Monitoring the temporal and spatio-temporal dynamics of the effective reproduction number from mortality data

L Roques, BioSP, INRAE

With :
Virgile Baudrot, Olivier Bonnefon, Samuel Soubeyrand, BioSP INRAE
Jacques Demongeot, UGA, Faculty of Medicine of Grenoble
Henri Berestycki, CAMS, EHESS

INRAE

Biostatistique BIOSP & Processus Spatiaux

Part I: Estimating R_t from mortality data, a simple formula

ELSEVIER

One Health

Available online 24 October 2020, 100187

In Press, Journal Pre-proof

Towards unified and real-time analyses of outbreaks at country-level during pandemics

Samuel Soubeyrand [a], Jacques Demongeot [b], Lionel Roques [a]

Show more

https://doi.org/10.1016/j.onehlt.2020.100187

Samuel Soubeyrand

Jacques Demongeot

Main ideas

- SIRD model with constant parameters over moving window $(t-\tau, t)$
- Work with daily mortality data (cumulated value $\hat{D}_t$)

$$\begin{cases} S'(s) = -\dfrac{\alpha_s}{N} S\, I \\ I'(s) = \dfrac{\alpha_s}{N} S\, I - (\beta+\gamma)\, I \\ R'(s) = \beta\, I \\ D'(s) = \gamma\, I \end{cases} \xrightarrow{5\cdot 10^{-4} \approx \gamma \ll \beta \approx 0.1} I'(s) = \beta\, I(s) \left(\frac{\alpha_s\, S}{\beta\, N} - 1 \right)$$

$\mathcal{R}_s = \alpha_s S(s)/(\beta\, N)$ is constant $(= \mathcal{R}_t)$ for $s \in (t-\tau, t)$

$$\begin{cases} I'(s) = \beta\, I(s)\, (\mathcal{R}_t - 1) \\ D'(s) = \gamma\, I \end{cases}$$

$$D(s) = D(t-\tau) + \gamma\, I(t-\tau) \int_{t-\tau}^{s} e^{\beta\, u\, (\mathcal{R}_t - 1)}\, du$$

$D(t-\tau) = \hat{D}_{t-\tau}$ and $D(t) = \hat{D}_t$

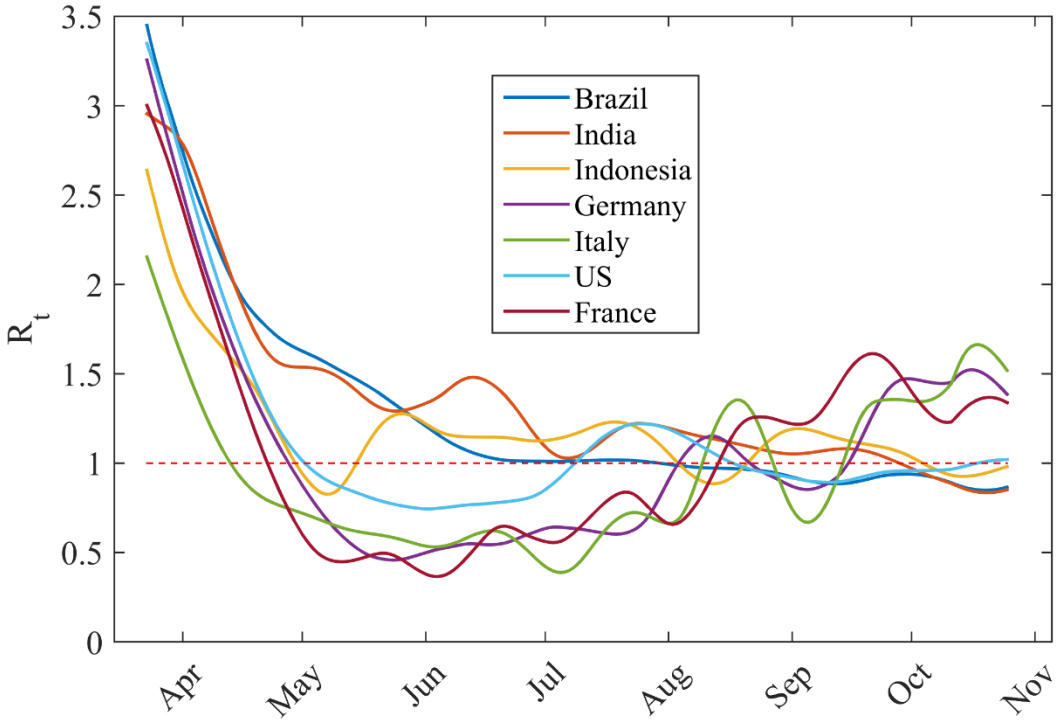
$$D(s) = \hat{D}_{t-\tau} + (\hat{D}_t - \hat{D}_{t-\tau}) \frac{e^{\beta\, (R_t - 1)\, (s - (t-\tau))} - 1}{e^{\beta\, \tau\, (R_t - 1)} - 1}, \text{ for all } s \in (t-\tau, t).$$

$\rightarrow$ $D(s)$ does not depend on the mortality rate γ and on $I(t-\tau)$

$$D(s) = \hat{D}_{t-\tau} + (\hat{D}_t - \hat{D}_{t-\tau}) \frac{e^{\beta\,(R_t - 1)\,(s-(t-\tau))} - 1}{e^{\beta\,\tau\,(R_t-1)} - 1}, \text{ for all } s \in (t-\tau, t).$$

For each t, $\mathcal{R}_t$ is computed with a standard (mean square) fitting procedure, over *smoothed* mortality data.

