

Spatio-temporal Point Process Compartment Modeling for Infectious Diseases

André V. Ribeiro Amaral*

andre.ribeiroamaral@kaust.edu.sa

King Abdullah University of Science and Technology
Geospatial Statistics and Health Surveillance Research Group

* Joint work with Jonatan A. González and Paula Moraga.

Introduction



Figure 1: Representation of susceptible (blue), infected (red), and recovered (green) individuals distributed in space at a given time point t .

Objective

We are also interested in modeling the infected individuals in space and time. And we will do this based on point pattern data observed at different time points.

Objective

We are also interested in modeling the infected individuals in space and time. And we will do this based on point pattern data observed at different time points.

In that way, our methodology addresses the epidemic spatio-temporal dynamics by modeling it in two steps

1. Fitting a temporal (compartment) model, and
2. Using the previous step acquired information as the mean of a log-Gaussian Cox process for the point pattern representing the infected individuals in the studied region and time interval

SIR modeling

The base-SIR model (Kermack and McKendrick, 1927) is described by the following diagram



and it has the following assumptions

1. Homogeneous population with uniform mixing,
2. Constant infectious and recovery rates over time, and
3. Preserved population mass,

SIR modeling

The base-SIR model (Kermack and McKendrick, 1927) is described by the following diagram



and it has the following assumptions

1. Homogeneous population with uniform mixing,
2. Constant infectious and recovery rates over time, and
3. Preserved population mass,

Many extensions of this model have been proposed (e.g., SEIR). Here, we will extend point (1.) of the base-SIR model to accommodate different age groups.

SIR modeling

For $t \in \mathcal{T} \subseteq \mathbb{R}_+$, let $\mathbf{S}_i(t)$, $\mathbf{I}_i(t)$, and $\mathbf{R}_i(t)$ denote the number of susceptible, infected, and recovered individuals, respectively, at time t for age-group i . Then,

$$\begin{aligned}\frac{d\mathbf{S}_i(t)}{dt} &= -\beta \mathbf{S}_i(t) \sum_{\text{all } j} C_{ij} \cdot \frac{\mathbf{I}_j(t)}{\mathbf{N}_j} \\ \frac{d\mathbf{I}_i(t)}{dt} &= +\beta \mathbf{S}_i(t) \sum_{\text{all } j} C_{ij} \cdot \frac{\mathbf{I}_j(t)}{\mathbf{N}_j} - \gamma \mathbf{I}_i(t) \\ \frac{d\mathbf{R}_i(t)}{dt} &= +\gamma \mathbf{I}_i(t),\end{aligned}\tag{1}$$

such that C_{ij} is a contact matrix, $\mathbf{N}_i(t) = \mathbf{N}_i$, $\forall t$, and $\beta, \gamma > 0$.

SIR modeling

For $t \in \mathcal{T} \subseteq \mathbb{R}_+$, let $\mathbf{S}_i(t)$, $\mathbf{I}_i(t)$, and $\mathbf{R}_i(t)$ denote the number of susceptible, infected, and recovered individuals, respectively, at time t for age-group i . Then,

$$\frac{d\mathbf{S}_i(t)}{dt} = -\beta \mathbf{S}_i(t) \sum_{\text{all } j} C_{ij} \cdot \frac{\mathbf{I}_j(t)}{\mathbf{N}_j} \quad (1)$$

$$\frac{d\mathbf{I}_i(t)}{dt} = +\beta \mathbf{S}_i(t) \sum_{\text{all } j} C_{ij} \cdot \frac{\mathbf{I}_j(t)}{\mathbf{N}_j} - \gamma \mathbf{I}_i(t)$$

$$\frac{d\mathbf{R}_i(t)}{dt} = +\gamma \mathbf{I}_i(t),$$

such that C_{ij} is a contact matrix, $\mathbf{N}_i(t) = \mathbf{N}_i$, $\forall t$, and $\beta, \gamma > 0$.

Model (1) will be solved numerically. That is, given $(\mathbf{S}_i(0), \mathbf{I}_i(0), \mathbf{R}_i(0))$, $\forall i$, we will define a solution at $\{t_k; k = 0, 1, \dots, n\}$, such that $t_k \in \mathcal{T}$, $\forall k$.

