Towards Epidemiological Forecasting with Mobility Data

Olga Mula

(U. Paris Dauphine PSL & Inria Paris)

Infectious Disease Outbreaks – 24/02/2021

Works with:

- Sorbonne Univesité (A. Bakhta, T. Boiveau, Y. Maday)
- Paris Sciences Lettres – Initiative Face Au Virus
  (J. Atif, O. Cappé, A. Kazakci, Y. Leo, L. Massoulié)

Motivation

My expertise:

- applied mathematician:
  - model reduction of parametrized PDEs
  - inverse problems
- I am not expert in medical science

Scientific goals:

- develop forecasting methods for epidemiology
- combine seemingly relevant data: health series, mobility, population density, data from waste water, vaccination data.

Warning: I can change my mind and my algorithms if I become aware of more data.

Plan of the talk

1) Epidemiological Forecasting using Model Reduction

- Paper: Epidemiological Forecasting with Model Reduction of Compartmental Models. Application to the COVID-19 Pandemic. A. Bakhta, T. Boiveau, Y. Maday and O. Mula. Biology. 2021.

2) Mobility Data from Facebook

- Report 2: Impact of mobility and population density on the Covid-19 outbreak (February-Nov 2020).
- Report 1: Feedback on mobility during the Covid-19 epidemic (February-May 2020).

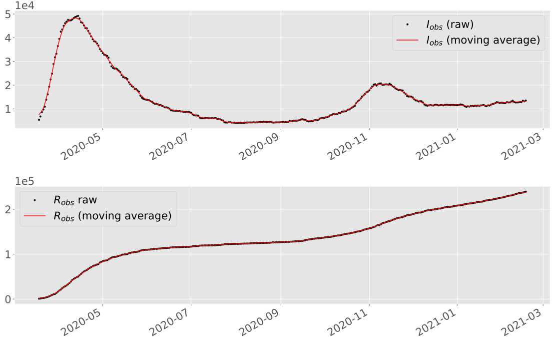
Epidemiological Forecasting using Model Reduction

Work with A. Bakhta, T. Boiveau, Y. Maday.

Our goal is to predict the series of:

- infected I_{obs} (taken as $15 H_{\text{obs}}$)
- removed R_{obs} (dead+recovered)

We test everything for the case of Paris.



Search in the family of SIR models

Consider the family of SIR dynamics with β and $\gamma \in L^\infty([0,T])$.

$$\frac{\mathrm{d}S}{\mathrm{d}t}(t) = -\beta(t)I(t)S(t)$$
$$\frac{\mathrm{d}I}{\mathrm{d}t}(t) = -\frac{\mathrm{d}S}{\mathrm{d}t}(t) - \gamma(t)I(t)$$
$$\frac{\mathrm{d}R}{\mathrm{d}t}(t) = \gamma(t)I(t),$$

with constant population $N = S(t) + I(t) + R(t)$, $\forall t \in [0,T]$.

Observation: The evolution $I_{\mathrm{obs}}(t)$ and $R_{\mathrm{obs}}(t)$ belongs to this family.

Proof: If we set β and γ to

$$\beta^*_{\mathrm{obs}}(t) := -\frac{N}{I_{\mathrm{obs}}(t)S_{\mathrm{obs}}(t)}\frac{\mathrm{d}S_{\mathrm{obs}}}{\mathrm{d}t}(t)$$
$$\gamma^*_{\mathrm{obs}}(t) := \frac{1}{I_{\mathrm{obs}}(t)}\left[\frac{\mathrm{d}I_{\mathrm{obs}}}{\mathrm{d}t}(t) - \frac{\beta^*_{\mathrm{obs}}(t)I_{\mathrm{obs}}(t)S_{\mathrm{obs}}(t)}{N}\right]$$

then the associated SIR solution satisfies

$$I(t) = I_{\mathrm{obs}}(t), \quad R(t) = R_{\mathrm{obs}}(t)$$

Implications

The family of SIR dynamics has perfect approximation properties:

If we search for $\beta, \gamma \in L^\infty$, we have:

- Perfect fitting for I_{obs}, R_{obs} in $[0, T]$
- Optimal forecasting power in $[T, T+\tau]$

L^∞ is too large for a practical algorithm:

We need to identify a smaller approximation class for β and γ.

