Joint Reproduction Number and Spatial Connectivity Structure Estimation via Graph Sparsity-Promoting Penalized Functional

E. Lasalle[1], B. Pascal[1,†]

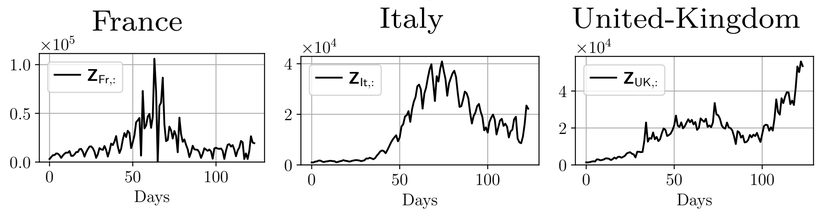
† Partially supported by Grant ANR-23-CE48-0009 "*OptiMoCSI*"

1. Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France firstname.lastname@ls2n.fr

Aims and contributions

During an *epidemic outbreak*, decision makers crucially need accurate and robust tools to monitor the pathogen propagation. The *effective reproduction number* R, defined as the expected number of secondary infections stemming from one contaminated individual, quantifies the epidemic intensity. Tracking the *spatiotemporal* evolution of R during COVID-19 pandemic raised unprecedented challenges due to the *poor quality* of reported infection counts (Pascal *et al.*, 2022, *IEEE Trans. Signal Process.*).

COVID-19 counts from Sept. 1, 2020 to Oct. 1, 2021



Source: https://coronavirus.jhu.edu/

Challenge: Infer and leverage epidemiological data's *spatial* structure to get accurate and robust R estimates.

Contributions: *joint* estimator of the reproduction number and connectivity structure

- procedure to generate *realistic* synthetic spatiotemporal epidemiological counts with *prescribed* connectivity pattern under a *scaled Poisson* model;
- spatial structure inferred via an *original* graph Laplacian estimation under Poisson *data-dependent* noise;
- *accuracy* of the joint estimator assessed through intensive numerical simulation on synthetic data, and illustration on *real* COVID-19 data.

Outcome: Data-driven algorithm to estimate *spatiotemporal reproduction numbers* from low quality counts https://github.com/elasalle/EpiJointSpatiotempEstim.

Spatiotemporal model

Epidemic monitoring: *spatiotemporal* framework

- country $c \in \{1, \dots, C\}$, $C \in \mathbb{N}^*$,
- day $t \in \{1, \dots T\}$, $T \in \mathbb{N}^*$,

reported infections: $\mathsf{Z}_{c,t} \in \mathbb{N} \Longrightarrow$ goal: infer $\mathsf{R}_{c,t}$

Scaled Poisson model: $\{\gamma_c > 0,\, c = 1, \dots, C\}$

$$\forall (c,t),\ \frac{\mathsf{Z}_{c,t}}{\gamma_c} | \mathsf{Z}_{c,1}, \dots, \mathsf{Z}_{c,t-1}; \mathsf{R}_{c,t} \sim \mathcal{P}\left(\frac{\mathsf{R}_{c,t}\Phi^{\mathsf{Z}}_{c,t}}{\gamma_c}\right)$$

where $\Phi^{\mathsf{Z}}_{c,t} = \sum_{s=1}^{\tau_\Phi} \phi(s)\mathsf{Z}_{c,t-s}$, ϕ: serial interval

$\{\mathbf{Z}_{c,\cdot},\, c = 1, \dots, C\}$ *independent* random vectors

(Cori *et al.*, 2013, *Am. J. Epidemiol.*; Abry *et al.*, 2020, *PLoS One*; Pascal *et al.*, 2025, *Signal Process.*)

COVID-19: ϕ fitted by a Gamma distribution

- of mean 6.6 days,
- and standard deviation 3.5 days.

Regularized spatiotemporal estimator:
enforces a realistic, hence smooth, behavior

- in time (Cori *et al.*, 2013),
- and across *connected* countries

given a *fixed* connectivity structure (Abry *et al.*, 2020, *PLoS One*; Pascal *et al.*, 2025, *Signal Process.*.)

