Early Detection of COVID-19 Hotspots Using Spatio-Temporal Data

Shixiang Zhu [1] **Alexander Bukharin** [1] **Liyan Xie** [1] **Shihao Yang** [1] **Pinar Keskinocak** [1] **Yao Xie** [1]

Abstract

Recently, the Centers for Disease Control and Prevention (CDC) has worked with other federal agencies to identify counties with increasing coronavirus disease 2019 (COVID-19) incidence (hotspots) and offers support to local health departments to limit the spread of the disease. Understanding the spatio-temporal dynamics of hotspot events is of great importance to support policy decisions and prevent large-scale outbreaks. This paper presents a spatio-temporal Bayesian framework for early detection of COVID-19 hotspots (at the county level) in the United States. We assume both the observed number of cases and hotspots depend on a class of latent random variables, which encode the underlying spatio-temporal dynamics of the transmission of COVID-19. Such latent variables follow a zero-mean Gaussian process, whose covariance is specified by a non-stationary kernel function. The most salient feature of our kernel function is that deep neural networks are introduced to enhance the model's representative power while enjoying great interpretability. Our model demonstrates superior hotspot-detection performance compared to other baseline methods.

1. Introduction

The ongoing global pandemic caused by the coronavirus disease (COVID-19) has spread rapidly over more than 200 countries in the world since its emergence in 2019. Even the largest economies' resources have been strained due to the spread of COVID-19. Predicting potential hotspots ahead of time can play a significant role in deploying targeted interventions, such as testing, tracing, and isolation, and slow down the disease spread (Oster et al., 2020b).

Large-scale, population-based testing can indicate regional hotspots, but at the cost of a delay between testing and actionable results. Accurately identifying changes in the infection rate requires sufficient testing coverage of a given population, which can be costly and requires substantial testing capacity. Regional variation in testing access can also hamper the ability of public health organizations to detect rapid changes in infection rates. Recent studies (Centers for Disease Control and Prevention, 2021) aimed at estimating the spread of COVID-19 by forecasting the number of confirmed cases or the number of deaths. However, these methods failed to provide a satisfactory case prediction accuracy. Therefore, there is a high unmet need for tools and methods that can facilitate the timely and accurate identification of infection hotspots and enable policymakers to act effectively with minimal delay (Rossman et al., 2020).

The Centers for Disease Control and Prevention (CDC) has worked with other federal agencies to identify counties with a significant increase in COVID-19 incidence (hotspots) (Oster et al., 2020b), which offers a unique opportunity to investigate the spatio-temporal dynamics between the identified hotspots. The identified hotspots indicate the relative temporal increases in confirmed cases and mark the onset of local outbreaks.

In this paper, we propose an effective COVID-19 hotspot detection framework that utilizes the hotspot data and multiple other data sources to enhance hotspot detection accuracy. We assume the hotspot and number of cases in the same location depending on common priori factors, represented by a latent spatio-temporal random variable. This latent variable is modeled by a Gaussian process, whose covariance is characterized by an interpretable non-stationary kernel. We note that the non-stationarity of our kernel plays a pivotal role in the success of our model because the spread of the virus shows heterogeneous spatial correlation across different regions. For example, the virus is likely to spread more slowly in a sparsely populated area such as rural Nebraska compared to a densely populated area such as New York City. We formulate our kernel function using carefully crafted feature functions incorporating neural networks, which provide greater flexibility in capturing the complex dynamics of the spread of COVID-19 while still being highly interpretable. To tackle the intractability of Gaussian process with a large-scale data set, we also derive a sparse model and fit the model efficiently via a variational learning strategy, which goes beyond the scope of this paper and will not be discussed.

*Equal contribution [1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, GA, USA. Correspondence to: Shixiang Zhu <shixiang.zhu@gatech.edu>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

2. COVID-19 Hotspot Detection Framework

This section presents our hotspot detection framework, consisting of two spatio-temporal models: confirmed cases and hotspots. Consider weeks $\mathcal{T} = \{t = 1, \ldots, T\}$ starting from March 15, 2020 to January 17, 2021 and locations (counties) $\mathcal{I} = \{i = 1, \ldots, I\}$, with latitude and longitude $s_i \in \mathcal{S} \subset \mathbb{R}^2$, $i \in \mathcal{I}$, where $\mathcal{S}$ represents the space of geographic coordinate system (GCS). The two models, respectively, focus on weekly confirmed cases $y_{it} \in \mathbb{Z}_+$ and identified hotspots $h_{it} \in \{0, 1\}$ of COVID-19 at location $i \in \mathcal{I}$ and time $t \in \mathcal{T}$, where $h_{it} = 1$ if there is a hotspot at location i and time t, and 0, otherwise.

CDC (Oster et al., 2020b) defined the hotspots based on relative temporal increases in the number of cases, i.e., the occurrence of the hotspots depends on the spatio-temporal correlation across different locations and over time, and *not* on the mean number of cases. Hence, we capture the correlation between y_{it} and h_{it} by connecting these two models in the spatio-temporal space (t, s_i) through a latent spatio-temporal random variable $f(t, s)$, characterized by a Gaussian process (GP) with zero mean and covariance specified by a kernel function k.