$\Rightarrow$ Find subspaces

$$\mathrm{B}_n = \text{span}\{\beta_i : [0, T+\tau] \to \mathbb{R}_+\}_{i=1}^n$$

and

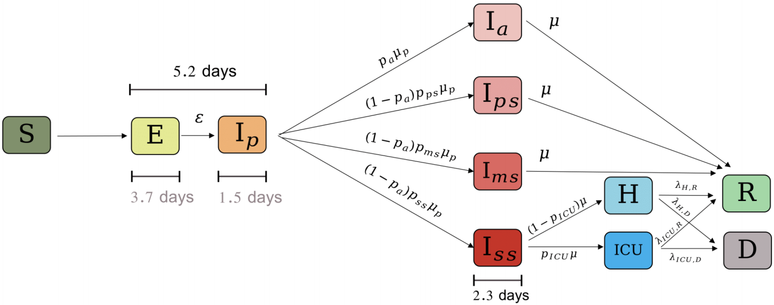
$$\mathrm{G}_n = \text{span}\{\gamma_i : [0, T+\tau] \to \mathbb{R}_+\}_{i=1}^n$$

of small dimension n.

Our strategy is based on model reduction of detailed epidemiological models.

Step 1: Generate virtual scenarios from detailed models

We consider detailed models with many compartments.
One example is SE5CHRD from Colizza et al.



K=11 compartments

p=10 parameters $\mu = (\varepsilon, p_a, \mu_p, p_{ps}, \dots) \in \mathbb{R}^p$
The parameters are not time-dependent.

Step 1: Generate virtual scenarios from detailed models

Notation:

- Parameters: $\mu = (\varepsilon, p_a, \mu_p, p_{ps}, \dots) \in \mathbb{R}^K$
- Solution:

$$u_\mu = (S_\mu, E_\mu, \dots, H_\mu, R_\mu, D_\mu) = \{u_\mu(t) \,:\, t \in [0, T+\tau]\} \in L^2([0, T+\tau])^p$$

We suppose that the parameters can take values in a certain range

$$\varepsilon \in [\varepsilon_{\min}, \varepsilon_{\max}]; \quad p_a \in [p_{a,\min}, p_{a,\max}]; \quad \dots$$

so that

$$\mu \in \mathcal{P} = [\varepsilon_{\min}, \varepsilon_{\max}] \times [p_{a,\min}, p_{a,\max}] \times \cdots \in \mathbb{R}^p$$

The set of solutions is

$$\mathcal{M} := \{u_\mu \,:\, \mu \in \mathcal{P}\}.$$

It provides a set of “virtual scenarios” with a detailed model.

Step 2: Collapse/Project into family of SIR dynamics

For a fixed μ, collapse the 11 compartments into the 3 compartments:

$$\begin{aligned} S_\mu^c &= S_\mu + E_\mu \\ I_\mu^c &= I_{p,\mu} + I_{a,\mu} + \dots \\ R_\mu^c &= R_\mu + D \end{aligned}$$

The associated dynamic is thus

$$\dot{S}_\mu^c = \dot{S}_\mu + \dot{E}_\mu = \text{formulas from Colizza et al.}$$

$$\dots$$

Step 2: We collapse into 3 compartments

In general, the collapsed variables will not follow a SIR dynamic because there are no $\beta_\mu(t)$, $\gamma_\mu(t)$ such that

$$\begin{aligned}
\dot{S}^c_\mu(t) &= -\beta_\mu(t) I^c_\mu(t) S^c_\mu(t) \\
\dot{I}^c_\mu(t) &= -\dot{S}^c_\mu(t) - \gamma_\mu(t) I^c_\mu(t) \\
\dot{R}^c_\mu(t) &= \gamma_\mu(t) I^c_\mu(t),
\end{aligned}$$

We search for $\beta_\mu(t)$, $\gamma_\mu(t)$ so that this relation is satisfied at best in a least squares sense

$$\begin{aligned}
(\beta_\mu(t), \gamma_\mu(t)) \in \operatorname*{arg\,min}_{\beta\geq 0, \gamma \geq 0} & \|\dot{S}^c_\mu(t) + \beta I^c_\mu(t) S^c_\mu(t)\|^2 \\
&+ \|\dot{I}^c_\mu(t) + \dot{S}^c_\mu(t) + \gamma I^c_\mu(t)\|^2 \\
&+ \|\dot{R}^c_\mu(t) - \gamma I^c_\mu(t)\|^2
\end{aligned}$$

We do the search for all $t \in [0, T + \tau]$.