Example with $\tau = 28$:

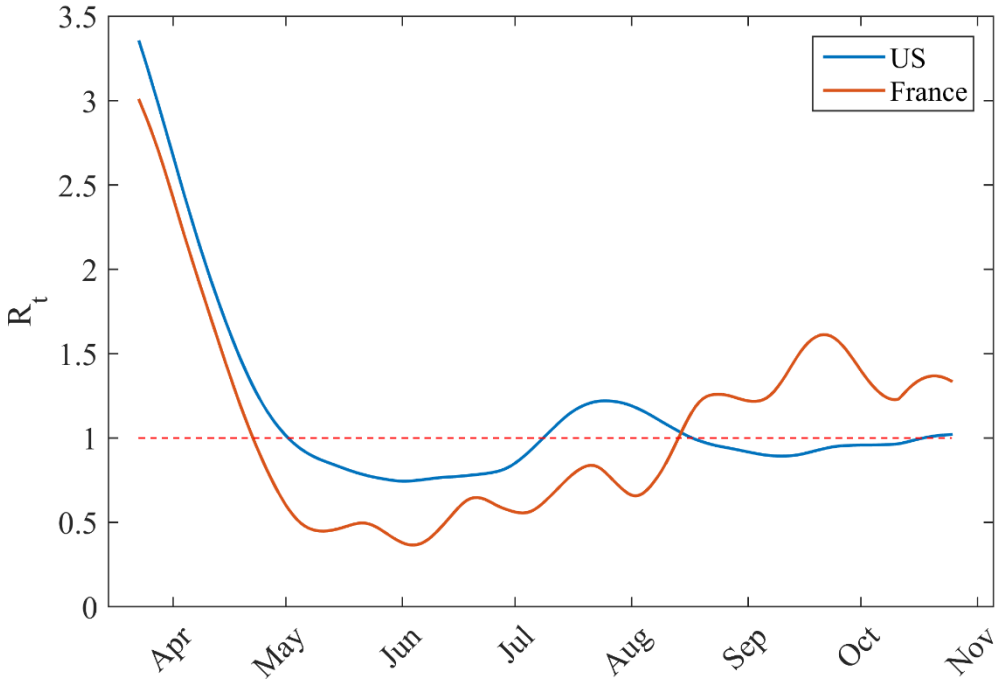


Data: JHU CSSE, 27 October 2020

$$D(s) = \hat{D}_{t-\tau} + (\hat{D}_t - \hat{D}_{t-\tau}) \frac{e^{\beta\,(R_t - 1)\,(s-(t-\tau))} - 1}{e^{\beta\,\tau\,(R_t-1)} - 1}, \text{ for all } s \in (t-\tau, t).$$

For each t, $\mathcal{R}_t$ is computed with a standard (mean square) fitting procedure, over *smoothed* mortality data.

Example with $\tau = 28$:



Data: JHU CSSE, 27 October 2020

October 27: $\mathcal{R}_t(\text{France}) \approx 1.3$ vs $\mathcal{R}_t(\text{US}) \approx 1.0$

$$PFR := \frac{number\ of\ deaths}{total\ population}$$

October 27: $PFR(\text{France}) = 51/100\,000$ vs $PFR(\text{US}) = 69/100\,000$

$$\rightarrow PFR(\text{US}) \approx 1.3 \times PFR(\text{France})$$

Higher immunity rate in the US explains the lower $\mathcal{R}_t$?

Part II: Spatio-temporal dynamics of COVID-19 in France

local effects vs global factors

Olivier Bonnefon

Virgile Baudrot

Samuel Soubeyrand

Henri Berestycki, CAMS, EHESS

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

HOME | ABOUT

Search

Comment on this paper

A parsimonious model for spatial transmission and heterogeneity in the COVID-19 propagation

Lionel Roques, Olivier Bonnefon, Virgile Baudrot, Samuel Soubeyrand, Henri Berestycki

