Bayesian SIR Models for COVID-19 in Texas

Jonathan Lindbloom
Advised by Dr. Alejandro Aceves

World Changers Shaped Here

SMU

1. Introduction

One major challenge in constructing a reliable model for observed COVID-19 cases, hospitalizations, and fatalities data is to adequately capture the observed time-varying transmission dynamics summarized by the effective reproduction number $\mathcal{R}_t$. Common approaches to address this modeling component include relating $\mathcal{R}_t$ to its covariates such as mobility and degree of social distancing, estimating a sequence of change-points for $\mathcal{R}_t$, and explicitly modeling the dynamics of the $\mathcal{R}_t$ process. An equally-important challenge is to express the inherent uncertainty in conclusions drawn from such a model. In this work, we detail the construction of an Bayesian SIR-like model capable of tackling both of these challenges, specifically in regards to the available Texas data.

2. Compartmental Models

To model the spread of COVID-19 we employ an SIR-like model, a mechanistic model which describes the flow of individuals between compartments representing susceptible (S), infectious (I), and recovered (R) individuals. A simplified form of our model is given by

$$\frac{dS}{dt} = -\frac{\beta(t)}{N}SI, \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta(t)}{N}SI - \mu I, \tag{2}$$

$$\frac{dR}{dt} = \mu I, \tag{3}$$

where $\beta(t)$ is the time-dependent transmission rate, μ is the recovery rate, and N is the population size. For this model we have $\mathcal{R}_t = \frac{\beta(t)}{N}$; to prescribe the function $\beta(t)$, we model $\mathcal{R}_t$ with a Gaussian process and set $\beta(t) = \mu\mathcal{R}_t$. In order to fit the model parameters to data, we discretize this system of ODEs with $\Delta t = 1$ day and compare the daily new cases $C_{\text{new}}(t) = \frac{\beta(t)}{N}SI$ with that reported in the JHU CSSE COVID-19 Data [1]. Our comparison (via a likelihood function) is identical to that used by [2], accounting for reporting delay, weekend reporting modulation, and a count-dependent reporting noise assumed to follow a t-distribution.

3. Bayesian Inference

The Bayesian approach to parameter inference from data is a powerful method for quantifying uncertainty in model predictions. Let $\mathcal{M}_\Theta = \{M_\theta : \theta \in \Theta\}$ denote a parametric family of models, where θ describes a vector of possible model parameters. We use Bayes' Theorem to write

$$p(\theta \,|\, \mathcal{D}) = \frac{p(\mathcal{D} \,|\, \theta)p(\theta)}{\int_\Theta p(\mathcal{D} \,|\, \theta)\, d\theta} \tag{4}$$

where $p(\theta)$ is a prior distribution over model parameters and $p(\theta \,|\, \mathcal{D})$ is the corresponding posterior distribution conditioned on the observed data. The function $p(\mathcal{D} \,|\, \theta)$, known as the likelihood function, encodes the relative probability of the data given any set of model parameters. In general we cannot evaluate the posterior density function directly and must resort to methods such as Markov Chain Monte Carlo (MCMC) to compute samples of the random variable $X \sim p(\theta \,|\, \mathcal{D})$. These samples can then be transformed into samples from any general quantity of interest $Y = f(X)$, e.g., $\mathcal{R}_t$ or cumulative fatalities by time $t = T$.

