COVID19 — From modeling to estimation using Kalman based estimators

Annabelle Collin[1,3], Mélanie Prague[1,2,3,4], Philippe Moireau[1,4]

Part of joint work with

Linda Wittkop[1,2,3,5], Dan Dutartre[1], Quentin Clairon[2,3,4], Rodolphe Thiébaut[1,2,3,4,5], Boris P. Hejblum[1,2,3,4]

1: Inria, 2:Université de Bordeaux - CNRS, 3: Inserm, 4:Vaccine Research Institute, 5: CHU Bordeaux, 6:Ecole Polytechnique - CNRS

History and disclaimer

- Since 2015
  - A longstanding collaboration between A. Collin, P. Moireau (Analysis,Num Anal,Scientific Comp) about Kalman based estimators for "large dimensional" (classically distributed in our case) systems.
- Since 2018
  - A joint work with M. Prague (Biostatistics) on alternative to non-linear mixed effect model approaches (NLME) for estimating mechanistic models in pharmacokinetics, epidemiology

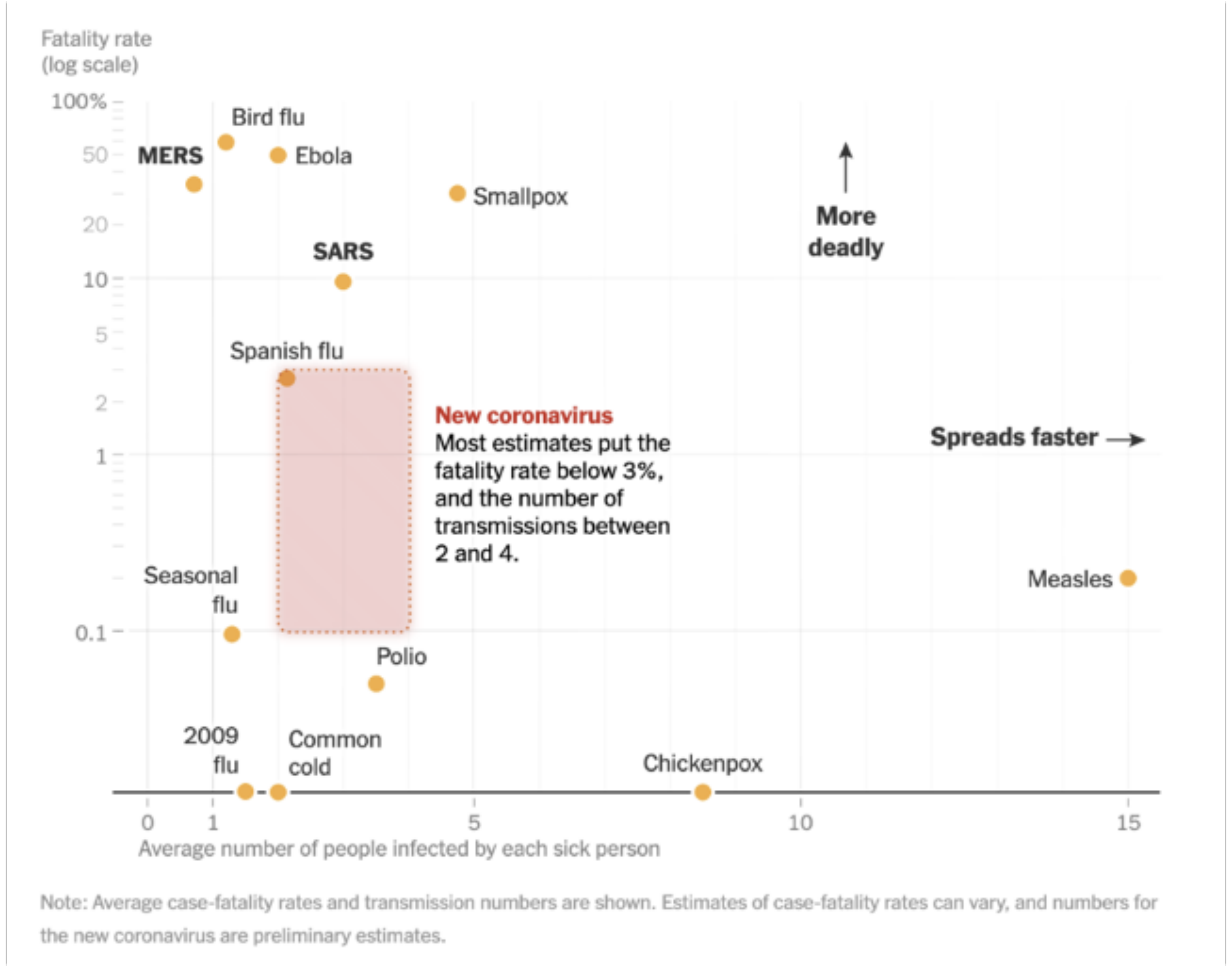
> *A. Collin, M. Prague, P. Moireau – Estimation for dynamical systems using a population-based Kalman filter – Applications to pharmacokinetics models hal-02869347 – Submitted.*

- End of march 2020
  - A. Collin, P. Moireau test kalman estimation for COVID models
  - While M. Prague is using SAEM methods for COVID predictions with B. P. Hejblum, L. Wittkop, R. Thiébaut, D. Dutartre, Q. Clairon.
- Early June 2020
  - Can we give some **supplementary** grounds with our Kalman-based approach about some of the modeling choices and estimation assumptions used the NLME approach?
- Early July 2020
  - Submission of a paper with supplementary materials

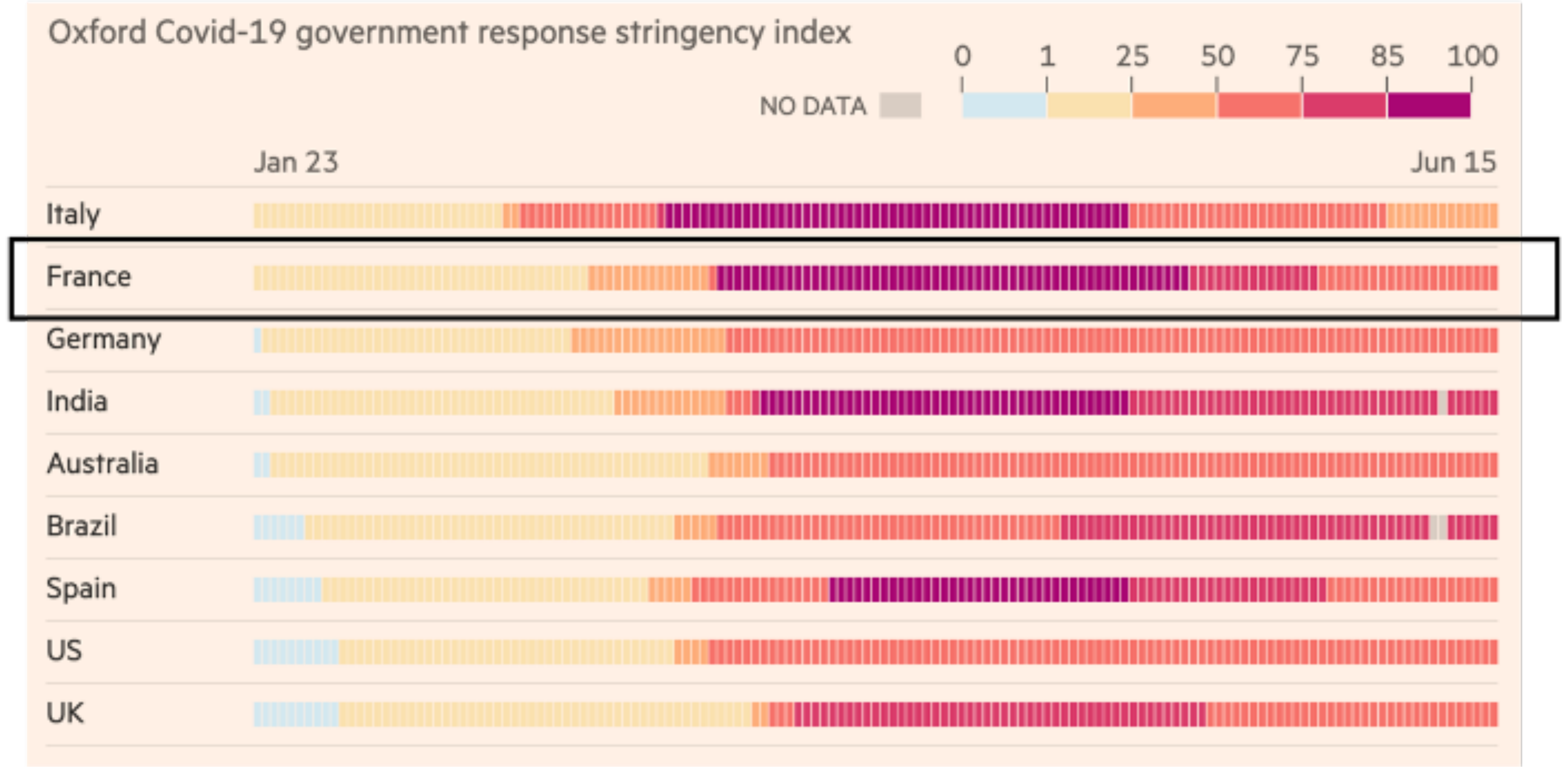
> *M. Prague, L. Wittkop, A. Collin, Dan Dutartre, Q. Clairon, P. Moireau, R. Thiébaut, B. P. Hejblum – Multi-level modeling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate – Submitted.*

A little bit of context

SARS-Cov-2 appeared in Wuhan (China) in December 2019
No Vaccine until December 11th 2020

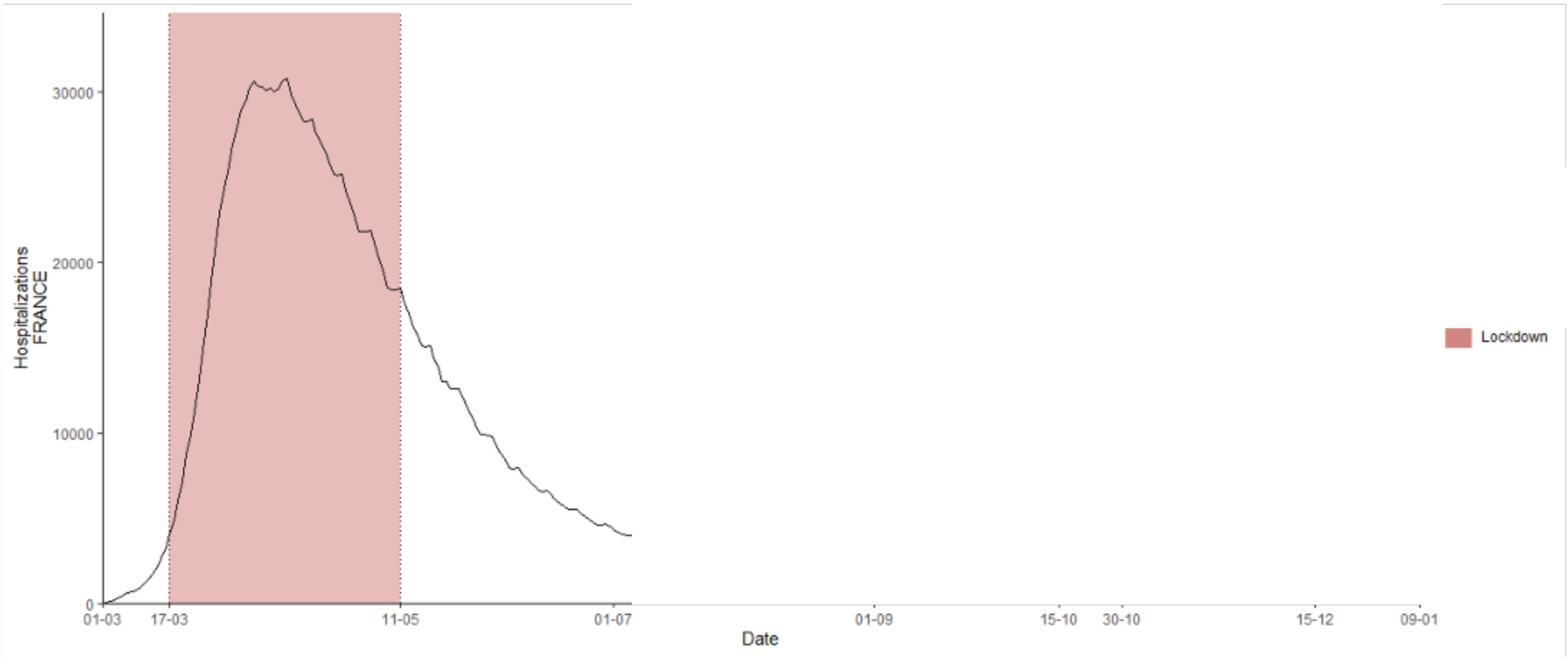


Worldwide implementation of **Non-pharmaceutical Intervention**
from less stringent (masks, hand washing…) to most stringent complete lock-down.



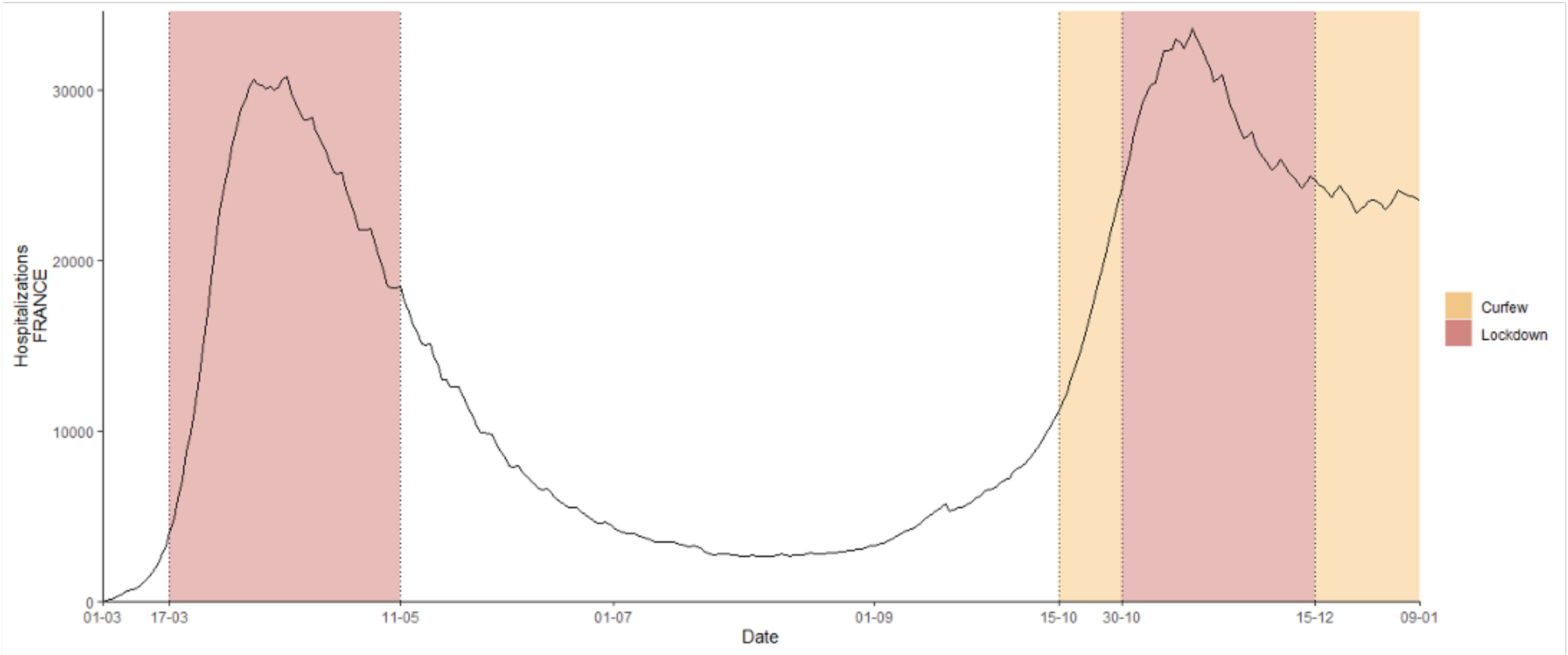
Main goals

How to **estimate the effect** of lock-down on first wave?



Main goals

Can it **inform and predict** the second, third … and so on … waves ?



