Development of a Spatially Regularized Detector for Emergent/Re-emergent Disease Outbreaks

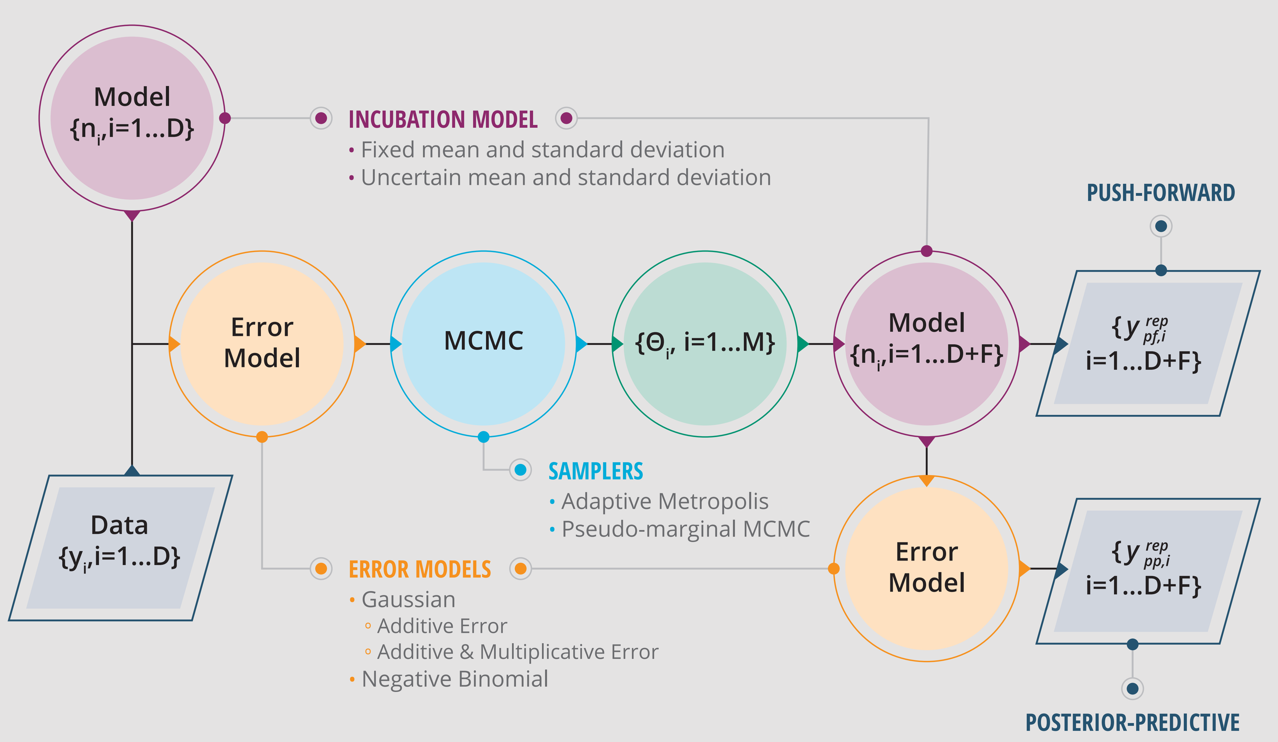
Cosmin Safta (csafta@sandia.gov) & Jaideep Ray (jairay@sandia.gov)
Data Sciences & Computing, Sandia National Laboratories, Livermore, CA 94550

Summary

We propose a Bayesian framework to detect outbreaks of emerging diseases. Our method triggers when the observed case-count data disagree significantly from forecasted levels. In contrast, forecasting allows us to use socioeconomic parameters and spatiotemporal data on disease prevalence to compensate for low-quality epidemiological information. The proposed framework assumes correlations between adjacent regions to account for spatio-temporal correlations in the disease spread rate. Posterior distributions are approximated via Markov-Chain Monte Carlo. Results computed with COVID-19 data from New Mexico, US, are used to demonstrate the method.

Epidemiological Model Inference Workflow

$$p(\Theta|\mathcal{D}) \propto p(\mathcal{D}|\Theta) \times p(\Theta)$$



Epidemiological Model

The epidemiological model consists of two components: an infection rate model and an incubation rate model. These models are combined through a convolution into a forecast of number of cases that turn symptomatic daily.

