

1 Title

2 **The influence of climate factors on COVID-19 transmission**
3 **in Malaysia: An autoregressive integrated moving average**
4 **(ARIMA) model**

5 **Authors**

6 Cho Naing^{1,2*}, Han Ni^{3¶}, Htar Htar Aung^{1¶}, Elaine Chan¹, Joon Wah Mak¹

7 **Affiliations**

8 1. International Medical University, Kuala Lumpur, Malaysia

9 2. Faculty of Tropical Health and Medicine, James Cook University, Queensland, Australia

10 3. SEGi University, Sibu Clinical Campus, Sarawak, Malaysia

11

12 *Corresponding author: Email: cho3699@gmail.com (CN)

13 ¶These authors contributed equally to this work.

14

15

16

17

18

19

20

21

22 **Abstract**

23 **Background**

24

25 A unique concern pertaining to the spread of COVID-19 across countries is the asymmetry of
26 risk and the irrational fear of a new pandemic and its possible serious consequences. This study
27 aimed to perform a time series analysis on the association between climate factors and daily
28 cases of COVID-19 in Malaysia up to 15 July 2020. The second objective was to predict daily
29 new cases using a forecasting technique. To address within-country variations, the analysis was
30 extended to the state level with Sarawak state as an example.

31 **Methodology/Principal Findings**

32

33 Datasets on the daily confirmed cases and climate variables in Malaysia and Sarawak state were
34 obtained from publicly accessible official websites. A descriptive analysis was performed to
35 characterize all the important variables over the study period. An autoregressive integrated
36 moving average (ARIMA) model was introduced using daily cases as the dependent variable and
37 climate parameters as the explanatory variables.

38 For Malaysia, the findings suggest that, *ceteris paribus*, the number of COVID-19 cases
39 decreased with increasing average temperature ($p=0.003$) or wind speed ($p=0.029$). However,
40 none of the climate parameters showed a significant relationship with the number of COVID-19
41 cases in Sarawak state. Forecasts from the ARIMA models showed that new daily COVID-19
42 cases had already reached the outbreak level and a decreasing trend in both settings. Holding

43 other parameters constant, a small number of new cases (approximately a single digit) is a
44 probable second wave in Sarawak state,

45 **Conclusions/Significance**

46
47 The findings suggest that climate parameters and forecasts are helpful for reducing the
48 uncertainty in the severity of future COVID-19 transmission. A highlight is that forecasts will be
49 a useful tool for making decisions and taking the appropriate interventions to contain the spread
50 of the virus in the community.

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68 **Introduction**

69 As of 15 June 2020, 7,823,291 confirmed cases and 431,541 deaths from the novel coronavirus
70 disease (COVID-19) have been reported from the WHO's six regions [1]. In Malaysia, the first
71 sporadic case was detected on 24th January 2020, and as of 16 July 2020, 8,739 confirmed cases
72 and 122 deaths have been reported [2]. A unique concern across countries is the asymmetry of
73 risk and the irrational fear of a new pandemic with its possible catastrophic consequences, as
74 documented with the 1918 Spanish flu that killed an estimated 50 million people worldwide. To
75 reduce uncertainty, governments need to accurately forecast the spread of confirmed cases.
76 According to current evidence, the COVID-19 virus ((severe acute respiratory syndrome
77 coronavirus 2 (SARS-CoV-2)) is primarily transmitted among people through respiratory
78 droplets and contact routes [3]. As such, climate factors such as temperature and/or relative
79 humidity may influence the transmission of coronavirus [4]- by affecting its survival. Previously,
80 studies have documented the relationship between weather parameters and severe acute
81 respiratory coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus
82 (MERS-CoV).

83 Time series studies are most relevant when data have been accumulated over a
84 considerably long period of time. A common challenge to time series methodology is the
85 availability of data. We were able to identify the weather parameters and daily COVID-19 cases
86 at the country level instead of the weekly, monthly or annual aggregates. This finer temporal
87 resolution will allow us to further characterize short-period variations. Despite the sustained

88 transmission of COVID-19 cases, the effects of climate factors on the population risk of COVID-
89 19 in Malaysia have not yet been assessed in detail.

