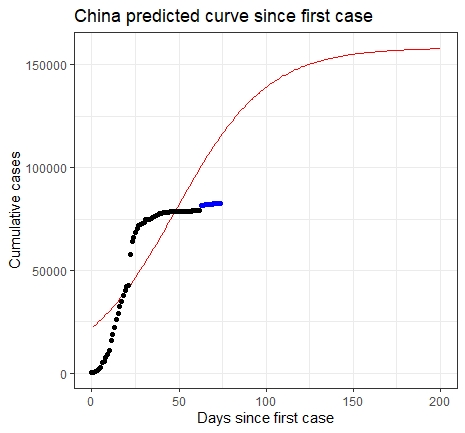
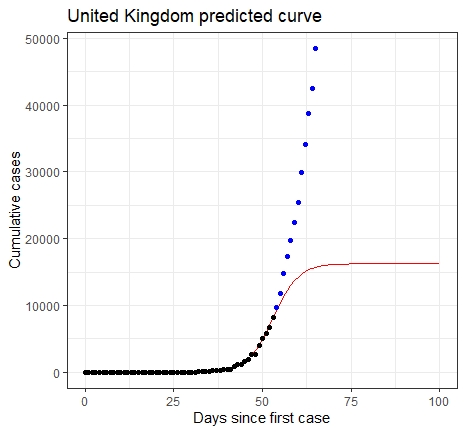
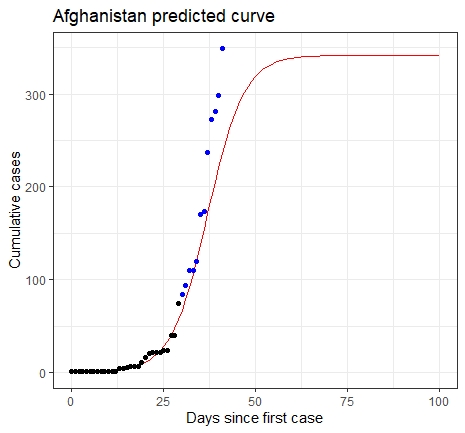
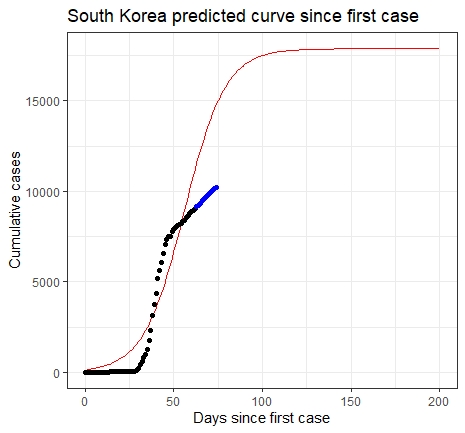
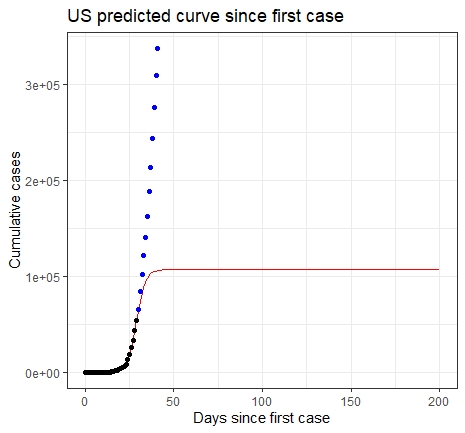
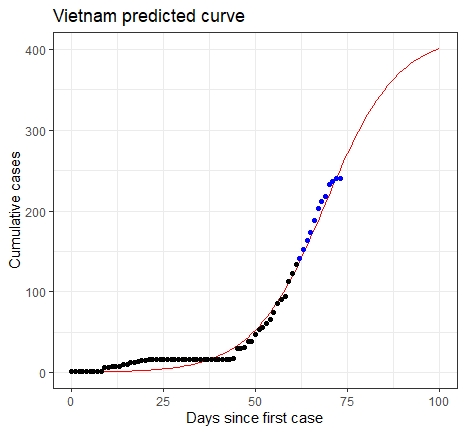
Discussion



Based on the figures, by March 23 the outbreak of COVID-19 had happened in China and Korea, with the decreasing trend of the growth rate indicating that this outbreak had been in control in these two regions. It is noticeable that the predicted number of cumulative cases and growth rate after March 23 are significantly smaller than the recorded ones, possibly attributed to their rapidly responding to and mitigating the spread of this epidemic. On the contrary, US and UK had also been experiencing the disease outbreak, but the growth rate after March 23 is much larger than the predicted one, which may be explained by the absence and inefficiency of public health interventions. In Vietnam and Afghanistan, a large number of recorded cases had not been seen by March 23, yet the increasing trend of the growth rate indicates an epidemic possibly going to happen, which is supported by the recorded data after March 23. These two regions’ efforts to control COVID-19 seems lack of proven effectivity.

The logistic growth model used in this report has limitations, especially for this dataset. The form of a logistic curve requires that t is the days since the first infection. However, it is likely that the first date recorded in our collected data is not the true date that the disease emerges in a particular region. For the dataset analyzed in this report, there exist regions whose large number of infected cases on the first day recorded suggests this first day may not be the day when the first infection happened. For example, the first date recorded for China is 1/22/20, however the recorded cases is 533, which means the exact date of the first infection should be earlier. This may account for the discrepancy between the fitted curve and the recorded number for China. Besides, the logistic growth model may miss some important information. The parameter a is the carrying capacity, and in this case, the maximum number of cases a region can reach, so a cannot be larger than the population of a certain region, but this information is not considered by the model. For c, the mid-point, we cannot add the effect of newly developed treatments and vaccinations, which may significantly shorten c, to this modeling process. Public health interventions and other social factors may also enlarge the difference between the trend provided by the model and the truth. Another factor needs to mention is that the data itself may not be accurate, the number of cases reported may be smaller than the truth, for which one possible explanation is that some cases have not been tested or they are tested falsely negative.

Several optimization algorithms were implemented when fitting the curve. The Newton–Raphson method was considered, and it was not easy to calculate the hessian matrix of the RSS for the original form, we transformed y and a to be the inverse. However, the starting values for this algorithm have a significant impact on the result, as the Newton method tends to find the local minimum instead of the global one, which is not suitable for non-linear least squares. To reduce the severity of this limitation, we considered adding the momentum when doing iteration. The final algorithm chosen is Adam, as it computes adaptive learning rates for each parameter and dampens oscillations. Adam may lead to the optimal solution, however it needs a large number of iterations which exceeds our computer capacity, thus our fitting may not minimize the RSS, resulting in an inaccurate prediction.