Step 3: Build the subspaces B_n and G_n (model reduction)

We consider the sets

$$B := \{\beta_\mu \in L^2([0, T+\tau]) \,:\, \mu \in \mathcal{P}\}$$
$$G := \{\gamma_\mu \in L^2([0, T+\tau]) \,:\, \mu \in \mathcal{P}\}$$

By applying model reduction techniques (SVD, NMF, Nonnegative Greedy), we find functions

$$b_1, \dots, b_n \in L^2([0, T+\tau]); \quad g_1, \dots, g_n \in L^2([0, T+\tau]).$$

We approximate for all $\mu \in \mathcal{P}$

$$\beta_\mu(t) \approx \beta_{\mu,n}(t) := \sum_{i=1}^{n} c_i(\mu) b_i(t) \qquad \in \mathrm{B}_n := \mathrm{span}\{b_1, \dots, b_n\}$$
$$\gamma_\mu(t) \approx \gamma_{\mu,n}(t) := \sum_{i=1}^{n} \tilde{c}_i(\mu) g_i(t) \qquad \in \mathrm{G}_n := \mathrm{span}\{g_1, \dots, g_n\}$$

We expect that the approximation error decays quickly as n grows so that we can take n small.

Step 4: Fitting and predictions using B_n and G_n

We next consider for $t \in [0, T+\tau]$

$$\begin{cases} \beta_n(t) & = \sum_{i=1}^n c_i b_i(t) & \in \mathrm{B}_n \\ \gamma_n(t) & = \sum_{i=1}^n \tilde{c}_i g_i(t) & \in \mathrm{G}_n \end{cases}$$

and the associated SIR dynamics

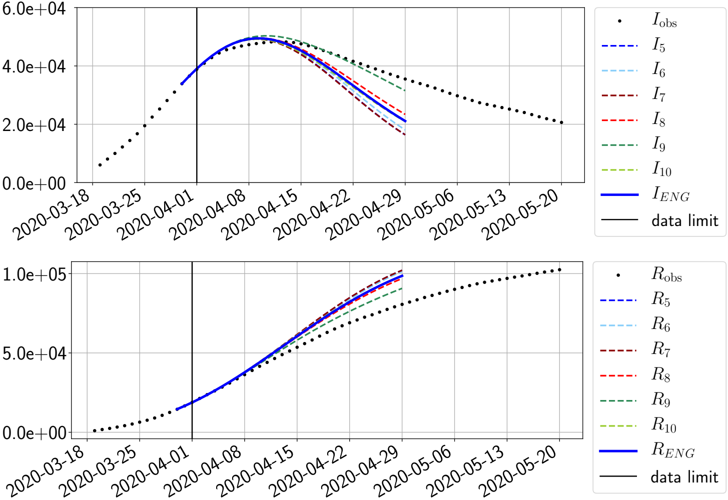
$$\begin{aligned} \dot{S}(t) &= -\beta_n(t) I(t) S(t) \\ \dot{I}(t) &= -\dot{S}(t) - \gamma_n(t) I(t) \\ \dot{R} &= \gamma_n(t) I(t), \end{aligned}$$

We search for the c_i and $\tilde{c}_i$ that satisfy at best the observations

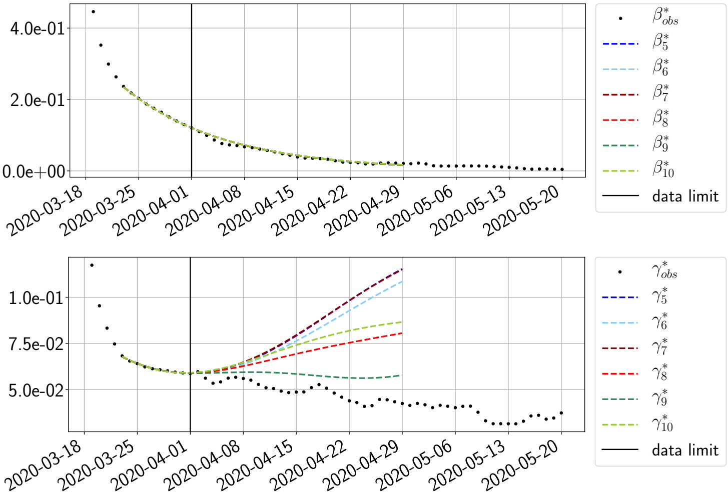
$$\min_{c_i, \tilde{c}_i} \int_0^T \left(\| I_{\text{obs}}(t) - I(t) \|^2 + \| R_{\text{obs}}(t) - R(t) \|^2 \right) \mathrm{dt}$$

