

1 **Title: An ensemble n -sub-epidemic modeling framework for short-term forecasting**
2 **epidemic trajectories: Application to the COVID-19 pandemic in the USA**

3
4 **Short title:** An ensemble n -sub-epidemic modeling framework for short-term forecasting
5 epidemics

6
7 Gerardo Chowell^{1,2}, Sushma Dahal¹, Amna Tariq¹, Kimberlyn Roosa³, James M. Hyman⁴,
8 Ruiyan Luo¹

9
10 ¹ Department of Population Health Sciences, School of Public Health, Georgia State University,
11 Atlanta, GA, USA.

12 ² Division of International Epidemiology and Population Studies, Fogarty International Center,
13 National Institutes of Health, Bethesda, MD, USA.

14 ³ National Institute for Mathematical and Biological Synthesis (NIMBioS), University of
15 Tennessee, Knoxville, TN, USA

16 ⁴ Department of Mathematics, Center for Computational Science, Tulane University, New
17 Orleans, LA, USA.

18 * Corresponding author (gchowell@gsu.edu)

19
20 **Abstract**

21
22 We analyze an ensemble of n -sub-epidemic modeling for forecasting the trajectory of epidemics
23 and pandemics. These ensemble modeling approaches, and models that integrate sub-epidemics
24 to capture complex temporal dynamics, have demonstrated powerful forecasting capability. This
25 modeling framework can characterize complex epidemic patterns, including plateaus, epidemic
26 resurgences, and epidemic waves characterized by multiple peaks of different sizes. We
27 systematically assess their calibration and short-term forecasting performance in short-term
28 forecasts for the COVID-19 pandemic in the USA from late April 2020 to late February 2022.
29 We compare their performance with two commonly used statistical ARIMA models. The best fit

30 sub-epidemic model and three ensemble models constructed using the top-ranking sub-epidemic
31 models consistently outperformed the ARIMA models in terms of the weighted interval score
32 (WIS) and the coverage of the 95% prediction interval across the 10-, 20-, and 30-day short-term
33 forecasts. In the 30-day forecasts, the average WIS ranged from 377.6 to 421.3 for the sub-
34 epidemic models, whereas it ranged from 439.29 to 767.05 for the ARIMA models. Across 98
35 short-term forecasts, the ensemble model incorporating the top four ranking sub-epidemic
36 models (Ensemble(4)) outperformed the (log) ARIMA model 66.3% of the time, and the
37 ARIMA model 69.4% of the time in 30-day ahead forecasts in terms of the WIS. Ensemble(4)
38 consistently yielded the best performance in terms of the metrics that account for the uncertainty
39 of the predictions. This framework could be readily applied to investigate the spread of
40 epidemics and pandemics beyond COVID-19, as well as other dynamic growth processes found
41 in nature and society that would benefit from short-term predictions.

42

43 **Summary**

44 The COVID-19 pandemic has highlighted the urgent need to develop reliable tools to forecast
45 the trajectory of epidemics and pandemics in near real-time. We describe and apply an ensemble
46 *n*-sub-epidemic modeling framework for forecasting the trajectory of epidemics and pandemics.
47 We systematically assess its calibration and short-term forecasting performance in weekly 10-30
48 days ahead forecasts for the COVID-19 pandemic in the USA from late April 2020 to late
49 February 2022 and compare its performance with two different statistical ARIMA models. This
50 framework demonstrated reliable forecasting performance and substantially outcompeted the
51 ARIMA models. The forecasting performance was consistently best for the ensemble sub-
52 epidemic models incorporating a higher number of top-ranking sub-epidemic models. The
53 ensemble model incorporating the top four ranking sub-epidemic models consistently yielded the

54 best performance, particularly in terms of the coverage rate of the 95% prediction interval and
55 the weighted interval score. This framework can be applied to forecast other growth processes
56 found in nature and society including the spread of information through social media.

57

58 **Introduction**

59
60 The coronavirus disease 2019 (COVID-19) pandemic has amplified the critical need for reliable
61 tools to forecast the trajectory of epidemics and pandemics in near real-time. During the early
62 stages of the COVID-19 pandemic, multiple modeling teams embarked on the challenging task
63 of producing short-term forecasts of the course of the COVID-19 pandemic in terms of the
64 trajectory for the number of new cases, hospitalizations, or deaths (e.g., [1-10]) . Soon after the
65 epidemic started, our research team published short-term forecasts of the pandemic during the
66 early outbreaks of the novel coronavirus in China [4] and subsequently focused on producing
67 weekly forecasts for the USA [11]. In a related effort, the US COVID-19 Forecasting Hub
68 brought together multiple research teams to synthesize weekly short-term forecasts of the
69 COVID-19 pandemic in the USA [12]. It is time to systematically and rigorously evaluate the
70 forecasting performance of these different pandemic forecasting efforts and document the
71 lessons learned to continue advancing our understanding of epidemic forecasting.

72
73 Ensemble modeling approaches and models that integrate sub-epidemics to capture complex
74 temporal dynamics have demonstrated powerful forecasting capability (e.g., [13] [14-17]). In
75 prior work, we developed a sub-epidemic modeling framework to characterize and improve
76 forecasting accuracy during complex epidemic waves [13]. This mathematical framework
77 characterizes epidemic curves by aggregating multiple asynchronous sub-epidemics and
78 outperforms simpler growth models at providing short-term forecasts of various infectious
79 disease outbreaks [13, 18]. It is possible to model sub-epidemics associated with transmission
80 chains that are asynchronously triggered and progress somewhat independently from the other

81 sub-epidemics. This framework supports a family of sub-epidemic models that yield similar fits
82 to the calibration data, but their corresponding forecasts could produce diverging trajectories.

83
84 Ensemble modeling aims to boost forecasting performance by systematically integrating the
85 predictive accuracy tied to individual models [16, 19-21]. Past work indicates that multimodel
86 ensemble approaches are powerful forecasting tools that frequently outperform individual
87 models in epidemic forecasts [14, 15, 22-27]. We extend prior sub-epidemic modeling work and
88 propose an ensemble sub-epidemic modeling framework for forecasting the trajectory of
89 epidemics and pandemics. In this model, the sub-epidemics can start at different time points and
90 may follow different growth rates, scaling of growth, and final sizes. Hence, this ensemble
91 modeling framework can characterize more diverse epidemic patterns which were impossible to
92 capture by earlier sub-epidemic models, including plateaus, epidemic resurgences, and epidemic
93 waves characterized by multiple peaks of different sizes.

94
95 Here, we systematically assess the calibration and short-term forecasting performance in weekly
96 10-30 day forecasts in the context of the COVID-19 pandemic in the USA from late April 2020
97 to late February 2022, including the Omicron-dominated wave. We then compare the
98 performance of the ensemble modeling framework with a set of Autoregressive Integrated
99 Moving Average (ARIMA) models, following the EPIFORGE 2020 guidelines to report
100 epidemic forecasts [28]. Our extended ensemble modeling framework substantially outperforms
101 individual top-ranking sub-epidemic models and the ARIMA models based on standard
102 performance metrics that account for the uncertainty of the predictions.

103

104 **Data**

105 We used daily COVID-19 deaths reported in the USA from the publicly available data tracking
106 system of the Johns Hopkins Center for Systems Science and Engineering (CSSE) from 27
107 February 2020 to 30 March 2022 [29]. The data is updated on the CSSE webpage once every
108 day at 23:59 (UTC) and is read from the daily case report. The data is also publicly available in
109 the GitHub repository [30].

110

111 ***n*-sub-epidemic model**

112

113 We model epidemic trajectories comprised of one or more overlapping and asynchronous sub-
114 epidemics. That is, the sub-epidemics are used as building blocks to characterize more complex
115 epidemic trajectories. The mathematical equation for the sub-epidemic building block is the 3-
116 parameter generalized-logistic growth model (GLM), which has performed well in short-term
117 forecasts of single outbreak trajectories for different infectious diseases, including COVID-19
118 [31-33]. This model is given by the following differential equation:

119

□

$$120 \quad \frac{dC(t)}{dt} = C'(t) = rC^p(t) \left(1 - \frac{C(t)}{K_0}\right),$$

121

122 where $\frac{dC(t)}{dt}$ describes the curve of daily deaths over time t . The cumulative curve at time t is
123 given by $C(t)$, while r is a positive parameter denoting the growth rate per unit of time, K_0 is
124 the final outbreak size, and $p \in [0,1]$ is the “scaling of growth” parameter which allows the
125 model to capture early sub-exponential and exponential growth patterns. If $p = 0$, this equation
126 describes a constant number of new deaths over time, while $p = 1$ indicates that the early

127 growth phase is exponential. Intermediate values of p ($0 < p < 1$) describe early sub-
128 exponential (e.g., polynomial) growth dynamics.

129

130 An n -sub-epidemic trajectory comprises n overlapping sub-epidemics and is given by the
131 following system of coupled differential equations:

132

$$\frac{dC_i(t)}{dt} = C_i'(t) = A_i(t)r_i C_i^{p_i}(t) \left(1 - \frac{C_i(t)}{K_{0_i}}\right),$$

133

134 Where $C_i(t)$ tracks the cumulative number of deaths for sub-epidemic i , and the parameters that
135 characterize the shape of the i_{th} sub-epidemic are given by (r_i, p_i, K_{0_i}) , for $i = 1, \dots, n$. Thus, the
136 1-sub-epidemic model is equivalent to the generalized growth model described above. When
137 $n > 1$, we model the onset timing of the $(i + 1)_{th}$ sub-epidemic, where $(i + 1) \leq n$, by
138 employing an indicator variable given by $A_i(t)$ so that the $(i + 1)_{th}$ sub-epidemic is triggered
139 when the cumulative curve of the i_{th} sub-epidemic exceeds C_{thr} .

140

141 The $(i + 1)_{th}$ sub-epidemic is only triggered when $C_{thr} \leq K_{0_i}$. Hence, we have:

142

$$A_i(t) = \begin{cases} 1, & C_{i-1}(t) > C_{thr} \\ 0, & \text{Otherwise} \end{cases} \quad i = 2, \dots, n,$$

144

145 where $A_1(t) = 1$ for the first sub-epidemic. Hence, the total number of parameters that are needed
146 to model an n -sub-epidemic trajectory is given by $3n + 1$. The initial number of deaths is given

147 by $C_1(0) = I_0$, where I_0 is the initial number of deaths in the observed data. The cumulative
148 curve of the n -sub-epidemic trajectory is given by:

$$C_{tot}(t) = \sum_{i=1}^n C_i(t).$$

149
150 The n -sub-epidemic wave model can characterize diverse epidemic patterns, including epidemic
151 plateaus where the epidemic stabilizes at a high level for an extended period, epidemic
152 resurgences where the number of cases increases again after a low incidence period, and
153 epidemic waves characterized by multiple peaks.

154 155 **Parameter estimation**

156
157 To reduce the noise in the original data due to artificial reasons such as the weekend effects, we
158 use the 7-day moving average of daily death series to fit the n -sub-epidemic model. Let

159
$$y_{t_j} = y_{t_1}, y_{t_2}, \dots, y_{t_{n_d}} \text{ where } j = 1, 2, \dots, n_d$$

160 denote the smoothed daily COVID-19 death series of the epidemic trajectory based on the
161 moving average. Here, t_j are the time points for the time series data, n_d is the number of
162 observations, and each y_{t_j} , $j=1,2,\dots,n_d$, is the average of the death counts at the neighboring
163 seven days $(t_{j-3}, t_{j-2}, t_{j-1}, t_j, t_{j+1}, t_{j+2}, t_{j+3})$. We will use this smoothed data to estimate a total
164 of $3n + 1$ model parameters, namely $\Theta = (C_{thr}, r_1, p_1, K_{0_1}, \dots, r_n, p_n, K_{0_n})$. Let $f(t, \Theta)$ denote
165 the expected curve of new COVID-19 deaths of the epidemic's trajectory. We can estimate
166 model parameters by fitting the model solution to the observed data via nonlinear least squares
167 [34] or via maximum likelihood estimation assuming a specific error structure [35]. For

168 nonlinear least squares, this is achieved by searching for the set of parameters $\hat{\Theta}$ that minimizes
169 the sum of squared differences between the observed data $y_{t_j} = y_{t_1}, y_{t_2}, \dots, y_{t_{n_d}}$ and the model
170 mean corresponds to $f(t, \Theta)$. That is, $\Theta = (C_{thr}, r_1, p_1, K_{0_1}, \dots, r_n, p_n, K_{0_n})$ is estimated by
171 $\hat{\Theta} = \arg \min \sum_{j=1}^{n_d} (f(t_j, \Theta) - y_{t_j})^2$.

172
173 This parameter estimation method weights each of the data points equally and does not require a
174 specific distributional assumption for y_t , except for the first moment $E[y_t] = f(t_i; \theta)$. That is,
175 the mean of the observed data at time t is equivalent to the expected count (e.g., number of
176 deaths) denoted by $f(t, \Theta)$ at time t [36]. This method yields asymptotically unbiased point
177 estimates regardless of any misspecification of the variance-covariance error structure. Hence,
178 the estimated model mean $f(t_i, \hat{\Theta})$ yields the best fit to observed data y_{t_i} in terms of squared L2
179 norm. In Matlab, we can use the *fmincon* function to set the optimization problem.

180
181 To quantify parameter uncertainty, we follow a parametric bootstrapping approach which allows
182 the computation of standard errors and related statistics in the absence of closed-form formulas
183 [37]. We generate B bootstrap samples from the best-fit model $f(t, \hat{\Theta})$, with an assumed error
184 structure, to quantify the uncertainty of the parameter estimates and construct confidence
185 intervals. Typically, the error structure in the data is modelled using a probability model such as
186 the Poisson or negative binomial distribution. Because the time-series data we are fitting to
187 involve large counts, the Poisson or negative binomial distribution can be well approximated by
188 a normal distribution for large numbers. So, using the best-fit model $f(t, \hat{\Theta})$, we generate B -
189 times replicated simulated datasets of size n_d , where the observation at time t_j is sampled from a

190 normal distribution with mean $f(t_j, \hat{\theta})$ and variance $\frac{\sum_{j=1}^{n_d} (f(t_j, \hat{\theta}) - y_{t_j})^2}{n_d - (3n + 1)}$. Next, we refit the model
191 to each of the B simulated datasets to re-estimate parameters for each. The new parameter
192 estimates for each realization are denoted by $\hat{\Theta}_b$ where $b = 1, 2, \dots, B$. Using the sets of re-
193 estimated parameters ($\hat{\Theta}_b$), it is possible to characterize the empirical distribution of each
194 estimate, calculate the variance, and construct confidence intervals for each parameter. The
195 resulting uncertainty around the model fit can similarly be obtained from $f(t, \hat{\Theta}_1)$,
196 $f(t, \hat{\Theta}_2), \dots, f(t, \hat{\Theta}_B)$.

197

198 *Model-based forecasts with quantified uncertainty*

199

200 Forecasting the model $f(t, \hat{\Theta})$, h days ahead provides an estimate for $f(t + h, \hat{\Theta})$. The
201 uncertainty of the forecasted value can be obtained using the previously described parametric
202 bootstrap method. Let

$$f(t + h, \hat{\Theta}_1), f(t + h, \hat{\Theta}_2), \dots, f(t + h, \hat{\Theta}_B)$$

203 denote the forecasted value of the current state of the system propagated by a horizon of h time
204 units, where $\hat{\Theta}_b$ denotes the estimation of parameter set Θ from the b_{th} bootstrap sample. We can
205 use these values to calculate the bootstrap variance as the measure of the uncertainty of the
206 forecasts and use the 2.5% and 97.5% percentiles to construct the 95% prediction intervals (PI).

207

208

209 **Model selection**

210

211 We considered a set of n -sub-epidemic models where $1 \leq n \leq 2$ and ranked them from best to
212 worst according to the AIC_c which is given by [38, 39]:

$$AIC_c = n_d \log(SSE) + 2m + \frac{2m(m+1)}{n_d - m - 1}$$

213

214 where $SSE = \sum_{j=1}^{n_d} (f(t_j, \hat{\Theta}) - y_{t_j})^2$, $m = 3n + 1$ is the number of model parameters, and n_d is
215 the number of data points. The AIC_c for the parameter estimation from the nonlinear least-
216 squares fit, which implicitly assumes normal distribution for error.

217

218 We selected the four top ranking sub-epidemic models for further analyses. We used them to
219 construct three ensemble sub-epidemic models, which we refer to as: Ensemble(2), Ensemble(3),
220 and Ensemble(4). The next section describes the process of constructing these ensemble models
221 from the top-ranking sub-epidemic models.

222

223

224 **Constructing Ensemble Models from top-ranking models**

225

226 Ensemble models that combine the strength of multiple models may exhibit significantly
227 enhanced predictive performance (e.g., [14-17]). Here we generate ensemble models from the
228 weighted combination of the highest-ranking sub-epidemic models as deemed by the AIC_{c_i} for
229 the i -th model where $AIC_{c_1} \leq \dots \leq AIC_{c_I}$ and $i = 1, \dots, I$. An ensemble derived from the top-
230 ranking I models is denoted by Ensemble(I) and illustrated in **Figure 1**. Thus, Ensemble(2) and
231 Ensemble(3) refer to the ensemble models generated from the weighted combination of the top-
232 ranking 2 and 3 models, respectively. We compute the weight w_i for the i -th model, $i = 1, \dots, I$,
233 where $\sum w_i = 1$ as follows:

234

235
$$w_i = \frac{\frac{1}{AIC_{c_i}}}{\frac{1}{AIC_{c_1}} + \frac{1}{AIC_{c_2}} + \dots + \frac{1}{AIC_{c_I}}} \text{ for all } i = 1, 2, \dots, I,$$

236

237 and hence $w_I \leq \dots \leq w_1$.

238

239 The estimated mean curve of daily COVID-19 deaths for the Ensemble(I) model is:

$$f_{ens(I)}(t) = \sum_{i=1}^I w_i f_i(t, \hat{\Theta}^{(i)})$$

240

241 where given the training data, $\hat{\Theta}^{(i)}$ denotes the set of estimated parameters, and $f_i(t, \hat{\Theta}^{(i)})$

242 denotes the estimated mean curve of daily COVID-19 deaths, for the i -th model. Accordingly,

243 we compute the weighted average and sample the bootstrap realizations of the forecasts for each

244 model to construct the 95% CI or PI using the 2.5% and 97.5% quantiles [16]. Our MATLAB

245 (The Mathworks, Inc) code for model fitting and forecasting is publicly available in the GitHub

246 repository [30].

247

248 **Figure 1.** Schematic diagram of the construction of the ensemble model from the weighted

249 combination of the highest-ranking sub-epidemic models as deemed by the AIC_{c_i} for the i -th

250 model where $AIC_{c_1} \leq \dots \leq AIC_{c_I}$ and $i = 1, \dots, I$. An ensemble derived from the top-ranking I

251 models is denoted by Ensemble(I).

252

253 As a sensitivity analysis, we also investigated how the ensemble sub-epidemic models performed

254 when the ensemble weights were proportional to the relative likelihood (l) rather than the

255 reciprocal of the AIC_c . Let AIC_{min} denote the minimum AIC from the set of models. The relative

256 likelihood of model i is given by $l_i = e^{((AIC_{min} - AIC_i)/2)}$ [40]. We compute the weight w_i for the
257 i -th model where $\sum w_i = 1$ as follows:

258

$$259 \quad w_i = \frac{l_i}{l_1 + l_2 + \dots + l_I} \quad \text{for all } i = 1, 2, \dots, I,$$

260

261 and hence $w_I \leq \dots \leq w_1$.

262

263 **Auto-regressive integrated moving average models (ARIMA)**

264

265 We also generated short-term predictions of the pandemic trajectory using ARIMA models to
266 compare their performance with that of the sub-epidemic models. ARIMA models have been
267 frequently employed to forecast trends in finance [41-43] and weather [44-46]. The ARIMA (p,
268 d, q) process is given by

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\epsilon_t$$

269 or equivalently as $\phi(B)(1 - B)^d (y_t - \mu t^d / d!) = \theta(B)\epsilon_t$, where p is the order of the AR
270 model, d is the degree of differencing, q is the order of the MA model, $\{\epsilon_t\}$ is a white noise
271 process with mean 0 and variance σ^2 , and B denotes the backshift operator. The p -order
272 polynomial $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ and the q -order polynomial $\theta(z) = 1 - \theta_1 z - \dots -$
273 $\theta_q z^q$ are assumed to have no roots inside the unit circle to ensure causality and invertibility. The
274 constant $c = \mu(1 - \phi_1 - \dots - \phi_p)$, and μ is the mean of $(1 - B)^d y_t$. When $d=0$, μ is the mean
275 of y_t .

276

277 The *auto.arima* function in the R package “forecast” is used to select orders and build the model
278 [47]. First, the degree of differencing $0 \leq d \leq 2$ is selected based on successive KPSS unit-
279 root tests [48], which test the data for a unit root; if the test result is significant, the differenced
280 data is tested for a unit root; and this procedure is repeated until the first insignificant result is
281 obtained. Then given d , the orders p and q are selected based on the AIC_c for the d -times
282 differenced data. For $d=0$ or $d=1$, a constant will be included if it improves the AIC_c value; for
283 $d>1$, the constant μ is fixed at 0 to avoid the model having a quadratic or higher order trend,
284 which is dangerous when forecasting. The final model is fitted using the maximum likelihood
285 estimation.

286

287 To guarantee the forecasted values and prediction intervals are above zero, we take the following
288 two strategies. In the first one, we conduct the ARIMA order selection and model fitting using
289 the log-transformed data. Then we take the exponential of the forecasted values and the PI
290 bounds to predict the incident death counts and get the PIs. We refer to this approach as the (log)
291 ARIMA throughout the manuscript. In the second case, the negative values are set as zero. Then,
292 it is possible that the actual coverage probability of such PIs can be smaller than the nominal
293 value (95%). We refer to this approach as ARIMA throughout the manuscript.

294

295 **Forecasting strategy and performance metrics**

296

297 We conducted short-term forecasts using the top-ranking n -sub-epidemic model ($1 \leq n \leq 2$)
298 and three ensemble models constructed with the top-ranking sub-epidemic models namely
299 Ensemble(2), Ensemble(3), and Ensemble(4). For comparison, we also generated short-term

300 forecasts using the previously described ARIMA models. Overall, we conducted 588 forecasts
301 across models.

302

303 Using a 90-day calibration period for each model, we conducted a total of 98 weekly sequential
304 10-day, 20-day and 30-day forecasts from 20 April 2020 to 28 February 2022, spanning five
305 pandemic waves. This range of forecasting horizons is comparable to that investigated in prior
306 COVID-19 forecasting studies [49]. This period covers the latter part of the early spring wave, a
307 summer wave in 2020, a fall-winter 2020/2021 wave, the summer-fall wave in 2021, and the
308 winter 2022 wave.

309

310 To assess the forecasting performance, we used four performance metrics: the mean absolute
311 error (MAE), the mean squared error (MSE), the coverage of the 95% prediction intervals, and
312 the mean interval score (MIS) [50]. The *mean absolute error* (MAE) is given by:

313

$$\text{MAE} = \frac{1}{N} \sum_{h=1}^N |f(t_h, \hat{\theta}) - \tilde{y}_{t_h}|.$$

314 Here \tilde{y}_{t_h} is the time series of the original death counts (unsmoothed) of the h -time units ahead
315 forecasts, where t_h are the time points of the time series data [51]. Similarly, the *mean squared*
316 *error* (MSE) is given by:

317

$$\text{MSE} = \frac{1}{N} \sum_{h=1}^N (f(t_h, \hat{\theta}) - \tilde{y}_{t_h})^2.$$

318 We also employed two metrics that account for prediction uncertainty: the *coverage rate of the*
319 *95% PI* e.g., the proportion of the observations that fall within the 95% PI as well as the
320 *weighted interval score* (WIS) [50, 52] which is a proper score. The WIS and the coverage rate
321 of the 95% PIs take into account the uncertainty of the predictions, whereas the MAE and MSE
322 only assess the closeness of the mean trajectory of the epidemic to the observations [53].

323
324 Recent epidemic forecasting studies have embraced the Interval Score (IS) for quantifying model
325 forecasting performance [18, 24, 49, 54]. The WIS provides quantiles of predictive forecast
326 distribution by combining a set of ISs for probabilistic forecasts. An IS is a simple proper score
327 that requires only a central $(1-\alpha)\times 100\%$ PI [50] and is described as

328

$$IS_{\alpha}(F, y) = (u - l) + \frac{2}{\alpha} \times (l - y) \times \mathbf{1}(y < l) + \frac{2}{\alpha} \times (y - u) \times \mathbf{1}(y > u).$$

329

330 In this equation $\mathbf{1}$ refers to the indicator function, meaning that $\mathbf{1}(y < l) = 1$ if $y < l$ and
331 0 otherwise. The terms l and u represent the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the forecast F . The IS
332 consists of three distinct quantities:

333

334 1. The sharpness of F , given by the width $u - l$ of the central $(1 - \alpha) \times$
335 100% PI.

336 2. A penalty term $\frac{2}{\alpha} \times (l - y) \times \mathbf{1}(y < l)$ for the observations that fall below
337 the lower end point l of the $(1 - \alpha) \times 100\%$ PI. This penalty term is

338 directly proportional to the distance between y and the lower end l of the
339 PI. The strength of the penalty depends on the level α .

340 3. An analogous penalty term $\frac{2}{\alpha} \times (y - u) \times \mathbf{1}(y > u)$ for the observations
341 falling above the upper limit u of the PI.

342
343 To provide more detailed and accurate information on the entire predictive distribution, we
344 report several central PIs at different levels $(1 - \alpha_1) < (1 - \alpha_2) < \dots < (1 - \alpha_K)$ along with
345 the predictive median, m , which can be seen as a central prediction interval at level $1 - \alpha_0 \rightarrow 0$.

346 This is referred to as the WIS, and it can be evaluated as follows:

347

$$WIS_{\alpha_{0:K}}(F, y) = \frac{1}{K + \frac{1}{2}} \cdot (w_0 \cdot |y - m| + \sum_{k=1}^K w_k \cdot IS_{\alpha_k}(F, y))$$

348
349 where, $w_k = \frac{\alpha_k}{2}$ for $k = 1, 2, \dots, K$ and $w_0 = \frac{1}{2}$. Hence, WIS can be interpreted as a measure of
350 how close the entire distribution is to the observation in units on the scale of the observed data
351 [10, 55].