$$\widehat{\mathbf{R}}^{\mathsf{fix-L}} = \underset{\mathbf{R} \in \mathbb{R}_+^{C \times T}}{\operatorname{argmin}} \sum_{c=1}^{C} \omega_c \mathsf{D}_{\mathsf{KL}}\left(\mathbf{Z}_{c,:} \middle| \mathbf{R}_{c,:} \odot \mathbf{\Phi}^{\mathbf{Z}}_{c,:}\right) + \sum_{c=1}^{C} \lambda_{\mathsf{T}} \|\mathsf{D}_2 \mathbf{R}_{c,:}\|_1 + \lambda_{\mathsf{S}} \sum_{t=1}^{T} \|\mathsf{B}\mathbf{R}_{:,t}\|_2^2$$

$$\mathsf{D}_{\mathsf{KL}}\left(\mathbf{Z}_{c,:} | \mathbf{R}_{c,:} \odot \mathbf{\Phi}^{\mathbf{Z}}_{c,:}\right) = \sum_{t=1}^{T} \mathsf{d}_{\mathsf{KL}}\left(\mathsf{Z}_{c,t} | \mathsf{R}_{c,t}\Phi^{\mathsf{Z}}_{c,t}\right)$$

with ideal weights $\omega_c > 0$: $\omega_c = 1/\gamma_c$

Heuristic: $\omega_c = 1/\mathrm{std}\left(\mathbf{Z}_{c,:}\right)$

$\mathsf{D}_2 \in \mathbb{R}^{T-2 \times T}$: second-order difference operator

$\mathsf{B} \in \mathbb{R}^{E \times C}$ incidence matrix: *connectivity*

$\Longrightarrow$ E *edges* between the C countries: $e = \{c, c'\}$ if $c < c'$, $\mathsf{B}_{e,c} = -\mathsf{B}_{e,c'} = \sqrt{\mathsf{W}_{c,c'}}$: weight of e

Example: $W_{c,c'} > 0$ iff terrestrial borders $c \sim c'$

Graph Total Variation: Laplacian $\mathsf{L} = \mathsf{B}^\top\mathsf{B}$

$$\text{at fixed time } t\text{: } \|\mathsf{B}\mathbf{R}_{:,t}\|_2^2 = \mathbf{R}_{:,t}^\top \mathsf{L} \mathbf{R}_{:,t}$$

smoothness of $c \mapsto \mathbf{R}_{c,t}$ in the graph's topology

If $\widetilde{\mathsf{B}} \in \mathbb{R}^{C \times C}$ s.t. $\widetilde{\mathsf{B}}^\top\widetilde{\mathsf{B}} = \mathsf{L}$: *unchanged* GTV.

Algorithmic resolution: convex, *non-smooth*

$$\text{minimize}\left\{F(\mathbf{R}) + G(\mathsf{K}\mathbf{R}) : \mathbf{R} \in \mathbb{R}_+^{C \times T}\right\}$$

- $F(\mathbf{R}) = \sum_{c=1}^{C} \omega_c \mathsf{D}_{\mathsf{KL}}\left(\mathbf{Z}_{c,:} \middle| \mathbf{R}_{c,:} \odot \mathbf{\Phi}^{\mathbf{Z}}_{c,:}\right)$
- $\mathsf{K}\mathbf{R} = (\mathbf{Q}_{\mathsf{T}}, \mathbf{Q}_{\mathsf{S}}) = \mathbf{Q}$
- $\mathbf{Q}_{\mathsf{T}} = \left(\mathsf{D}_2\mathbf{R}_{1,:}, \dots, \mathsf{D}_2\mathbf{R}_{C,:},\right)$
- $\mathbf{Q}_{\mathsf{T}} = \left(\mathsf{B}\mathbf{R}_{:,1}, \dots, \mathsf{B}\mathbf{R}_{:,T}\right)$
- $G(\mathbf{Q}_{\mathsf{T}}, \mathbf{Q}_{\mathsf{S}}) = \lambda_{\mathsf{T}}\|\mathbf{Q}_{\mathsf{T}}\|_1 + \lambda_{\mathsf{S}}\|\mathbf{Q}_{\mathsf{S}}\|_2^2$

$\Longrightarrow$ Chambolle-Pock scheme until: precision ε or $k_{\max}$ iterations (Chambolle *et al.*, 2011, *J. Math. Imag. Vis.*; Pascal *et al.*, 2022, *IEEE Trans. Signal Process.*; Du *et al.*, 2023, *GRETSI*)

Joint estimation strategy

Regularized and sparsity-inducing objective function: $\widehat{\mathbf{R}}^{\mathsf{Joint}}, \widehat{\mathsf{L}}^{\mathsf{Joint}} \in \underset{\mathbf{R} \in \mathcal{R}, \mathsf{L} \in \mathcal{L}}{\operatorname{Argmin}}\, J(\mathbf{R}, \mathsf{L})$ with

$$J(\mathbf{R}, \mathsf{L}) = \sum_{c=1}^{C} \omega_c \mathsf{D}_{\mathsf{KL}}\left(\mathbf{Z}_{c,:} \middle| \mathbf{R}_{c,:} \odot \mathbf{\Phi}^{\mathbf{Z}}_{c,:}\right) + \lambda_{\mathsf{T}} \sum_{c=1}^{C} \|\mathsf{D}_2\mathbf{R}_{c,:}\|_1 + \lambda_{\mathsf{S}} \sum_{t=1}^{T} \mathbf{R}_{:,t}^\top \mathsf{L} \mathbf{R}_{:,t} + \lambda_{\mathsf{L}} \|\mathsf{L}\|_{\mathrm{Fro}}^2;\ \|\mathsf{L}\|_{\mathrm{Fro}}^2 = \sum_{c,c'} \mathsf{L}_{c,c'}^2$$

with admissible sets $\mathcal{R} = \{\mathbf{R} \in \mathbb{R}^{C \times T},\ \text{s.t. } \forall c, t\ \mathbf{R}_{c,t} \geq 0\}$: *non-negativity of reproduction numbers*,

