

Enseigner l'épidémiologie mathématique aux biologistes avec la COVID-19

Frédéric Hamelin

L'institut Agro de Rennes

Séminaire "Covid, statistique et enseignement supérieur"

Lyon, le 24 novembre 2021

JE NE SUIS PAS STATISTICIEN

ET ENCORE MOINS SPÉCIALISTE DE SANTÉ PUBLIQUE

Je suis enseignant-chercheur à L'institut Agro de Rennes
Nous formons des ingénieurs en sciences du vivant

Thématiques d'enseignement et de recherche :

- ▶ Épidémiologie **végétale** quantitative
- ▶ Écologie/évolution mathématique/**théorique**

Enseigner l'épidémiologie mathématique aux biologistes :

- ▶ Combiner **modèles mécanistes** et données
- ▶ Utiliser R* pour tirer parti des cours de statistiques

Les données sont l'interface pédagogique entre maths et bio

* Je suis débutant en R !

MODÉLISATION PÉDAGOGIQUE DE LA COVID-19

L'épidémie de Covid-19 est un puissant outil pédagogique

Les données officielles sont disponibles en ligne

Des modèles mécanistes simples permettent de répondre à des **questions ouvertes au début de l'épidémie**, telles que :

- ▶ quel est le taux de létalité du virus ?
- ▶ quelle est la fraction de la population immunisée ?

parmi beaucoup d'autres (qui font autant de TD).

L'inconnue principale est le nombre d'individus infectés (beaucoup d'asymptomatiques).

Ces TD permettent de montrer aux étudiants l'intérêt de la modélisation mathématique des épidémies.

FOCUS SUR LE PREMIER CONFINEMENT

Cet exposé se concentre sur la question suivante :

Quelle est la proportion de la population qui est immunisée à l'issue du premier confinement ?

Il s'agit d'estimer la distance à l'immunité collective (si elle existe) à une époque où les vaccins n'existent pas encore et où la proportion d'asymptomatiques est largement inconnue.

A cette époque, les données concernent les décès, les tests, et les admissions à l'hôpital.

Les tests de dépistage étaient relativement peu répandus.

Les **données hospitalières** pouvaient paraître plus fiables*.

*Cependant d'autres études/TD montrent qu'il était possible de se baser sur les données de tests (Roques et al 2020a,b).

ORIGINE DES DONNÉES

[Accueil](#) > [Jeux de données](#) > [SIVIC](#)

Données hospitalières relatives à l'épidémie de COVID-19 SIVIC

☆ 57 favoris

Mis à jour le 2 septembre 2021 — [Licence Ouverte / Open Licence version 2.0](#)[Producteur](#)[Métadonnées](#)[Actions](#)

Santé publique France

Santé publique France est l'agence nationale de santé publique. Créée en mai 2016 par ordonnance et décret, c'est un établissement public administratif sous tutelle du ministère chargé de la Santé. Sa mission : améliorer et protéger la santé des populations. Cette mission s'articule autour de...