352
353

354 **Results**

355
356 **Quality of the sub-epidemic model fits**

357
358 The best fit sub-epidemic model and three ensemble models constructed using the top-ranking
359 sub-epidemic models (Ensemble(2), Ensemble(3), Ensemble(4)) yielded similar quality fits to 98

360 sequential weekly calibration periods from 20-April-2020 to 28-February-2022 (Figure 2, Table
 361 1). For instance, the average WIS was ~247 with little variation across models (Table 1). The
 362 coverage rate of the 95% PIs averaged 97% and ranged from 91% to 100% during the study
 363 period. Moreover, all performance metrics displayed similar temporal trends (Figure 2).
 364

Model	Mean absolute error (MSE)	Mean squared error (MAE)	Percentage coverage of the 95% prediction interval	Weighted Interval Score (WIS)
Best fit sub-epidemic model	309260.00	394.74	97.06	247.28
Ensemble(2) model	308300.00	394.91	97.30	246.93
Ensemble(3) model	308620.00	395.24	97.46	247.09
Ensemble(4) model	309160.00	396.17	97.46	247.33

365 *The Ensemble(*i*) model incorporates the top *i* ranked sub-epidemic models in the ensemble as
 366 described in the text.

367
 368 **Table 1.** Mean performance metrics quantifying the quality of model fits across 98 sequential
 369 weekly calibration periods of the daily time series of COVID-19 deaths in the USA from 20-
 370 April-2020 through 22-February 2022.

371
 372 **Figure 2.** Performance metrics quantifying the quality of sub-epidemic model fits to 98
 373 sequential weekly calibration periods of the daily time series of COVID-19 deaths in the USA
 374 from 20-April-2020 through 22-February 2022. The best fit sub-epidemic model and three
 375 ensemble models constructed using the top-ranking sub-epidemic models (Ensemble(2),
 376 Ensemble(3), Ensemble(4)) yielded similar quality fits.

377
378 Representative fits of the top-ranking sub-epidemic models to the daily curve of COVID-19
379 deaths in the USA from 27-Feb-2020 to 20-April-2020 are shown in **Figure 3**. Although these
380 sub-epidemic models fit the data well, each of them results from the aggregation of two sub-
381 epidemics characterized by different growth rates, scaling of growth, and outbreak sizes as
382 shown in **Figure 4**.

383
384 **Figure 3.** Representative fits of the top-ranking sub-epidemic models to the daily curve of
385 COVID-19 deaths in the USA from 27-Feb-2020 to 20-April-2020. The sub-epidemic models
386 capture well the entire epidemic curve, including the latter plateau dynamics, by considering
387 models with two sub-epidemics. The best model fit (solid red line) and 95% prediction interval
388 (dashed red lines) are shown in the left panels. The cyan curves correspond to the associated
389 uncertainty from individual bootstrapped curves. The sub-epidemic profiles are shown in the
390 center panels, where the red and blue curves represent the two sub-epidemics and the grey curves
391 are the estimated epidemic trajectories. For each model fit, the residuals are also shown (right
392 panels). Black circles correspond to the data points.

393
394 **Figure 4.** Parameter estimates for the first (top panel) and the second sub-epidemics (bottom
395 panels) were derived for the top-ranking sub-epidemic model after fitting the sub-epidemic
396 modeling framework to the daily curve of COVID-19 deaths in the USA from 27-Feb-2020 to
397 20-April-2020 (see also **Figure 2**). Parameter estimates for both sub-epidemics are well
398 identified, as indicated by their relatively narrow bootstrap confidence intervals.

399

400 **Short-term forecasting performance**

401
 402 The best fit sub-epidemic model and three ensemble models constructed using the top-ranking
 403 sub-epidemic models (Ensemble(2), Ensemble(3), Ensemble(4)) consistently outperformed the
 404 ARIMA models in terms of the weighted interval score (WIS) and the coverage of the 95%
 405 prediction interval across the 10, 20 and 30 day short-term forecasts (Table 2). For instance, for
 406 30-day forecasts, the average WIS ranged from 377.6 to 421.3 for the sub-epidemic models,
 407 whereas, it ranged from 439.29 to 767.05 for the ARIMA models. Across 98 short-term
 408 forecasts, the Ensemble(4) outperformed the (log) ARIMA model 66.3% of the time and the
 409 ARIMA model 69.4% of the time in 30-day ahead forecasts in terms of the WIS (Figure 5 &
 410 Figure 6). Similarly, the coverage of the 95% PI ranged from 82.2% to 88.2% for the sub-
 411 epidemic models, whereas it ranged from 58% to 60.3% for the ARIMA models in 30-day
 412 forecasts. In terms of the coverage of the 95% PI, the Ensemble(4) outperformed the (log)
 413 ARIMA model 89.8% of the time and the ARIMA model 91.8% of the time (Figure 5 & Figure
 414 6). Forecasting performance generally improved as the number of top-ranking sub-epidemic
 415 models included in the ensemble increased (Table 1). The Ensemble(4) model consistently
 416 yielded the best performance in terms of the metrics that account for the uncertainty of the
 417 predictions.

418

Model	Mean absolute error (MSE)	Mean squared error (MAE)	Percentage coverage of the 95% prediction interval	Weighted Interval Score (WIS)
10 days ahead				
Top-ranked sub-epidemic model	551740.00	535.16	87.14	352.00
Ensemble(2) model	504560.00	516.44	88.88	331.83

Ensemble(3) model	491020.00	513.39	89.29	328.00
Ensemble(4) model	491740.00	513.14	89.39	326.56
(log) ARIMA model	424880.00	458.72	42.45	365.19
ARIMA model	430070.00	467.18	43.06	380.47
20 days ahead				
Top-ranked sub-epidemic model	646880.00	570.34	85.15	382.90
Ensemble(2) model	576700.00	544.35	88.57	354.04
Ensemble(3) model	558890.00	540.71	89.59	350.73
Ensemble(4) model	557130.00	539.30	89.44	346.83
(log) ARIMA model	591980.00	536.22	51.07	422.41
ARIMA model	538690.00	528.87	55.05	404.92
30 days ahead				
Top-ranked sub-epidemic model	749560.00	613.75	82.18	421.29
Ensemble(2) model	670740.00	586.52	87.35	383.36
Ensemble(3) model	650790.00	584.20	88.20	382.79
Ensemble(4) model	644270.00	579.77	88.16	377.64
(log) ARIMA model	818530.00	621.58	57.99	767.05
ARIMA model	656480.00	591.93	60.34	439.29

419 *The Ensemble(*i*) model incorporates the top *i* ranked sub-epidemic models in the ensemble as
 420 described in the text.

421
 422 **Table 2.** Mean forecasting performance metrics for the sub-epidemic models (ensemble weights
 423 are proportional to the reciprocal of the AICc) and the ARIMA models across 98 sequential
 424 weekly calibration periods of the daily time series of COVID-19 deaths in the USA from 20-
 425 April-2020 through 22-February 2022. Values highlighted in bold correspond to the best
 426 performance metrics.

427

428 **Figure 5.** Forecasting performance metrics for the (log) ARIMA model and the Ensemble(4)
429 model across 98 30-day forecasts. The symbol (^) indicates weekly forecasts where the
430 Ensemble(4) model outperformed the (log) ARIMA model. For example, the Ensemble(4)
431 outperformed the (log) ARIMA model 66.3% of the time in terms of the WIS and 89.8% of the
432 time in terms of the coverage rate of the 95% PI (Figure 4 & Figure 6).

433

434 **Figure 6.** Forecasting performance metrics for the ARIMA model and the Ensemble(4) model
435 across 98 30-day forecasts. The symbol (^) indicates weekly forecasts where the Ensemble(4)
436 model outperforms the ARIMA model. For instance, the Ensemble(4) outperformed the ARIMA
437 model 69.4% of the time in terms of the WIS and 91.8.8% of the time in terms of the coverage
438 rate of the 95% PI (Figure 4 & Figure 6).

439

440 In terms of the metrics based on point estimate information, the ARIMA models showed lower
441 overall MSE or MAE compared to the sub-epidemic models in 10 and 20-day forecasts, but the
442 Ensemble(4) achieved the best forecasting performance in 30-day forecasts (Table 2). Overall,
443 the forecasting performance deteriorated at longer forecasting horizons across all models
444 considered in our study.

445

446 Representative 30-day forecasts of the top-ranking sub-epidemic models to the daily curve of
447 COVID-19 deaths in the USA from 20-April-2020 to 20-May-2022 are shown in Figure 7. The
448 corresponding sub-epidemic profiles of the forecasts are shown in Figure 8. These models
449 support forecasts with diverging trajectories even though they yield similar fits to the calibration

450 period. For instance, the top-ranked sub-epidemic model predicts a decline in the mortality
451 curve, whereas the second-ranked model predicts a stable pattern during the next 30 days (Figure
452 7). The corresponding forecasts generated from three ensemble models (Ensemble(2),
453 Ensemble(3), Ensemble(4)) built from the top-ranking sub-epidemic models are shown in Figure
454 9. The individual 30-day ahead predictions across 98 forecasting periods generated by the
455 Ensemble(4) and the ARIMA models are available in the GitHub repository [30].

456
457 **Figure 7.** Representative 30-day forecasts of the top-ranking sub-epidemic models to the daily
458 curve of COVID-19 deaths in the USA from 20-April-2020 to 20-May-2020. The model fit
459 (solid line) and 95% prediction interval (shaded area) are also shown. The vertical line indicates
460 the start time of the forecast. Circles correspond to the data points. These four top-ranking
461 models support forecasts with diverging trajectories even though they yield similar fits to the
462 calibration period. For instance, the 1st ranked sub-epidemic model predicts a decline in the
463 mortality curve whereas the 2nd ranked model predicts a stable pattern during the next 30 days.

464
465 **Figure 8.** Representative sub-epidemic profiles of the forecasts derived from the top-ranking
466 sub-epidemic models to the daily curve of COVID-19 deaths in the USA from 20-April-2020 to
467 20-May-2022. The model fit (solid line) and 95% prediction interval (shaded area) are also
468 shown. Black circles correspond to the calibration data. Blue and red curves represent different
469 sub-epidemics of the epidemic wave profile. Gray curves correspond to the overall epidemic
470 trajectory obtained by aggregating the sub-epidemic curves. The vertical line indicates the start
471 time of the forecast.

472

473 **Figure 9.** Representative sub-epidemic ensemble model forecasts (Ensemble(2), Ensemble(3),
 474 Ensemble(4)) of COVID-19 deaths in the USA from 20-April-2020 to 20-May-2022. Circles
 475 correspond to the data points. The model fits (solid line) and 95% prediction intervals (shaded
 476 area) are shown. Circles correspond to the data points. The vertical line indicates the start time of
 477 the forecast

478
 479 In sensitivity analyses, defining ensemble weights as proportional to the relative likelihood did
 480 not achieve better performance relative to the ensemble models generated using weights
 481 proportional to the reciprocal of the AIC_c . Moreover, the rank of the ensemble models was not
 482 affected by the type of weights (**Table 3**).

483

Model	Mean absolute error (MSE)	Mean squared error (MAE)	Percentage coverage of the 95% prediction interval	Weighted Interval Score (WIS)
10 days ahead				
Top-ranked sub-epidemic model	551740.00	535.16	87.14	352.00
Ensemble(2) model	548540.00	534.14	87.25	348.66
Ensemble(3) model	547220.00	533.51	87.25	347.99
Ensemble(4) model	546350.00	533.23	87.35	347.60
(log) ARIMA model	424880.00	458.72	42.45	365.19
ARIMA model	430070.00	467.18	43.06	380.47
20 days ahead				
Top-ranked sub-epidemic model	646880.00	570.34	85.15	382.90
Ensemble(2) model	640240.00	567.90	85.71	377.27

Ensemble(3) model	640960.00	568.45	85.71	376.67
Ensemble(4) model	639280.00	567.74	85.56	376.36
(log) ARIMA model	591980.00	536.22	51.07	422.41
ARIMA model	538690.00	528.87	55.05	404.92
30 days ahead				
Top-ranked sub-epidemic model	749560.00	613.75	82.18	421.29
Ensemble(2) model	744130.00	612.63	82.65	414.72
Ensemble(3) model	745230.00	613.21	82.59	414.54
Ensemble(4) model	743020.00	612.48	82.52	414.16
(log) ARIMA model	818530.00	621.58	57.99	767.05
ARIMA model	656480.00	591.93	60.34	439.29

484

485

486 **Table 3.** Mean forecasting performance metrics for the sub-epidemic models (ensemble weights
 487 were based on the relative likelihood) and the ARIMA models across 98 sequential weekly
 488 calibration periods of the daily time series of COVID-19 deaths in the USA from 20-April-2020
 489 through 22-February 2022. Values highlighted in bold correspond to the best performance
 490 metrics.

491

492

493 **Discussion**

494

495 Our ensemble sub-epidemic modeling approach outperformed individual top-ranking sub-
 496 epidemic models and a set of ARIMA models in weekly short-term forecasts covering the
 497 national trajectory of the COVID-19 pandemic in the USA from the early growth phase up until

498 the Omicron-dominated wave. This framework has demonstrated reliable forecasting
499 performance across different pandemic phases from the early growth phase characterized by
500 exponential or sub-exponential growth dynamics to plateaus and new disease surges driven by
501 the relaxation of social distancing policies or the emergence of new variants. Importantly, we
502 found that forecasting performance consistently improved for the ensemble sub-epidemic models
503 that incorporated a higher number of top-ranking sub-epidemic models. The ensemble model
504 incorporating the top four ranking sub-epidemic models consistently yielded the best
505 performance, particularly in terms of the coverage rate of the 95% prediction interval and the
506 weighted interval score.

507
508 Our findings support the power of ensemble modeling approaches (e.g.,[14-17]). Our ensemble
509 modeling framework derived from a family of sub-epidemic models demonstrated improved
510 performance as the number of top-ranking sub-epidemic models included in the ensemble
511 increased. Prior studies have documented the potential of ensemble models to enhance
512 forecasting performance during multi-epidemic periods [14]. For instance, in the context of
513 influenza, one study utilized “weighted density ensembles” for predicting timing and severity
514 metrics and found that the performance of the ensemble model was comparable to that of the top
515 individual model, albeit the ensemble’s forecasts were more stable across influenza seasons [17].
516 In the context of dengue in Puerto Rico, another study found that forecasts derived from
517 Bayesian averaging ensembles outperformed a set of individual models [25]. Results from the
518 US COVID-19 Forecasting Hub CDC were consistent with our findings in that a multimodel
519 ensemble frequently outperformed the set of individual models.

520

521 We also evaluated short-term forecasting performance by a set of ARIMA models, as prior
522 studies have underscored the value of ARIMA models in epidemic forecasting [56], by providing
523 a relatively simple and transparent approach to forecasting. For instance, in the context of
524 influenza-like-illness in the USA, a set of ARIMA models provided reasonably accurate short-
525 term forecasts during the 2016/17 influenza season [57]. In another forecasting study during
526 multiple seasons of influenza in the USA, an ARIMA model yielded similar short-term
527 forecasting performance compared to other models based on the mechanistic SIR modeling
528 framework [58]. ARIMA models have also been used for spatial prediction of the COVID-19
529 epidemic [59, 60]. Another study [61] showed that the ARIMA model is more effective than the
530 Prophet time series model for forecasting COVID-19 prevalence. Finally, it is worth noting that
531 the US COVID-19 Forecast Hub did not include an ARIMA model in its set of evaluated models
532 [49]. Therefore, it is interesting to assess how ARIMA models perform in the context of the
533 COVID-19 pandemic in the US.

534
535 Prior work has underscored the need to assess alternative ways of constructing ensembles from a
536 set of individual models [14, 16]. We explored two ways of constructing the ensembles by
537 relying on the AIC_c or the relative likelihood associated with the individual models. We found
538 that the short-term forecasting performance achieved by the ensemble models was not
539 significantly affected by the type of ensemble weights used to construct them although
540 performance using ensemble weights based on the reciprocal of the AIC_c was slightly better.
541 Further research could explore how different weighting strategies influence the forecasting
542 performance of ensemble modeling approaches.

543

544 Short-term forecasting is an essential attribute of the models. As prior studies have underscored,
545 longer-term forecasts are of value, but their dependability varies inversely with the time horizon.
546 Our 20 and 30-day forecasts are most valuable for monitoring, managing, and informing the
547 relaxing of social distancing requirements. The early detection of potential disease resurgence
548 can signal the need for strict distancing controls, and the reports of cases can identify the
549 geographic location of incubating sub-epidemics.

550

551 Our study is not exempt of limitations. Our analysis relied on daily time series data of COVID-
552 19 deaths in the USA, which is inherently noisy due to heterogeneous data reporting at fine
553 spatial scales (i.e., county-level) [62]. Noisy data complicate the ability of any mathematical
554 model to identify meaningful signals about the impact of transmission dynamics and control
555 interventions. To deal with the high noise levels in the data, we fitted the models to smoothed
556 time series rather than the actual daily series, as described in the parameter estimation section.
557 Other forecasting studies, including the US COVID-19 Forecasting Hub, have relied on weekly
558 death counts to address this issue [49]. Beyond the COVID-19 pandemic, there is a need to
559 establish benchmarks to systematically assess forecasting performance across a diverse catalog
560 of mathematical models and epidemic datasets involving multiple infectious diseases, social
561 contexts, and spatial scales.

562

563 While our analysis demonstrated the accuracy of our ensemble sub-epidemic modeling
564 framework in forecasting the COVID-19 pandemic, the same framework could be readily used to
565 forecast other epidemics irrespective of the type of disease and spatial scale involved. Beyond
566 infectious diseases, this framework could also be used to forecast other biological and social

567 growth processes, such as the epidemics of lung injury associated with e-cigarette use or vaping
568 and the viral spread of information through social media platforms.

569
570 In summary, our ensemble sub-epidemic models provided reliable short-term forecasts of the
571 trajectory of the COVID-19 pandemic in the USA involving multiple waves and outcompeted a
572 set of ARIMA models. The forecasting performance of the ensemble models improved with the
573 number of top-ranking sub-epidemic models included in the ensemble. This framework could be
574 readily applied to investigate the spread of epidemics and pandemics beyond COVID-19 and in a
575 range of problems in nature and society that would benefit from short-term predictions.

576
577 **Author contributions:** Conceptualization, G.C.; methodology, G.C., J.H, R.L, validation, G.C.,
578 R.L; formal analysis, G.C., R.L; investigation, G.C., R.L; resources, G.C., ; data curation, G.C.,
579 S.D.; writing—original draft preparation, G.C., R.L; writing, review, and editing, A.T., G.C.,
580 S.D., K.R., R.L., J.M.H., ; visualization, G.C, R.L; supervision, G.C., R.L; project
581 administration, G.C.; funding acquisition, G.C. All authors have read and agreed to the published
582 version of the manuscript.

583
584 **Funding:** G.C. is partially supported from NSF grants 1610429 and 1633381 and R01 GM
585 130900. A.T. and S.D. are supported by a 2CI fellowship from Georgia State University.

586
587 **Conflicts of Interest:** The authors declare no conflict of interest.

588
589 **References**

590

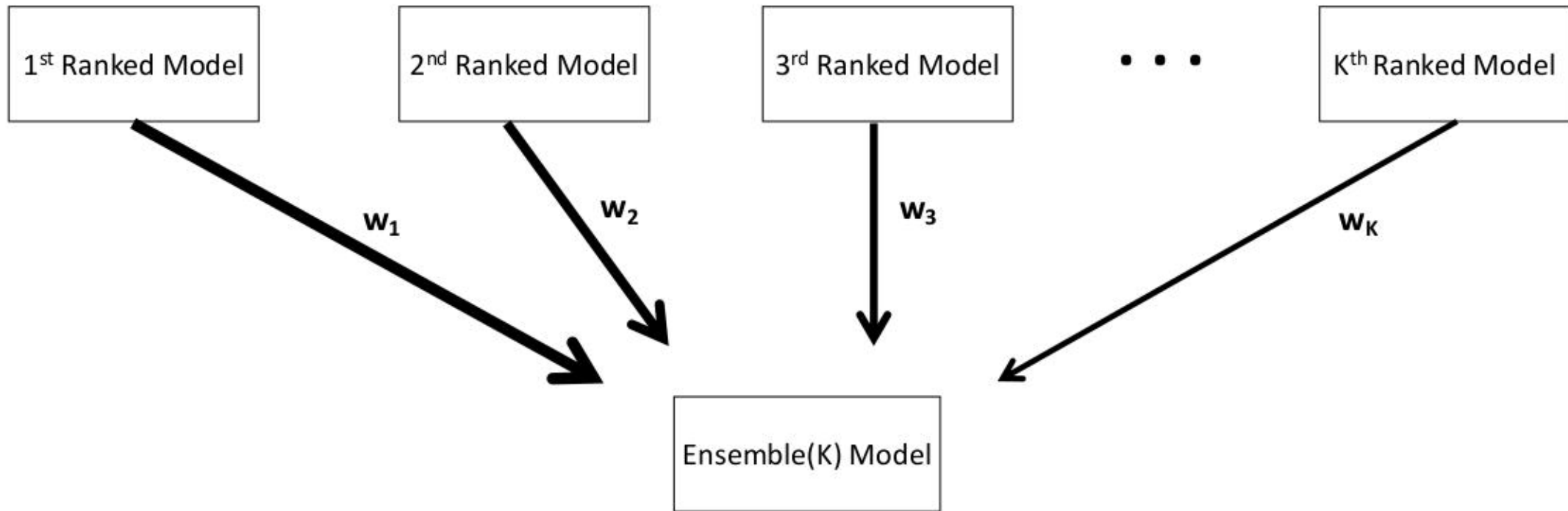
- 591 1. Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and
592 forecasting the spread of COVID-19. *Proc Natl Acad Sci U S A*. 2020;117(29):16732-8. Epub
593 2020/07/04. doi: 10.1073/pnas.2006520117. PubMed PMID: 32616574; PubMed Central
594 PMCID: PMCPMC7382213.
- 595 2. Farcomeni A, Maruotti A, Divino F, Jona-Lasinio G, Lovison G. An ensemble approach to
596 short-term forecast of COVID-19 intensive care occupancy in Italian regions. *Biometrical
597 Journal*. 2021;63(3):503-13. doi: <https://doi.org/10.1002/bimj.202000189>.
- 598 3. Tariq A, Undurraga EA, Laborde CC, Vogt-Geisse K, Luo R, Rothenberg R, et al.
599 Transmission dynamics and control of COVID-19 in Chile, March-October, 2020. *PLoS Neg Trop
600 Dis*. 2021;15(1):e0009070. doi: 10.1371/journal.pntd.0009070.
- 601 4. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Real-time forecasts of
602 the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect Dis Model*.
603 2020;5:256-63. doi: <https://doi.org/10.1016/j.idm.2020.02.002>.
- 604 5. Paireau J, Andronico A, Hozé N, Layan M, Crépey P, Roumagnac A, et al. An ensemble
605 model based on early predictors to forecast COVID-19 health care demand in France.
606 *Proceedings of the National Academy of Sciences*. 2022;119(18):e2103302119. doi:
607 doi:10.1073/pnas.2103302119.
- 608 6. Drews M, Kumar P, Singh RK, De La Sen M, Singh SS, Pandey AK, et al. Model-based
609 ensembles: Lessons learned from retrospective analysis of COVID-19 infection forecasts across
610 10 countries. *Science of The Total Environment*. 2022;806:150639. doi:
611 <https://doi.org/10.1016/j.scitotenv.2021.150639>.
- 612 7. Zhang S, Ponce J, Zhang Z, Lin G, Karniadakis G. An integrated framework for building
613 trustworthy data-driven epidemiological models: Application to the COVID-19 outbreak in New
614 York City. *PLOS Computational Biology*. 2021;17(9):e1009334. doi:
615 10.1371/journal.pcbi.1009334.
- 616 8. Watson GL, Xiong D, Zhang L, Zoller JA, Shamshoian J, Sundin P, et al. Pandemic velocity:
617 Forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental
618 model. *PLOS Computational Biology*. 2021;17(3):e1008837. doi: 10.1371/journal.pcbi.1008837.
- 619 9. Català M, Alonso S, Alvarez-Lacalle E, López D, Cardona P-J, Prats C. Empirical model for
620 short-time prediction of COVID-19 spreading. *PLOS Computational Biology*.
621 2020;16(12):e1008431. doi: 10.1371/journal.pcbi.1008431.
- 622 10. Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Castro Rivadeneira AJ, et al.
623 Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the
624 United States. *Proc Natl Acad Sci U S A*. 2022;119(15):e2113561119. Epub 2022/04/09. doi:
625 10.1073/pnas.2113561119. PubMed PMID: 35394862.
- 626 11. Chowell G, Tariq A, Dahal S, Roosa K. Forecasts of national COVID-19 incidence in the
627 United States Georgia State University, School of Public Health. Epidemic Forecasting Center:
628 GSU; 2022 [cited 2022 May 3]. Available from:
629 <https://publichealth.gsu.edu/research/coronavirus/>.
- 630 12. CDC. The COVID-19 forecast hub 2021 [cited 2021 November 20]. Available from:
631 <https://covid19forecasthub.org/>.
- 632 13. Chowell G, Tariq A, Hyman JM. A novel sub-epidemic modeling framework for short-
633 term forecasting epidemic waves. *BMC Med*. 2019;17(1):164. doi: 10.1186/s12916-019-1406-6.