90 Hence, two research questions arise: “*Is there a role of climate parameters on the*
91 *possible viral spread in Malaysia?*” and “*what is a forecast of regional transmission of new*
92 *cases?*”. Taken together, the objective of the present study was to perform a time series analysis
93 on the association between climate factors and daily cases of COVID-19 in Malaysia between
94 January 2020 and 15 July 2020. The second objective was to predict daily new cases using a
95 forecasting technique. Further, to address within-country variations, we extend our analysis to
96 the state level with Sarawak state as an example.

97 We anticipate that our findings will be useful not only for policymakers in Malaysia
98 but also for countries that will experience inclement weather conditions in 2020 to prepare the
99 necessary public health response for the potential spread of COVID-19.

100

101 **Materials and Methods**

102 We performed the present study following a prior protocol. A description of the protocol is
103 available from the corresponding author on request.

104 **Study area**

105 Malaysia is located between 2° and 7° N, bordering Thailand to the north and Singapore to the
106 south. Sarawak state (1.5533°N, 110.3592°E) is an administrative region of Malaysia located on
107 Borneo Island.

108 **Data collection**

109 **COVID-19 incidence**

110 Publicly accessible datasets of the daily confirmed cases and deaths in Malaysia were obtained
111 from the official website of the Malaysian Ministry of Health (MOH). These data were cross-
112 checked with the John Hopkins University Coronavirus Resource Center Repository
113 (<https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>) up to 15th July 2020

114 These data originated from the active and passive case surveillance system in Malaysia.
115 We extracted the total daily counts of confirmed cases up to 15th July 2020. The first COVID-19
116 case was announced in China in late December 2019, and the first case in Malaysia was
117 announced on 23rd January 2020. This finding implies that COVID-19 had spread to Malaysia
118 within one month.

119 **Climate factors**

120 The daily mean temperature (°F), relative humidity (%), precipitation (%) and wind velocity
121 (mph) over the study period were collected from an archive of weather data for Malaysia
122 (<https://weatherspark.com/history/>). For the country level, we used the means of the climate
123 parameters across the reported states.

124 **Statistical analysis**

125 First, a descriptive analysis was performed to characterize all the variables over the study
126 period in Malaysia and Sarawak state. Then, we introduced an autoregressive integrated moving
127 average (ARIMA) model, using daily cases as the dependent variable and the climate parameters
128 as the explanatory variables.

129 A structural ARIMA model was introduced to examine the association between the
130 number of COVID-19 cases (daily counts) and the various explanatory variables (i.e., the mean
131 temperature, wind speed, precipitation and relative humidity). We computed various
132 permutations of the order of the autoregressive (AR), order of integration (I) and order of moving

133 average (MA) terms and chose the optimal combination of parameters using the mean squared
134 error. The correlograms were visualized to help in determining the orders of the MA and AR
135 terms to include in the model. Before modeling the daily confirmed cases, we checked the
136 stationarity of the data series by the augmented Dickey-Fuller (ADF) test [5]. A model was first
137 applied to quantify the aggregate country-level associations between the climate variables and
138 the daily cases of COVID-19. Another model was applied to quantify the state-specific
139 association for Sarawak state, Malaysia.

140 We applied the *auto.arima(1,1,0)* function to forecast new cases. The best order of an
141 ARIMA process was determined through a unit root test to identify the appropriate degree of
142 differencing and by the minimization of the corrected Akaike information criterion (AIC) [5].
143 We then conducted fourteen-day-ahead point forecasts and calculated the prediction intervals,
144 taking into account the incubation period. Two analytical models were introduced depending on
145 the situation of COVID-19 in Malaysia in general and Sarawak state in particular.

146 The forecast accuracy was checked with four accuracy measures: the mean absolute error
147 (MAE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE), and root
148 mean square error (RMSE) [6]. We performed the Ljung-Box test, for which a p-value greater
149 than 0.05 means that the residuals for the current time series model are independent [5].

150 The basic ARIMA forecasting estimated equation is

151
$$\Delta \hat{y}_t = \lambda_1 \Delta y_{t-1} + \dots + \lambda_p \Delta y_{t-p} + \psi_1 \Delta e_{t-1} + \dots + \psi_q \Delta e_{t-q} + e_t \dots \dots \dots (1)$$

152 where

153 Δ = degree of nonseasonal differences,

154 \hat{Y} = time series that will be predicted at time t,

155 P = lag order of AR,

156 Λ =the coefficient of each parameter,

157 Q =lag order of MA,

158 Ψ =coefficient of each parameter q , and

159 et =the residuals of the errors at time t .