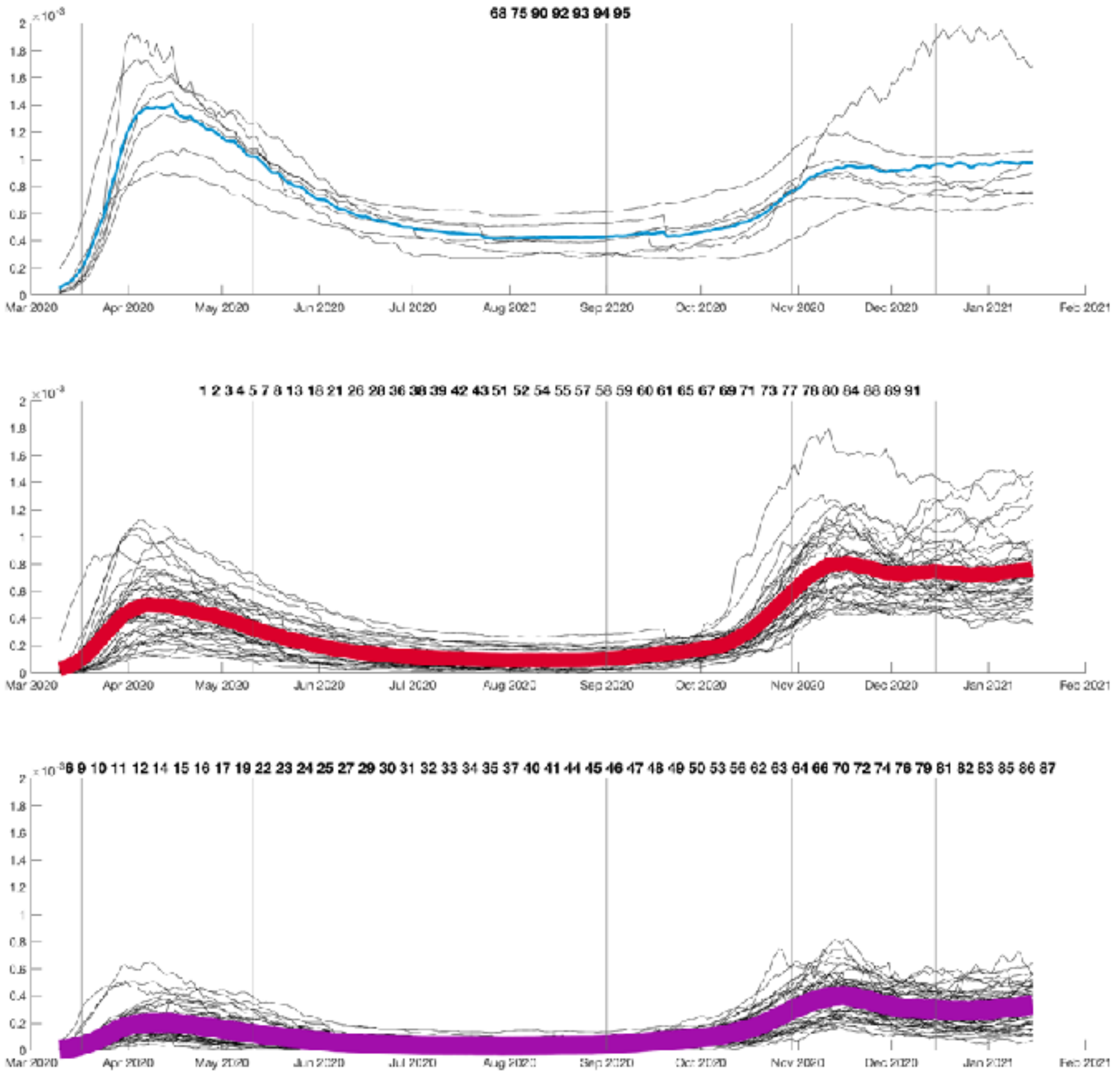
Available data

- **Statistical analysis** only so good as the **data :**

  « Garbage in, garbage out »

- **Data is crucial**, but hardly a hospital priority when there are not enough respirators…
- **Public / Semi-Public data on Covid-19:**
  - Multiple sources
  - Multiple formats
  - Multiple geographical resolutions
  - Multiple interpretation (phase 2 vs phase 3, cf # of tests…)
  - Background noise

Percent of hospitalization data by departments
Means clustering with 3 clusters

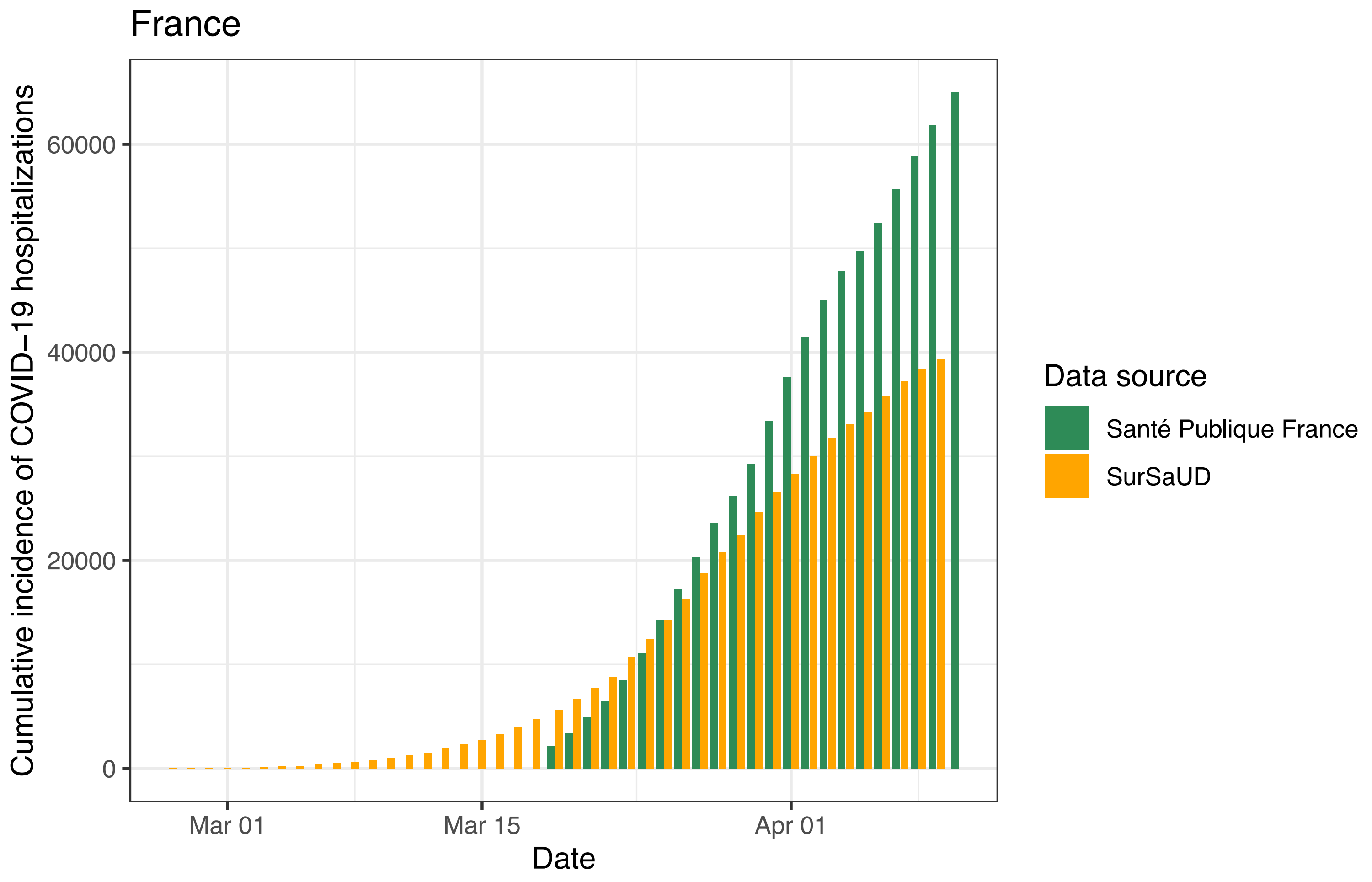


Available data

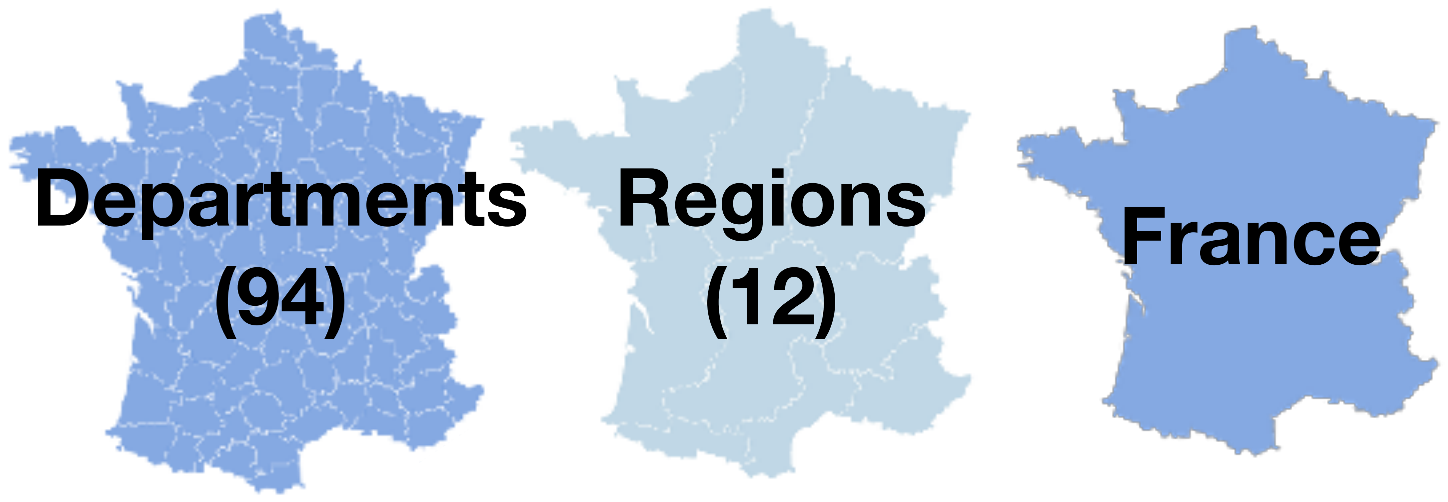
- **Infection Data (# positive tests)**
  - Sante Publique France: March 1st -> March 25th
  - SI-dep: May 15th -> now
- **Infection Data (Sentinelle)**
  - Number of individual seen by GP with COVID-19 related symptoms
- **Hospital Data (SurSaUD)**
  - Admission through urgent care
- **Hospital Data (SI-VIC available June 1st)**
  - Hospitalisation Admission
  - ICU
  - Death

Partial data collection with change of sources & change of collection mode

Multiple sources & format With inconsistencies



Multiple geographic levels



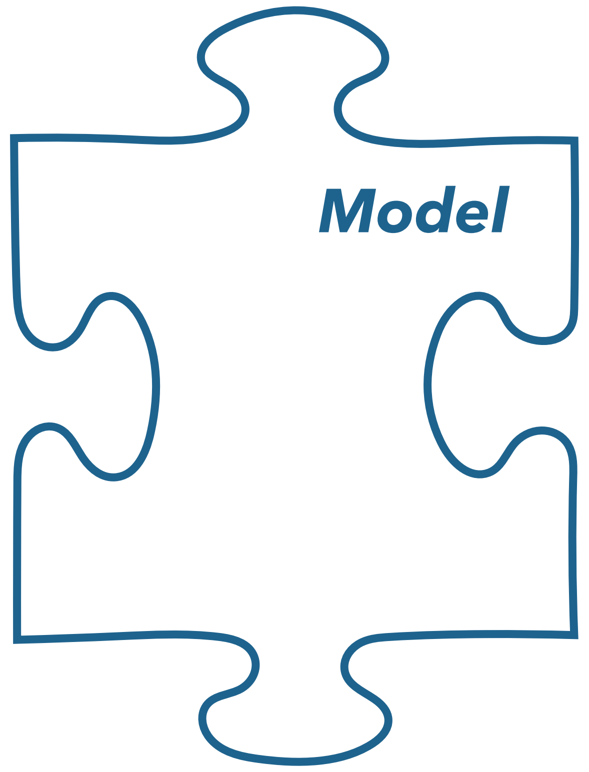
Estimation paradigms

State
Exposed (E)
Infectious (I, A)
Removed (R)
Hospitalized (H)
–> **ODE system**

State: $x = \begin{pmatrix} E \\ I \\ \vdots \\ H \end{pmatrix}$

Augmented state:

$$z = \begin{pmatrix} x \\ \theta \end{pmatrix}$$



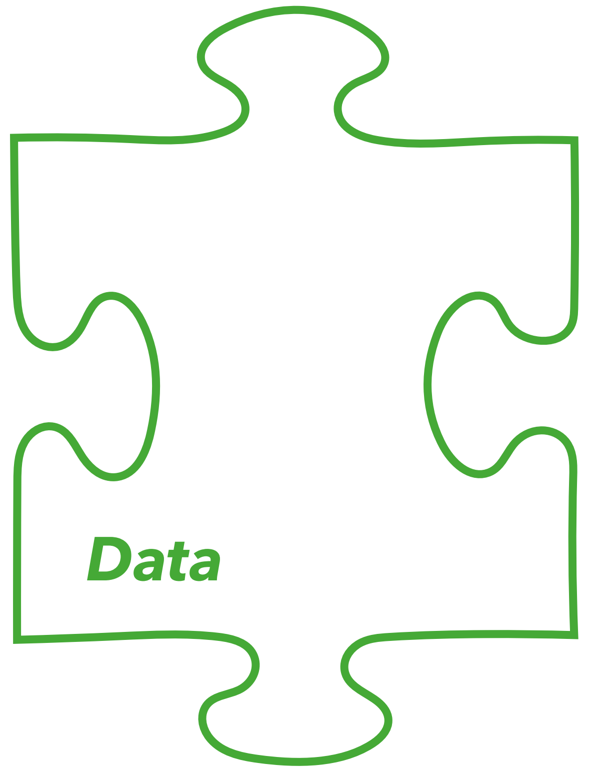
Uncertainties on parameters $\boldsymbol{\theta}$ and initial conditions …

$$\dot{x} = f(x, \theta, t) + B_x v$$
$$\dot{\theta} = 0$$
$$x(0) = x_0 + \xi_x$$
$$\theta(0) = \theta_0 + \xi_\theta$$

$$\dot{z} = F(z, t) + Bv$$
$$z(0) = z_0 + \xi$$

Strategies for coupling the informations
Variational, Sequential

Goal: minimize a discrepancy comparing *z* and *y*



Observations y

Observation operator h

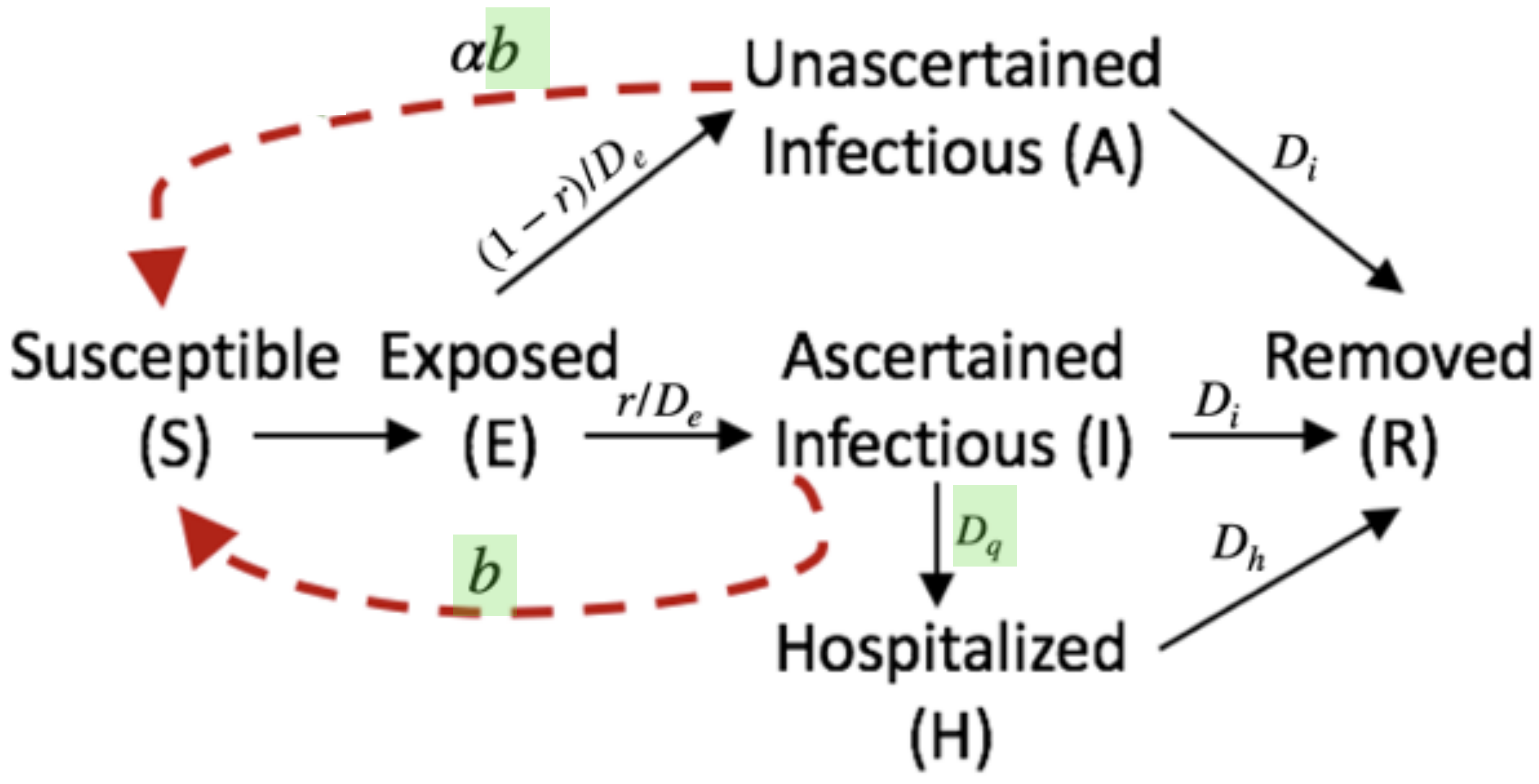
$$y = h(z, t) + \chi$$

Mixed effect dynamical model

The mathematical model

- The SEIRAH model : An extended Susceptible-Exposed-Infectious-Recovered (SEIR) model

$$
\begin{cases}
\dot{S} = -\dfrac{b(t)S(I+aA)}{N} \\
\dot{E} = \dfrac{b(t)S(I+aA)}{N} - \dfrac{E}{D_e} \\
\dot{I} = \dfrac{rE}{D_e} - \dfrac{I}{D_q} - \dfrac{I}{D_i}, \\
\dot{R} = \dfrac{I+A}{D_i} + \dfrac{H}{D_h} \\
\dot{A} = \dfrac{(1-r)E}{D_e} - \dfrac{A}{D_i} \\
\dot{H} = \dfrac{I}{D_q} - \dfrac{H}{D_h}.
\end{cases}
$$