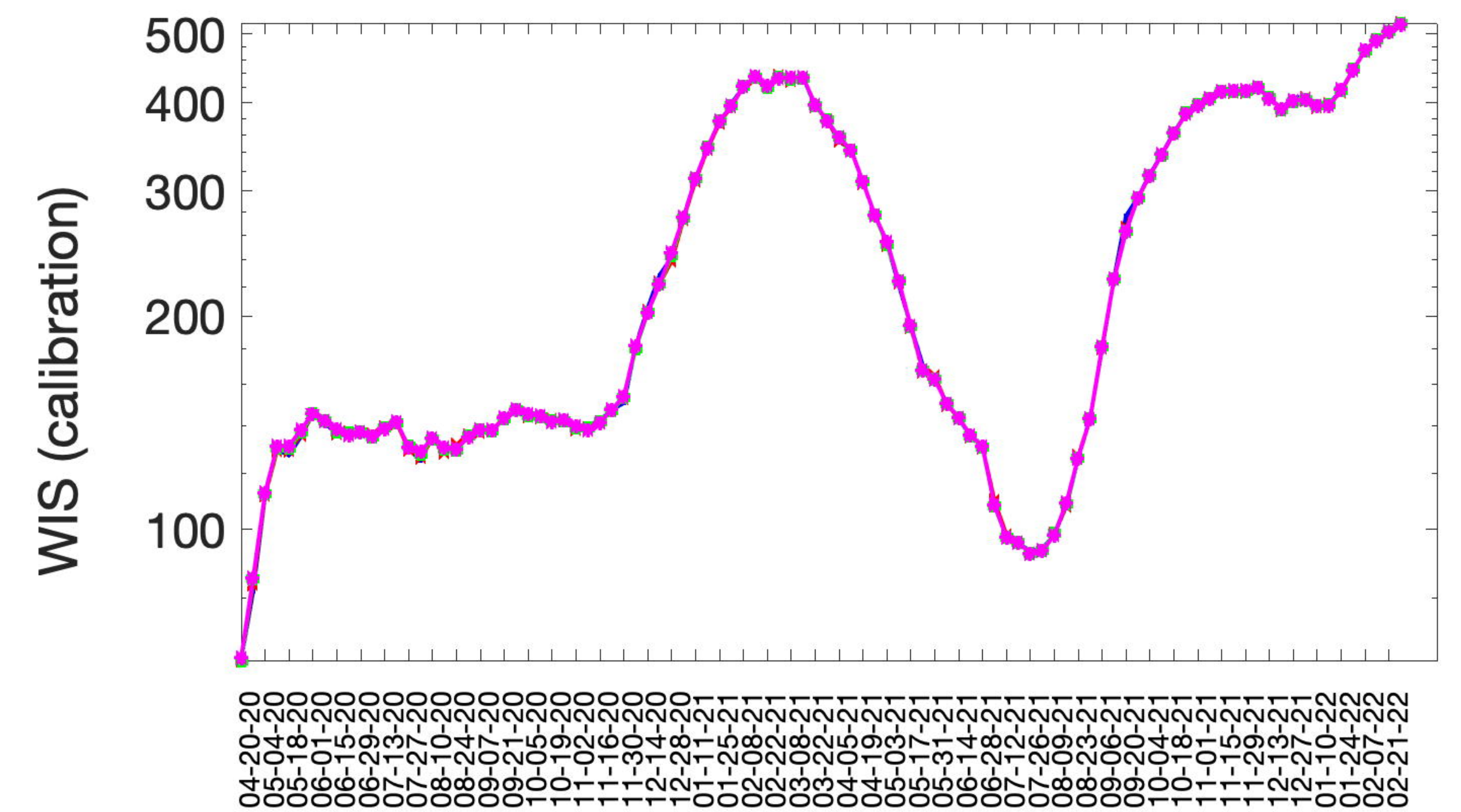
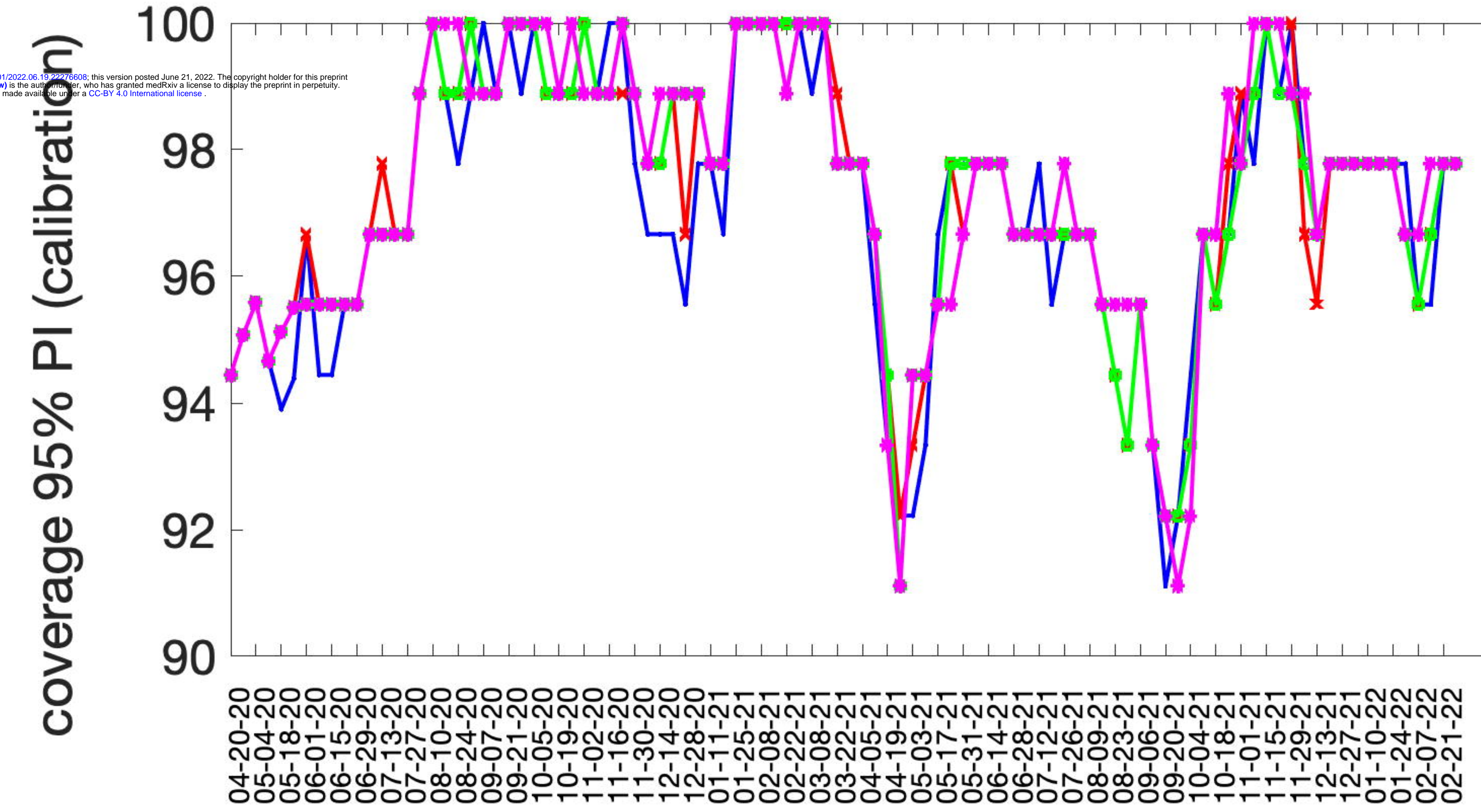
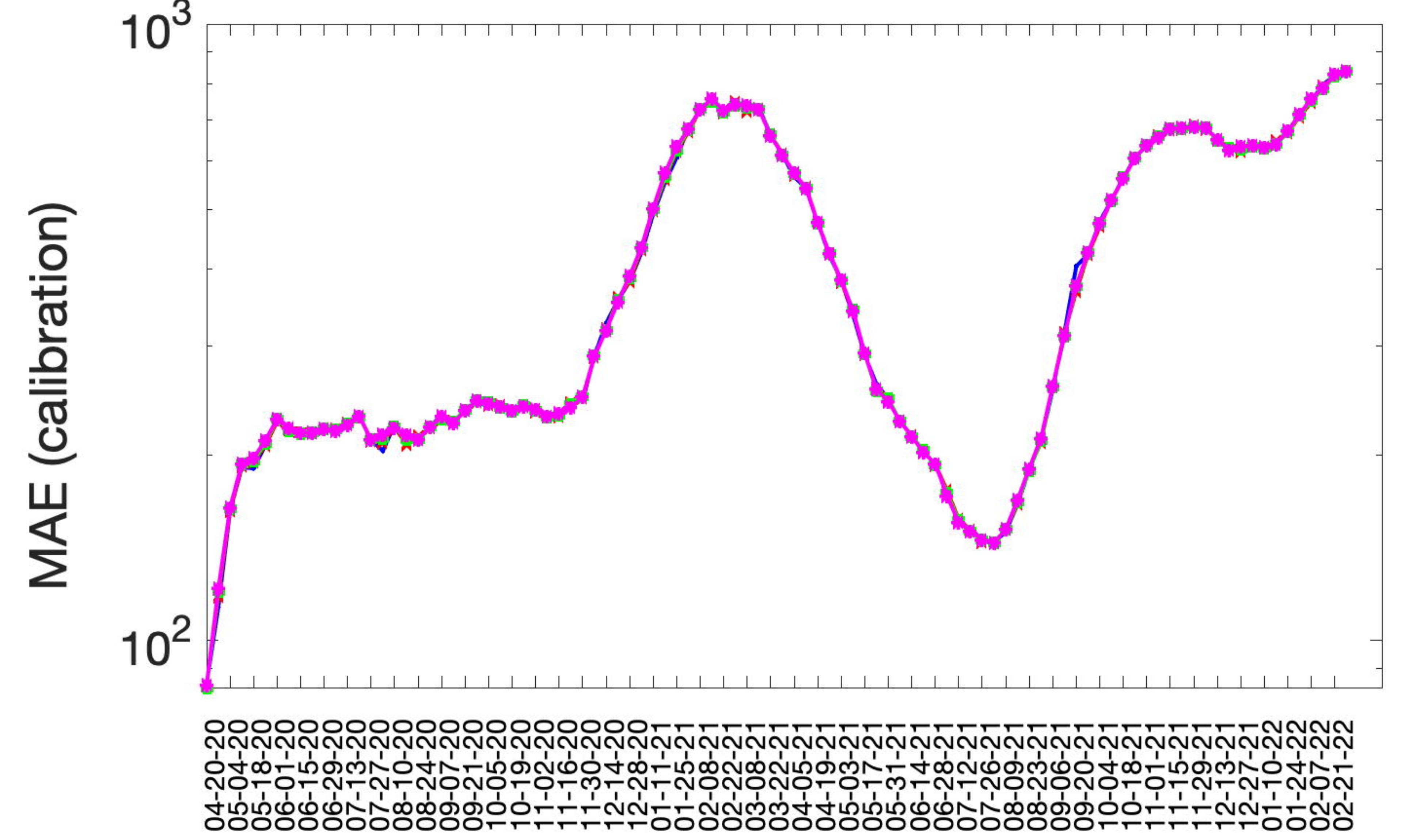
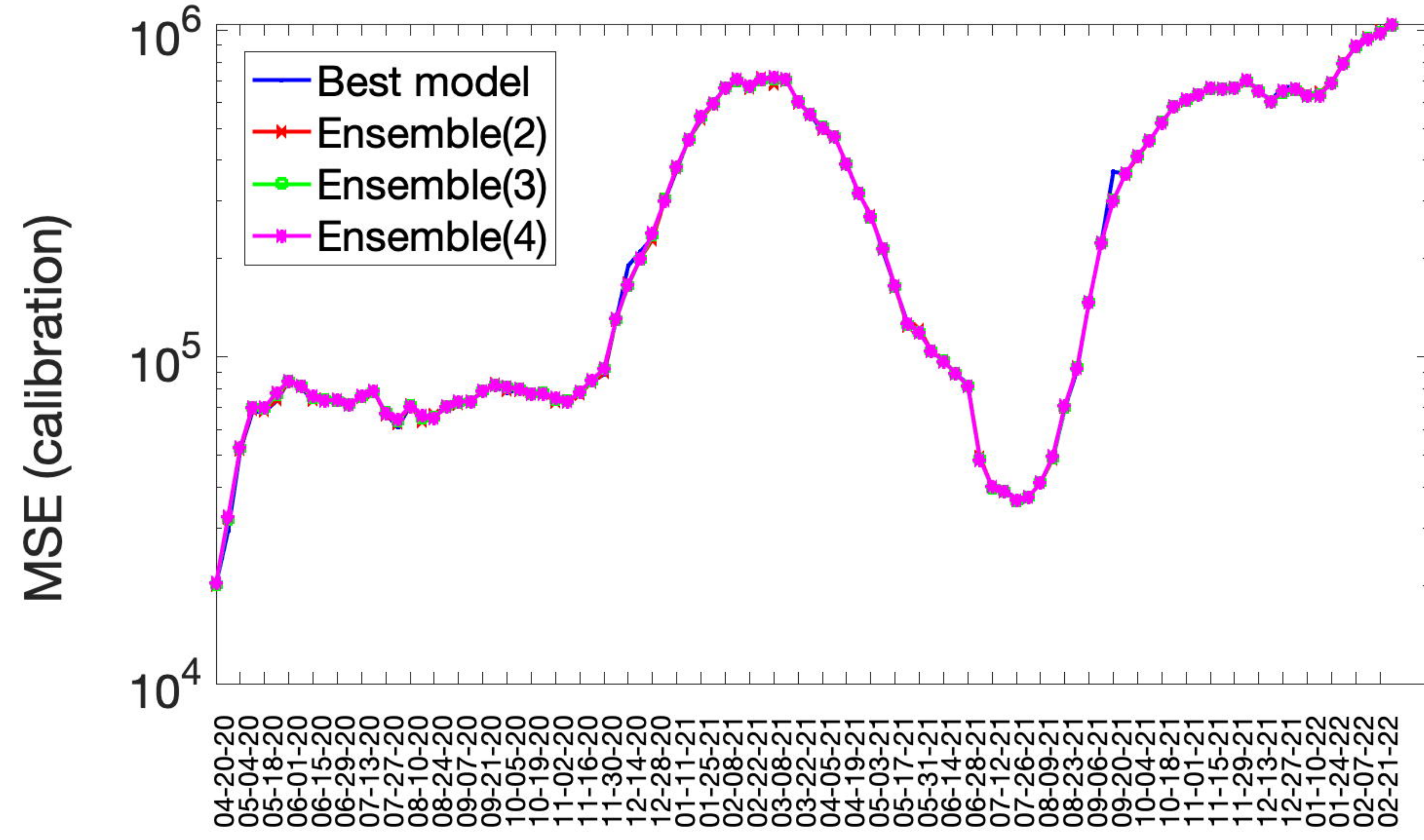
- 634 14. Chowell G, Luo R, Sun K, Roosa K, Tariq A, Viboud C. Real-time forecasting of epidemic
635 trajectories using computational dynamic ensembles. *Epidemics*. 2020;30:100379. doi:
636 <https://doi.org/10.1016/j.epidem.2019.100379>.
- 637 15. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, et al. The RAPIDD ebola
638 forecasting challenge: Synthesis and lessons learnt. *Epidemics*. 2018;22:13-21. Epub
639 2017/09/30. doi: 10.1016/j.epidem.2017.08.002. PubMed PMID: 28958414; PubMed Central
640 PMCID: PMC5927600.
- 641 16. Chowell G, Luo R. Ensemble bootstrap methodology for forecasting dynamic growth
642 processes using differential equations: application to epidemic outbreaks. *BMC Medical
643 Research Methodology*. 2021;21(1):34. doi: 10.1186/s12874-021-01226-9.
- 644 17. Ray EL, Reich NG. Prediction of infectious disease epidemics via weighted density
645 ensembles. *PLoS Comput Biol*. 2018;14(2):e1005910. Epub 2018/02/21. doi:
646 10.1371/journal.pcbi.1005910. PubMed PMID: 29462167; PubMed Central PMCID:
647 PMC5834190.
- 648 18. Tariq A, Chakhaia T, Dahal S, Ewing A, Hua X, Ofori SK, et al. An investigation of spatial-
649 temporal patterns and predictions of the coronavirus 2019 pandemic in Colombia, 2020-2021.
650 *PLoS Negl Trop Dis*. 2022;16(3):e0010228. Epub 2022/03/05. doi:
651 10.1371/journal.pntd.0010228. PubMed PMID: 35245285; PubMed Central PMCID:
652 PMC8926206.
- 653 19. Tebaldi C, Knutti R. The use of the multi-model ensemble in probabilistic climate
654 projections. *Philos Trans A Math Phys Eng Sci*. 2007;365(1857):2053-75. Epub 2007/06/16. doi:
655 10.1098/rsta.2007.2076. PubMed PMID: 17569654.
- 656 20. Lindström T, Tildesley M, Webb C. A Bayesian ensemble approach for epidemiological
657 projections. *PLoS Comput Biol*. 2015;11(4):e1004187. Epub 2015/05/01. doi:
658 10.1371/journal.pcbi.1004187. PubMed PMID: 25927892; PubMed Central PMCID:
659 PMC4415763.
- 660 21. Smith T, Ross A, Maire N, Chitnis N, Studer A, Hardy D, et al. Ensemble modeling of the
661 likely public health impact of a pre-erythrocytic malaria vaccine. *PLoS Med*.
662 2012;9(1):e1001157. Epub 2012/01/25. doi: 10.1371/journal.pmed.1001157. PubMed PMID:
663 22272189; PubMed Central PMCID: PMC3260300 employed by the PATH Malaria
664 Organization which was supporting the development of RTS,S, the vaccine which is the focus of
665 this paper. AB left PATH prior to any collaboration on this paper. All other authors have
666 declared no competing interests. The views expressed are those of the authors.
- 667 22. McGowan CJ, Biggerstaff M, Johansson M, Apfeldorf KM, Ben-Nun M, Brooks L, et al.
668 Collaborative efforts to forecast seasonal influenza in the United States, 2015-2016. *Sci Rep*.
669 2019;9(1):683. Epub 2019/01/27. doi: 10.1038/s41598-018-36361-9. PubMed PMID: 30679458;
670 PubMed Central PMCID: PMC6346105 consulting for S.K. Analytics. The remaining authors
671 declare no competing interests.
- 672 23. Johansson MA, Apfeldorf KM, Dobson S, Devita J, Buczak AL, Baugher B, et al. An open
673 challenge to advance probabilistic forecasting for dengue epidemics. *Proc Natl Acad Sci U S A*.
674 2019;116(48):24268-74. Epub 2019/11/13. doi: 10.1073/pnas.1909865116. PubMed PMID:
675 31712420; PubMed Central PMCID: PMC6883829.

- 676 24. Roosa K, Tariq A, Yan P, Hyman JM, Chowell G. Multi-model forecasts of the ongoing
677 Ebola epidemic in the Democratic Republic of Congo, March 2013-October 2019. *J R Soc*
678 *Interface*. 2020;17(169):20200447. doi: doi:10.1098/rsif.2020.0447.
- 679 25. Yamana TK, Kandula S, Shaman J. Superensemble forecasts of dengue outbreaks. *J R Soc*
680 *Interface*. 2016;13(123). Epub 2016/10/14. doi: 10.1098/rsif.2016.0410. PubMed PMID:
681 27733698; PubMed Central PMCID: PMC5095208.
- 682 26. Novaes de Amorim A, Deardon R, Saini V. A stacked ensemble method for forecasting
683 influenza-like illness visit volumes at emergency departments. *PLOS ONE*. 2021;16(3):e0241725.
684 doi: 10.1371/journal.pone.0241725.
- 685 27. Kim J-S, Kavak H, Züfle A, Anderson T. COVID-19 ensemble models using representative
686 clustering. *SIGSPATIAL Special*. 2020;12(2):33–41. doi: 10.1145/3431843.3431848.
- 687 28. Pollett S, Johansson MA, Reich NG, Brett-Major D, Del Valle SY, Venkatramanan S, et al.
688 Recommended reporting items for epidemic forecasting and prediction research: The EPIFORGE
689 2020 guidelines. *PLOS Medicine*. 2021;18(10):e1003793. doi: 10.1371/journal.pmed.1003793.
- 690 29. CSSE Covid-19 Timeseries [Internet]. 2022 [cited May 20, 2022]. Available from:
691 [https://github.com/CSSEGISandData/COVID-](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirm_ed_US.csv)
692 [19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirm](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirm_ed_US.csv)
693 [ed_US.csv](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirm_ed_US.csv).
- 694 30. n-subepidemic ensemble modeling framework [Internet]. 2022. Available from:
695 [https://github.com/atariq2891/An-ensemble-n-sub-epidemic-modeling-framework-for-short-](https://github.com/atariq2891/An-ensemble-n-sub-epidemic-modeling-framework-for-short-term-forecasting-epidemic-trajectories)
696 [term-forecasting-epidemic-trajectories](https://github.com/atariq2891/An-ensemble-n-sub-epidemic-modeling-framework-for-short-term-forecasting-epidemic-trajectories)
- 697 31. Shanafelt DW, Jones G, Lima M, Perrings C, Chowell G. Forecasting the 2001 Foot-and-
698 Mouth Disease Epidemic in the UK. *Ecohealth*. 2017. Epub 2017/12/15. doi: 10.1007/s10393-
699 017-1293-2. PubMed PMID: 29238900.
- 700 32. Chowell G, Hincapie-Palacio D, Ospina J, Pell B, Tariq A, Dahal S, et al. Using
701 Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final
702 Burden of Zika Epidemics. *PLoS Curr*. 2016;8. Epub 2016/07/02. doi:
703 10.1371/currents.outbreaks.f14b2217c902f453d9320a43a35b9583. PubMed PMID: 27366586;
704 PubMed Central PMCID: PMC4922743.
- 705 33. Pell B, Kuang Y, Viboud C, Chowell G. Using phenomenological models for forecasting
706 the 2015 Ebola challenge. *Epidemics*. 2018;22:62-70. Epub 2016/12/04. doi:
707 10.1016/j.epidem.2016.11.002. PubMed PMID: 27913131.
- 708 34. Banks HT, Hu S, Thompson WC. Modeling and inverse problems in the presence of
709 uncertainty: CRC Press; 2014.
- 710 35. Roosa K, Luo R, Chowell G. Comparative assessment of parameter estimation methods
711 in the presence of overdispersion: a simulation study. *Math Biosci Eng*. 2019;16(5):4299-313.
712 Epub 2019/09/11. doi: 10.3934/mbe.2019214. PubMed PMID: 31499663.
- 713 36. Myung IJ. Tutorial on maximum likelihood estimation. *Journal of Mathematical*
714 *Psychology*; 2003. p. 90-100.
- 715 37. Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning : Data Mining,*
716 *Inference, and Prediction*. New York, NY.: Springer-Verlag New York; 2009.
- 717 38. Sugiura N. Further analysts of the data by akaike' s information criterion and the finite
718 corrections. *Communications in Statistics-theory and Methods*. 1978;7:13-26.