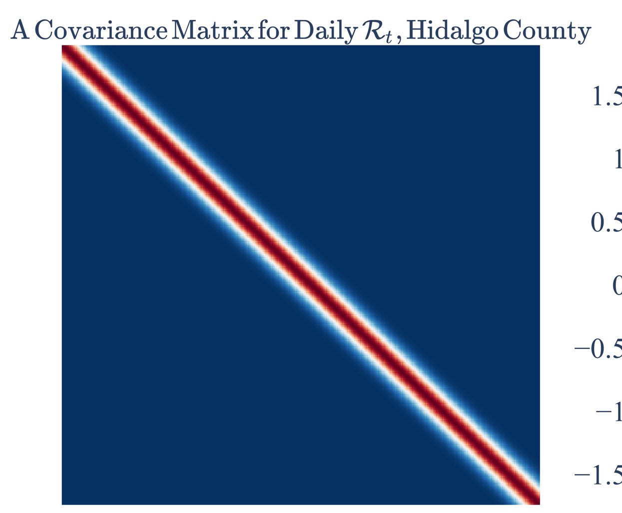
4. Gaussian Processes

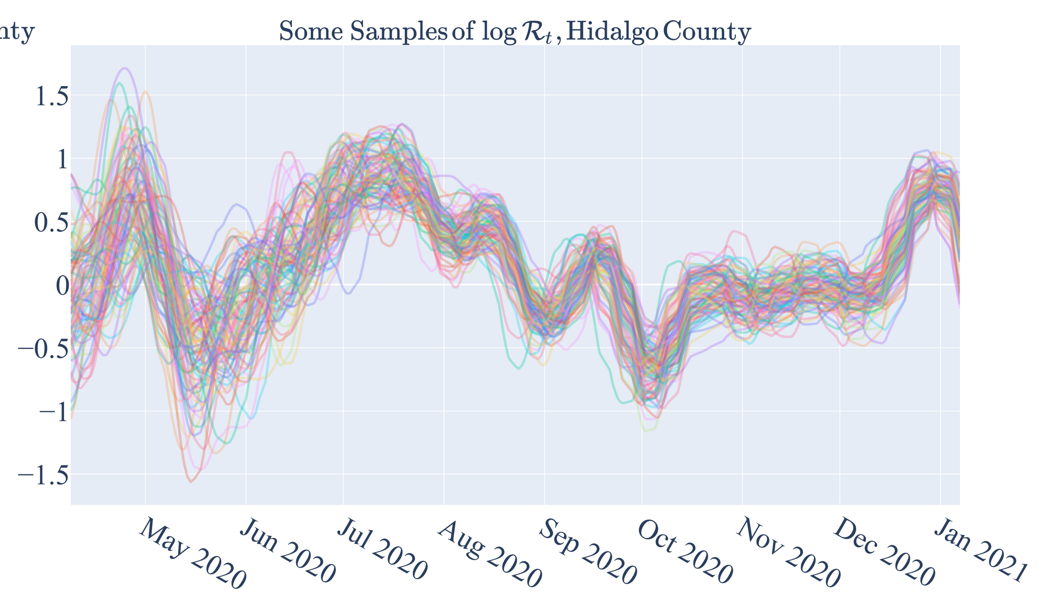
A Gaussian process is a stochastic process for which every finite-dimensional distribution is multivariate normal. Gaussian processes are specified by assigning a mean function $m(t)$ and a covariance kernel $k(t, t')$ which prescribes the covariance matrix of any finite-dimensional distribution. In this work, we use Gaussian processes to define prior distributions on the $\mathcal{R}_t$ process where

