

# Supplementary information for “Sequential time-window learning with ABC”

João Pedro Valeriano<sup>1\*</sup>, Pedro Henrique Cintra<sup>2</sup>, Gustavo Libotte<sup>3</sup>, Igor Reis<sup>4</sup>, Felipe Fontinele<sup>5</sup>, Renato Silva<sup>3</sup>  
and Sandra Malta<sup>3</sup>

<sup>1</sup>Instituto de Física Teórica, Universidade Estadual Paulista, Brazil.

<sup>2</sup>Instituto de Física “Gleb Wataghin”, Universidade Estadual de Campinas, Brazil.

<sup>3</sup>Laboratório Nacional de Computação Científica, Brazil.

<sup>4</sup>Instituto de Física, Universidade de São Paulo, Brazil.

<sup>5</sup>Department of Physics, University of Alberta, Canada.

\*Corresponding author(s). E-mail(s): [joaopedrovm98@gmail.com](mailto:joaopedrovm98@gmail.com);

Contributing authors: [pedrohpc@ifi.unicamp.com](mailto:pedrohpc@ifi.unicamp.com);

[glibotte@lncc.br](mailto:glibotte@lncc.br); [igorreis.0203@gmail.com](mailto:igorreis.0203@gmail.com);

[feradofogo@gmail.com](mailto:feradofogo@gmail.com); [rssr@lncc.br](mailto:rssr@lncc.br); [smcm@lncc.br](mailto:smcm@lncc.br);

In this supplementary text, we present some relevant information to complement the results shown in the main text. In section [S1](#), we present the parameters of the ABC-SMC implementation used and also the prior distributions considered for each model parameter. In section [S2](#) we compare the results of considering fixed or adaptive window sizes, showing that the adaptive window size leads to very similar results (when not better), allowing the user not to worry about choosing an ideal value for the window size. Section [S3](#) presents how the quality of both fit and prediction via the flat prior approach increase as one consider more samples in the ABC-SMC algorithm. We conclude that, although increasing the sample number leads to better fitting results, it does not compare to using the information available from past time-windows. In section [S4](#), we can see in detail the results of fit and prediction for each time-window of the COVID-19 epidemic curve in Brazil. Lastly, in section [S5](#), we can see the same results presented for Brazil in the main text, now for other

six countries: Germany, India, Japan, South Korea, United States and United Kingdom. This indicates the robustness of the method to the dataset considered, as all these countries had different responses to the development of the COVID-19 pandemic.

## S1 ABC-SMC parameters

We used uniform priors for every parameter of the SEIRD model for inference via ABC-SMC. The bounds of each prior are detailed in table S1. Notice that  $\tau$  is the incubation period, that is  $\tau = 1/c$ . Also, notice that  $N$  is the only parameter whose prior distribution varies from one country to another, in order to account for the different population sizes. We have previously investigated the effect of informative priors in epidemic curve fitting, obtained from empirical data, and concluded that this information did not provide considerable advantage [1]. Therefore, we use uniform (flat) priors.

We employ a Sequential Monte Carlo of three posteriors, each with  $n$  samples to be accepted, and the algorithm generates samples until  $n$  are accepted. The tolerance for each posterior of the ABC-SMC algorithm is set on the median of the NRMSD over the samples accepted in the last posterior. To choose the first tolerance, we generate  $Mn$  samples, and choose the percentile  $100/M$  of the NRMSD over these samples. In our case, we choose  $n = 1000$  and  $M = 10$ .

Parameter	Lower bound	Upper bound	Units
$N$	0.1% pop.	20% pop.	individuals
$\beta_I$	0	1	days <sup>-1</sup>
$\beta_E$	0	1	days <sup>-1</sup>
$\mu$	0	1	days <sup>-1</sup>
$\gamma$	0	1	days <sup>-1</sup>
$\tau$	0	30	days
$c_R$	0	1	dimensionless
$c_E$	0	2	dimensionless

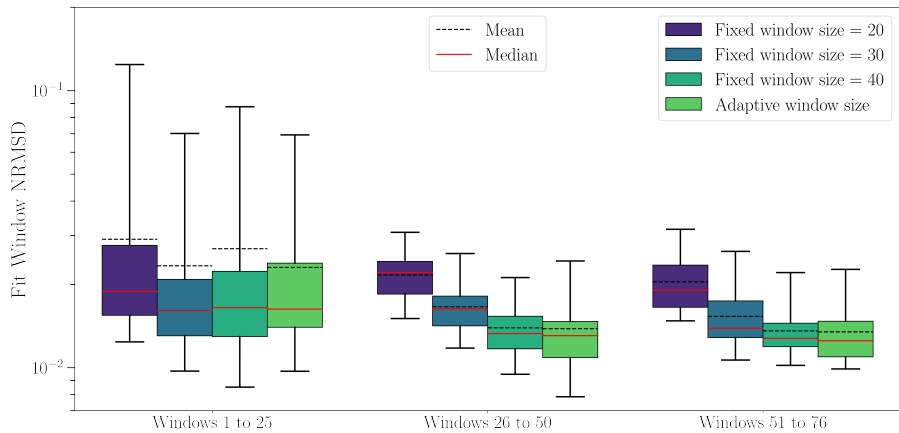
**Table S1** Bounds of uniform priors used in our fitting procedure.

## S2 Comparing fixed and adaptive window sizes

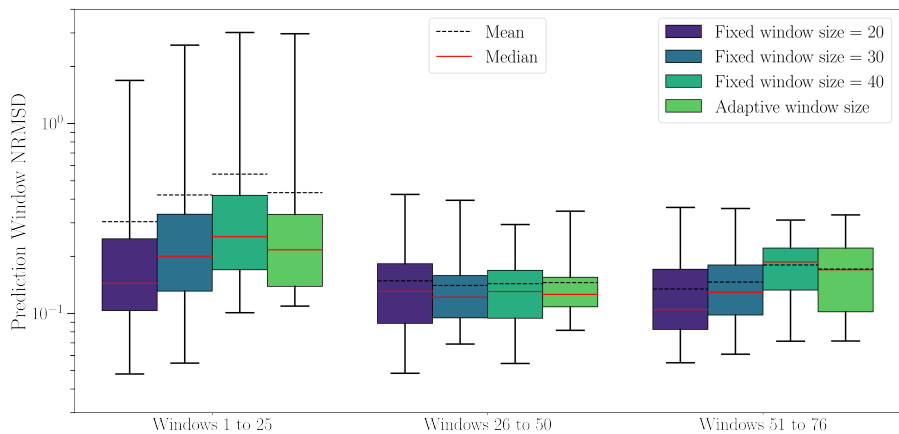
In Figures S1 and S2, we respectively compare the fit and the prediction for data on COVID-19 in Brazil, obtained with window sizes fixed on different values or with adaptive values. We divide the epidemic curve in three sets of time-windows, so that we can perceive the difference between the initial region – where the NRMSD is higher as this region is harder to fit and the algorithm has not accumulated enough information to produce useful priors yet – and the rest of the curve, where the NRMSD value seems to stabilize, even though it still shows fluctuation, as there are regions that are harder to be described by a SEIRD model with parameters constant through the time-window.



We can see that the adaptive window size provides very similar results to the different fixed window sizes. The adaptive window size algorithm makes the method independent of the arbitrariness of choosing a specific value for the window size.



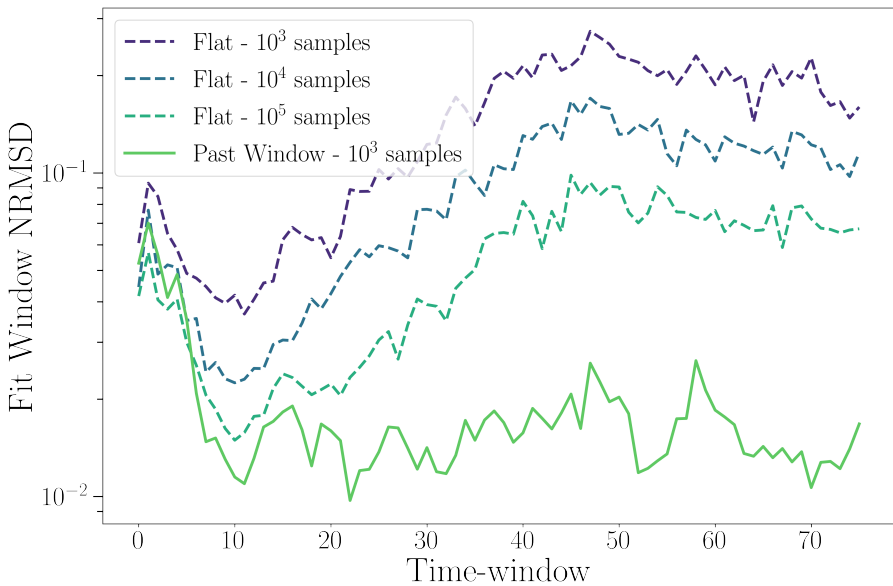
**Fig. S1** NRMSD of the fit of the COVID-19 epidemic curve in Brazil, for different fixed window sizes, and for the adaptive window size. Box-plots take into account the values of the NRMSD over multiple windows of three separate regions over the epidemic curve, and over ten different executions of the algorithm, to take its stochasticity into account.



**Fig. S2** NRMSD of the prediction for the COVID-19 epidemic curve in Brazil, for different fixed window sizes, and for the adaptive window size. Box-plots take into account the values of the NRMSD over multiple windows of three separate regions over the epidemic curve, and over ten different executions of the algorithm.

