Short-term forecasting of total number of reported COVID-19 cases in South Africa - A Bayesian temporal modelling approach

Belay B. Yimer and Ziv Shkedy

Centre for epidemiology Versus arthritis

University of Manchester

SUSAN Conference 2021, Kenya

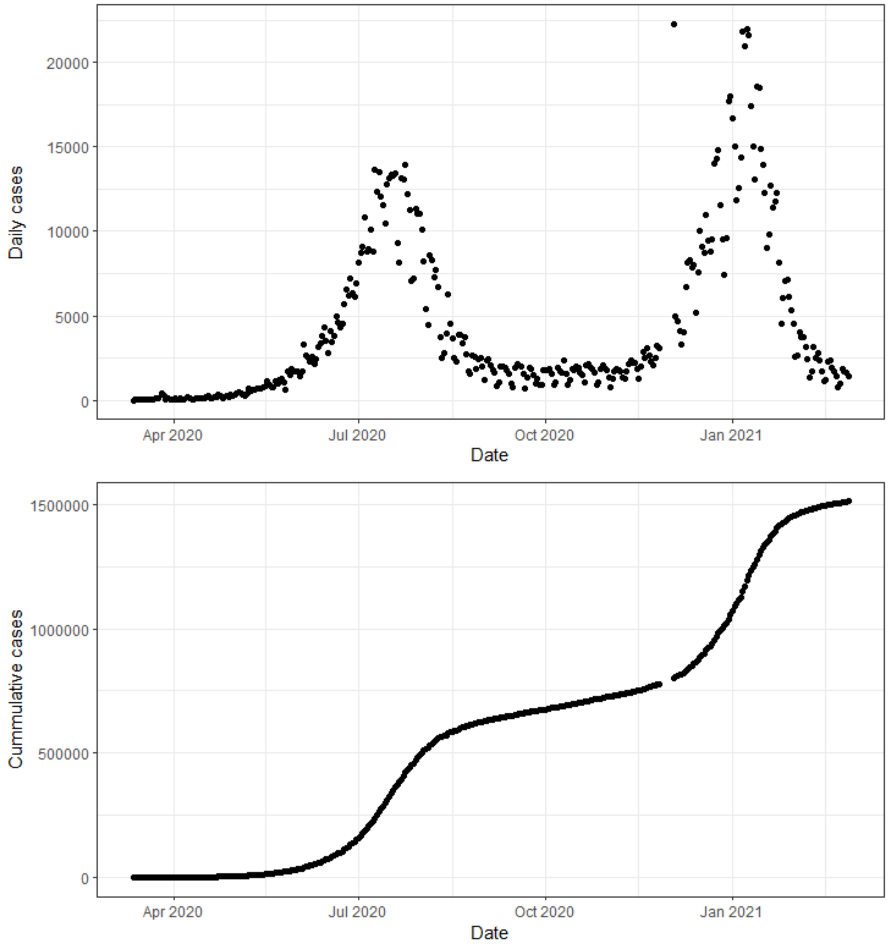
MANCHESTER 1824

The University of Manchester

Introduction

- Reliable and accurate short-term forecasts of COVID-19 cases are critical
    - to understand the progress of the pandemic in a country and
    - to evaluate the impact of intervention measures in controlling the COVID-19 outbreak.