The goal is to find the optimal pair of these two models that best predict the hotspots and the cases for one week ahead. We refer to the proposed framework as the spatio-temporal Gaussian process (`STGP`).

2.1. Spatio-Temporal Gaussian Process (STGP) Models

For the notational simplicity, we first denote the spatio-temporal coordinate (t, s_i) by $\mathbf{x}_{it} \in \mathcal{X}$, $\mathcal{X} \coloneqq \mathcal{T} \times \mathcal{S} \subset \mathbb{R}^3$. For any subset $\mathbf{X} \subseteq \mathcal{X}$ with N spatio-temporal coordinates, the set of function variables $\mathbf{f} \coloneqq \{f(\mathbf{x}_{it})\}_{\mathbf{x}_{it} \in \mathbf{X}}$ has joint zero-mean Gaussian distribution

$$p(\mathbf{f}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}_{XX}), \tag{1}$$

where $\mathbf{K}_{XX}$ is a $N \times N$ matrix and its entries are pairwise evaluations of $k(\mathbf{x}, \mathbf{x}')$, $\forall \mathbf{x}, \mathbf{x}' \in \mathbf{X}$.

Case model. We define a spatio-temporal model for the confirmed cases in the following form:

$$y_{it} = \mu_{it} + f(\mathbf{x}_{it}) + \epsilon_{it},\ i \in \mathcal{I}, t \in \mathcal{T},$$

where $\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is assumed to be i.i.d. normally distributed; μ_{it} is the mean of number of confirmed cases at week t in location i.

We assume the mean of the number of confirmed cases in a certain location relates to covariates in other locations according to an underlying undirected graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$, where $\mathcal{I}$ is the set of vertices representing all the locations, and $\mathcal{E} \subseteq \{(i, j) \in \mathcal{I}^2\}$ is a set of undirected edges representing the connections between locations. There is an edge between two vertices whenever the corresponding locations are geographically adjacent. Let $\boldsymbol{\eta}_{it} \coloneqq [\eta_{it1}, \ldots, \eta_{itl}, \ldots, \eta_{itL}]^\top \in \mathbb{R}^L$ denote the data of location $i \in \mathcal{I}$ at time $t \in \mathcal{T}$ and $\boldsymbol{\omega}_{it} \coloneqq [\omega_{it1}, \ldots, \omega_{itl}, \ldots, \omega_{itL}]^\top \in \mathbb{R}^L$ denote the parameters of the corresponding location; L denotes the number of features. Here, in practice, we use the number of confirmed cases, the number of deaths, and six community mobilities variables in the past two weeks as the input features with $L = 16$. Formally, we define μ_{it} as

$$\mu_{it} = \sum_{\tau \in \mathcal{H}_t} \sum_{j:(i,j) \in \mathcal{E}} \boldsymbol{\eta}_{j\tau} \boldsymbol{\omega}_{j\tau},\ \forall i \in \mathcal{I}, t \in \mathcal{T}, \tag{2}$$

where $\mathcal{H}_t = \{\tau : t - d \leq \tau < t\}$ represent the recent history with memory depth $d < T$.

For a set of N observed spatio-temporal coordinates $\mathbf{X}$, we denote the number of confirmed cases and their estimated means as $\mathbf{y} \coloneqq \{y_{it}\}_{(i,t):\mathbf{x}_{it} \in \mathbf{X}}$ and $\boldsymbol{\mu} \coloneqq \{\mu_{it}\}_{(i,t):\mathbf{x}_{it} \in \mathbf{X}}$, respectively. Then we can express the conditional probability distribution of $\mathbf{y}$ as

$$p(\mathbf{y}|\mathbf{f}) = \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}, \mathbf{K}_{XX} + \sigma_\epsilon \mathbf{I}), \tag{3}$$

where $\mathbf{I}$ is a $N \times N$ identity matrix.

Hotspot model. We express the conditional probability of the hotspots $\mathbf{h}$ for a subset of spatio-temporal coordinates $\mathbf{X}$ as:

$$p(\mathbf{h}|\mathbf{f}) = \prod_{\mathbf{x}_{it} \in \mathbf{X}} \mathcal{B}(h_{it}|\phi(f(\mathbf{x}_{it}))), \tag{4}$$

where $\mathcal{B}(h_{it}|\phi(f(\mathbf{x}_{it}))) = \phi(f(\mathbf{x}_{it}))^{h_{it}}(1 - \phi(f(\mathbf{x}_{it})))^{1-h_{it}}$ is the likelihood for the Bernoulli distribution and ϕ is a sigmoid function.