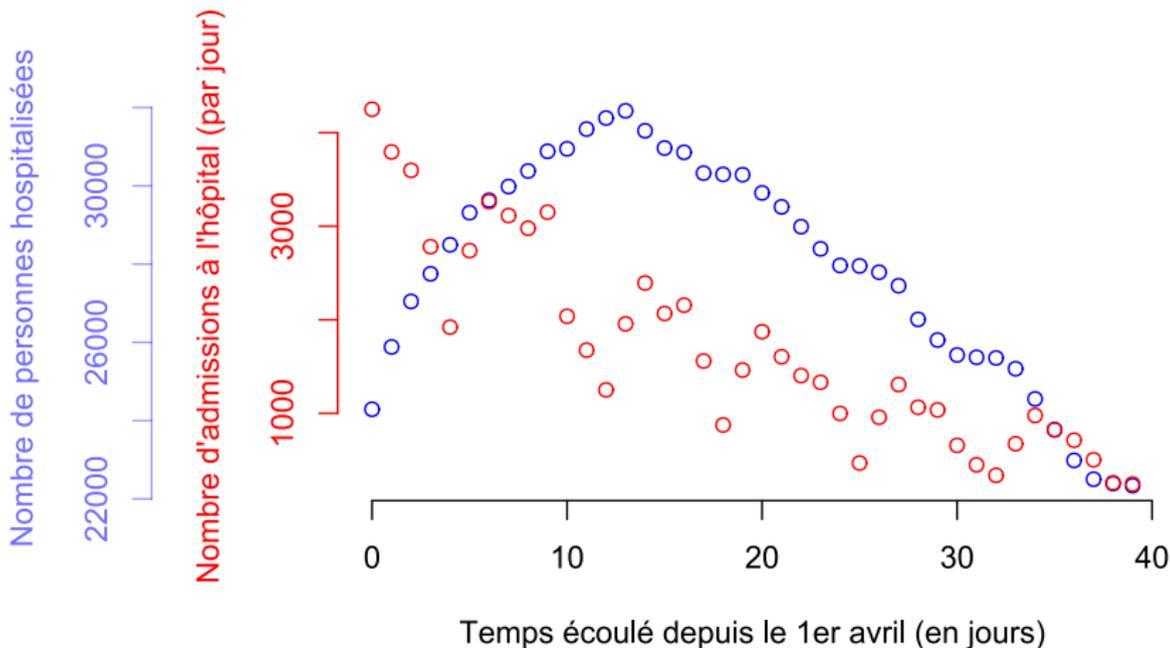
20 jeux de données

SOMMAIRE

- [Description](#)
- [Ressources](#) ⁽¹²⁾
- [Ressources communautaires](#) ⁽³⁸⁾
- [Réutilisations](#) ⁽¹⁴³⁾

DYNAMIQUE DES HOSPITALISATIONS

DURANT LE PREMIER CONFINEMENT (AVRIL-MAI 2020)

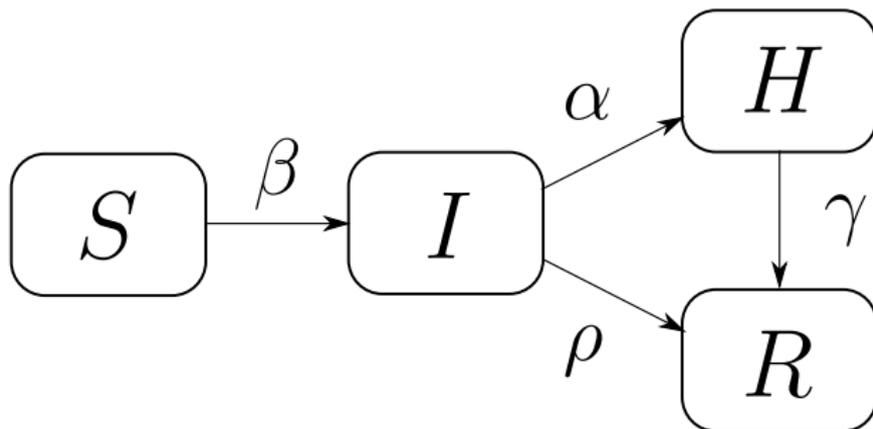


BASES DU MODÈLE MÉCANISTE

Hypothèses volontairement grossières pour simplifier l'exercice :

- ▶ la population française est **spatialement homogène** en France métropolitaine,
- ▶ les hommes et les femmes sont égaux vis-à-vis du virus,
- ▶ la structure en âge de la population peut être ignorée pour modéliser l'épidémie,
- ▶ les individus hospitalisés ne transmettent plus l'infection,
- ▶ la guérison confère une immunité permanente,
- ▶ **paramètres épidémiologiques constants** sur la période considérée (le premier confinement),
- ▶ le modèle est **déterministe**.

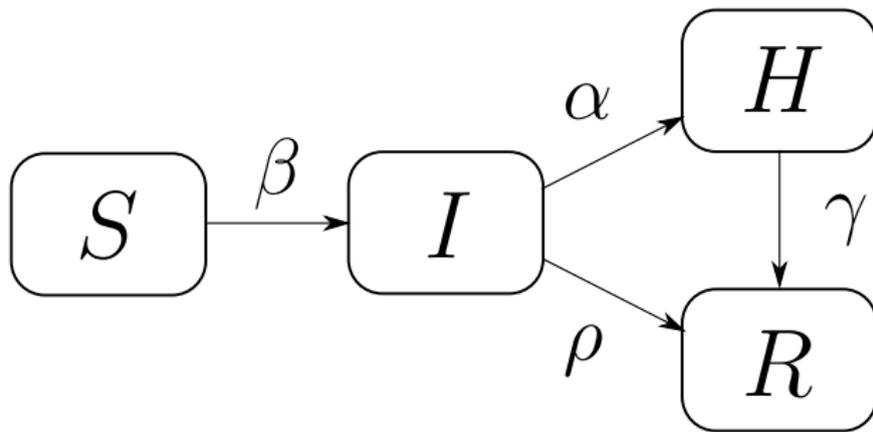
SCHÉMA ET NOTATIONS 1/2



Variables :

