Short-term forecasting of total Number of reported COVID-19 cases in South Africa - A Bayesian temporal modeling approch

Belay Birlie Yimer , , Ziv Shkedy

Abstract

To be updated.

Author summary

To be updated.

Introduction

To be updated

Methods

Data

We downloaded data from Coronavirus COVID-19 (2019-nCoV) Data Repository for South Africa maintained by Data Science for Social Impact research group at the University of Pretoria [ref]. The data repository captures the daily number of new cases, number of tests, number of deaths and recoveries. Our primary outcome of interest was the daily number of newly diagnosed COVID-19 cases and the unit of time used in modelling was a day. We used the daily case reports from March 12, 2020, until February 15, 2021, in our analysis.

Statistical analysis

We considered two widely used temporal models to model the daily number of newly diagnosed COVID-19 cases. We let $Y(t)$ denote the daily number of newly diagnosed COVID-19 cases at time t and $\mu(t)$ represent the expected number of cases at time t. We considered a Negative binomial distribution for $Y(t)$ to account for possible overdispersion. That is, $Y(t) \sim NB(\mu(t), \delta)$, where δ is the overdispersion parameter. We considered two temporal models to capture the trend over time: a random walk of order one ($RW(1)$) and an autoregressive model of order one ($AR(1)$) [ref]. The two models were chosen becuse

The $AR(1)$ model [1] is given by,

$$
\begin{aligned}
Y(t) &\sim NB(\mu(t), \delta) \quad t = 1, \ldots, n, \\
log(\mu(t)) &= \alpha + u_t, \\
u_1 &\sim N(0, \tau_u(1-\rho^2)^{-1}), \\
u_t &= \rho u_{t-1} + \epsilon_t, \quad t = 2, \ldots, n, \\
\epsilon_t &\sim N(0, \tau_\epsilon),
\end{aligned}
$$

where, α is an intercept, ρ a temporal correlation term (with $|\rho| < 1$) and ϵ_t is a Gaussian error term with zero mean and precision τ_u.

Similarly, the $RW(1)$ model [1] is given by,

$$
\begin{aligned}
Y(t) &\sim NB(\mu(t), \delta) \quad t = 1, \ldots, n, \\
log(\mu(t)) &= \alpha + u_t, \\
u_t - u_{t-1} &\sim N(0, \tau_u), \quad t = 2, \ldots, n,
\end{aligned}
$$

where α is the intercept term as before and τ_u is the precison parameter.

The two models were fitted within the Bayesian framework using *inla* [2]. To complete the specification of both models, we assume the following priors. For the $AR(1)$ model, we denote $\theta_1 = log(\tau_u(1-\rho^2))$ where $\Gamma(10, 100)$ prior is specified for θ_1, and we denote $\theta_2 = log\frac{1+\rho}{1-\rho}$ and assume a $N(0, 0.15)$ prior for θ. Similarly, we represent the precison parameter of $RW(1)$, τ_u, as $\theta = log(\tau_u)$ and assume a $\Gamma(10, 100)$ prior for θ. To assess the models' accuracy in predicting cases, we present the forecast period's actual observed values and the predicted values.Additionally, the model fits were evaluated by using DIC (Deviance information criteria). The computer code that we used for our analyses is avaliable at https://github.com/belayb/COVIDincidenceSA/tree/master/COVIDincidenceSA.

Results

Figure 1 presents the daily number of reported COVID-19 cases from 12 March 2020 to 19 February 2021. Similar to elsewhere in the world, South Africa pass through a two-wave pandemic. The pandemic's first peak was on 07 July 2020, where up to 13944 new COVID-19 cases reported, followed by a second peak in January 2021, where more than 21,000 daily cases reported. Figure 2 presents the cumulative number of new reported COVID-19 cases and tests performed. To date, 8,838,937 tests have been conducted, and a total of 1,500,677 cases reported.