Learning objective. We aim to detect the hotspot while taking advantage of the information that have been recorded in the number of confirmed cases. To this end, we find the optimal model parameters by solving the following combined objective:

$$\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}) \coloneqq \ell_h(\boldsymbol{\theta}) + \delta \ell_y(\boldsymbol{\theta}), \tag{5}$$

where δ controls the ratio between two objectives and $\boldsymbol{\theta} \in \Theta$ is the set of parameters defined in GP. The $\ell_y(\boldsymbol{\theta}) \coloneqq \log p(\mathbf{y})$ denotes the log marginal likelihood of observed confirmed cases and $\ell_h(\boldsymbol{\theta}) \coloneqq \log p(\mathbf{h})$ denotes the log marginal likelihood of observed hotspots. We note that log marginal likelihood of cases in the second term plays a key role in "regularizing" the model by leveraging the information in the case records as shown in Fig. 7 (Appendix D). We also present the k-fold cross validation that quantitatively measures the F_1 score of the hotspot detection and the mean square error of the case prediction with different δ in Fig. 8 (Appendix D). The result confirms that the appropriate choice of δ can significantly improve the performance of hotspot detection.

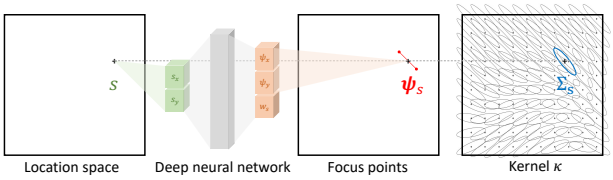


Figure 1. An illustration of the deep neural network that maps an arbitrary spatial location s to its covariance Σ_s and the corresponding weight w_s.

2.2. Spatio-Temporal Deep Neural Kernel

We discuss the choice of the kernel function k in this subsection. Standard GP models use a stationary covariance, in which the covariance between any two points is a function of Euclidean distance. However, stationary GPs fail to adapt to variable smoothness in the function of interest. This is of particular importance in geophysical and other spatial data sets, in which domain knowledge suggests that the function may vary more quickly in some parts of the input space than in others. For example, COVID-19 is likely to be spreading slower than in sparsely versus densely populated regions. Here, we consider the following non-stationary spatio-temporal kernel:

$$k(t,t',s,s') = \nu(t,t') \cdot \left(\sum_{r=1}^{R} w_{s'}^{(r)} \upsilon^{(r)}(s,s') \right), \qquad (6)$$

where $\nu(t,t')$ is a stationary kernel that captures temporal correlation between time t and t'; $\upsilon^{(r)}(s,s')$ is a component of the non-stationary spatial kernel which evolves over the space and $w_{s'}^{(r)}$ is the corresponding weight satisfying $\sum_{r=1}^{R} w_{s'}^{(r)} = 1$. R is the number of components considered. By likening the relationship between the spatial kernel component to that of the Gaussian component in Gaussian mixture, we seek to enhance the representative power of our kernel by adding more independent components to the spatial kernel.

Stationary temporal kernel. We define the kernel function that characterizes the temporal correlation between $t, t' \in \mathcal{T}$ as an stationary Gaussian function:

$$\nu(t,t') = \exp\left\{ -\frac{1}{2\sigma_\nu^2} ||t-t'||^2 \right\},$$

where $\sigma_\nu \in \mathbb{R}_+$ is the bandwidth parameter. This kernel function hypothesizes that the level of virus' transmission is highly related to its recent history and their correlation will decay exponentially over time.

Non-stationary spatial kernel. To account for non-stationarity, we now allow the smoothing kernel to depend on spatial location s. For ease of discussion and simplicity of notation, we omit the superscript r in $\upsilon^{(r)}(s,s')$ and $w_{s'}^{(r)}$, and present the structure of a single non-stationary spatial kernel component. We use $\kappa_s(\cdot)$ to denote a kernel which

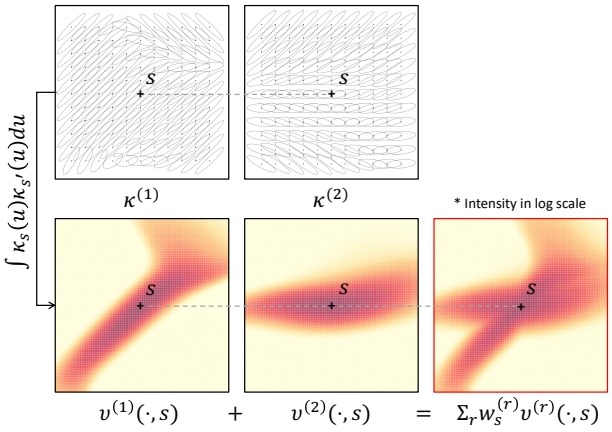


Figure 2. An examples of the spatial kernel with two components $\sum_r w_s^{(r)} \upsilon^{(r)}(\cdot,s)$ evaluated at the same location s. This instance is constructed using two different kernel κ, which are parameterized by two randomly generated φ_1 and φ_2.

is centered at the point s and whose shape is a function of location s. Once $\kappa_s(\cdot)$ is specified for all $s \in \mathcal{S} \subseteq \mathbb{R}^2$, the correlation between two points s and s' is then

$$\upsilon(s,s') \propto \int_{\mathbb{R}^2} \kappa_s(u)\kappa_{s'}(u)du. \qquad (7)$$