### S3 Increasing the number of samples in the flat prior approach

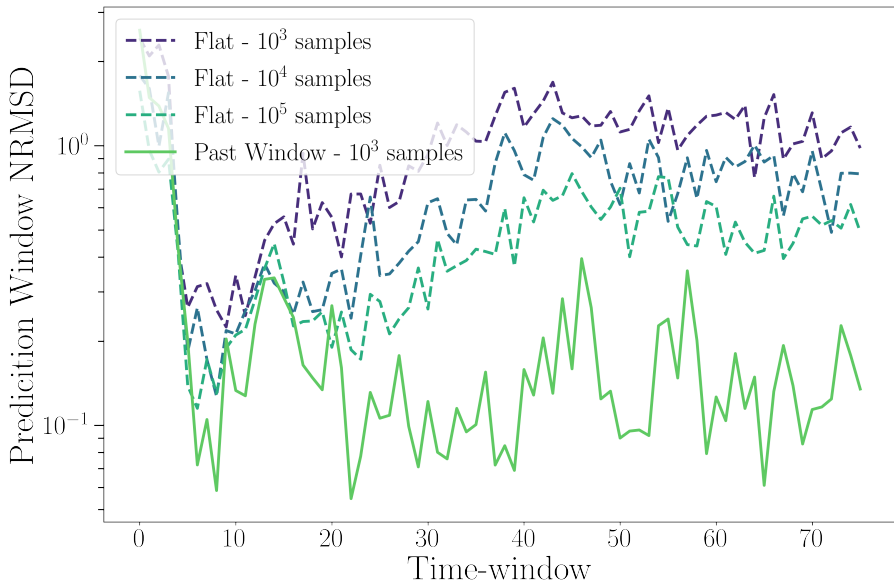
As one can see in the main text, there is a dramatic difference in the quality of both fit and prediction between the flat prior and the past window approaches. It is clear this is due to the more informative prior used for each window in the past window approach. One may expect to improve the quality of results of the flat prior approach by increasing the number of samples to be accepted, leading to a better approximation of the posterior distributions. We tested for the improvement in both the fit and the prediction by increasing the number of samples to be accepted (it is important to remember that the algorithm will generate as many sample as needed to obtain the desired amount with lower NRMSD than the chosen tolerance) ten and a hundred times, for each ABC-SMC posterior, for a total of three posteriors.



**Fig. S3** NRMSD of the fit of the SEIRD model over time-windows of cumulative cases and deaths of COVID-19 in Brazil, for different numbers of samples in the flat prior approach, compared to the past window approach.

We present the results of this test for fit and prediction NRMSDs in Figures S3 and S4, respectively, considering the SEIRD model applied to data on COVID-19 cumulative cases and deaths in Brazil. The curves represent the average over ten different executions of the algorithm. It is clear that increasing the number of samples does lead to an improvement of the model fit and of the predictions, but it is still considerably far from the results of considering information obtained fitting the past windows. This shows how useful it is to

take into account the information provided by data in time-windows before the current one.



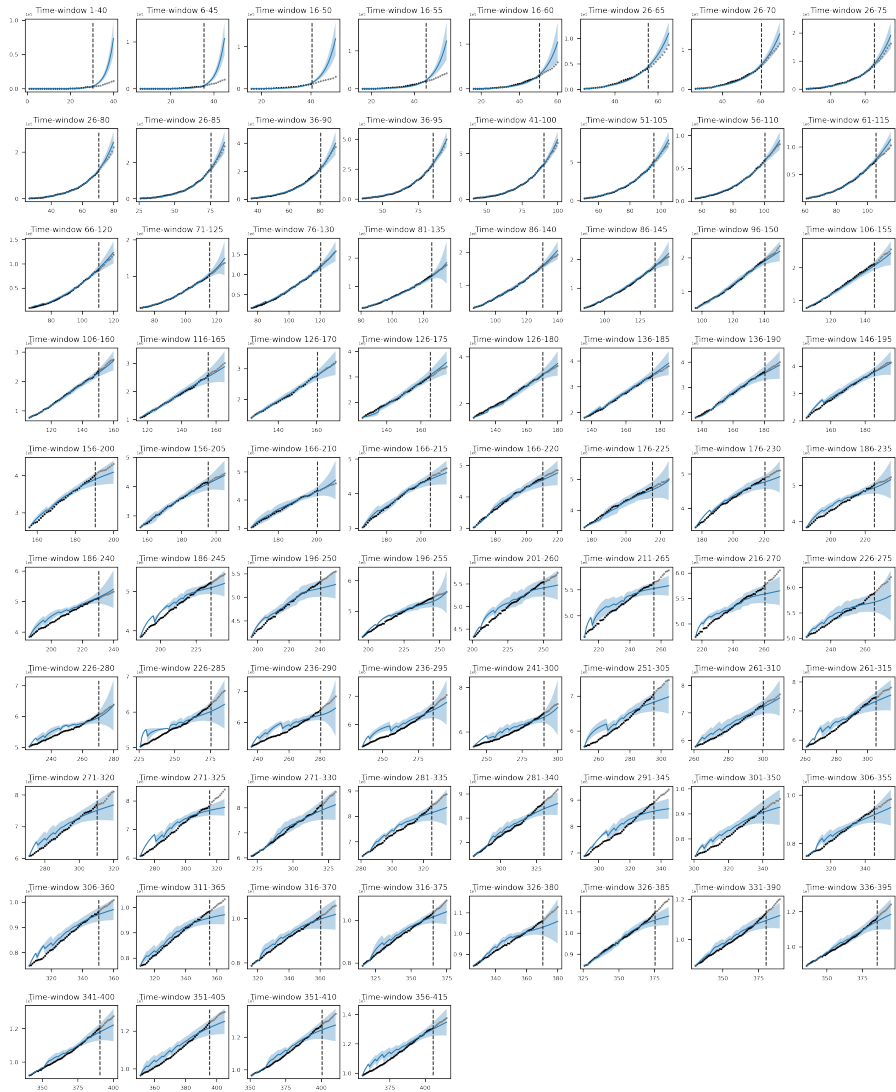
**Fig. S4** NRMSD of the predictions of the SEIRD model over time-windows of cumulative cases and deaths of COVID-19 in Brazil, for different numbers of samples in the flat prior approach, compared to the past window approach.

## S4 Window by window results for Brazil

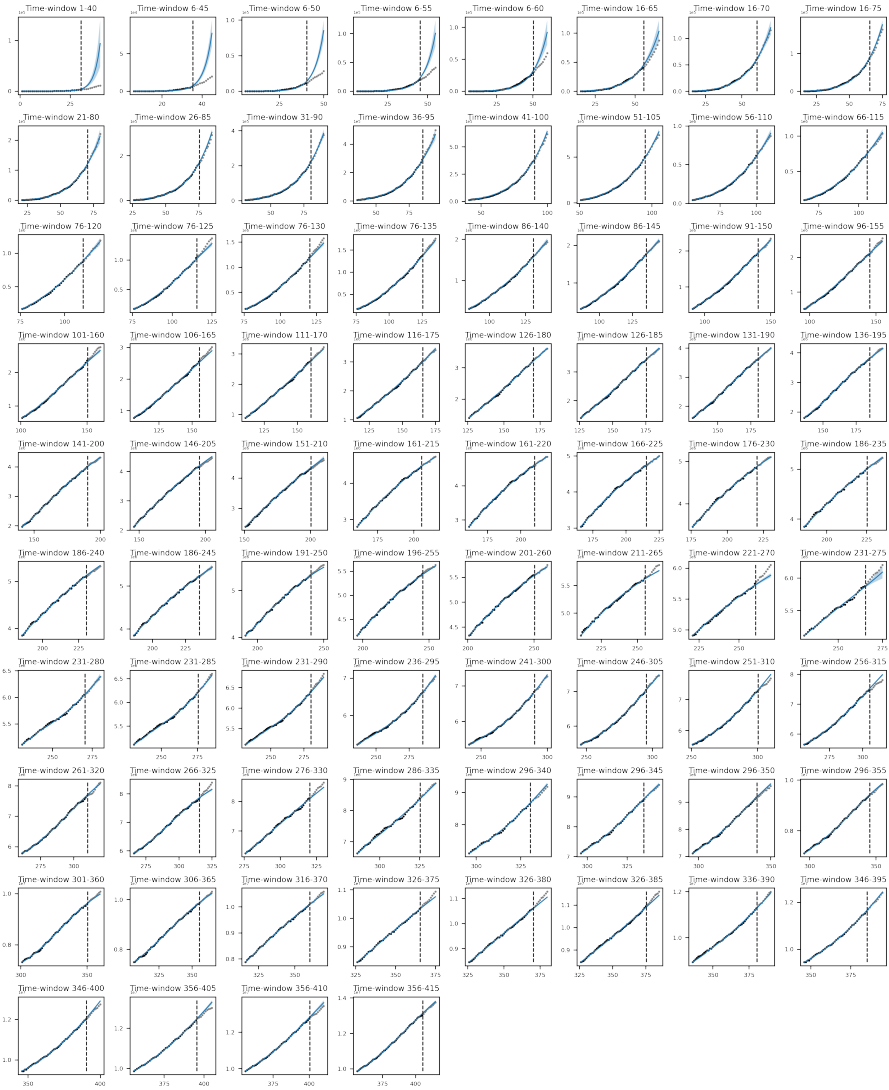
Here, we can see the fit and the prediction obtained for each time-window of the epidemic curve of Brazil, for both the flat prior and the past window posterior approaches. The data considered for the fit, and the data following it, to be compared to the forecast, are presented respectively in black and gray. In blue, we have the results of the fitted model, to be compared to both fit and prediction ranges.

We use the 100 best sets of parameters, that is, the sets of parameters that minimize the fit NRMSD. For each parameter set, we compute prediction curves. On each day of the prediction curve, we have a distribution of predictions for that day. The blue curve describes the average behavior of the predictions, and the blue shade around it is the interval between 2.5th and 97.5th percentiles from the distribution for each day.

