\ &+ \beta^{(str)} \cdot \mathrm{Stringency}_t + \beta^{(w)} \cdot \mathrm{Weekend}_t \\ &+ \beta^{(v)} \cdot \mathrm{Vaccination}_{t-7} + \gamma^{(\lambda)} \log(N_i) + u_i^{(ar)} \end{aligned} \quad (3)$$

Neighborhood Component (θ_{it}):

$$\begin{aligned} \log(\theta_{it}) = \beta_0^{(ne)} &+ \beta^{(w)} \cdot \mathrm{Weekend}_t + \beta^{(v)} \cdot \mathrm{Vaccination}_{t-7} \\ &+ \gamma^{(\theta)} \log(N_i) + u_i^{(ne)}, \end{aligned} \quad (4)$$

Spatial weights and random effects

Spatial weights w_{ji} were defined using a power-law decay based on adjacency order:

$$w_{ji} = \frac{o_{ji}^{-\gamma}}{\sum_{k \neq i} o_{ki}^{-\gamma}}, \quad j \neq i$$

where o_{ji} is the adjacency-based distance between provinces j and i, and γ is an estimated decay parameter. Random intercepts $u_i^{(comp)} \sim \mathcal{N}(\beta_0^{(comp)}, \sigma_{comp}^2)$ were included for each component to allow region-specific deviations.

Covariates

N_i^*: Population of province i, scaled per 1,000,000 (offset). $\mathrm{Variant}_{k,t}$: Binary indicator for Beta, Delta, or Omicron variant dominance. $\mathrm{Stringency}_t$: Oxford COVID-19 Government Response Stringency Index. $\mathrm{Weekend}_t$: Binary indicator for weekends. $\mathrm{Vaccination}_{t-7}$: Complement of coverage (1 – coverage), lagged by 7 days. $\mathrm{Seasonality}_t$: Fourier terms with 3 harmonics to capture annual seasonality. t: Linear time trend (in days).

Model development and validation

Model construction followed a sequential strategy, starting with a baseline specification that included only intercept terms in each of the endemic, autoregressive, and neighbourhood components. Covariates were then added stepwise based on prior epidemiological knowledge and

exploratory analyses. The inclusion of covariates and selection of final model specification were guided by the Akaike Information Criterion (AIC) [19], Bayesian Information Criterion (BIC), and the mean logarithmic score (MLS), along with their epidemiological plausibility in the South African context. Predictive performance was evaluated through simulation-based forecasts over selected 14-day periods. Internal validation involved comparing predicted and observed values, examining the calibration and coverage of 95% prediction intervals, and computing root mean square error (RMSE) and mean absolute error (MAE). A complete summary of model selection and validation results is provided in the Results section. All models were implemented using the `hhh4()` function from the `surveillance` package [20] version 1.24.1 under R version 4.4.0. Parameter estimation relied on penalized likelihood methods while accounting for overdispersion at the provincial level.

Sensitivity analysis

To explore the sensitivity of the model to assumptions about the delay between vaccination and immune response, we considered three lag structures: 7, 14, and 21 days. In each case, the complement of vaccination coverage (i.e., $1 - \text{coverage}$) was entered as a covariate in all three components of the model: endemic, autoregressive, and neighbourhood. Each specification was estimated separately, and model fit was compared using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and mean logarithmic score (MLS).

Results

Descriptive data overview

We analyzed daily confirmed COVID-19 case counts from South Africa's nine provinces between March 5, 2020, and July 25, 2022, covering 862 complete days. Daily new cases were derived by subtracting cumulative case totals. Several relevant covariates were also compiled, including variant dominance, lagged vaccination coverage, and a time-varying stringency index, each aligned with case data to explore potential drivers of transmission.

Exploratory data analysis

Figure 1 presents the national and provincial trajectories of daily new COVID-19 cases. At the national level (Fig. 1a), five distinct waves are visible, each characterized by rapid rises and sharp declines in case counts. Periodic dips during weekends suggest reporting delays, motivating the inclusion of a weekend effect in the modeling framework. At the provincial level (Fig. 1b), epidemic waves show notable heterogeneity in both timing and magnitude. Gauteng, KwaZulu-Natal, and Western Cape recorded the highest peaks, especially during Beta and Omicron waves. In contrast, provinces such as Limpopo and Northern Cape exhibited lower peaks. These differences highlight the importance of including province-specific and spatial interaction terms in the endemic-epidemic model.

Summary statistics and peak case counts

Between March 2020 and July 2022, COVID-19 case counts varied substantially across South Africa's nine provinces, both in cumulative totals and peak daily cases. Gauteng recorded the highest overall burden, with over 1.3 million cumulative cases and daily peaks exceeding 16,000 during both the third and fourth waves. KwaZulu-Natal and the Western Cape followed, with peak daily cases surpassing 11,000 and 9000 respectively. In contrast, provinces such as Limpopo and the Northern Cape experienced considerably lower case loads, with cumulative totals well below 200,000 and maximum daily counts rarely exceeding 2000.

The standard deviation of reported cases was notably high in provinces with large urban centers, reflecting pronounced fluctuations across successive waves. By mid-2022, peak daily case numbers had declined in all provinces, with most reporting fewer than 5000 daily cases during the fifth wave. This downward trend likely reflects the combined effect of increased population-level immunity, prior infections, and changes in testing patterns. These observations underscore the importance of modeling approaches that can capture both spatial heterogeneity and temporal dynamics in COVID-19 transmission.

In addition, because the model included population offsets (cases per million residents), we note that provinces with large urban populations, particularly Gauteng and the Western Cape, also exhibited higher per-capita incidence, whereas sparsely populated provinces such as Limpopo and the Northern Cape had markedly lower per-capita burdens. This helps place the raw case counts in the proper context by showing how they relate to the size of each province's population.

Spatio-temporal endemic-epidemic model analysis

After conducting exploratory analysis, we fitted a series of progressively complex models within the endemic-epidemic framework to capture the spatio-temporal dynamics of COVID-19 across South Africa's nine provinces. Model selection was guided by predictive performance, epidemiological interpretability, and parsimony, using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and mean logarithmic score (MLS) for comparison.