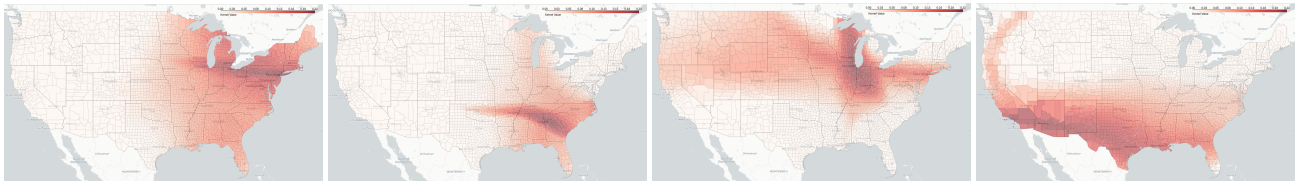
Because of the constructive formulation under the moving average specification, the resulting correlation function $\upsilon(s,s')$ is certain to be positive definite. We favor working with the kernels $\kappa_s(\cdot)$ rather than working directly with the correlation function $\upsilon(s,s')$ since this makes it difficult to ensure symmetry and positive definiteness for all s and s'. Following the idea of (Bernardo et al., 1998; Zhu et al., 2021), we define each $\kappa_s(\cdot)$ to be a normal kernel centered at s with spatially varying covariance matrix Σ_s. In this case given the parameterized Σ_s and $\Sigma_{s'}$, the correlation function is given by an easy to compute formula

$$\upsilon(s,s') \propto \frac{|\Sigma_s + \Sigma_{s'}|^{-\frac{1}{2}}}{2\pi} \exp\left\{ -\frac{(s'-s)^\top (\Sigma_s + \Sigma_{s'})^{-1}(s'-s)}{2} \right\}.$$

The derivation of this formula can be found in Appendix B.

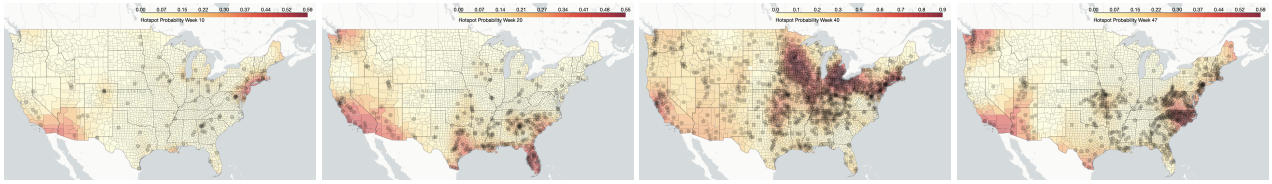
To assure that the kernel $\{\kappa_s(\cdot)\}$ vary smoothly over space $\mathcal{S}$, we parameterize Σ_s and then allow the parameters to evolve with location. For this paper we will focus on a geometrically based specification which readily extends beyond the use of the Gaussian kernel considered here.

There is a one-to-one mapping from a bivariate normal distribution to its one standard deviation ellipse, so we define a spatially varying family of ellipses which, in turn, defines the spatial distribution for Σ_s. Let the two focus points in $\Psi \subset \mathbb{R}^2$ denoted by $\boldsymbol{\psi}_s := (\psi_x(s), \psi_y(s)) \in \Psi$ and $-\boldsymbol{\psi}_s := (-\psi_x(s), -\psi_y(s)) \in \Psi$ define an ellipse centered at s with fixed area A. This then corresponds to the Gaussian



(a) New York (b) Atlanta (c) Chicago (d) Los Angeles

Figure 3. Examples of the learned spatial kernel $\sum_{r=1}^{R} w_s^{(r)} \upsilon^{(r)}(\cdot, s)$ with four components evaluated at four major metropolitan areas in the U.S.. These maps show the spatial influence of these area to other region of the U.S.. The color depth indicates the intensity of the kernel value; the darker the color the higher the kernel's value.



(a) April 12, 2020 (b) August 9, 2020 (c) October 4, 2020 (d) December 6, 2020

Figure 4. Spatial view of one-week-ahead and county-wise hotspot probability $p(\mathbf{h}_*)$ suggested by our fitted model ($\delta = 10^{-5}$) using real COVID-19 data. This figure presents examples at four particular weeks, where the color depth indicates the probability of predicted hotspots and the black circles represent the hotspots given in the data.

kernel with covariance matrix Σ_s defined by

$$\Sigma_s = \lambda^2 \begin{pmatrix} Q + \frac{\|\boldsymbol{\psi}_s\|^2}{2} \cos 2\alpha & \frac{\|\boldsymbol{\psi}_s\|^2}{2} \sin 2\alpha \\ \frac{\|\boldsymbol{\psi}_s\|^2}{2} \sin 2\alpha & Q - \frac{\|\boldsymbol{\psi}_s\|^2}{2} \cos 2\alpha \end{pmatrix}, \quad (8)$$

where $\alpha = \tan^{-1}(\psi_y(s)/\psi_x(s))$, $Q = \sqrt{4A^2 + \|\boldsymbol{\psi}_s\|^4 \pi^2}/2\pi$, and λ is a scaling parameter that controls the overall intensity of the covariance. This demonstrates how the spatially distributed pairs $\boldsymbol{\psi}_s := (\psi_x(s), \psi_y(s))$ give rise to a spatially distributed covariance matrix Σ_s. The derivation of (8) can be found in Appendix C.