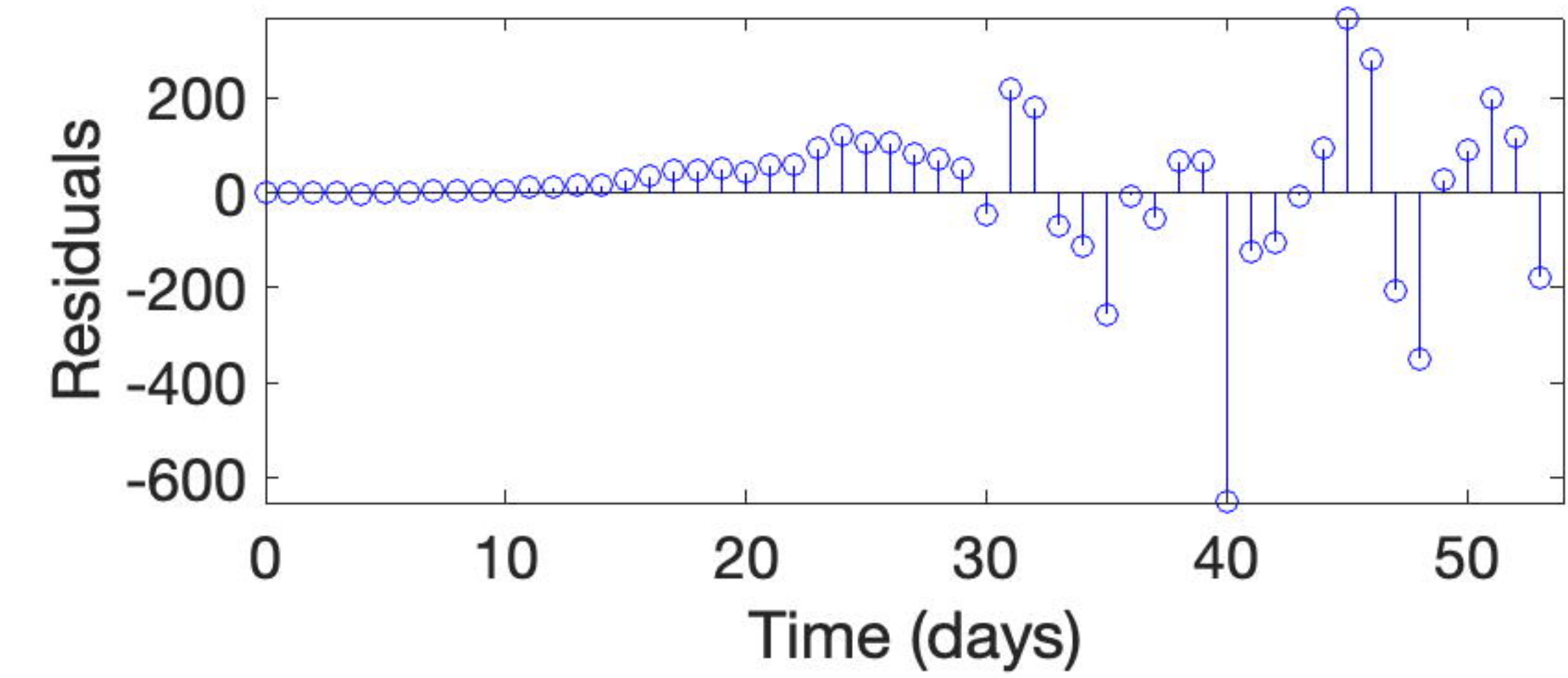
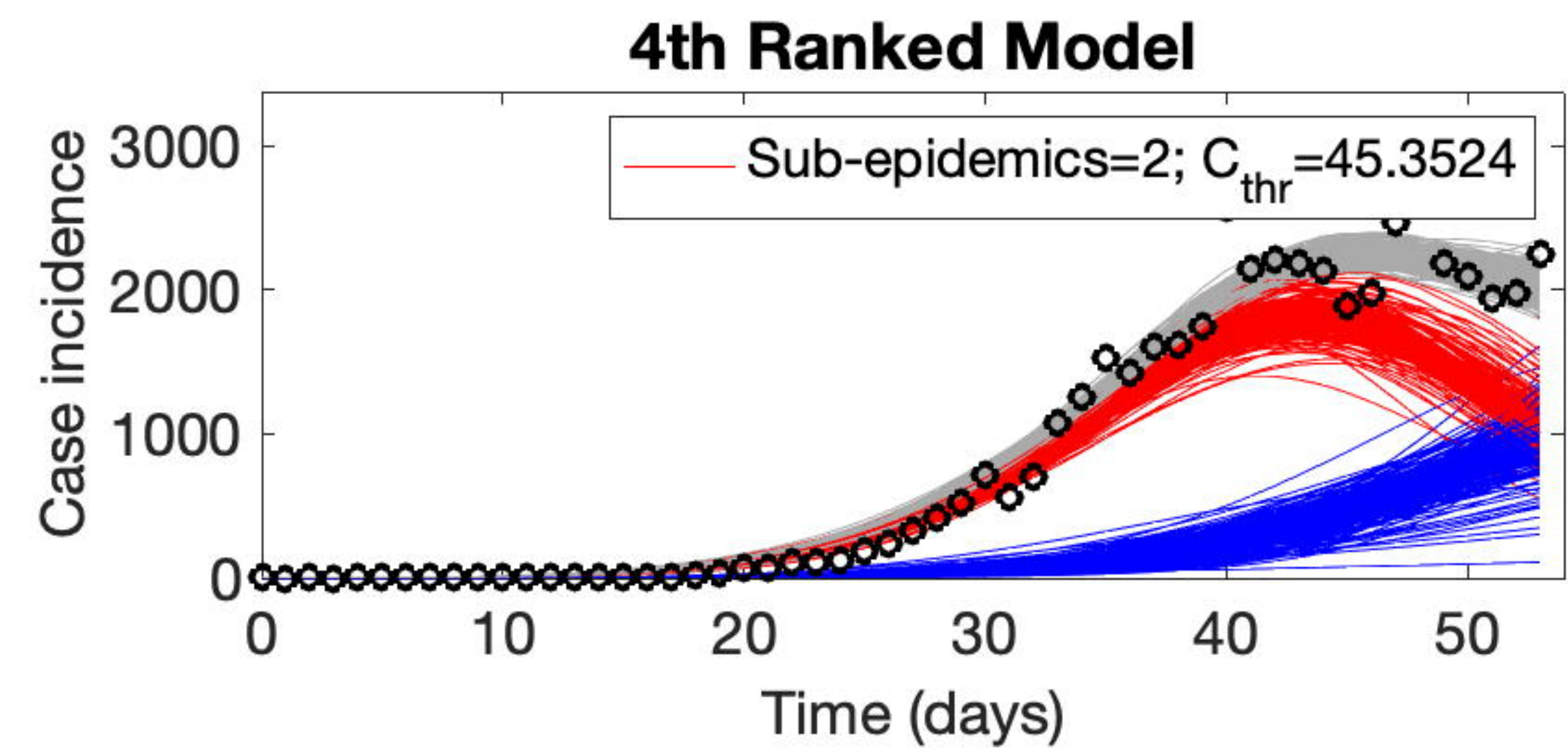
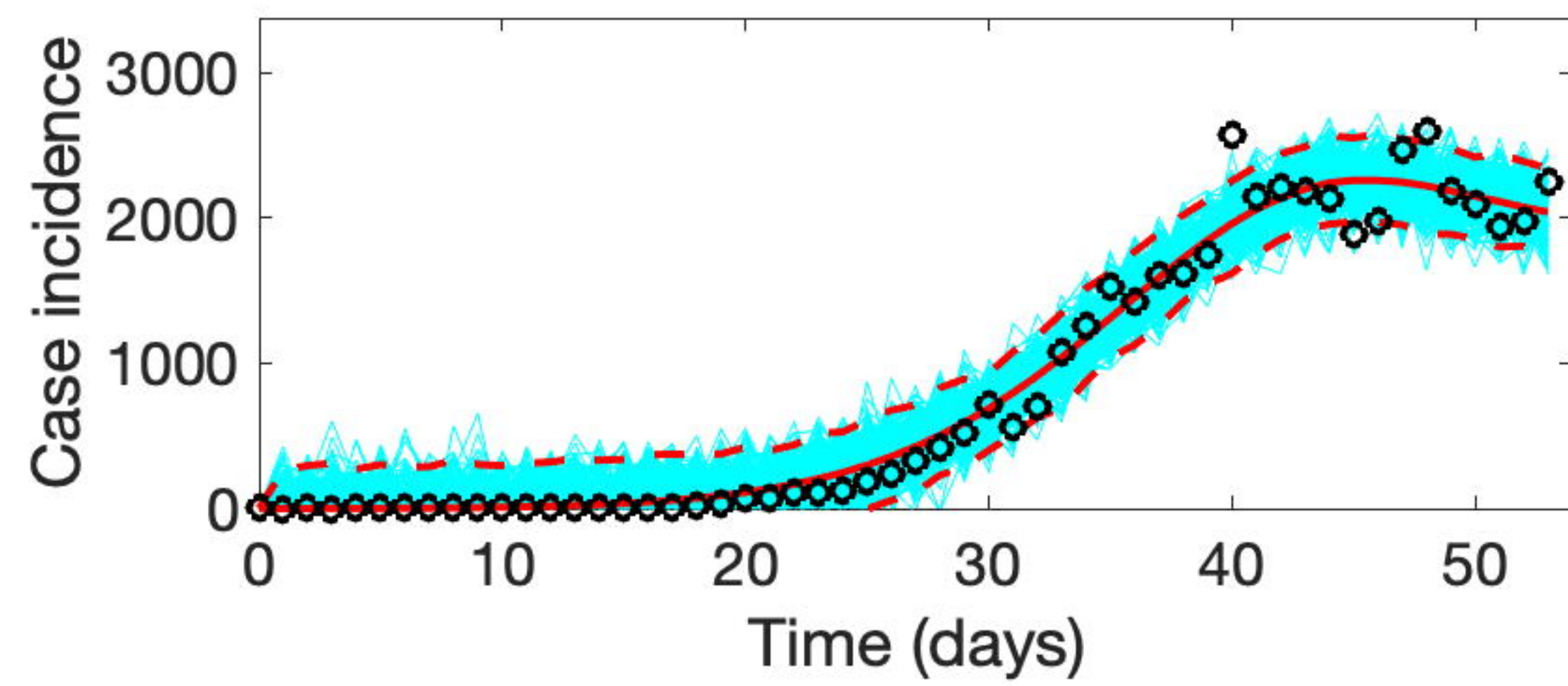
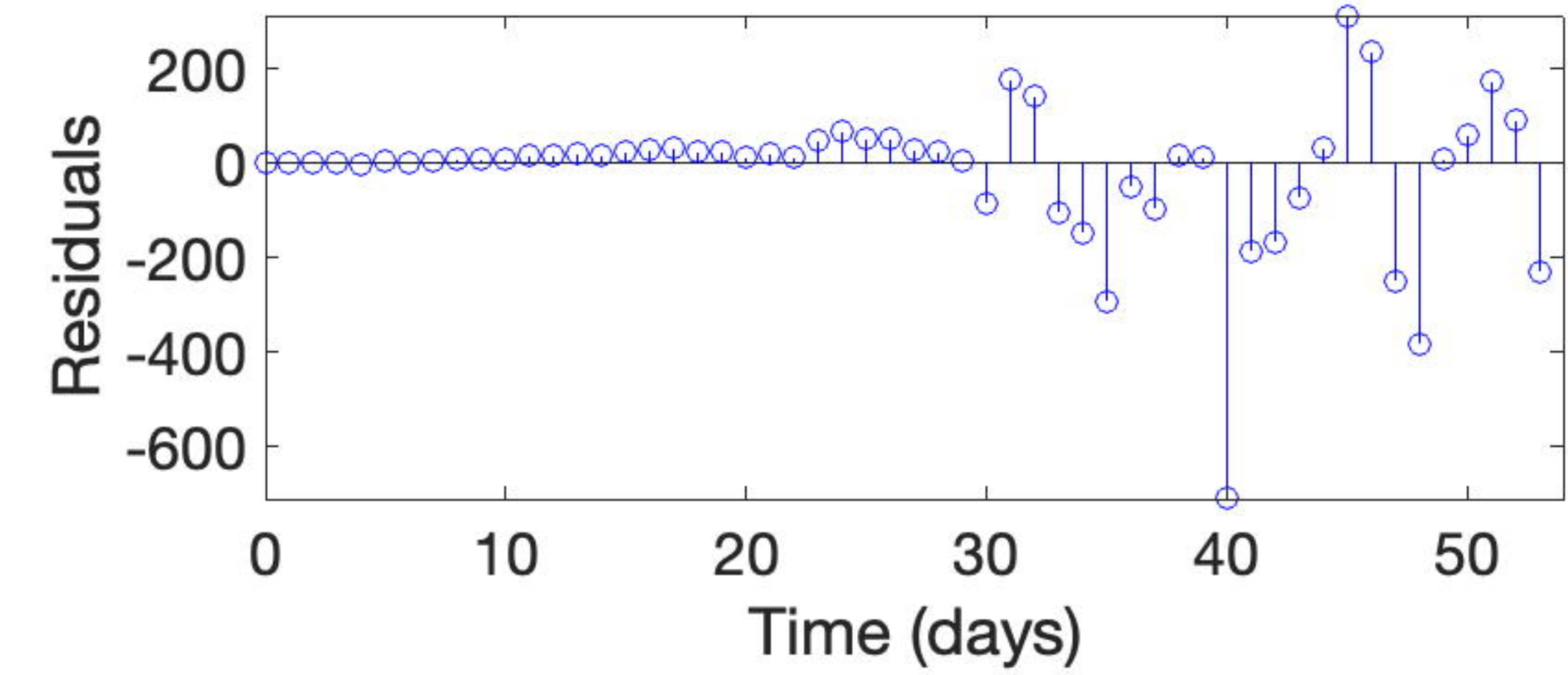
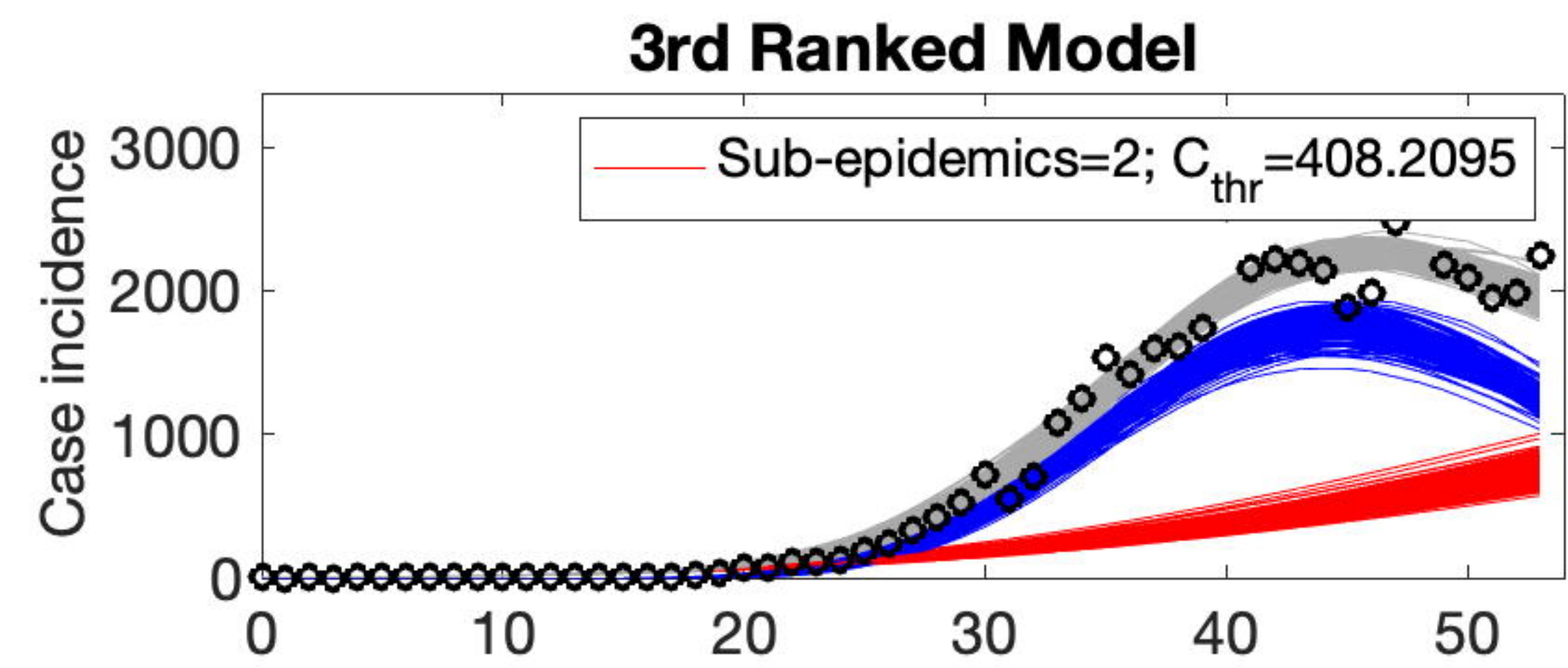
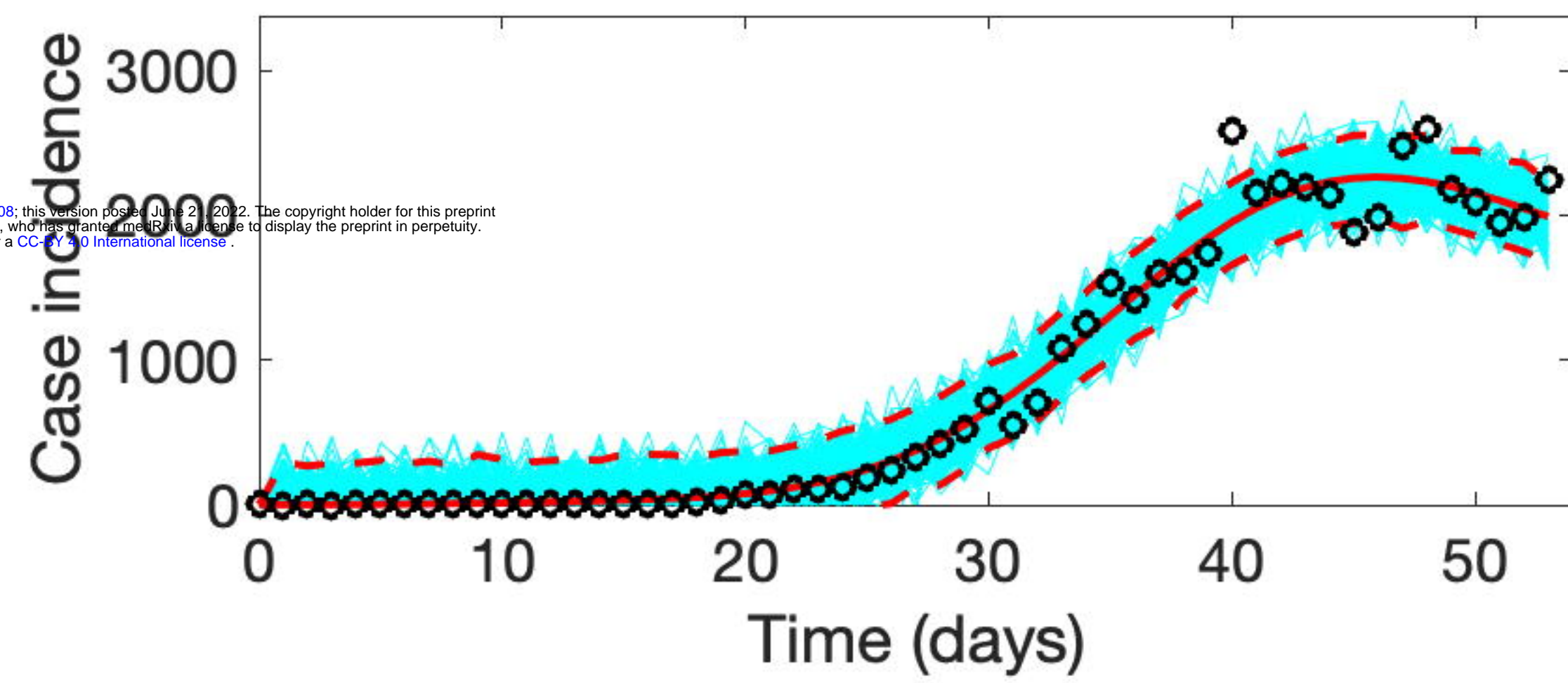
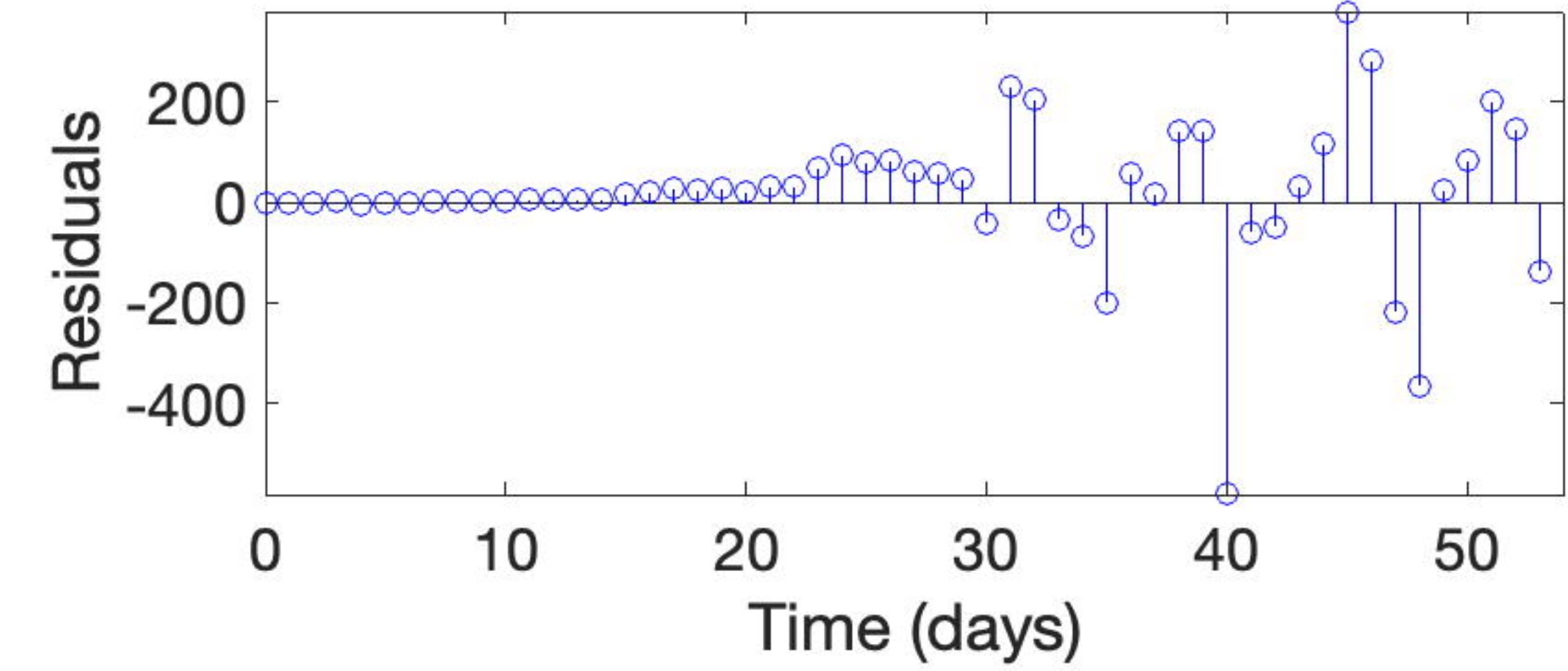
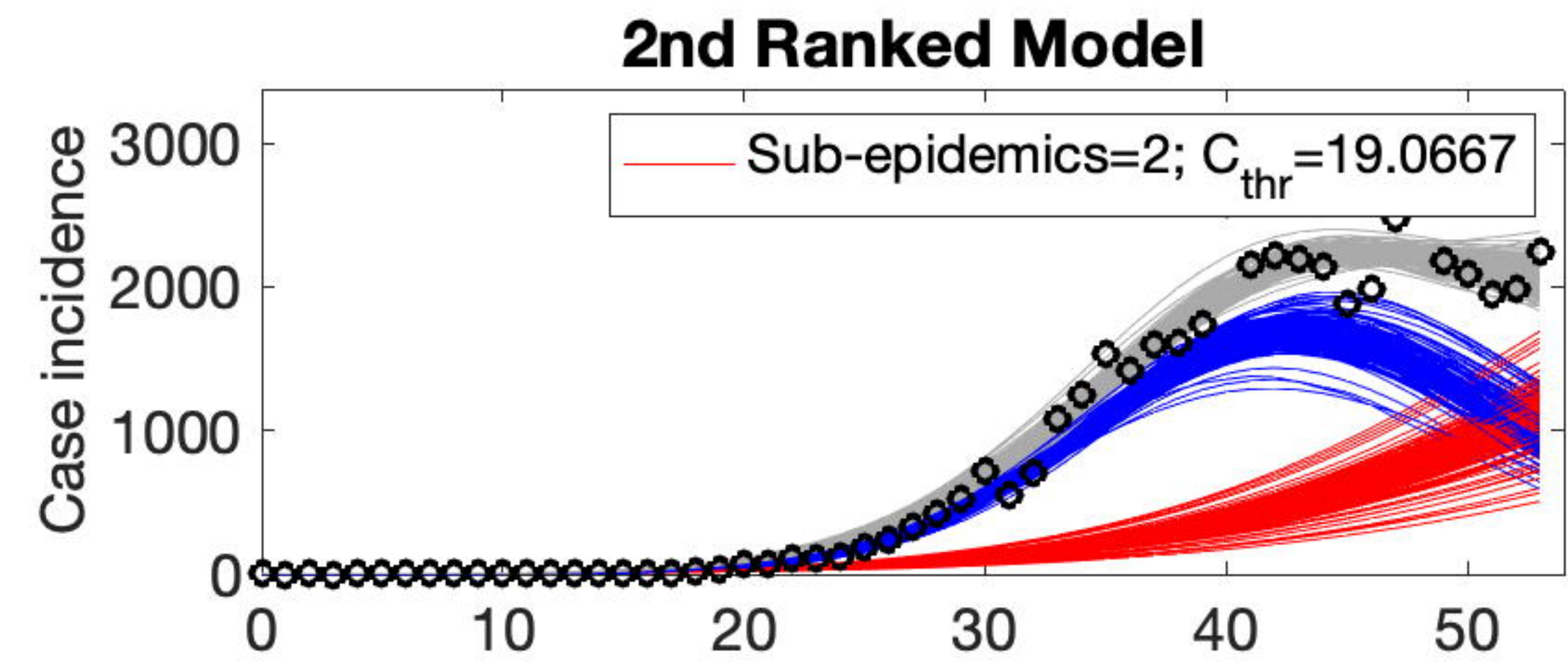
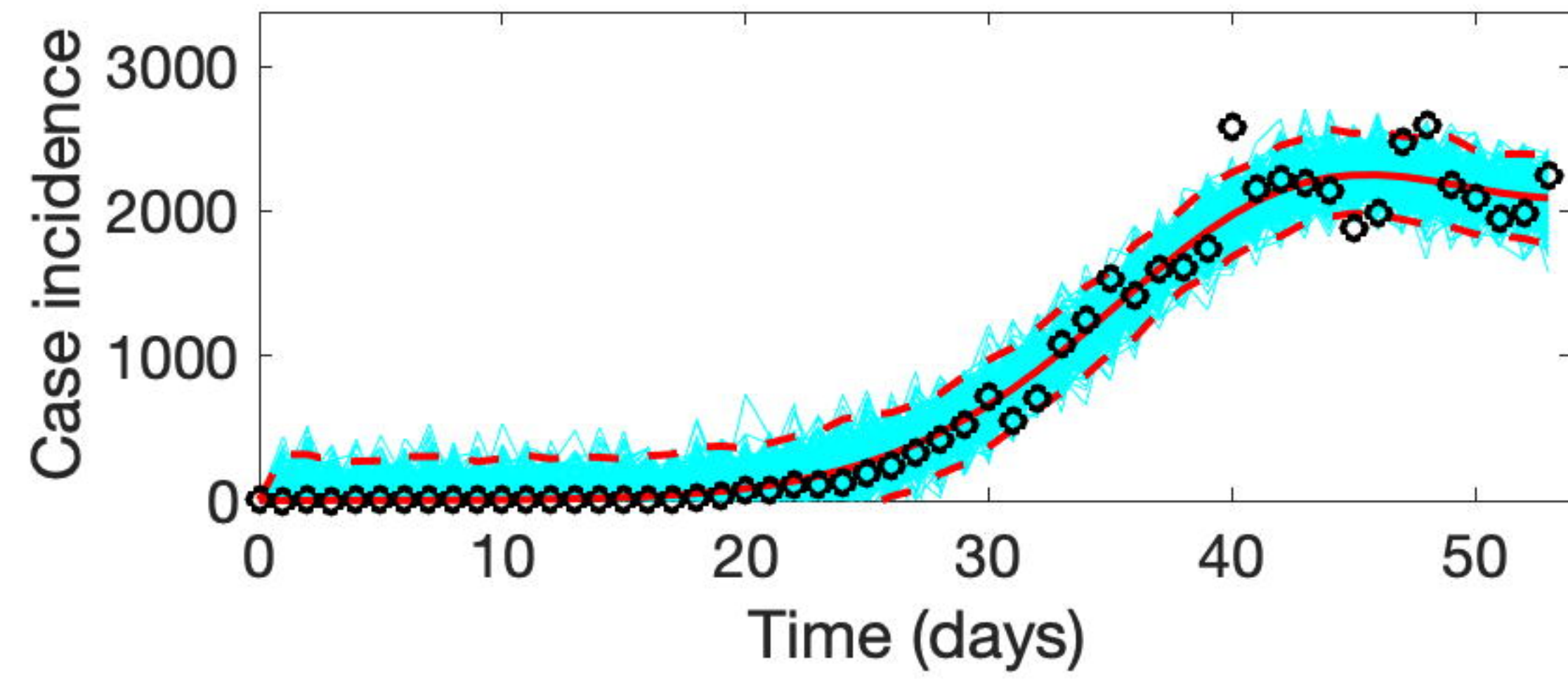
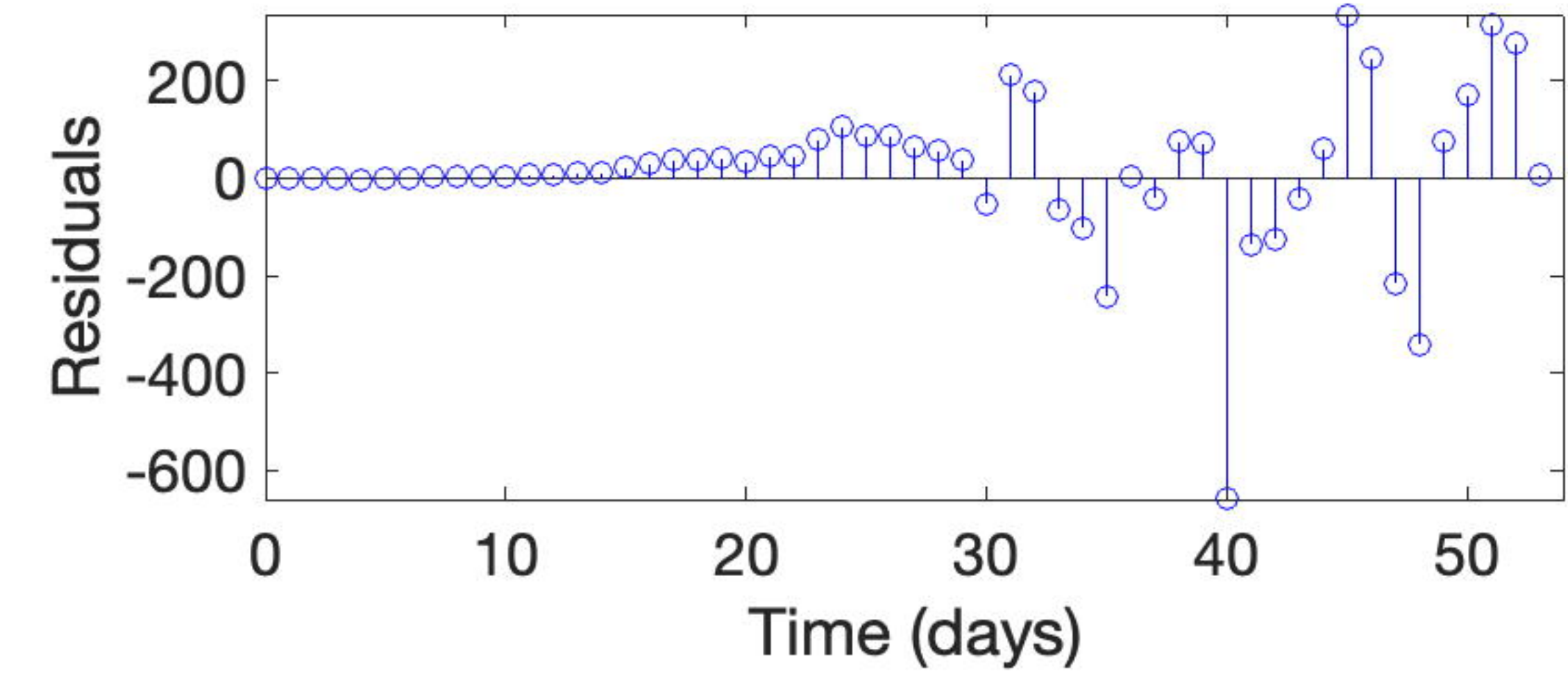
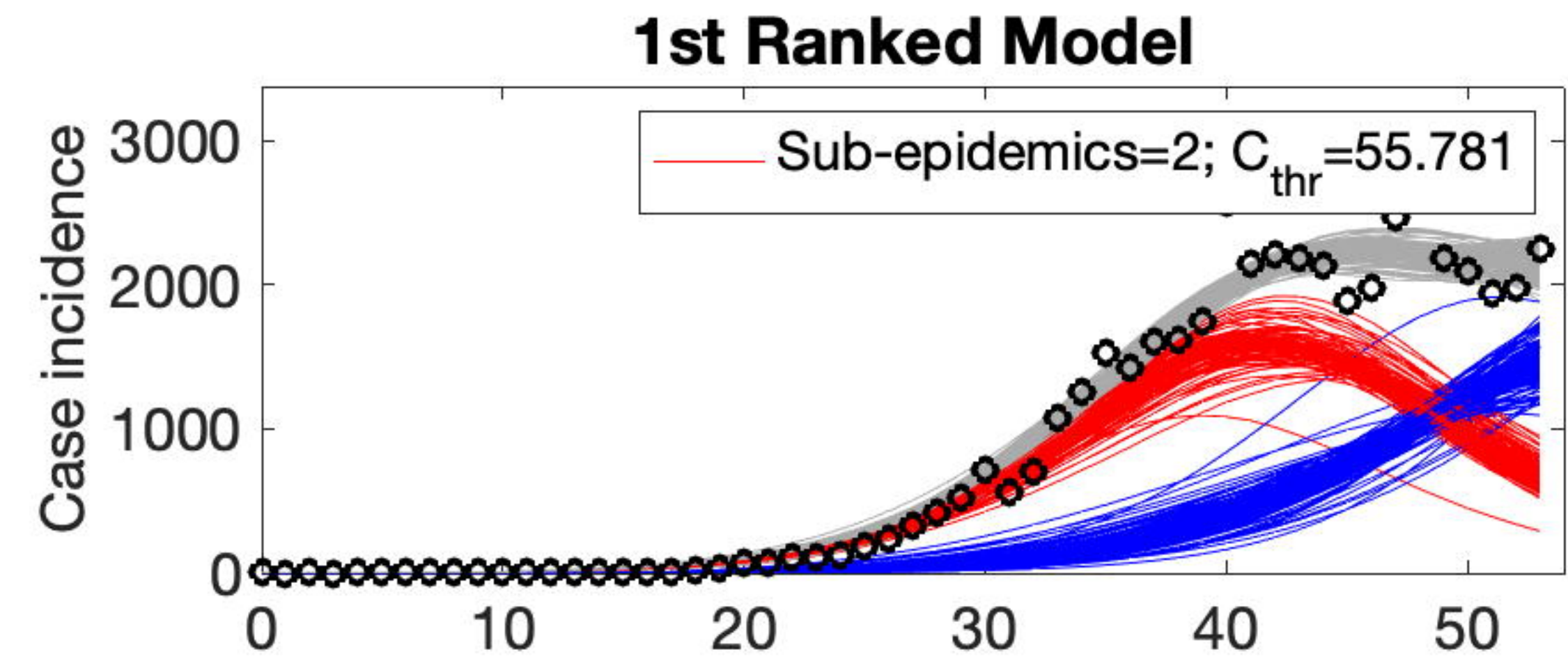
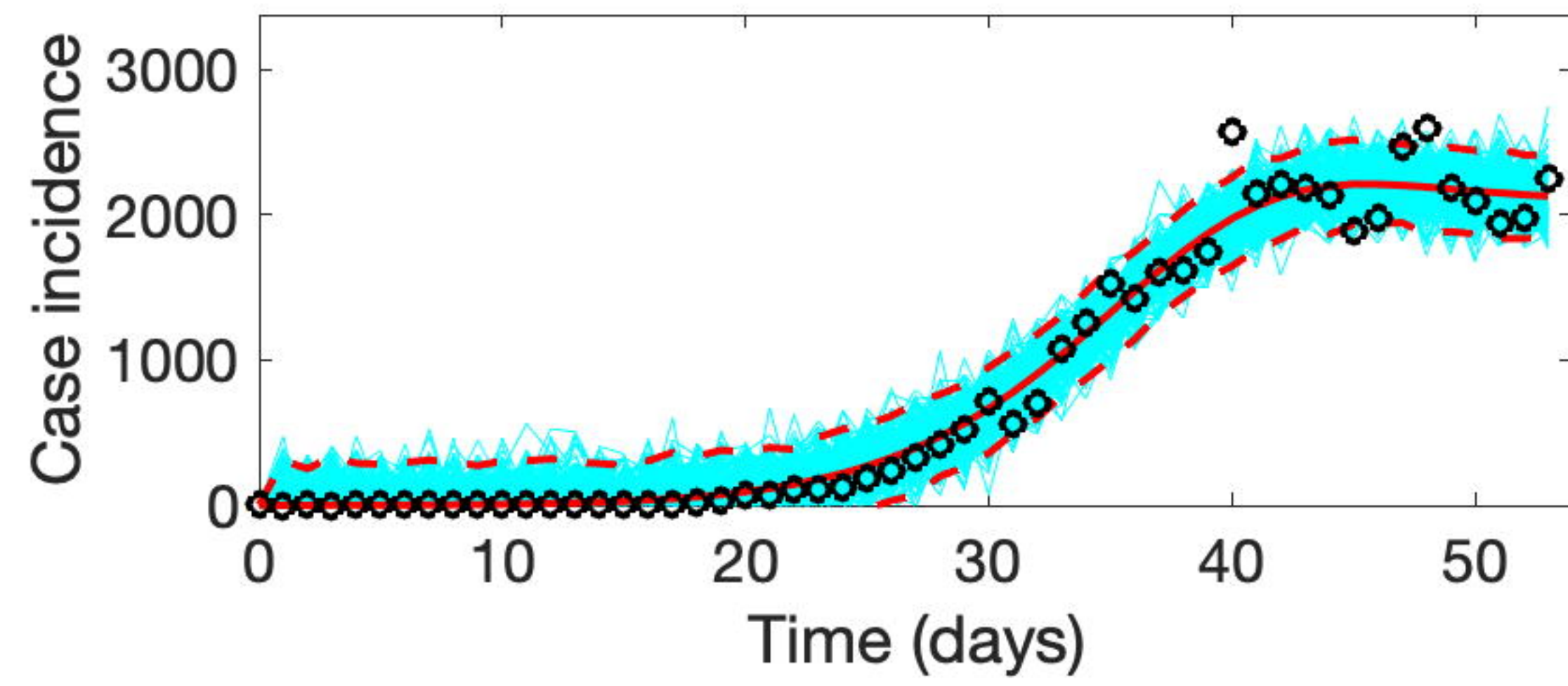
- 719 39. Hurvich CM, Tsai C-L. Regression and time series model selection in small samples.
720 *Biometrika*. 1989;76:297-307.
- 721 40. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical
722 information-theoretic approach. 2 ed: Springer-Verlag, New York, NY; 2002. p. 488.
- 723 41. Prapanna M, Shit L, Goswami. S. Study of effectiveness of time series modeling (ARIMA)
724 in forecasting stock prices. *International Journal of Computer Science, Engineering and*
725 *Applications*. 2014;4.2(13).
- 726 42. Adebisi AA, Adewumii A, Ayo C. Stock price prediction using the ARIMA model. *UKSim-*
727 *AMSS 16th International Conference on Computer Modelling and Simulation: IEEE*; 2014.
- 728 43. Almasarweh M, Alwadi S. ARIMA model in predicting banking stock market data.
729 *Modern Applied Science* 2018;12(11):309.
- 730 44. Tektaş M. Weather Forecasting Using ANFIS and ARIMA MODELS. *Environmental*
731 *Research, Engineering and Management*. 2010;51(1):5-10.
- 732 45. Shamsnia SA, Shahidi N, Liaghat A, Sarraf A, Vahdat SF. Modeling of weather parameters
733 using stochastic methods (ARIMA model)(case study: Abadeh Region, Iran). *International*
734 *conference on environment and industrial innovation* 2011.
- 735 46. Dimri T, Ahmad S, Sharif M. Time series analysis of climate variables using seasonal
736 ARIMA approach. *Journal of Earth System Science*. 2020;129(1):149. doi: 10.1007/s12040-020-
737 01408-x.
- 738 47. Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for
739 R. *Journal of Statistical Software*. 2008;27(3):1 - 22. doi: 10.18637/jss.v027.i03.
- 740 48. Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. Testing the null hypothesis of stationarity
741 against the alternative of a unit root: How sure are we that economic time series have a unit
742 root? *Journal of Econometrics*. 1992;54(1):159-78. doi: [https://doi.org/10.1016/0304-](https://doi.org/10.1016/0304-4076(92)90104-Y)
743 [4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y).
- 744 49. Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval
745 format. *PLoS Comput Biol*. 2021;17(2):e1008618. doi: 10.1371/journal.pcbi.1008618.
- 746 50. Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal*
747 *of the American Statistical Association*. 2007;102(477):359-78. doi:
748 10.1198/016214506000001437.
- 749 51. Kuhn M, Johnson K. *Applied predictive modeling*: New York: Springer; 2013.
- 750 52. M4Competition. *Competitor's Guide: Prizes and Rules*. 2018. Available from:
751 <https://www.m4.unic.ac.cy/wp-content/uploads/2018/03/M4-Competitors-Guide.pdf>.
- 752 53. Funk S, Camacho A, Kucharski AJ, Lowe R, Eggo RM, Edmunds WJ. Assessing the
753 performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region
754 of Sierra Leone, 2014-15. *PLoS Comput Biol*. 2019;15(2):e1006785. Epub 2019/02/12. doi:
755 10.1371/journal.pcbi.1006785. PubMed PMID: 30742608.
- 756 54. Hwang E. Prediction intervals of the COVID-19 cases by HAR models with growth rates
757 and vaccination rates in top eight affected countries: Bootstrap improvement. *Chaos Solitons*
758 *Fractals*. 2022;155:111789-. Epub 2022/01/03. doi: 10.1016/j.chaos.2021.111789. PubMed
759 PMID: 35002103.
- 760 55. Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval
761 format. *PLoS Comput Biol*. 2021;17(2):e1008618. doi: 10.1371/journal.pcbi.1008618.