Seven models were fitted, each sequentially incorporating seasonal trends, variant indicators, policy-related covariates, weekend reporting effects, vaccination

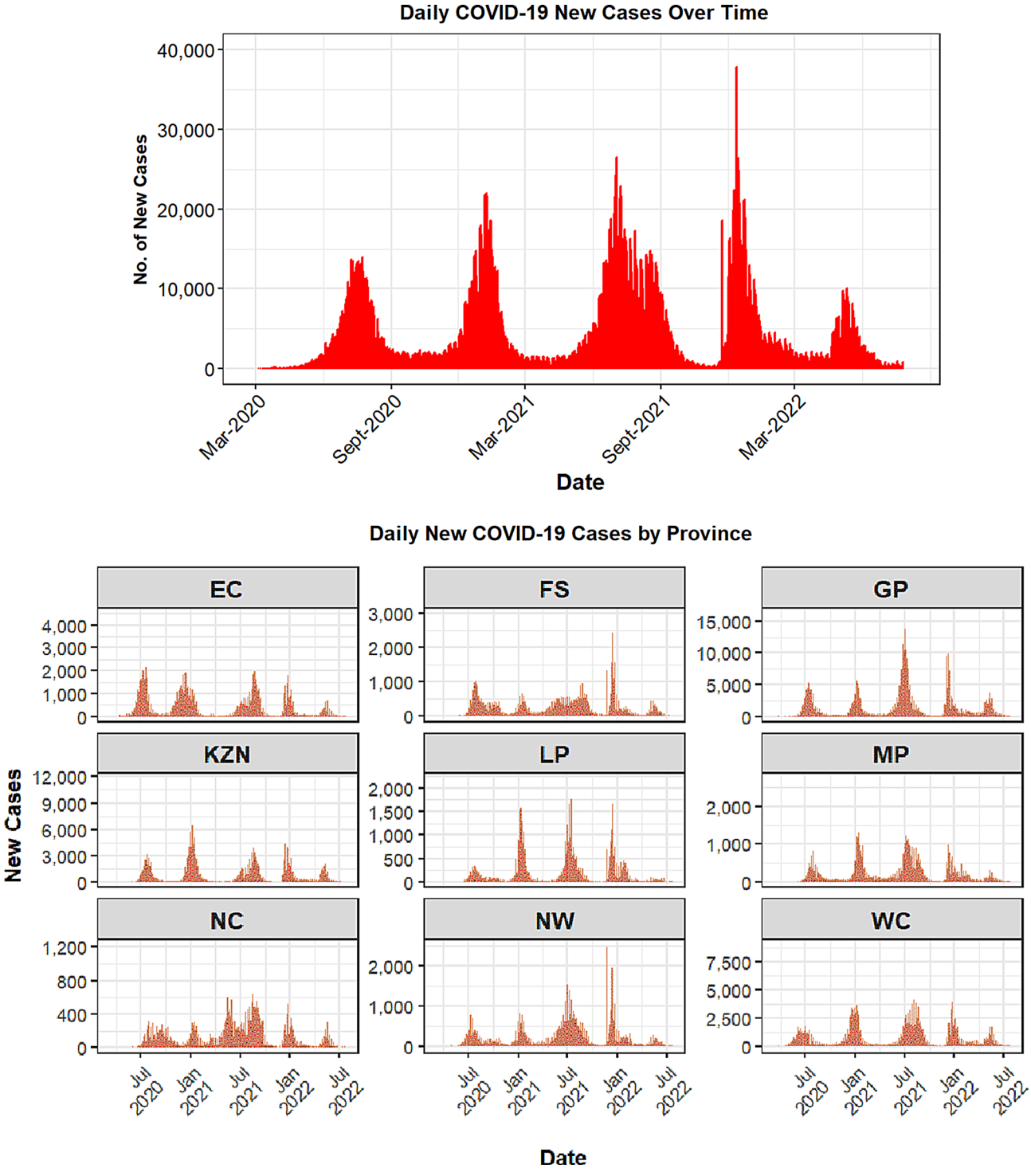

Fig. 1 COVID-19 case trajectories in South Africa at national and provincial levels

coverage, and province-specific random intercepts. As summarized in Table 1, the final model (Model 7) provided the strongest overall performance (AIC = 90,425.35, BIC = 90,725.18, MLS = 1243.45). Evidence of overdispersion in the case counts (dispersion estimate $\hat{\psi} = 0.6648$, 95% CI: 0.6439–0.6857) supported the use of a Negative Binomial distribution.

The estimated range of the dominant epidemic eigenvalue for Model 7 (0.45–1.22) indicated alternating periods of growth and containment, aligning with the observed pattern of repeated epidemic waves. Based on

Table 1 Model comparison using AIC, BIC, log score, and dominant eigenvalue

Model	AIC	BIC	Log Score	Eigenvalue Range
Model 1	97043.07	97070.89	6051.96	0.93
Model 2	91318.24	91491.92	1973.28	0.68–1.22
Model 3	91134.43	91370.91	1722.27	0.58–1.24
Model 4	90821.90	91074.14	1286.28	0.53–1.17
Model 5	90758.78	91015.78	1310.21	0.50–1.17
Model 6	90429.45	90705.48	1326.32	0.46–1.24
Model 7	90425.35	90752.18	1243.45	0.45–1.22

Table 2 Exponentiated coefficient estimates (rate ratios) with 95% confidence intervals for the final endemic-epidemic model

Covariate	Endemic	Autoregressive	Neighbourhood
	RR [95% CI]	RR [95% CI]	RR [95% CI]
Weekend	0.66 [0.58, 0.73]	0.80 [0.80, 0.90]	0.61[0.76, 1.03]
Vaccination (lag 7)	1.01[1.00, 1.01]	0.99 [0.99, 1.00]	1.00 [0.99, 1.01]
Stringency Index	1.03 [1.02, 1.04]	0.99 [0.99, 1.01]	–
Population	–	0.95 [0.91, 0.99]	5.62 [2.17, 14.55]
Beta variant	1.02[0.43, 2.38]	1.02 [0.97, 1.07]	–
Delta variant	1.34 [0.47, 3.77]	1.17 [1.11, 1.23]	–
Omicron variant	2.83[1.64, 4.86]	0.61 [0.54, 0.69]	–
Model Fit Metrics			
AIC = 90425.35 BIC = 90725.18 Mean Logarithmic Score (logS) = 1243.45			

Note: RR = rate ratio; CI = confidence interval. Values are exponentiated coefficients from the final model. Population as a covariate was included only in the autoregressive and neighbourhood components

its predictive strength and epidemiological coherence, Model 7 was selected as the basis for subsequent inference and forecasting.

Interpretation of the final model

Model 7 was selected based on its strong balance of fit, interpretability, and epidemiological relevance. It produced the lowest AIC (90,425.35), BIC (90,725.18), and mean log score (MLS = 1243.45) among all candidate models. The dominant epidemic eigenvalue ranged from 0.45 to 1.22, suggesting fluctuating but largely controlled transmission patterns across provinces. The model converged successfully using the `Nelder-Mead` optimizer, with stable estimates for both fixed effects and variance parameters. Table 2 summarizes the fixed-effect estimates for the endemic, autoregressive, and neighbourhood components. To make the results easier to interpret, it is helpful to explain what the regression coefficients mean in the endemic-epidemic model. The estimates are given on a logarithmic scale, and converting them into rate ratios illustrates how the expected case counts change when a covariate increases by one unit, assuming all other terms in the model remain constant. This interpretation is similar to what is used in standard Poisson or Negative Binomial regression, and differs from the log-odds ratios seen in logistic regression.

Autoregressive component (AR)

Local transmission was highest during the Delta wave and significantly lower during Omicron dominance. Population size (log scale) was inversely related to autoregressive transmission, while weekend indicators were associated with reduced case counts, likely due to reporting delays. The stringency index and lagged vaccination coverage showed limited influence in this component.

Neighbourhood component (NE)

Spatial spread was primarily driven by population size, reinforcing the role of densely populated provinces as sources of interprovincial transmission. Weekend effects again appeared to reduce apparent transmission, and neither vaccination coverage nor variant indicators had a strong influence on spatial spread in this component.