Neural network representation for focus points. Here we represent the mapping $\varphi : \mathcal{S} \to \Psi \times [0, 1]$ from the location space $\mathcal{S}$ to the joint space of focus point Ψ and the weight $[0, 1]$ using a deep neural network. To be specific, the input of the network is the location s and the output of the networks is the concatenation of the corresponding focus points $\boldsymbol{\psi}_s$ of that location and the weight w_s defined in (6). The architecture of the neural network has been described in Fig. 1. In Fig. 2, we also demonstrate two specific instances of the resulting spatial kernel υ given two different κ. This implies that the neural network φ encodes the non-homogeneous geographical information across the region that plays a key role in virus' transmission.

3. Results

Model interpretation. To intuitively interpret the learned spatial kernel, we visualize the kernel evaluation given one of its inputs, i.e., $\sum_{r=1}^{R} w_s^{(r)} \upsilon^{(r)}(s, \cdot)$. Such kernel evaluation represents the spatial correlation (or sphere of influence) of a particular location spreading the virus. Fig. 3 shows four examples of the spatial kernel for the latitude and longitude of New York, Atlanta, Chicago, and Los Angeles, respectively. We observe that these major metropolitan areas have a substantially different spatial correlation with their neighboring regions due to the non-stationarity of the spatial kernel. For example, as one of the nation's major economic and transportation hubs, New York has a significant impact on the entire Eastern United States, while Atlanta only has a regional influence in the Southeastern United States. Chicago and Los Angeles, the second and third most populous cities in the United States, can extend their influences to the entire north and south of the country, respectively. Increasing the number of spatial kernel components could increase the flexibility and the interpretability of the model (Appendix E); however, due to the need for additional parameters in the neural networks, the computational time dramatically increases when $R \geq 3$, with minimal performance improvement.

Hotspot detection. In Fig. 4, we visualize the prediction results on the map to intuitively examine the predictive performance from the spatial perspective. We select four particular weeks that represent different stages of the COVID-19 pandemic in 2020. As we can see, the hotspot probability resulting from our model is considerably high whenever a genuine hotspot occurs and considerably low otherwise, which confirms the effectiveness of our framework. Our method can also capture the spatial occurrence of these hotspots nicely, in which regions with sparsely distributed hotspots usually have a lower probability. In comparison, other regions with densely distributed hotspots have a higher probability. We emphasize that our hotspot detection framework can provide a realistic prediction that varies smoothly over time and space due to our GP assumption. This can be extremely useful when we try to make a continuous prediction or estimate the likelihood of a hotspot to happen at an arbitrary spatio-temporal coordinate. We also compare the performance of our method and six other baseline approaches (Appendix F).

References

Bernardo, J., Berger, J., Dawid, A., Smith, A., et al. Non-stationary spatial modeling. 1998.

Centers for Disease Control and Prevention. COVID-19 forecasts: Deaths, 2021. URL https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html#state-forecasts.

Google. Covid-19 community mobility reports, 2020. URL https://www.google.com/covid19/mobility/.

Oster, A. M., Caruso, E., DeVies, J., Hartnett, K. P., and Boehmer, T. K. Transmission dynamics by age group in covid-19 hotspot counties—united states, april–september 2020. *Morbidity and Mortality Weekly Report*, 69(41): 1494, 2020a.

Oster, A. M., Kang, G. J., Cha, A. E., Beresovsky, V., Rose, C. E., Rainisch, G., Porter, L., Valverde, E. E., Peterson, E. B., Driscoll, A. K., et al. Trends in number and distribution of covid-19 hotspot counties—united states, march 8–july 15, 2020. *Morbidity and Mortality Weekly Report*, 69(33):1127, 2020b.

Rossman, H., Keshet, A., Shilo, S., Gavrieli, A., Bauman, T., Cohen, O., Shelly, E., Balicer, R., Geiger, B., Dor, Y., et al. A framework for identifying regional outbreak and spread of covid-19 from one-minute population-wide surveys. *Nature Medicine*, 26(5):634–638, 2020.

Shalev-Shwartz, S. and Ben-David, S. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

Times, T. N. Y. We're sharing coronavirus case data for every u.s. county, 2020. URL https://www.nytimes.com/article/coronavirus-county-data-us.html.

Zhu, S., Li, S., Peng, Z., and Xie, Y. Imitation learning of neural spatio-temporal point processes. *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2021. doi: 10.1109/TKDE.2021.3054787.

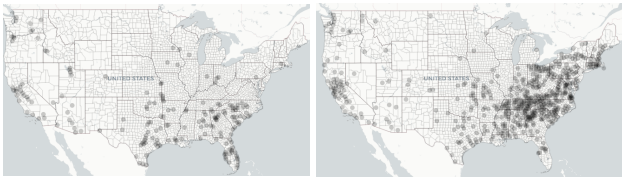
A. COVID-19 data description

The data sets we used in our study includes number of cases and deaths, COVID-19 hotspots identified by Centers for Disease Control and Prevention (CDC), and community mobility provided by Google. The study period is from March 15, 2020 to January 17, 2021, consisting of 50 weeks and 3,144 US counties. We exclude the data after February 2021 when large-scale COVID-19 vaccine rollout had been launched across the United States, which effectively shifted the dynamics of the COVID-19 spread.