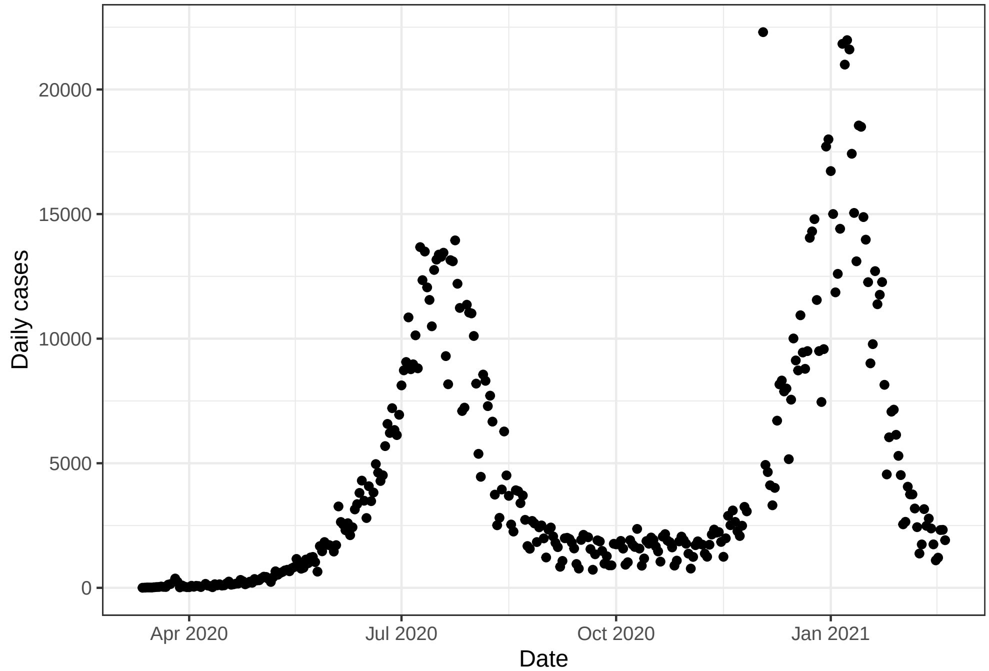


Fig 1. Daily number of COVID-19 cases in South Africa from 12/03/2020-19/02/2021.

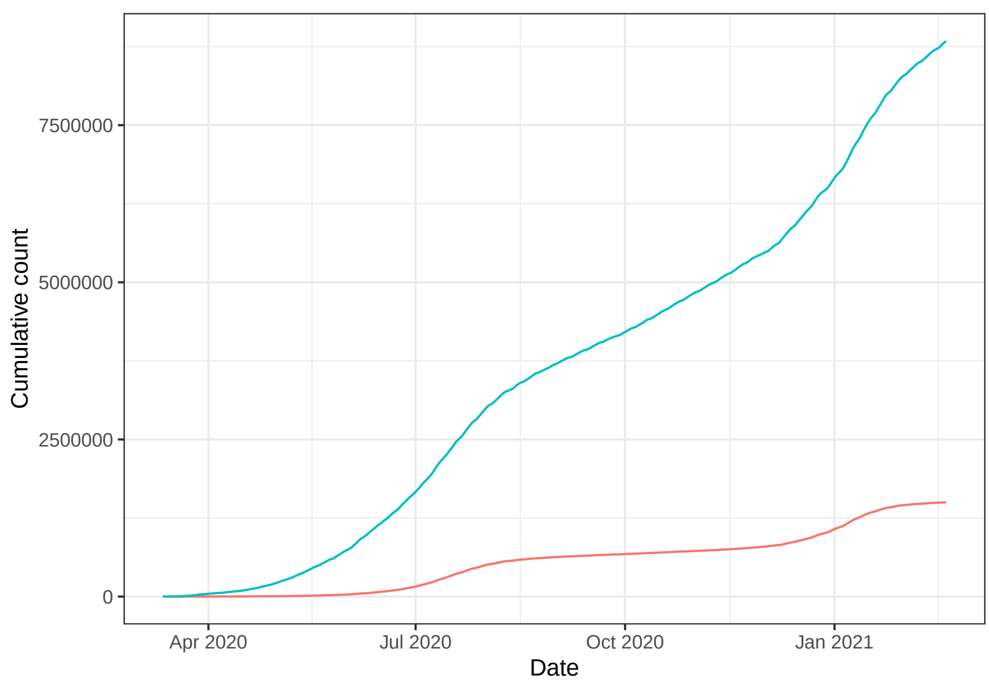
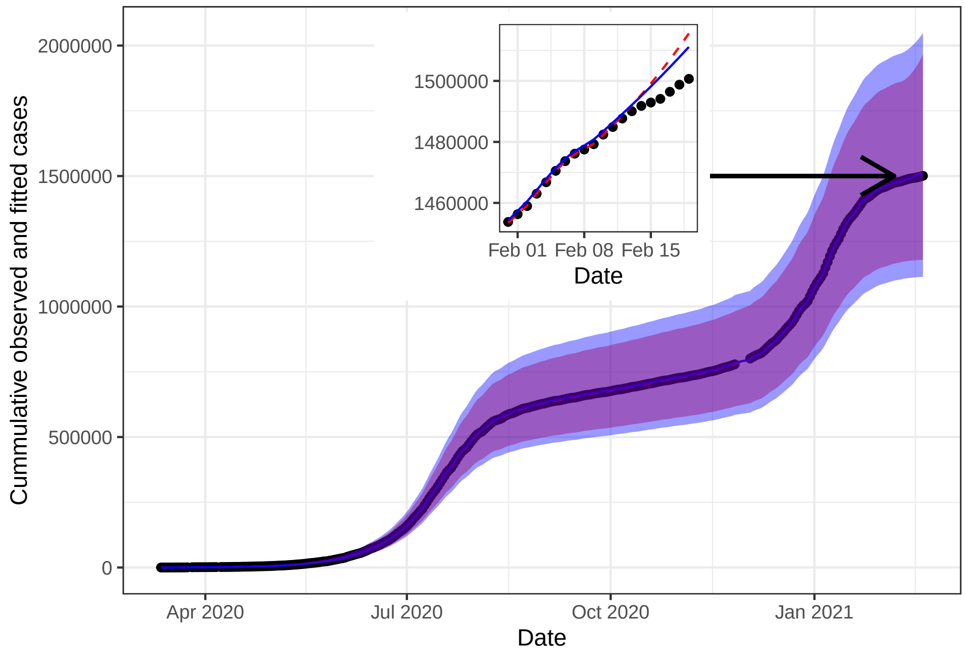