160 STATA 16.1 (StataCorp, Texas, USA) was used to calculate the descriptive statistics, and

161 the “*Forecast*” package in *R* (R Core team) was used to forecast the confirmed COVID-19 cases.

162 **Results**

163 Table 1 presents the characteristics of the variables for Malaysia and Sarawak state.

164 Table 1. Characteristics of the weather variables and new reported cases

Malaysia from 13 March 2020-15 July 2020		
Variable	Mean (SD)	Minimum-maximum
Temperature (°F)	89.6 (0.57)	88-90
Precipitation (%)	6.15 (0.21)	5.9-6.6
Humidity (%)	99.5 (0.8)	98- 100
Wind speed (mph)	5.3 (0.8)	4.1-7.3
Cases	50 (62)	0-277
Deaths	1 (1.5)	0-9
Sarawak state, Malaysia, from 13 March 2020-15 July 2020		
Description	Mean (SD)	Minimum-maximum
Temperature (°F)	88.0 (1.01)	86-89

Precipitation (%)	5.7 (0.16)	5.5-6
Humidity (%)	100 (0)	100-100
Wind speed (mph)	3.3 (0.2)	2.9-3.6
Cases	5 (6)	0-32

165
166 During the period from 24th January 2020 to 15th July 2020, the mean daily cases in Malaysia
167 was 55, while the highest daily cases of 277 was recorded on 6th April 2020. During the period
168 from 13th March 2020 to 15 July 2020, the mean daily cases in Sarawak state in Malaysia was 5,
169 while the highest daily cases of 32 was recorded on 1st April 2020. Figure 1 shows the total
170 number of cases stratified by the states of Malaysia.

171 **Fig 1. Total number of confirmed COVID-19 cases stratified by the states of Malaysia**

172

173 **Relationship with climate parameters**

174 For Malaysia, the aggregate mean temperature, wind speed, precipitation and relative humidity
175 were 89.8 °F, 5.3 mph, 6.1% and 99.4%, respectively (Table 1). The differenced daily COVID-
176 19 cases in Malaysia over time from 24th January 2020 to 15th July 2020 are presented in Fig 2,
177 which shows the time series data with an impression of stationary.

178 **Fig 2. Differenced daily COVID-19 cases in Malaysia from 24 January 2020 to 15 July**
179 **2020.**

180 The estimated coefficients of the ARIMA(1,1,1) model for Malaysia are presented in Table
181 2. Temperature (-58.99 (-97.6 to -20.4, $p=0.003$ and wind speed (-77.03 95% CI:-221.5 to 67.5,
182 $p=0.029$).

183 were significantly and inversely related to Covid-19 confirmed cases in Malaysia during the
184 study period.

185 Table 2 Estimates of the regression coefficients of the ARIMA model for Malaysia between 24
186 January 2020 and 15 July 2020

Parameter	Coefficient (95% CI)	<i>p</i> -value
Intercept	-1.05 (-5.7 to 3.58)	0.65
Temperature (°F)	-58.99 (-97.6 to -20.4)	0.003
Precipitation (%)	132.1 (-0.66 to 264.8)	0.051
Humidity (%)	20.9 (-3.9 to 45.9)	0.099
Wind speed (mph)	-77.03 (-221.5 to 67.5)	0.029

187 The partial autocorrelation function (PACF) plot (S1 Fig) had a significant spike only at lag 1,
188 reflecting that all the higher-order autocorrelations are effectively explained by the lag-1
189 autocorrelation.

190 For Sarawak state from 13th March 2020 to 15th July 2020, we followed the same procedures
191 performed for the model of Malaysia (S2 Fig). None of the climate variables were significantly
192 associated with the number of confirmed COVID-19 cases (Table 3).