- ▶ $S(t)$: nombre de **sensibles** (non-infectés) à la date t
- ▶ $I(t)$: nombre d'**infectieux** non-hospitalisés à la date t
- ▶ $H(t)$: nombre d'infectés **hospitalisés** à la date t
- ▶ $R(t)$: nombre de **rétirés** de l'épidémie à la date t

SCHÉMA ET NOTATIONS 2/2



Paramètres :

- ▶ β : taux de **transmission** par unité de temps
- ▶ ρ : taux de **guérison** par unité de temps
- ▶ α : taux d'**admission** à l'hospital par unité de temps
- ▶ γ : taux de **sortie** de l'hôpital par unité de temps

MISE EN ÉQUATIONS

Variante du modèle SIR de base en épidémiologie humaine :

$$\frac{dS}{dt} = -\beta \frac{S}{N} I,$$

$$\frac{dI}{dt} = +\beta \frac{S}{N} I - (\alpha + \rho) I,$$

$$\frac{dH}{dt} = \alpha I - \gamma H,$$

$$\frac{dR}{dt} = \rho I + \gamma H,$$

où la taille de la population est

$$N = S + I + H + R = \text{constante}.$$

PARAMÈTRES IMPORTANTS

Reproductivité du virus (durant le confinement) :

$$\mathcal{R} = \frac{\beta}{\alpha + \rho}.$$

C'est le nombre moyen d'infections secondaires générées par un individu infecté dans la population :

- ▶ si $\mathcal{R} > 1$ une vague épidémique se forme
- ▶ si $\mathcal{R} < 1$ l'épidémie s'éteint progressivement

La **probabilité d'être hospitalisé** suite à l'infection est

$$\rho = \frac{\alpha}{\alpha + \rho}.$$

IDENTIFIABILITÉ STRUCTURELLE ET PRATIQUE

Le modèle mécaniste est **structurellement** observable et identifiable (H. et al 2021) pourvu que l'on fixe la valeur du paramètre ρ (le taux de guérison) qui est à peu près connu :

$$\rho = \frac{1}{10} \text{ par jour .}$$

Un individu non-hospitalisé guérit après 10 jours en moyenne.

En **pratique**, nous pourrions estimer $R(t) - R(0)$ (le nb de personnes immunisées pendant le confinement), mais pas $R(0)$. Néanmoins, nos résultats seront indépendants de la valeur donnée à $R(0)$ (le nombre initial d'immunisés) : 1%.

MODÈLE D'OBSERVATION

(LE PLUS SIMPLE QUI SOIT)

Soit $A(t)$ le nombre d'admissions théorique à la date t :

$$A(t) = \alpha I(t).$$

Le nombre d'admissions observé à la date t est :

$$\underline{A}(t) \sim \text{Poisson}(A(t)).$$

Idem pour le nombre d'hospitalisés observé à la date t :

$$\underline{H}(t) \sim \text{Poisson}(H(t)).$$

MAXIMISATION DE LA VRAISEMBLANCE

Appelons θ le vecteur des paramètres à estimer :