Point Process modeling

For the previous discretization $\{t_k\}_k$, let $\xi(t_k)$ be a log-Gaussian Cox process driven by $\Lambda(\mathbf{u}; t_k)$.

Point Process modeling

For the previous discretization $\{t_k\}_k$, let $\xi(t_k)$ be a log-Gaussian Cox process driven by $\Lambda(\mathbf{u}; t_k)$.

In particular, we will define it as

$$\Lambda(\mathbf{u}; t_k) = \mu(\mathbf{u}; t_k) \cdot \exp\{\zeta(\mathbf{u}; t_k)\} \quad (2)$$

where $\zeta(\mathbf{u}; t_k)$ is a stationary Gaussian process with $\mathbb{E}(\zeta(\mathbf{u}; t_k)) = -\sigma^2/2$, $\forall k$ and \mathbf{u} , and $\text{Cov}(\zeta(\mathbf{u}_1; t_k), \zeta(\mathbf{u}_2; t_k)) = \sigma^2 \rho(h; t_k)$, such that σ^2 is the variance and h is the Euclidean distance between \mathbf{u}_1 and \mathbf{u}_2 .

As a remark, for specific choices of $\mu(\mathbf{u}; t_k)$, Model (2) was previously discussed by Diggle (2006).

Methodology

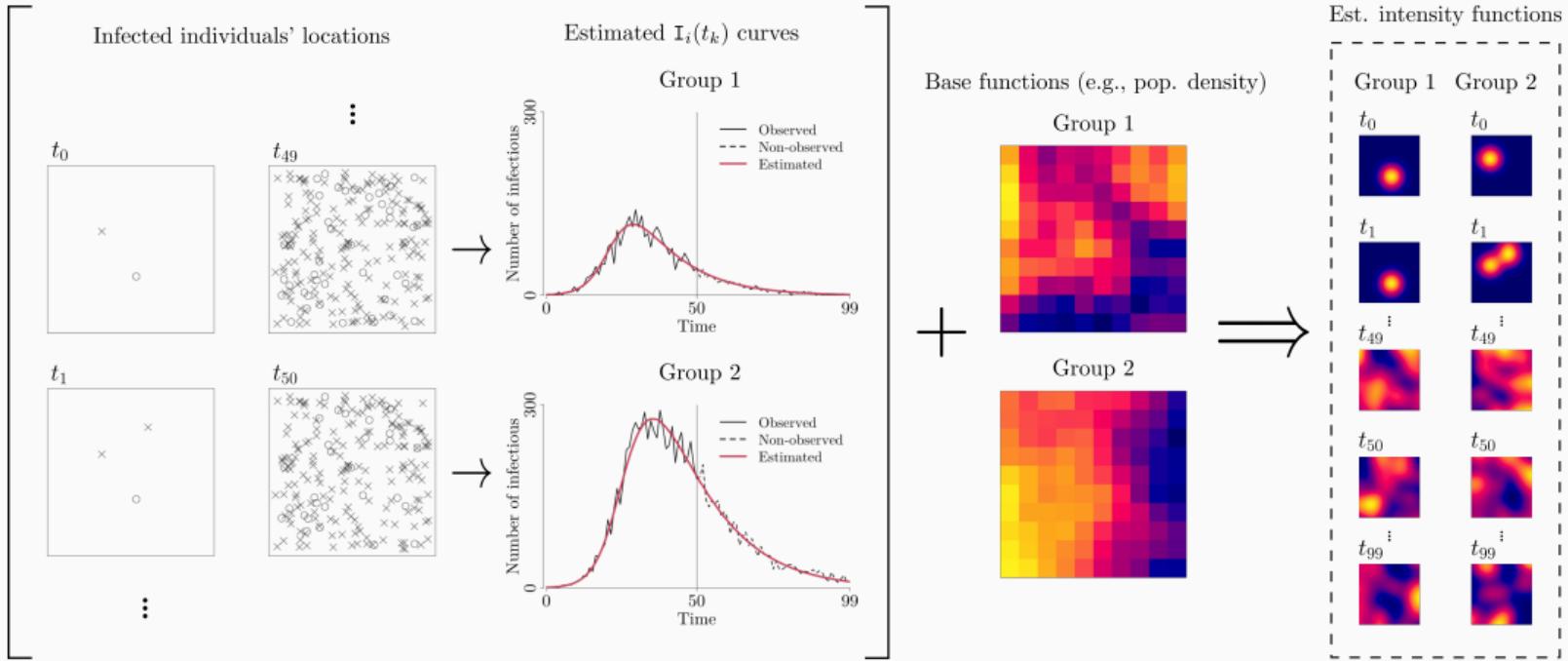


Figure 2: Two-step spatio-temporal modeling approach for infectious in all groups.

Temporal modeling

Recall that, for a set of initial values $(S_i(0), I_i(0), R_i(0))$, $\forall i$, and initial guesses for β and γ , we can solve the system of ODEs for $S_i(t_k)$, $I_i(t_k)$, and $R_i(t_k)$ with some numerical method.

For later reference, we will name these solutions $S_i^{\text{ODE}}(t_k)$, $I_i^{\text{ODE}}(t_k)$, and $R_i^{\text{ODE}}(t_k)$, respectively.

Temporal modeling

Now, suppose that we have obtained $\mathbf{i}_i(t_k)$, $\forall i, k$.

One way to model such data is assuming that they come from a certain probability distribution with mean given by the ODE solution $\mathbb{I}_i^{\text{ODE}}(t_k)$, $\forall i, k$. In particular, we will assume

$$\mathbb{I}_i(t_k) \sim \text{Negative Binomial}(\mathbb{I}_i^{\text{ODE}}(t_k), \varphi), \quad (3)$$