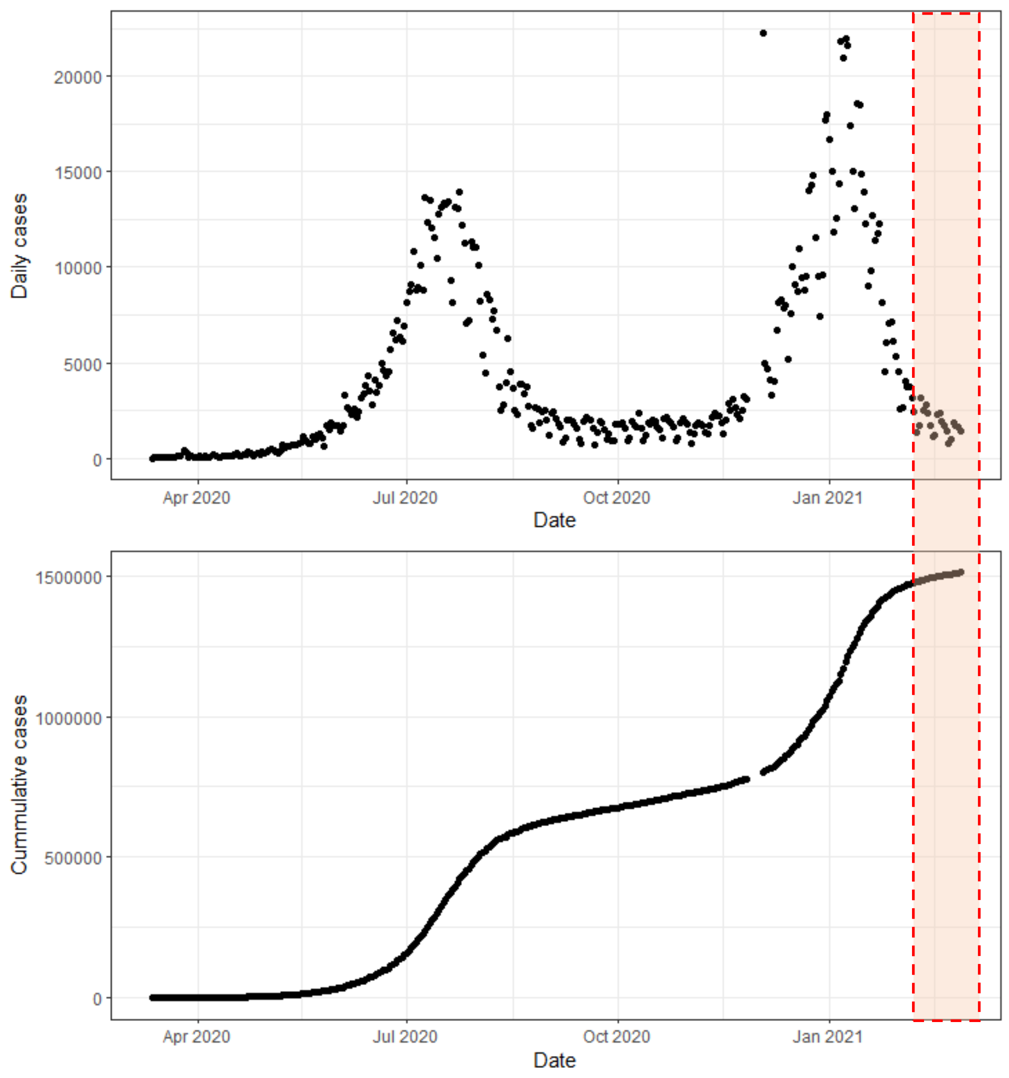
Reported COVID 19 cases in South Africa (12 March 2020-27 February 2021



https://raw.githubusercontent.com/dsfsi/covid19za/master/data/

Aim

- Can we predict the cumulative number of cases k days a head?
- How can we measure the performance/accuracy of our prediction?



Modelling approach

- We consider a Negative binomial distribution for modelling the number of daily COVID 19 cases to account for possible overdispersion.

$$Y_t \sim \mathrm{NB}(\mu_t, \delta)$$

- We consider four temporal models to capture the trend over time
- Models were fitted within the Bayesian framework using *INLA* assuming non informative priors

Model	
$AR(1)$	$\log(\mu_t) = \alpha + u_t,$ $u_1 \sim N(0, \tau_u(1-\rho^2)^{-1}),$ $u_t = \rho u_{t-1} + \epsilon_t, \quad t = 2, \ldots, T,$ $\epsilon_t \sim N(0, \tau_\epsilon),$
$AR(2)$	$\log(\mu_t) = \alpha + u_t,$ $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t, \quad t = 2, \ldots, T,$ $\epsilon_t \sim N(0, \tau_\epsilon),$
$RW(1)$	$\log(\mu_t) = \alpha + u_t,$ $u_t - u_{t-1} \sim N(0, \tau_u), \quad t = 2, \ldots, T,$
$RW(2)$	$\log(\mu_t) = \alpha + u_t,$ $u_t - 2u_{t-1} + u_{t-2} \sim N(0, \tau_u), \quad t = 3, \ldots, T.$

Model selection

Algorithm 1 Rolling origin cross-validation (ROCV).

