Short-term real-time prediction of total number of reported COVID-19 cases in South Africa - A Bayesian Temporal Modeling Approch

Belay Birlie Yimer [,*], Ziv Shkedy

1 Department 1, Street, City, State, Zip
2 Department 2, Street, City, State, Zip

* Corresponding author: belaybirlie.yimer@manchester.ac.uk

Abstract

To be updated.

Author summary

To be updated.

Introduction

To be updated Here are two sample references: [1,2].

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 [ref] 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 [ref] 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* [ref]. 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.

Results

Figure 1 (Top panel) presents the daily number of reported COVID-19 cases from 5 March 2020 to 15 February 2021. Similar to elsewhere in the world, South Africa pass through a two-wave pandemic. The first peak of the epidemic was on —-, followed by a second peak in January 2021. Figure 1 (Bottom Panel) presents the cumulative number of new reported COVID-19 cases and tests performed. To date, 1,353,176 tests have been conducted, corresponding to a testing rate of 22.816 per 1000 population. There was a significant correlation between the number of cases detected and the number of tests performed daily (Rho = 0.7759, p-value$<$ 0.001).

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

The models described in the previous section were all considered for modeling cumulative cases.The parameter estimates for the different models are presented in Supplementary Table 1. As mentioned in the previous section, our main interest is to produce a short term forecast for the number of reported cases and deaths. As depicted from the short-term forecasts for the three models fitted to cases (see Fig. 5, Table 2), all three models appear to fit the observed data (within the estimation period) well with the 3 parameter and 4 parameter logistic models providing very similar predictions over the 30-day ahead period.

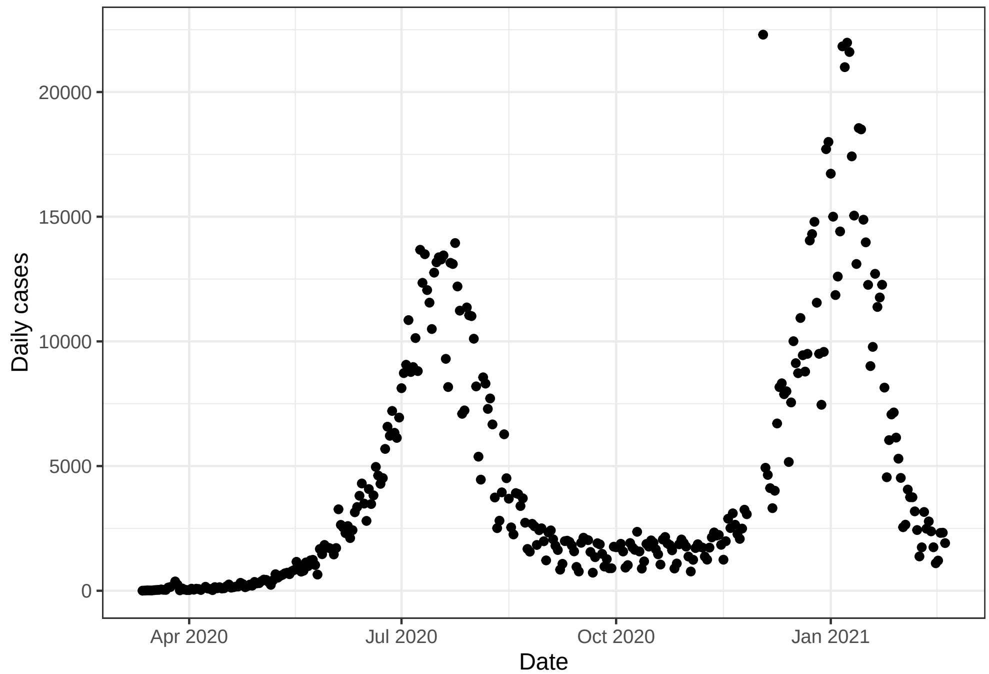

Fig 1. Daily new COVID-19 cases.