such that φ is the overdispersion parameter. This implies that $\mathbb{E}(\mathbb{I}_i(t_k)) = \mathbb{I}_i^{\text{ODE}}(t_k)$ and $\text{Var}(\mathbb{I}_i(t_k)) = \mathbb{I}_i^{\text{ODE}}(t_k)(1 + \mathbb{I}_i^{\text{ODE}}(t_k) \cdot \varphi)$.

Temporal modeling

For such an approach, notice that this is an iterative procedure. That is,

1. Set initial values for β , γ , and φ .
2. Solve Model (1) for $S_i(t_k)$, $I_i(t_k)$, and $R_i(t_k)$.
3. Plug the $I_i^{\text{ODE}}(t_k)$ curve into the mean component of Model (3) to evaluate its likelihood function.
4. Update β , γ , and φ , and get back to (2.) until reach convergence.

Here, we adopt a Bayesian framework and use **RStan** (Stan Development Team, 2021) to estimate the posterior distribution of $\boldsymbol{\theta} = (\beta, \gamma, \varphi)^{\top}$.

Spatio-temporal modeling

Given $\mathbb{I}_i(t_k)$ curves, we will have the following model

$$\mathcal{N}_i(t_k) | \Lambda_i(\mathbf{u}; t_k) = \lambda_i(\mathbf{u}; t_k) \sim \text{Poisson} \left(\int_{\mathcal{U}} \lambda_i(\mathbf{u}; t_k) d\mathbf{u} \right),$$

and the corresponding intensity functions will be described by

$$\Lambda_i(\mathbf{u}; t_k) = \mu_i(\mathbf{u}; t_k) \cdot \exp\{\zeta_i(\mathbf{u}; t_k)\}, \quad (4)$$

such that

$$\mu_i(\mathbf{u}; t_k) = \lambda_{0,i}(\mathbf{u}; t_k) \cdot \mathbb{I}_i(t_k) \cdot \exp\{\omega_{1,i} x_{1,i}(\mathbf{u}; t_k) + \cdots + \omega_{p,i} x_{p,i}(\mathbf{u}; t_k)\}, \quad (5)$$

where $\lambda_{0,i}(\mathbf{u}; t_k) \geq 0$, $\forall \mathbf{u}, k$, and $\int_{\mathcal{U}} \lambda_{0,i}(\mathbf{u}; t_k) = 1$. $(x_{1,i}(\mathbf{u}; t_k), \dots, x_{p,i}(\mathbf{u}; t_k))$ is a vector of p spatio-temporal covariates with coefficients $(\omega_{1,i}, \dots, \omega_{p,i})^\top$.