Infection Rate:

$$f_\Gamma(t;k,\theta,t_0) = \theta^{-k}(t-t_0)^{k-1}\exp(-(t-t_0)/\theta)/\Gamma(k)$$

Incubation Model:

$$F_{LN}(t;\mu,\sigma) = \frac{1}{2}\mathrm{erfc}\left(-\frac{\log t - \mu}{\sigma\sqrt{2}}\right)$$

Daily Symptomatic Cases for day *i*:

$$n_i = \int_{t_0}^{t_i}\left(\sum_{j=1}^{K} N_j f_\Gamma(\tau - t_0 - \Delta t_j; k_j, \theta_j)\right) \times \left(F_{LN}(t_i-\tau;\mu,\sigma) - F_{LN}(t_{i-1}-\tau;\mu,\sigma)\right)d\tau$$

Likelihood Model

Focus on spatial correlations:

$$\text{Model}: Y_i^{(p)} = \underbrace{\{n_{i,1}, n_{i,2}, \ldots\}}_{\text{No. of regions}}$$

$$\mathcal{L}_{\mathcal{D}} = \prod_{i=1}^{N_d} \frac{1}{(2\pi)^{N_r/2}\det\Sigma_i^{1/2}}\exp\left(-\frac{1}{2}(Y_i^{(o)} - Y_i^{(p)})\Sigma_i^{-1}(Y_i^{(o)} - Y_i^{(p)})^T\right)$$

Account for proximity between regions and employ an additive-multiplicative error model:

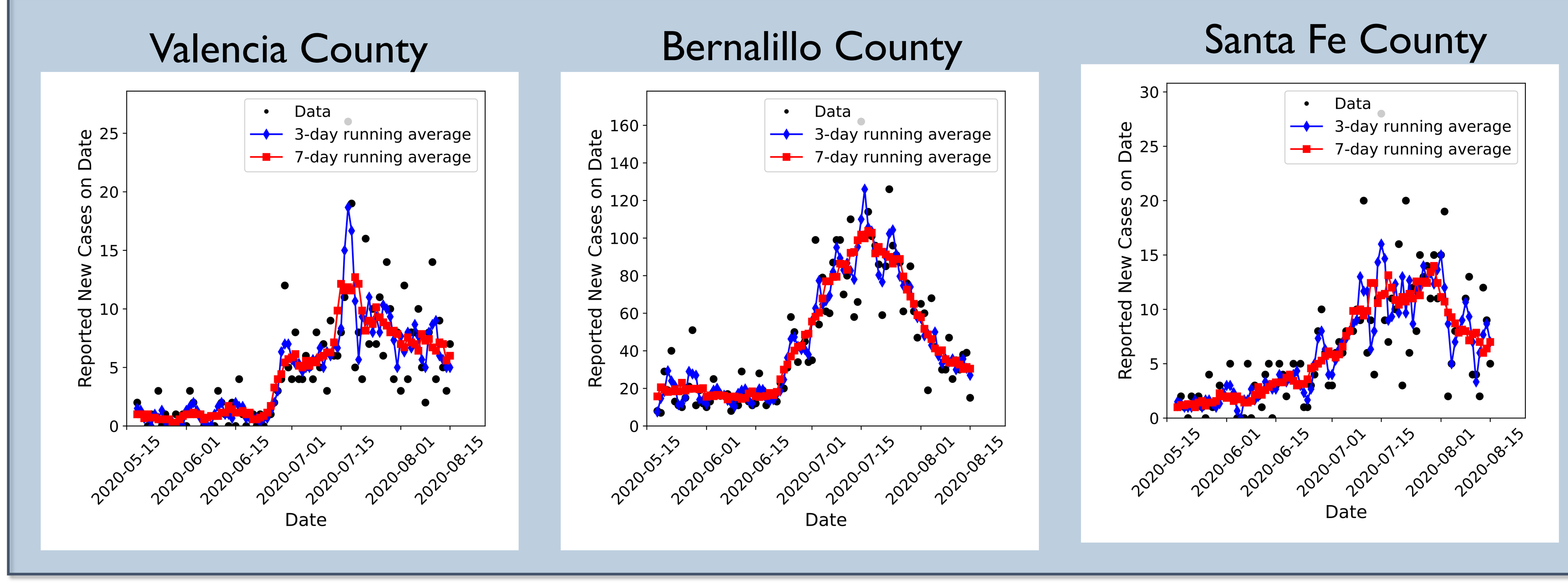
$$\Sigma_i = \frac{\tau_\Phi^2}{1-\lambda_\Phi^2}\left[I + \lambda_\Phi W\right] + \mathrm{diag}\left(\sigma_a + \sigma_m Y_i^{(p)}\right)^2$$