- 762 56. Rguibi MA, Moussa N, Madani A, Aaroud A, Zine-Dine K. Forecasting Covid-19
763 Transmission with ARIMA and LSTM Techniques in Morocco. SN Comput Sci. 2022;3(2):133-
764 Epub 2022/01/14. doi: 10.1007/s42979-022-01019-x. PubMed PMID: 35043096.
- 765 57. Kandula S, Shaman J. Near-term forecasts of influenza-like illness: An evaluation of
766 autoregressive time series approaches. *Epidemics*. 2019;27:41-51. doi:
767 <https://doi.org/10.1016/j.epidem.2019.01.002>.
- 768 58. Reich NG, Brooks LC, Fox SJ, Kandula S, McGowan CJ, Moore E, et al. A collaborative
769 multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proc*
770 *Natl Acad Sci U S A*. 2019;116(8):3146-54. Epub 2019/01/17. doi: 10.1073/pnas.1812594116.
771 PubMed PMID: 30647115; PubMed Central PMCID: PMC6386665.
- 772 59. Roy S, Bhunia GS, Shit PK. Spatial prediction of COVID-19 epidemic using ARIMA
773 techniques in India. *Model Earth Syst Environ*. 2021;7(2):1385-91. Epub 2020/08/25. doi:
774 10.1007/s40808-020-00890-y. PubMed PMID: 32838022; PubMed Central PMCID:
775 PMC6363688.
- 776 60. Jacques Demongeot KO, Mustapha Rachdi, Lahoucine Hobbad, Mohamed Alahiane,
777 Siham Iggui, Jean Gaudart, Idir Ouassou, . he application of ARIMA model to analyze COVID-19
778 incidence pattern in several countries. *J Math Comput Sci*. 2021;12.
- 779 61. Naresh Kumar aSS. COVID-19 pandemic prediction using time series forecasting models.
780 11th International Conference on Computing, Communication and Networking Technologies
781 (ICCCNT): IEEE; 2020.
- 782 62. Taylor KS, Taylor JW. Interval forecasts of weekly incident and cumulative COVID-19
783 mortality in the United States: A comparison of combining methods. *PLOS ONE*.
784 2022;17(3):e0266096. doi: 10.1371/journal.pone.0266096.
- 785

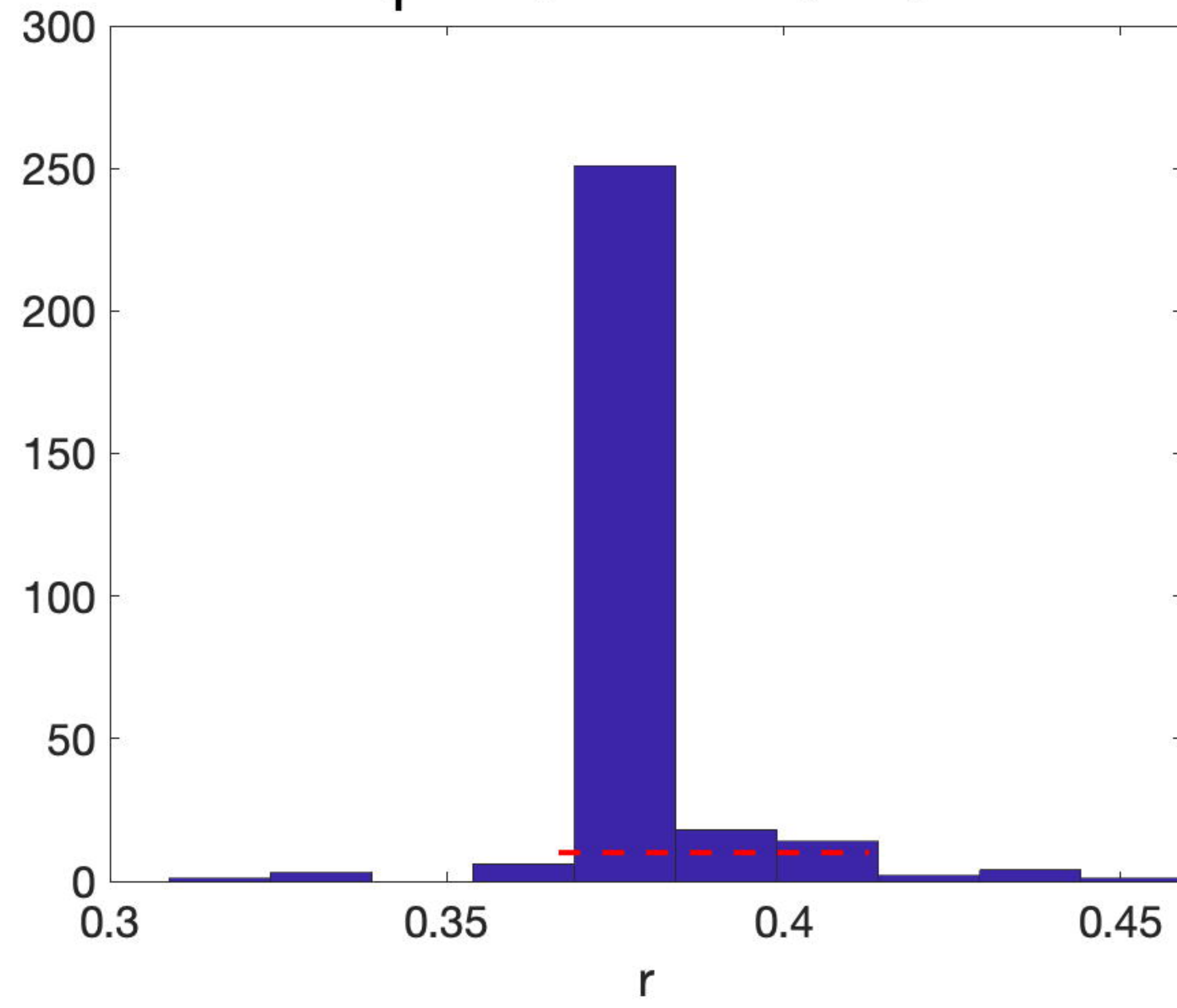




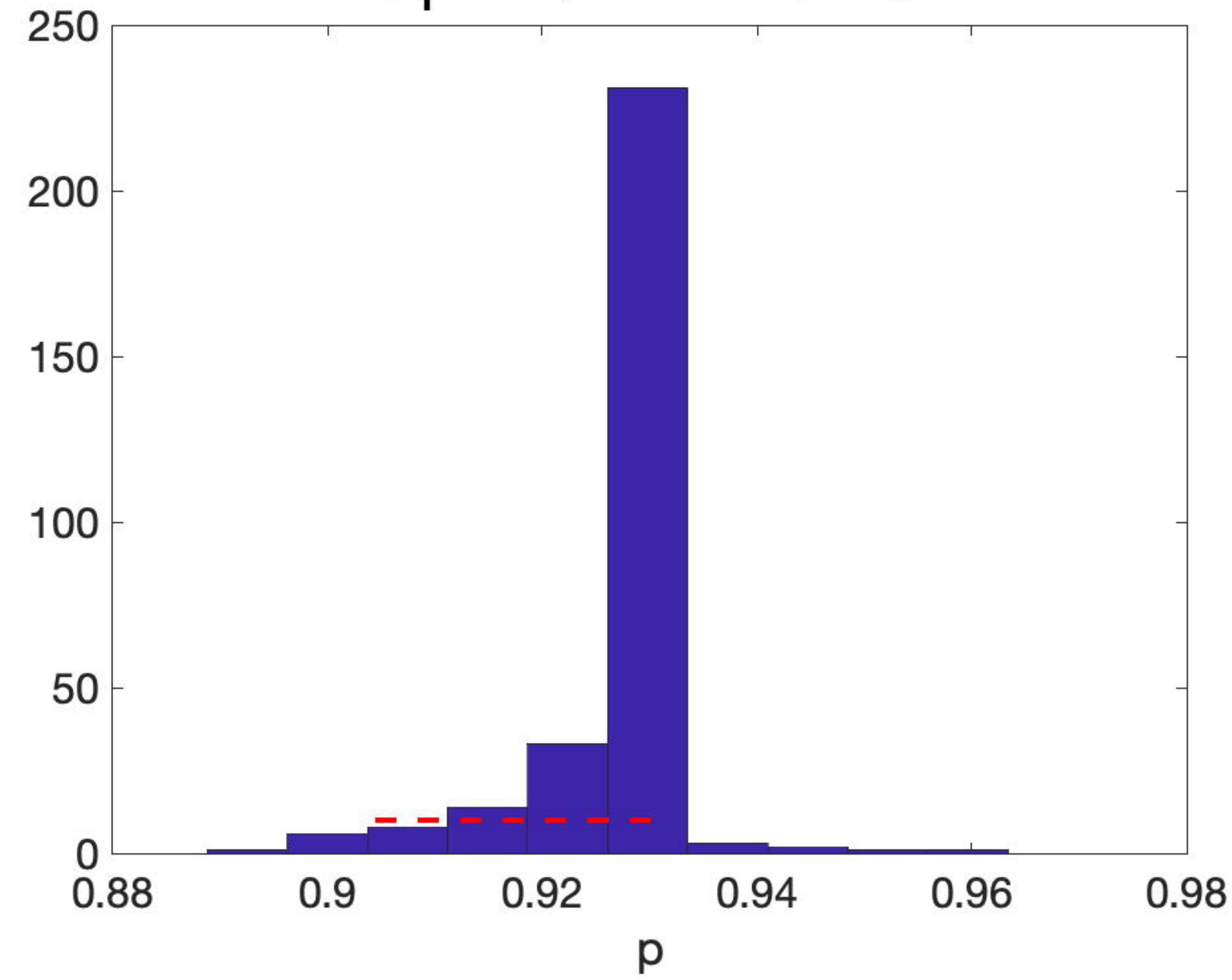
medRxiv preprint doi: <https://doi.org/10.1101/2022.06.13.22271608>; this version posted June 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.