## 6 Supplementary information for “Sequential time-window learning with ABC”



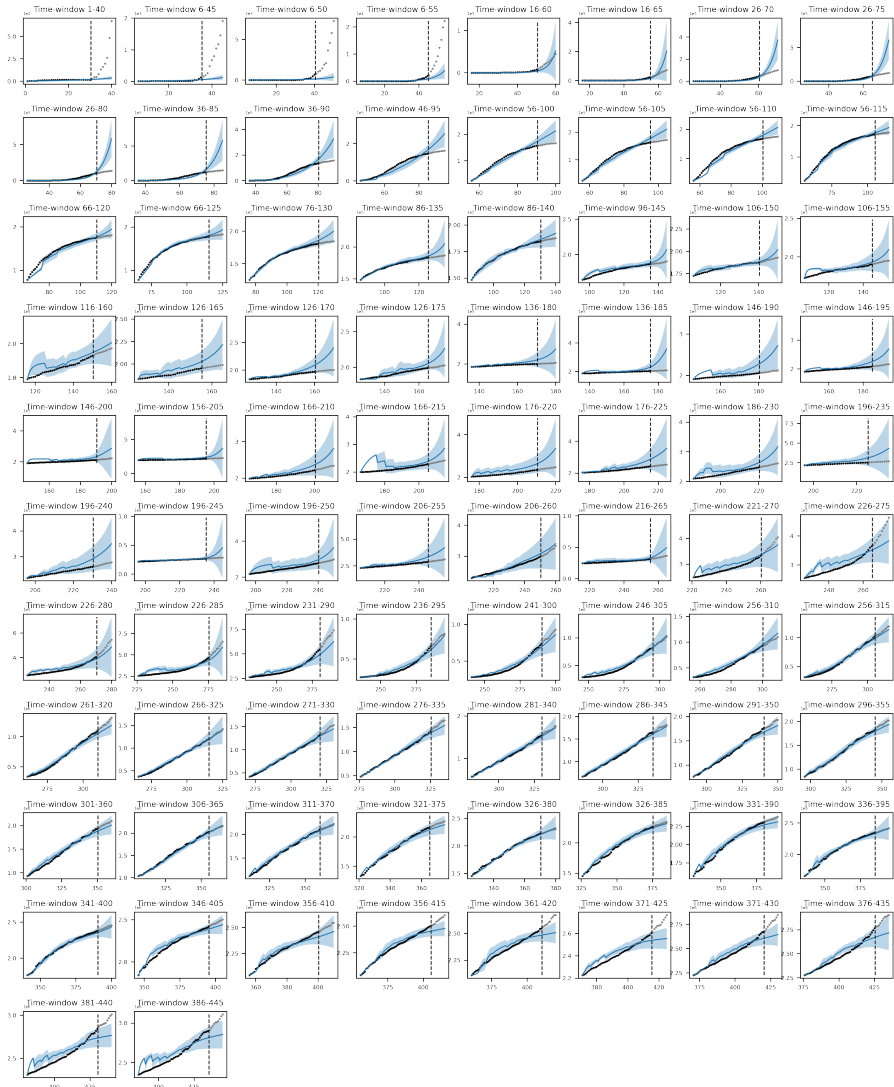
**Fig. S5** Fitting and prediction for each time window during the whole epidemic curve for Brazil using the flat prior approach.



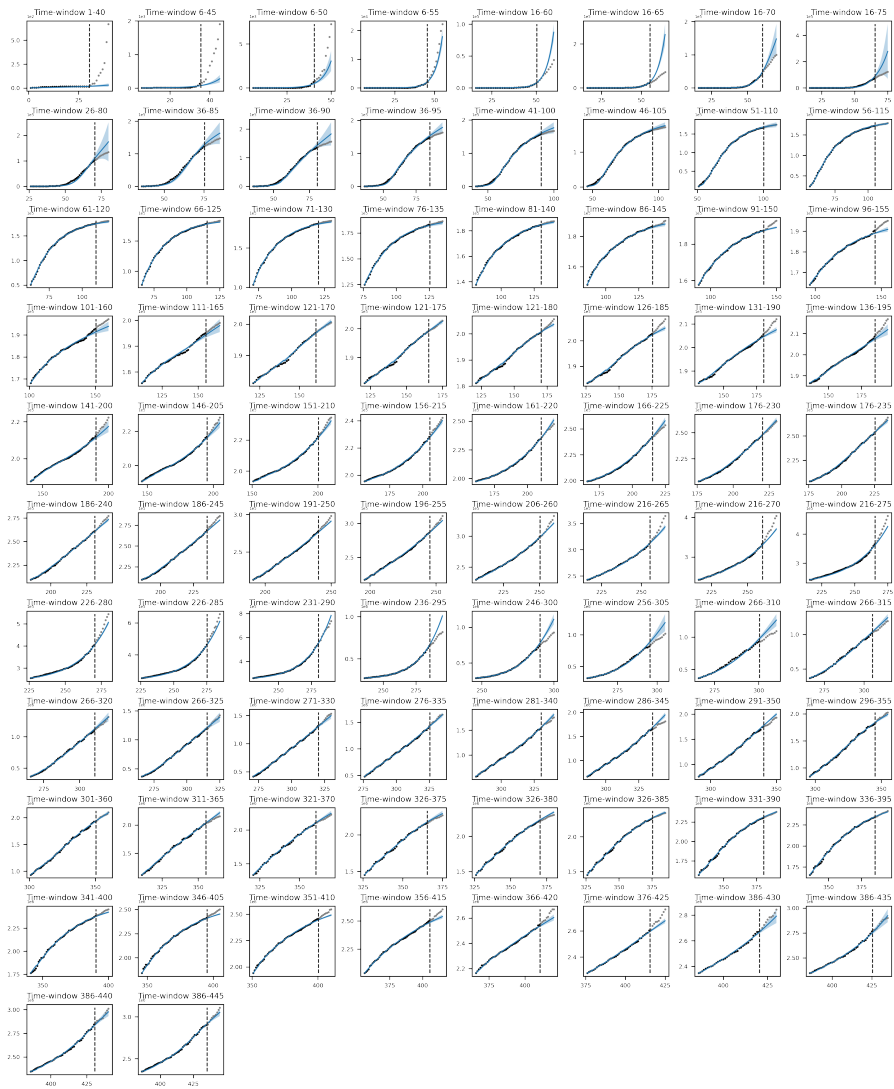
**Fig. S6** Fitting and prediction for each time window during the whole epidemic curve for Brazil using the past window approach.

## S5 Results for other countries

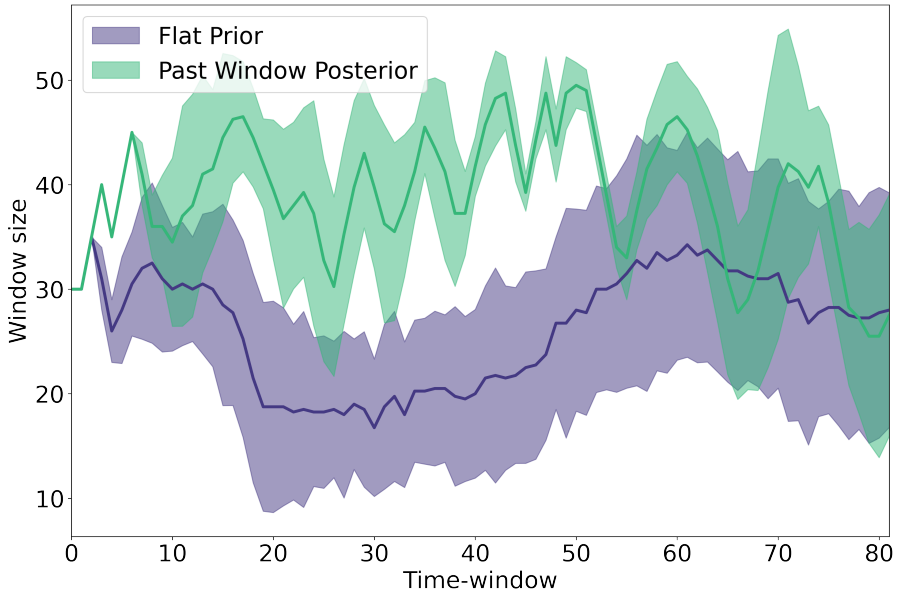
### S5.1 Germany



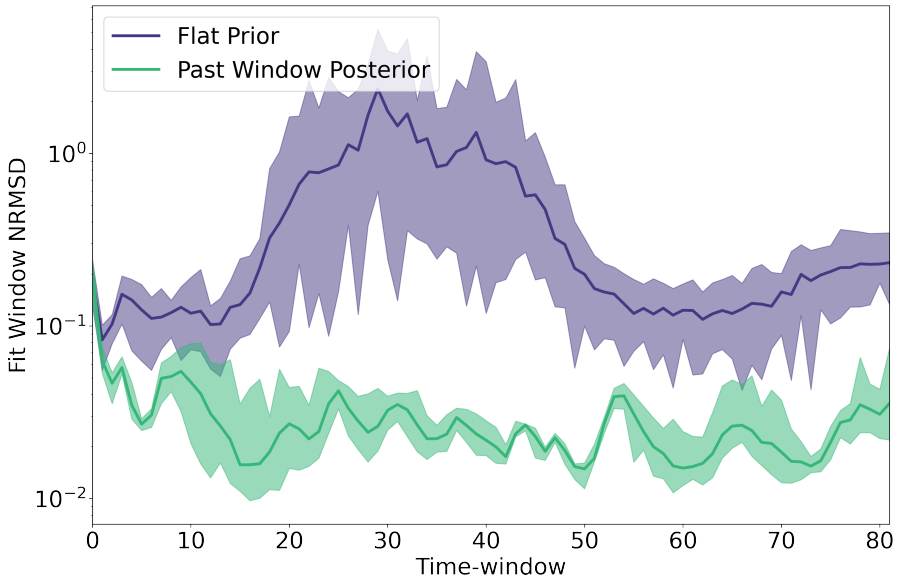
**Fig. S7** Fitting and prediction for each time window during the whole epidemic curve for Germany using the flat prior approach.



**Fig. S8** Fitting and prediction for each time window during the whole epidemic curve for Germany using the past window approach.