Spatial correlations:

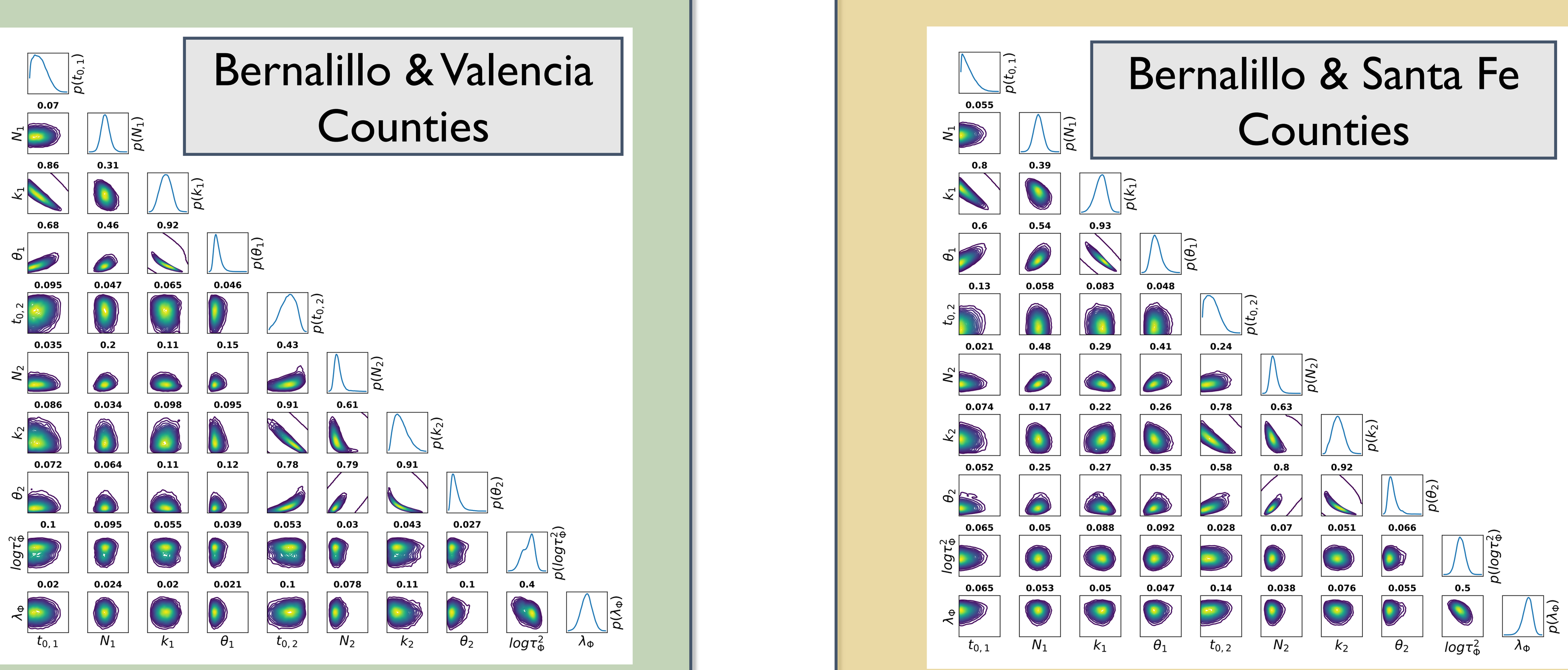
$$W: w_{kk} = 0, \quad w_{kl} \in [0,1]$$