[1] *Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W. et al. (2020), Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science 368 (6490), 489–493.*

[2] *Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R. et al. (2020), The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. Annals of Internal Medicine 172 (9), 577–582.*

[3] *Wang, C., Liu, L., Hao, X., Guo, H., Wang, Q., Huang, J. et al. (2020). Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China. medRxiv 2020.03.03.20030593. doi: 10.1101/2020.03.03.20030593.*

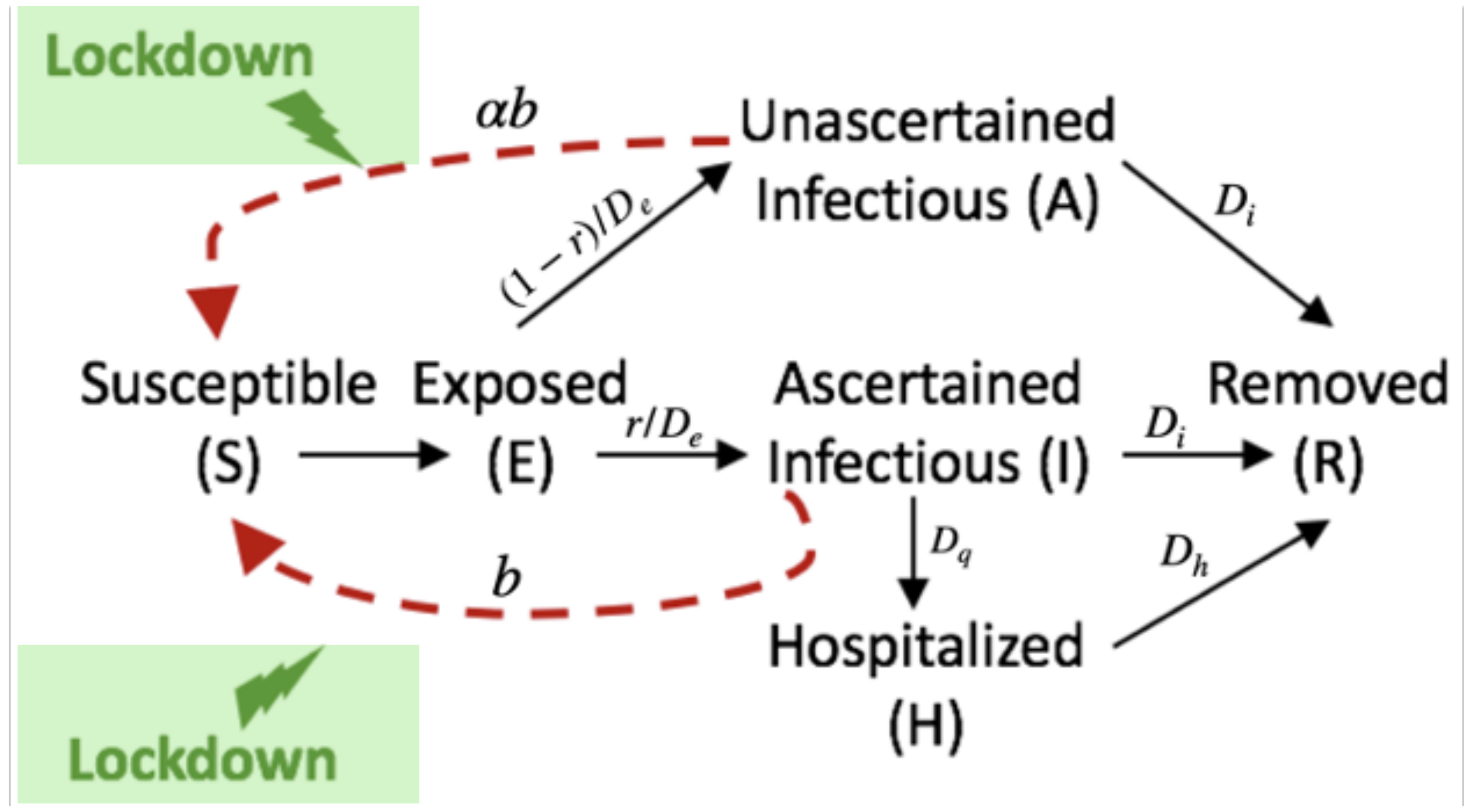
[4] *INED (2020). Insee: Recensements de population, Estimations de population. https://www.ined.fr/fichier/s_rubrique/159/estim.pop.nreg.sexe.gca.1975.2020.fr.xls. Accessed: 2020-03-25.*

Parameter	Interpretation	Value	References
b	Transmission rate of ascertained cases	Region Specific	**Estimated**
r	Ascertainment rate	Region Specific	*Réseau Sentinelle*
a	Ratio of transmission between A and I	1.5	[1]
D_e	Latent (incubation) period (days)	5.2	[2]
D_i	Infectious period (days)	2.3	[1,3]
D_q	Duration from I onset to H (days)	Region Specific	**Estimated**
D_h	Hospitalization period (days)	30	[1,3]
N	Population size	Region Specific	[4]

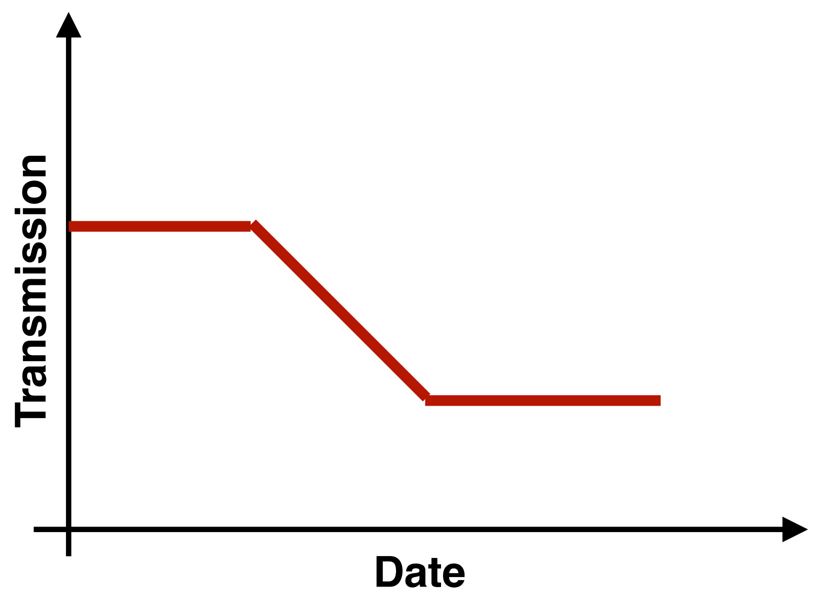
Definition of the transmission change over time

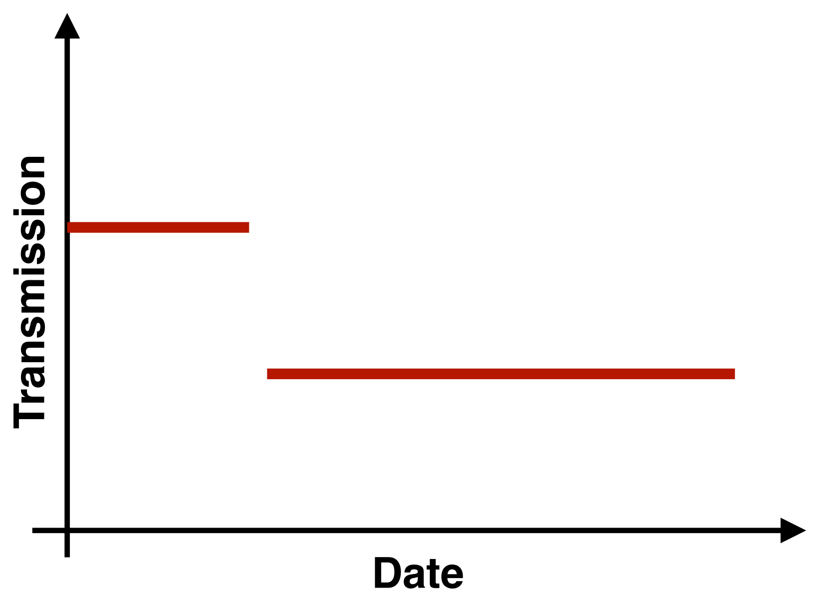
- We assume that non-pharmaceutical intervention (NPI) reduces the transmission b.

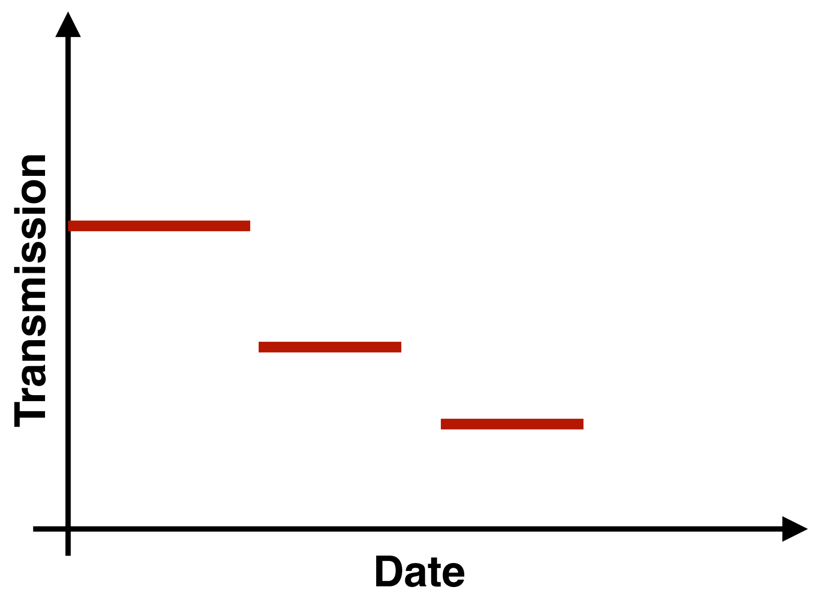
$$\begin{cases} \dot{S} = -\frac{b(t)S(I+\alpha A)}{N} \\ \dot{E} = \frac{b(t)S(I+\alpha A)}{N} - \frac{E}{D_e} \\ \dot{I} = \frac{rE}{D_e} - \frac{I}{D_q} - \frac{I}{D_i}, \\ \dot{R} = \frac{I+A}{D_i} + \frac{H}{D_h} \\ \dot{A} = \frac{(1-r)E}{D_e} - \frac{A}{D_i} \\ \dot{H} = \frac{I}{D_q} - \frac{H}{D_h}. \end{cases}$$

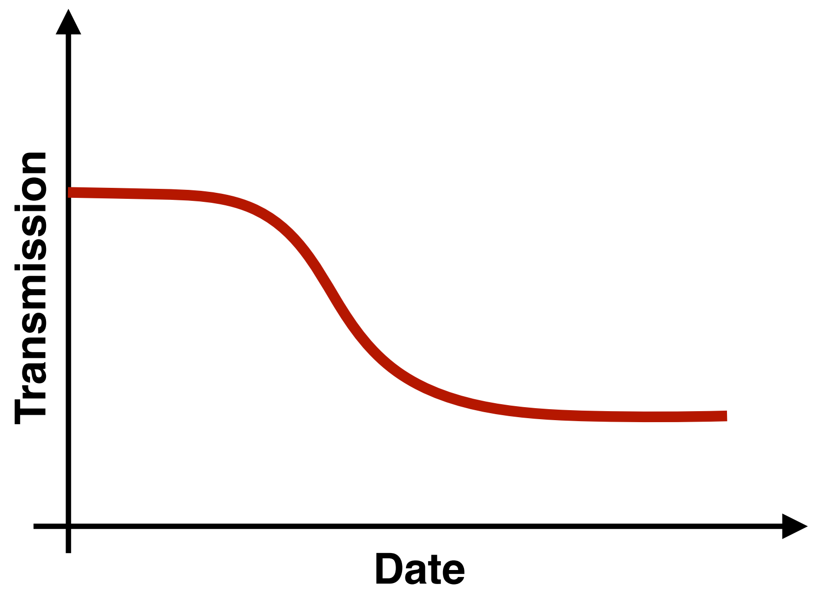


- **Major contribution of this work consist in using sequential methods to inform the parametric shape** of the effect of NPI







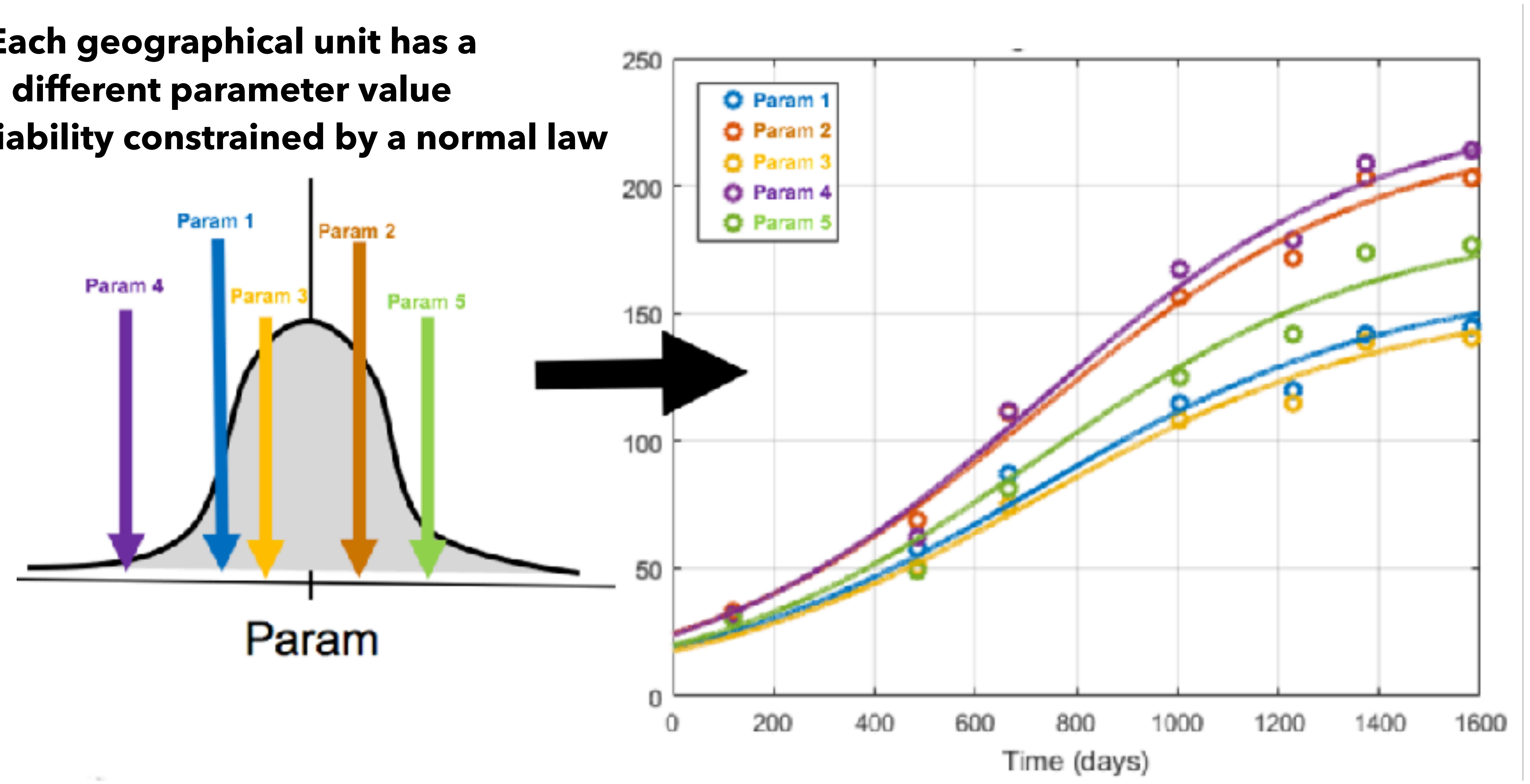


The statistical model

- We use a mixed effect model on parameters to account for **inter-department / inter-region variability (denoted i)**.

$$\log(b_i(t)) = g(t) + u_i^b, \qquad \text{with } u_i^b \simeq \mathcal{N}(0, \sigma_b^2)$$
$$\log(D_{q_i}) = D_{q0} + u_i^{D_q}, \qquad \text{with } u_i^{D_q} \simeq \mathcal{N}(0, \sigma_{D_q}^2)$$
$$\log(E_i(0)) = E(0) + u_i^{E_0}, \qquad \text{with } u_i^{E_0} \simeq \mathcal{N}(0, \sigma_{E_0}^2)$$
$$\log(I_i(0)) = I(0) + u_i^{I_0}, \qquad \text{with } u_i^{I_0} \simeq \mathcal{N}(0, \sigma_{I_0}^2)$$
$$\log(H_i(0)) = H(0) + u_i^{H_0}, \qquad \text{with } u_i^{H_0} \simeq \mathcal{N}(0, \sigma_{H_0}^2)$$

Each geographical unit has a different parameter value with variability constrained by a normal law