193 Table 3 Estimates of the regression coefficients of the ARIMA model for Sarawak state of
194 Malaysia

Variable	Coefficient (95% CI)	<i>p</i> -value
Temperature (°F)	-0.79 (-5.9 to 4.3)	0.76
Precipitation (%)	-0.16 (-19.0 to 18.7)	0.99
Wind speed (mph)	-7.02 (-26.8 to 12.8)	0.49
Constant	-0.04 (-0.37 to 0.3)	0.83

195

196 **Forecasting of the confirmed COVID-19 cases**

197 For Malaysia, ARIMA forecasting with the *auto.arima*(1, 1,1) function showed the best order
198 of an ARIMA process with the appropriate degree of differencing and identification of the AR
199 and MA parameters. This finding was also applicable to the *auto.arima*(1,1,1) model for
200 Sarawak state. The forecast results for new daily COVID-19 cases in Malaysia and Sarawak state
201 (prediction of 14 days) are presented in Fig 3 and Fig 4, respectively. We produced a final set of
202 forecasts and prediction intervals using the most recent data, up until 15th July 2020. We
203 estimated two levels of uncertainty (80% and 95% CIs). Ljung–Box tests showed that the null
204 hypothesis of the serial independence of the residuals for the current time series models could
205 not be rejected ($p > 0.63$ for Malaysia, $p > 0.87$ for Sarawak state).

206 **Fig 3 Forecasts of new daily COVID-19 cases in Malaysia (prediction of 14 days).**

207 **Fig 4 Forecasts of new daily COVID-19 cases in Sarawak state (prediction of 14 days).**

208 As seen in the figures, the spreads were characterized by a declining trend, indicating that
209 the COVID-19 transmission has probably already reached its peak. However, the slowdown in
210 the new COVID-19 cases seems to require shorter time for Sarawak state compared to the whole
211 country. On the other hand, a small spike (around a single digit) was seen in Sarawak state; this
212 finding cannot rule out the possibility of a new wave (with small intensity) if the current
213 recommendations/standard operating procedures such as social/physical distancing, the use of face
214 masks, and hand washing, as issued by the Malaysian MOH is not followed properly. The
215 forecast accuracies were validated with the MAE, MAPE, MASE, and RMSE (S1 Table).

216

217

218 Discussion

219 Summary of the main results

220 The findings of this study were based on COVID-19-related data and climate parameters
221 available for Malaysia in aggregate and Sarawak state in particular. The major observations of
222 the current analysis are as follows:

223 1. For both Malaysia (24th January 2020 to 15th July 2020) and Sarawak state (13th March to
224 15th July 2020), time series ARIMA models were introduced.

225 2. For Malaysia, the findings suggest that, *ceteris paribus*, as the average temperature or
226 wind speed increased, the number of COVID-19 cases decreased.

227 3. None of the climate parameters showed a significant relationship with the number of
228 COVID-19 cases in Sarawak state.

229 4. Forecasts from the ARIMA models showed that new daily COVID-19 cases have already
230 reached the outbreak level, and a decreasing trend was observed in both settings. *Ceteris paribus*,
231 a small number of new cases (approximately a single digit) would be indicative of a probable
232 second wave in Sarawak state.

233 Our findings were supported by a study in Turkey, which also showed an inverse
234 relationship between temperature and the number of COVID-19 cases [7].

235 A study in Singapore reported that an average maximum temperature increase of 1 °C decreased
236 the incidence rate by -7.5% (95% CI = [-12.3; -2.6]) on the same day [8]. In the 2003. SARS
237 epidemic, a study of Beijing and Hong Kong showed an inverse relationship between
238 temperature and the number of cases [9]. Although an exact cause is not yet known, a possible
239 reason may be that higher temperature influences the viability of the severe acute respiratory

240 syndrome coronavirus 2 virus [7]. A higher wind speed was significantly and positively related
241 to the number of new COVID-19 cases reported in a study in Turkey [7], which was
242 contradictory to the current findings in Malaysia.

243 Temperature is a fundamental factor in the human living environment that can play an
244 important role in public health in terms of epidemic development, prevention, and control [10].
245 Unpublished studies have also reported that low temperatures are beneficial to viral transmission
246 [11,12].

247 Bill Gates commented on COVID-19 transmission saying that *“I hope it’s not that bad,*
248 *but we should assume it will be until we know otherwise”* [13]. It is believable that forecasts and
249 their associated uncertainty can and should be an integral part of the decision-making process,
250 especially in high-risk cases [14]. In this study, univariate time series models which assume that
251 the data are accurate and past patterns will continue to apply were used [15]. Significant,
252 consistent forecast errors (potentially spanning outside the prediction intervals) should be
253 associated with changes in the observed patterns and the need for additional actions and
254 measures encountered.