Off-diagonal entries account for mixing between adjacent regions (counties)

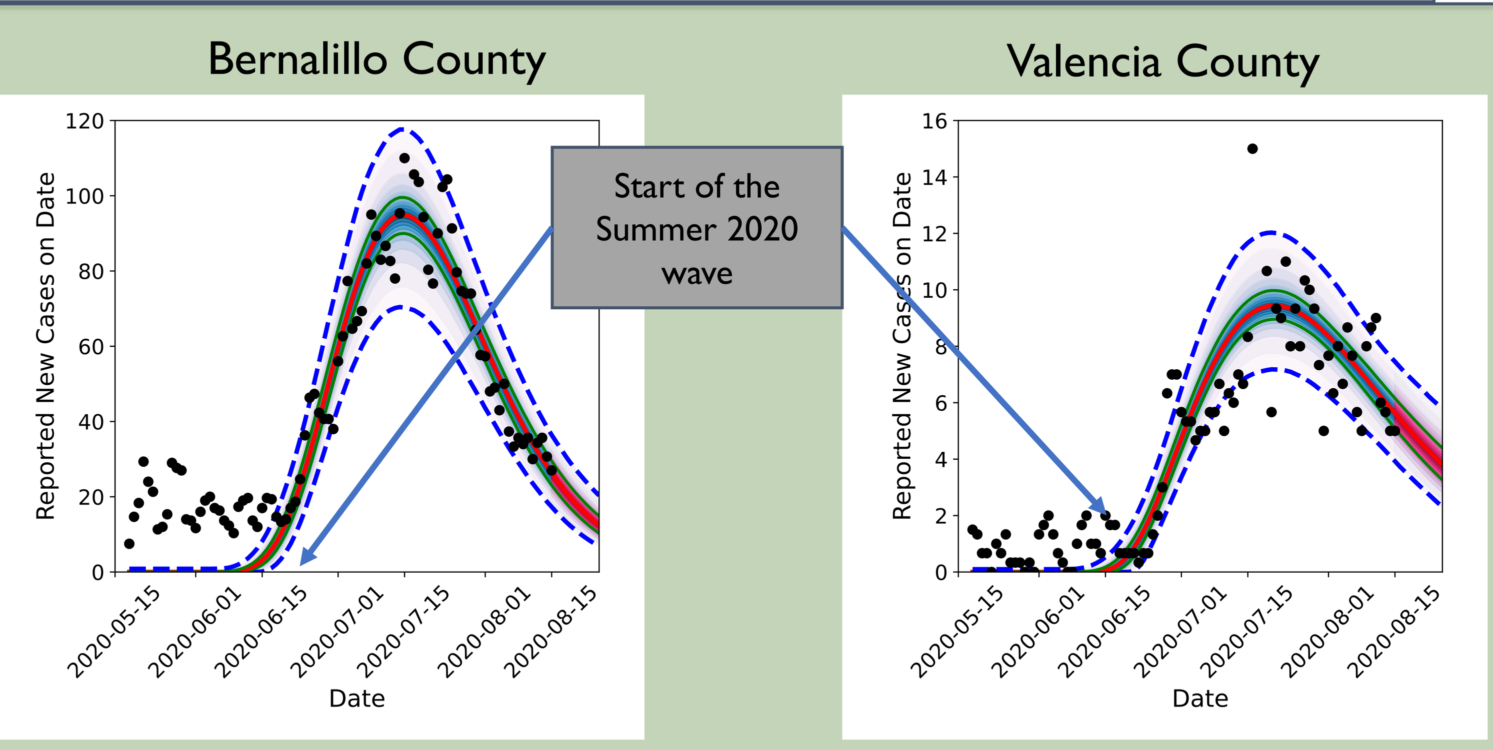
Data $\mathcal{D}$



Marginal Posterior Densities

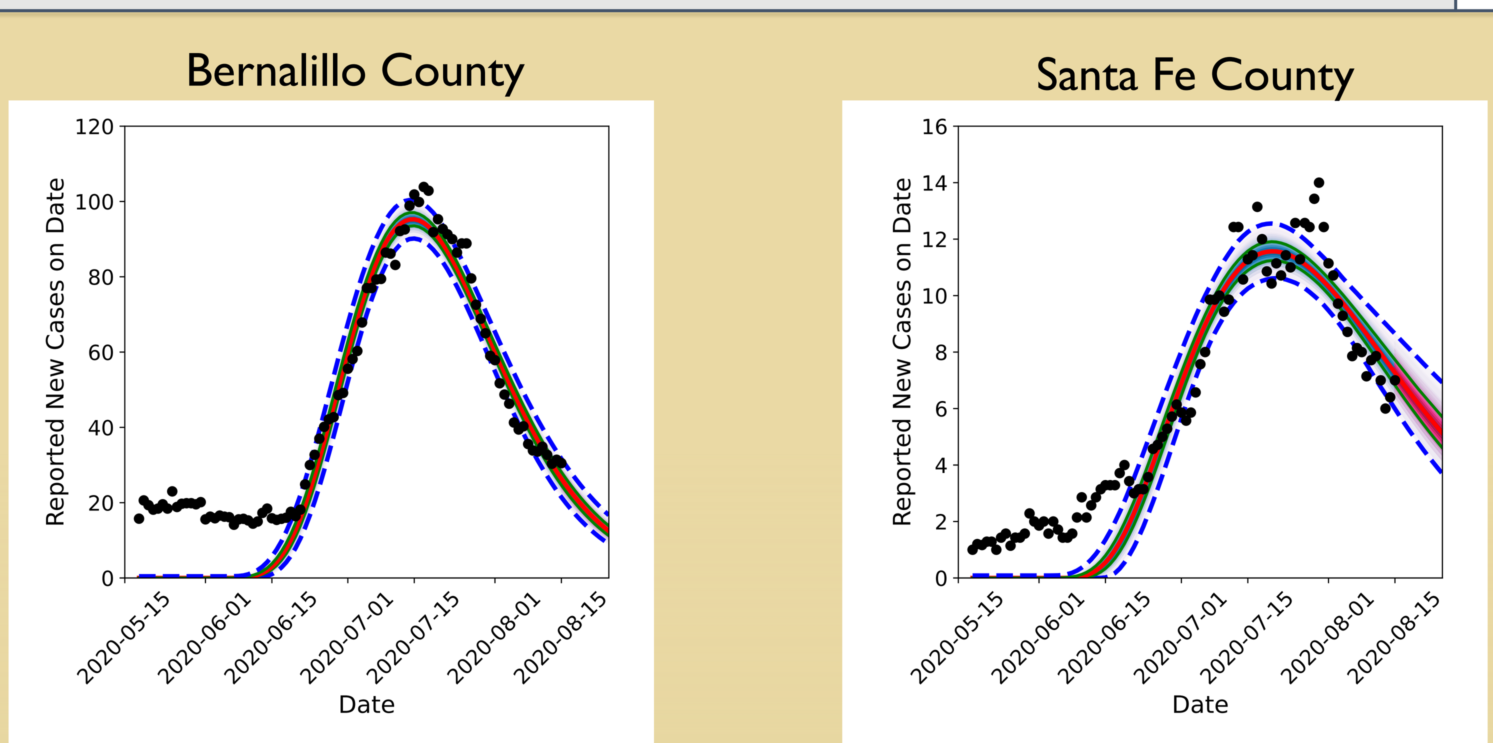


Posterior Predictive Distributions: Bernalillo & Valencia Counties



- 3-day averaged counts

Posterior Predictive Distributions: Bernalillo & Santa Fe Counties



- 7-day averaged counts

References:
- P. Blonigan, J. Ray, C. Safta (2021). Forecasting Multi-Wave Epidemics Through Bayesian Inference. *Archives of Computational Methods in Engineering*, doi:10.1007/s11831-021-09603-9
- C. Safta, J. Ray, K. Sargsyan (2020). Characterization of partially observed epidemics through Bayesian inference: Application to COVID-19. *Computational Mechanics*, doi:10.1007/s00466-020-01897-z

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. 2022-8401C