The observation model

We observe five variables :

Y_1: Incident number of cases tested positive

$$Y_1(t) = \frac{rE}{D_e}$$

Y_2: Incident number of hospitalized

$$Y_2(t) = \frac{I}{D_q}$$

Y_3: Prevalent number of hospitalized

$$Y_3(t) = H$$

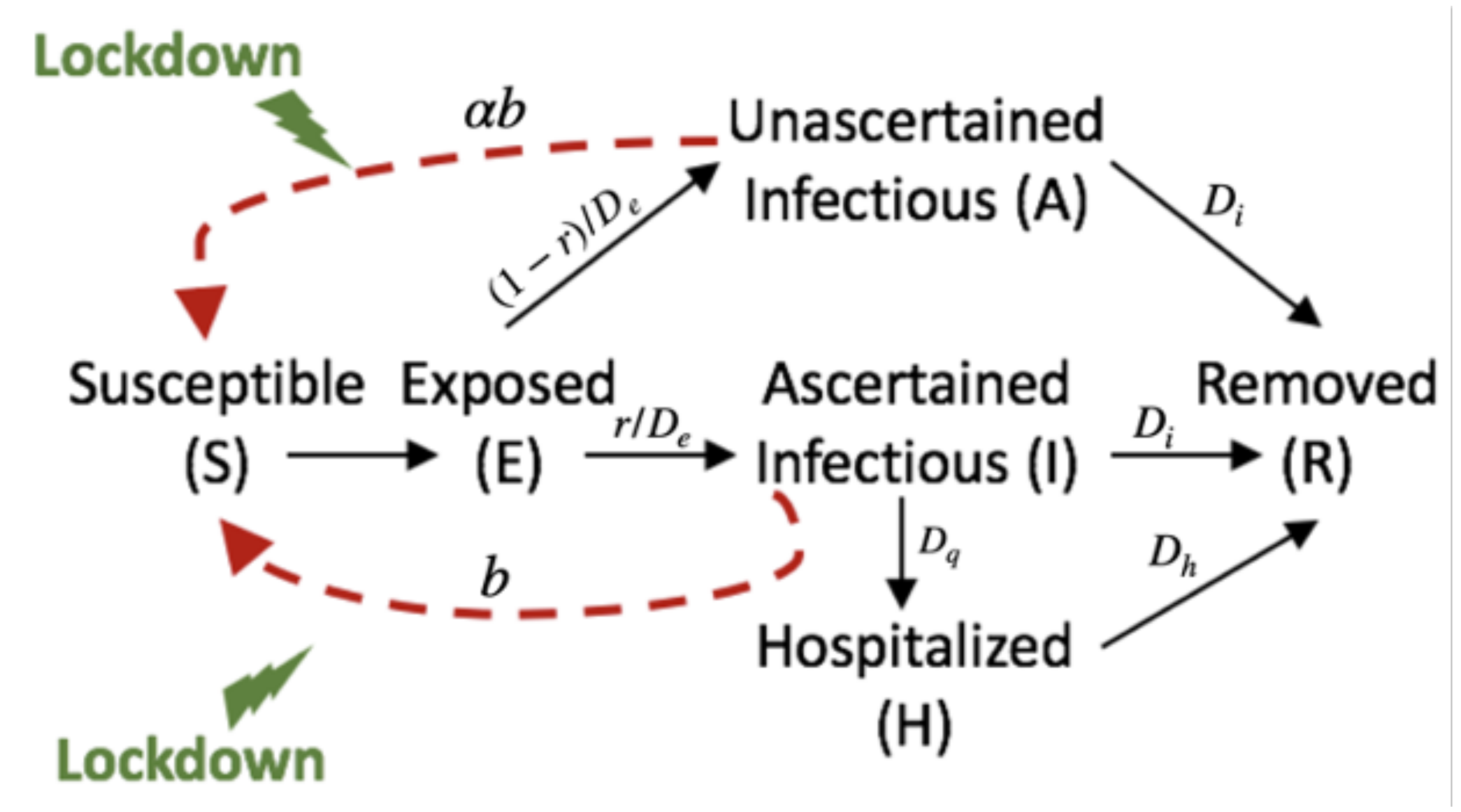
Y_4: Prevalent number of cases in ICU

$$Y_4(t) = 0.25 \times H$$

Y_5: Number of deaths

$$Y_5(t) = 0.005 \times R$$

$$\begin{cases} \dot{S} = -\frac{b(t)S(I + aA)}{N} \\ \dot{E} = \frac{b(t)S(I + aA)}{N} - \frac{E}{D_e} \\ \dot{I} = \frac{rE}{D_e} - \frac{I}{D_q} - \frac{I}{D_i}, \\ \dot{R} = \frac{I + A}{D_i} + \frac{H}{D_h} \\ \dot{A} = \frac{(1 - r)E}{D_e} - \frac{A}{D_i} \\ \dot{H} = \frac{I}{D_q} - \frac{H}{D_h}. \end{cases}$$



Reduced-order population Unscented Kalman filter

Maximum likelihood estimation (variational approach)

- Minimize the criterion with respect to the uncertainties under the constraint of the model dynamics

$$(\hat{\xi}, \hat{v}) = \text{argmax}(\log \mathscr{L}_T((\zeta, v); y(t_1), \dots, y(t_{N_T})))$$

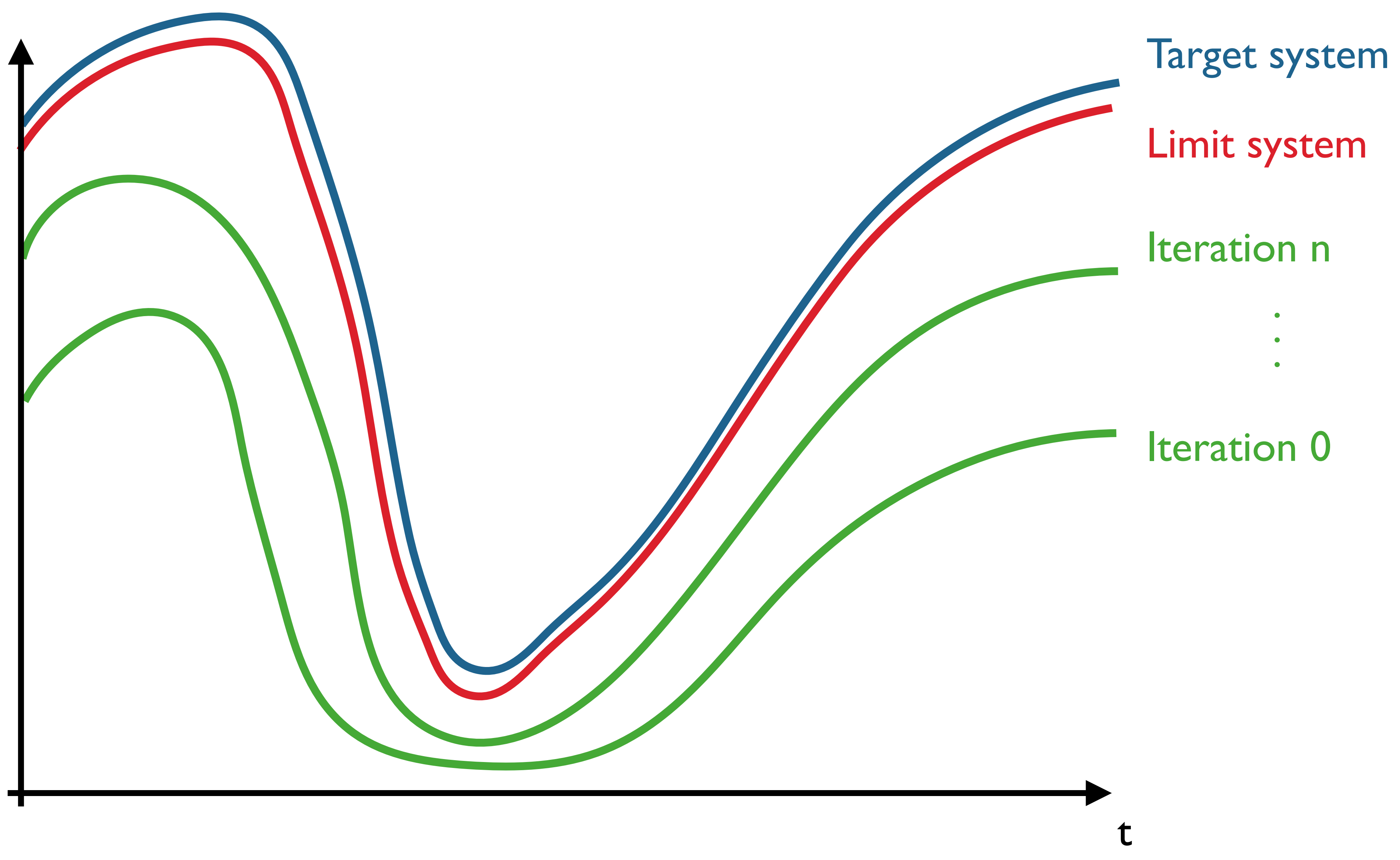
ζ Uncertainties (parameters / initial conditions)
v Model error
y Observations

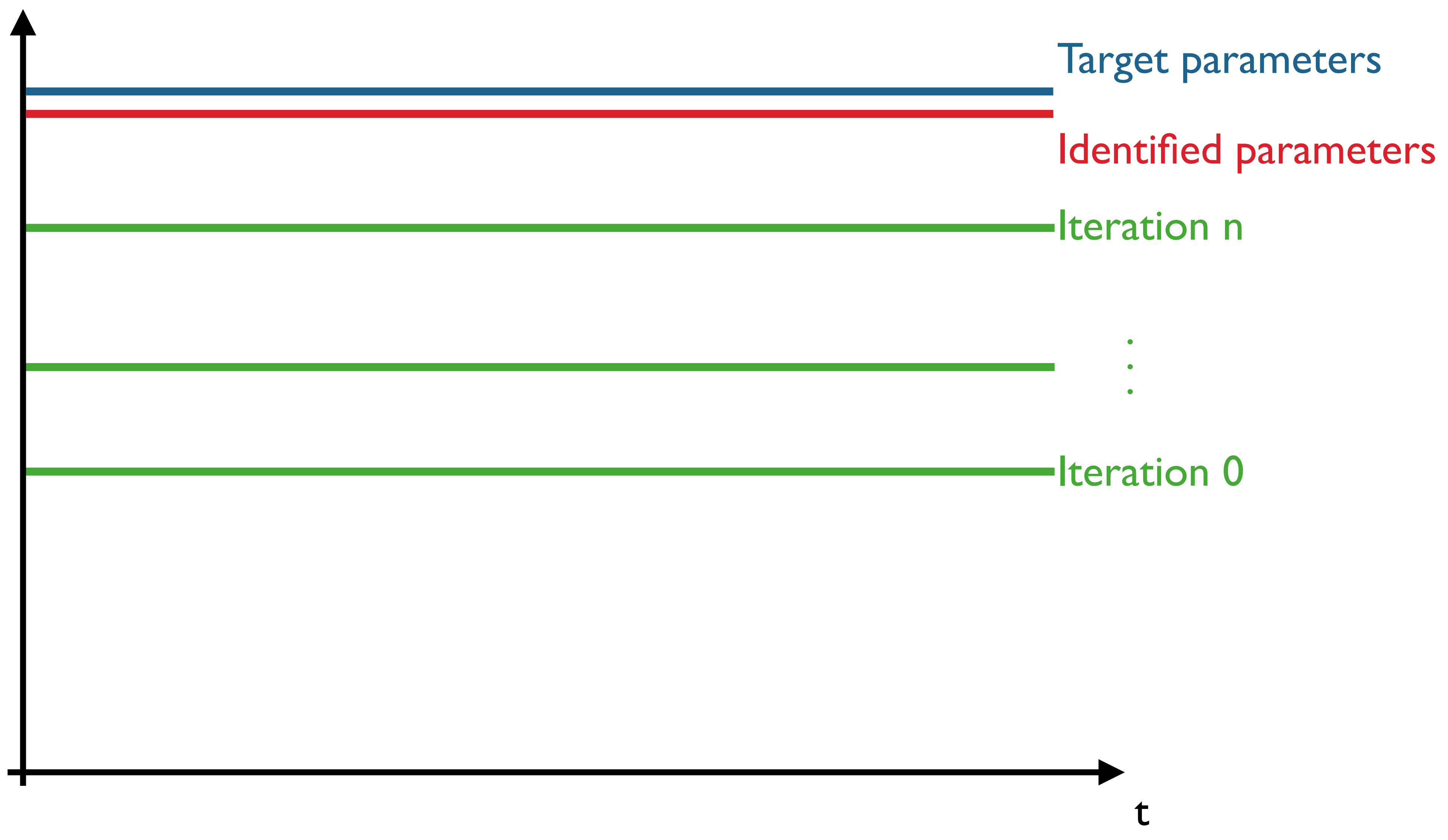
- Corresponds to least-square minimisation when Gaussian laws are considered

$$\min_{\xi, v} \left\{ \mathscr{J}_T(\xi, v) = \underbrace{\frac{1}{2}\langle \xi, P_\diamond^{-1} \xi \rangle}_{\text{uncertainties priors}} + \underbrace{\frac{1}{2}\int_0^T \langle v(t), Q(t)^{-1} v(t) \rangle \, dt}_{\text{model error}} + \frac{1}{2} \sum_{k=0}^{N_{T,\text{obs}}} \overbrace{\left[\langle y_k - h(z(t_k)), (W_k)^{-1} (y_k - h(z(t_k))) \rangle \right]}^{\text{comparison between } y \text{ and } z} \right\}$$

with $\dot{z} = F(z, t) + Bv$
$z(0) = z_0 + \xi$

Rk: It is possible to rewrite this functional in order to take into account mixed-effects model (SAEM algorithm)





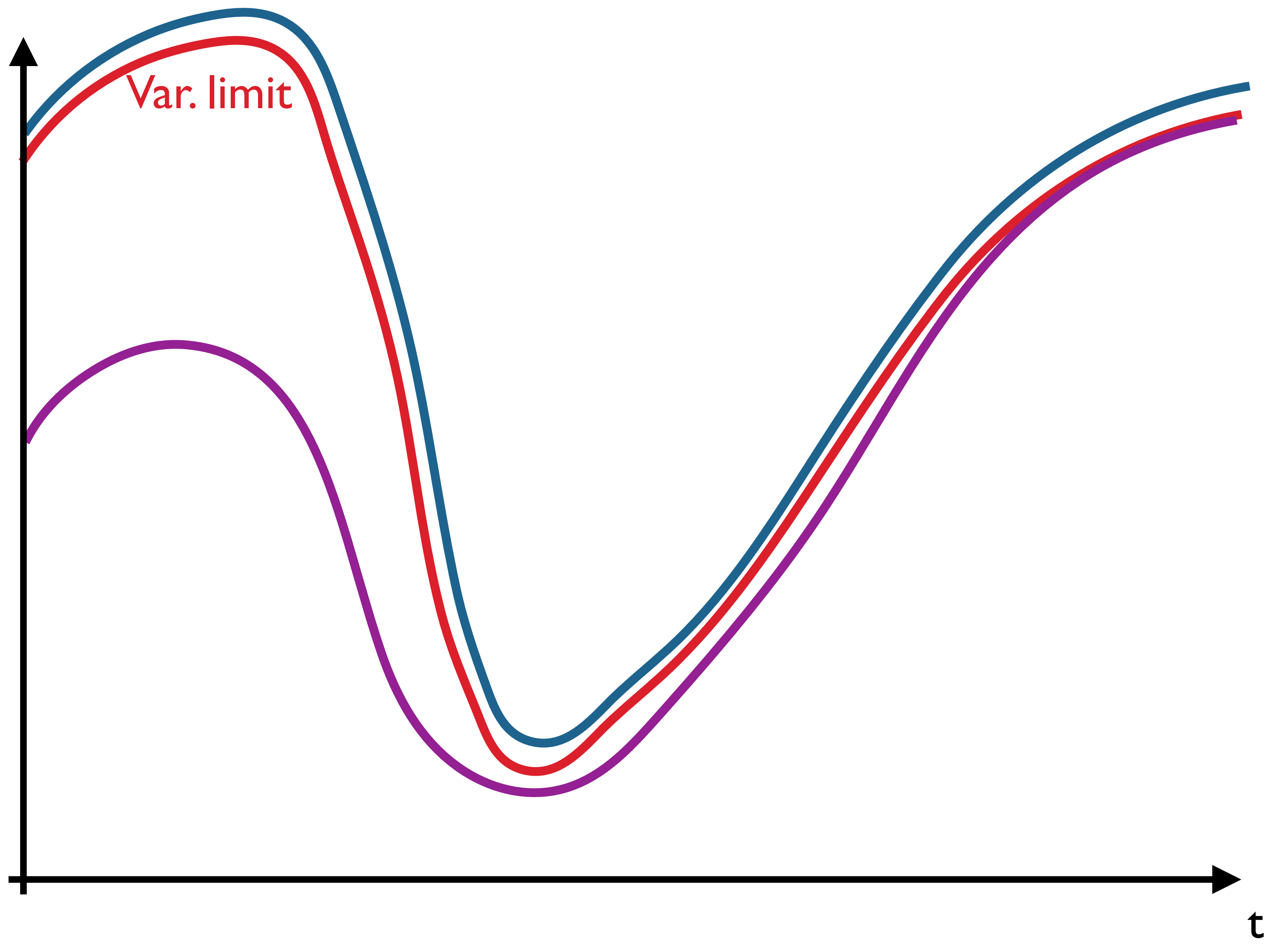
Sequential Approach

- Correct the dynamics by a feedback based on the discrepancy combining the data and the model state