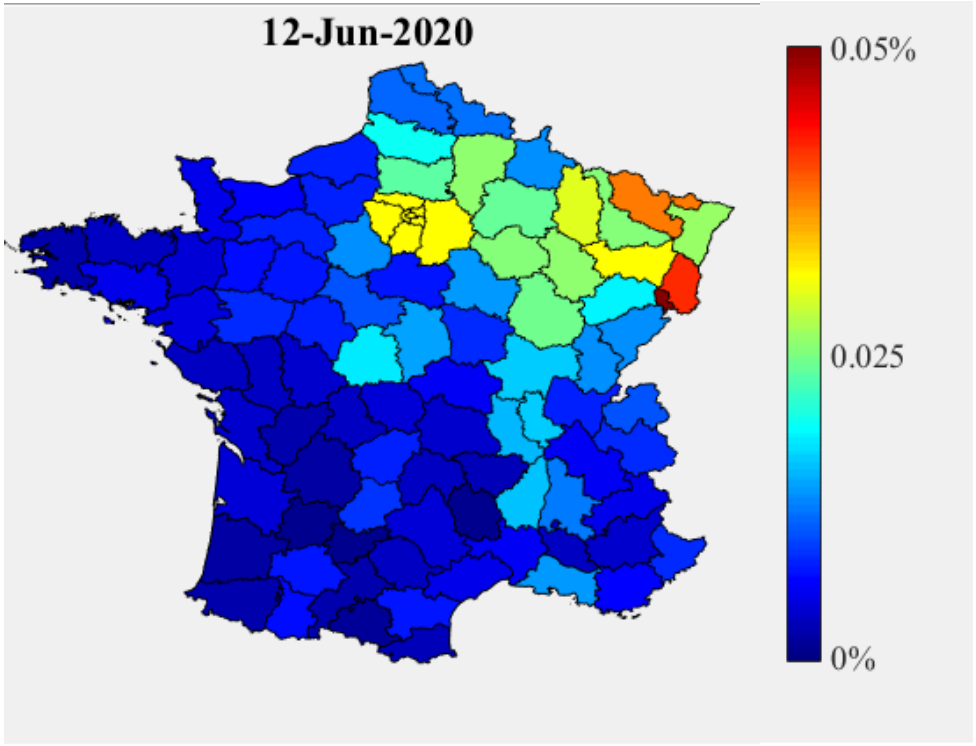
doi: https://doi.org/10.1101/2020.07.15.20154740

This article is a preprint and has not been certified by peer review [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

January 24, 2020: 1st reported cases in France. Some cases already present in December 2019 (Retrospective PCR tests, [Deslandes et al. 2020]).

In June, the spatial pattern of the disease spread seems to have kept track of the first introductions

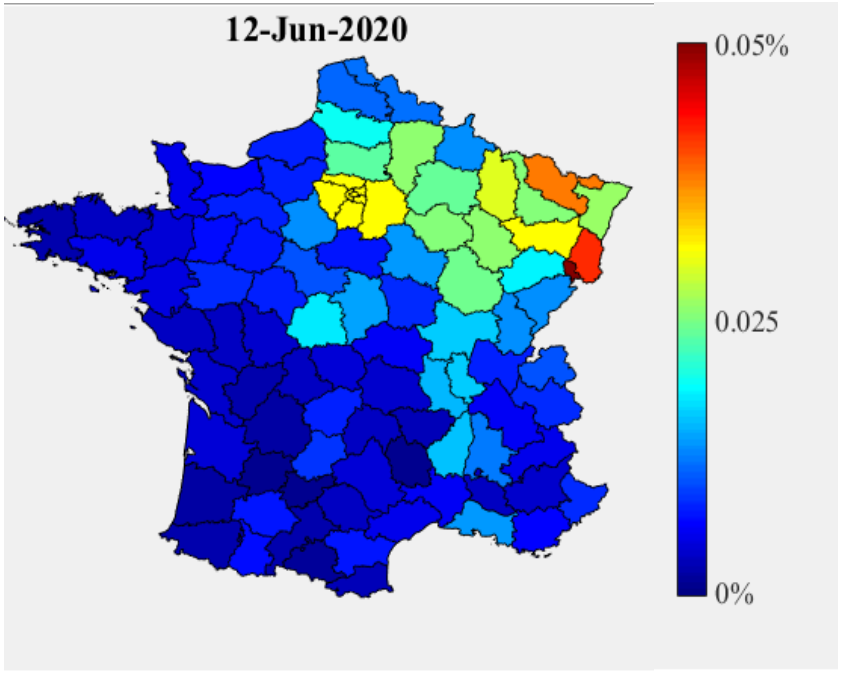
Local PFR at county scale



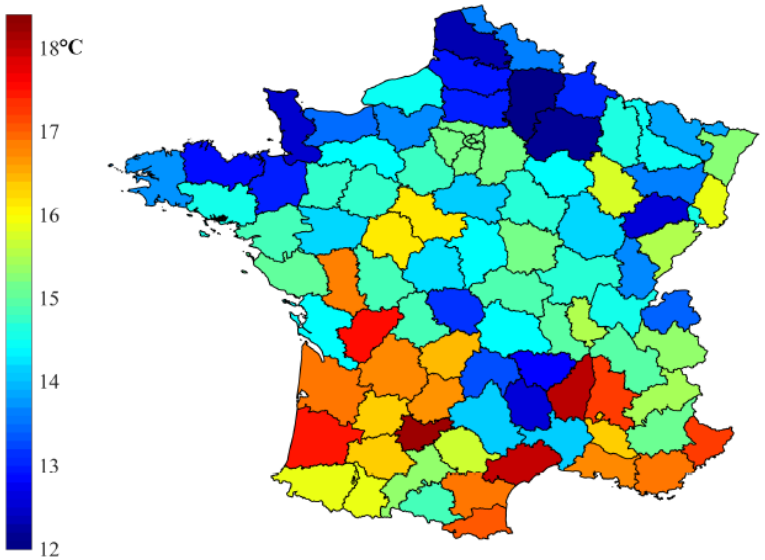
$$PFR := \frac{number\ of\ deaths}{total\ population}$$