$$\log \mathcal{R}_t \sim \mathcal{GP}(0, k), \tag{5}$$

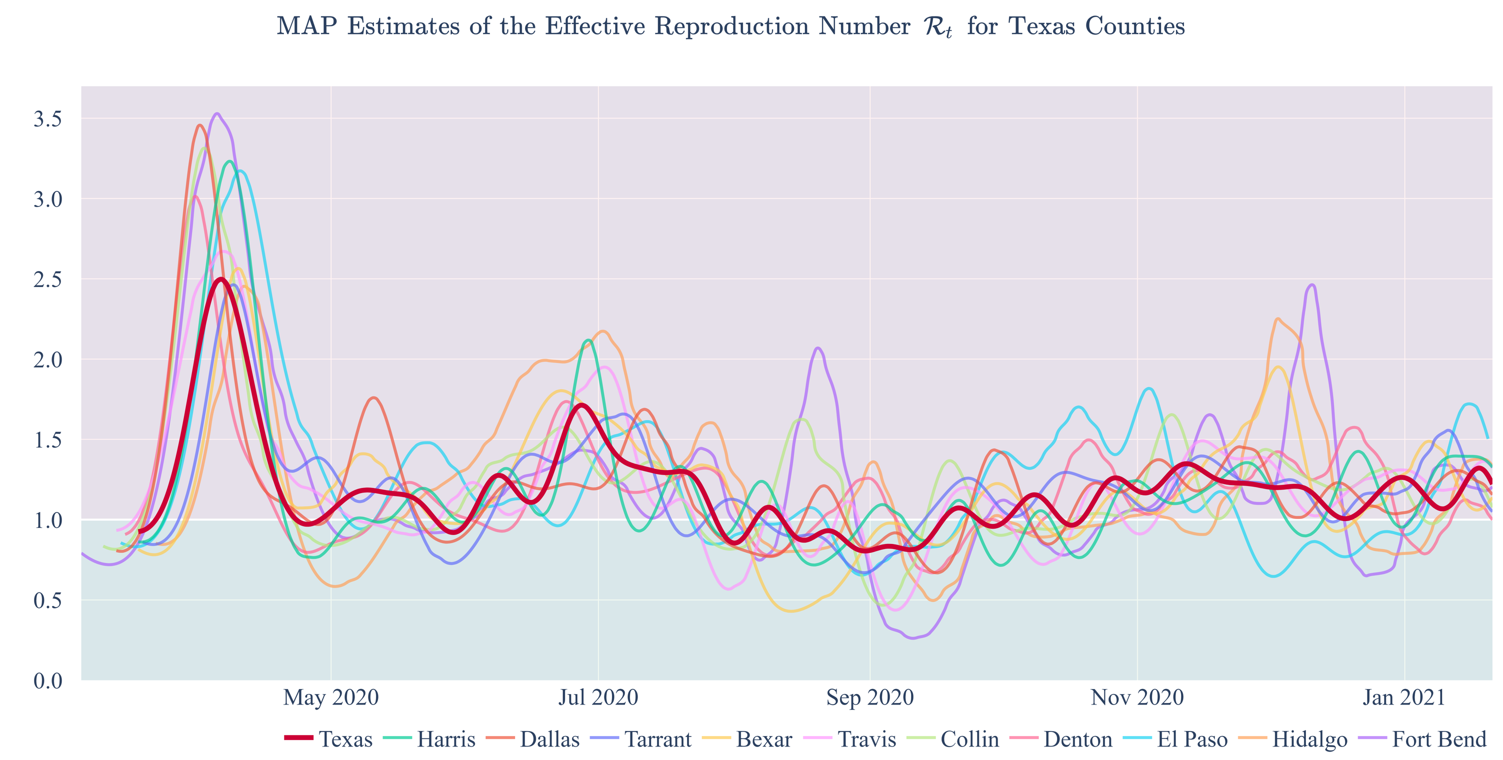
$$k(t, t') = \eta^2 \exp\left[-\frac{(t - t')^2}{2l^2}\right]. \tag{6}$$