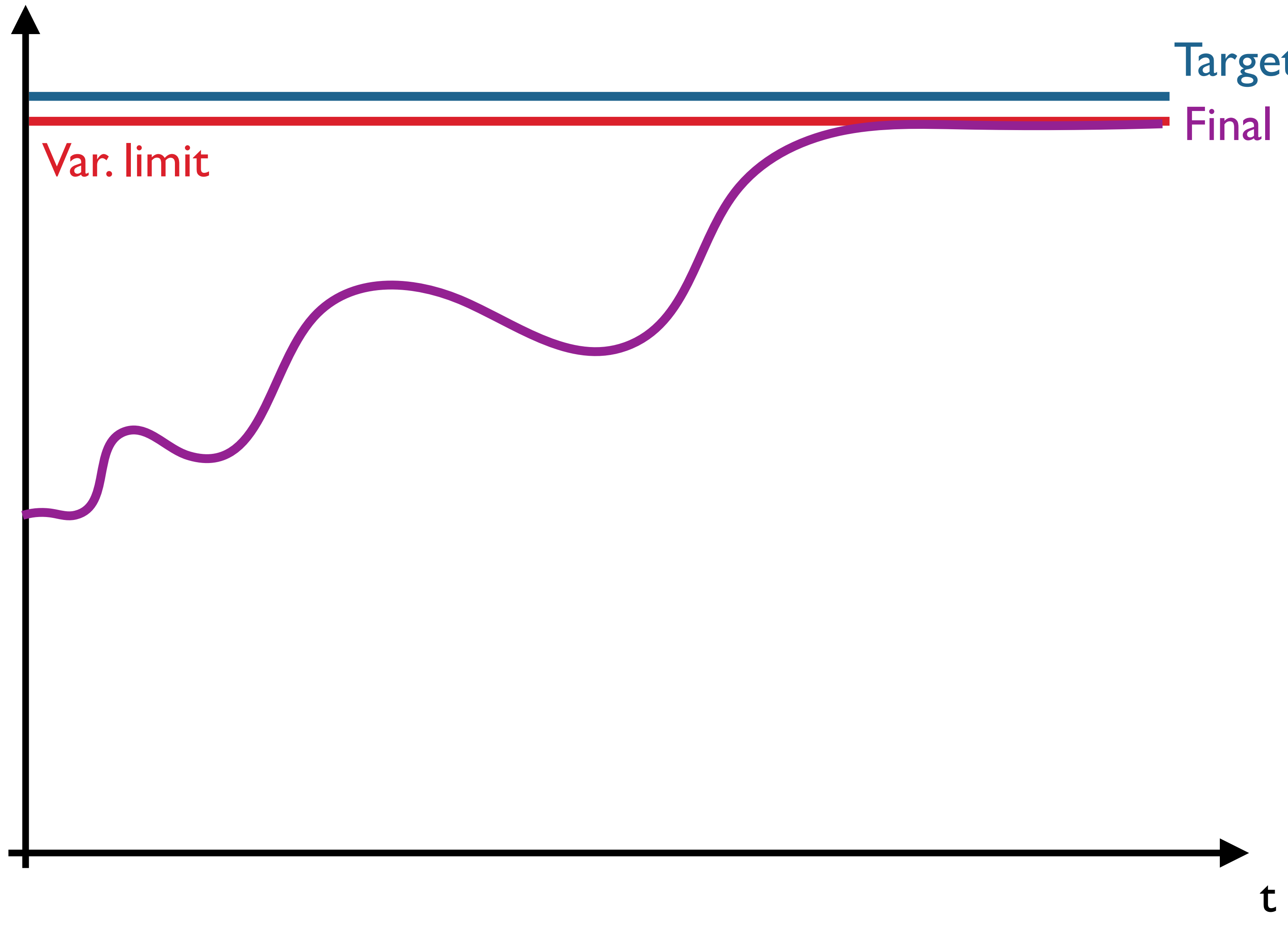
$$\dot{z} = f(z,t) + Bv + k(y,x)$$
$$z(0) = z_0 + \xi$$

with $z = \begin{pmatrix} x \\ \theta \end{pmatrix}$ state / parameters

- Rk: Allow also to estimate initial conditions and model error.



The parameters have a dynamics too !



- More precisely we consider a **reduced-order** version of a **population Unscented Kalman** filter:
  i) Define a **population** criterion: we need a least square criterion
  ii) Introduction to **Kalman** filter in a linear context
  iii) Presentation of the **Unscented** Kalman filter (UKF)
  iv) Presentation of the **reduced-order** version

i) A population criterion

- Population approach: compensate the lack of data by an available population
- A population made of groups indexed by i

$$\xi^i(t) = \xi^0_\diamond + \tilde{\xi}^i, \qquad \xi = \begin{pmatrix} \xi^1 \\ \vdots \\ \xi^{N_P} \end{pmatrix} \in (\mathcal{Z})^{N_P} \simeq \mathbb{R}^{N_z},$$

Mixed-effect approach: pooling all the patients together and estimating a global distribution of the model parameters in the population.

  - Criterion:

$$\min_{\xi,v} \left\{ \mathscr{J}_T(\xi,v) = \frac{1}{2}\langle \xi_\diamond, (P^0_\diamond)^{-1}\, \xi_\diamond \rangle + \sum_{i=1}^{N_P} \left[\frac{1}{2}\langle (\xi^i - \xi_\diamond), (\tilde{P}^i_\diamond)^{-1}\,(\xi^i - \xi_\diamond)\rangle + \frac{1}{2}\int_0^T \langle v^i(t), Q^i(t)^{-1} v^i(t)\rangle\, \mathrm{d}t + \frac{1}{2}\sum_{k=0}^{N_{T,\mathrm{obs}}} \langle y^i_k - C(z^i(t_k)), (W^i_k)^{-1}(y^i_k - C(z^i(t_k)))\rangle \right] \right\}$$

  - Assuming that the population share the same weighted mean value: $\xi^0_\diamond = \mathbb{E}_{a,N_P}(\xi) \stackrel{\text{def}}{=} \sum_{i=1}^{N_P} a^i \xi^i$ with $\sum_{i=1}^{N_P} a^i = 1,$

$$\min_{\xi,v} \left\{ \mathscr{J}_T(\xi,v) = \frac{1}{2}\langle \mathbb{E}_{a,N_P}(\xi), (P^0_\diamond)^{-1}\, \mathbb{E}_{a,N_P}(\xi) \rangle + \sum_{i=1}^{N_P} \left[\frac{1}{2}\langle (\xi^i - \mathbb{E}_{a,N_P}(\xi)), (\tilde{P}^i_\diamond)^{-1}\,(\xi^i - \mathbb{E}_{a,N_P}(\xi))\rangle + \frac{1}{2}\int_0^T \langle v^i(t), Q^i(t)^{-1} v^i(t)\rangle\, \mathrm{d}t + \frac{1}{2}\sum_{k=0}^{N_{T,\mathrm{obs}}} \langle y^i_k - C(z^i(t_k)), (W^i_k)^{-1}(y^i_k - C(z^i(t_k)))\rangle \right] \right\}$$

- Rewriting

$$\langle \mathbb{E}_{a,N_P}(\xi), (P_\diamond)^{-1}\, \mathbb{E}_{a,N_P}(\xi) \rangle + \sum_{i=1}^{N_P} \langle (\xi^i - \mathbb{E}_{a,N_P}(\xi)), (\tilde{P}^i_\diamond)^{-1}\,(\xi^i - \mathbb{E}_{a,N_P}(\xi))\rangle = \sum_{i,j=1}^{N_P} \left\langle \xi^i, \left[a^i a^j (P_\diamond)^{-1} + \delta_{ij}(\tilde{P}^i_\diamond)^{-1} - (a^i(\tilde{P}^j_\diamond)^{-1} + a^j(\tilde{P}^i_\diamond)^{-1}) + a^i a^j \sum_{\ell=1}^{N_P} (\tilde{P}^\ell_\diamond)^{-1} \right] \xi^j \right\rangle.$$

$$= (\mathbf{P}_0^{-1})_{i,j}$$

i) A population criterion

- We obtain a classical estimation framework: Minimize

$$\min_{\xi,v}\left\{\mathscr{J}_T(\xi,v)=\sum_{i=1}^{N_P}\left[\frac{1}{2}\langle\xi^i,\mathbf{P}_0^{-1}\xi^i\rangle+\frac{1}{2}\int_0^T\langle v^i(t),Q^i(t)^{-1}v^i(t)\rangle\,\mathrm{d}t+\frac{1}{2}\sum_{k=0}^{N_{T,\mathrm{obs}}}\langle y_k^i-C(z^i(t_k)),(W_k^i)^{-1}(y_k^i-C(z^i(t_k)))\rangle\right]\right\}.$$

with

$$\begin{cases}\dot z(t)=F(z(t),t)+B(t)v(t), & \forall t\in[0,T]\\ z(0)=z_0,\end{cases}\qquad z=(z^1\cdots z^{N_P})^\intercal\in(\mathcal{Z})^{N_P}$$

$$y_k=H_k(z(t_k))+\chi_k,\quad 1\le k\le N_{T,\mathrm{obs}},$$

with

$$F(z(t),t)=\begin{pmatrix}F(z^1(t),t)\\ \vdots\\ F(z^{N_P}(t),t)\end{pmatrix}\text{ and }z_0=z_\diamond\times\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}+\xi,$$

$$H_k(z(t_k))=\begin{pmatrix}h_k(z^1)\\ \vdots\\ h_k(z^{N_P})\end{pmatrix}$$

- The key of our **uncertainty modeling** is that $\mathbf{P}_0$ couples the population members since indeed

$$\mathbf{P}_0\neq\begin{pmatrix}P_0 & & 0\\ & \ddots & \\ 0 & & P_0\end{pmatrix}$$

ii) The Kalman and Bucy filter (linear context)

- Kalman and Bucy in 1961 have shown that the minimizer of the following least-square minimisation

$$\min_{\xi,v}\left\{\mathscr{J}_T(\xi,v)=\frac{1}{2}\langle\xi,P_\diamond^{-1}\,\xi\rangle+\frac{1}{2}\int_0^T\langle v(t),Q(t)^{-1}v(t)\rangle\,\mathrm{dt}+\frac{1}{2}\sum_{k=0}^{N_{T,\mathrm{obs}}}\langle y_k-h\big(z(t_k)\big),(W_k)^{-1}(y_k-h\big(z(t_k)\big)\rangle\Big]\right\}$$

(population or not criterion)

with $\dot{z}=F(z,t)+Bv$

$z(0)=z_0+\xi$

when the model and the observer operator are **linear** corresponds to the solution of (time discrete version!)

Model

$$\left(\; z_{k+1}=F_{k+1|k}z_k \;\right)$$

discrete transition operator

Observer model

$$\begin{cases}\hat{z}_{k+1} &= F_{k+1|k}\hat{z}_k+K_k(y_k-H_k\hat{z}_k)\\ P_{k+1} &= F_{k+1|k}P_kF^T_{k+1|k}-K_kH_kP_k,\end{cases}$$

Covariance matrix (full matrix of size N_z x N_z)

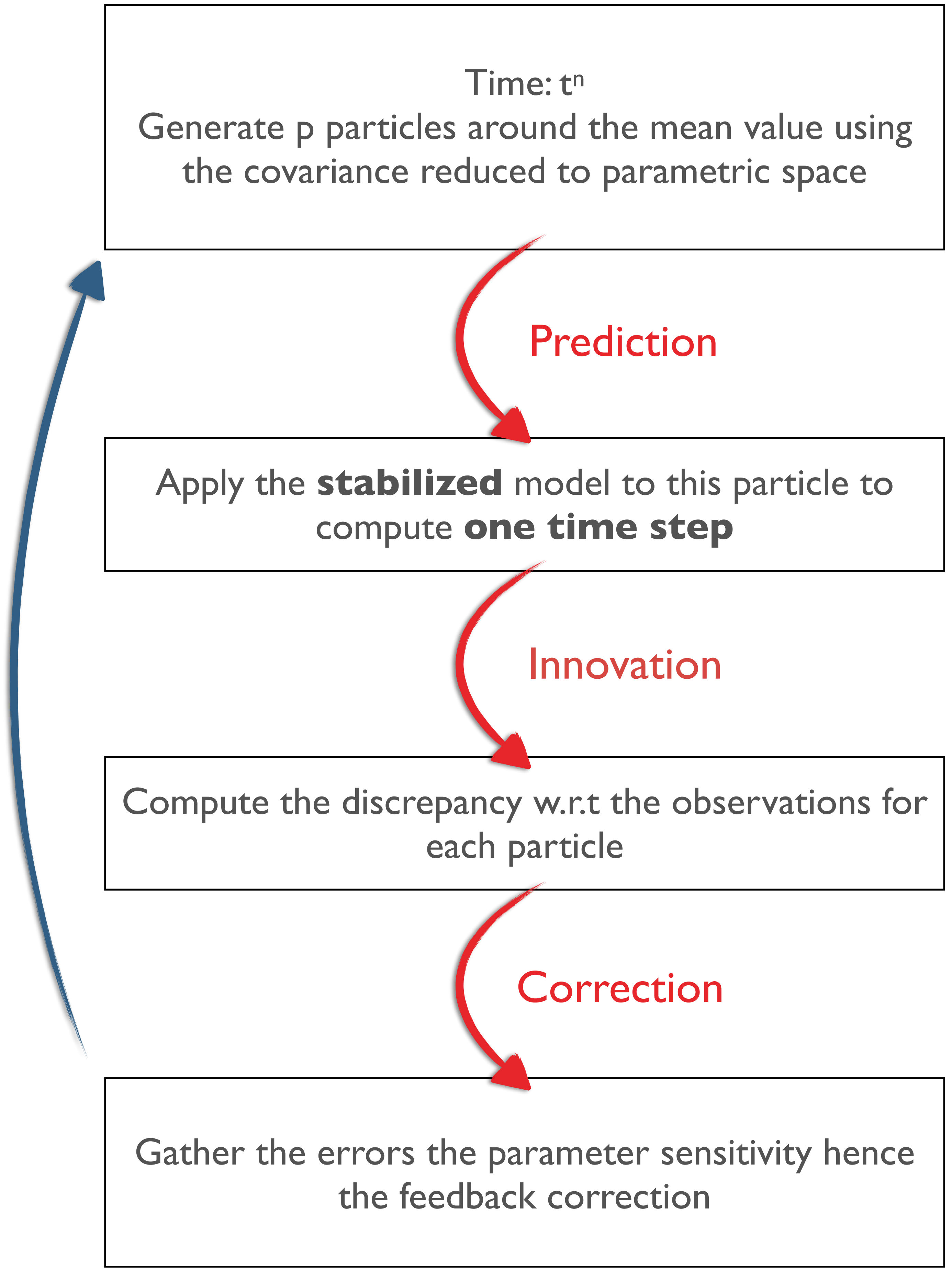
with $K_k=P_kH_k^T(H_kP_kH_k^T+W)^{-1}$

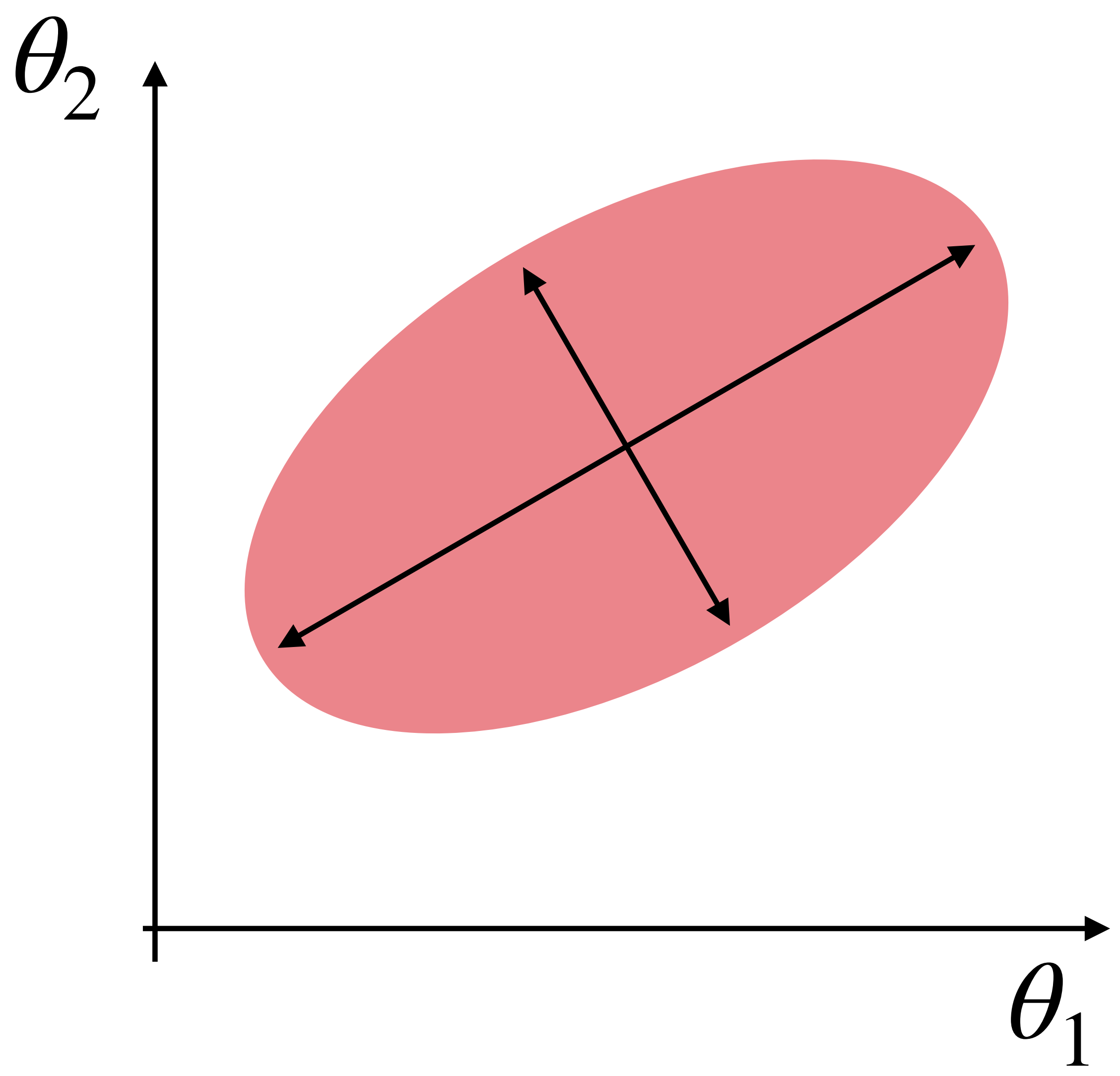
z state and parameters
H observation operator
y observations
W covariance matrix of the observation error
P covariance matrix of the estimation error

- **And if it is not linear?** It is classical to rely on approximate optimal sequential estimator based on the generalization of the Kalman filter to non-linear operators.

iii) The Unscented Kalman Filter

- Here we consider an *Unscented Kalman Filter*.
- The non linear operators are replaced by finite difference approximations based on sampling points.
- Sampling points can be seen as well-chosen “interpolation points” which propagate the mean and covariance of a random variable





S. Julier, J. Uhlmann and H. Durrant-Whyte, A new approach for filtering nonlinear systems, in American Control Conference (1995).

iv) A reduced order version

- The main idea behind the reduced order strategy is to consider a SVD decomposition of the covariance matrix P of the form
$$P = LU^{-1}L^T$$
with U an invertible matrix of small size and L an extension operator.