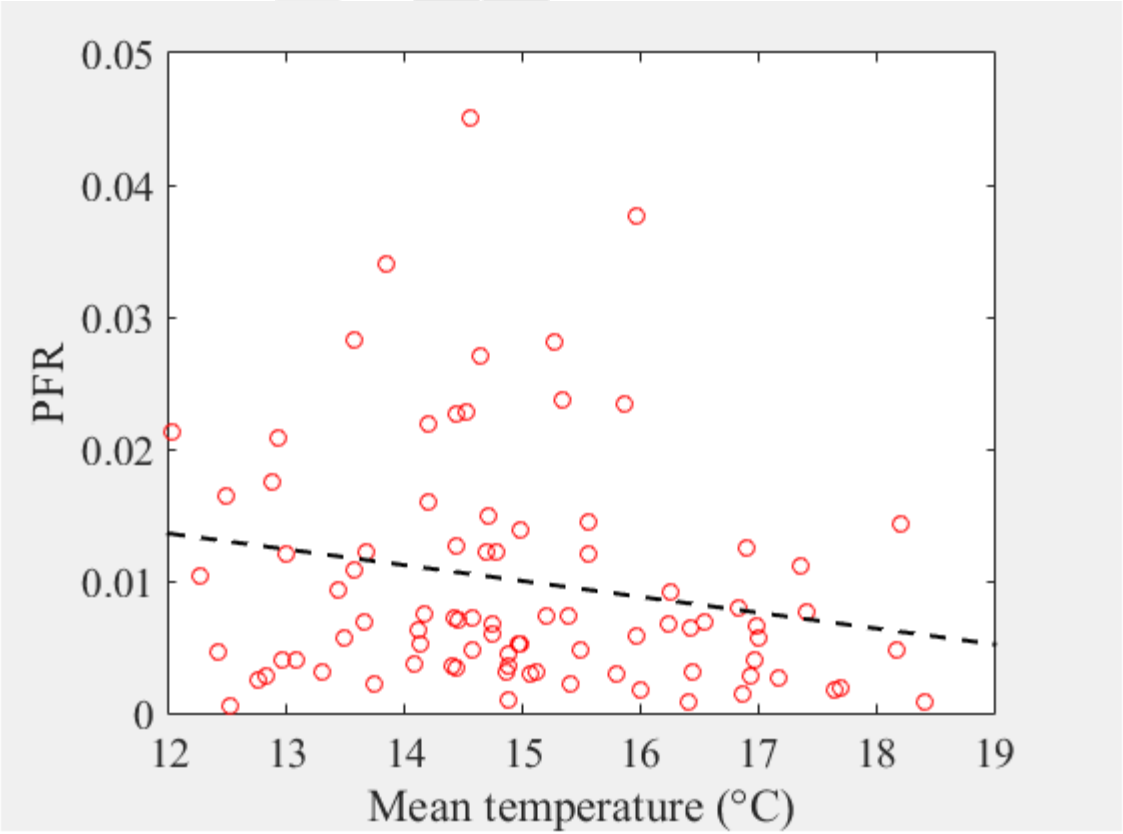
This spatial pattern may also be correlated with covariates such as climate [Demongeot et al. 2020]

PFR



Mean temperature (30 March -11 June)





Pearson correlation coefficient: -0.21.

Is this pattern the consequence of the heterogeneity of spatial covariates?

Does it simply reflect the initial heterogeneity?

Data

Mainland France (excluding Corsica island) = 94 counties called 'départements'.

Daily number of hospital deaths excluding nursing homes at the county scale are available from Santé Publique France

Models

Model $\mathcal{M}_0$: Standard mean field SIRD model at country scale

$$\begin{cases} S'(t) = -\dfrac{\alpha(t)}{N}\, S\, I, \\ I'(t) = \dfrac{\alpha(t)}{N}\, S\, I - (\beta + \gamma)\, I, \\ R'(t) = \beta\, I, \\ D'(t) = \gamma\, I, \end{cases}$$

- $\alpha(t)$: the contact rate (to be estimated)
- $1/\beta$ is the mean time until an infectious becomes recovered: 10 days

  $\rightarrow$ Median period of viral shedding: 20 days

  $\rightarrow$ infectiousness tends to decay before the end of this period: starts from 2.5 days before symptom onset and declines within 7 days of illness onset.
- γ: hospital death rate of the infectious. The IFR corresponds to the fraction of the infected who die, that is: $\gamma/(\gamma + \beta)$. Thus, a value of 0.5% [Roques et al, 2020, Salje et al, 2020] for the IFR implies a value $\gamma = 5 \cdot 10^{-4}$.

Model $\mathcal{M}_1$: SIRD model at the 'département' scale with globally constant contact rate and no spatial transmission. The model $\mathcal{M}_0$ is applied at the scale of each département k, leading to compartments S_k, I_k, R_k, D_k, with N replaced by N_k, the total population in the département k.