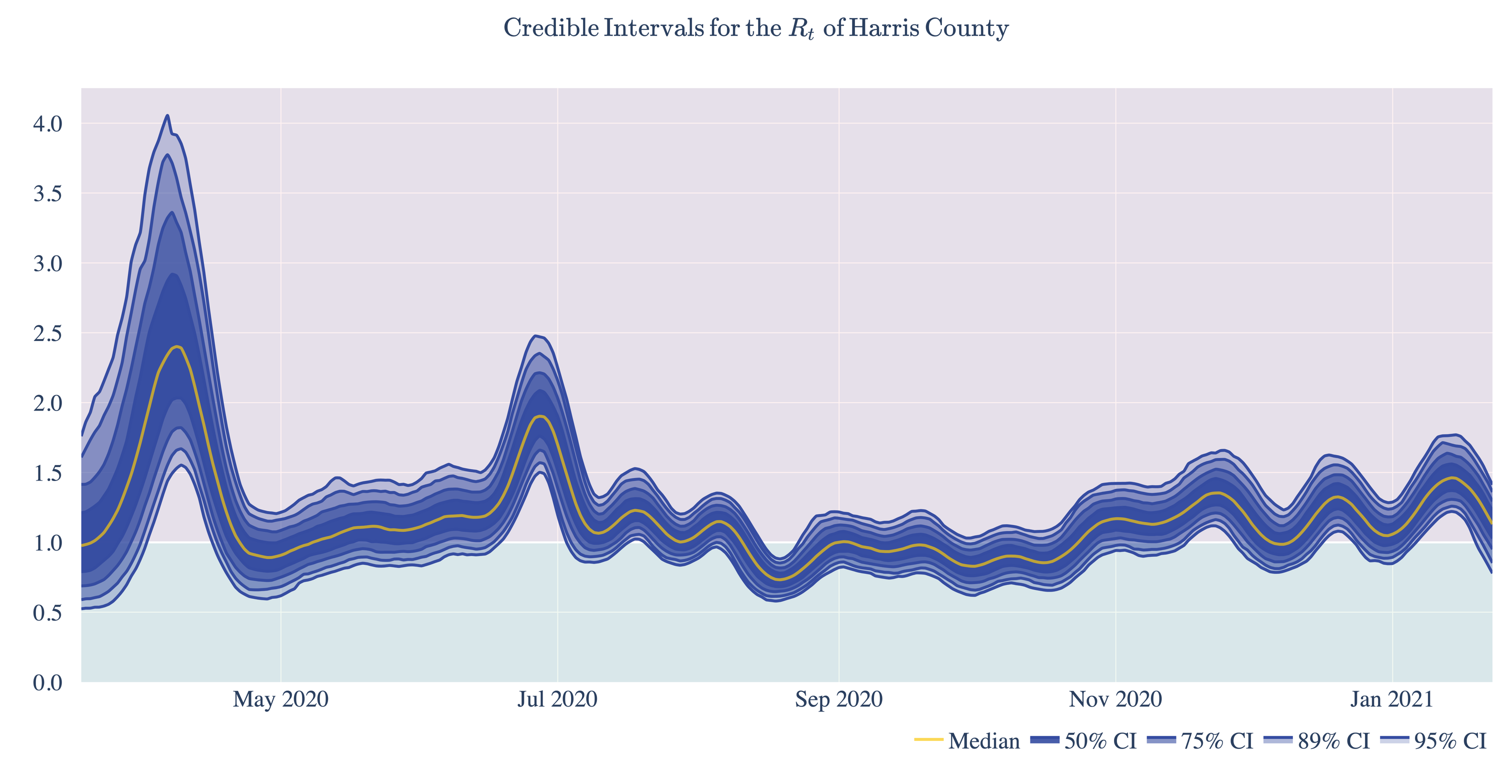
Here the parameter η describes the magnitude of $\mathcal{R}_t$'s deviance away from 1.0, and l controls the characteristic length scale over which $\mathcal{R}_t$ varies. We set $l = 1$ week to capture weekly changes in $\mathcal{R}_t$, and allow η to be inferred during inference.