- For linear operators, this decomposition is stable over time and the equation on P leads to the two following systems with admissible computational times:
$$\dot{L} = AL \text{ and } \dot{U} = L^T C^T W^{-1} C L$$

- In non-linear cases, extensions of these two systems have been developed.

P. Moireau, D. Chapelle, Reduced-Order Unscented Kalman Filtering with Application to Parameter Identification in Large-Dimensional Systems — COCV 2011.

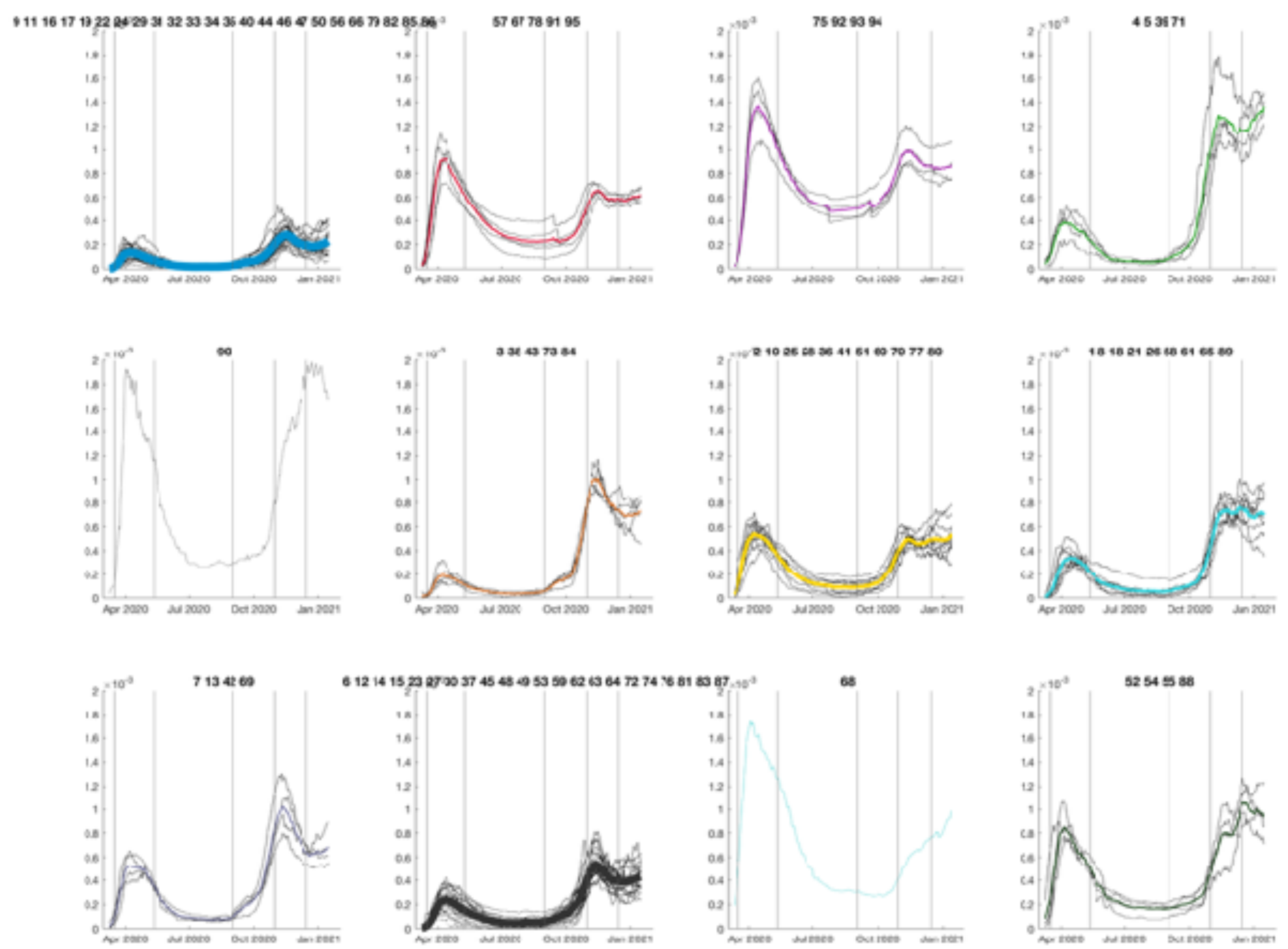
- Missing step: define the initial reduced covariance matrix U_0^{-1} and the initial extension matrix L_0 from the initial covariance matrix P_0. Our strategy is based on a clustering approach applied to the observations sequence using the **k-means algorithm**.

$$(U_0^{-1})_{r,s} = \frac{1}{n_{c_r} n_{c_s}} \sum_{i \in c_r} \sum_{j \in c_s} (P_0)_{i,j}$$

$$(L_0)_{i,r} = \beta_{i,r} \mathbb{1}_{N_z, N_z}$$

how much each region/ department i belongs to cluster c_r

Means clustering with 12 clusters on H/Npop for 94 departments



Application to the COVID crisis

Details about the estimation

- "State" variable for the dynamics with a Backward-Euler time scheme by regions without the S variables

$$\begin{pmatrix} x^i_{n+1} \\ \log b^i_{n+1} \\ \theta^i_{n+1} \end{pmatrix} = \begin{pmatrix} x^i_n + \tau f(x^i_n, b^i_n, \theta^i_n) \\ \log b^i_n + \tau g^i(t_n, \theta^i_n) \\ \theta^i_n \end{pmatrix} + \begin{pmatrix} 0_5 \\ \boxed{1 \quad v_n} \\ 0_{N_p} \end{pmatrix} \qquad x = (E, I/Dq, R, A, H)^\intercal \in \mathbb{R}^5$$

as the shape of b is undetermined

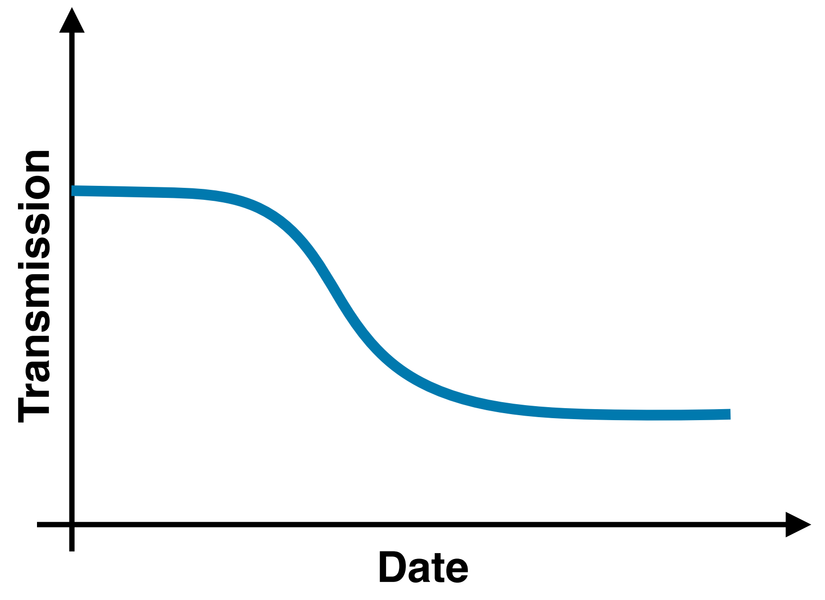
- $b(t)$ modeled by a logistic function during the first lockdown

$$b(t) = G(t) \overset{\text{def}}{=} b_M - \frac{(b_M - b_m)}{1 + e^{-\frac{(t - t_\ell)}{\tau}}}$$

dynamics →

$$\mathrm{d}(\log b) = g(t)\mathrm{d}t + \mathrm{d}v(t)$$

Wiener process



- State transformation ("Twisted" UKF)

$$\psi(x) = \mathrm{logit}\left(\frac{E(0)}{N}\right), \mathrm{logit}\left(\frac{I(0)}{D_q N}\right), \mathrm{logit}\left(\frac{R(0)}{N}\right), \mathrm{logit}\left(\frac{A(0)}{N}\right), \mathrm{logit}\left(\frac{H(0)}{N}\right).$$

- Parameters

$$\theta = (\log(D^i_q), \log(b_M), \log(b_m), \log(\tau), \log(t_\ell), \mathrm{logit}(E(0)), \mathrm{logit}(I(0)), \mathrm{logit}(H(0)))^\intercal$$

- Observations: Incident number of cases tested positive rE/D_E and Incident number of hospitalized I/D_q for each region

For estimation of initial conditions

$$P^i_0 = \begin{pmatrix}
\sigma^2_{E_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{E_0} & 0 & 0 \\
0 & \sigma^2_{I_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{I_0} & 0 \\
0 & 0 & \sigma^2_{R_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^2_{A_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^2_{H_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{H_0} \\
0 & 0 & 0 & 0 & 0 & \sigma^2_{b_M} & 0 & \sigma^2_{b_M} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{D_q} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma^2_{b_M} & 0 & \sigma^2_{b_M} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{b_m} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\tau} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{t_\ell} & 0 & 0 & 0 \\
\sigma^2_{E_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{E_0} & 0 & 0 \\
0 & \sigma^2_{I_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{I_0} & 0 \\
0 & 0 & 0 & 0 & \sigma^2_{H_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{H_0}
\end{pmatrix}$$

Covariance per department / region

Estimation using our Kalman filter in 3 steps (first lockdown + regions)

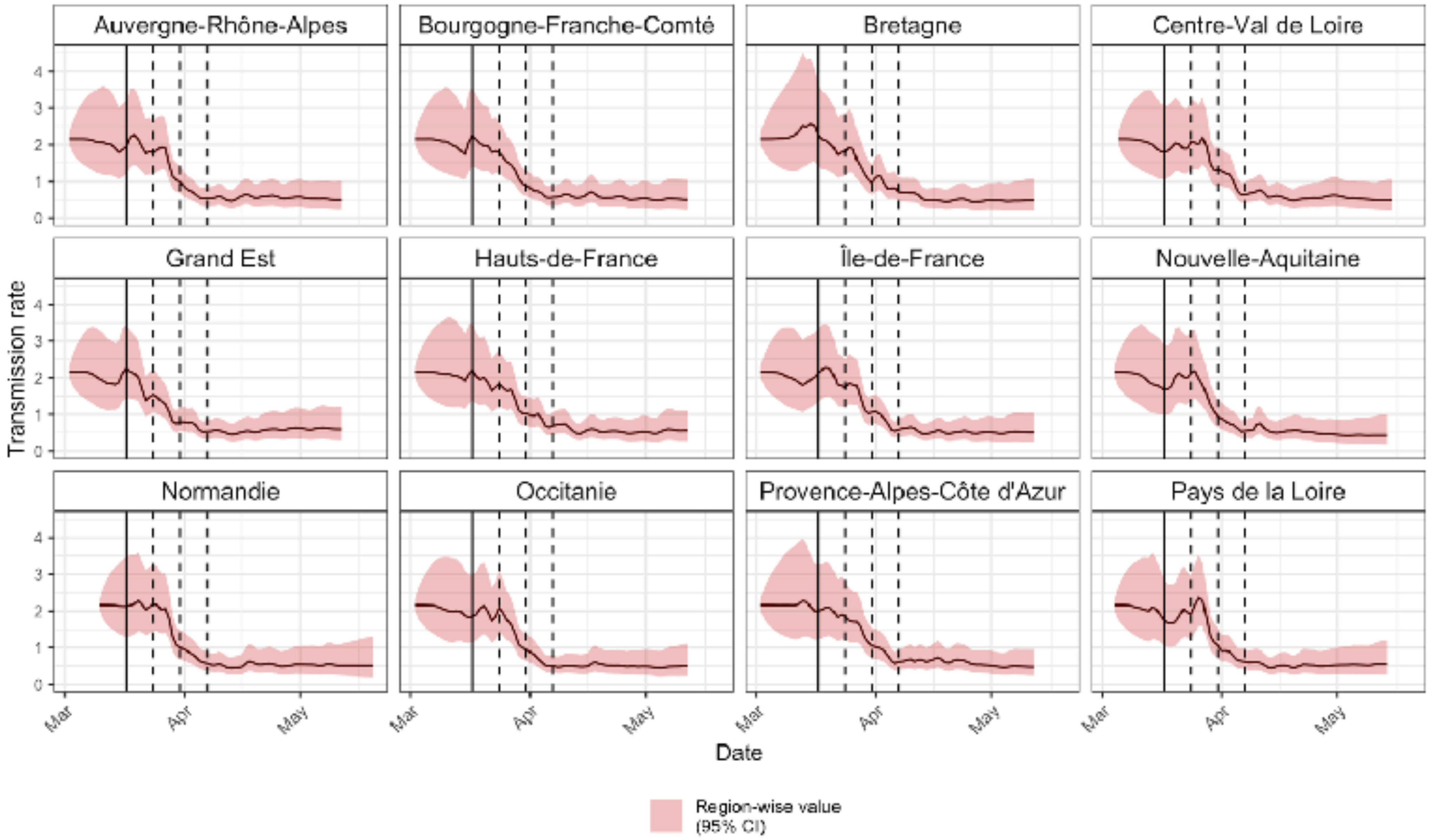
1. Estimation of region-wise model parameters and national weighted averages using logistic function b

$$\mathrm{d}(\log b) = g(t)\mathrm{d}t$$

Region	b_M	b_m	τ	t_ℓ	D_q	I_0	H_0	E_0
Auvergne-Rhône-Alpes	–	0.54 [0.50;0.58]	2.49 [1.51;3.47]	22.22 [20.91;23.53]	–	4 [1;14] (vs 9)	9 [2;34] (vs 9)	2157 [1667; 2790]
Bourgogne-Franche-Comté	–	0.56 [0.52;0.60]	2.56 [1.58;3.53]	21.15 [19.81;22.49]	–	4 [1;12] (vs 8)	2 [0; 8] (vs 2)	1292 [944; 1767]
Bretagne	–	0.53 [0.48;0.58]	3.82 [2.81;4.82]	20.30 [18.67;21.94]	–	5 [2;12] (vs 13)	5 [1;19] (vs 5)	45 [16; 128]
Centre-Val de Loire	–	0.55 [0.50;0.60]	2.80 [1.82;3.77]	21.91 [20.61;23.20]	–	1 [0; 3] (vs 2)	4 [1;15] (vs 4)	790 [611; 1021]
Grand Est	–	0.55 [0.50;0.59]	2.68 [1.76;3.60]	20.03 [18.70;21.36]	–	6 [2;18] (vs 12)	24 [6;95] (vs 25)	6752 [5733; 7951]
Hauts-de-France	–	0.56 [0.52;0.60]	3.31 [2.34;4.27]	21.91 [20.51;23.30]	–	4 [1;14] (vs 9)	4 [1;15] (vs 4)	1794 [1315; 2448]
Île-de-France	–	0.52 [0.48;0.56]	3.10 [2.19;4.00]	22.61 [21.47;23.73]	–	16 [5;49] (vs 34)	4 [1;15] (vs 4)	8506 [6669;10849]
Normandie	–	0.49 [0.44;0.55]	2.18 [1.23;3.14]	16.33 [14.77;17.88]	–	4 [1;11] (vs 9)	8 [2;34] (vs 9)	1335 [1105; 1611]
Nouvelle-Aquitaine	–	0.52 [0.46;0.56]	2.34 [1.33;3.33]	21.11 [19.77;22.45]	–	2 [1; 7] (vs 5)	1 [0; 4] (vs 1)	789 [546; 1141]
Occitanie	–	0.49 [0.44;0.54]	2.42 [1.44;3.40]	22.10 [20.81;23.37]	–	3 [1; 9] (vs 6)	3 [1;11] (vs 3)	906 [594; 1382]
Pays de la Loire	–	0.51 [0.46;0.56]	2.30 [1.37;3.23]	22.21 [21.10;23.31]	–	3 [1;10] (vs 7)	2 [0; 8] (vs 2)	575 [256; 1290]
Provence-Alpes-Côte d'Azur	–	0.59 [0.55;0.63]	3.02 [2.05;4.00]	22.05 [20.63;23.47]	–	4 [1;13] (vs 9)	2 [0; 8] (vs 2)	1124 [753; 1677]
France	2.16 [2.02;2.31]	0.53 [0.51;0.56]	2.78 [2.29;3.27]	21.49 [20.82;22.16]	0.26 [0.20;0.34]			