In this approach, the contact rate $\alpha(t)$ is assumed to be the same in all of the départements.

Model $\mathcal{M}_2$: SIRD model at the 'département' scale with spatially heterogeneous contact rate and no spatial transmission.

The model $\mathcal{M}_1$ is extended by assuming that the contact rate $\alpha_k(t)$ depends on the considered département.

Model $\mathcal{M}_3$: Département scale model with globally constant contact rate and spatial transmission [Kendall, 1956]. The model $\mathcal{M}_1$ is extended to take into account disease transmission events between the départements:

$$\begin{cases} S_k'(t) = -\dfrac{\rho(t)}{N_k}\, S_k \displaystyle\sum_{j=1}^{n_d} w_{j,k}\, I_j, \\ I_k'(t) = \dfrac{\rho(t)}{N_k}\, S_k \displaystyle\sum_{j=1}^{n_d} w_{j,k}\, I_j - (\beta+\gamma)\, I_k, \\ R_k'(t) = \beta\, I_k, \\ D_k'(t) = \gamma\, I_k. \end{cases}$$

- Implicitly assumes that infectious individuals may transmit the disease to susceptible individuals in other départements, but eventually return to their département of origin.
- Power law decay with the distance: $w_{j,k} = \frac{1}{1+(\mathrm{dist}(j,k)/d_0)^\delta}$, with $\mathrm{dist}(j,k)$ the geographic distance (in km) between the centroids of départements j and k
- Supported by analysis of the short-time dispersal of bank notes in the US [Brockmann et al, 2006]
- Standard choice in infectious disease modelling [e.g., Merler and Ajelli 2009, for influenza]

Summary of the models

	Heterog. initial data	Heterog. contact rate	Intercounty transmission	Nb. parameters
$\mathcal{M}_0$	no	no	no	n_t
$\mathcal{M}_1$	yes	no	no	n_t
$\mathcal{M}_2$	yes	yes	no	$n_d \times n_t$
$\mathcal{M}_3$	yes	no	yes	$n_t + 2$

Table 1: Main characteristics of the four models. The quantity $n_t = 74$ corresponds to the number of days of the observation period and $n_d = 87$ corresponds to the number of administrative units.

Observation model. $\overline{D}_k(t)$: expected cumulative number of deaths given by the model, in département k. Daily number of new observed deaths $\hat{\mu}_{k,t}$ in département k:

$$\hat{\mu}_{k,t} \sim \text{Poisson}(\overline{D}_k(t) - \overline{D}_k(t-1)).$$

With the mean-field model $\mathcal{M}_0$, $\overline{D}_k(t) = D(t)\, N_k/N$.

Real-time monitoring of the parameters and data assimilation procedure.

-Smooth out the effect of small variations in the data

-Avoid identifiability/computational issues due to the large number of parameters, while keeping the temporal dependence of the parameters

$\rightarrow$ we fit auxiliary problems with time-constant parameters over moving windows $(t-\tau/2, t+\tau/2)$:

$$\begin{cases} \tilde{S}_k'(s) = -\dfrac{\tilde{\alpha}}{N_k}\,\tilde{S}_k\,\tilde{I}_k, \\ \tilde{I}_k'(s) = \dfrac{\tilde{\alpha}}{N_k}\,\tilde{S}_k\,\tilde{I}_k - (\beta+\gamma)\,\tilde{I}_k, \quad \text{for } s \in (t-\tau/2, t+\tau/2), \qquad (\tilde{\mathcal{M}}_{1,t}) \\ \tilde{R}_k'(s) = \beta\,\tilde{I}_k, \\ \tilde{D}_k'(s) = \gamma\,\tilde{I}_k, \end{cases}$$

Initial conditions at the date $t-\tau/2$ are computed iteratively from the solution of the model $\mathcal{M}_i$.

The parameters of the models $\tilde{\mathcal{M}}_i$ are estimated with a maximum likelihood procedure

Inference procedure: maximum likelihood estimation over $(t-\tau/2, t+\tau/2)$ for each t.

Probability of the observations $\hat{\mu}_k$, given the model:

$$f_{\overline{D}_k,\hat{\mu}_k}(s) := \frac{(\overline{D}_k(s) - \overline{D}_k(s-1))^{\hat{\mu}_{k,s}}}{\hat{\mu}_{k,s}!}\, e^{-(\overline{D}_k(s)-\overline{D}_k(s-1))}.$$

Models $\tilde{\mathcal{M}}_{0,t}$, $\tilde{\mathcal{M}}_{1,t}$: $\mathcal{L}(\tilde{\alpha}_t) = \prod_{k=1}^{n_d} \prod_{s=t-\tau/2}^{t+\tau/2} f_{\overline{D}_k,\hat{\mu}_k}(s)$.

MLE $\tilde{\alpha}_t^*$, we set $\alpha(t) = \tilde{\alpha}_t^*$ in models $\mathcal{M}_0,\ \mathcal{M}_1$.