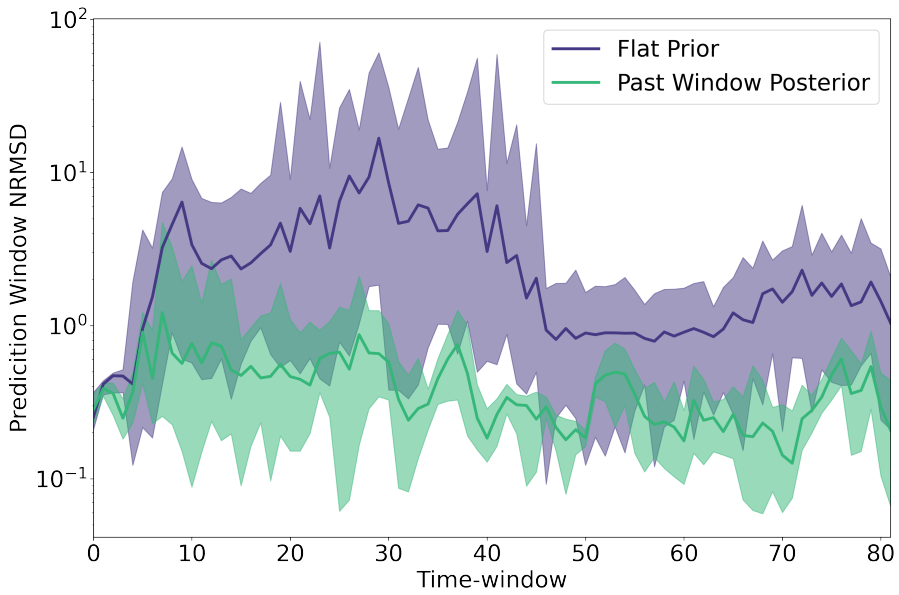


**Fig. S9** Windows' sizes selected over five executions, for the case past window posteriors.

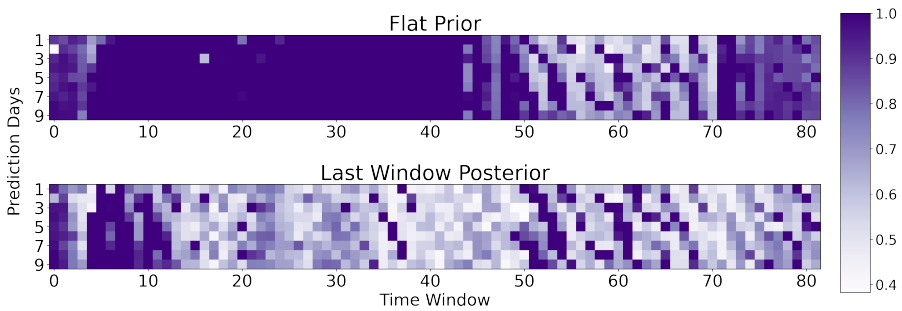


**Fig. S10** Normalized RMSD over the fitting window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



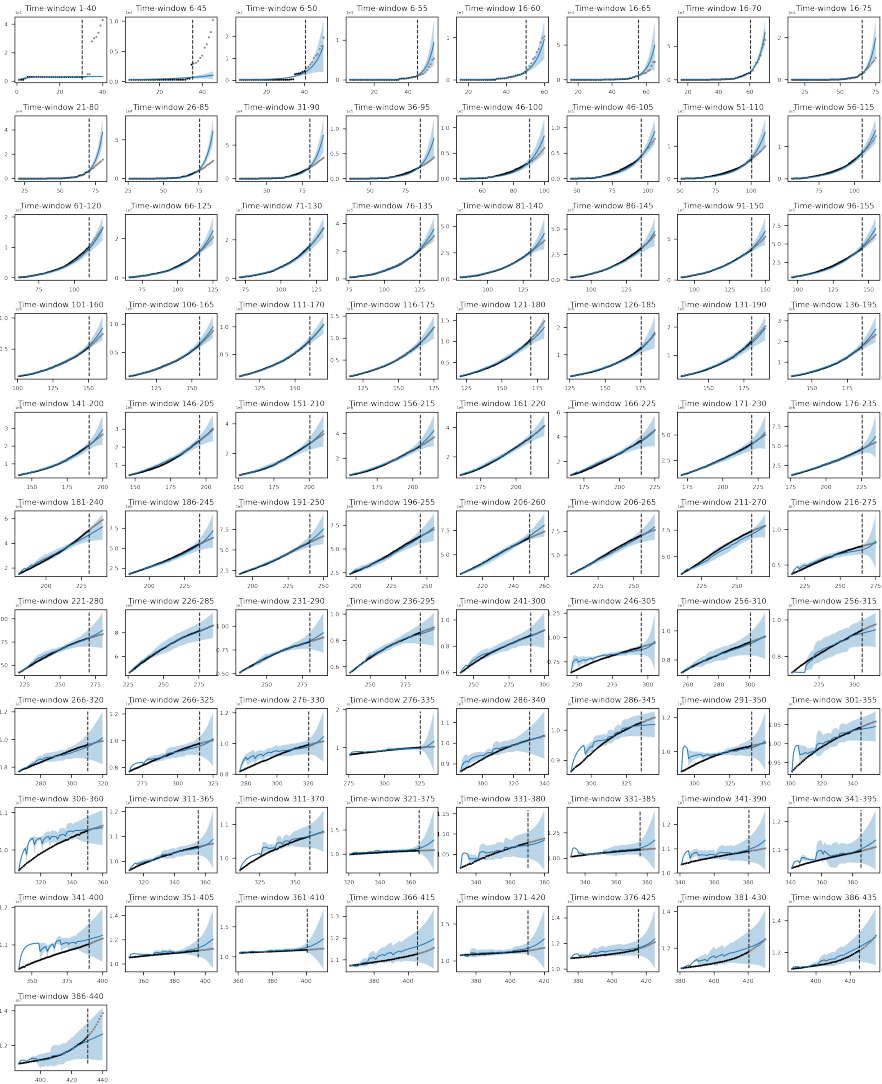


**Fig. S11** Normalized RMSD over the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



**Fig. S12** Normalized RMSD for every day of the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

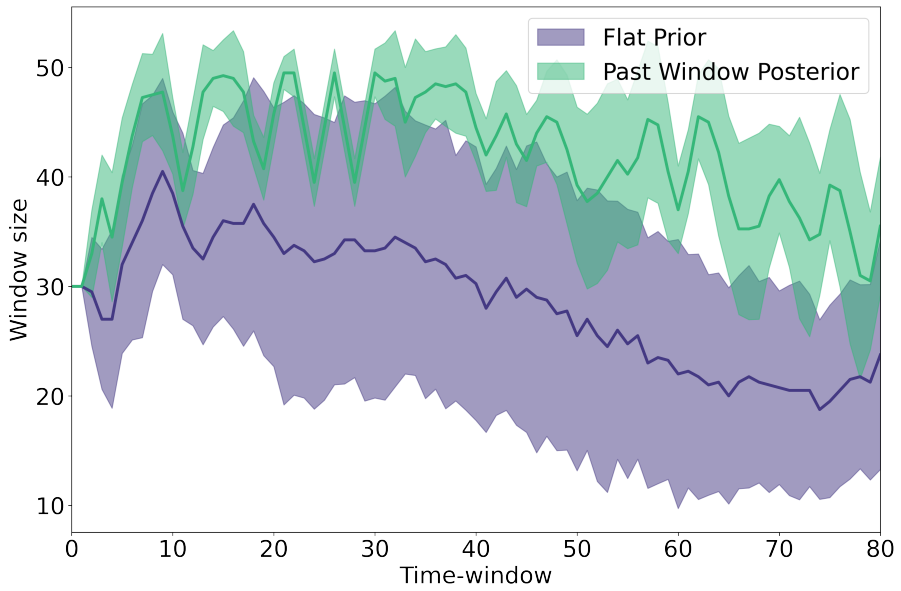
## S5.2 India



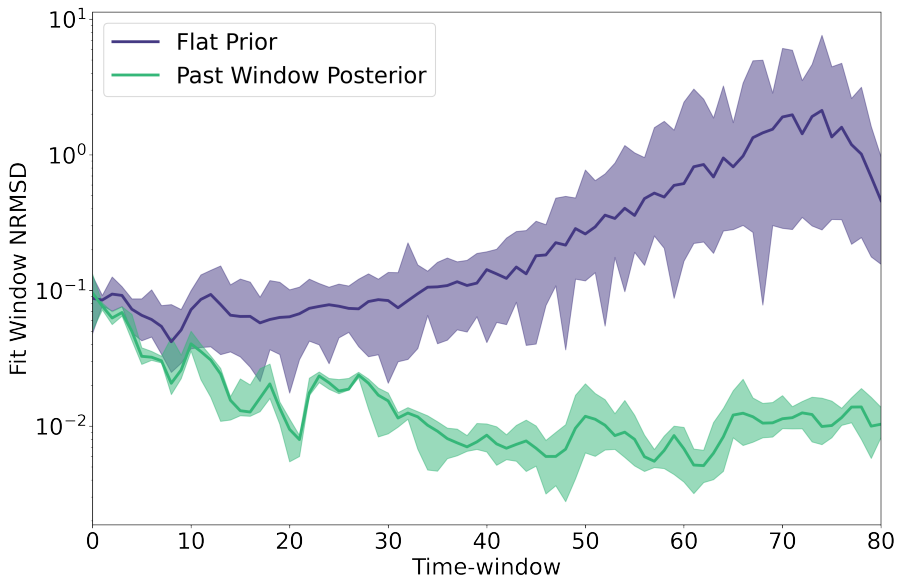
**Fig. S13** Fitting and prediction for each time window during the whole epidemic curve for India using the flat prior approach.



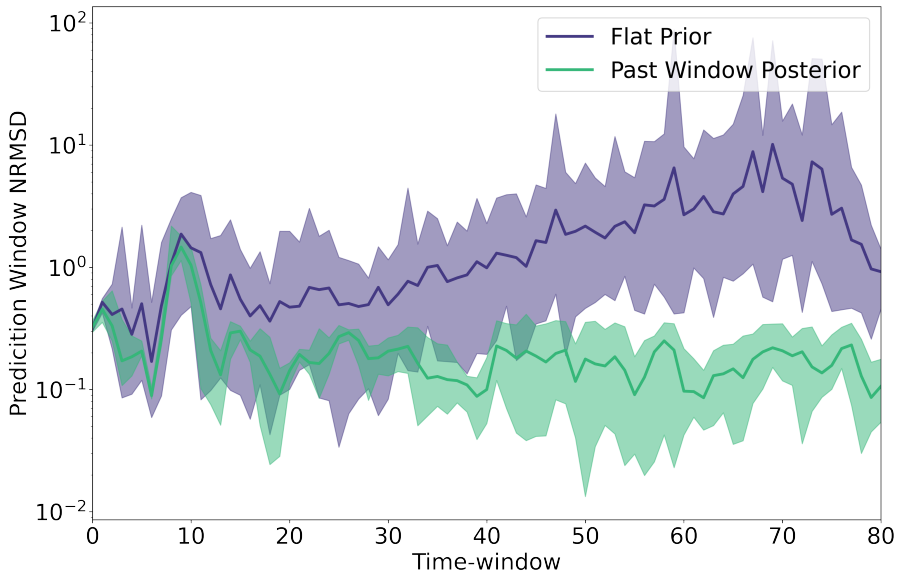
**Fig. S14** Fitting and prediction for each time window during the whole epidemic curve for India using the past window approach.