Spatio-temporal modeling

The final model is specified as follows

$$\mathcal{N}_i(t_k) | \Lambda_i(\mathbf{u}; t_k) = \lambda_i(\mathbf{u}; t_k) \sim \text{Poisson} \left(\int_{\mathcal{U}} \lambda_i(\mathbf{u}; t_k) d\mathbf{u} \right), \quad \forall i, k \quad (6)$$

$$\Lambda_i(\mathbf{u}; t_k) = \mu_i(\mathbf{u}; t_k) \cdot \exp\{\zeta_i(\mathbf{u}; t_k)\}$$

$$\mu_i(\mathbf{u}; t_k) = \lambda_{0,i}(\mathbf{u}; t_k) \cdot \mathbb{I}_i(t_k)$$

$$\zeta_i(\mathbf{u}; t_k | \boldsymbol{\eta}_i) \sim \text{Gaussian Process}(\beta_{0,i}, \phi_i(h; t_k | \boldsymbol{\eta}_i))$$

$$\boldsymbol{\eta}_i \sim \text{priors},$$

such that $\phi_i(h; t_k | \boldsymbol{\eta}_i)$ is a covariance function that depends on the structure of $\zeta_i(\mathbf{u}; t_k)$, and $\boldsymbol{\eta}_i$ is the vector of parameters and hyperparameters.

Model (6) will be fitted using R-INLA (Rue et al., 2009).

Data Simulation

We consider as a study region an area of approx. 3 km^2 in São Paulo, Brazil. For such a region, we divided people into three age groups: 0–19, 20–59, 60+.

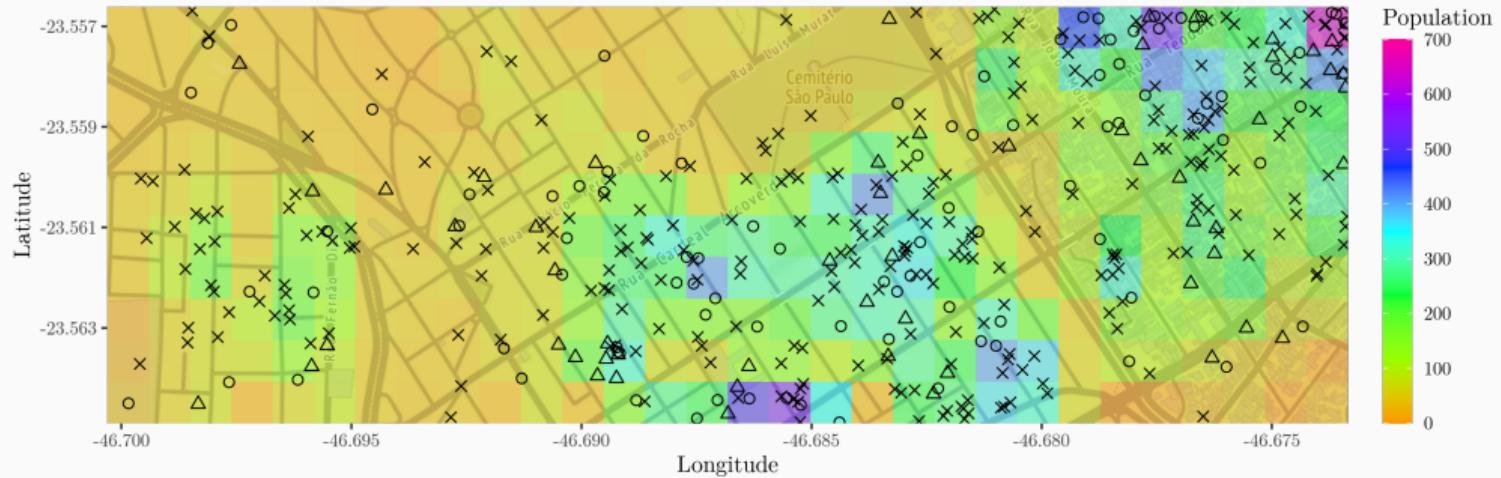


Figure 3: Studied region in São Paulo (Brazil) with the overlapped grid for the estimated population and infected individuals' locations.