Model $\tilde{\mathcal{M}}_{2,t}$, esimated parameters $\tilde{\alpha}_k$ (independently in each département).

$$\mathcal{L}_k(\tilde{\alpha}_k) = \prod_{s=t-\tau/2}^{t+\tau/2} f_{\overline{D}_k,\hat{\mu}_k}(s).$$

MLE $\tilde{\alpha}_{k,t}^*$, we set $\alpha_k(t) = \tilde{\alpha}_{k,t}^*$

Model $\mathcal{M}_3$, use the estimate obtained with model $\mathcal{M}_1$ with $\rho(t) = C\,\alpha(t)$. Estimate 3 parameters:

C, the proximity scale d_0 and the exponent δ (ML estimation).

Results: model fit

	Heterog. initial data	Heterog. contact rate	Intercounty transmission	Nb. parameters
$\mathcal{M}_0$	no	no	no	n_t
$\mathcal{M}_1$	yes	no	no	n_t
$\mathcal{M}_2$	yes	yes	no	$n_d \times n_t$
$\mathcal{M}_3$	yes	no	yes	$n_t + 2$

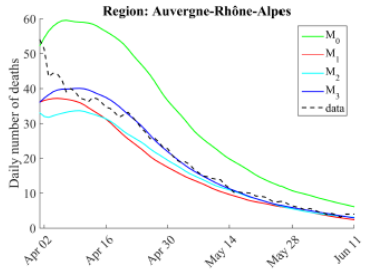
Table 1: Main characteristics of the four models. The quantity $n_t = 74$ corresponds to the number of days of the observation period and $n_d = 87$ corresponds to the number of administrative units.

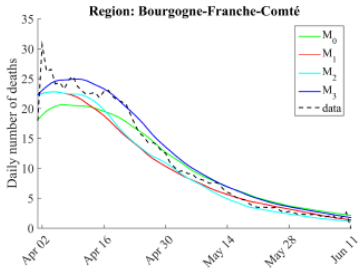
Model	AIC	BIC	Log-likelihood	ΔAIC
$\mathcal{M}_0$	$2.68 \cdot 10^4$	$2.73 \cdot 10^4$	$-13.4 \cdot 10^3$	$-9.57 \cdot 10^3$
$\mathcal{M}_1$	$1.74 \cdot 10^4$	$1.79 \cdot 10^4$	$-8.62 \cdot 10^3$	-220
$\mathcal{M}_2$	$2.97 \cdot 10^4$	$7.36 \cdot 10^4$	$-8.41 \cdot 10^3$	$-1.25 \cdot 10^4$
$\mathcal{M}_3$	$1.72 \cdot 10^4$	$1.77 \cdot 10^4$	$-8.52 \cdot 10^3$	0

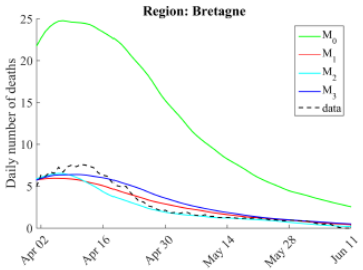
Table 2: Log-likelihood, AIC and BIC values for the four models. The last column ΔAIC corresponds to the difference with the AIC value of the best model (here $\mathcal{M}_3$).

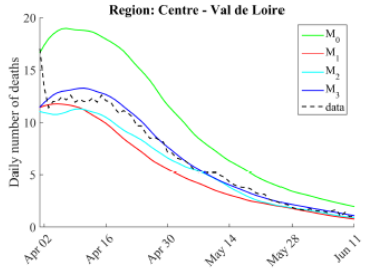
Statisticians often argue that AIC chooses the best predictive model, and BIC attempts to select the 'true' process that generated the data.