Confirmed cases and deaths. We used the data set from The New York Times (Times, 2020) which includes two parts: (i) *confirmed cases* are counts of individuals whose coronavirus infections were confirmed by a laboratory test and reported by a federal, state, territorial, or local government agency; (ii) *confirmed deaths* are individuals who have died and meet the definition for a confirmed COVID-19 case. In practice, we have observed periodic weekly oscillations in daily reported cases and deaths, which could have been caused by testing bias (higher testing rates on certain days of the week). To reduce such bias, we aggregate the number of cases and deaths of each county *by week*.

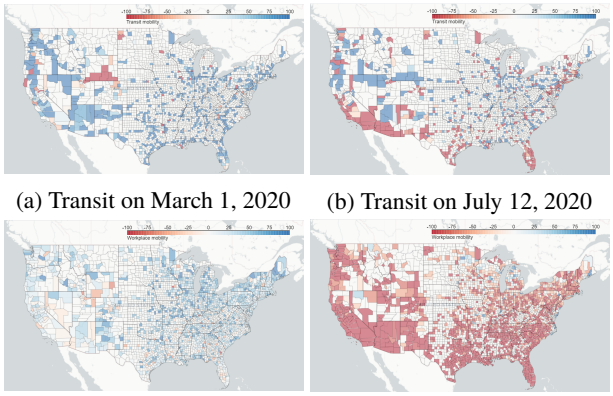
Hotspots. On May 7, 2020, CDC, along with other federal agencies, began identifying counties with increasing COVID-19 hotspots to better understand transmission dynamics and offer targeted support to health departments in affected communities. CDC identified hotspots daily starting on January 22, 2020 among counties in U.S. states and the District of Columbia by applying standardized criteria developed through a collaborative process involving multiple federal agencies (Oster et al., 2020b;a); in general, hotspots were defined based on relative temporal increases in number of cases. To match the temporal resolution with the number of cases and deaths, we expand the definition of a hotspot from daily-level to weekly-level, where a week is identified as a hotspot if it contains at least one hotspot day identified by CDC. The weekly number of counties meeting hotspot criteria peaked in early April, decreased and stabilized during mid-April–early June, then increased again during late June–early July. The percentage of counties in the South and West Census regions meeting hotspot criteria increased from 10% and 13%, respectively, during March–April to 28% and 22%, respectively, during June–July. Fig. 5 gives snapshots of the identified hotspots at four particular weeks.

Community mobility. The COVID-19 Community Mobility Reports (Google, 2020) record people's movement by county daily, across various categories such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential. The data shows how visitors to (or time spent in) categorized places change compared to the baseline days (in percentage). The negative percentage means that the level of mobility is lower than the baseline,



(a) June 28, 2020 (b) December 20, 2020

Figure 5. Snapshots of hotspot identified by CDC. The black circles indicate the counties that have been identified as hotspot in that week.



(a) Transit on March 1, 2020 (b) Transit on July 12, 2020

(c) Workplace on March 1, 2020 (d) Workplace on July 12, 2020

Figure 6. Overview of Google mobility data in two selected categories: workplace and transit on two different days. Counties in red and blue indicate their mobility is lower and higher than the normal level, respectively. The mobility level varies over time and space due to local government policy change in response to COVID-19.

and the positive percentage represents the opposite. The mobility on a baseline day represents a normal value for that day of the week. This mobility report sets the baseline as the median value from the five weeks from January 3rd to February 6th, 2020. Similar to the aforementioned two data sets, we aggregate each county's mobility data by week. Examples of two categories, transit stations and workplaces, are shown in Fig. 6.

B. Proof of non-stationary kernel

Assume two independent bivariate Gaussian random variables X_s, $X_{s'}$ centered at locations s, s', respectively, with $\Sigma_s, \Sigma_{s'}$ parameterized by

$$\Sigma_s = \begin{bmatrix} a^2 & \rho ab \\ \rho ab & b^2 \end{bmatrix}, \ \Sigma_{s'} = \begin{bmatrix} a'^2 & \rho' a' b' \\ \rho' a' b' & b'^2 \end{bmatrix}.$$

Given two independent Gaussian random variables X and Y and their probability density functions f_X and f_Y, the distribution f_Z of $Z = X + Y$ equals the convolution of f_X

and f_Y, i.e.,

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) dx$$

Denote the probability density function of X_s and $X_{s'}$ as $\kappa_s(\cdot), \kappa_{s'}(\cdot)$, we have

$$f_{X_s+X_{s'}}(x) = \int_{\mathbb{R}^2} \kappa_s(u)\kappa_{s'}(x-u) du.$$

We also have the following equation due to the property of the Gaussian function:

$$\kappa_{s'}(2s'-u) = \kappa_{s'}(u),\ \kappa_s(2s-u) = \kappa_s(u).$$

Let $x = 2s'$ or $x = 2s$, we therefore have

$$f_{X_s+X_{s'}}(2s') = f_{X_s+X_{s'}}(2s) = \int_{\mathbb{R}^2} \kappa_s(u)\kappa_{s'}(u) du = \upsilon(s,s'),$$

which leads to (7).