Our final output is: $I(t)$, $R(t)$ for $t \in [0, T+\tau]$

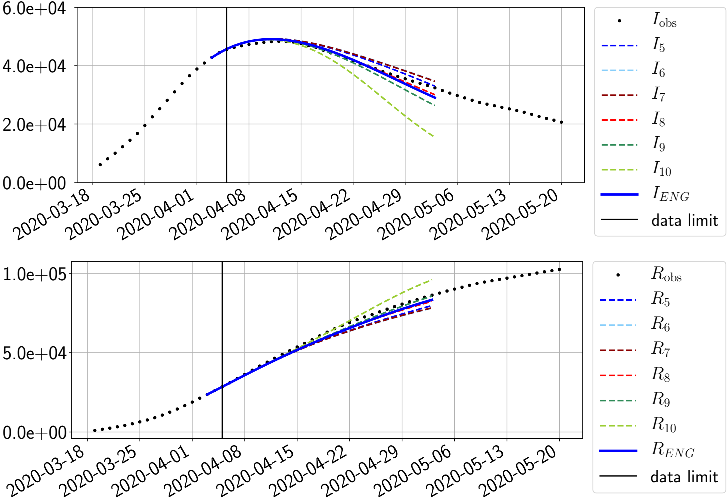
Results. T=April 1st.



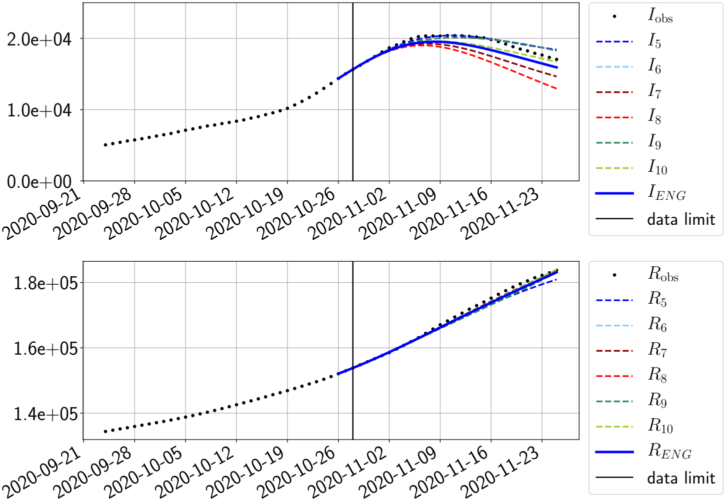
Results. T=April 1st.



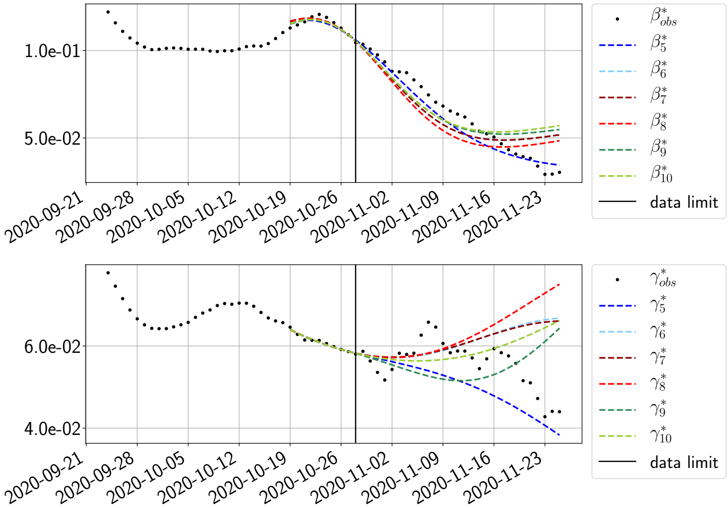
Results. T=April 5.