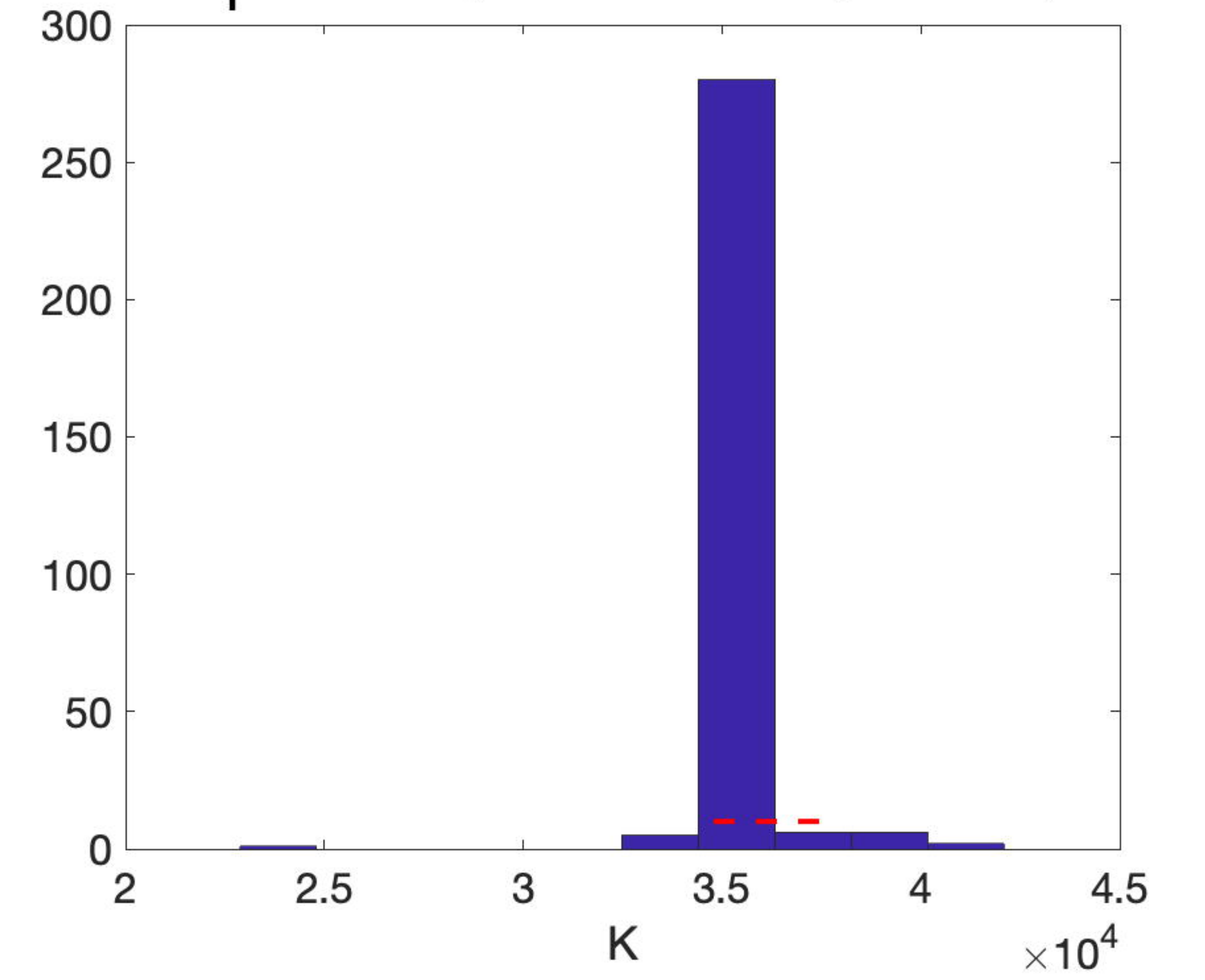
$r_1=0.38(95\%CI:0.37,0.41)$



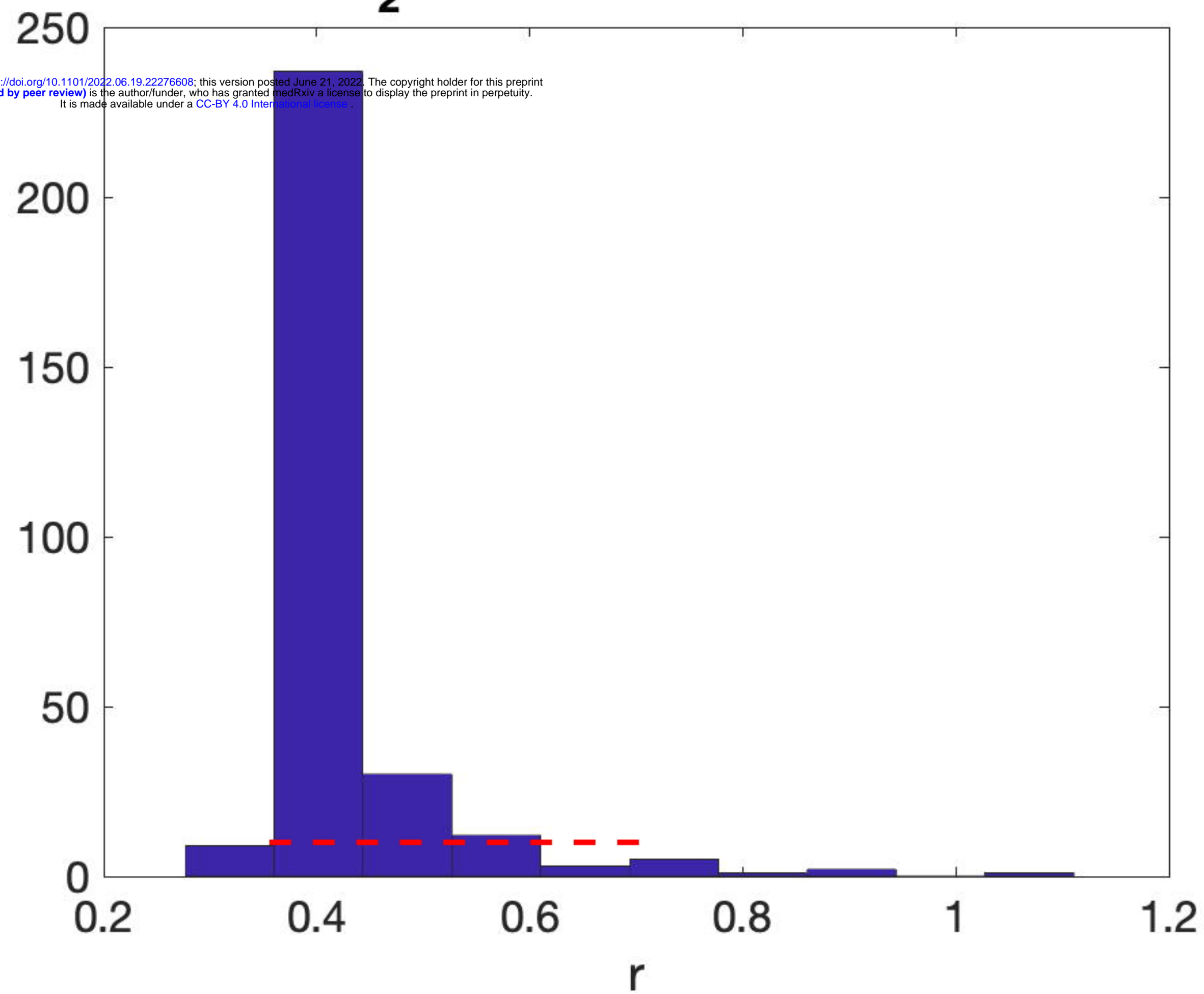
$p_1=0.93(95\%CI:0.9,0.93)$



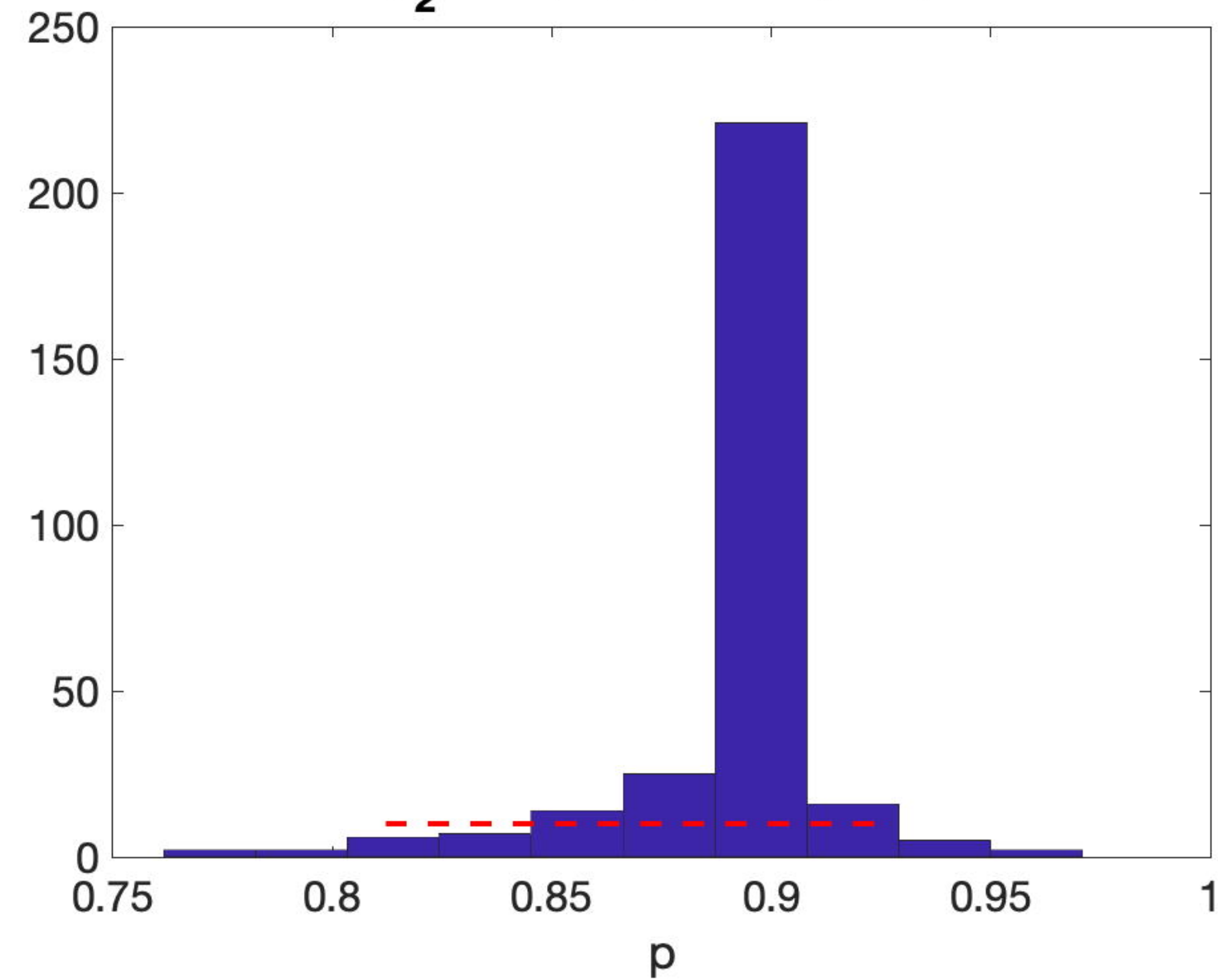
$K_1=3.52e+04(95\%CI:3.48e+04,3.79e+04)$



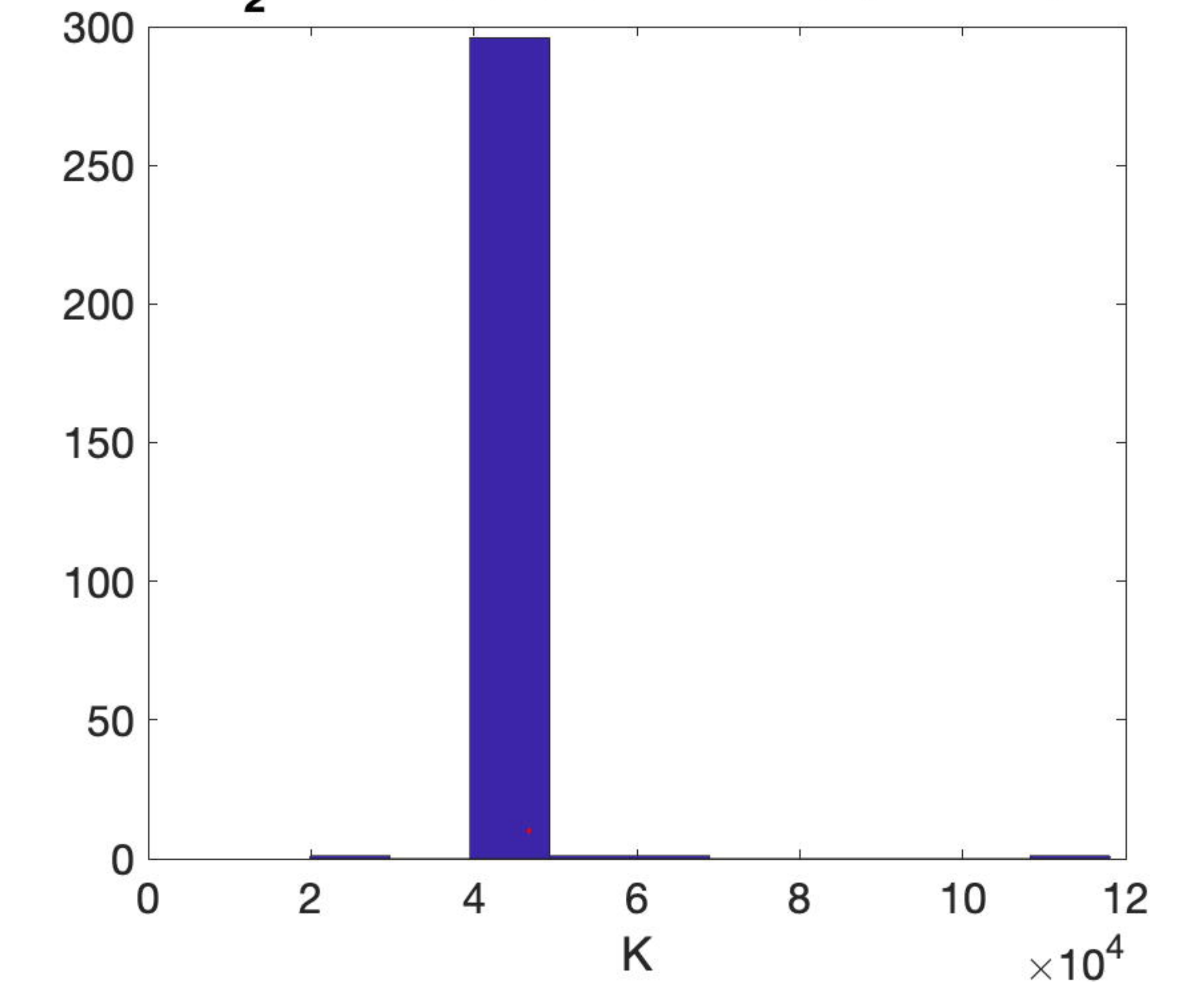
$r_2=0.44(95\%CI:0.36,0.72)$



$p_2=0.89(95\%CI:0.81,0.92)$

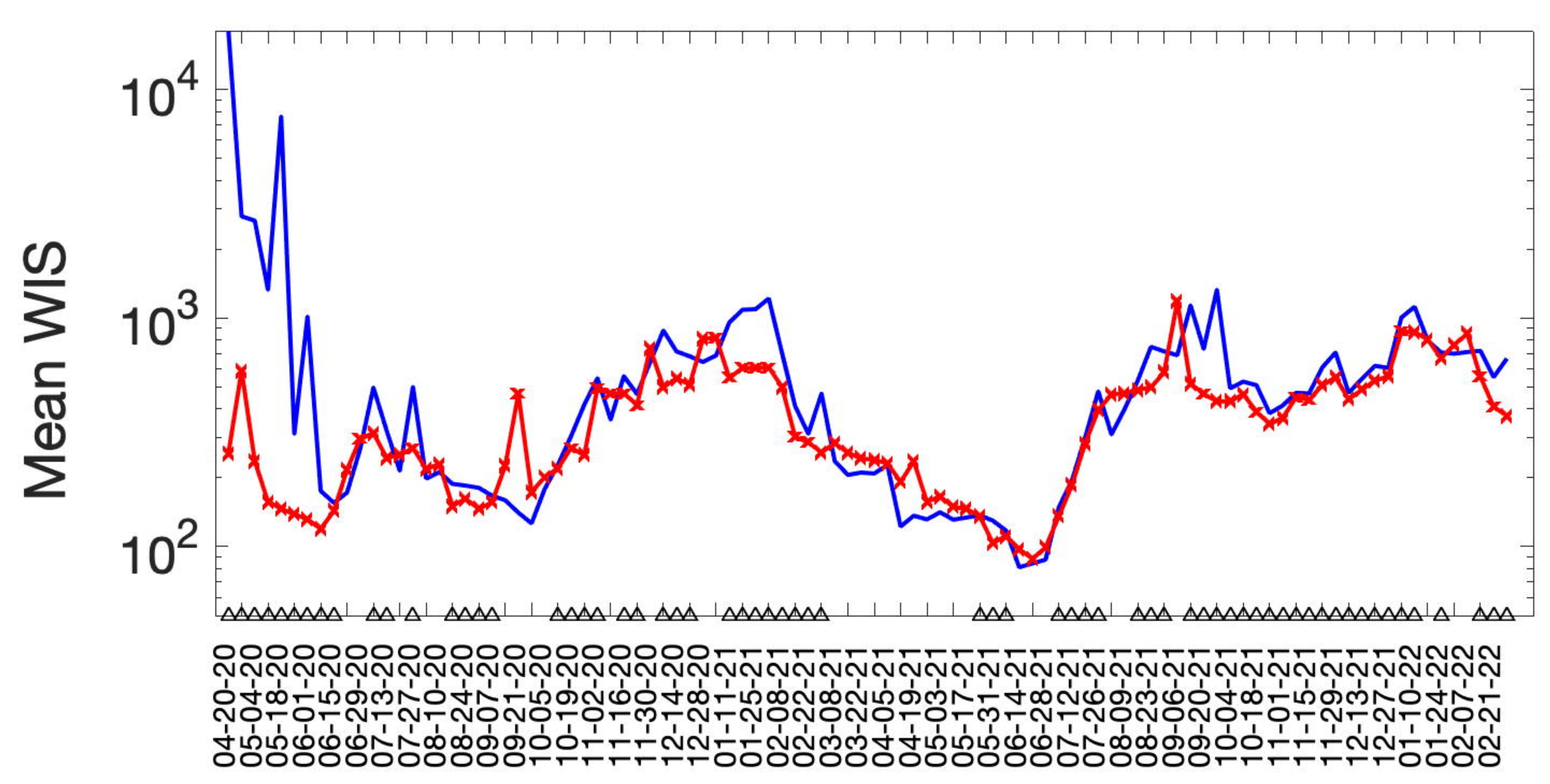
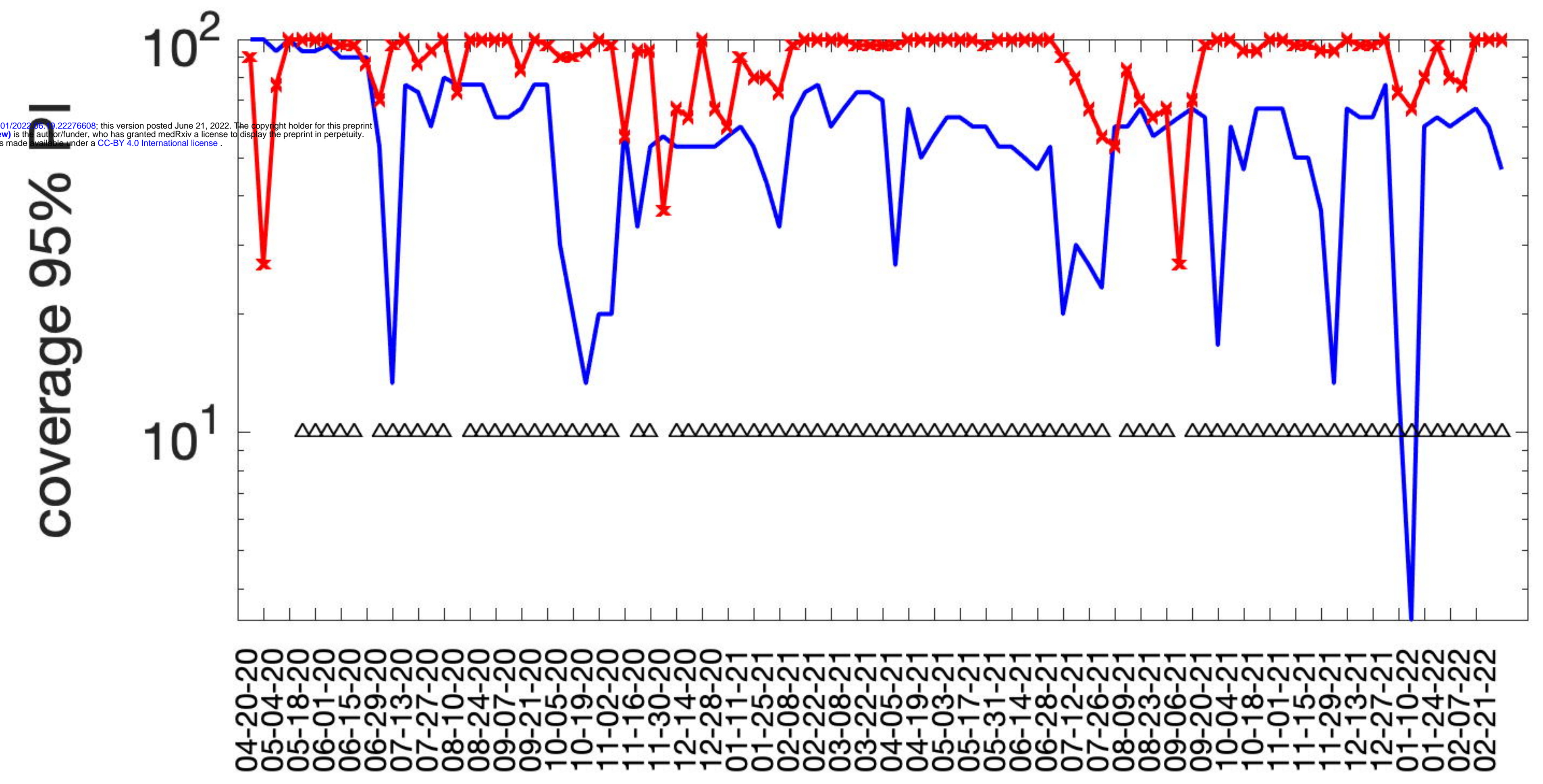
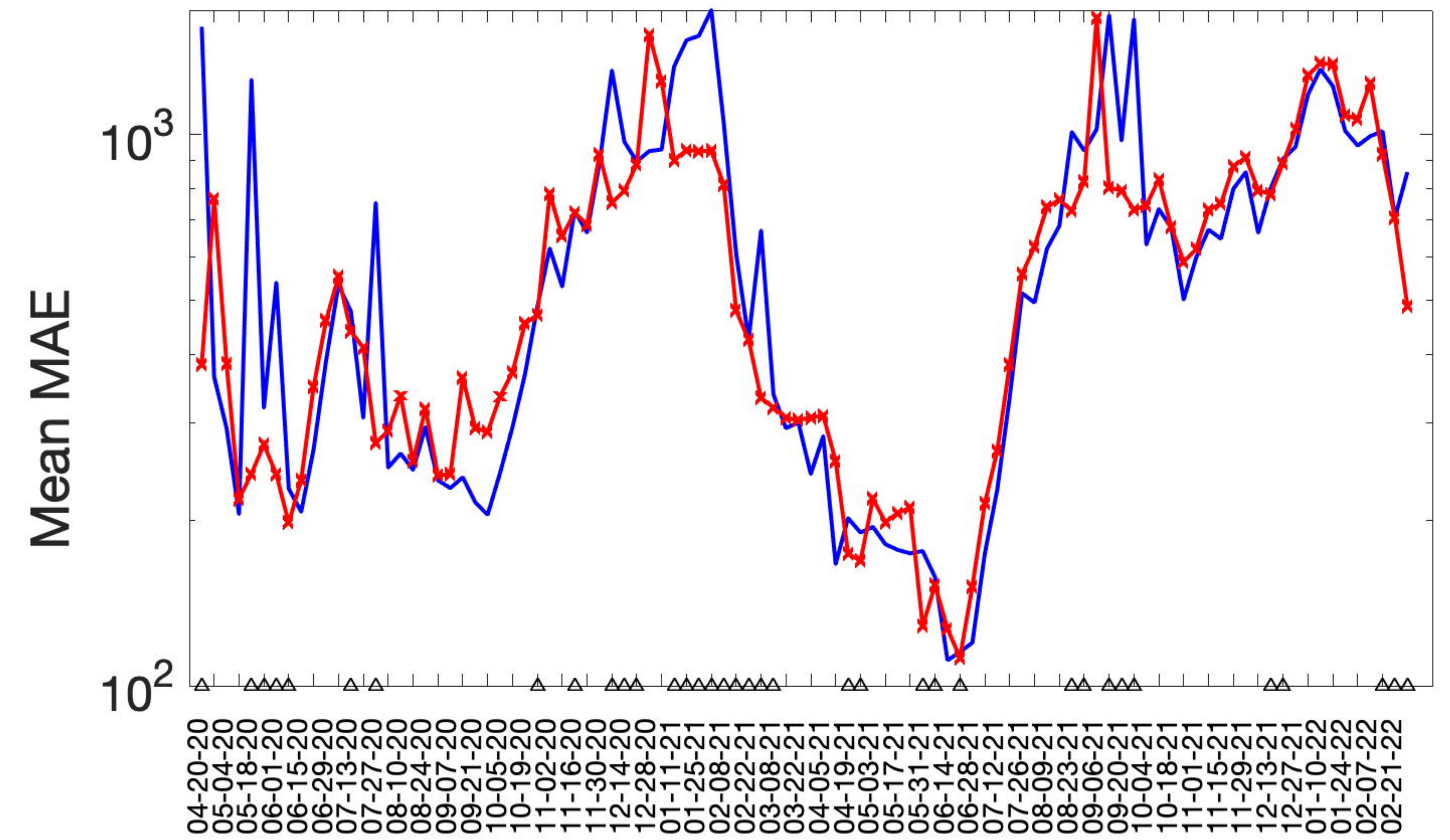
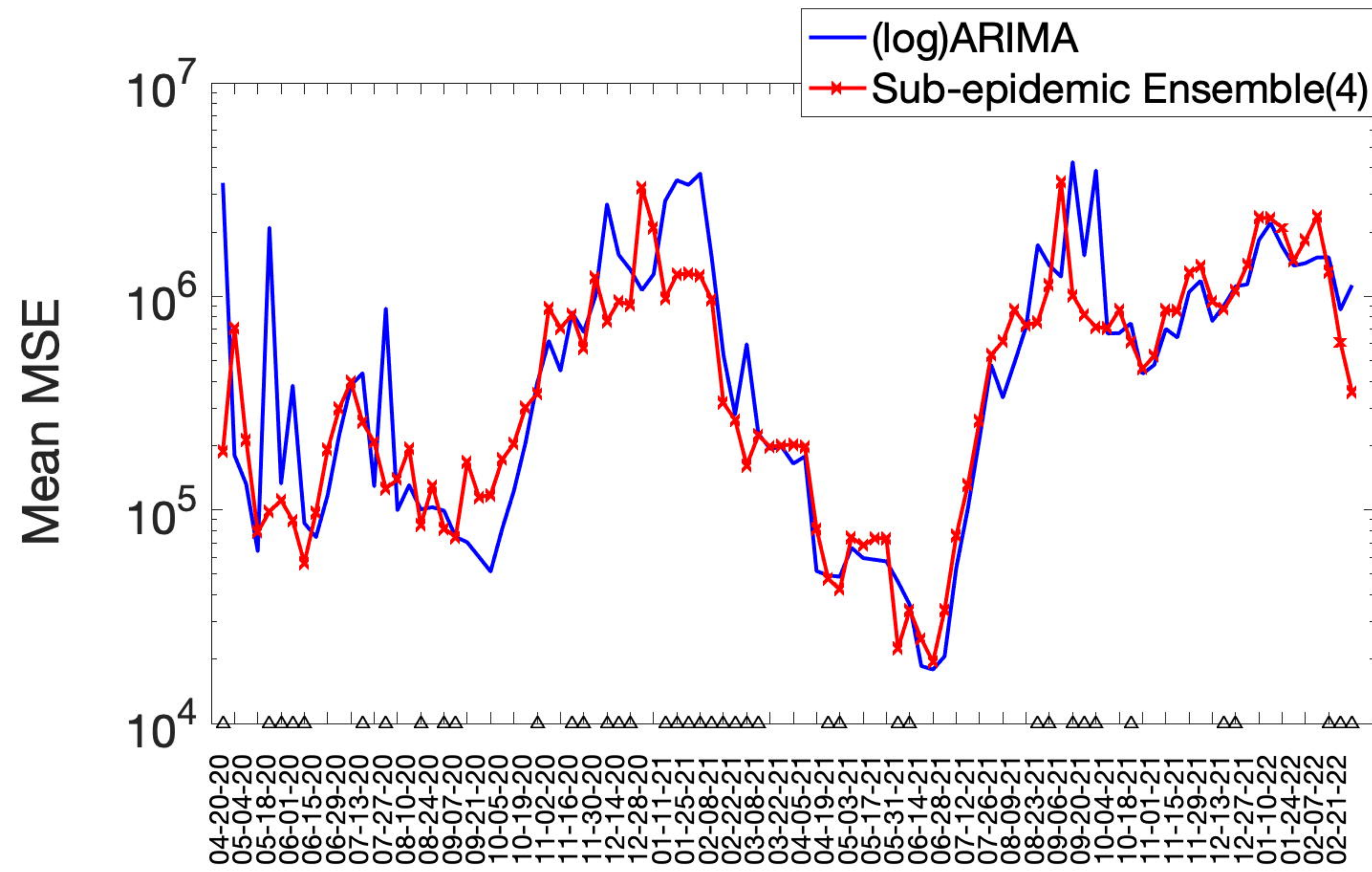