Data Simulation

For a contact matrix and a vector of frequencies for people in each group, both previously estimated, we can, conditional on β , γ , and φ , simulate the $I_i(t_k)$ curves. We will do this for two scenarios, namely EP (Early Peak) and FC (Flat Curve).

Table 1: SIR model parameters for the two scenarios, namely EP and FC, as we can see in Figure 4.

	Scenario EP	Scenario FC
β	0.04	0.0175
γ	0.2	0.1
$1/\varphi$	0.01	0.01

Data Simulation

Then, based on Table 1, we can generate curves for the two scenarios.

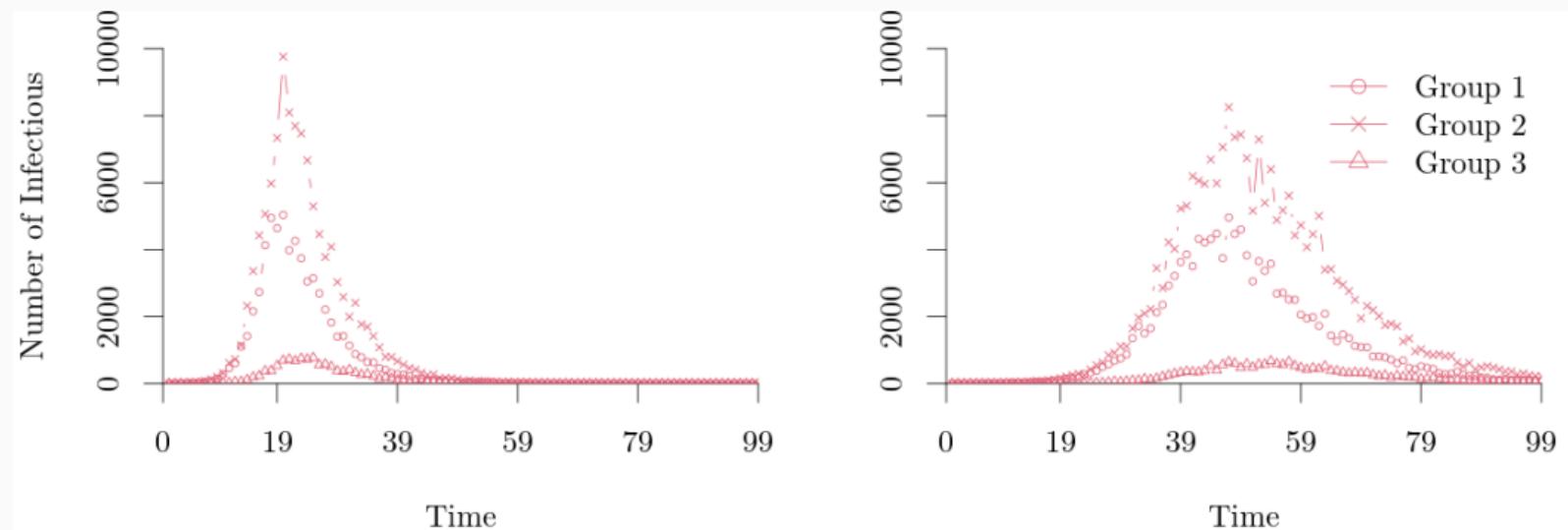


Figure 4: Simulated $I_i(t_k)$ curves, for all age groups 1, 2, and 3 (which corresponds to ages 0–19, 20–59, 60+, respectively). Scenarios EP (left) and FC (right).