2. Estimation of b without the logistic function *a priori*

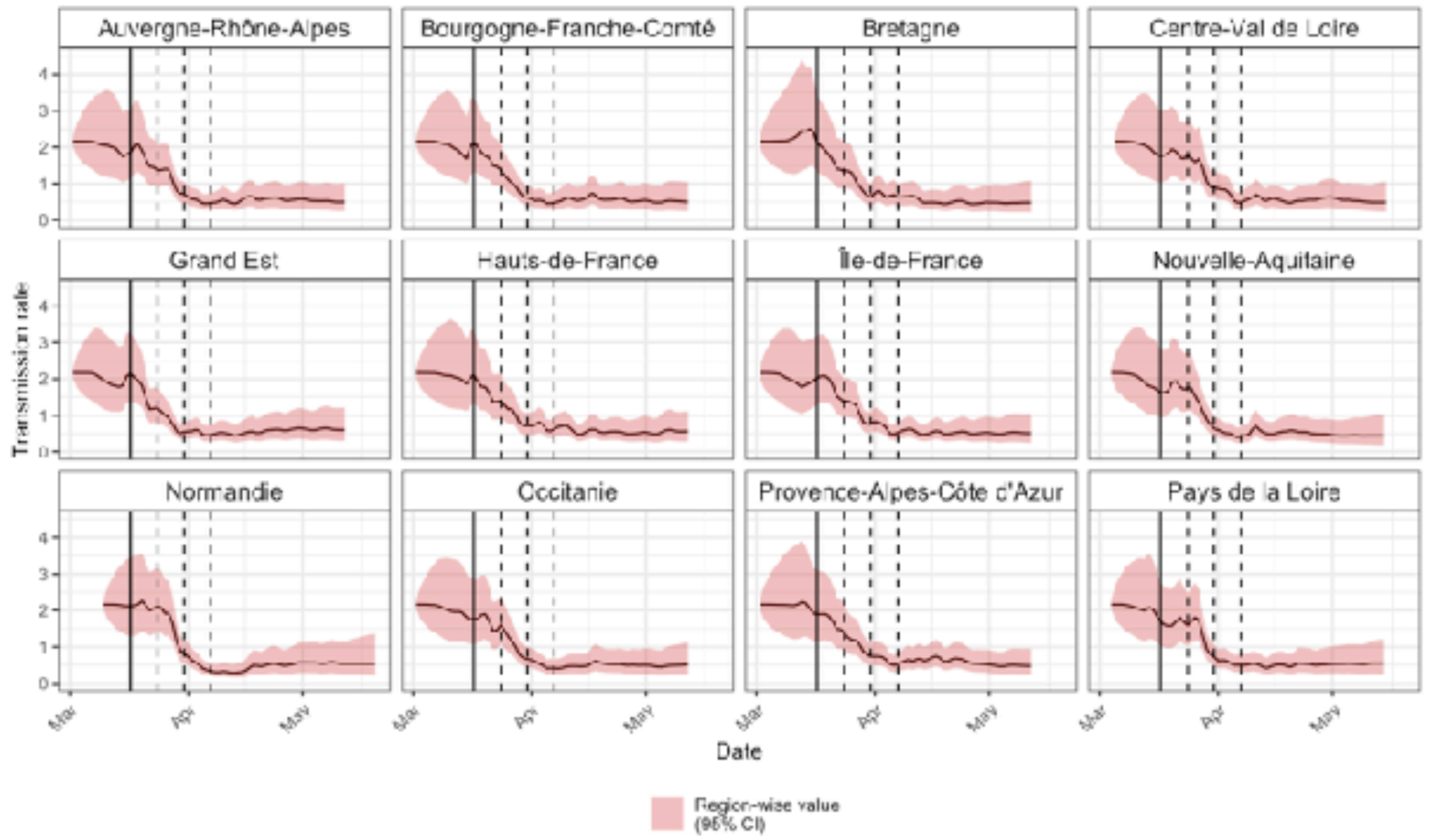
$$\mathrm{d}(\log b) = \mathrm{d}v(t)$$

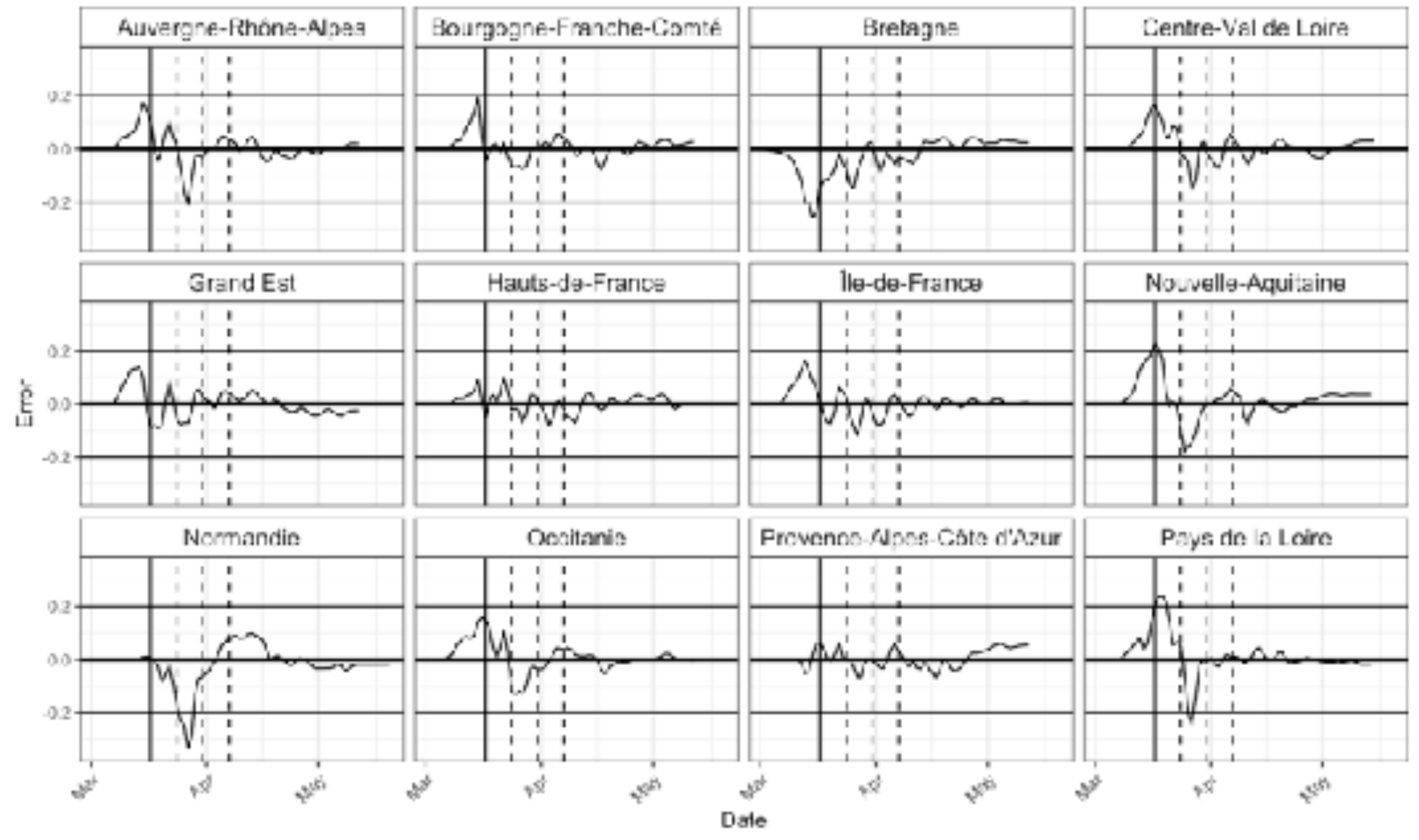


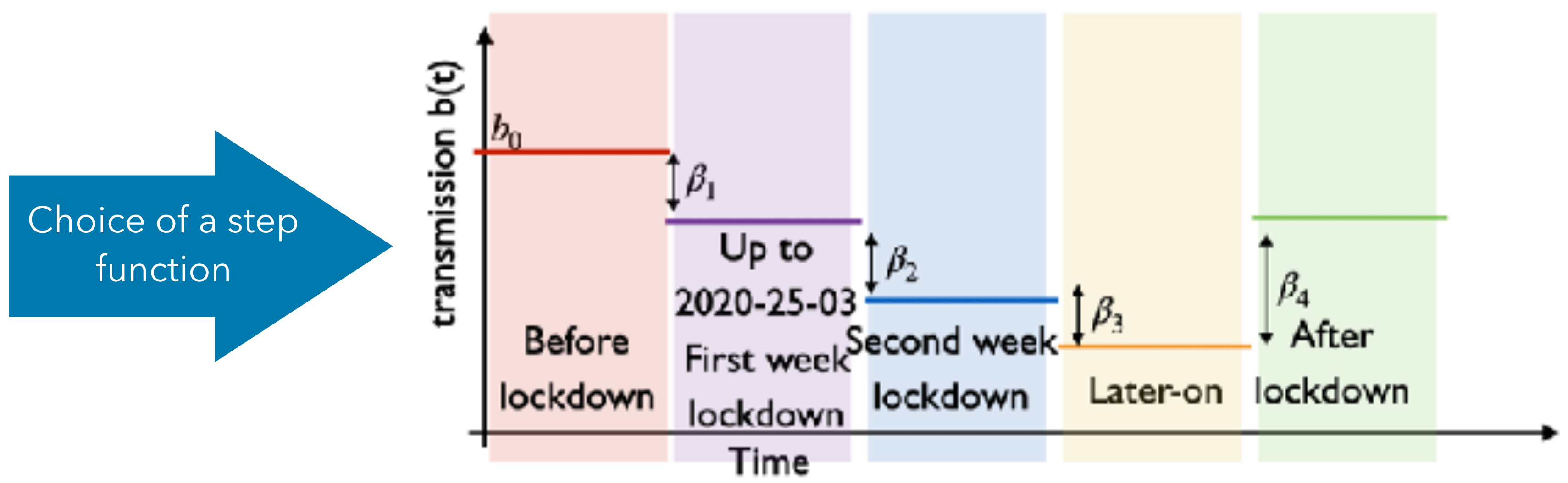
Estimation using our Kalman filter in 3 steps (first lockdown + regions)

3. Estimation of region-wise model parameters and national weighted averages using logistic function b

$$d(\log b) = g(t)dt + dv(t)$$







➡ *Final prediction of effective reproductive number*

$$R_e(t,\xi_i) = \frac{D_I b_i}{A(t,\xi_i) + I(t,\xi_i)} \left(\alpha A(t,\xi_i) + \frac{D_{q_i} I(t,\xi_i)}{D_I + D_{q_i}} \right)$$

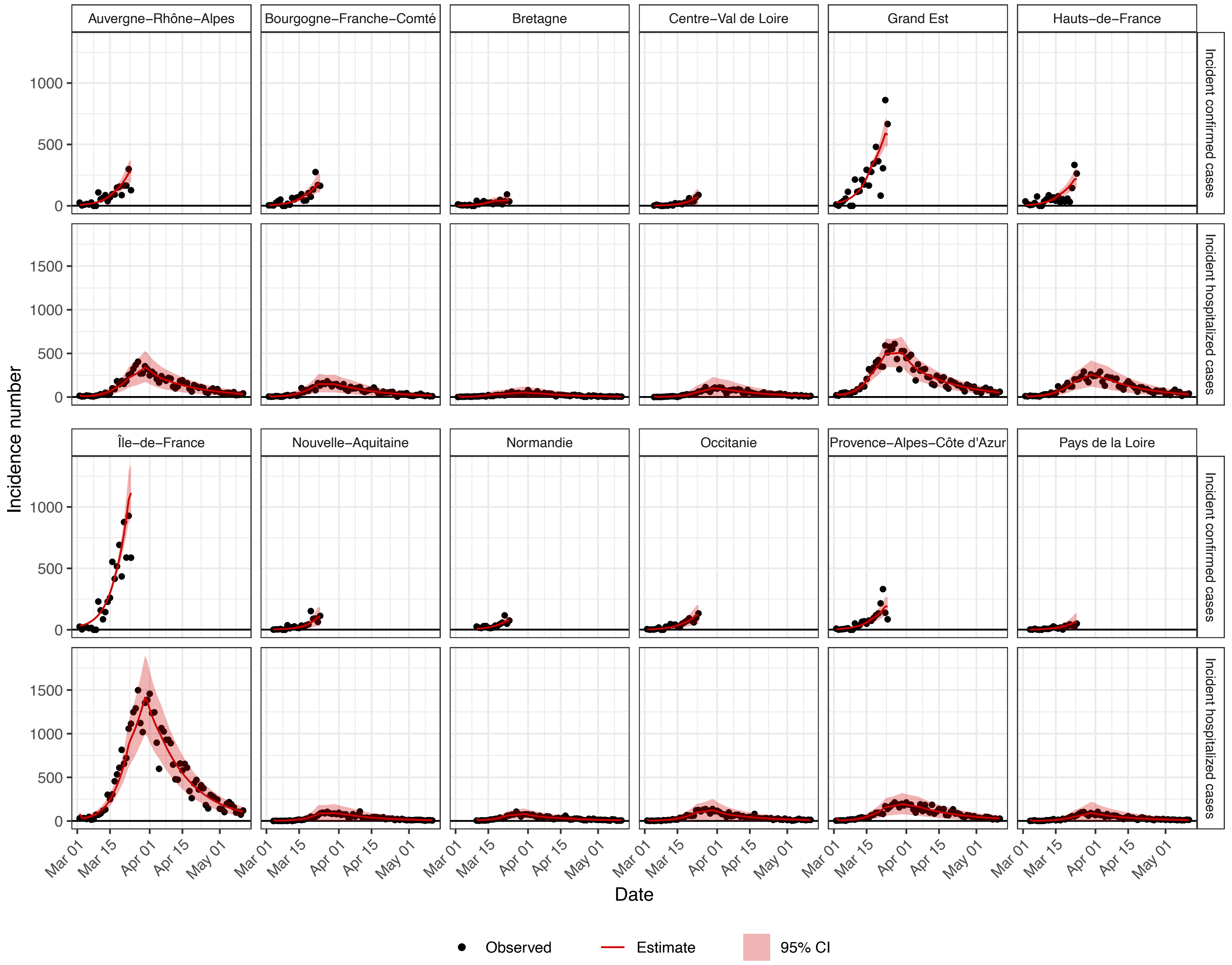
and attack rates = people at risk / population

Region Geographical region	R_e (before lockdown) on 2020-03-15	R_e (during lockdown) on 2020-03-29	R_e (end lockdown) on 2020-05-11	Attack rates: Infected proportion of the population in % computed as (E+I+A+H+R)/N 2020-03-17	 2020-05-11
Auvergne-Rhône-Alpes	2.20 [1.68;2.87]	1.01 [0.83;1.23]	0.62 [0.44;0.90]	0.28 [0.21;0.36]	3.85 [3.71;3.99]
Bourgogne-Franche-Comté	2.12 [1.63;2.77]	0.93 [0.76;1.15]	0.63 [0.43;0.92]	0.47 [0.37;0.60]	5.52 [5.32;5.75]
Bretagne	3.15 [2.40;4.11]	0.99 [0.79;1.24]	0.61 [0.40;0.91]	0.17 [0.12;0.23]	1.52 [1.45;1.60]
Centre-Val de Loire	2.38 [1.8;3.14]	1.33 [1.08;1.64]	0.62 [0.43;0.88]	0.20 [0.15;0.26]	3.74 [3.59;3.89]
Grand Est	2.30 [1.81;2.93]	0.71 [0.58;0.87]	0.76 [0.53;1.08]	1.19 [0.99;1.44]	9.47 [9.15;9.83]
Hauts-de-France	2.35 [1.81;3.04]	0.97 [0.79;1.19]	0.71 [0.50;1.00]	0.35 [0.27;0.45]	4.39 [4.23;4.56]
Île-de-France	2.38 [1.85;3.06]	1.02 [0.84;1.23]	0.65 [0.49;0.93]	0.74 [0.60;0.93]	9.94 [9.65;10.3]
Normandie	2.68 [2.14;3.35]	1.45 [1.19;1.75]	0.68 [0.46;0.99]	0.12 [0.11;0.14]	2.11 [2.02;2.21]
Nouvelle-Aquitaine	2.23 [1.67;2.97]	1.11 [0.91;1.36]	0.55 [0.37;0.83]	0.10 [0.08;0.14]	1.43 [1.37;1.49]
Occitanie	2.29 [1.77;2.96]	0.96 [0.79;1.18]	0.64 [0.43;0.95]	0.17 [0.14;0.22]	1.83 [1.76;1.91]
Pays de la Loire	2.55 [1.95;3.31]	1.18 [0.97;1.44]	0.70 [0.48;1.01]	0.16 [0.13;0.21]	2.10 [2.01;2.19]
Provence-Alpes-Côte d'Azur	2.59 [1.99;3.36]	1.11 [0.90;1.37]	0.60 [0.42;0.85]	0.33 [0.25;0.44]	4.10 [3.95;4.26]
France	2.40 [1.85;3.10]	1.04 [0.85;1.27]	0.65 [0.45;0.94]	0.41 [0.33;0.52]	4.90 [4.74;5.08]