1: Store the data starting day 1 to day T.
2: Estimation period: initialize the number of initial observation for estimating the model (k)
3: Prediction period: Set the number of days a head where the prediction is sought (w)
4: **while** $k + w \leq T$ **do**
5: Fit the proposed models using observations $t = 1, ..., t = k$
6: Hold out the next $k + 1, ..., k + w$ observations
7: Discard the remaining $T - (k + w)$ observations
8: Compute predictive error metrics

a) Compute DIC and WAIC in the estimation period,

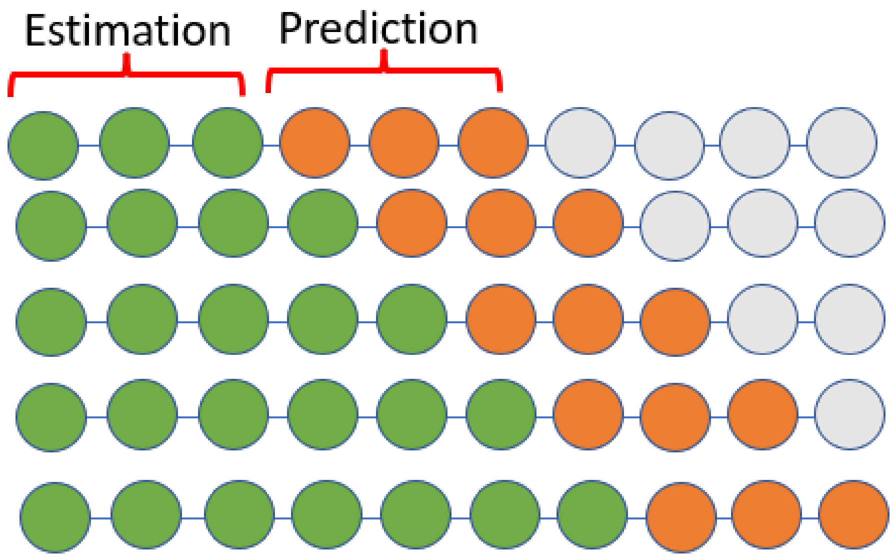
b) $MAE = \frac{\sum_{t=k+1}^{k+w} |C_t - C_t^*|}{w}$,

c) $MAPE = \frac{1}{w} \sum_{t=k+1}^{k+w} |\frac{C_t - C_t^*}{C_t}|$,

d) $Chi - Squared = \sum_{t=k+1}^{k+w} \frac{(C_t - C_t^*)^2}{C_t}$,

where C_t and C_t^* denotes the observed and predicted cumulative cases at time t, respectively.

9: $k = k + 1$
10: **end while**



Model selection

Algorithm 2 Modified rolling origin cross-validation (MROCV).

```
Store the data starting day 1 to day T.
Estimation period: initialize the number of initial observation for estimating the model (k)
while w ≤ 10 do
    while k + w ≤ T do
        Fit the proposed models using observations t = 1, ..., t = k
        Prediction period: Hold out the next k + 1, ..., k + w observations
        Discard the remaining T − (k + w) observations
        Store predicted and observed cumulative cases (C_t, C*_t)
        AE:
```