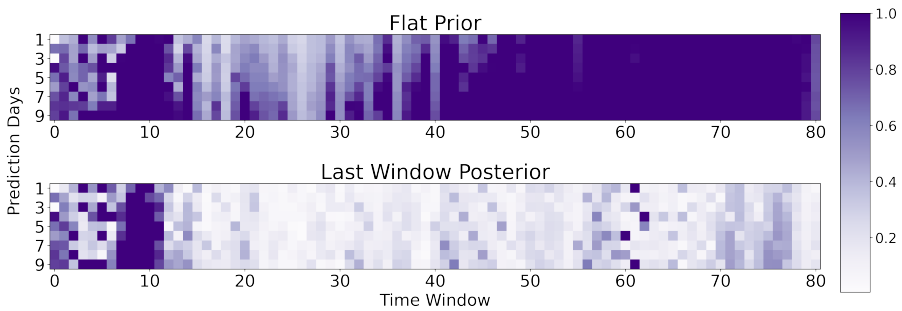
**Fig. S15** Windows' sizes selected over five executions, for the case past window posteriors.



**Fig. S16** Normalized RMSD over the fitting window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

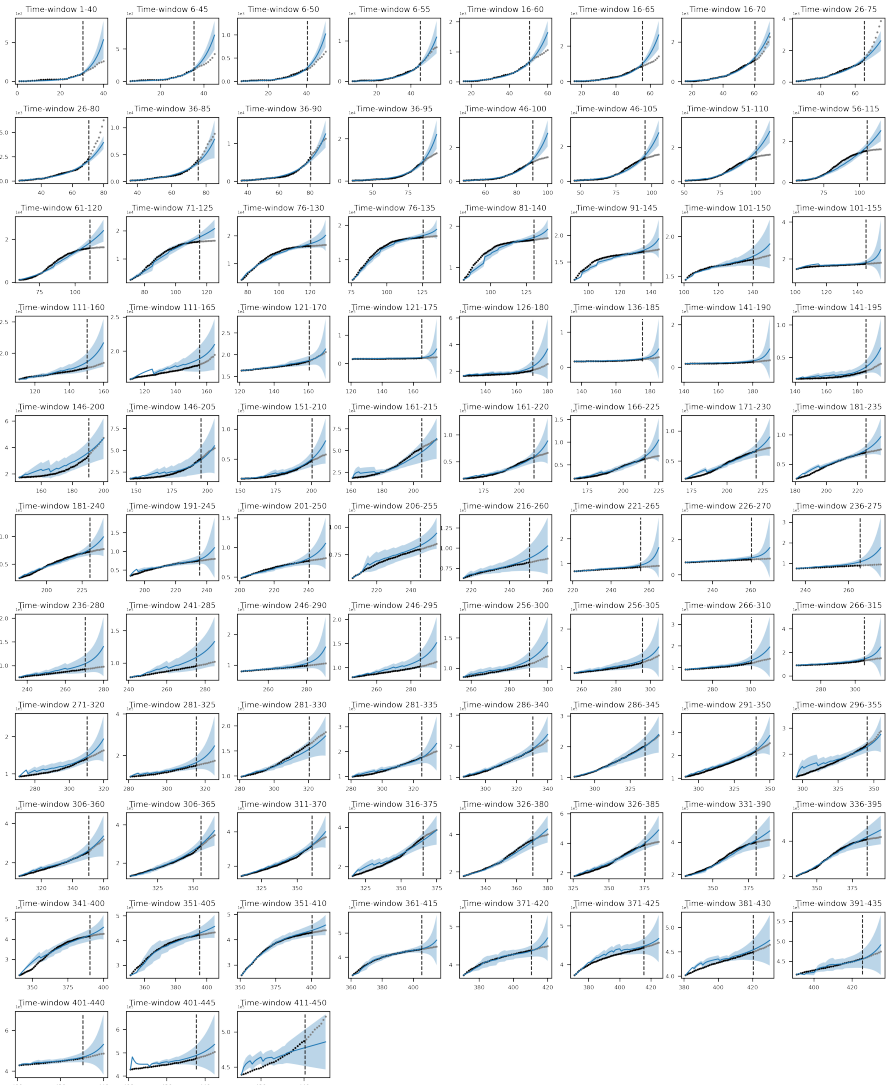


**Fig. S17** Normalized RMSD over the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

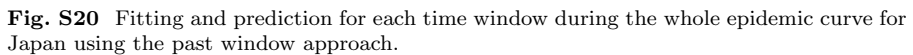


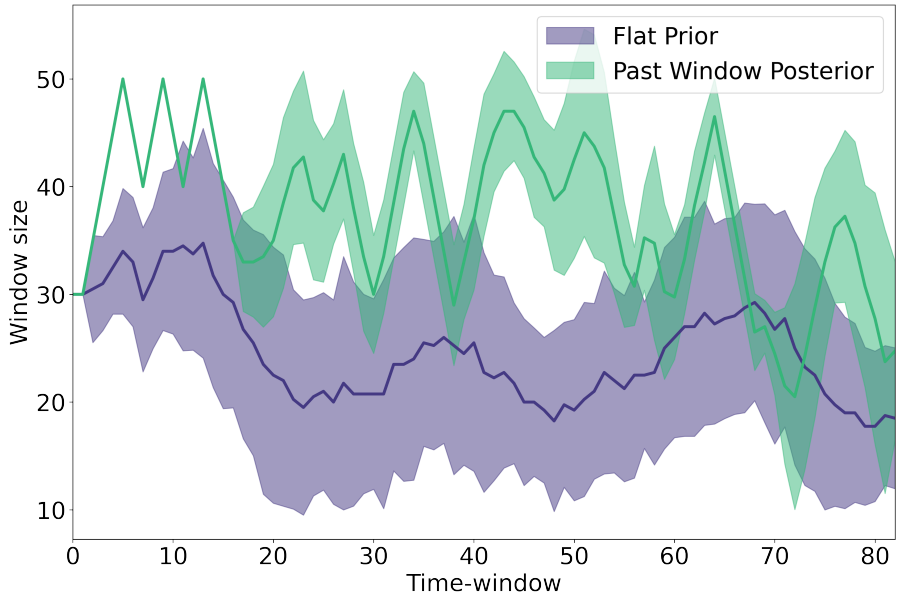
**Fig. S18** Normalized RMSD for every day of the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

### S5.3 Japan

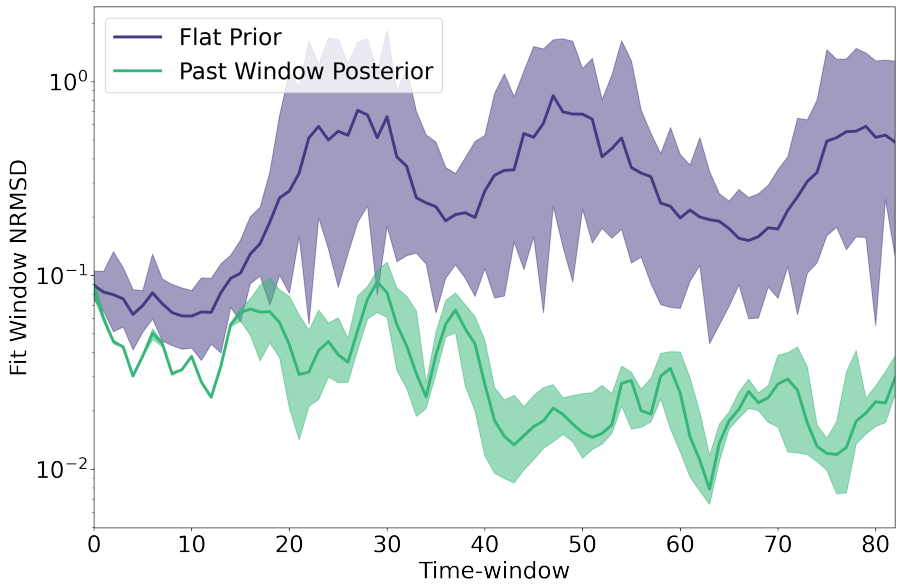


**Fig. S19** Fitting and prediction for each time window during the whole epidemic curve for Japan using the flat prior approach.