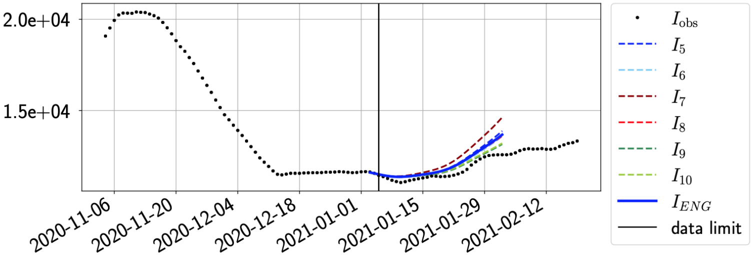
Results. T=October 28.



Results. T=October 28.



Results. T=January 5, 2021.



Muti-Regional Extension

Roadmap:

1. Multi-regional version of detailed models.
   $\rightarrow$ Need for interregional population fluxes
2. Collapse into multi-regional SIR
3. Build reduced models (several possible options here)
4. Fit and predict as before

What do/can we expect from this extension?

- Better long term forecasts
- Better understand spatial propagation and rise of clusters
- Study mobility scenarios

Mobility data from Facebook

Works with colleagues from Face au Virus.

Nature of the data

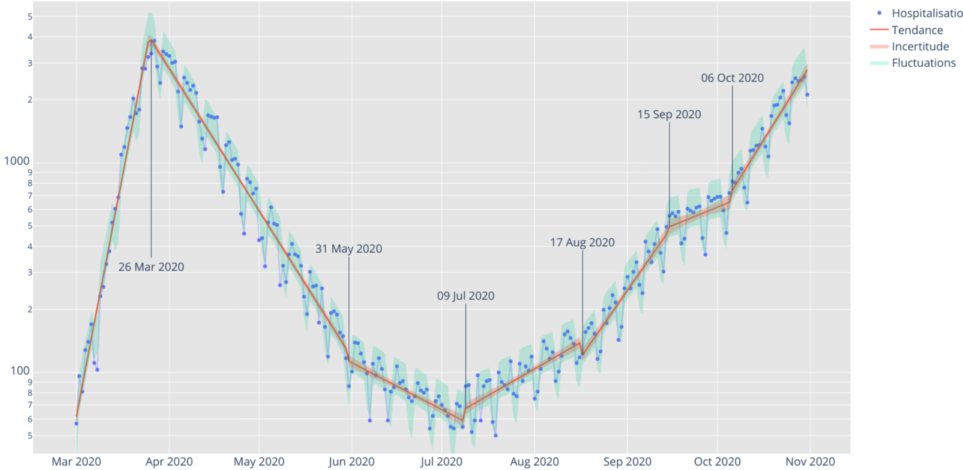
Facebook:

- **Raw data:** GPS data from users that accept geolocalisation.
- **Anonymisation:** Aggregation on tiles $\geq$ 200 x 200 m^2.

At PSL University, we are given anonymized data.

Remark: Many other companies collect geoloc data.

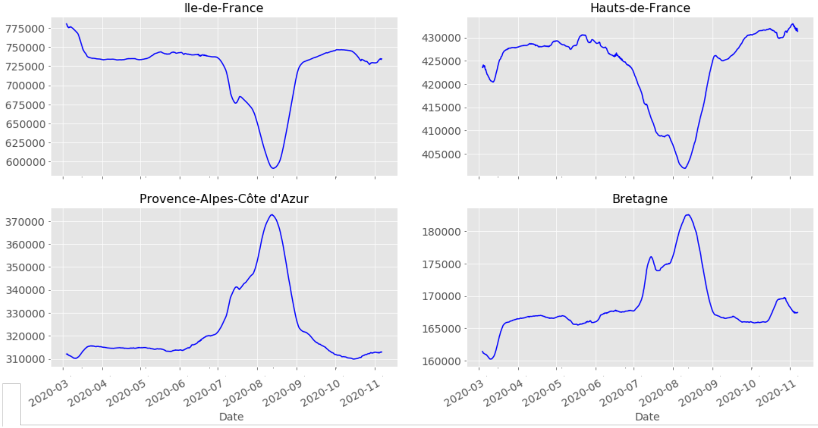
Between waves: Statistical fitting of hospitalized series



Breakpoints in summer cannot be linked to sanitary measures.

Population mobility and mixing could be an explanatory factor.