Endemic component (END)

Omicron dominance was associated with a significant increase in background transmission, while Beta and Delta showed weaker, non-significant effects. Both the stringency index and vaccination coverage were positively associated with endemic incidence, possibly reflecting increased testing or indirect behavioral factors during more intense policy periods.

This component-specific structure reveals that covariates exert different effects depending on the route of transmission, reinforcing the importance of decomposing transmission pathways in spatial-temporal models. For example, in our results, the Omicron indicator shows different effects across components. In the autoregressive component (local transmission), the estimated rate ratio is 0.61, indicating that, during the Omicron-dominant period, the influence of recent local cases on new infections was about 39% lower than in periods without a dominant variant. In the endemic component (background risk), the corresponding rate ratio is approximately 2.83, indicating that background transmission during the Omicron period was nearly three times higher than during the baseline period. These further illustrate that covariates can influence each component of the model in distinct ways, underscoring the value of decomposing transmission into endemic, local, and spatial contributions.

Seasonality

Three seasonal harmonics were included to allow flexible within-year variation in the endemic component. As reported in Table 3 and visualized in Fig. 2, the estimated pattern displays small rise around mid January, a modest rise in late March-early April and a pronounced seasonal peak around September-November. These fluctuations reflect data-dnotion that seasonality plays a significant role rather than direct alignment with school

Table 3 Seasonal harmonic estimates ($S=3$) in the endemic component

Term	Estimate	Std. Error	95% CI Lower	95% CI Upper
Amplitude A_1	3.1310	0.2109	2.7178	3.5442
Phase shift φ_1	1.6530	0.0503	1.5540	1.7520
Amplitude A_2	2.5590	0.1133	2.3369	2.7811
Phase shift φ_2	−1.8930	0.0411	−1.9735	−1.8125
Amplitude A_3	1.0250	0.0902	0.8479	1.2021
Phase shift φ_3	0.7884	0.0585	0.6737	0.9031

Note: Amplitudes and phase shifts describe the seasonal variation captured using a third-order Fourier series in the endemic component. Phase shift is in radians and governs peak timing within the calendar year

breaks, festive periods or lockdown intervals, which were not expected to coincide with the harmonic peaks. Incorporating these harmonics improved overall model fit and allowed the endemic component to capture broad seasonal patterns in background transmission.

Random effects

Province-specific random intercepts were estimated for all three components results are presented in Table 4. The neighbourhood component showed the highest variability ($\hat{\sigma}^2 = 2.03$), indicating substantial differences in provinces' likelihood to seed infections in other regions. The endemic component displayed moderate variation ($\hat{\sigma}^2 = 0.88$), while the autoregressive component was more uniform ($\hat{\sigma}^2 = 0.003$), suggesting consistent within-province dynamics once fixed effects were accounted for.

Table 5 provides province-level random intercept estimates by model component. Gauteng and Western Cape had the highest endemic effects, likely reflecting their large urban populations and greater connectivity. These provinces, along with KwaZulu-Natal, also exhibited the strongest neighbourhood random effects, indicating higher spatial export potential. Limpopo and Northern Cape consistently showed the lowest values in both components, consistent with more rural populations and lower mobility.

Table 4 Estimated variance of random intercepts for each model component

Component	Parameter	Variance Estimate
Autoregressive (AR)	σ^2_{ar}	0.0029
Neighbourhood (NE)	σ^2_{ne}	2.0263
Endemic (END)	σ^2_{end}	0.8763

Note: Variance estimates reflect unobserved heterogeneity across provinces after accounting for fixed covariates. Spatial variability was highest in the neighbourhood component

Table 5 Estimated province-specific random effects by model component (model 7)

Province	Endemic RE	AR RE	NE RE
Gauteng (GP)	1.20	0.05	2.40
Western Cape (WC)	0.90	0.02	1.80
KwaZulu-Natal (KZN)	0.75	0.01	1.50
Eastern Cape (EC)	0.60	0.00	0.95
Mpumalanga (MP)	0.55	0.00	0.80
Free State (FS)	0.48	−0.01	0.70
North West (NW)	0.45	−0.02	0.68
Limpopo (LP)	0.30	−0.01	0.35
Northern Cape (NC)	0.25	0.00	0.20

Note: Higher random effect values reflect greater deviation from baseline log-risk in each component

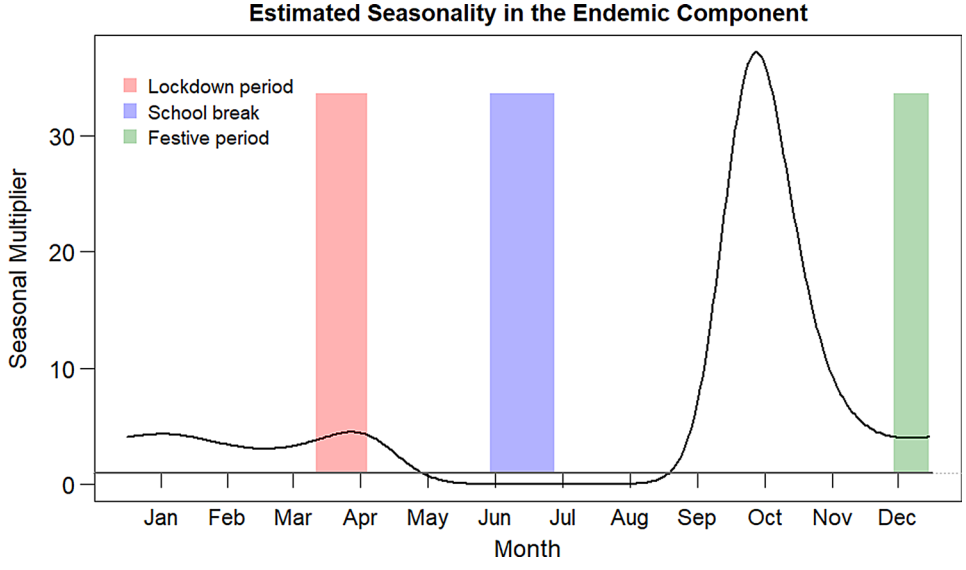

Fig. 2 Estimated seasonality in the endemic component over the calendar year. This figure visually conveys the seasonal timing and magnitude across the year. These seasonal patterns align with behavioral and climatic events (e.g., holidays, school closures), highlighting the importance of time-varying background risk in endemic transmission

Autoregressive random effects showed little variation across provinces, consistent with the low estimated variance for this component. Figure 3 displays the spatial distribution of province-specific random intercepts, where darker shading represents stronger deviations from the baseline log-risk.