255 The modeling quality is dependent on whether the assumptions are violated or not. In this
256 study, the model’s diagnostic tests and a small percentage of error terms ensure the forecasting
257 validity. The assumptions required for the prediction intervals of ARIMA models, such as
258 uncorrelated and normally distributed residuals, were not violated. Moreover, the accuracy,
259 robustness, coherency and uncertainty mitigation that the models imply have made them a
260 popular forecasting framework, highlighting the potential benefits of multiple temporal
261 aggregation in decision making [16]. Forecasts of deaths and hospitalizations (the number of new

262 cases in this case) will help inform public health decision making by projecting the likely impact
263 of the COVID-19 pandemic.

264 **Study limitations**

265 There were limitations to the current analyses that need to be acknowledged. As the results are
266 based on aggregate data for the country and specific to only one particular state, in light of
267 geographic differences, the evaluation power could have been affected. Additionally, the results
268 may have been affected by the quality of the data in relation to the accuracy of the diagnostic
269 tests for COVID-19. Furthermore, there might be other factors (e.g., the quality of the case
270 surveillance system) other than weather variables that can influence the transmission potential of
271 COVID-19. Only a few weather variables were available for inclusion in the current study.

272 Nevertheless, there are strengths in the current study. The source data for COVID-19
273 cases originated from the national database, which is identical to the John Hopkins's database,
274 thus ensuring the accuracy of the data. The findings were, to a certain extent, comparable to
275 those from other published studies. For instance, an inverse relationship between the average
276 temperature and the number of new cases was in line with the findings from other countries. As
277 we hope the presented forecasts clearly show, epidemic growth is a highly nonlinear process,
278 where every day lost to inaction is too much [17]. On the other hand, a small spike
279 (approximately a single digit) was seen in Sarawak state; this finding cannot rule out the
280 possibility of new (small intensity) waves. This finding implies a potential small wave is
281 expected in Sarawak state, if people fail to comply with the measures recommended by the
282 Malaysian MOH such as social/physical distancing, use of face masks and hand washing. In this
283 context, forecasts will be a useful tool for governments and individuals to make decisions and
284 take appropriate actions to contain the spread of the virus [15].

285 **Conclusions**

286 The findings suggest that it is important to consider climate parameters and forecasts to reduce
287 the uncertainty in the severity of future COVID-19 transmission. Our results highlight that
288 forecasts will be a useful tool for making decisions and taking the appropriate interventions to
289 contain the spread of the virus in the community.

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307 **Acknowledgments**

308 We thank our institutions for allowing us to perform this study. We acknowledge Dr. Norah Htet
309 for assisting with the earlier draft.

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327 **References**

- 328 1. WHO. Coronavirus disease (COVID-19) Pandemic.2020a.
- 329 2. Malaysian Ministry of Health (MOH), 2020.
- 330 Available at <https://www.worldometers.info/coronavirus/country/malaysia/>
- 331 3. WHO. Models of transmission of virus causing COVID-19: implications for IPC
- 332 precaution recommendations. 2020b.
- 333 Available at
- 334 [https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-](https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations)
- 335 [causing-covid-19-implications-for-ipc-precaution-recommendations.](https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations)
- 336
- 337 4. Casanova LM, Jeon S, RutalaWA, et al. Effects of air temperature and relative humidity
- 338 on coronavirus survival on surfaces. *Appl Environ Microbiol.* 2010; 76:2712-7.
- 339 5. Peter B, Richard D. *Introduction to Time Series and Forecasting* (2nd ed.). Springer. 2002;
- 340 35–38.
- 341 6. Hyndman RJ, Anne B K. Another look at measures of forecast accuracy. *International*
- 342 *Journal of Forecasting*, 2006; 22(4):679-688 [doi:10.1016/j.ijforecast.2006.03.001](https://doi.org/10.1016/j.ijforecast.2006.03.001).
- 343 7. Şahin M. Impact of weather on COVID-19 pandemic in Turkey. *Sci Total Environ.*
- 344 2020;728:138810.
- 345 8. Tobías A, Molina T. Is temperature reducing the transmission of COVID-19?. *Environ*
- 346 *Res.* 2020; 186:109553. doi:10.1016/j.envres.2020.109553.
- 347 9. Bi 2007 Bi P, Wang J, Hiller JE. Weather: driving force behind the transmission of severe
- 348 acute respiratory syndrome in China? *Intern Med J.* 2007;37(8):550-554.