$\mathcal{L} = \left\{\mathsf{L} \in \mathbb{R}^{C \times C},\ \text{s.t. } \mathrm{Trace}(\mathsf{L}) = C,\ \forall c \neq c'\ \mathsf{L}_{c,c'} = \mathsf{L}_{c',c} \leq 0 \text{ and } \mathsf{L}\mathbf{1}_C = \mathbf{0}_C\right\}$:
proper Laplacian matrix with fixed total weight, excluding trivial empty graph.

J continuous, lower-bounded, coercive in $\mathbf{R}$ and defined on a compact set $\mathcal{L}$ for L has a *minimizer*.

Minimization w.r.t. $\mathbf{R}$ at fixed L: $\widetilde{\mathsf{B}} \in \mathbb{R}^{C \times C}$ s.t. $\widetilde{\mathsf{B}}^\top\widetilde{\mathsf{B}} = \mathsf{L}$, $\widehat{\mathbf{R}}^{\mathsf{fix-L}}$ from Chambolle-Pock scheme

Minimization w.r.t. L at fixed $\mathbf{R}$:

$$\widehat{\mathsf{L}}^{\mathsf{fix-R}} = \underset{\mathsf{L}}{\operatorname{argmin}}\ \lambda_{\mathsf{L}} \sum_{c,c'=1}^{C} \mathsf{L}_{c,c'}^2 + \lambda_{\mathsf{S}} \sum_{c,c'=1}^{C} (\mathbf{R}_{c,:}^\top \mathbf{R}_{c',:}) \mathsf{L}_{c,c'},$$

$$\text{s.t. } \mathrm{Trace}(\mathsf{L}) = C,\ \mathsf{L}_{c,c'} = \mathsf{L}_{c',c} \leq 0,\ \forall c \neq c',\ \mathsf{L}\mathbf{1}_C = \mathbf{0}_C.$$

- non-negative, continuous, coercive, strictly convex $\Longrightarrow$ *unique minimizer*
- quadratic program, linear constraints: *interior point*

CVXOPT: Andersen *et al.*, 2013, *cvx.org*

Alternating optimization on $\mathbf{R}$, L:

Input: $\mathbf{Z} \in \mathbb{R}^T$, $(\lambda_{\mathsf{T}}, \lambda_{\mathsf{S}}, \lambda_{\mathsf{L}}) \in \mathbb{R}_+^3$, $n_{\max} \in \mathbb{N}^*$, $\mathbf{R}^{(0)}$;
$\mathsf{L}^{(0)}, \mathsf{M}^{(0)} \leftarrow \mathtt{minimizeL}\left(\mathbf{R}^{(0)}, \lambda_{\mathsf{S}}, \lambda_{\mathsf{L}}\right)$
$\widetilde{\mathsf{B}}^{(0)} \leftarrow \mathtt{Cholesky}(\mathsf{L}^{(0)})$; define K using $\widetilde{\mathsf{B}}^{(0)}$
$\mathbf{Q}^{(0)} \leftarrow \mathsf{K}\mathbf{R}^{(0)}$
for $n = 0$ **to** $n_{\max} - 1$ **do**
$\mathbf{R}^{(n+1)}, \mathbf{Q}^{(n+1)} = \mathtt{minR}\left(\mathbf{Z}, \lambda_{\mathsf{T}}, \lambda_{\mathsf{S}}, \widetilde{\mathsf{B}}^{(n)}, \mathbf{R}^{(n)}, \mathbf{Q}^{(n)}\right)$
$\mathsf{L}^{(n+1)}, \mathsf{M}^{(n+1)} = \mathtt{minL}\left(\mathbf{R}^{(n+1)}, \lambda_{\mathsf{S}}, \lambda_{\mathsf{L}}, \mathsf{L}^{(n)}, \mathsf{M}^{(n)}\right)$
$\widetilde{\mathsf{B}}^{(n+1)} = \mathtt{Cholesky}(\mathsf{L}^{(n+1)})$
end for
Output: number $\mathbf{R}^{(n_{\max})}, \mathsf{L}^{(n_{\max})}$

Numerical experiments

Synthetic multivariate scaled Poisson counts: strictly respect a *prescribed* spatial structure L