Since $X_s + X_{s'}$ follows a Gaussian distribution $X_s + X_{s'} \sim \mathcal{N}(s+s', \Sigma_s + \Sigma_{s'})$, the non-stationary kernel $\upsilon(s,s')$ can be written as

$$\begin{aligned}
&\upsilon(s,s') \\
&= f_{X_s+X_{s'}}(2s') \\
&= \frac{1}{2\pi|\Sigma_s+\Sigma_{s'}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(s'-s)^\top(\Sigma_s+\Sigma_{s'})^{-1}(s'-s)\right\}.
\end{aligned}$$

Let $P = (\rho^2-1)b^2$ and $P' = (\rho'^2-1)b'^2$, we have

$$\upsilon(s,s') \propto \frac{1}{q_1} \exp\left\{-\frac{1}{q_2}(s-s')^\top W(s-s')\right\},$$

where

$$\begin{aligned}
W &= \begin{bmatrix} b^2+b'^2 & -(\rho ab+\rho' a'b') \\ -(\rho ab+\rho' a'b') & a^2+a'^2 \end{bmatrix}, \\
q_1 &= 2\pi\sqrt{-2\rho\rho' aa'bb' - a^2(P-b'^2) - a'^2(P'-b^2)}, \\
q_2 &= -2(2\rho\rho' aa'bb' + a^2(P-b'^2) + a'^2(P'-b^2)).
\end{aligned}$$

C. Reparametrization of Gaussian distribution

Assume an ellipse centered at the origin with area A has two focus points $(\psi_x, \psi_y), (-\psi_x, -\psi_y)$ in $\mathbb{R}^2$ where $\psi_x, \psi_y \in \mathbb{R}$. We define the semi-major and semi-minor axis of the ellipse as σ_1, σ_2. According to the ellipse formula we have:

$$\begin{cases} \pi\sigma_1\sigma_2 &= A, \\ \sigma_1^2 - \sigma_2^2 &= \psi_x^2 + \psi_y^2 = \|\psi\|^2. \end{cases}$$

By solving the above linear equation system, we have

$$\begin{aligned}
\sigma_1 &= \left(\frac{\sqrt{4A^2+\|\psi\|^4\pi^2}}{2\pi} + \frac{\|\psi\|^2}{2}\right)^{\frac{1}{2}}, \\
\sigma_2 &= \left(\frac{\sqrt{4A^2+\|\psi\|^4\pi^2}}{2\pi} - \frac{\|\psi\|^2}{2}\right)^{\frac{1}{2}}.
\end{aligned}$$

Since the rotation angle α of the ellipse is $\alpha = \tan^{-1}(\psi_y/\psi_x)$, a bivariate normal random variable X can be defined as

$$X = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix},$$

where Z_1 and Z_2 are two independent random variables with variance σ_1^2 and σ_2^2, respectively. Here we introduce a kernel scale parameter τ_z and derive the covariance of X as follows:

$$\Sigma = \tau_z^2 \begin{bmatrix} \sigma_1^2\cos^2\alpha + \sigma_2^2\sin^2\alpha & (\sigma_1^2-\sigma_2^2)\cos\alpha\sin\alpha \\ (\sigma_1^2-\sigma_2^2)\cos\alpha\sin\alpha & \sigma_1^2\sin^2\alpha + \sigma_2^2\cos^2\alpha \end{bmatrix}$$

Substitute the solution of σ_1 and σ_2 into the above equation, we have

$$\begin{aligned}
&\sigma_1^2\cos^2\alpha + \sigma_2^2\sin^2\alpha \\
&= \frac{\sqrt{4A^2+\|\psi\|^4\pi^2}}{2\pi}(\cos^2\alpha+\sin^2\alpha) + \frac{\|\psi\|^2}{2}(\cos^2\alpha-\sin^2\alpha) \\
&= \frac{\sqrt{4A^2+\|\psi\|^4\pi^2}}{2\pi} + \frac{\|\psi\|^2}{2}\cos(2\alpha),
\end{aligned}$$

and similarly

$$\begin{aligned}
\sigma_1^2\sin^2\alpha + \sigma_2^2\cos^2\alpha &= \frac{\sqrt{4A^2+\|\psi\|^4\pi^2}}{2\pi} - \frac{\|\psi\|^2}{2}\cos(2\alpha), \\
(\sigma_1^2-\sigma_2^2)\cos\alpha\sin\alpha &= \|\psi\|^2\cos\alpha\sin\alpha = \frac{\|\psi\|^2}{2}\sin 2\alpha.
\end{aligned}$$

Thereby we obtain the matrix shown in Equation (8).

D. Cross-validation for δ

This section presents the cross-validation result for δ in (5) defined in Section 2.1. Fig. 7 gives several examples of the predictions using different δ. Fig. 8 summarizes the k-fold cross validation that quantitatively measures the F_1 score of the hotspot detection and the mean square error of the case prediction with different δ.

E. Model comparison with different R in the spatial kernel

This section presents the comparison of our model using different number of spatial components in the kernel function. Fig. 9 gives an example of visualized kernel evaluation centered at Chicago. It shows that the representative power of the kernel can be greatly enhanced by increasing the number of spatial components R.