Data Simulation

Once we have simulated the $\mathbb{I}_i(t_k)$ curves, $\forall i$, the true generated intensity functions will be sampled from the following model

$$\Lambda_i(\mathbf{u}; t_k) = \lambda_{0,i}(\mathbf{u}; t_k) \cdot \mathbb{I}_i(t_k) \cdot \exp\{\zeta_i(\mathbf{u}; t_k)\}, \text{ for each } i \text{ and } k = 0, 1, 2, \dots, 99,$$

such that $\lambda_{0,i}(\mathbf{u}; t_k)$ is the normalized populational grid and $\zeta_i(\mathbf{u}; t_k)$ is defined as before.

Model Fitting

Given the data, we can fit a null model (\mathcal{M}_0) and an alternative model (\mathcal{M}_1). In particular, for $\xi_i(t_k)$ modeled as follows

$$\mathcal{N}_i(t_k) | \Lambda_i(\mathbf{u}; t_k) = \lambda_i(\mathbf{u}; t_k) \sim \text{Poisson} \left(\int_{\mathcal{U}} \lambda_i(\mathbf{u}; t_k) d\mathbf{u} \right),$$

the null model (\mathcal{M}_0) will be given by

$$\Lambda_i(\mathbf{u}; t_k) = \exp\{\omega_{0,i} + \zeta_i(\mathbf{u}; t_k)\},$$

such that $\omega_{0,i}$ is an unknown intercept, and the alternative model (\mathcal{M}_1) will be given by

$$\Lambda_i(\mathbf{u}; t_k) = \mu_i(\mathbf{u}; t_k) \cdot \exp\{\zeta_i(\mathbf{u}; t_k)\},$$

where $\mu_i(\mathbf{u}; t_k)$ is defined as in Equation (5). $\zeta_i(\mathbf{u}; t_k)$ is defined accordingly.

Implementation and Results

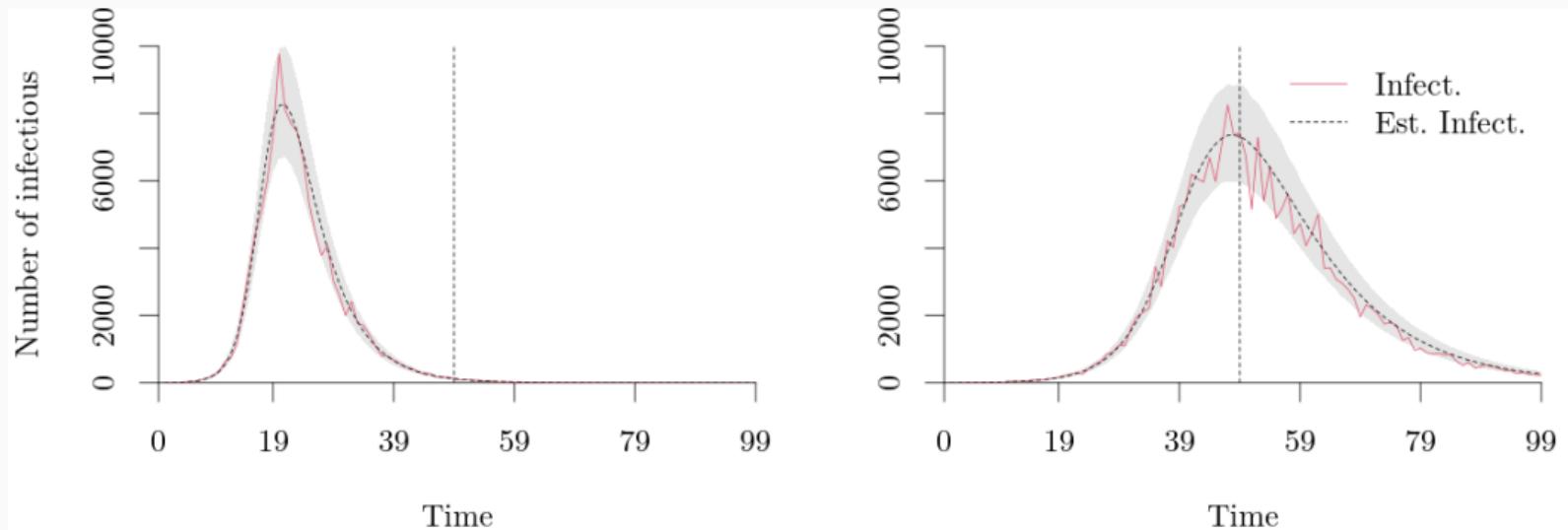


Figure 5: Estimated $I_i(t_k)$ curves for the age group 20–59 in the two scenarios, namely EP (left) and FC (right). 95% prediction intervals are also presented. Models were fitted with data up to t_{49} (vertical dashed line).