$$AE_{kw} = |C_{k+w} - C^*_{k+w}|,$$

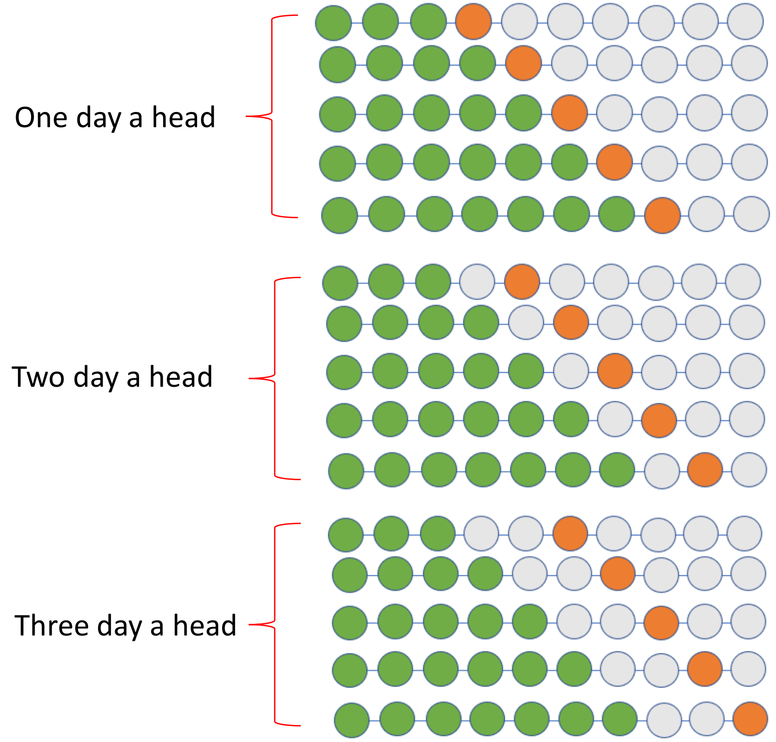
```
        APE:
```

$$APE_{kw} = \left|\frac{C_{k+w} - C^*_{k+w}}{C_{k+w}}\right|,$$

```
        OE2:
```

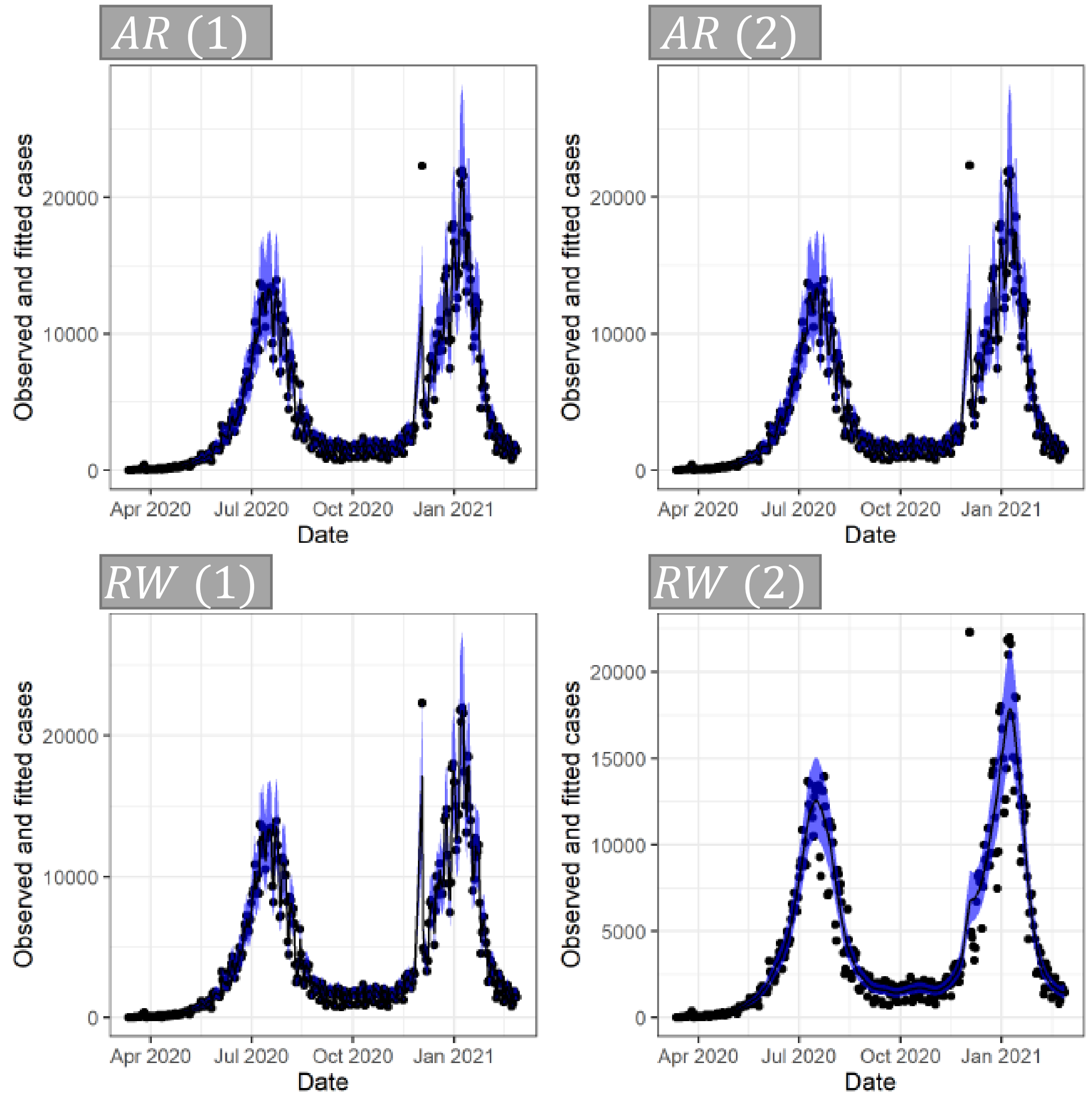
$$OE2_{kw} = \frac{(C_{k+w} - C^*_{k+w})^2}{C_{k+w}}$$

```
        k = k + 1
    end while
    Compute MAE_w = ΣAE_kw / K, MAPE_w = (1/K) Σ APE_kw, and Chi − Squared_w = Σ OE2_kw, where K is the number of estimation period.
    w = w + 1
end while
```