- 349 10. Qi H, Xiao S, Shi R, Ward MP, Chen Y, Tu W, et a Su Q, Wang W, Wang X, Zhang Z.
350 COVID-19 transmission in Mainland China is associated with temperature and humidity:
351 A time-series analysis. *Sci Total Environ.* 2020 1;728:138778.
- 352 11. Triplett M. Evidence that higher temperatures are associated with lower incidence of
353 COVID-19 in pandemic state, cumulative cases reported up to March 27. 2020. medRxiv
354 preprint.
- 355 12. Wang J. High temperature and high humidity reduce the transmission of COVID-19.
356 SSRN Electronic Journal. 2020 doi: 10.2139/ssrn.3551767.
- 357 13. Bill Gate. How to respond to COVID-19. Feb 2020. Available at
358 <https://www.gatesnotes.com/Health/How-to-respond-to-COVID-19>.
- 359 14. Makridakis S, Kirkham R, Wakefield A, Papadaki M, Kirkham J, a Long L. Forecasting,
360 uncertainty and risk; perspectives on clinical decision-making in preventive and curative
361 medicine,. *International Journal of Forecasting.* 2017;35(2):659-666
- 362 15. Petropoulos 2020 Petropoulos F, Makridakis S. Forecastingthe novel coronavirus COVID-
363 19.PLoS One.2020; 15(3):e0231236
- 364 16. Spiliotis E, Petropoulos F, Assimakopoulos V. Improving the forecasting performance of
365 temporal hierarchies. *PLoS One.* 2019;14(10):e0223422.
- 366 17. Perc M, Miksic NG, Stozer A. Forecasting COVID-19. *Frontiers in Physics.*2020; 8:127.

367

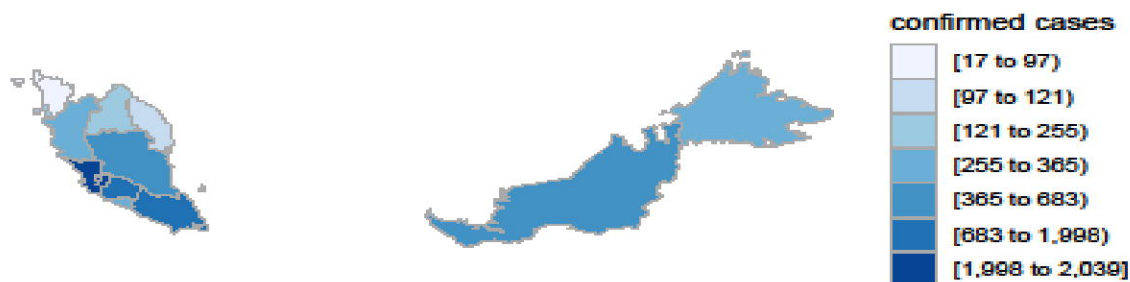
368 **Supporting information**

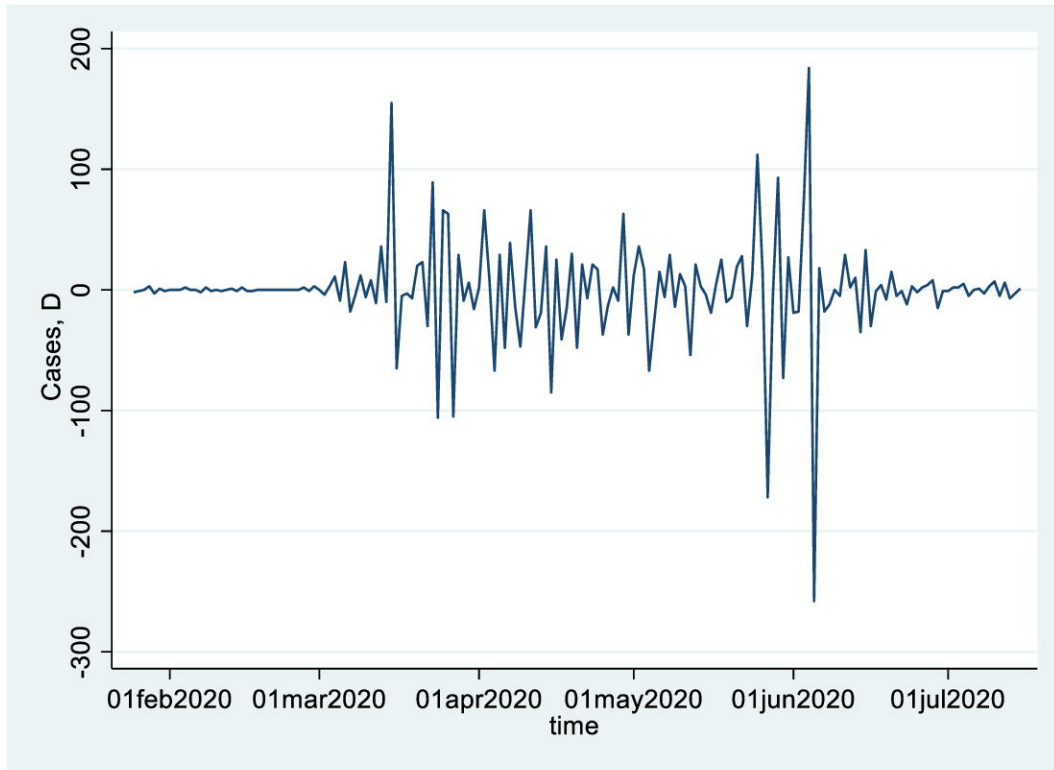
369 S1 Fig Autocorrelation (ACF) and partial autocorrelation of COVID-19 time series data in
370 Malaysia