Implementation and Results

Based on a point estimate for $\mathbf{I}_i(t_k)$, as we have just obtained, now we can employ our modeling approach in space-time; that is, as before

$$\mathcal{N}_i(t_k) | \Lambda_i(\mathbf{u}; t_k) = \lambda_i(\mathbf{u}; t_k) \sim \text{Poisson} \left(\int_{\mathcal{U}} \lambda_i(\mathbf{u}; t_k) d\mathbf{u} \right), \quad \forall i, k$$

$$\Lambda_i(\mathbf{u}; t_k) = \mu_i(\mathbf{u}; t_k) \cdot \exp\{\zeta_i(\mathbf{u}; t_k)\}$$

$$\mu_i(\mathbf{u}; t_k) = \lambda_{0,i}(\mathbf{u}; t_k) \cdot \mathbf{I}_i(t_k)$$

$$\zeta_i(\mathbf{u}; t_k | \boldsymbol{\eta}_i) \sim \text{Gaussian Process}(\beta_{0,i}, \phi_i(h; t_k | \boldsymbol{\eta}_i))$$

$$\boldsymbol{\eta}_i \sim \text{priors},$$

And all fitting procedure will be performed based on data observed up to t_{49} .

Model Assessment

To compare the models, we will compute the Mean Arctangent Absolute Percentage Error (MAAPE) (Kim and Kim, 2016). For each group i and at t_k ,

$$\text{MAAPE}_{i,k} = \frac{1}{J} \sum_{j=1}^J \arctan \left(\left| \frac{(f_{i,j,k} - \hat{f}_{i,j,k})}{f_{i,j,k}} \right| \right).$$

such that J is the total number of cells, and $f_{i,j,k}$ and $\hat{f}_{i,j,k}$ correspond to the true and predicted number of infectious in group i , cell c_j , and at t_k , respectively.

Model Assessment

MAAPE for null and alternative models for EP (left) and FC (right) scenarios.