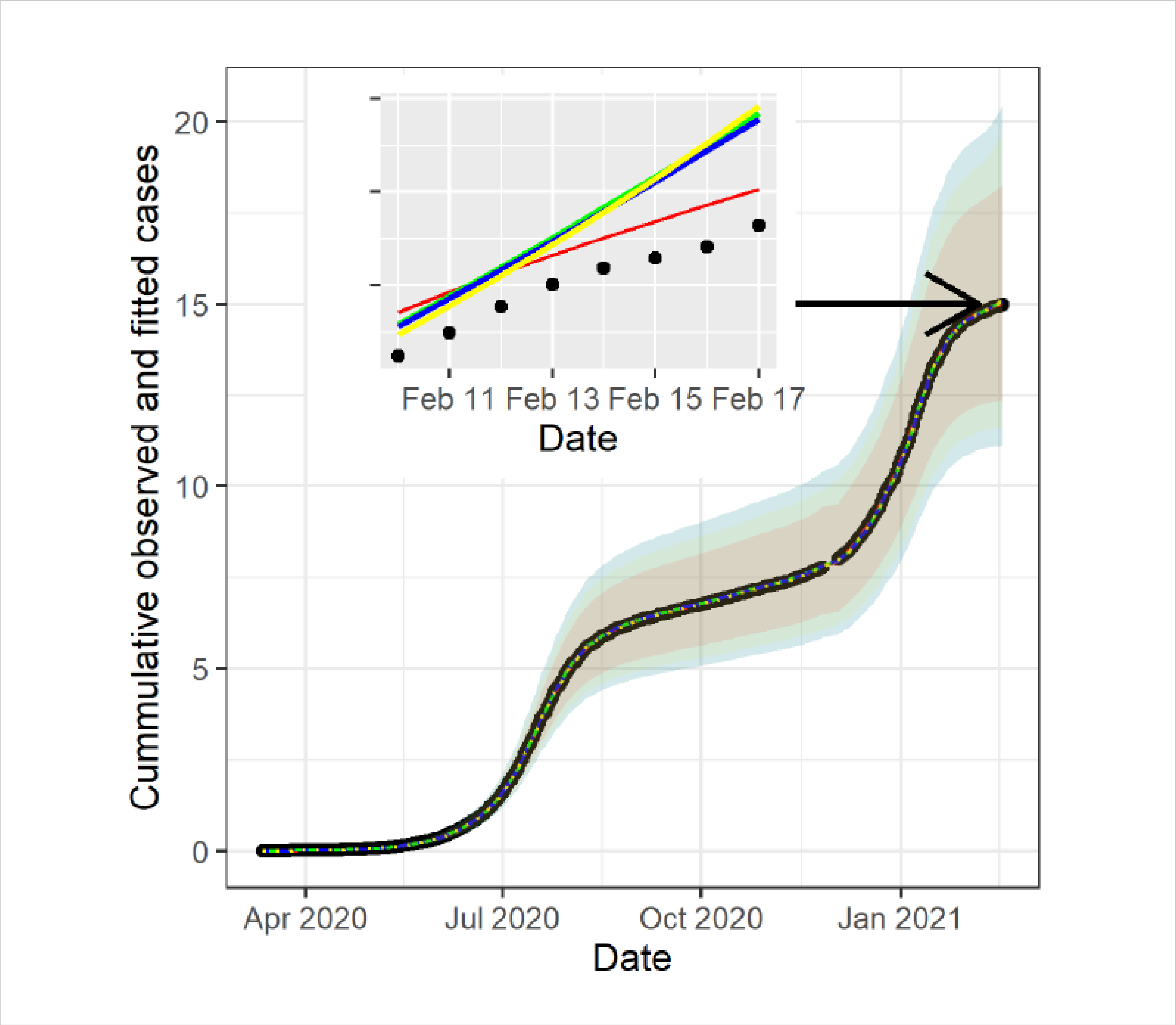
Application

- The four models described in the previous section were fitted to the daily reported new COVID-19 cases.
  - Estimation period fixed from 2/03/2020 to 07/02/2021
  - All models considered appear to fit the observed data (within the estimation period) well.
  - $AR(1)$, $AR(2)$, and $RW(1)$ models tend overfit the data
  - $RW(2)$ model produces a smooth line as predicted model