To evaluate the performance of the final model, we compared fitted case counts to the observed data across all nine South African provinces. As shown in Fig. 4, the fitted values (in blue and orange) closely follow the observed daily case trends (black), particularly during peak transmission periods. The model accurately captures both the timing and magnitude of major waves while accommodating provincial-level variation. Minor deviations during low-incidence periods are expected in dynamic models. This visual assessment complements the favorable AIC, mean log score, and epidemic eigenvalue results, confirming that Model 7 reliably

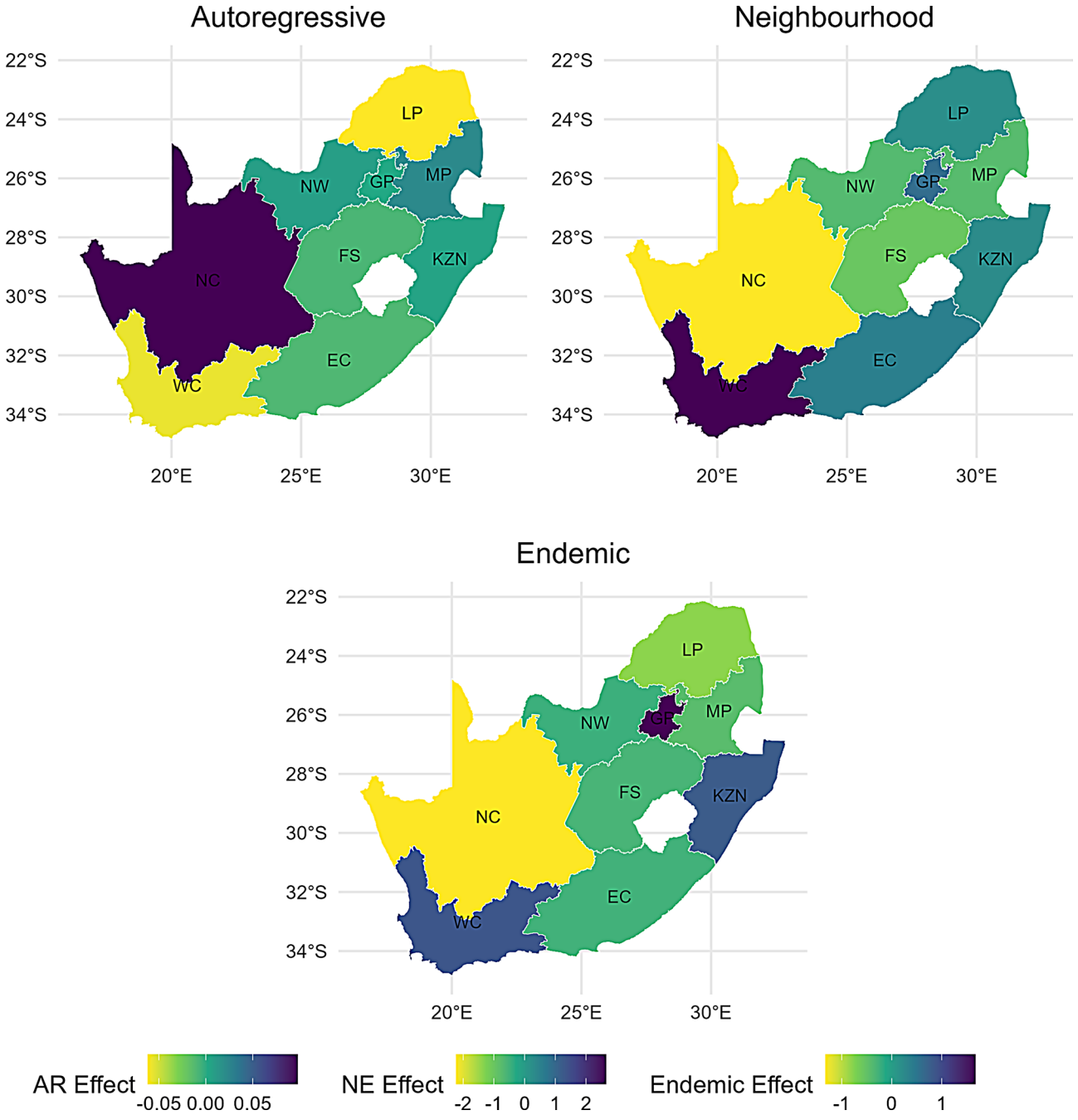

Fig. 3 Spatial distribution of province-level random effects for the AR, NE, and endemic components. Darker shades represent higher deviations from the component-specific baseline

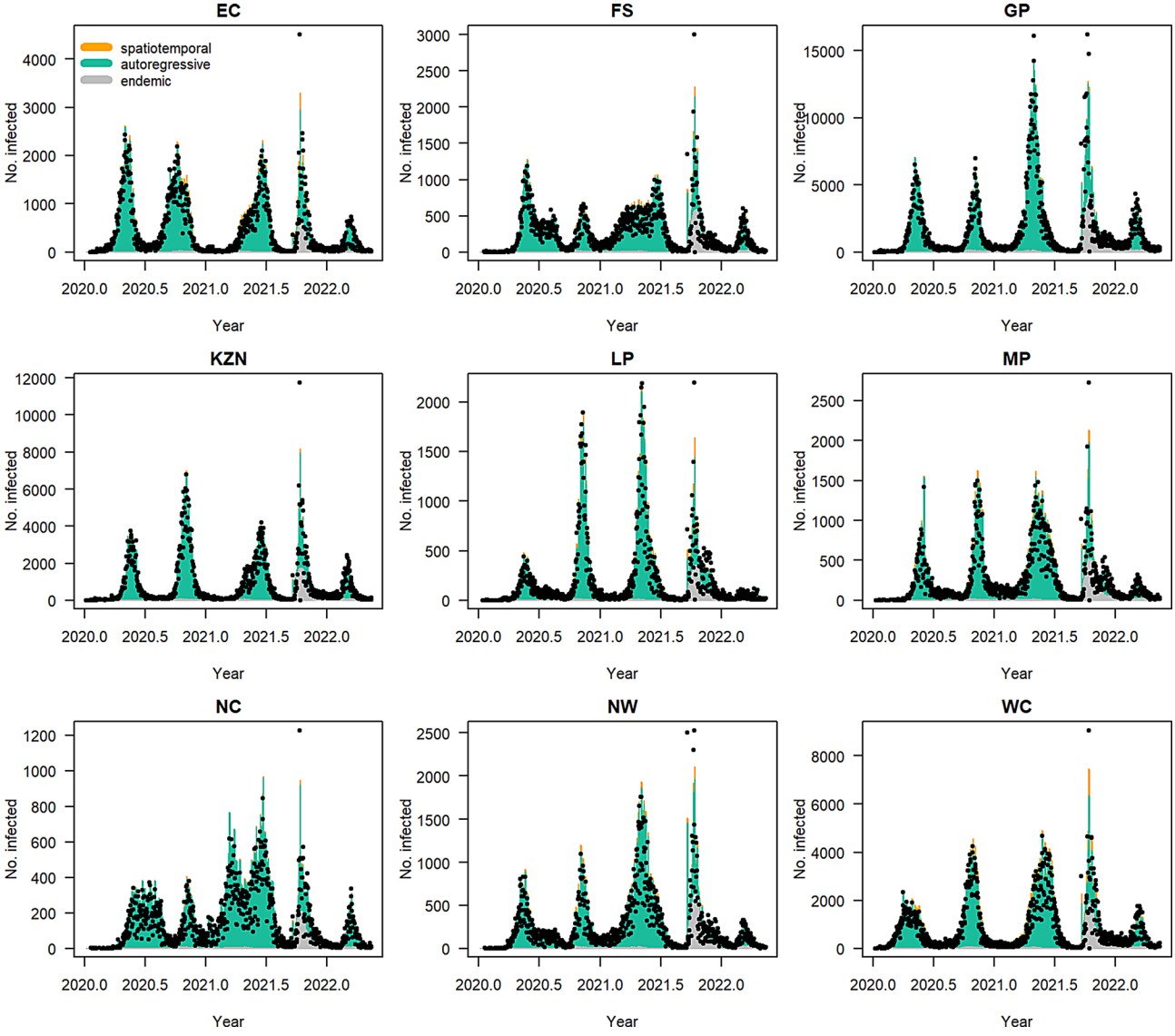

Fig. 4 Observed and fitted COVID-19 cases by province based on Model 7. This figure compares observed and fitted daily cases across provinces. Model 7 accurately captured major wave peaks and their timing, validating its use for short-term forecasting