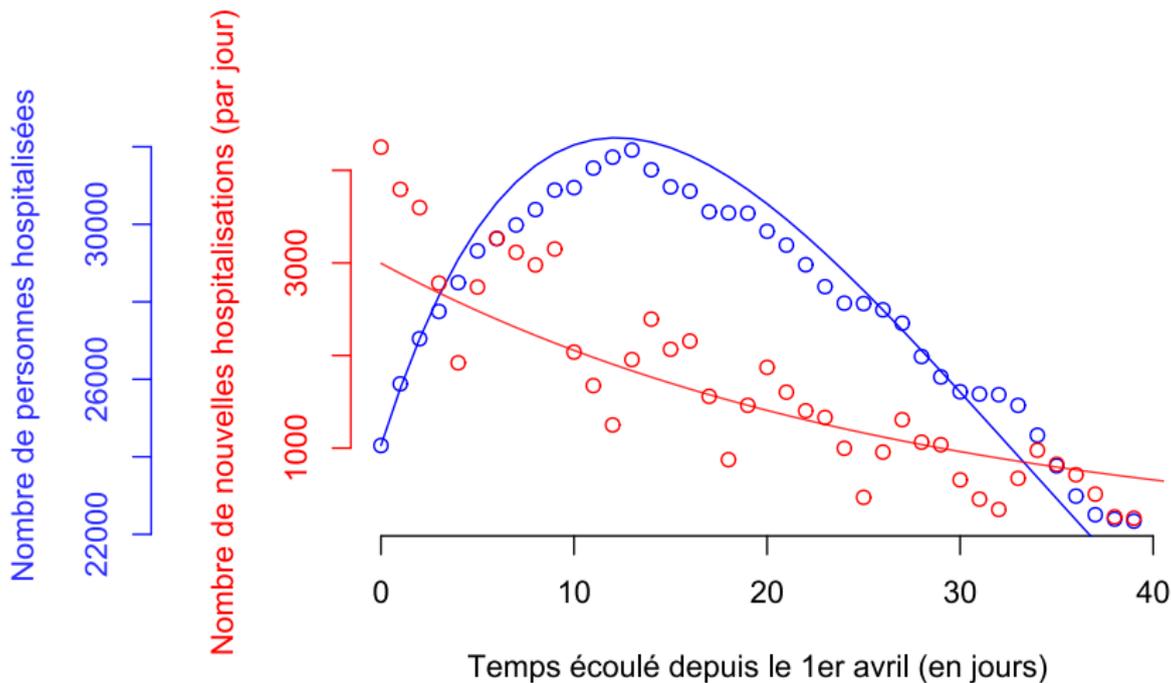
$$\theta = \{\alpha, \beta, \gamma, S(0), I(0), H(0)\}.$$

Sous l'hypothèse que les observations sont indépendantes conditionnellement au modèle épidémiologique, la vraisemblance s'écrit :

$$\mathcal{L}(\theta) = \prod_{t=0}^T \Pr(\underline{A}(t) | A_{\theta}(t)) \times \Pr(\underline{H}(t) | H_{\theta}(t)),$$

où $\underline{A}(t)$, $\underline{H}(t)$ sont les observations, et $A_{\theta}(t)$, $H_{\theta}(t)$ les valeurs théoriques (qui dépendent de θ).

CONFRONTATION DU MODÈLE AUX DONNÉES



ESTIMATIONS DES PARAMÈTRES

On trouve au passage :

- ▶ Reproductivité (en confinement) : $\mathcal{R} \approx 0,65$
- ▶ Probabilité d'être hospitalisé suite à l'infection : $p \approx 0,07$
- ▶ Temps moyen d'hospitalisation : $\gamma^{-1} \approx 17$ jours

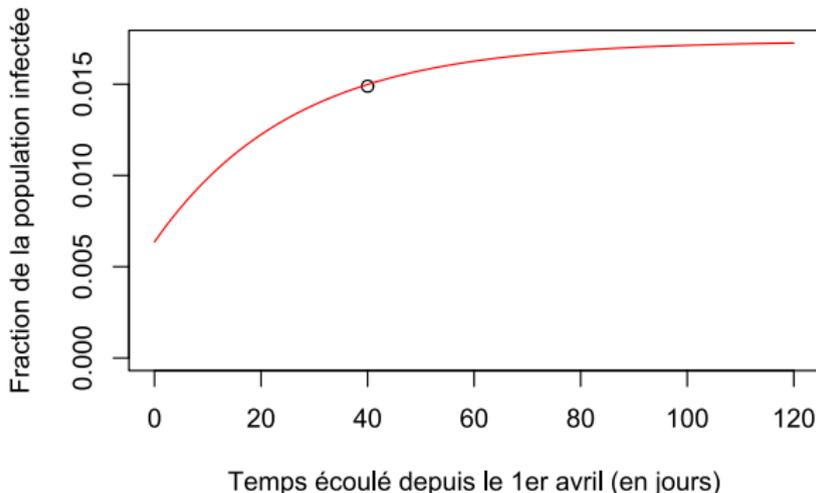
La valeur de \mathcal{R} est comparable avec celles estimées par

- ▶ Salje et al 2020 ($0,67 \pm 0,01$) avec données hospitalières,
 - ▶ Roques et al 2020b ($0,47 \pm 0,03$) avec tests et décès,
- et n'est pas surprenante en période de confinement ($\mathcal{R} < 1$).

FRACTION DE LA POPULATION IMMUNISÉE 1/2

La fraction de la population qui a été infectée durant le confinement au temps t est

$$F(t) = \frac{N - S(t) - R(0)}{N}.$$



FRACTION DE LA POPULATION IMMUNISÉE 2/2

D'après le modèle, 1,5% de la population est immunisée pendant le 1er confinement (960 000 personnes environ).