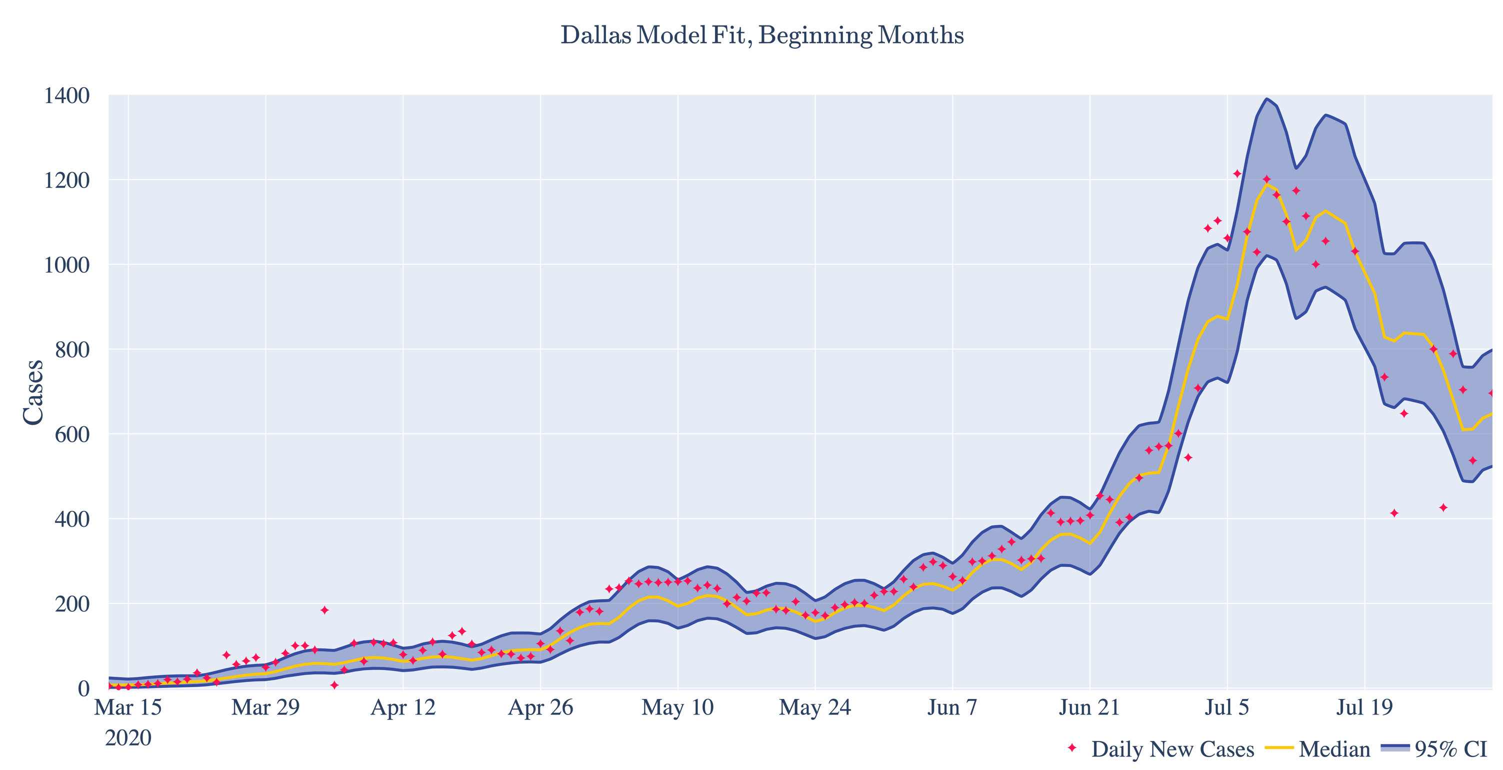


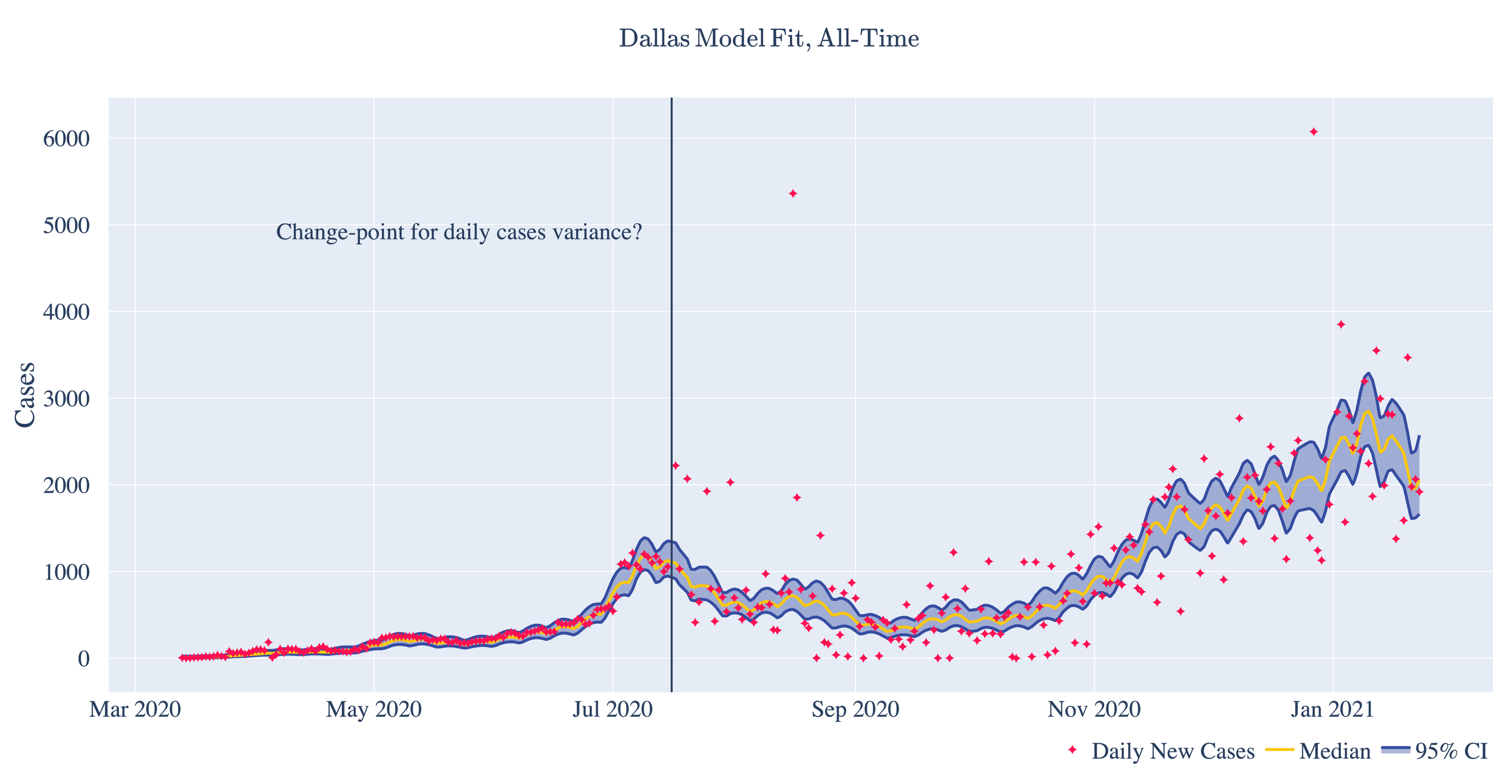


5. Model Fit

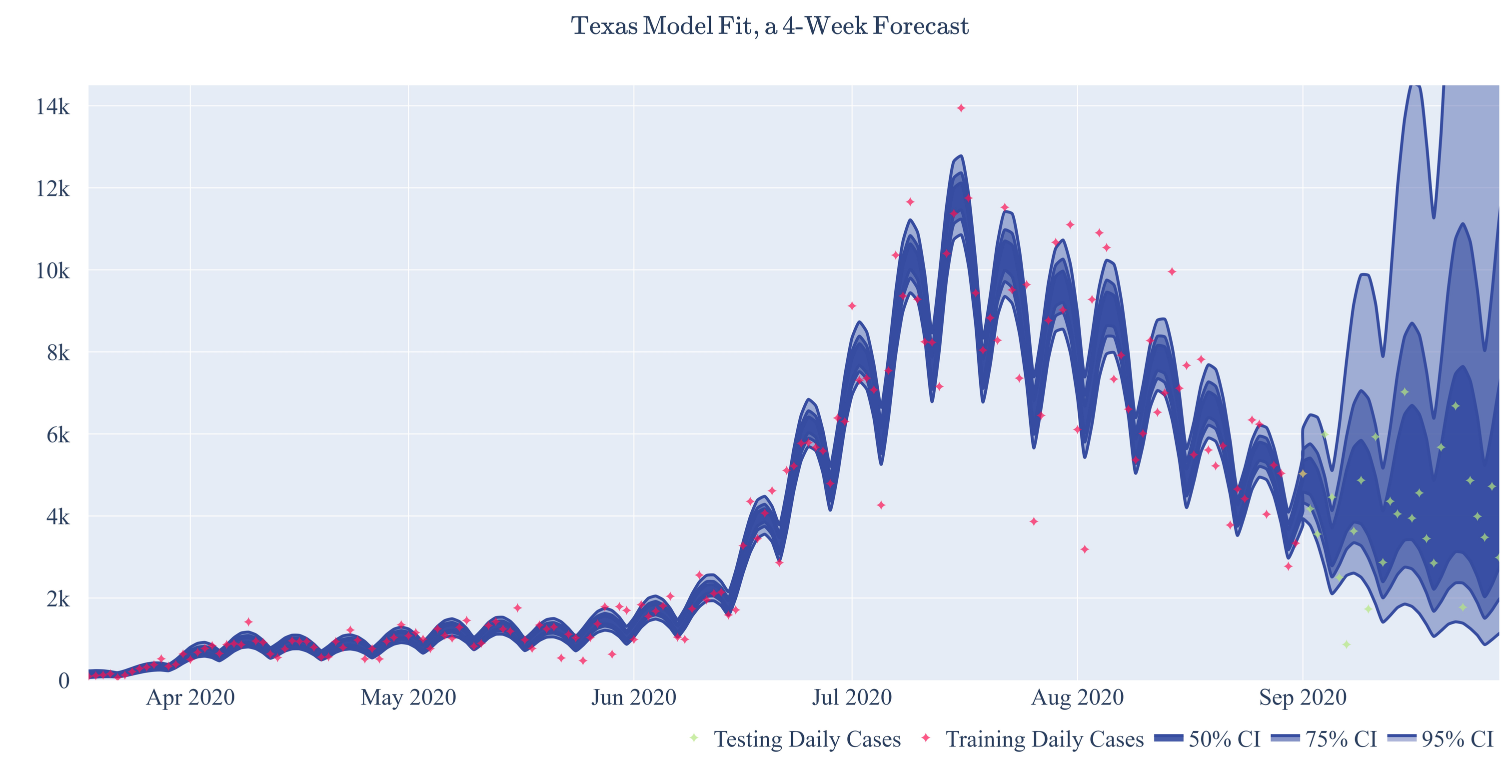








6. Model Forecasting



7. Discussion

Model Strengths:

- There is no need to specify priors over change-points, which can be cumbersome. Using a Gaussian process prior makes the model very flexible and thus capable of tracking changes in $\mathcal{R}_t$.
- The Bayesian approach allows us to directly quantify model uncertainty arising from both observational noise and parametric uncertainty.
- Although we are allowing $\mathcal{R}_t$ to vary daily, the use of a Gaussian process prior with a fixed length scale helps to enforce smooth changes in $\mathcal{R}_t$ and to avoid over-fitting.
- The model requires no modifications for the idiosyncrasies of the data of individual counties; the same model has been reliably fit to each of the 10 largest counties in Texas.

Future Work and Improvement:

- For some counties such as Dallas, it is evident that the model would benefit from the inclusion of a time-varying (and not simply count-dependent) observational noise variance. Additionally, the model would likely become easier to fit were we to first approximate the noise variance and keep it fixed during sampling.
- In the case that we believe that the dynamics of the $\mathcal{R}_t$ process have changed over time, it is fairly straightforward to construct a modified covariance function to function this (e.g., change-point kernel, warped kernel, etc.).
- Our model currently assumes a mean function for the Gaussian process priors that is identically 0, implying a regression of $\mathcal{R}_t$ to a mean of 1.0. This may be unwarranted, and it is worth exploring alternatives to better calibrate forecasts.
- We would like to extend the usefulness of this model to also fit the data for fatalities, hospitalizations, and vaccinations.
- To assess forecasting quality, we plan to compare our model's performance with the performance of the models submitted to the CDC according to scoring rules.

Acknowledgements

This model was inspired by the U.S. Army Engineer Research and Development Center (ERDC) SEIR model [3]. Experimental code is available here, which is adapted from the Priesemann Group's change-point model available here. This research was supported by the SMU Office of Engaged Learning and NSF-RTG grant DMS-1840260.

References

1. Dong E, Du H, Gardner L. (2020). *An interactive web-based dashboard to track COVID-19 in real time.* Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1.
2. Dehning, J., Zierenberg, J., Spitzner, F., Wibral, M., Neto, J., Wilczek, M., Priesemann, V. (2020). *Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions.* Science, Vol. 369, no. 6500, p. eabb9789. doi: 10.1126/science.abb9789.
3. Parno, Matthew. *The ERDC-SEIR Model for COVID-19.* (2020). Presentation for Dartmouth Math 76.