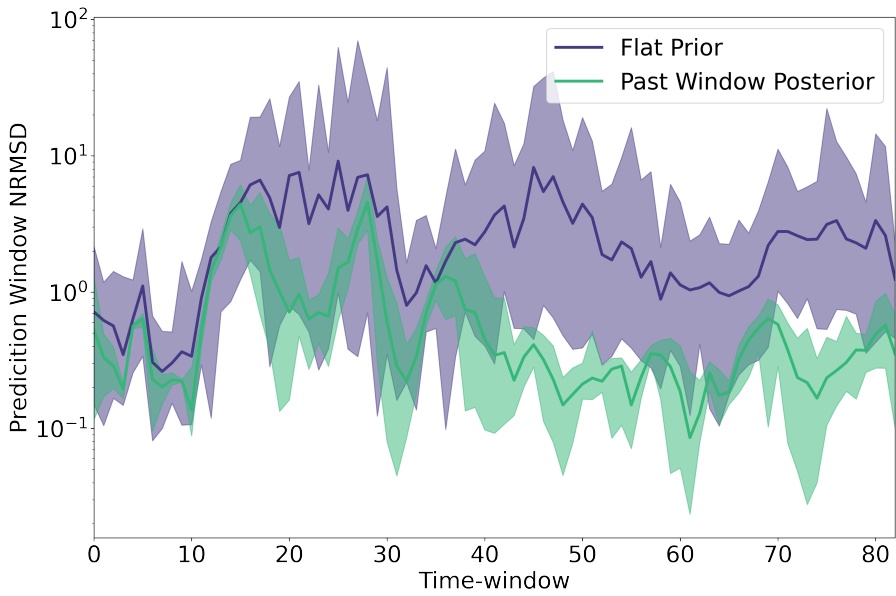


**Fig. S21** Windows' sizes selected over five executions, for the case past window posteriors.

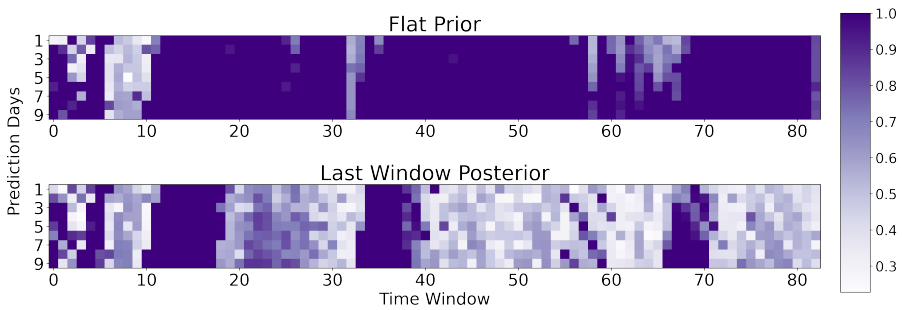


**Fig. S22** Normalized RMSD over the fitting window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



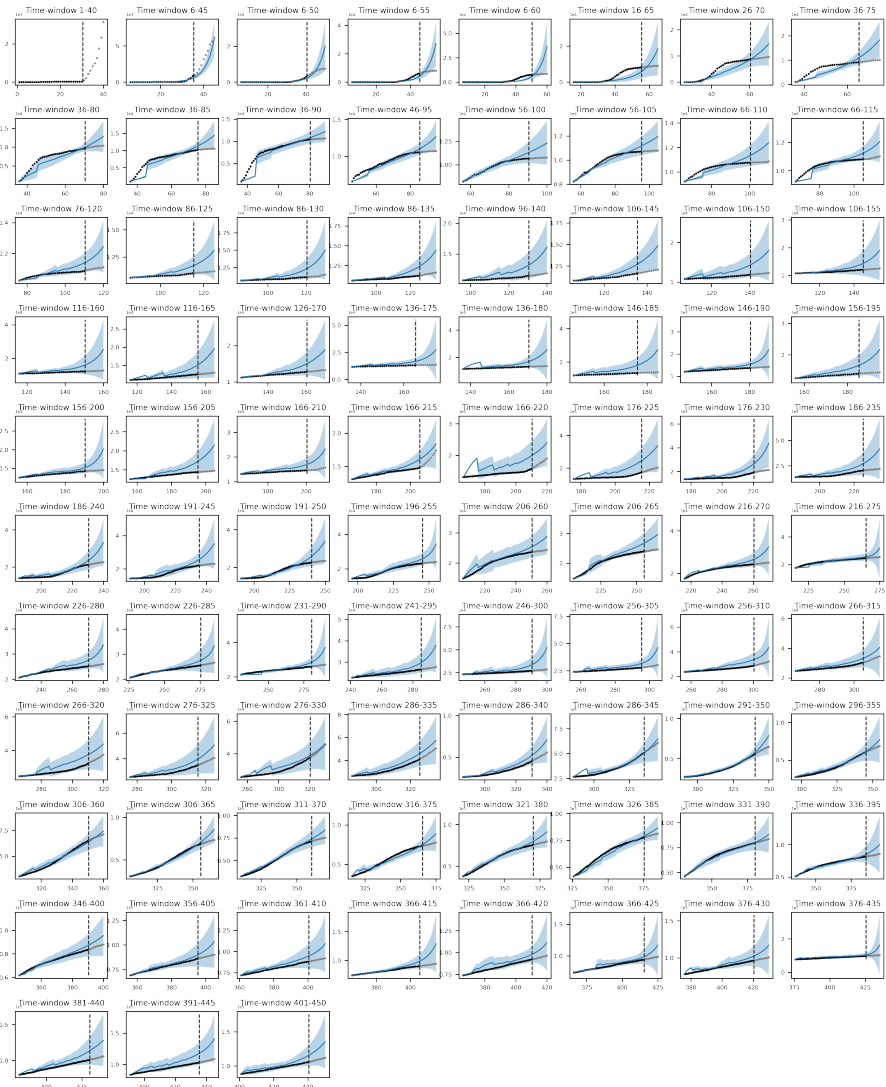


**Fig. S23** Normalized RMSD over the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



**Fig. S24** Normalized RMSD for every day of the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

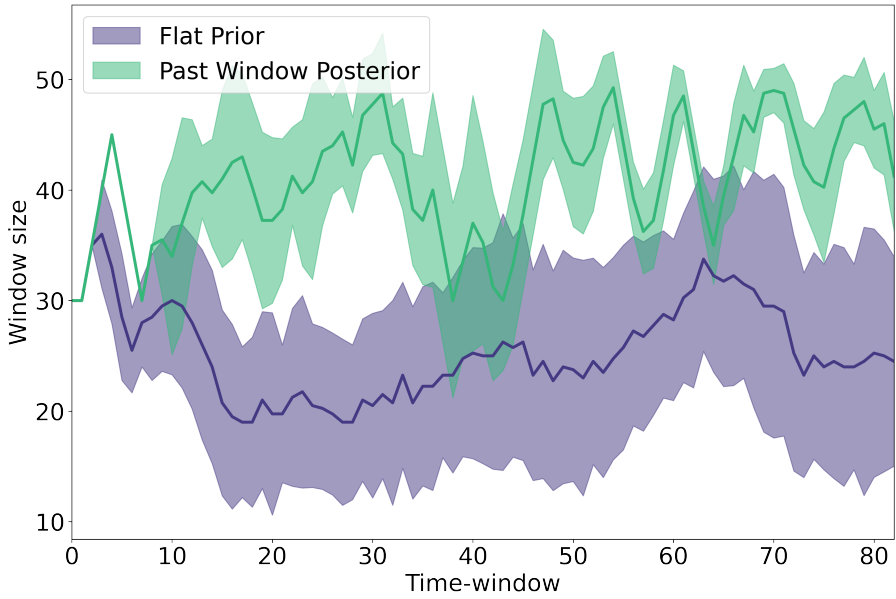
## S5.4 South Korea



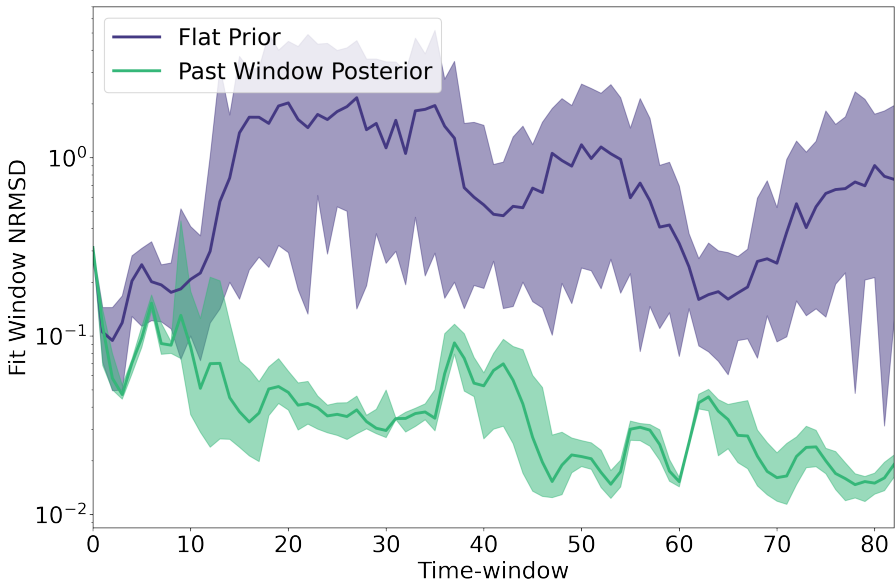
**Fig. S25** Fitting and prediction for each time window during the whole epidemic curve for South Korea using the flat prior approach.



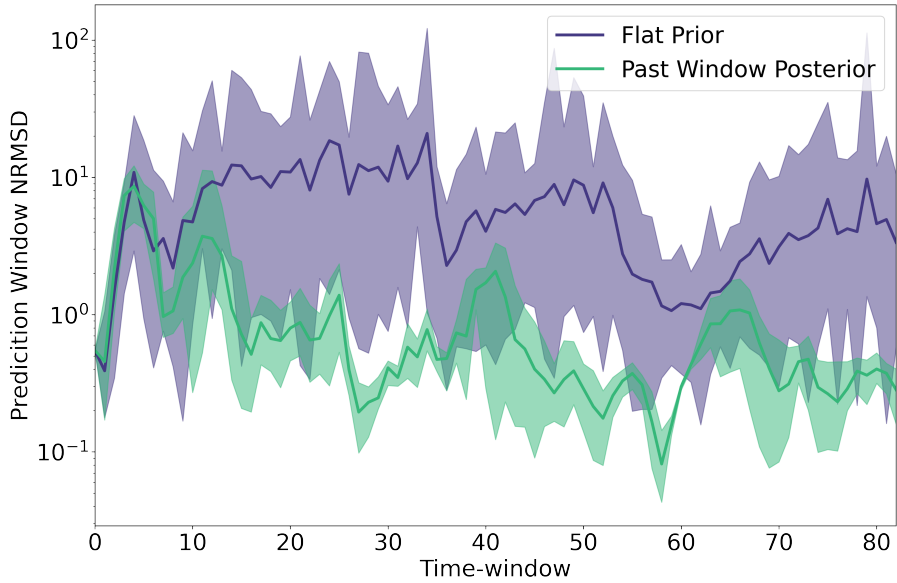
**Fig. S26** Fitting and prediction for each time window during the whole epidemic curve for South Korea using the past window approach.



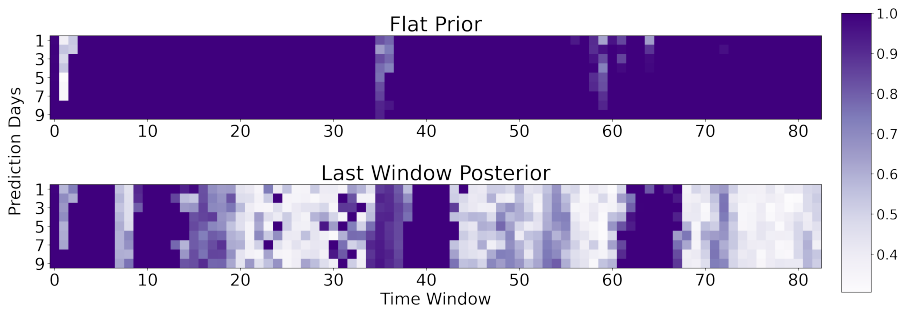
**Fig. S27** Windows' sizes selected over five executions, for the case past window posteriors.