Cette valeur est comparable avec celles estimées par

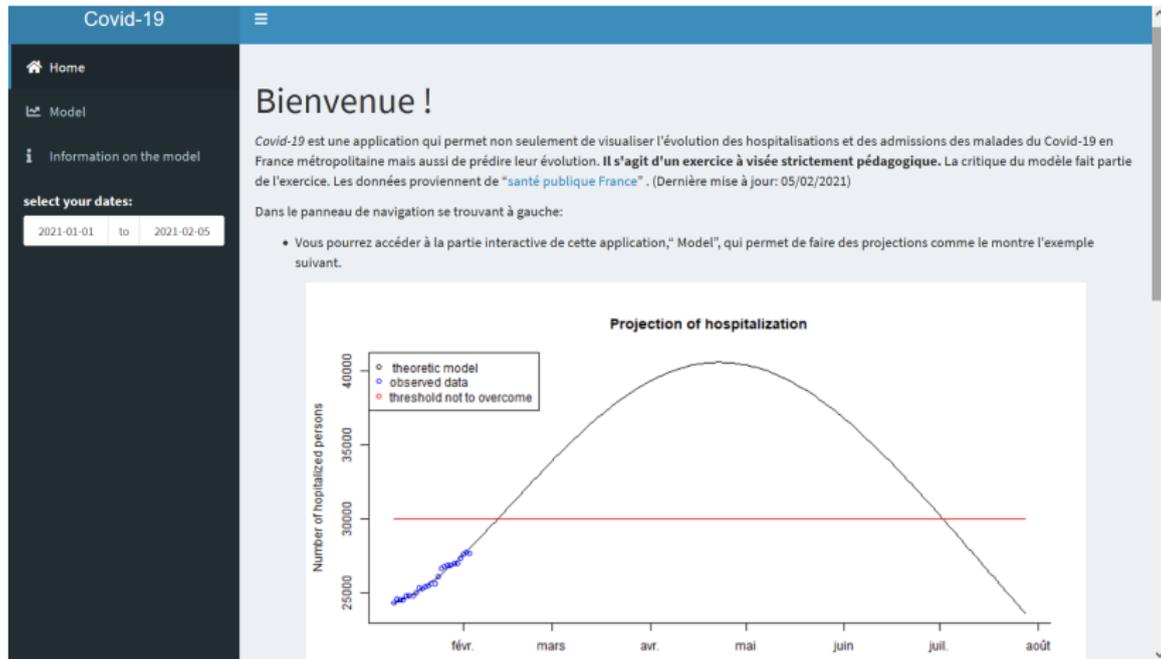
- ▶ Salje et al 2020 : 1% (0,5–2%) avec données hospitalières,
- ▶ Roques et al 2020b : 3,7% (3–4,8%) avec tests et décès.

En combinant avec leur étude de la première vague, Salje et al 2020 estiment une fraction totale d'immunisés de 5,3% (entre 3,3 et 9,3%) en sortie de confinement (encore loin de l'immunité de groupe espérée à l'époque : 66%).

POUR ALLER PLUS LOIN : APPLICATIONS WEB SHINY

COVID-19 : APPLI RÉALISÉE PAR MAIMOUNA DIARRA (M1 AGRO)

<https://demecology.shinyapps.io/Covid-19/>



RÉFÉRENCES

Lien vers d'autres TD et applis Shiny :

<https://fmhamelin.github.io/covidagro/>

Hamelin, F., Iggidr, A., Rapaport, A., Sallet, G., & Souza, M. (2021). Identifiability and observability of the SIR model with quarantine. arXiv preprint arXiv:2103.04614

Roques, L., Klein, E. K., Papaïx, J., Sar, A., & Soubeyrand, S. (2020). Impact of lockdown on the epidemic dynamics of COVID-19 in France. *Frontiers in medicine*, 7, 274.

Salje, H., Kiem, C. T., Lefrancq, N., Courtejoie, N., Bosetti, P., Paireau, J., ... & Cauchemez, S. (2020). Estimating the burden of SARS-CoV-2 in France. *Science*, 369(6500), 208-211.

MERCI ! DES QUESTIONS ?

Merci aux étudiants de l'Agro de Rennes (M1 2021) et du Master Modélisation en écologie (MODE), ici des M2 2021 :