Proposed generation: C *abstract* countries partitioned into I *clusters* $\mathcal{C}_1, \dots, \mathcal{C}_I$, e.g., $C = 9$, $I = 3$;

- for each cluster $\mathcal{C}_i$, one realistic synthetic COVID-19 reproduction number $\mathbf{R}^{\dagger}_{i,:} \in \mathbb{R}_+^T$, (Du *et al.*, 2023, *GRETSI*; Du *et al.*, 2024, *EUSIPCO*)
- to enforce spatial structure: in all countries of $\mathcal{C}_i$ *ground truth* $\forall c \in \mathcal{C}_i$, $\mathbf{R}^{\star}_{c,:} = \mathbf{R}^{\dagger}_{i,:}$,
- initial counts $\mathbf{Z}_{:,0}$ fixed, and scale parameters $\gamma_c = 0.01 \times \mathsf{Z}^0_c$ allow variability *inside clusters*,

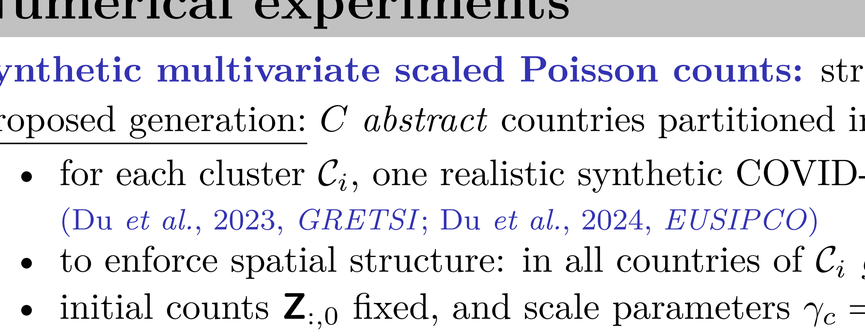
$\Longrightarrow$ synthesize scaled Poisson spatiatemporal counts $\mathbf{Z}_{c,:}$ with ground truth $\mathbf{R}^{\star}_{c,:}$ independently on c.

Experimental protocol: mean Relative Squared Error $\mathsf{mRSE}(\mathbf{R}) = \dfrac{1}{CT} \displaystyle\sum_{c=1}^{C} \sum_{t=1}^{T} \left(\frac{\widehat{\mathsf{R}}_{c,t} - \mathsf{R}^{\star}_{c,t}}{\mathsf{R}^{\star}_{c,t}}\right)^2$

- naive Maximum Likelihood estimator ML of scaled Poisson model,
- state-of-the-art epidemiological estimator $\mathsf{EpiEstim}$, (Cori *et al.*, 2013, *Am. J. Epidemiol.*)
- $\mathsf{fix-L}$ with: an empty graph $\mathsf{L}_\emptyset$, *blurred* version L_{b} of the true graph, and the true graph $\mathsf{L}^\star$,

Regularization parameters: optimized through *grid search*.

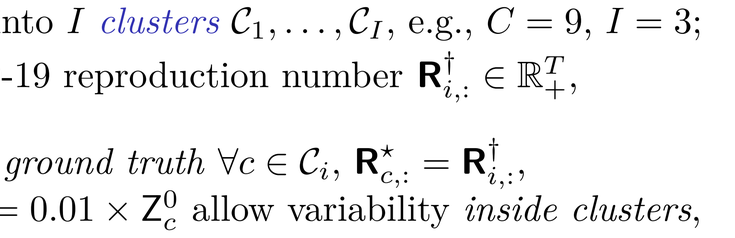
Performance: average and 95% Gaussian confidence intervals of mRSE over 20 *realizations*.



ML	EpiEstim	fix−$\mathsf{L}_\emptyset$	fix−L_{b}	fix−$\mathsf{L}^\star$	Joint (ours)
113.2(±4.5)	27.7(±1.2)	9.0(±1.1)	5.4(±0.6)	**2.6(±0.4)**	**2.7(±0.4)**

$$\|\widehat{\mathsf{L}}^{\mathsf{Joint}} - \mathsf{L}^\star\|_{\mathrm{Fro}}^2 / \|\mathsf{L}^\star\|_{\mathrm{Fro}}^2 = 2.3(\pm 1.0) \times 10^{-13}$$

Real COVID-19 counts: public data from *Johns Hopkins University repository*
https://coronavirus.jhu.edu/



- 39 countries in Europe and Africa
- from Sept. 1, 2020 to Oct. 1, 2021
- manual fine-tuning of hyperparameters

Conclusions & perspectives

- joint estimator of reproduction numbers and spatial connectivity structure proposed
- numerical experiments assessing accuracy of reproduction number and spatial structure estimates
- ▶ future exploration of its robustness to changes in monitored territories and/or phase of the pandemic