Our results are comparable to the results obtained by Salje et al. (Pasteur institute); Crepey et al. (EHESP); Alizon et al. (ETE CNRS)

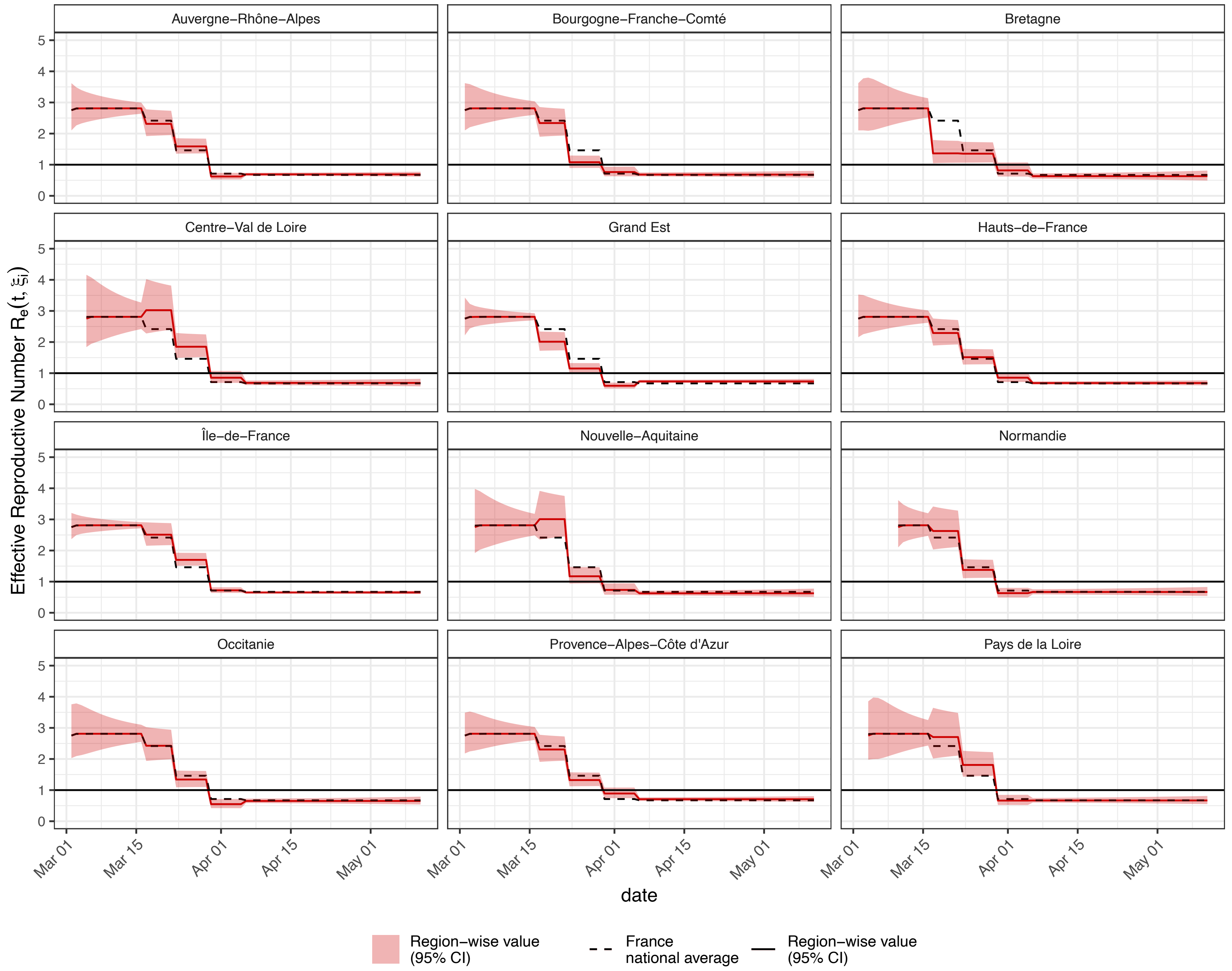
Estimation using SAEM approach (first lockdown + regions)

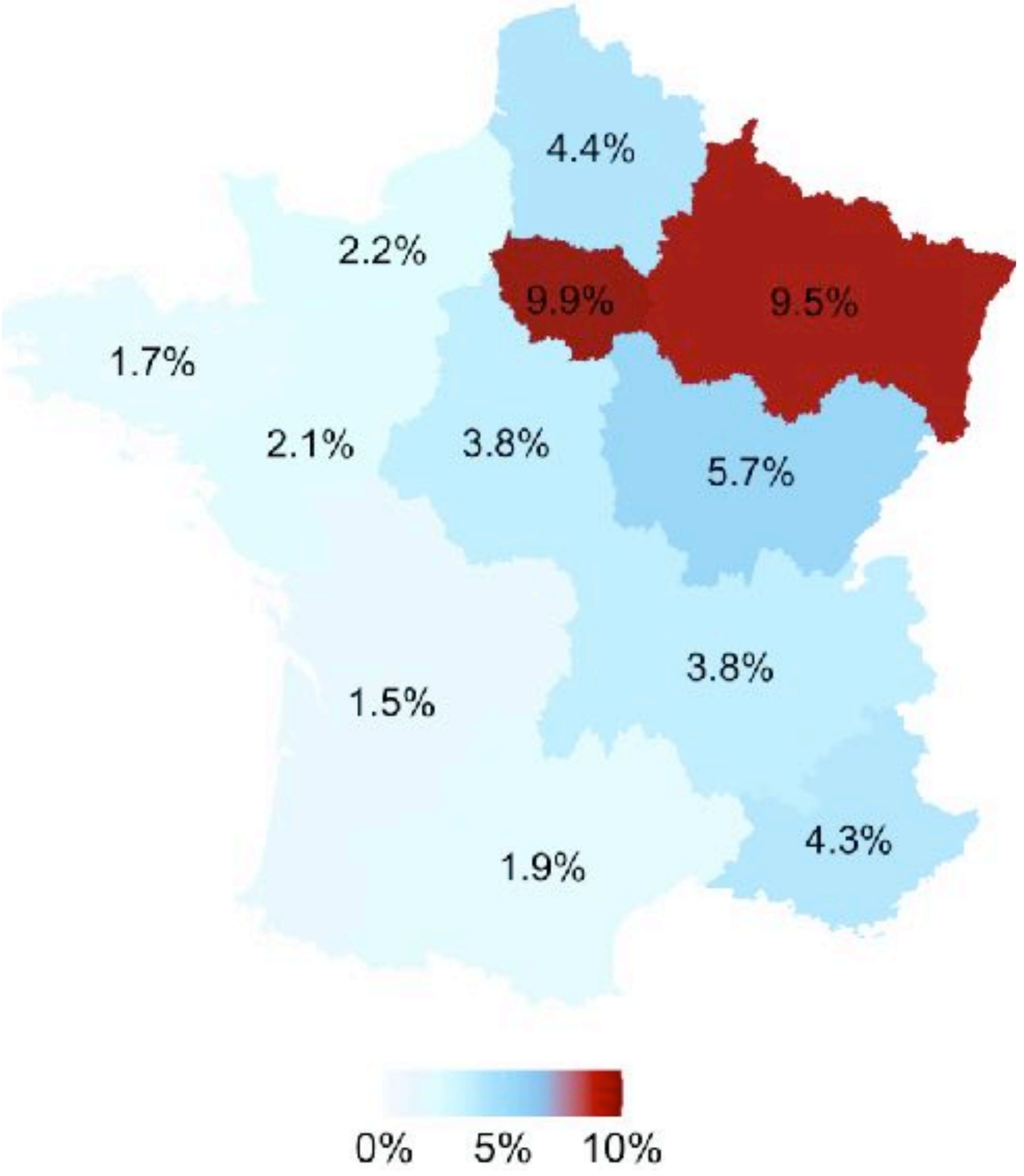
- In order to validate our Kalman estimation approach, an estimation using SAEM approach with a step function for *b* has been done.



Fitting curves of incident number of cases tested positive (rE/D_E) and Incident number of hospitalized (I/D_q) for each region with the SEIRAH model

Region specific Re compared to the national average. Lockdown started on March 17 .



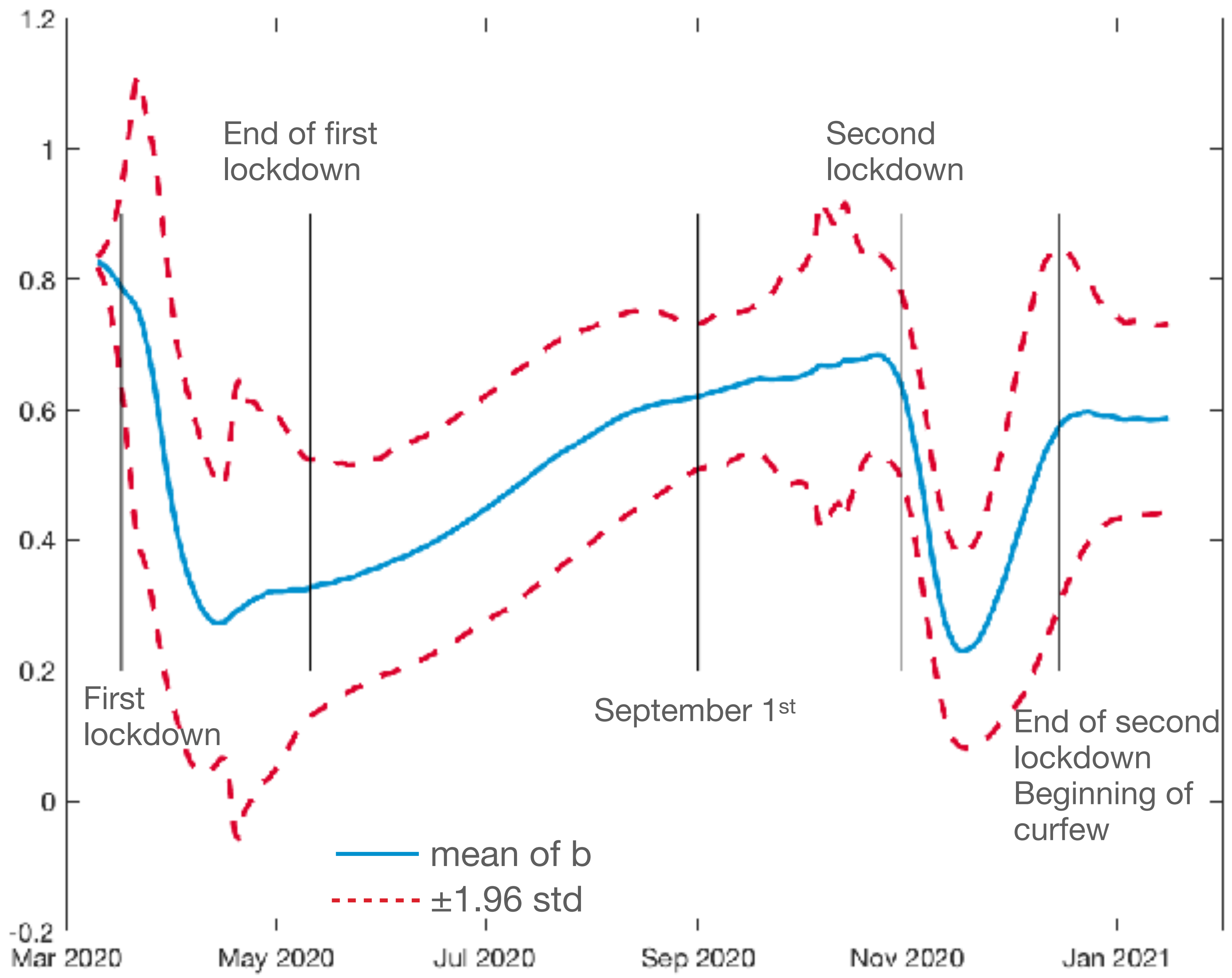


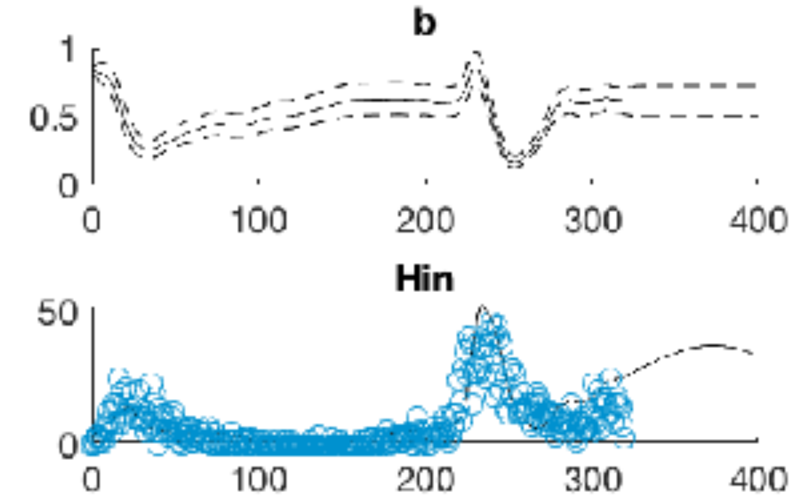
Model estimation for the proportion of Immunized individual in the population (deaths not taken into account), also referred as attack rate, on May 11th, 2020 (end of the first lockdown)

Our results are comparable to the results obtained by Salje et al. (Pasteur institute); Crepey et al. (EHESP); Alizon et al. (ETE CNRS)

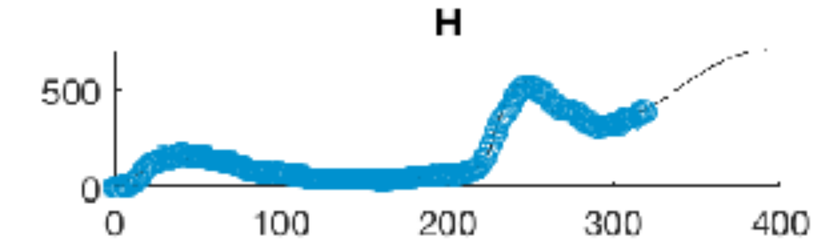
Departments results until today

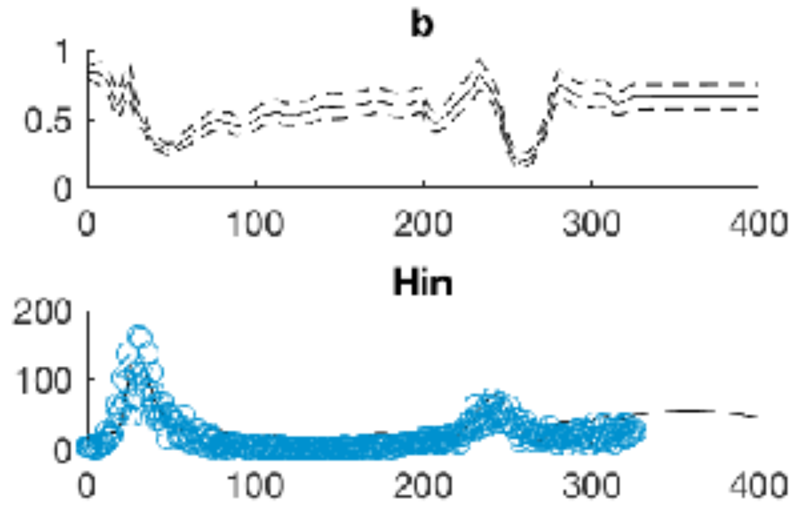
- Since the first lockdown, new data are available (Hospitalizations H and Incident number of hospitalized Hin)
- Refined level (94 departments instead of 12 regions)
- Reduced-order version of our population Unscented Kalman filter (on 12 clusters)
- Prediction assuming *b* stays constant (different values for each department) …



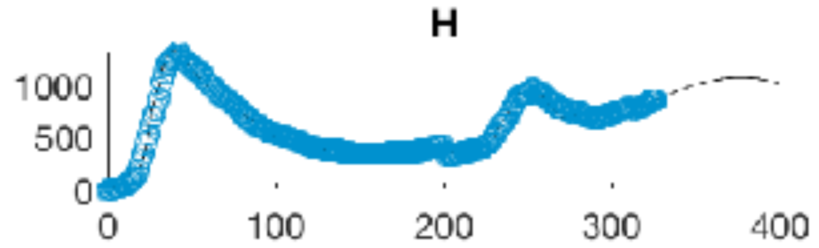


Ain (Auvergne-Rhône-Alpes)





Essone (Île-de-France)



	January 25th -> February 1st	January 25th -> February 8th
Number of hospitalisations	+2378	+4987

Conclusions and Perspectives

- Conclusions
  - Validation of our reduced-order population-based Kalman filter
    - Able to deal with mixed effects
    - Able to deal with error model
    - Very efficient in terms of computational times

  - Very interesting results obtained on the first lockdown in terms of immunity and effective reproducer number
  - Main limitation: dot not take into account travels between regions / departments

- Perspectives for the application
    - Evaluate the acceptability of lockdown strategy
    - Evaluate the predicting of second- and third- waves of COVID-19 using only the information of first wave
    - Account for vaccination and waning immunity

- Perspective for methodological developments
    - Deal with large systems (ODE or even PDE systems)