$K_2=4.69e+04(95\%CI:4.66e+04,4.7e+04)$

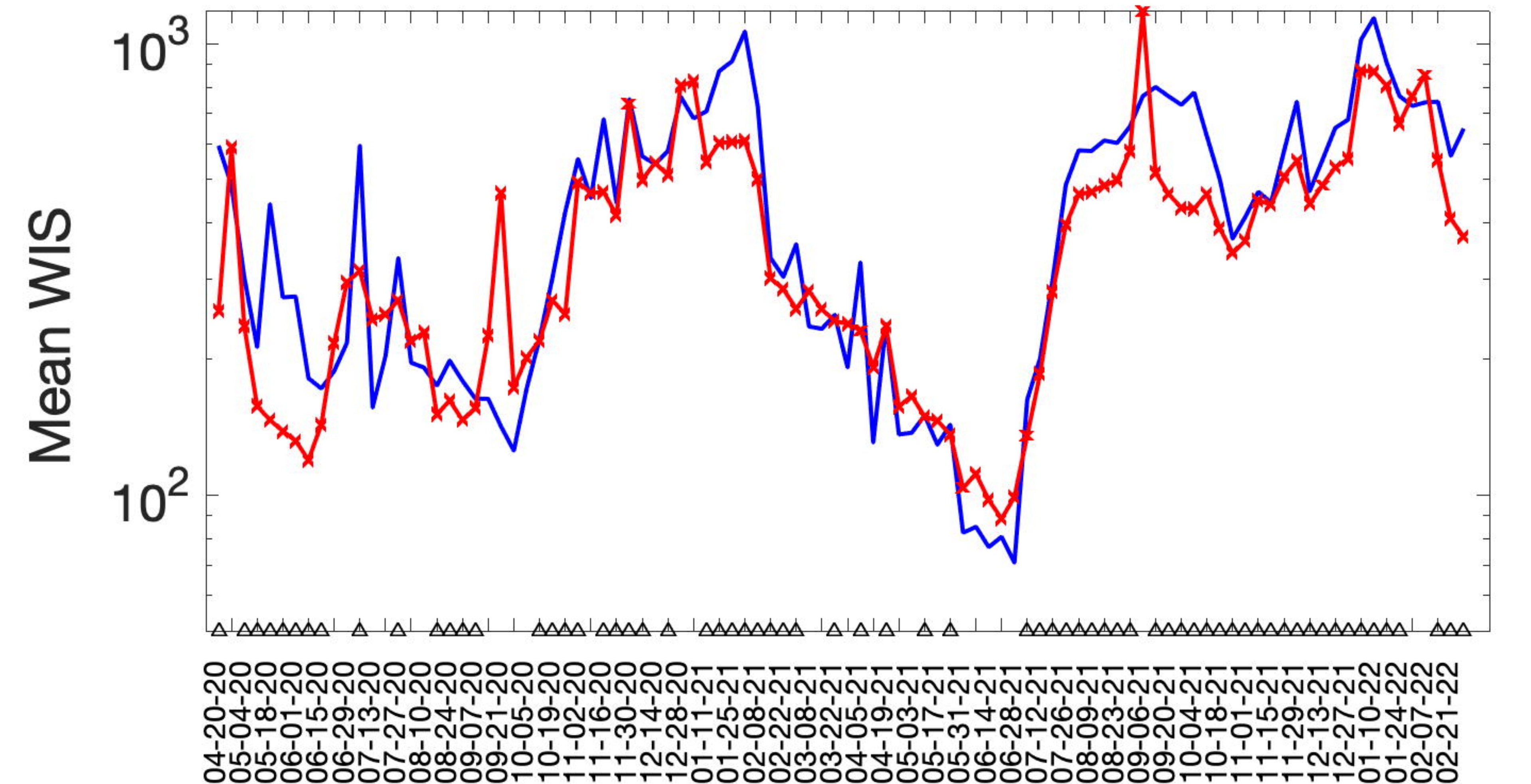
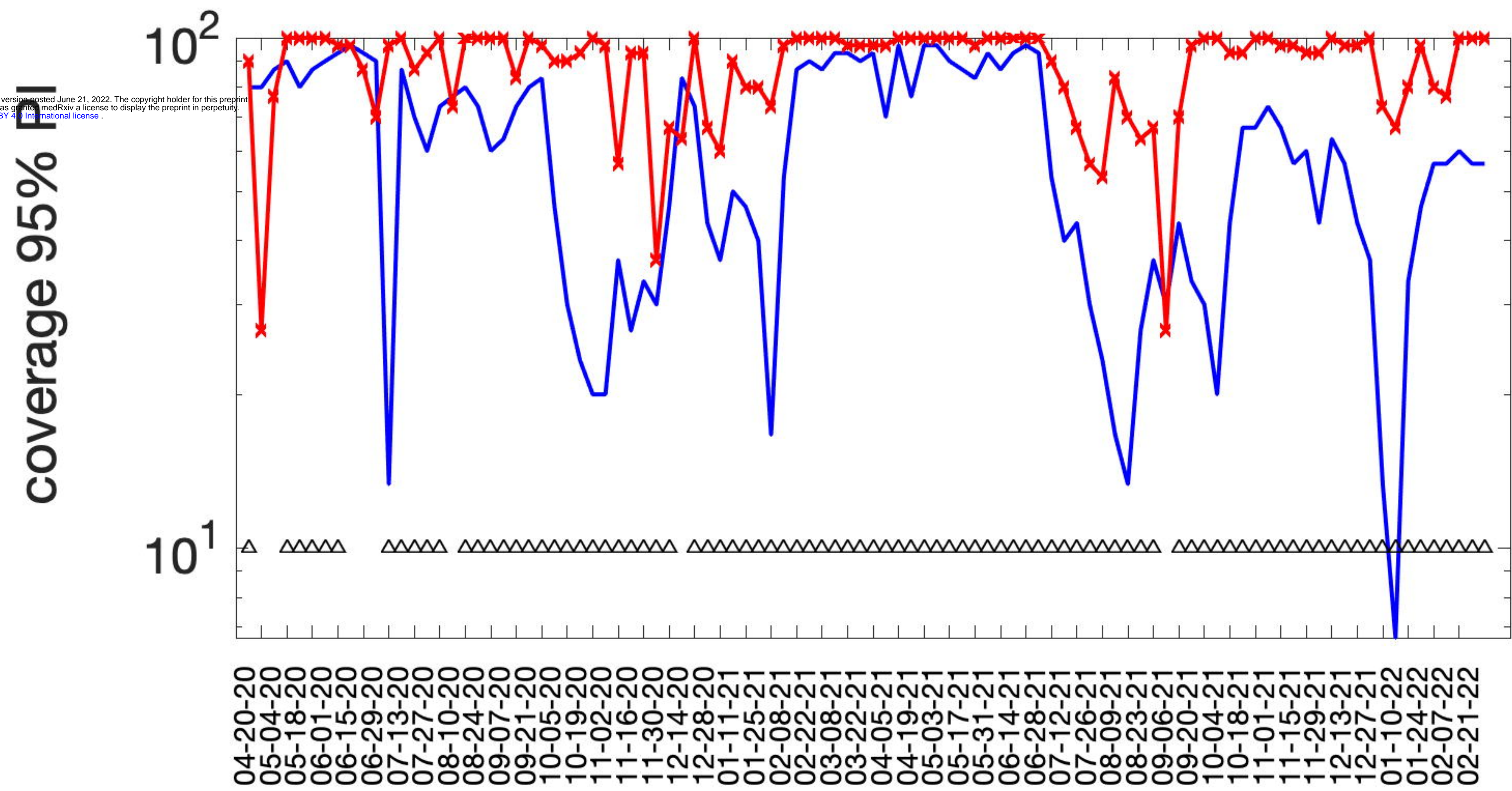
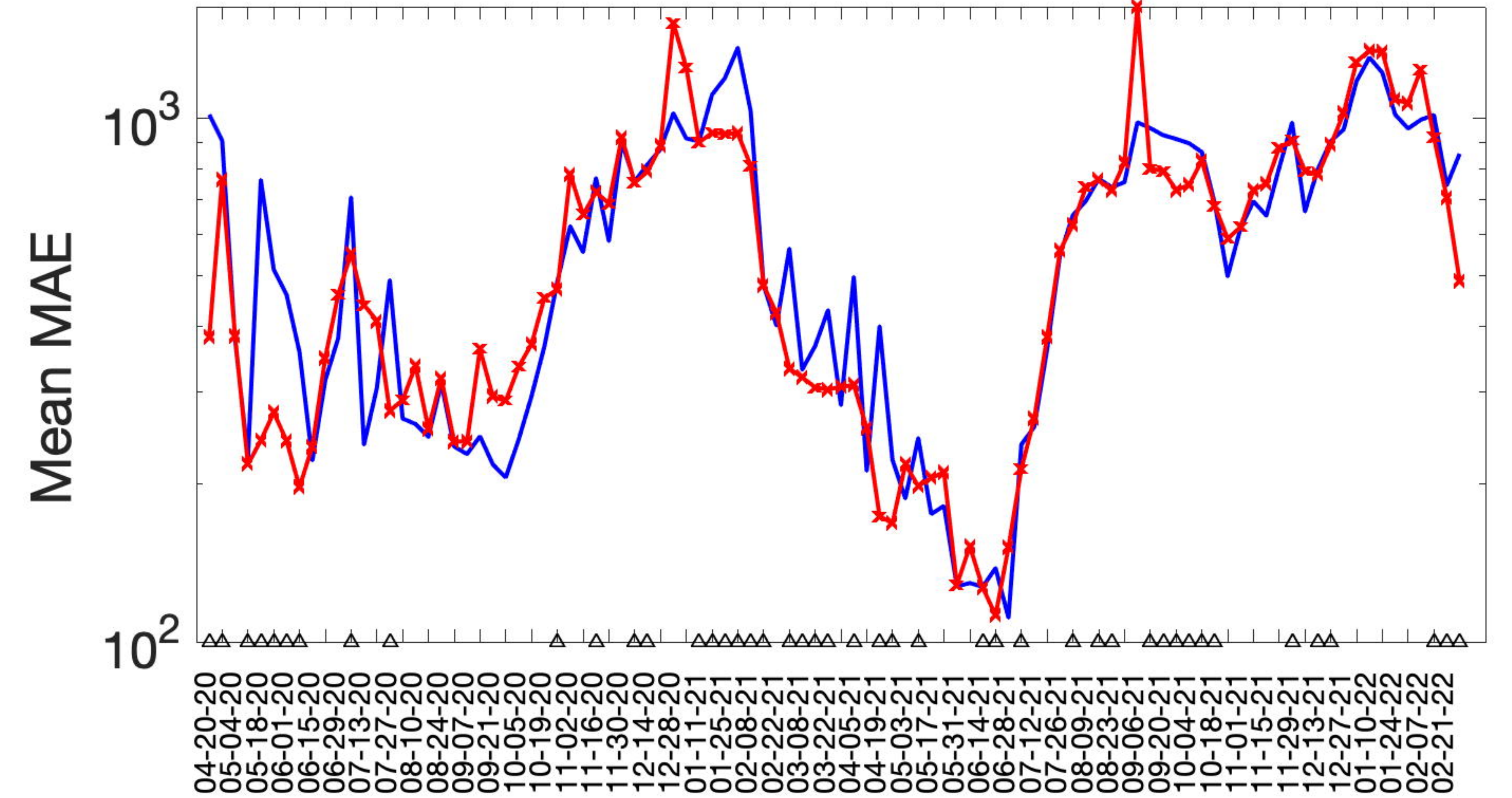
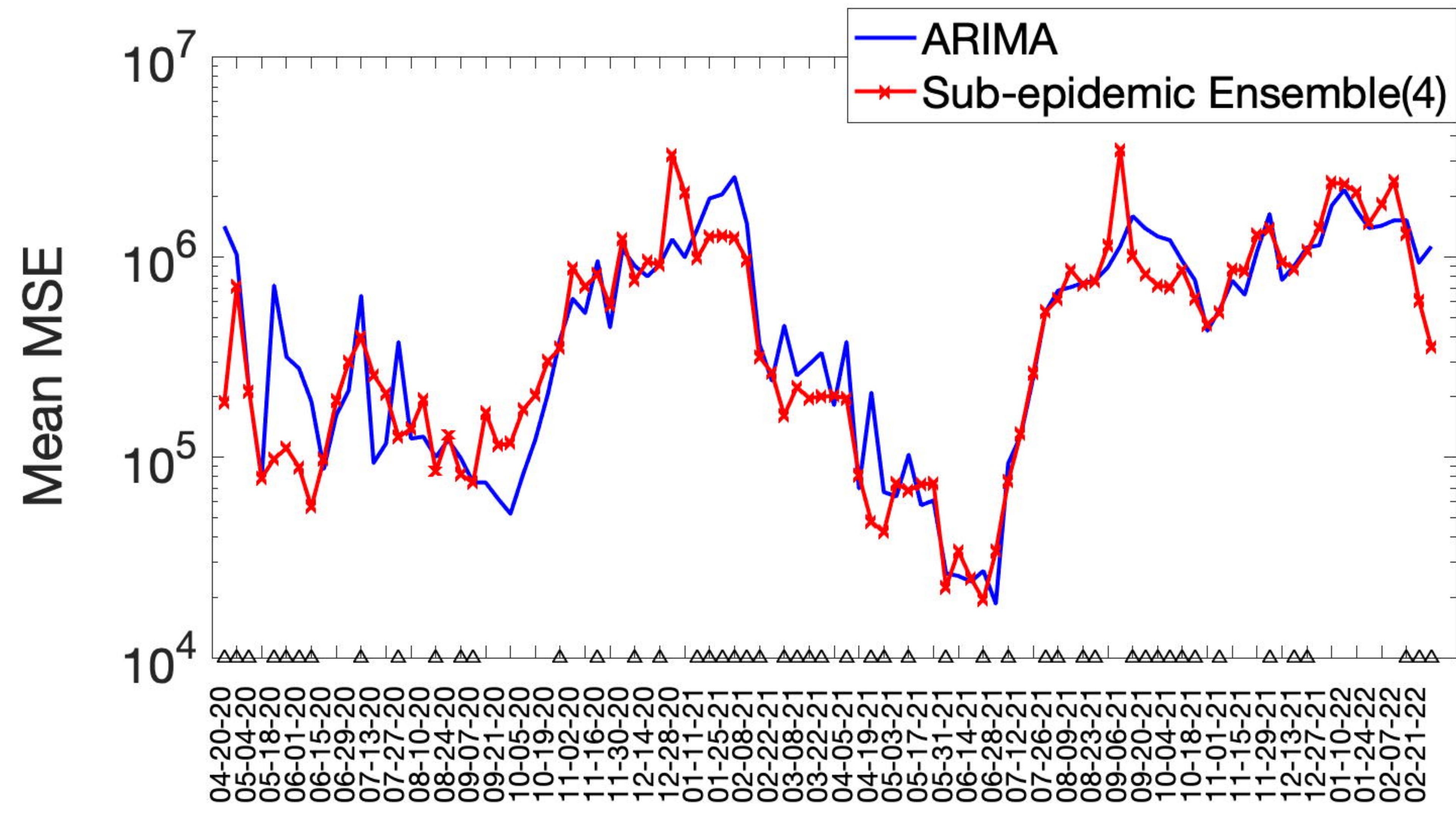


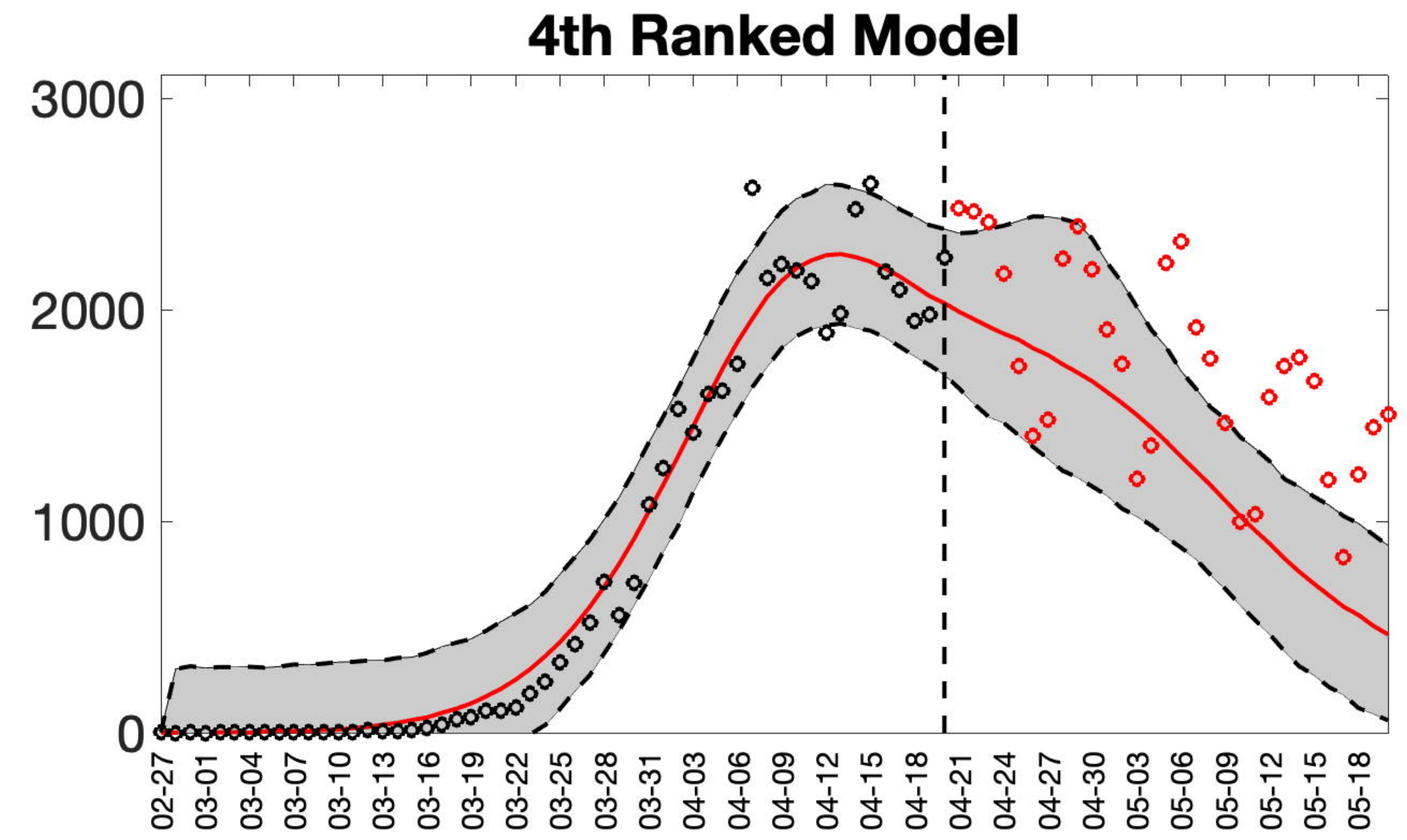
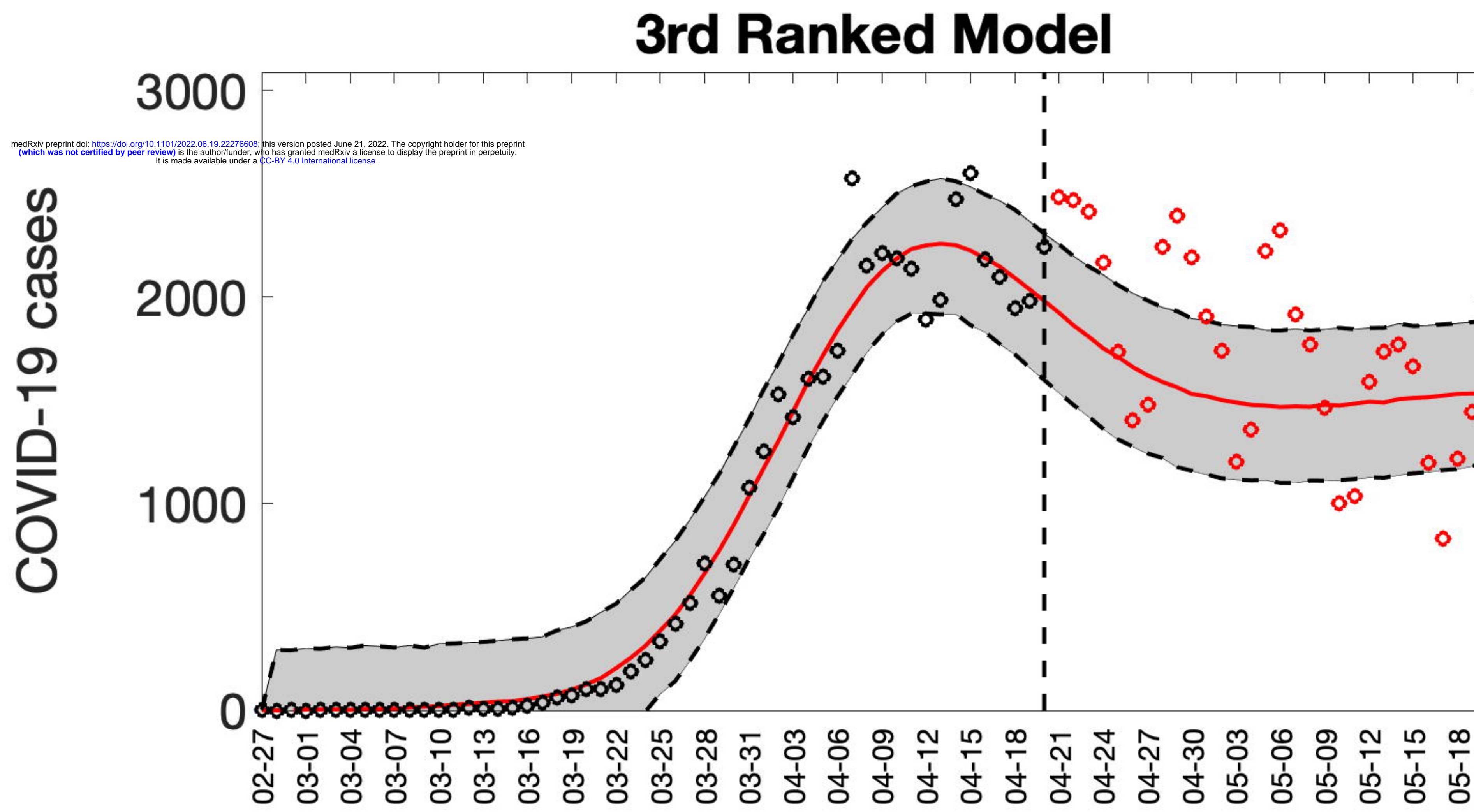
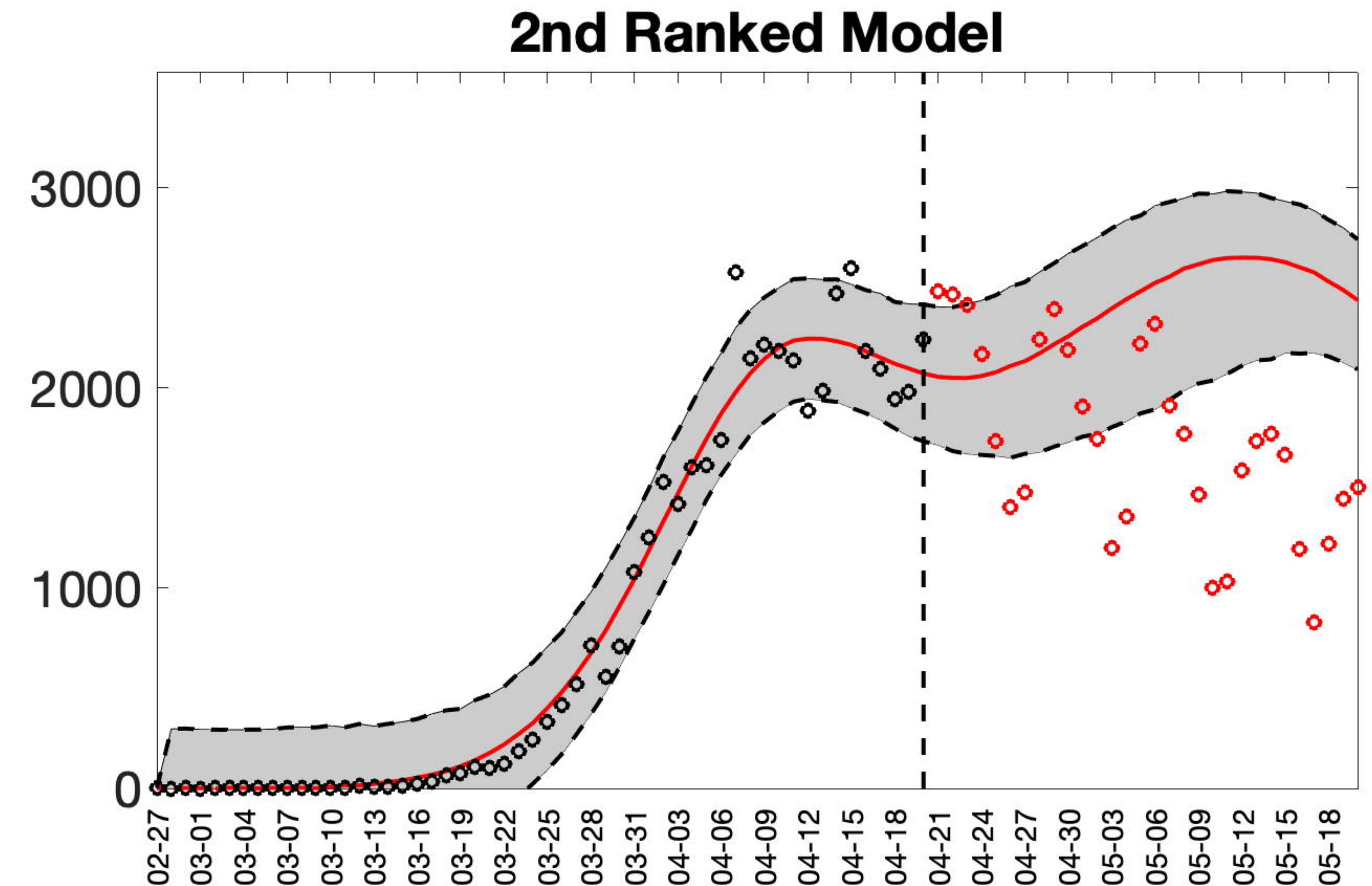
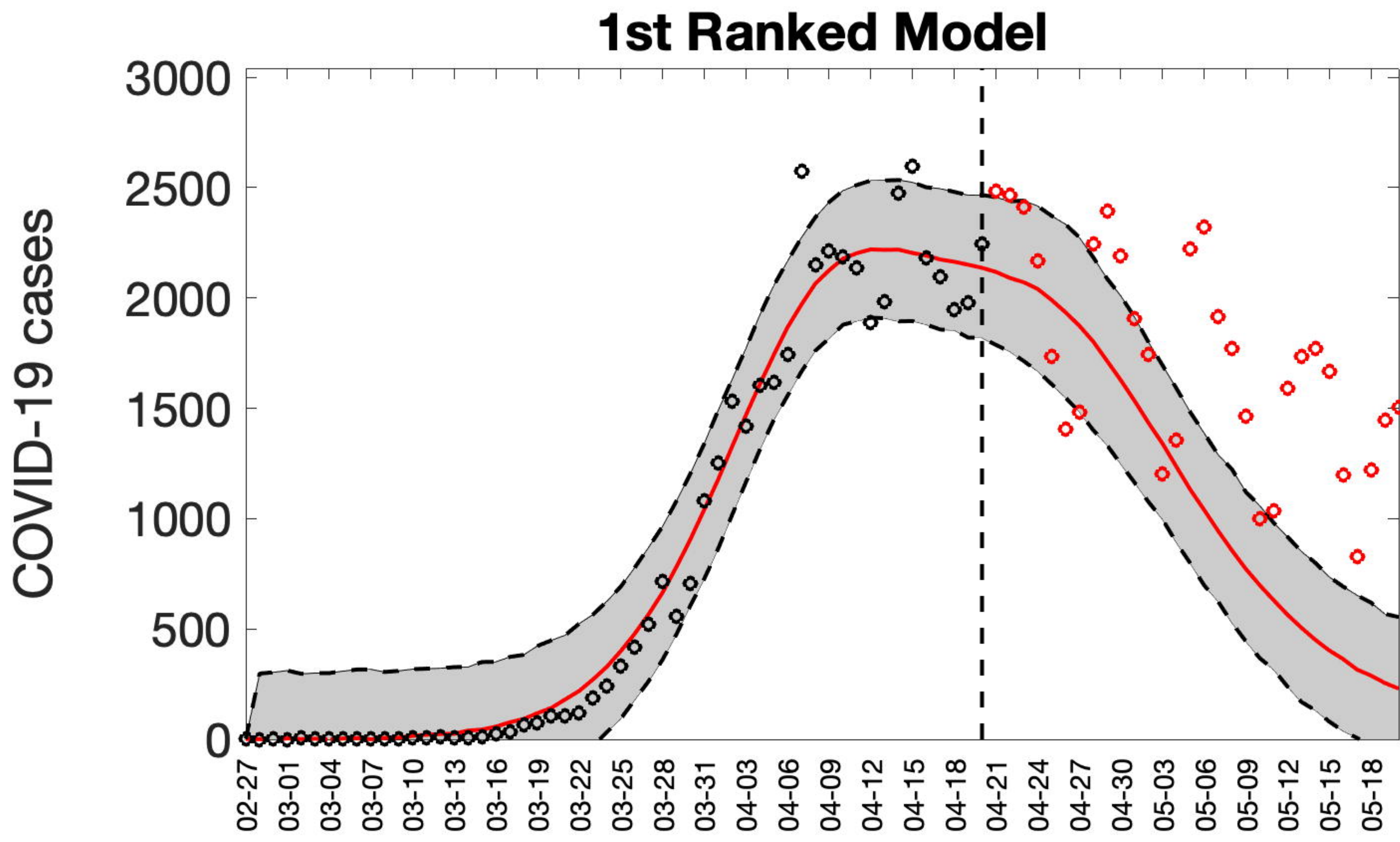
Frequency

medRxiv preprint doi: <https://doi.org/10.1101/2022.06.19.22276608>; this version posted June 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