**Fig. S28** Normalized RMSD over the fitting window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

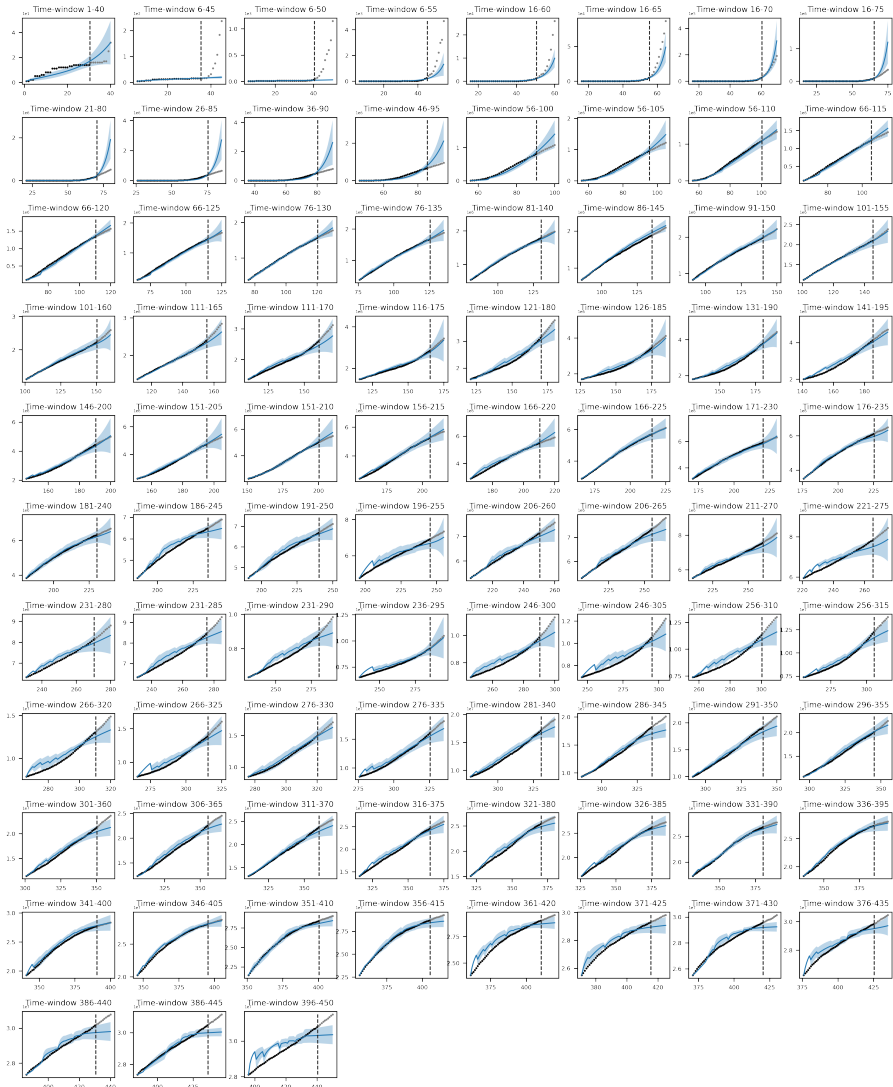


**Fig. S29** Normalized RMSD over the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



**Fig. S30** Normalized RMSD for every day of the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

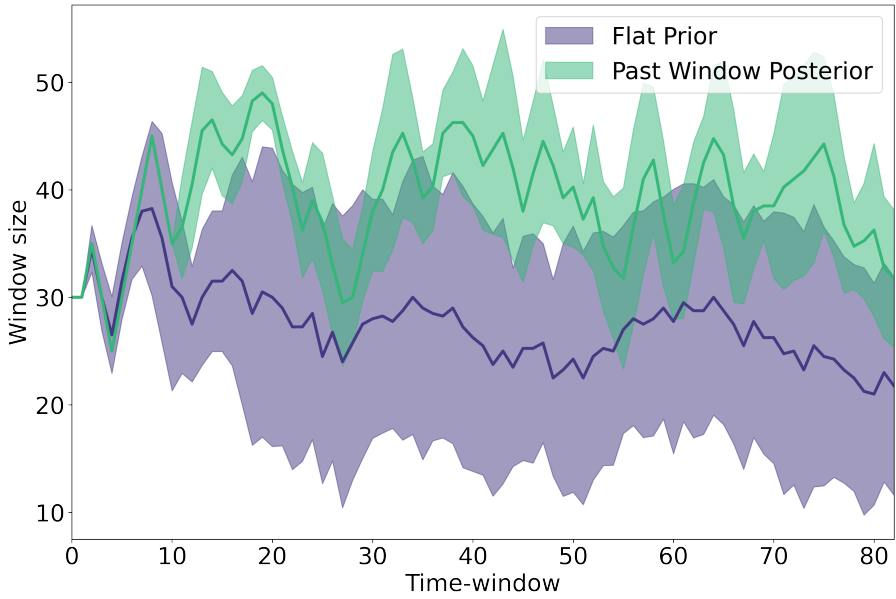
## S5.5 United States



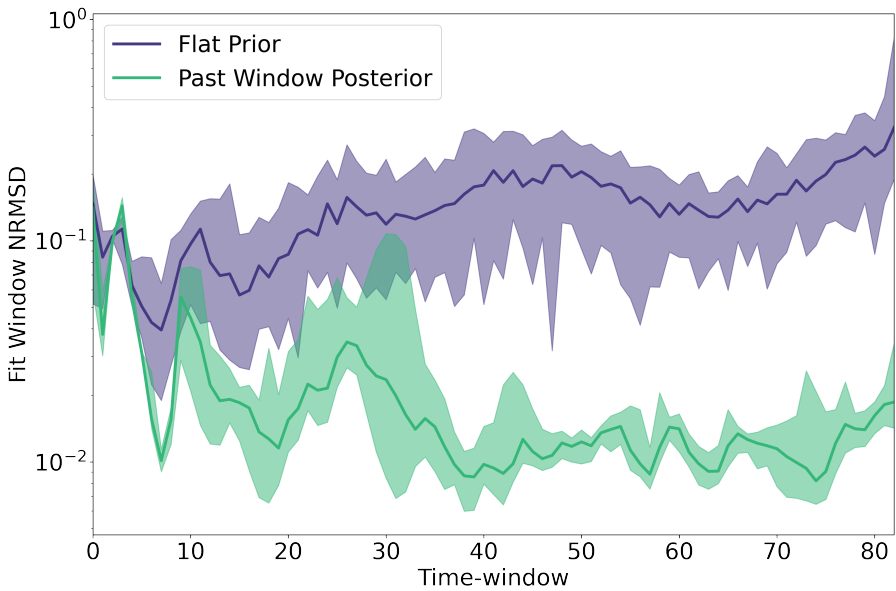
**Fig. S31** Fitting and prediction for each time window during the whole epidemic curve for United States using the flat prior approach.



**Fig. S32** Fitting and prediction for each time window during the whole epidemic curve for United States using the past window approach.