371 S 2 Fig Autocorrelation (ACF) and partial autocorrelation of COVID-19 time series data in
372 Sarawak state

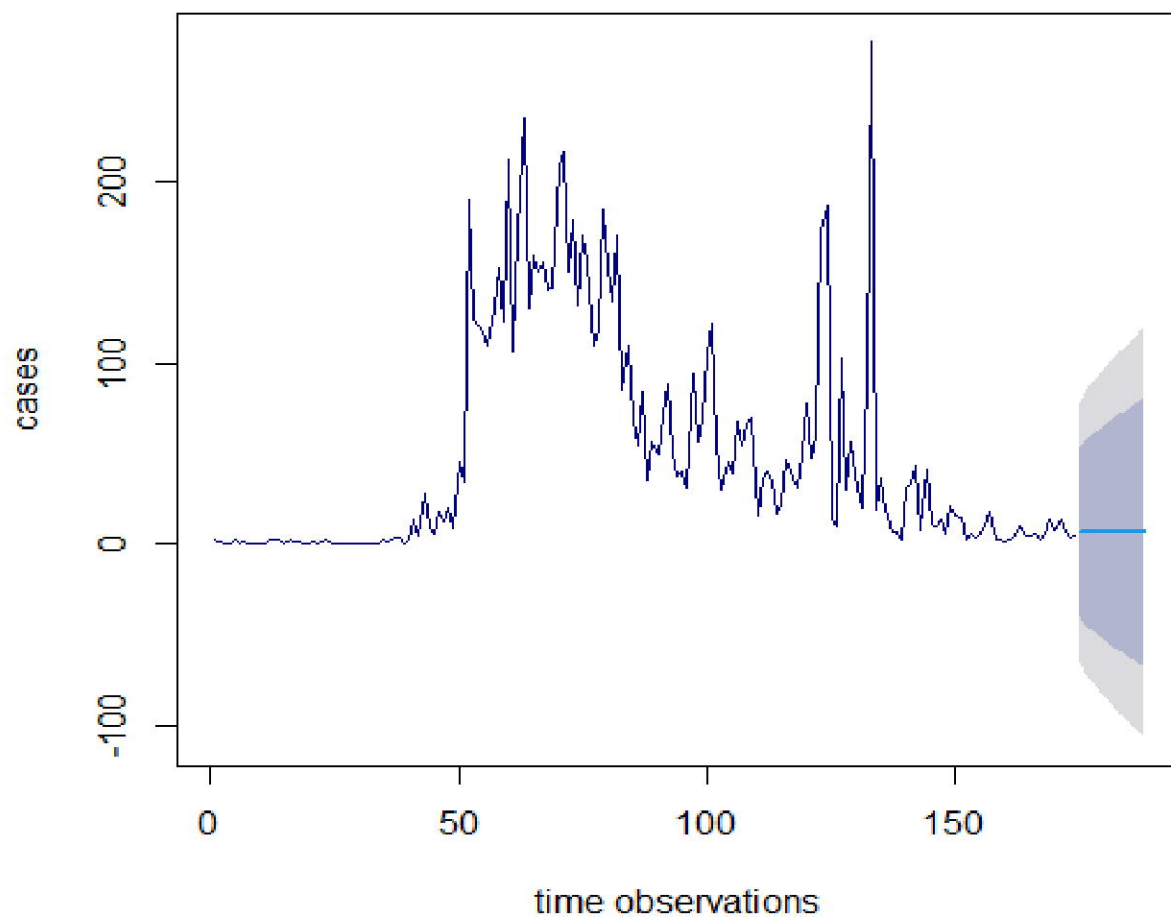
373 S1 Table Accuracy measures

374



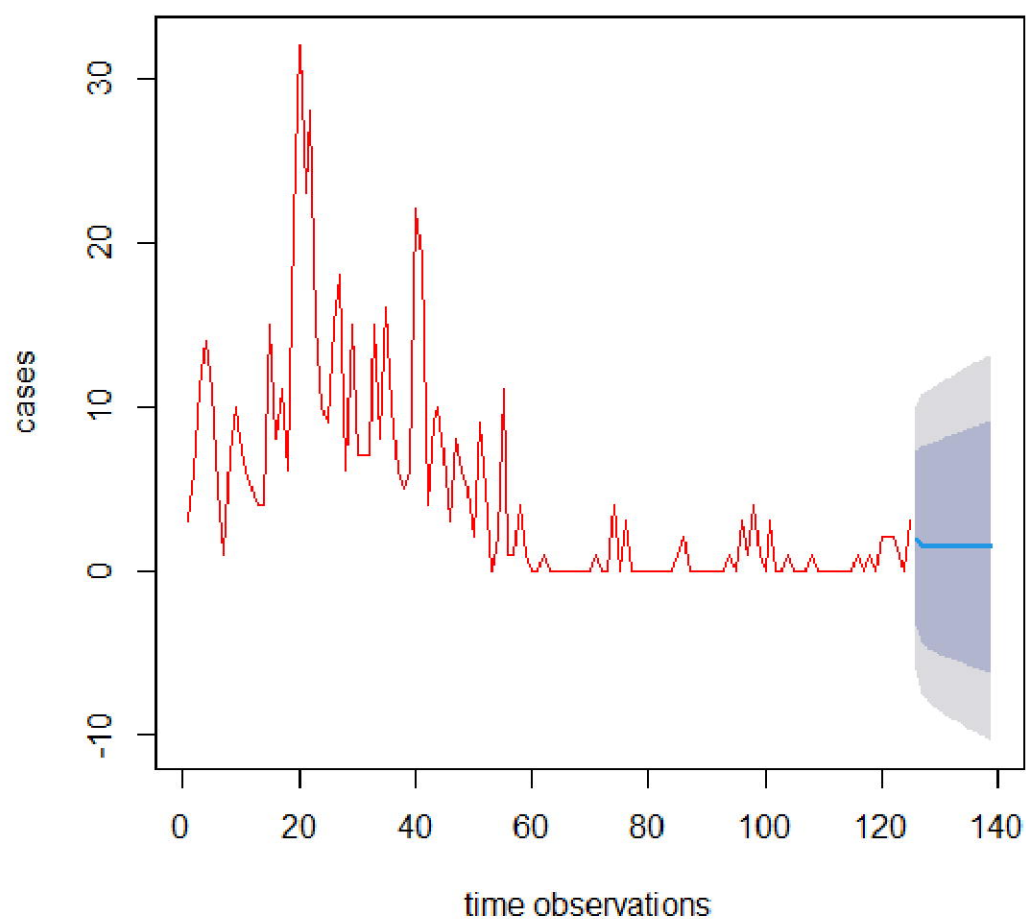


Forecasts from ARIMA(1,1,1)



The solid navy blue line denotes the actual data, which were last updated on July 15th. From this date onward, we have the predicted values, where the solid blue line denotes the continuation of the trend from the past 14 days, i.e., if nothing changes. The dark gray boundaries denote the 95% CI, while the light gray boundaries represent the 80% CI.

Forecasts from ARIMA(1,1,1)



The solid red solid line denotes the actual data, which were last updated on July 15th. From this date onward, we have the predicted values, where the solid blue line denotes the continuation of the trend from the past 14 days, i.e., if nothing changes. The dark gray boundaries denote the 95% CI, while light gray boundaries represent the 80% CI.