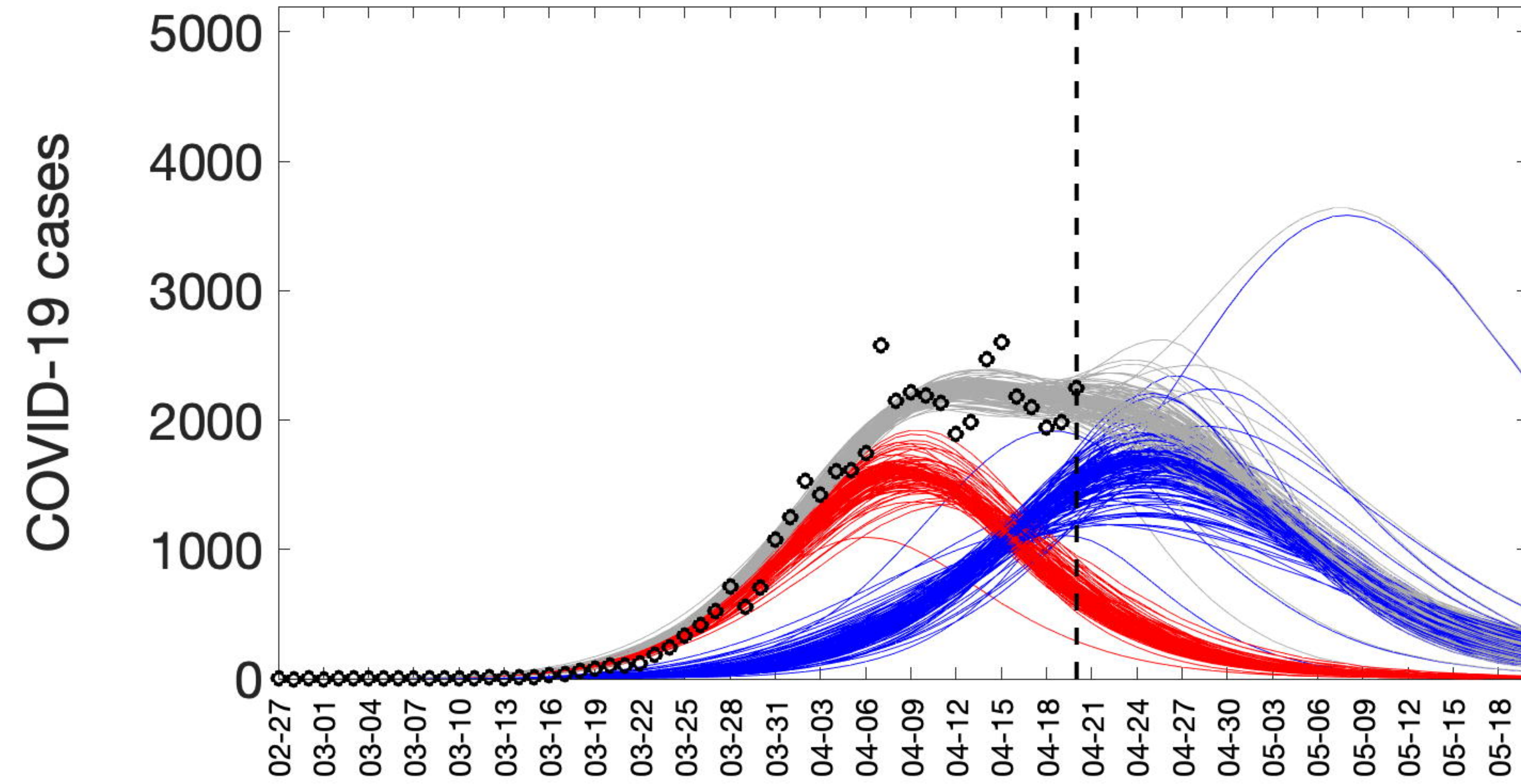


medRxiv preprint doi: <https://doi.org/10.1101/2022.06.21.22271608>; this version posted June 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

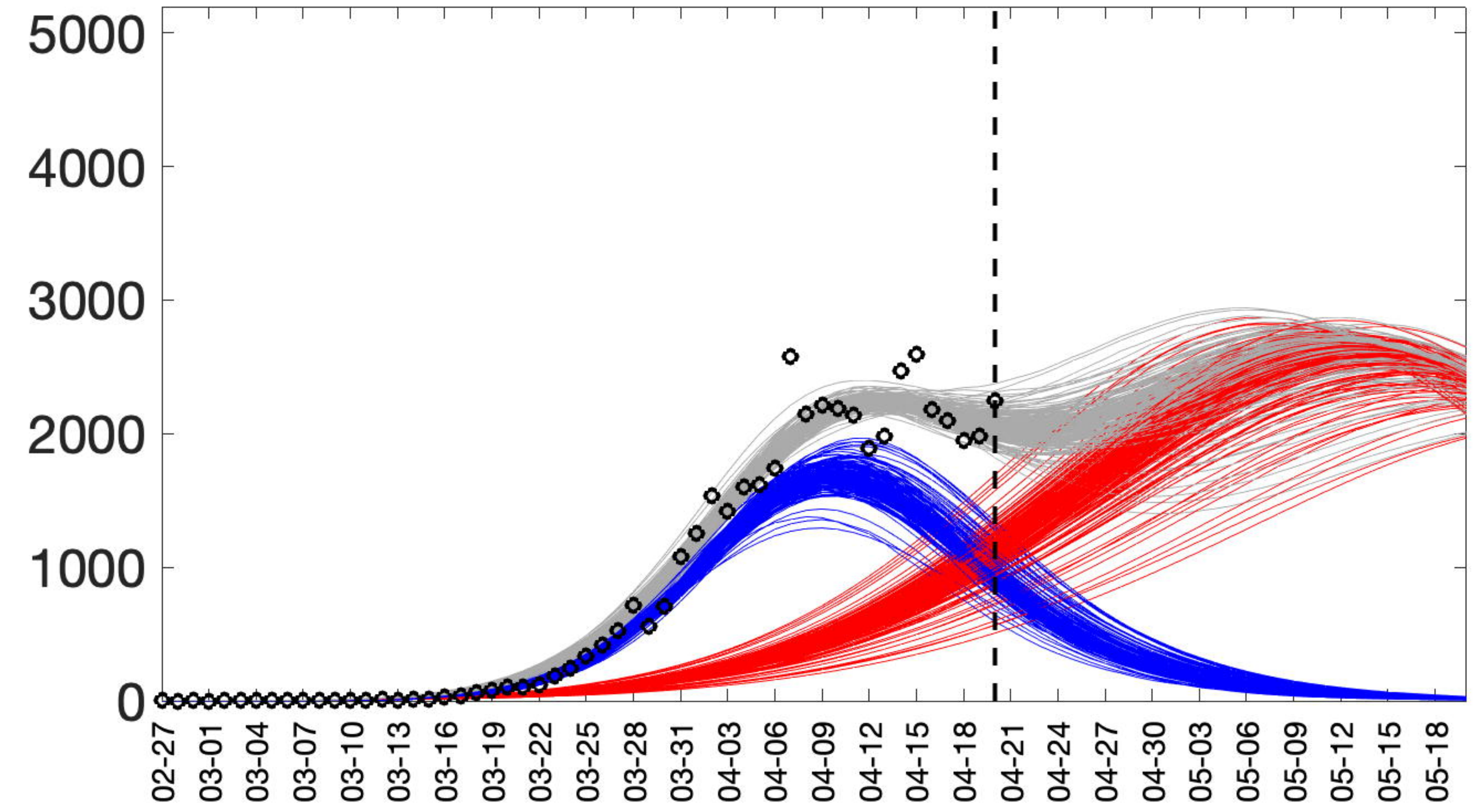




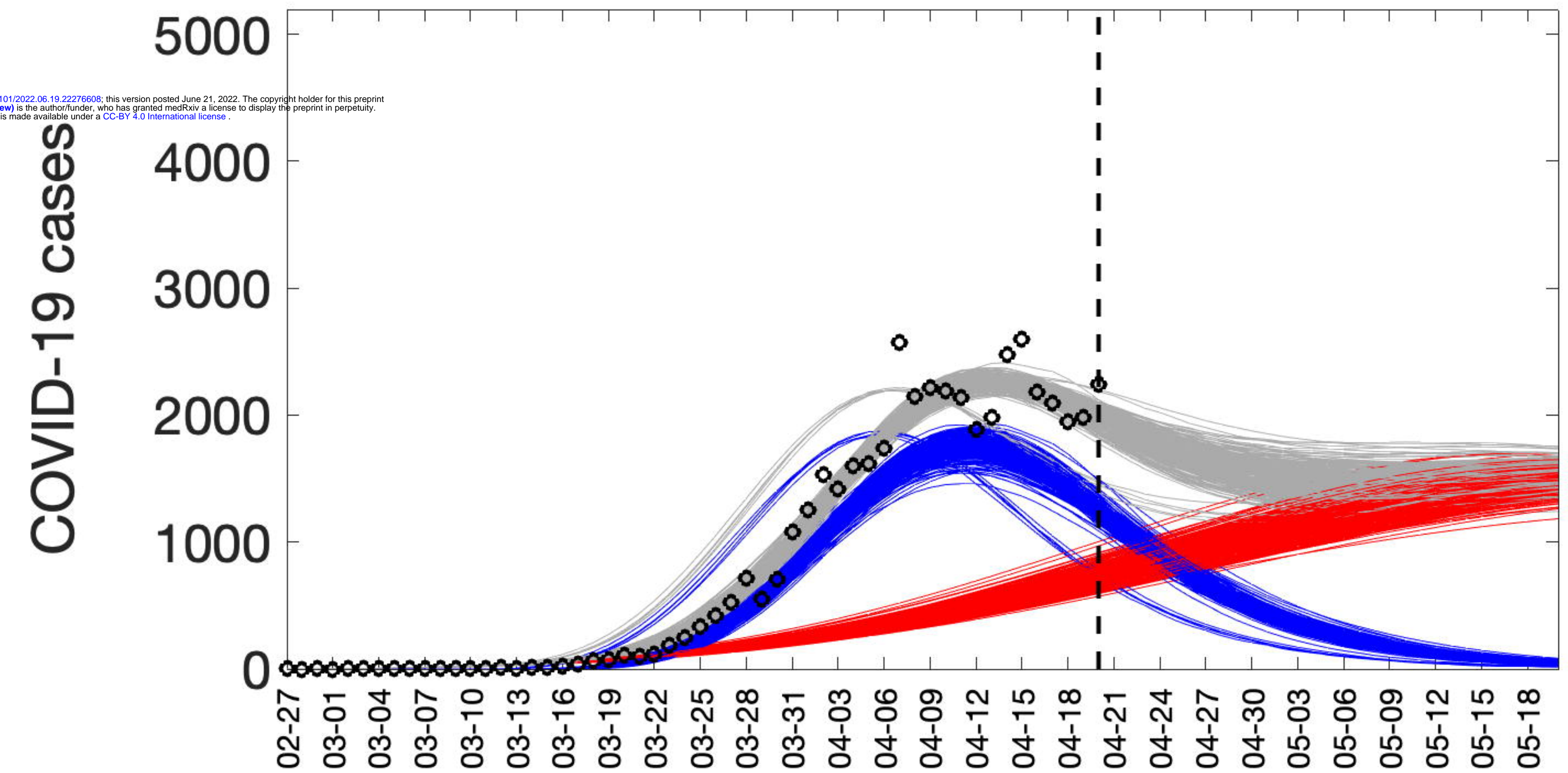
1st Ranked Model



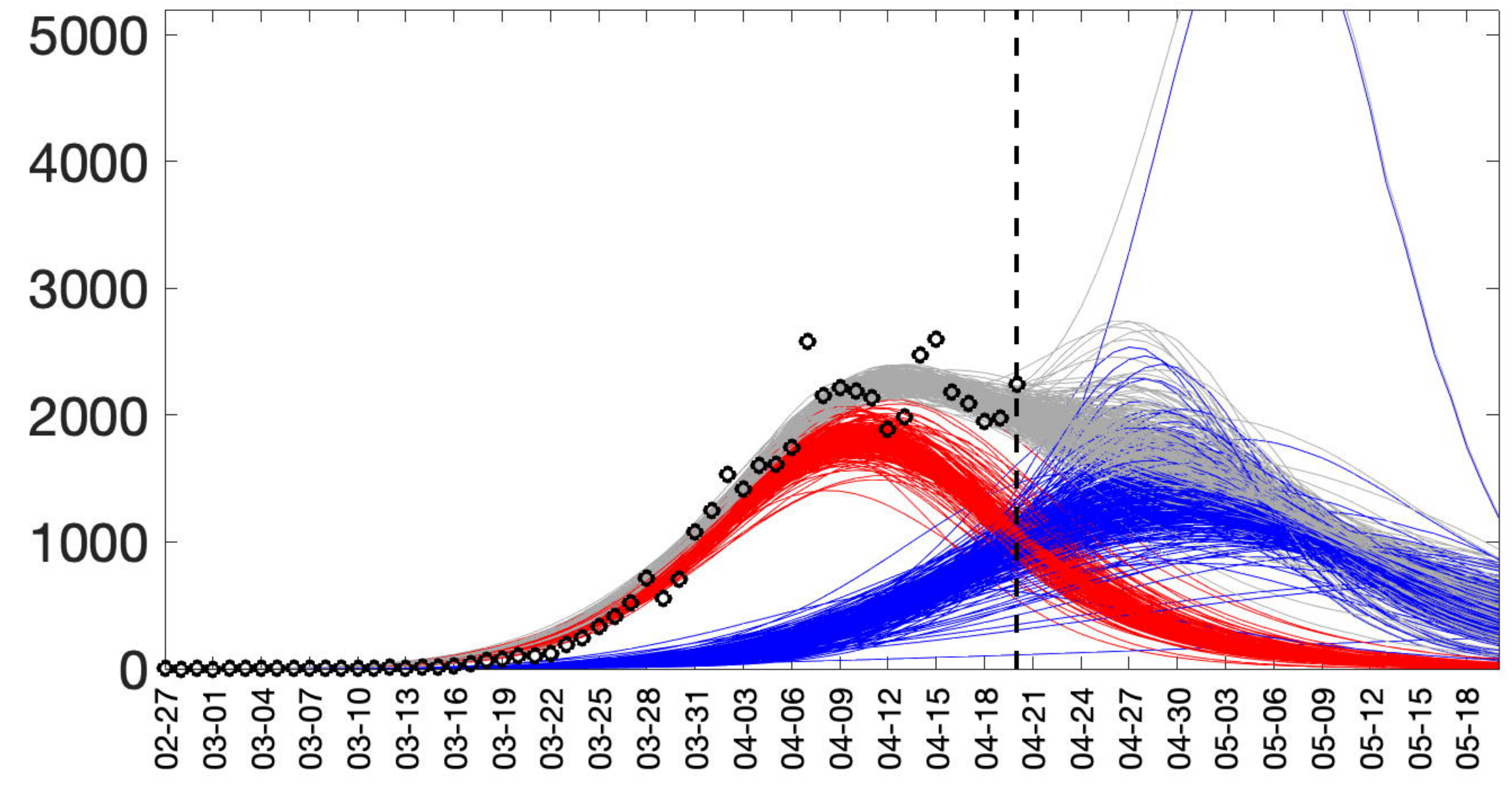
2nd Ranked Model



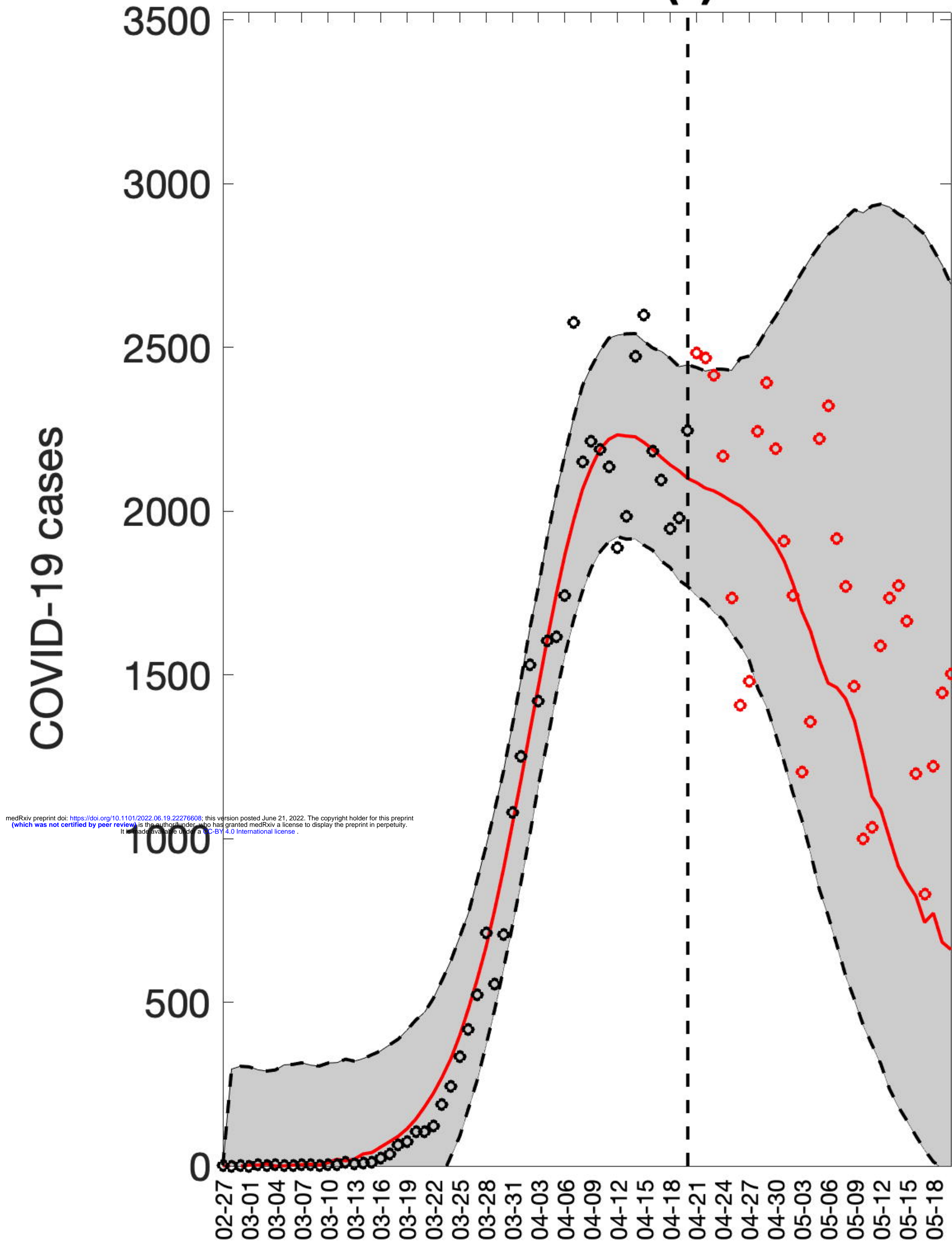
3rd Ranked Model



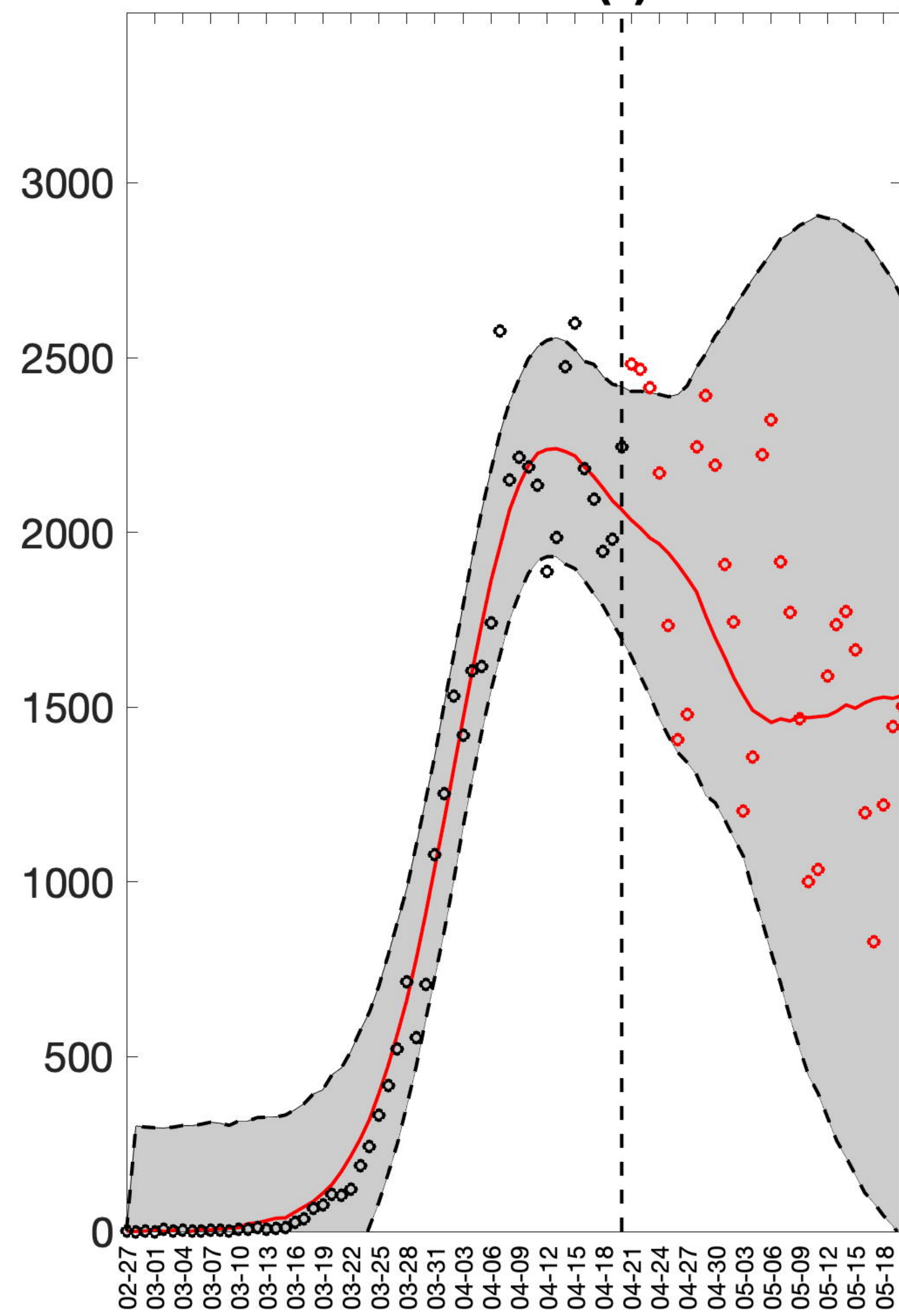
4th Ranked Model



Ensemble(2)



Ensemble(3)



Ensemble(4)