Table 6 Proportional contribution (%) of each model component to total predicted COVID-19 cases by province

Province	Endemic (%)	AR (%)	NE (%)
Eastern Cape (EC)	7.30	86.80	5.80
Free State (FS)	11.10	84.70	4.10
Gauteng (GP)	13.90	83.20	2.90
KwaZulu-Natal (KZN)	10.90	85.60	3.50
Limpopo (LP)	9.70	82.80	7.60
Mpumalanga (MP)	10.70	85.00	4.30
Northern Cape (NC)	7.10	91.70	1.20
North West (NW)	12.70	83.20	4.10
Western Cape (WC)	11.30	80.00	8.70
South Africa	**11.50**	**83.80**	**4.70**

reproduces the observed spatio-temporal transmission patterns of COVID-19 in South Africa. These results support its use in short-term forecasting and policy evaluation for multiwave epidemic settings.

Component contribution to province-level transmission

Table 6 shows the proportional contributions of the endemic, autoregressive (AR), and neighbourhood (NE) components to the total expected cases by province. Across all provinces, local transmission (AR) accounted for the largest share, with a national average of 83.8%. The NE component contributed most prominently in the Western Cape (8.7%) and Limpopo (7.6%), while Gauteng had the highest endemic share (13.9%). Northern Cape

had the most localized profile, with 91.7% of predicted cases attributed to within-province transmission.

Figure 5 illustrates the geographic variation in dominant transmission modes across provinces. In provinces like the Western Cape, the contribution of all three components was more balanced. In contrast, Northern Cape and Mpumalanga were strongly dominated by local transmission dynamics, with minimal contribution from spatial or background sources.

Internal validation and forecasting performance

To evaluate the predictive performance of the final model (Model 7), we carried out both internal validation and simulation-based forecasting using observed case counts withheld from model fitting.

Validation of short-term predictions

The model's ability to predict COVID-19 trends was tested over two distinct windows: Days 770–784 (in-sample validation) and Days 849–862 (true hold-out). As

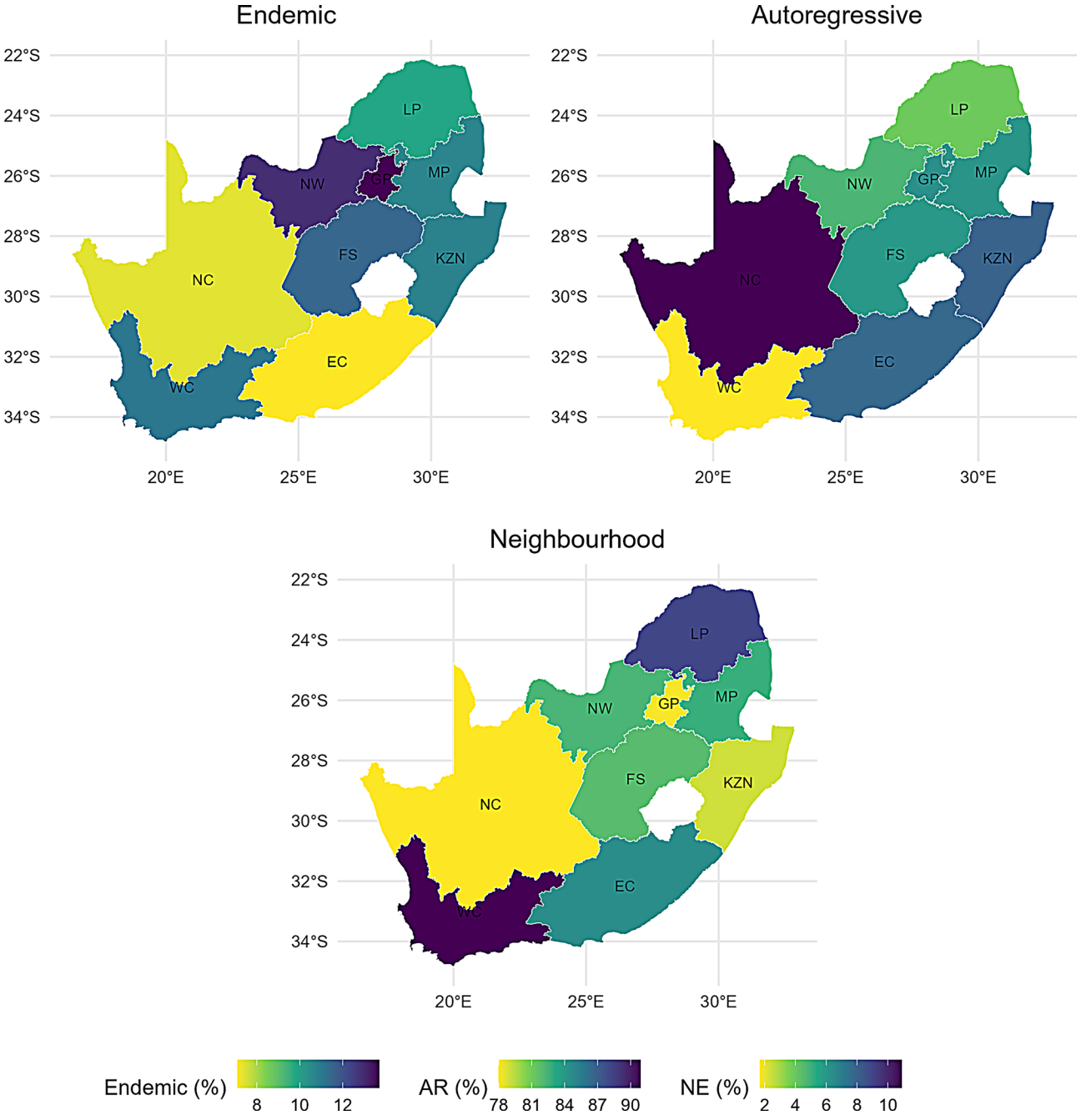

Fig. 5 Spatial distribution of the percentage contribution of each model component to total expected incidence across provinces

shown in Fig. 6, predicted case counts closely matched observed values in most provinces. In high-incidence provinces such as Gauteng and KwaZulu-Natal, the model tracked daily trends accurately for several days, while in provinces with smaller outbreaks (e.g., Free State, Limpopo), prediction intervals widened appropriately, reflecting greater uncertainty.