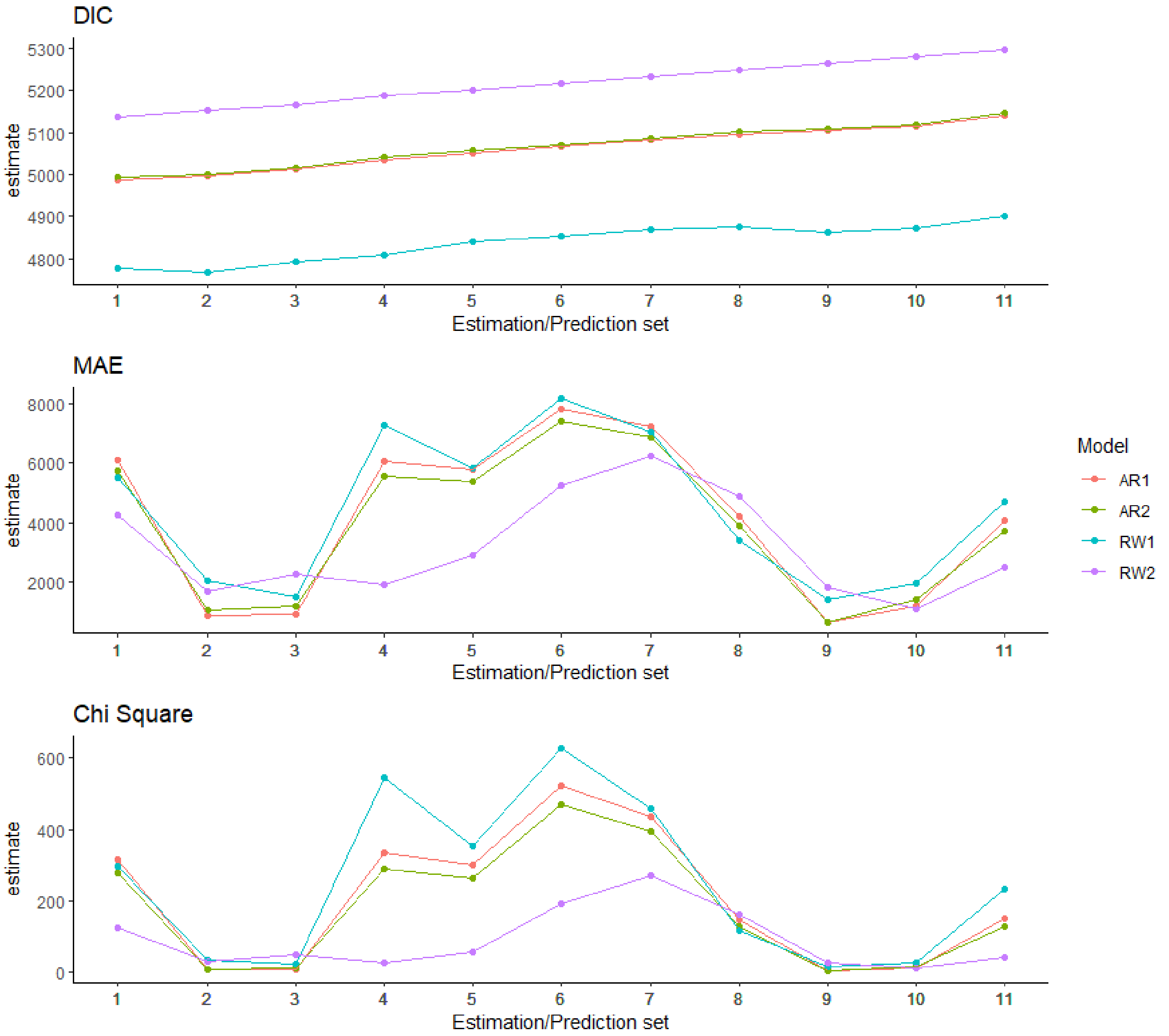
Application

- We produce 10-days ahead forecasting, up to 17/02/2021, of the cumulative COVID-19 cases for each model.
    - The $AR(1)$, $AR(2)$, and $RW(1)$ models performed well for the first three forecasting days and overestimated the cumulative cases from day three onward.
    - The $RW(2)$ model performed well, showing a consistent prediction performance throughout the forecasting period