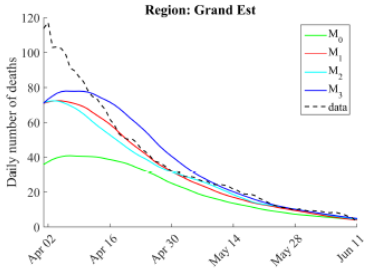
Results: model fit, regional scale

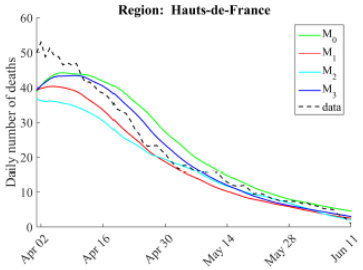


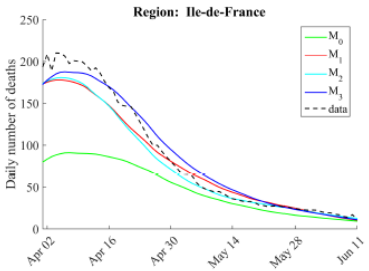


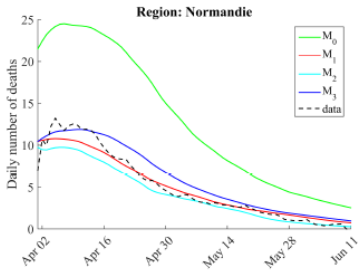


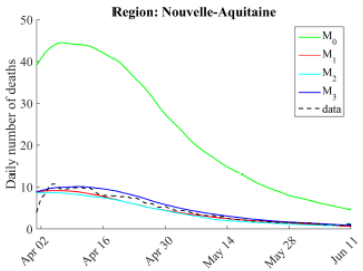


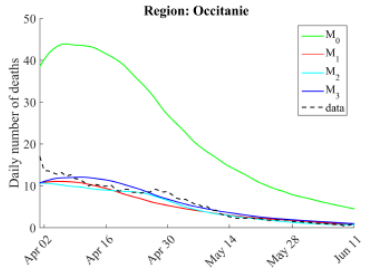


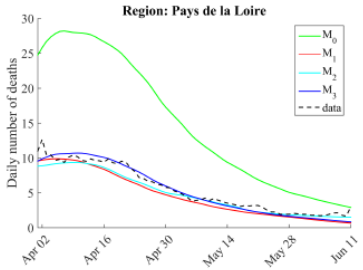


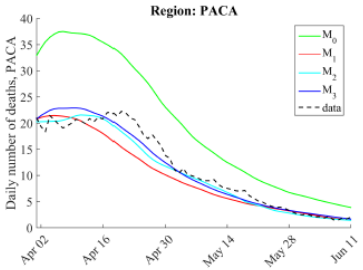












Results: model fit, parameter values

For model $\mathcal{M}_3$,
MLE: $C = 0.87$, $d_0 = 2.16$ km and $\delta = 1.85$

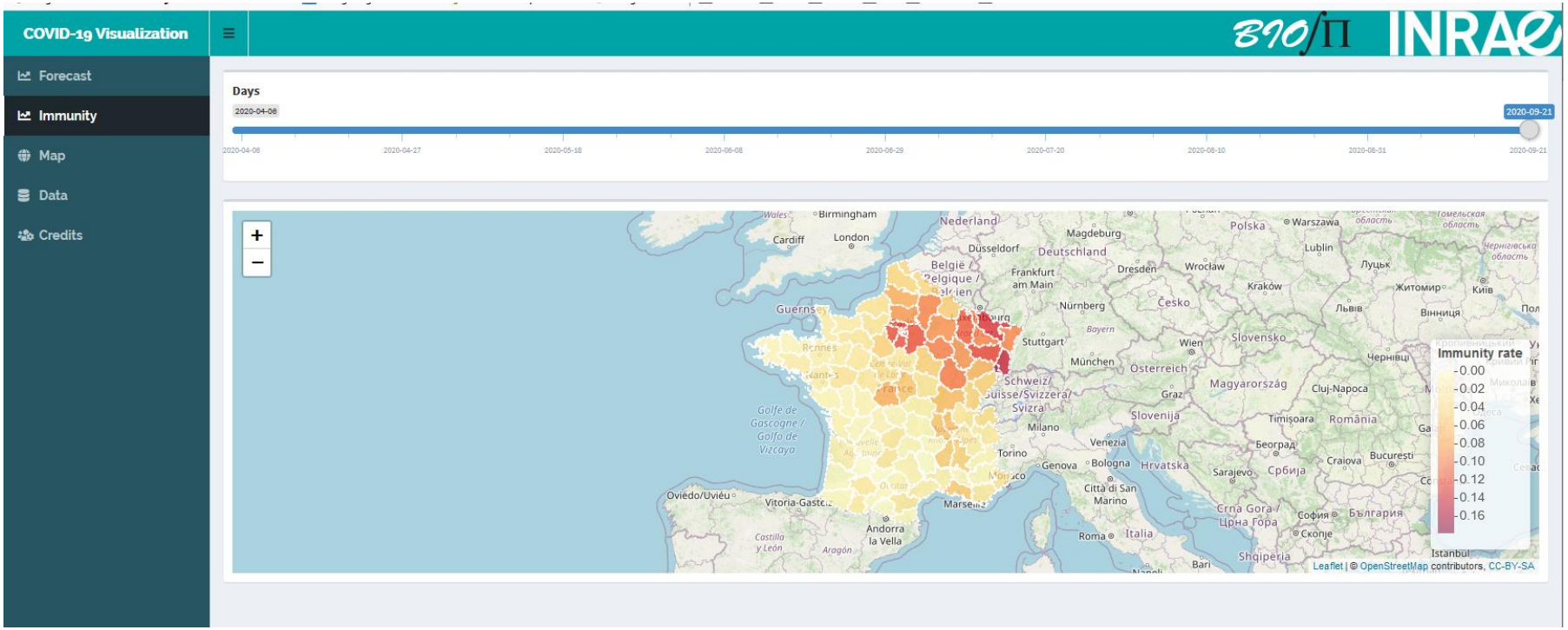
- nearly quadratic decay of the weights with the distance
- non-local contagion plays a secondary role compared to within-county contagion: the minimum distance between two counties is 36 km $\rightarrow$ weight 5.5/1000, to be compared with the weight 1 for within-county contagion.
- C is significantly smaller than 1 thus non-local contagion term plays an important role on the spreading of the epidemic.