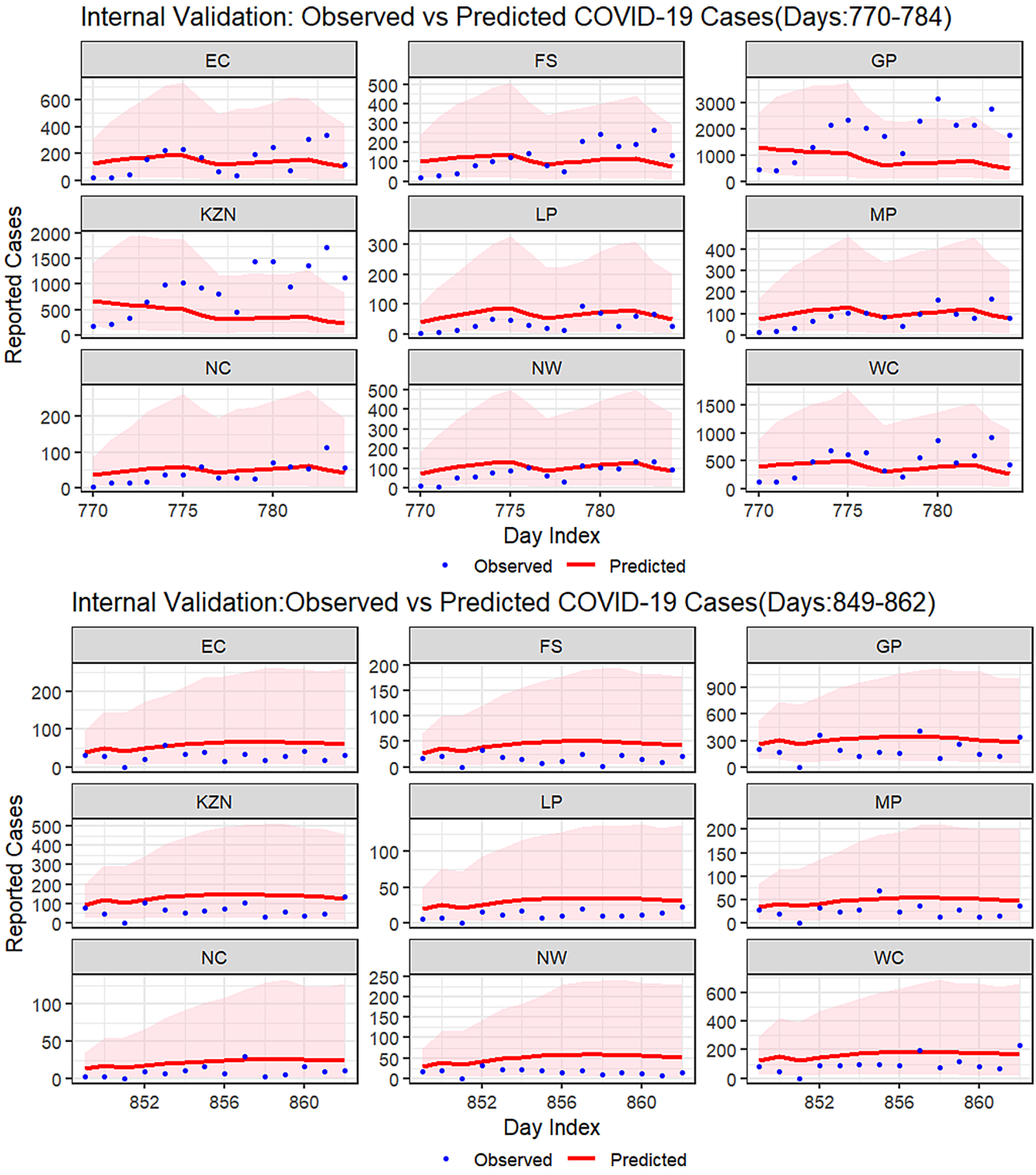

Fig. 6 Comparison of observed and predicted COVID-19 cases across South African provinces for two validation periods. Top: internal validation over Days 770–784 using model-fitted data. Bottom: hold-out validation over Days 849–862, not used in model training. Shaded areas represent 95% prediction intervals

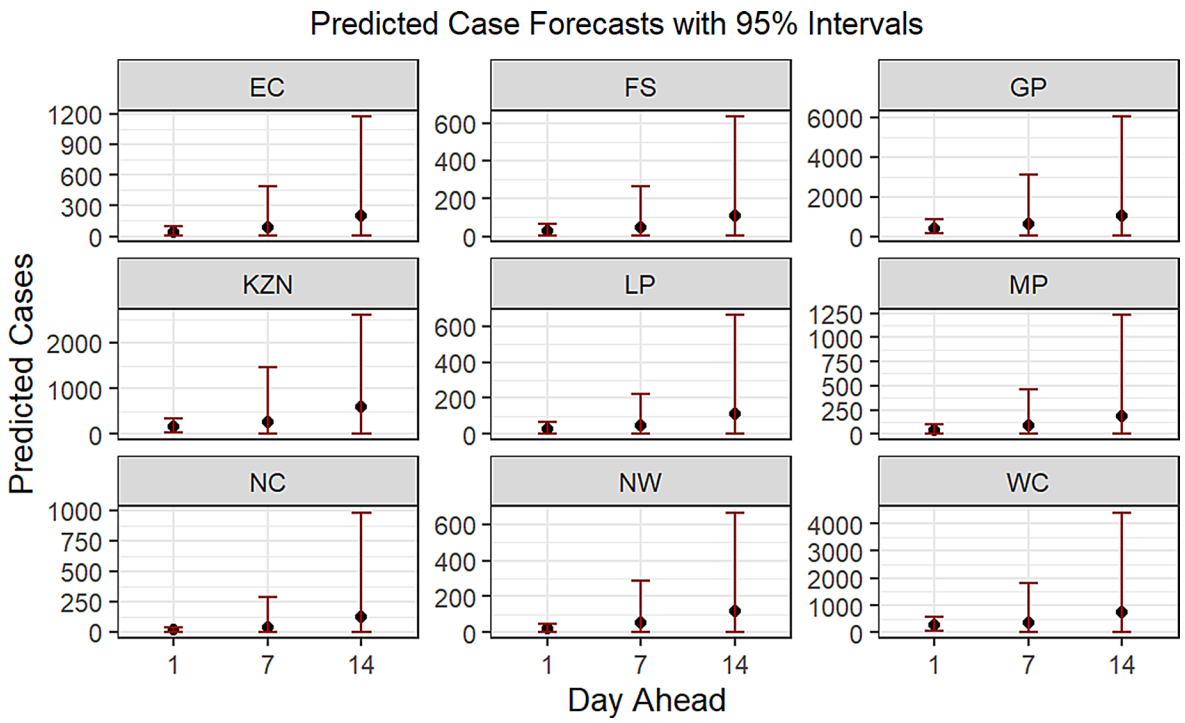

Fig. 7 Forecasted COVID-19 case counts across provinces for 1, 7, and 14 days ahead, based on stochastic simulations from the final model. Shaded bands represent 95% prediction intervals. Forecast uncertainty increases with horizon, especially in lower-incidence provinces

Forecast error remained low: root mean square error (RMSE) and mean absolute error (MAE) were 63.12 and 30.78 during Days 770–784, and 63.83 and 32.50 during Days 849–862, respectively, each lower than the overall standard deviation of 88.5.