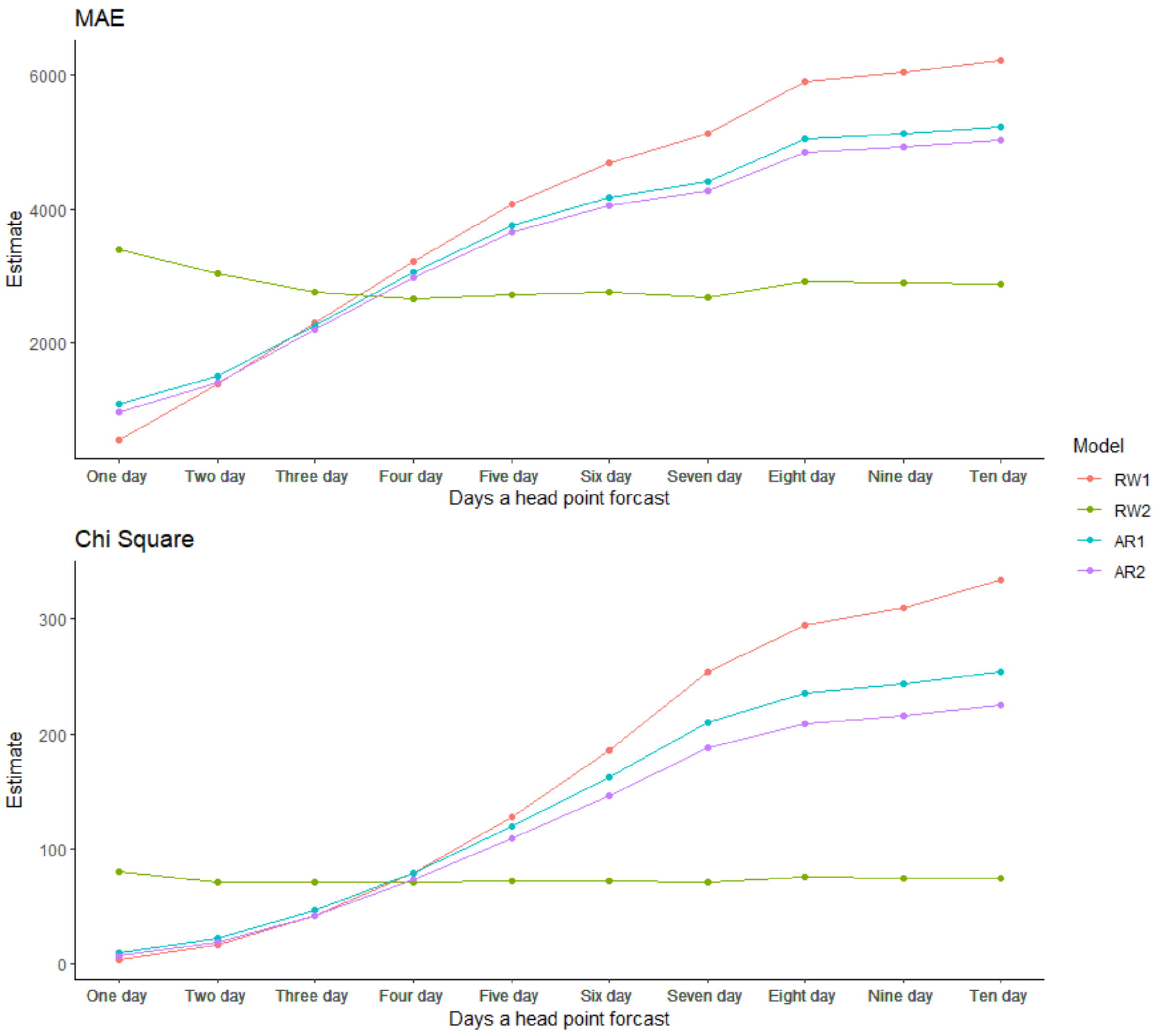


Prediction performance – Algorithm (1)

Estimation/ Prediction set	Estimation Period	Prediction Period
1	12/03/2020-07/02/2021	08/02/2021-17/02/2021
2	12/03/2020-08/02/2021	09/02/2021-18/02/2021
3	12/03/2020-09/02/2021	10/02/2021-19/02/2021
4	12/03/2020-10/02/2021	11/02/2021-20/02/2021
5	12/03/2020-11/02/2021	12/02/2021-21/02/2021
6	12/03/2020-12/02/2021	13/02/2021-22/02/2021
7	12/03/2020-13/02/2021	14/02/2021-23/02/2021
8	12/03/2020-14/02/2021	15/02/2021-24/02/2021
9	12/03/2020-15/02/2021	16/02/2021-25/02/2021
10	12/03/2020-16/02/2021	17/02/2021-26/02/2021
11	12/03/2020-17/02/2021	18/02/2021-27/02/2021



Prediction performance – Algorithm (2)



Conclusion

- We modelled COVID-19 cases in South Africa at the national level using publicly available data from 12 March 2020 to 27 February 2021.
- We have evaluated four widely used temporal models for forecasting confirmed cases of COVID-19 for South Africa.
- The analysis was based on readily accessible, publicly available data that is updated in real-time.
- The statistical methods applied are implemented using o-the-shelf open-source software and are not dependent on any assumptions regarding COVID-19 transmission dynamics.
- We have shown the usefulness of established temporal models to provide short term forecasts of the cumulative COVID-19 cases.
- Such models could help in decision-making when knowledge regarding factors affecting transmission-dynamics of the disease is limited.

R code

The source code for producing the results presented is available at

belayb/COVIDincidenceSA: Modelling COVID-19 incidence and forecasting - South Africa (github.com)