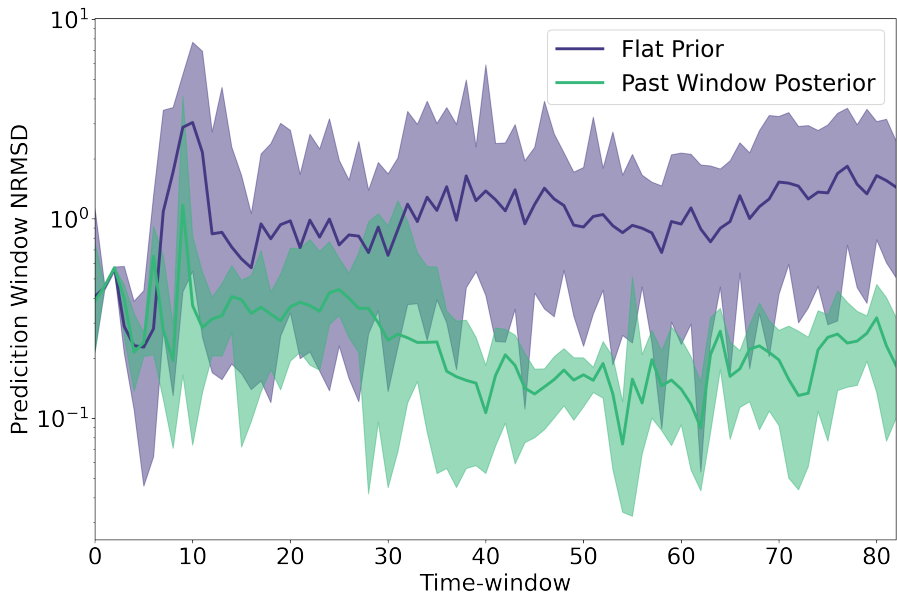


**Fig. S33** Windows' sizes selected over five executions, for the case past window posteriors.

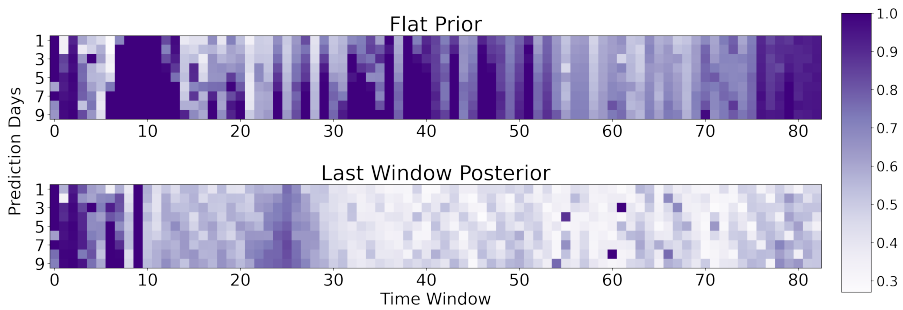


**Fig. S34** Normalized RMSD over the fitting window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



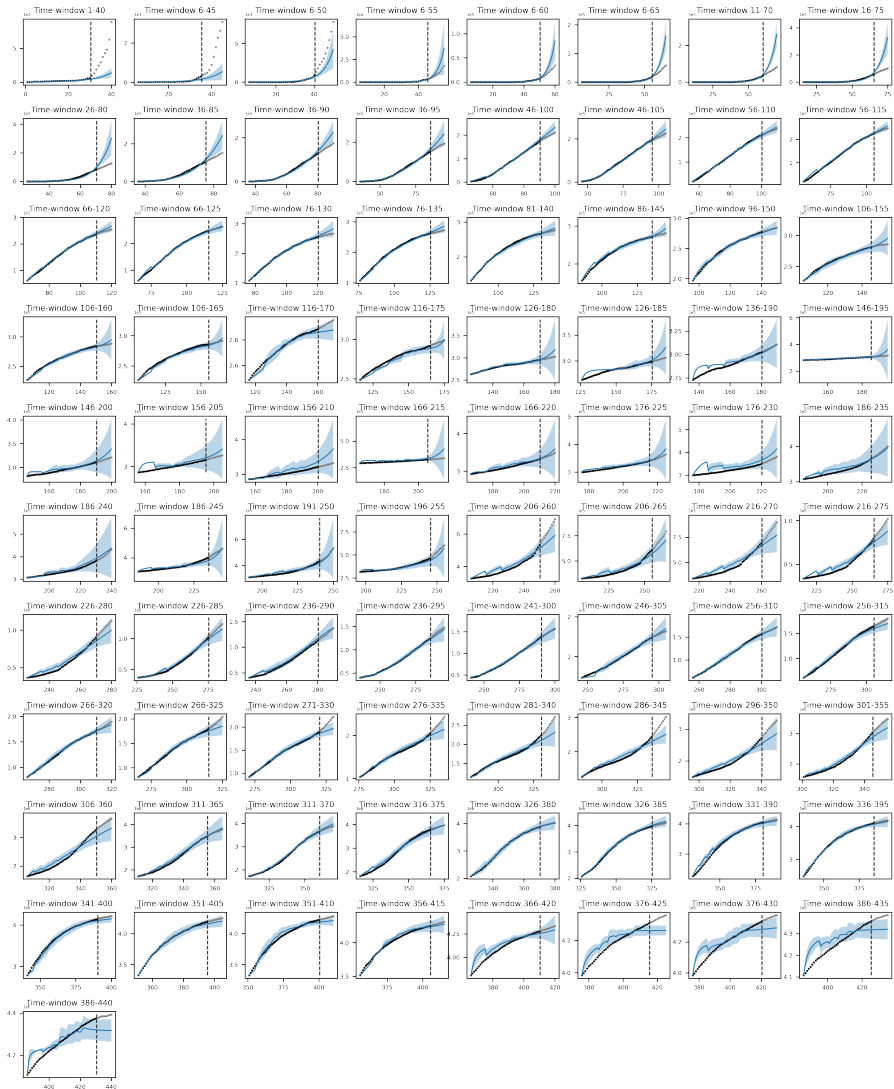


**Fig. S35** Normalized RMSD over the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



**Fig. S36** Normalized RMSD for every day of the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

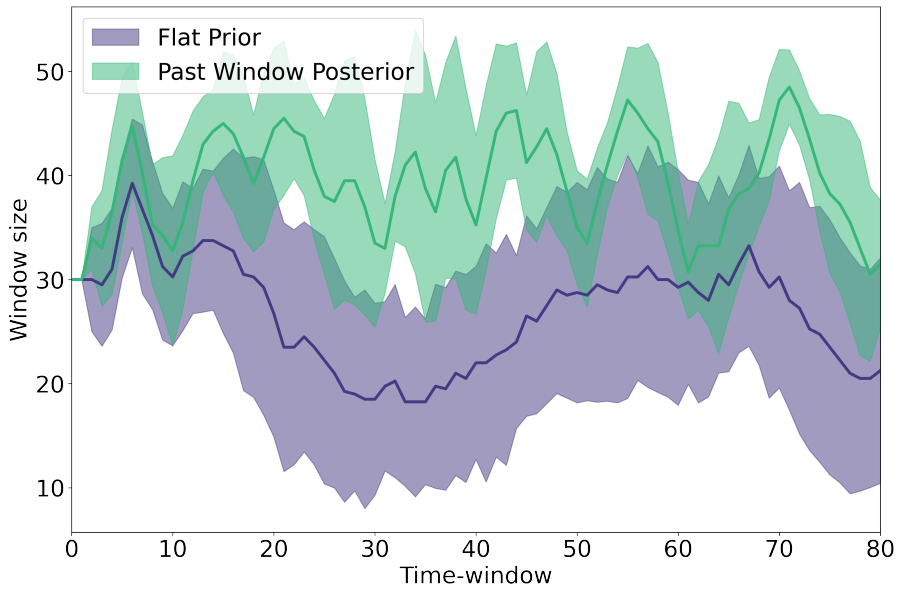
## S5.6 United Kingdom



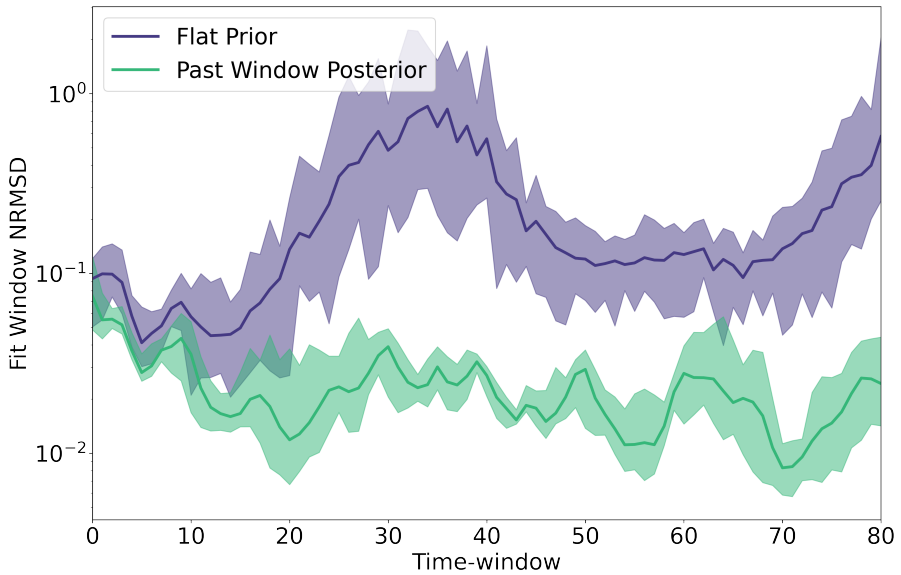
**Fig. S37** Fitting and prediction for each time window during the whole epidemic curve for United Kingdom using the flat prior approach.



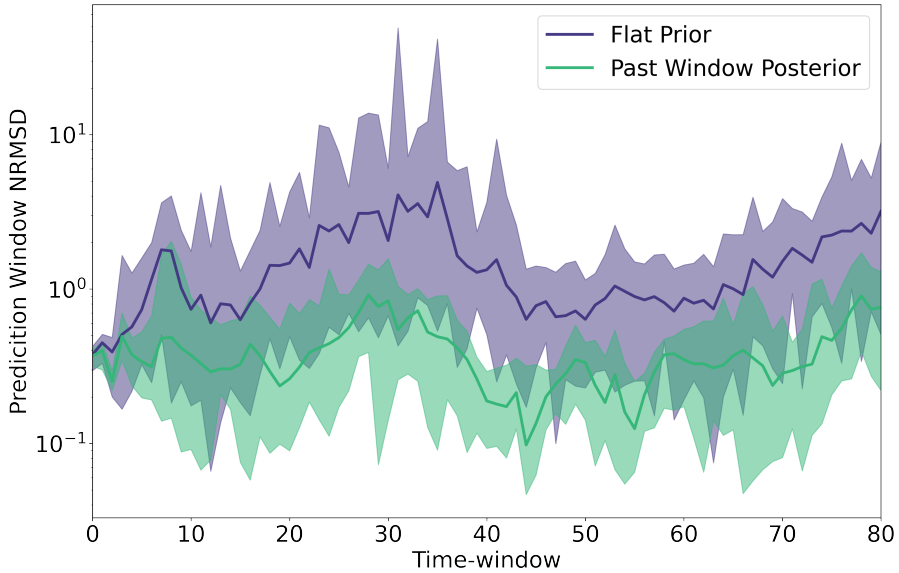
**Fig. S38** Fitting and prediction for each time window during the whole epidemic curve for United Kingdom using the past window approach.



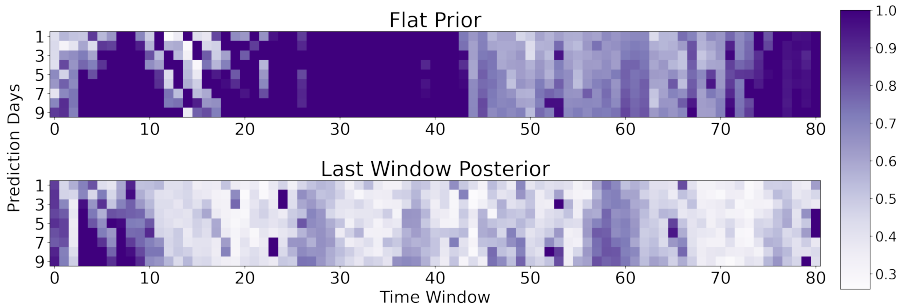
**Fig. S39** Windows' sizes selected over five executions, for the case past window posteriors.



**Fig. S40** Normalized RMSD over the fitting window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



**Fig. S41** Normalized RMSD over the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.



**Fig. S42** Normalized RMSD for every day of the prediction window, for both cases of using flat priors or past window posteriors, considering five executions for each case.

## References

- [1] Nunes, F.F., Miranda, J.P.V., Pinheiro, P.H., Cintra, I.R., de Lima, L.R., de Sousa Alves, T.L.: Analysis of informative priors’ effects on epidemic curve fitting. In: Proceedings of the XIV Academic Meeting on Computational Modeling - EAMC 2021, Petrópolis, RJ, Brazil, pp. 71–82 (2021). Available at: [http://www.eamc.lncc.br/PastEditions/Proceedings/Proceedings\\_EAMC2021.pdf](http://www.eamc.lncc.br/PastEditions/Proceedings/Proceedings_EAMC2021.pdf)