Forecasting and uncertainty visualization

Figure 7 presents simulated predictions for Days 1, 7, and 14 ahead. Across all provinces, forecast intervals widened with longer horizons. In provinces with high transmission (e.g., Western Cape), the model remained sharply calibrated, while more variable predictions were observed in low-incidence settings (e.g., Northern Cape).

The visual consistency between predicted intervals and observed data supports the model's use for short-term forecasting, with well-quantified uncertainty. These results further reinforce the model's suitability for epidemic monitoring and real-time planning.

Discussion

This study applied a spatio-temporal endemic–epidemic modeling framework to characterize COVID-19 transmission dynamics across nine South African provinces. The model decomposed observed case counts into three transmission components: background risk (endemic), within-province transmission (autoregressive, AR), and cross-province transmission (neighbourhood, NE). Covariates such as SARS-CoV-2 variant dominance, government stringency measures, vaccination coverage, and seasonal trends were incorporated. Province-specific random effects and power-law spatial weights were used to reflect heterogeneity in baseline risk and interprovincial connectivity.

The final model demonstrated good fit and predictive performance, with the AR component consistently accounting for the majority of cases across provinces often exceeding 70%, indicating that sustained local transmission was the primary driver of epidemic dynamics. This is consistent with findings from Rwanda and Belgium, where similar endemic-epidemic models also identified within-region transmission as the dominant mechanism [11, 21]. The neighbourhood component varied notably by province: Western Cape and Limpopo exhibited higher NE contributions, suggesting stronger spatial connectivity or population movement to and from these regions, while provinces such as Northern Cape and Mpumalanga displayed more self-contained transmission patterns. Variant-specific effects were also component-dependent. The Omicron variant was associated with a reduction in AR transmission (rate ratio = 0.61), possibly reflecting its rapid geographic spread and lower generation intervals, but simultaneously led to elevated background risk (rate ratio ≈2.83). In contrast, Delta modestly increased AR transmission (rate ratio = 1.17), aligning with its higher viral load and transmissibility. Interestingly, vaccination and stringency measures had limited effects on case dynamics, which may reflect uneven rollout, behavioral offsets, or the impact of

immune-evasive variants. Vaccination coverage, included as $(1 - \text{coverage})$ in all model components, showed little measurable effect on transmission. This may reflect uneven rollout, limited effectiveness against transmission during Omicron, or offsetting behavioral changes. To test the impact of lag assumptions, we conducted a sensitivity analysis using 7-, 14-, and 21-day delays. Model estimates and fit metrics (AIC, BIC, and mean log score) remained nearly unchanged, confirming that our conclusions were not sensitive to the choice of lag structure.

These findings suggest that within-province persistence was the dominant mechanism driving COVID-19 spread in South Africa. This highlights the importance of timely, locally targeted interventions, such as mask mandates, community-specific lockdowns, and improved access to testing, to contain outbreaks before they amplify across regions. In provinces with higher spatial influence, coordinated cross-provincial planning may also be essential. The model's strong short-term predictive performance during internal validation supports its potential utility for early warning and real-time forecasting. As such, it can inform resource allocation and policy response during both ongoing and future epidemic waves.

Limitations and future work

This study has several limitations that should be acknowledged. First, the data were available at the provincial level, which may obscure important within-province heterogeneity in transmission dynamics and intervention effects. This spatial resolution, though necessary and a significant improvement from national-level modeling, due to data availability, limits the ability of the model to capture more localized outbreaks or fine-scale variation in seeding patterns. Previous studies have shown that using finer spatial units, such as districts or municipalities, can improve the spatial resolution of transmission dynamics and enhance model accuracy. For example [7], applied the endemic–epidemic framework to district-level measles data in Germany [8]; demonstrated the benefits of local-level modeling in a metapopulation context; and [11] modeled COVID-19 transmission across Belgian municipalities, incorporating spatial variation and travel data.

Second, the model employed static power-law spatial weights, which do not reflect temporal changes in population mobility due to behavioral shifts or policy interventions. As a result, the geographical spread of infection may have been misrepresented in periods with unusual mobility patterns, such as during lockdowns or festive seasons. Incorporating dynamic mobility data, such as updated travel patterns between regions or anonymised mobile phone location data, could make the spatial transmission component more realistic over time. This would help the model reflect changes in movement due to lockdowns, holidays, or other policy interventions [9].

Third, variant dominance was modeled using binary indicators based on predefined prevalence thresholds. Although this approach provided a practical classification of dominant strains, it likely oversimplifies real-world transitions, where multiple variants can co-circulate and shift gradually over time. Modeling variant dynamics as continuous proportions or with smoothing techniques could improve biological realism and allow more nuanced inference on immune escape or transmissibility.

Fourth, variant indicators were excluded from the neighbourhood (NE) component to ensure model convergence and avoid overparameterization. The NE component captures a relatively weak spatial signal compared to the autoregressive (AR) and endemic (END) components, and including multiple binary variant indicators led to instability, particularly for Omicron. Excluding these indicators avoids the incorrect assumption that the effects of the omitted variants are equal to the baseline. However, this also means that variant-specific effects on spatial transmission were not directly modeled. Their influence was instead incorporated through AR and END, where statistical estimation was more stable. This could be addressed in future work by using more robust spatial structures or incorporating hierarchical shrinkage priors to allow partial pooling of variant effects while preserving identifiability [20].

Finally, the model assumes a fixed reporting structure and does not explicitly adjust for changes in testing rates, reporting delays, or underreporting factors that are known to affect the accuracy of infectious disease surveillance data. These could be addressed in future work by integrating observation models or latent reporting processes, especially when data from different periods or provinces vary substantially in data quality.

From a methodological perspective, further extensions could include the use of fully Bayesian hierarchical models implemented through the *Integrated Nested Laplace Approximation* (INLA) framework, which allows flexible specification of random effects, incorporation of prior knowledge, and improved uncertainty quantification [20]. Another direction involves the *Time-series Susceptible-Infectious-Recovered* (TSIR) model, a semi-mechanistic approach that incorporates epidemiological structure and can better account for non-linear transmission dynamics, stochasticity, and the effects of interventions. Furthermore, modeling both case and mortality data jointly through a bivariate framework would enable simultaneous inference on transmission and severity outcomes, especially useful in monitoring the impact of emerging variants or healthcare burden.

In addition, the model tended to underestimate the very highest peaks in provinces such as Gauteng and

KwaZulu-Natal. These regions are highly urbanised, have large and mobile populations, and historically show more abrupt surges that are difficult to fully capture with surveillance-based models. This behaviour has also been noted in related modeling work, particularly when rapid increases occur over short periods Reddy et al. [5]. Overall, these limitations reflect constraints common in surveillance-based modeling during public-health emergencies, and they highlight areas where improved data streams and methodological extensions could strengthen future applications of the endemic–epidemic framework. Importantly, the methodological approach presented here is adaptable to other infectious diseases where localized transmission and spatial dynamics are central, including mpox, measles, and potential future coronavirus outbreaks.