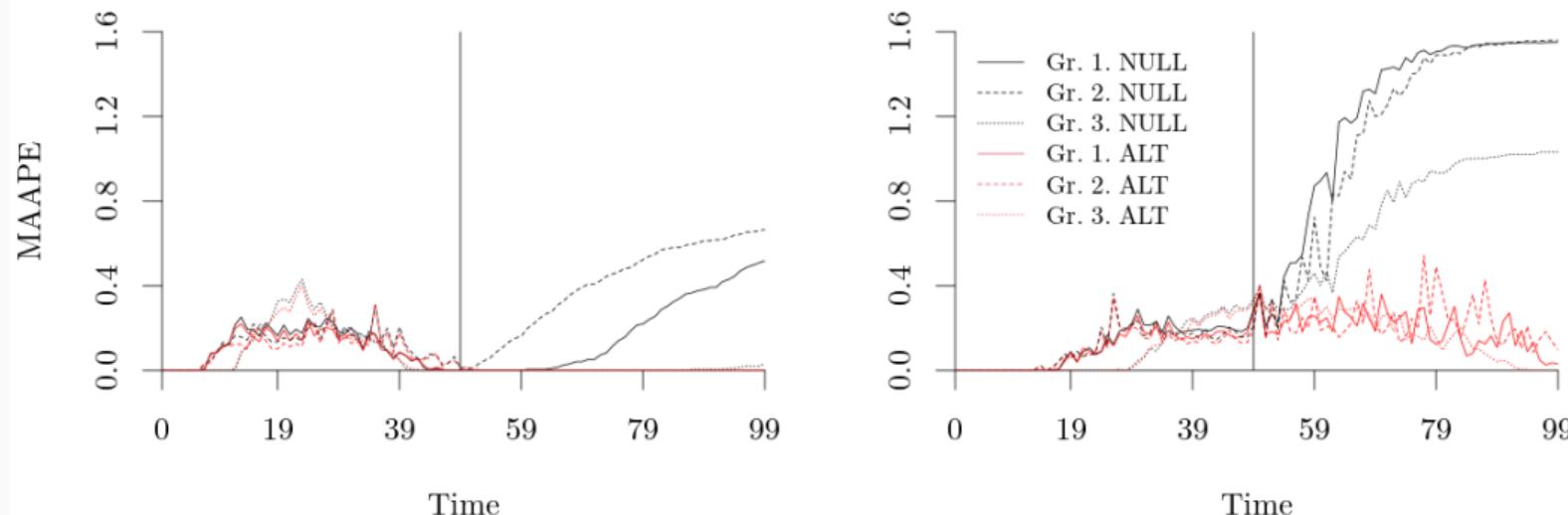


Figure 6: Computed MAAPEs for groups 1, 2, and 3 (0–19, 20–59, 60+, respectively). Models were fitted with data up to t_{49} (vertical solid line).

Discussion

We proposed a two-step framework for modeling data on the infectious locations and times as follows: first, using a compartment model; and second, incorporating such an information to a log-Gaussian Cox process model.

Discussion

We proposed a two-step framework for modeling data on the infectious locations and times as follows: first, using a compartment model; and second, incorporating such an information to a log-Gaussian Cox process model.

By doing this, provided the temporal model in the first step is well specified, the spatio-temporal modeling will produce reasonably good predictions.

Discussion

We proposed a two-step framework for modeling data on the infectious locations and times as follows: first, using a compartment model; and second, incorporating such an information to a log-Gaussian Cox process model.

By doing this, provided the temporal model in the first step is well specified, the spatio-temporal modeling will produce reasonably good predictions.

If SIR assumptions do not hold, step one can be modified to different statistical models. In this case, long-term predictions might not be as good, though.

Discussion

We proposed a two-step framework for modeling data on the infectious locations and times as follows: first, using a compartment model; and second, incorporating such an information to a log-Gaussian Cox process model.

By doing this, provided the temporal model in the first step is well specified, the spatio-temporal modeling will produce reasonably good predictions.

If SIR assumptions do not hold, step one can be modified to different statistical models. In this case, long-term predictions might not be as good, though.

Also, we used the exact locations and times of infected individuals. These data may be challenging to obtain. However, if good quality data is available, we think the proposed model can help understand infectious disease spreading.

References

- Diggle, P. J. (2006). Spatio-temporal point processes: methods and applications. *Monographs on Statistics and Applied Probability*, 107:1.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721.
- Kim, S. and Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.
- Stan Development Team (2021). RStan: the R interface to Stan. R package version 2.21.3.

Omitted Details

From Equation (4), $\zeta_i(\mathbf{u}; t_k)$ can be defined in different ways. However, throughout this presentation, we considered the following

$$\zeta_i(\mathbf{u}; t_k) = \beta_{0,i} + \vartheta_i(\mathbf{u}; t_k) + v_i(t_k) + \varepsilon_i(\mathbf{u}; t_k), \quad \forall i, \quad (7)$$

where ϑ_i is a zero-mean Gaussian process with covariance function $\sigma_i^2 \rho(h; t_k)$, such that $\rho(h; t_k)$ will be given by the Matérn model,

$$v_i(t_k) = \varrho_i(t_k), \text{ for } k = 0$$

$$v_i(t_k) = \alpha_i v_i(t_{k-1}) + \varrho_i(t_k), \text{ for } k = 1, \dots, n,$$

where $|\alpha_i| < 1$, $\forall i$, and $\varrho_i(t_k)$ is a zero-mean temporally independent Gaussian process with variance $\sigma_{i,\varrho}^2$, and $\varepsilon_i(\mathbf{u}; t_k)$ is a zero-mean temporally independent Gaussian process with variance $\sigma_{i,\varepsilon}^2$. In that case, $\mathbb{E}(\zeta_i(\mathbf{u}; t_k)) = \beta_{0,i}$.