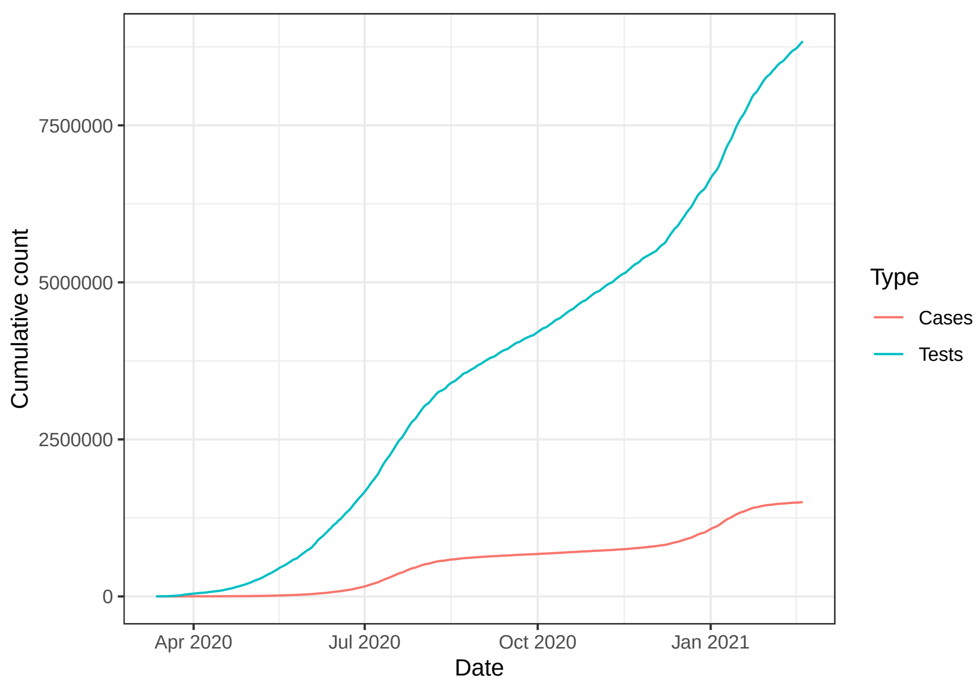

Fig 2. cummulative cases and cummulative tests.

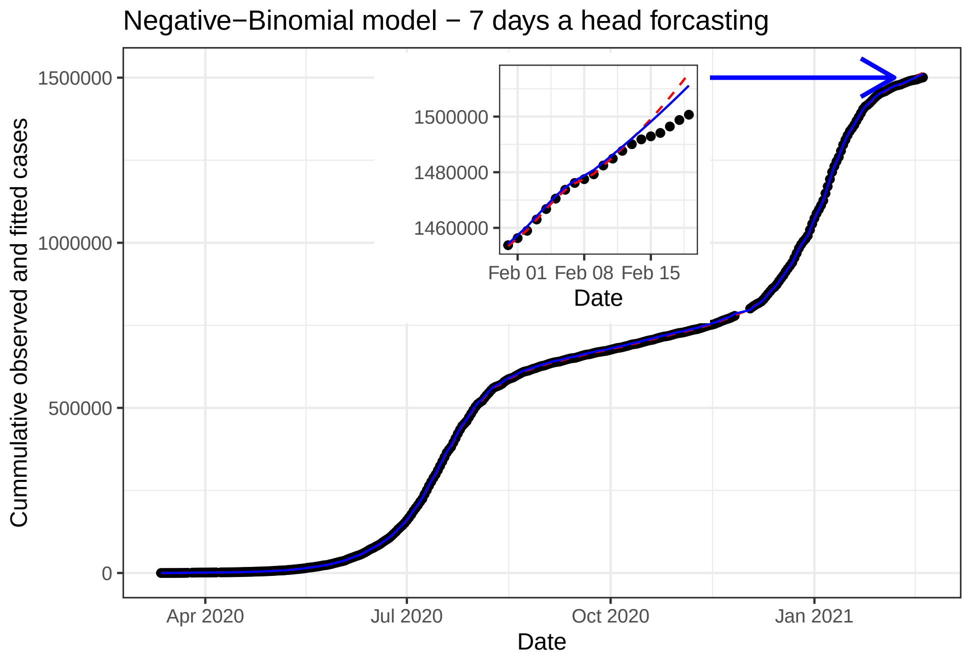

Fig 3. Fitted and observed data

Table 1. Short-term predictions of total number of reported cases at the national level under the RW1 model. Estimation period 05/03/2020-08/02/2021

	Date	Total	Prediction	Prediction Interval
329	2021-02-11	1484900	1485425	(1172933.36-1870556.26)
330	2021-02-12	1487681	1488625	(1174169.39-1877507.47)
331	2021-02-13	1490063	1491972	(1175202.53-1885836.24)
332	2021-02-14	1491807	1495472	(1176089.35-1895548.55)
333	2021-02-15	1492909	1499134	(1176864-1906676.65)
334	2021-02-16	1494119	1502964	(1177549.26-1919266.52)
335	2021-02-17	1496439	1506970	(1178161.06-1933372.71)
336	2021-02-18	1498766	1511162	(1178711.39-1949056.38)
337	2021-02-19	1500677	1515548	(1179209.55-1966383.94)

Table 2. Short-term predictions of total number of reported cases at the national level under the AR1 model. Estimation period 05/03/2020-08/02/2021

	Date	Total	Prediction	Prediction Interval
329	2021-02-11	1484900	1486312	(1107420.98-1978162.82)
330	2021-02-12	1487681	1489173	(1108639.61-1983981.29)
331	2021-02-13	1490063	1492102	(1109681.58-1990659.34)
332	2021-02-14	1491807	1495099	(1110591.96-1998160.29)
333	2021-02-15	1492909	1498166	(1111398.75-2006463.55)
334	2021-02-16	1494119	1501303	(1112121.57-2015558.08)
335	2021-02-17	1496439	1504510	(1112774.74-2025437.2)
336	2021-02-18	1498766	1507790	(1113369.03-2036097.59)
337	2021-02-19	1500677	1511143	(1113912.84-2047538.12)

References

1. Feynman RP, Vernon Jr. FL. The theory of a general quantum system interacting with a linear dissipative system. Annals of Physics. 1963;24: 118–173. doi:10.1016/0003-4916(63)90068-X

2. Dirac PA. The lorentz transformation and absolute time. Physica. 1953;19: 888–896. doi:10.1016/S0031-8914(53)80099-6