Evolution of FB users per region

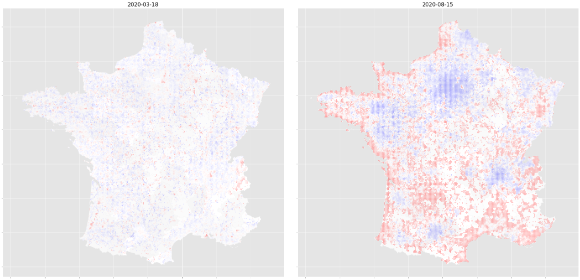


Key observation:
Migration to summer regions.

$\rho(d,k)$: density on day d at tile k.

$r(d,k) = \frac{\rho(d,k)-\rho(\bar{d},k)}{\rho(\bar{d},k)}$: Relative density variation.

Relative Density Variation ($\pm 25\%$).
Reference $\bar{d}$: Sunday March 22, 2020.



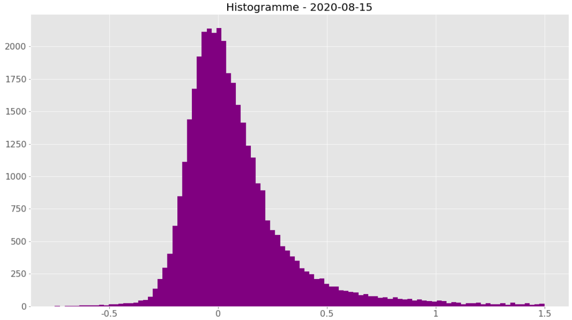
2020-03-18 2020-08-15

Relative Change of Density

Density Histogram at day d:

$$H(d) = \text{Histogram}\{r(d,k) \; : \; k \in 1, \dots, \#\text{Tiles}\}$$

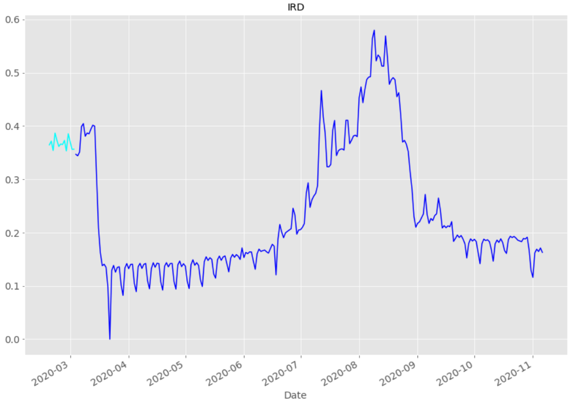
Note that $H(\bar{d}) = \delta_0$.



Relative Change of Density

We define the Relative Change of Density index as

$$\mathrm{RCD}(d) = W_2(H(d), H(\bar{d})) \approx \sigma(d)$$



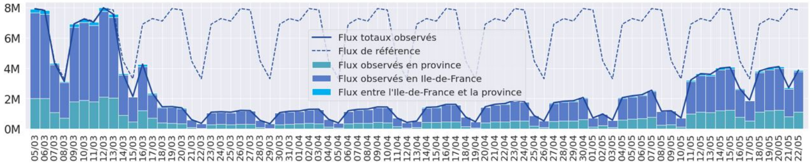
Key observation:
Important increase of RCD in the summer.

Conclusions

- Our forecasting method has shown good predictive properties in one single region.
- We are building a more automatic code to apply it to different regions.
- We are building extensions to include:
  - population mobility data (from FB)
  - data from waste waters

Appendix

Number of daily trips between French departments



Key numbers:

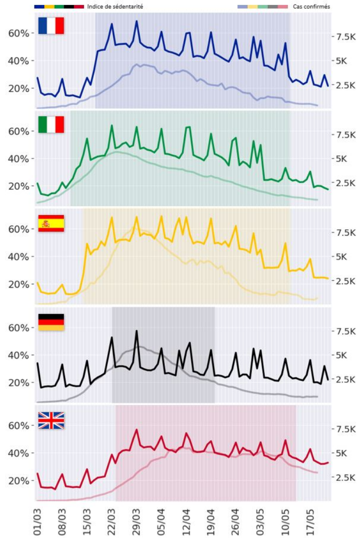
-80% trips during lockdown

50% trips immediately after lockdown

First wave: Stay-at-Home Index in Europe

From March to June 2020

Evolution of Stay-at-Home Index



Saty-at-Home Index: Nb. people having spent the day in a 600 m by 600 m tile.

FR, IT, ESP:

- 15% before lockdown
- 50% during lockdown
- 30% after

UK, DE: More moderate effect.