Conclusion

This study shows how the endemic–epidemic modeling approach can be applied to follow and interpret the progression of COVID-19 across regions that are closely linked. By separating routine background infection pressure, short-range spread within a province, and transmission driven by neighbouring provinces, the framework offers a clearer understanding of how outbreaks develop and why certain waves intensify. Bringing information on policies, behaviour, and circulating variants into the model also helped capture changes in transmission during the course of the pandemic, making the framework suitable for ongoing situational awareness. The value of this approach goes beyond the methodological insight. In health systems where resources are stretched, and response capacity varies across locations, recognizing early signs of local resurgence or changes in dominant transmission patterns is crucial. The findings from this work illustrate that routinely collected surveillance data, when structured through this type of model, can support timely and well-focused public-health action. Although COVID-19 now circulates at lower levels, sporadic flare-ups continue to occur, and the possibility of new variants remains. Approaches that distinguish different transmission pathways and use data already available to public-health authorities can therefore play an important role in strengthening preparedness. Taken together, the analysis demonstrates that the endemic–epidemic framework is a practical and adaptable tool for tracking epidemic trends, informing locally relevant interventions, and supporting longer-term planning in South Africa and other settings where health systems are decentralized or resource-constrained.

Acknowledgements
The author would like to thank Prof Niel Hens and Dr Tarylee Reddy for their valuable supervision and feedback throughout this work. We also acknowledge the support from Hasselt University and the South African Medical Research Council (SAMRC) during the internship period.

Author contributions
R.A. conceptualized the study, conducted the analysis, interpreted the results, and drafted the manuscript. N.H. and T.R. supervised the study, contributed to model development, and critically revised the manuscript. All authors approved the final version of the manuscript.

Funding
This research was funded by the VLIR-UOS scholarship programme under the Master of Statistics and Data Science at Hasselt University.

Data availability
The datasets analyzed during the current study are publicly available from the following sources: Daily cumulative confirmed COVID-19 case counts: National Institute for Communicable Diseases (NICD) provincial dashboards and epidemiological briefs [13]. Vaccination data (individuals with at least one dose): South African Department of Health [14]. Government response stringency index: Oxford COVID-19 Government Response Tracker [15]. SARS-CoV-2 variant prevalence data: NICD genomic surveillance reports [16] and GISAID sequence database [17].All datasets used are publicly accessible. Derived and processed data as used in the models, along with R scripts, are available from the corresponding author upon reasonable request.

Materials availability
Not applicable.

Code availability
All code used to implement the endemic-epidemic models in R (using the `surveillance` package) is available upon request from the corresponding author.

Declarations

Ethics approval and consent to participate
Not applicable. This study uses publicly available and anonymized COVID-19 case data.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 7 October 2025 / Accepted: 6 February 2026
Published online: 21 February 2026

References

1. Scarpino SV, Petri G. On the predictability of infectious disease outbreaks. Nat Commun. 2019;10(1):898. https://doi.org/10.1038/s41467-019-08616-0.
2. Chitnis N, Custer B, Fish D, Sattar A. The challenges of modeling and forecasting COVID-19. Epidemics. 2020;33:100403. https://doi.org/10.1016/j.epidem.2020.100403.
3. Karim SSA, de Oliveira T. New SARS-CoV-2 variants—clinical, public health, and vaccine implications. N Engl J Med. 2021;384(19):1866–68. https://doi.org/10.1056/NEJMc2100362.
4. Pulliam JRC, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. Science. 2022;376(6593):eabn4947. https://doi.org/10.1126/science.abn4947.
5. Reddy TP, Reddy T, Shkedy Z, Janse van Rensburg C, Mwambi H, Debba P, Zuma K, et al. Short-term real-time prediction of total number of reported COVID-19 cases and deaths in South Africa: a data driven approach. BMC Med Res Methodol 2021;21(1):15. https://doi.org/10.1186/s12874-020-01165-x.
6. Schoeman H, et al. Forecasting hospital admissions using machine learning models in South Africa. BMC Med Inf Decis Mak. 2023;23(1):112. https://doi.org/10.1186/s12911-023-02265-7.

7. Held L, Höhle M, Hofmann M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. Stat Modell. 2005;5(3):187–99. https://doi.org/10.1191/1471082X05st098oa.
8. Paul M, Held L. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. Stat Med. 2011;30(10):1118–36. https://doi.org/10.1002/sim.4177.
9. Meyer S, Held L. Power-law models for infectious disease spread. Ann Appl Stat. 2014;8(3):1612–39. https://doi.org/10.1214/14-AOAS743.
10. Giuliani D, Dickson MM, Espa G, Santi F. Modelling and predicting the spread of coronavirus (COVID-19) infection in Northern Italy using spatial regression models. Reg Sci Policy Pract. 2020;12(6):1047–70. https://doi.org/10.1111/rsp3.12371.
11. Nguyen MH, et al. The impact of national and international travel on spatio-temporal transmission of SARS-CoV-2 in Belgium in 2021. BMC Infect Dis. 2023;23(1):428. https://doi.org/10.1186/s12879-023-08368-9.
12. Tuyisenge L, et al. Spatio-temporal dynamic of the COVID-19 epidemic and the impact of imported cases in Rwanda. BMC Public Health. 2023;23(1):939. https://doi.org/10.1186/s12889-023-16256-6.
13. National Institute for Communicable Diseases. Daily hospital surveillance reports. Available at: 2022. https://www.nicd.ac.za/diseases-a-z-index/covid-19/surveillance-reports/. Accessed 01 Dec 2023.
14. South African Department of Health. COVID-19 vaccination statistics. Available at: 2022. https://sacoronavirus.co.za/latest-vaccine-statistics/. Accessed 01 Dec 2023.
15. Hale T, et al.). Oxford COVID-19 government response tracker. Available at: 2020. https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker. Blavatnik School of Government.
16. National Institute for Communicable Diseases. SARS-CoV-2 genomic surveillance reports. Available at: 2022. https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/sars-cov-2-genomic-surveillance/. Accessed 01 Dec 2023.
17. GISAID Initiative. SARS-CoV-2 sequence database. Available at: 2022. https://www.gisaid.org/. Accessed 01 Dec 2023.
18. Statistics South Africa. Mid-year population estimates, 2020 (statistical release P0302). Available at: 2020. https://www.statssa.gov.za/publications/P0302/P03022020.pdf. Pretoria: Stats SA.
19. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csáki F, editors. Proc. 2nd int. Symp. Information theory. Budapest: Akadémiai Kiadó; 1973. p. 267–81.
20. Meyer S, Held L, Höhle M. Spatio-temporal analysis of epidemic phenomena using the R package surveillance. J Stat Softw 2017;77(11):1–55. https://doi.org/10.18637/jss.v077.i11
21. Uwineza A, Niyirora G, Uwimana A, Niyonzima A, Nsanzimana S, Mugunga JC, et al. Spatio-temporal modelling of COVID-19 incidence and transmission dynamics in Rwanda. Sci Afr. 2023;19:e01515.

Publisher's Note