Fig 2. The cummulative number of COVID-19 cases and Cummulative number of tests in South Africa from 12/03/2020-19/02/2021. Red-line denote the number of cases and blue-line denotes the number of tests.

Short-term prediction of the total number of reported COVID-19 cases

We fit the two models described in the previous section to the daily reported new COVID-19 cases. The parameter estimates for the two models are presented in Supplementary Table 1. As depicted in Figure 3, two models fitted to the data appear to fit the observed data (within the estimation period) well with a narrow confidence interval obtained for the RW(1) model. The RW(1) and AR(1) models also provide very similar predictions over the 9-day ahead period.

\begin{figure}[H]



\caption{Predicted cummulative COVID-19 cases in South Africa under the RW(1) and AR(1) model. Estimation period 12/03/2020-10/02/2021. The black dots are the observed cummulative cases. The red dashed lines are for RW(1) and the blue line for AR(1).The shaded bands are the 95% prediction intervals for the predictions with the two models.} \end{figure}

Table 1. Short-term predictions of total number of reported cases at the national level under the RW1 model. Estimation period 12/03/2020-10/02/2021

Date	Total	Prediction	Prediction Interval	Total - Prediction
2021-02-11	1484900	1485429	(1172894.86-1870619.43)	-528.5228
2021-02-12	1487681	1488629	(1174130.71-1877572.01)	-947.5700
2021-02-13	1490063	1491976	(1175163.66-1885902.75)	-1912.5901
2021-02-14	1491807	1495476	(1176050.29-1895617.69)	-3669.3800
2021-02-15	1492909	1499138	(1176824.75-1906749.13)	-6229.1891
2021-02-16	1494119	1502969	(1177509.82-1919343.11)	-8849.6675
2021-02-17	1496439	1506976	(1178121.45-1933454.25)	-10536.8634
2021-02-18	1498766	1511168	(1178671.6-1949143.78)	-12402.2333
2021-02-19	1500677	1515555	(1179169.59-1966478.16)	-14877.6590

Table 2. Short-term predictions of total number of reported cases at the national level under the AR1 model. Estimation period 12/03/2020-10/02/2021

Date	Total	Prediction	Prediction Interval	Total - Prediction
2021-02-11	1484900	1486312	(1107420.98-1978162.82)	-1412.194
2021-02-12	1487681	1489173	(1108639.61-1983981.29)	-1491.764
2021-02-13	1490063	1492102	(1109681.58-1990659.34)	-2038.527
2021-02-14	1491807	1495099	(1110591.96-1998160.29)	-3291.921
2021-02-15	1492909	1498166	(1111398.75-2006463.55)	-5256.654
2021-02-16	1494119	1501303	(1112121.57-2015558.08)	-7183.538
2021-02-17	1496439	1504510	(1112774.74-2025437.2)	-8071.444
2021-02-18	1498766	1507790	(1113369.03-2036097.59)	-9024.272
2021-02-19	1500677	1511143	(1113912.84-2047538.12)	-10465.946

References

1. Gómez-Rubio V. Bayesian inference with inla. CRC Press; 2020.

2. Martins TG, Simpson D, Lindgren F, Rue H. Bayesian computing with inla: New features. Computational Statistics & Data Analysis. Elsevier; 2013;67: 68–83.