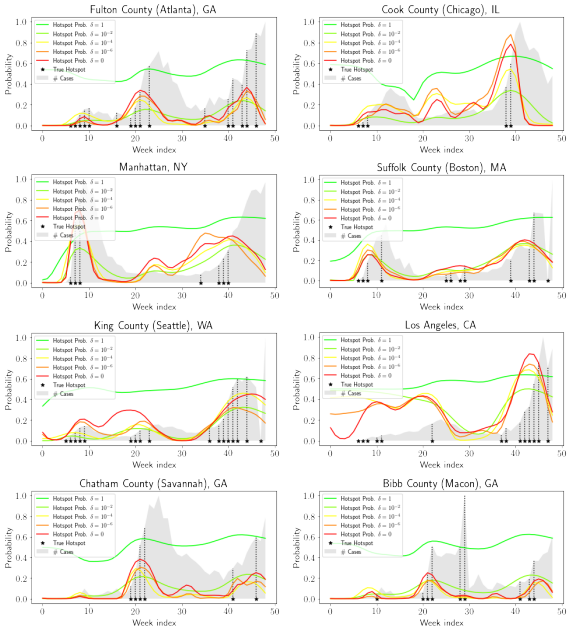


Figure 7. Comparison of one-week-ahead and county-wise hotspot probability $p(\mathbf{h}_*)$ using different δ. The model with $\delta = 10^{-5}$ attains the best performance in F_1 score.

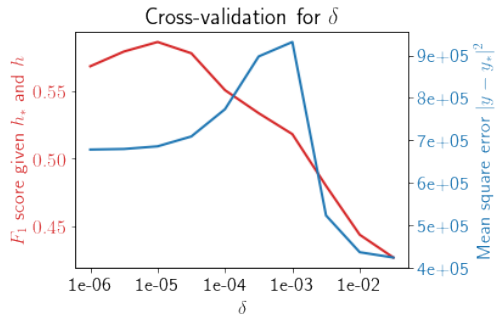
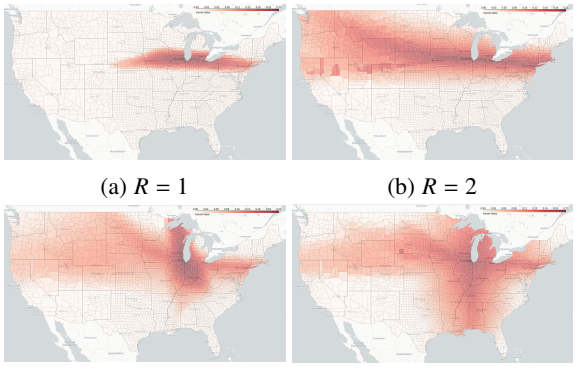


Figure 8. Cross-validation result for δ displaying the mean-square error (blue) and F_1-score (red).



(a) $R = 1$ (b) $R = 2$

(c) $R = 4$ (d) $R = 6$

Figure 9. visualization of spatial kernel at Chicago for different R.

Table 1. F_1 score of out-of-sample hotspot detections.

	Precision	Recall	F_1 score
`Perceptron`	0.424	0.242	0.308
`Logistic`	0.564	0.178	0.270
`Linear SVM`	0.622	0.064	0.117
k−NN	0.517	0.398	0.450
`Kernel SVM`	0.599	0.360	0.450
`Decision Tree`	0.537	0.293	0.340
`STGP` ($\delta = 10^{-5}$)	0.457	0.968	0.621

F. Comparison with baselines

We adopt standard performance metrics, including precision, recall, and F_1 score. This choice is because hotspot detection can be viewed as a binary classification problem. We aim to identify a hotspot for a particular location at a particular week in the data. Define the set of all identified hotspots as U, the set of detected hotspots using our method as V. Then precision P and recall R are defined as $P = |U \cap V|/|V|$, $R = |U \cap V|/|U|$, where $|\cdot|$ is the number of elements in the set. The F_1 score combines the *precision* and *recall*: $F_1 = 2PR/(P+R)$ and the higher F_1 score the better. Since numbers of hotspots in real data are highly sparse (comparing to the total number of spatio-temporal coordinates), we do not use the ROC curve (true positive rate versus false-positive rate) in our setting. The evaluation procedure is described as follows. Given the observed hotspot and other covariates (cases, deaths, and mobility) until week t, we perform detection for all the locations at week $t+1$. If the detected hotspot were indeed identified as a true hotspot by CDC, then it is a success. Otherwise, it is a misdetection. In our data, there are $50 \times 3144 = 157,200$ spatio-temporal coordinates in total, and 12,000 of them were identified as true hotspots.

We compare the hotspot detection results of our proposed method and several standard methods in binary classification, including perceptron, logistic regression, linear support vector machine (SVM), k-nearest neighbor (k-NN), kernel SVM with Gaussian kernel, and decision tree; see (Shalev-Shwartz & Ben-David, 2014) for a detailed review of those machine learning algorithms. Table 1 shows the F_1 score for the out-of-sample prediction at county-level using our method. The result confirms that our model significantly outperforms other baseline methods.