Byproduct of the estimation of the parameter $\alpha(t)$ (resp. $\alpha_k(t)$) of model $\mathcal{M}_1$ effective reproduction number in each département, which is given by the formula:

$$\mathcal{R}_t^k = \frac{\alpha(t)}{\beta+\gamma}\frac{S_k(t)}{N_k},$$

Other byproduct: immunity rate R_k/N_k (model $\mathcal{M}_3$)

http://covid19-forecast.biosp.org/



Results: limiting movement vs limiting the probability of transmission per contact

We start from the state of the epidemic at June 11.
During 10 days, with model $\mathcal{M}_3$:

- 'local' contact rate $\rho(t) \approx 0.05 \rightarrow \rho(t) = 0.1$ $(\times 2)$
- $d_0 \approx 2$ km $\rightarrow d_0 = 20$ km $(\times 10)$

We test four strategies (during 30 days):

- Strategy 1: no restriction. $\rho(t) = 0.11$ and $d_0 = 20$ km;
- Strategy 2: restriction on intercounty movement. $\rho(t) = 0.11$, $d_0 = 2.16$ km, corresponding
- Strategy 3: reduction of the contact rate within each département. $\rho(t) = 0.05$ and $d_0 = 20$ km;
- Strategy 4: reduction of the contact rate and restriction on intercounty movement: $\rho(t) = 0.05$ and $d_0 = 2.16$ km.

We test four strategies:

- Strategy 1: no restriction.
- Strategy 2: restriction on intercounty movement.
- Strategy 3: reduction of the contact rate within each département.
- Strategy 4: reduction of the contact rate and restriction on intercounty movement

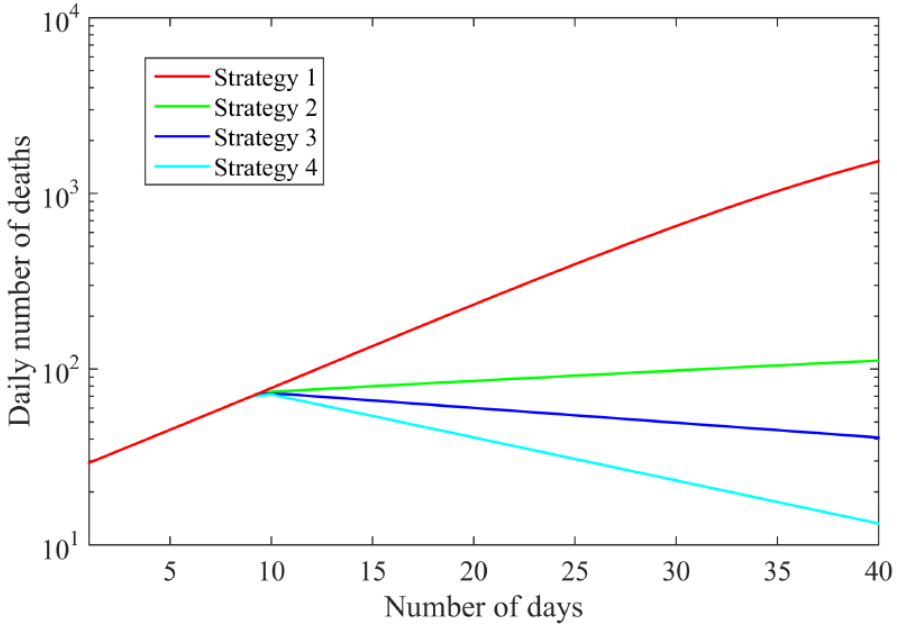


Figure 4: Daily number of deaths due to a new outbreak in logarithmic scale; comparison between four management strategies. The number of deaths is computed over the whole country.

Final number of deaths and $\mathcal{R}_t$ over the whole country estimated by fitting $I(t)$ with an exponential function ($I'(t) = (\beta + \gamma)I(\mathcal{R}_t - 1)$).

- Strategy 1: no restriction. 17 271 deaths and $\mathcal{R}_t \approx 2$.
- Strategy 2: restriction on intercounty movement. 81% decrease in the cumulative number of deaths (3 281 deaths) and $\mathcal{R}_t \approx 1.2$;
- Strategy 3: reduction of the contact rate within each département. 88% decrease (strategy 3, 2 139 deaths) and $\mathcal{R}_t \approx 0.8$;
- Strategy 4: reduction of the contact rate and restriction on intercounty movement. 91% decrease (1 503 deaths) and $\mathcal{R}_t \approx 0.4$.

To be compared with the 3 232 deaths during the last 30 days.

Conclusions

- Once the epidemic is established, the effect of global processes such as **restriction policies and sanitary measures overwhelms the effect of local factors.**

  $\rightarrow$ France is a geographically middle size centralised country

- Initial conditions and spatial diffusion are the main drivers of the spatial pattern of the COVID-19 epidemic

  $\rightarrow$ Covariates might play a major role in the emergence of the disease, but our work focuses on the disease dynamics after the emergence.

- Herd immunity (without masks) $1 - 1/\mathcal{R}_0 \approx 70\%$ of the population has been infected. Most impacted dépatements: the immunity rate is 16%, whereas it is less than 1% in less affected counties.

- Model $\mathcal{M}_3$ much more parsimonious than fully heterogeneous model $\mathcal{M}_2$

  $\rightarrow$ better suited to isolating key features of the epidemiological dynamics.

Thank you!