

Un apport de l'IA aux crises épidémiologiques Kiss the Covid

P. Mathieu^{a,1,*}, N. Mauhé^a

^a*Univ. Lille, CNRS, Centrale Lille, UMR 9189 – CRISTAL Lab, F-59000 Lille, France*

Abstract

Lors d'une crise sanitaire liée à l'apparition d'une maladie infectieuse, il est impératif d'avoir des systèmes d'aide à la décision des politiques publiques (évolution de la maladie, immunité collective, nombres dates et durées de périodes de confinement).

L'évolution d'une maladie infectieuse peut-être modélisée de différentes façons (approche centrée-individus, réseaux sociaux, equations mathématiques, modèles de flux).

Toutes ces modélisations necessite l'usage de paramètres pour calibrer le modèle. Plus il y a de paramètres, plus il est facile de coller aux données (tout comme il existe toujours un polynome de degrés n qui passe par n points).

Certains modélisateurs pensent que plus les paramètres sont précis (et nombreux), plus le modèle sera fiable. C'est un course sans fin.

L'important est de trouver l'essence du problème, le modèle le plus économe en paramètres (parcimonieux) rendant compte des phénomènes observés.

L'IA est d'une grande aide pour éviter les biais et régler au mieux ces modèles.

Keywords: Infectious disease, SIR modelling, Decision support system, artificial intelligence

2010 MSC: 00-01, 99-00

*Corresponding author

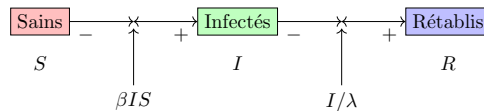
Email addresses: philippe.mathieu@univ-lille.fr (P. Mathieu),
nicolas.mauhe@univ-lille.fr (N. Mauhé)

1. Introduction (ce qui se fait en épidémiologie)

Depuis exactement un siècle et le papier séminal de Kermack et McKendrick, la modélisation épidémiologique a pris une part prépondérante dans l'étude de ces phénomènes complexes. La crise du H1N1 en 2008 a nettement amplifié ce phénomène, faisant apparaître un nombre important de modèles plus complexes les uns que les autres. Dans ce papier nous montrons que les techniques d'intelligence artificielle Utilisées à bon escient, Permettent d'une part de concevoir des modèles plus faciles à comprendre et à utiliser par les thématiques et d'autre part de faciliter grandement la conception d'outils d'aide à la décision pour les politiques sanitaires.

Le modèle présenté s'appuie sur des taux de contagion adaptable chaque jour, dont la valeur est automatiquement calculée par apprentissage génétique.

Le modèle SIR de Kermack et McKendrick [1] a été proposé en 1927 pour expliquer la hausse et la baisse rapides du nombre de patients infectés observées lors d'épidémies telles que la peste (à Londres en 1665, à Bombay en 1906) ou le choléra (à Londres en 1865). Il a été remis à l'avant-plan à la fin des années 1970 par Anderson et May [2].



Poisson process : ref ferguson.

2. L'apport de l'informatique

L'usage de l'IA est une méthode nouvelle pour l'épidémiologie, mais ancienne pour l'informatique.

Certains pensent que si on complexifie le modèle, on s'approche de la réalité : or plus il y a de paramètres, plus c'est facile de "fitter" : Problématique de l'overfitting (mettre une belle courbe qui passe par n points alors qu'en fait c'est une droite qui généralise le mieux).

Article	Type	Statut	Zone	Paramètres	Objectif
Benvenuto et al. [1] (ARIMA)	Statistical	Published	China	...	Prediction de l'incidence et de la prévalence
Zlojutro et al. [2]	Stochastic metapopulation epidemic model	Published	World	X cities, 4 compartments each	Number of cases and cities at risk
Chinazzi et al. [3] (GLEAM)	Stochastic metapopulation epidemic model	China	Published	3200 subpopulations, 4 compartments each	Impact of travel restriction
Scire et al. [4]	Poisson process	Published	Switzerland	...	Estimation of the reproductive number R
Li et al. [5]	Stochastic metapopulation epidemic model	Published	China	375 cities, 5 compartments each	Estimation of the prevalence and contagiousness
Flaxman et al. [6]	Statistical	Preprint	Europe	/	R0, Part de contaminés, Impact du confinement, Prévisions des décès
Ferguson et al. [7]	Agent-based	Preprint	Great Britain	58 millions agents ²	Prediction of ICU and deaths according to lockdown scenarios
Di Domenico et al. [8]	Compartmental epidemic model	Preprint	Paris region	11 compartments, 3 age classes	Estimation of propagation, Prediction of ICU, Deaths according to scenarios
Wilder et al. [9]	Agent-based model	Preprint	Hubei, Lombardy	10 millions agents	Impact of age distribution and familial household contacts on transmission
Simha et al. [10]	Compartment model	Preprint	Inde	3 compartments	Prediction of cases, Lockdowns impact
Roux et al. [11]	Compartmental model	Preprint	France	8 compartments x 17 age groups	Impact of lockdown on cases, deaths
Salje et al. [12]	...	Preprint	France	6 compartments x 8 age x 2 sex groups	Estimation du R0, et de la part de la pop infectée / décédée, impact du lockdown sur R0.
Team and Murray [13]	Statistical	Preprint	Europe & USA	...	Prediction of deaths and hospital demand
Luo [14]	Compartmental model	Preprint	Singapour, USA, Italie, Bresil	3 compartiments	Prediction du nombre de cas
Branas et al. [15]	Metapopulation epidemic model	Preprint	USA	3108 Counties, 5 compartments each	Prediction of hospital and ICU beds, and deaths
Keskinocak et al. [16]	Agent-based	Preprint	Georgia (USA)	1 million agents	Impact of Social Distancing sur décès et infections
Wang et al. [17] (STEM)	Compartmental model	Preprint	USA	3 compartments x 3,104 counties	Prediction of deaths and cases
Li et al. [18] (DELPHI)	Compartmental model	Documentation	World	11 compartments	Impact mesures, prédiction nombre de cas et nombre de morts
Woody et al. [19]	Statistical	Preprint	USA	50 states	Predict the number of deaths
Platen [20]	Compartment model	Preprint	Australia	4 compartments	Propagation number, Predicted number of deaths and cases
Arenas et al. [21]	Metapopulation epidemic model	Preprint	Spain	19 regions, 7 compartments each	Prediction of number of cases
Sanche et al. [22]	Poisson model & Compartmental model	Preprint	Wuhan	7 compartments	Estimate epidemiological parameters (R0, incubation period, etc.)
Aleta et al. [23]	Agent-based	Preprint	Boston	85k agents	Impact of distancing measures
REF TUOMISTO	Agent-based	Preprint	Helsinki	...	Impact of simulated mobility restriction on hospitalized, ICU and Deaths

Table 1: Description des modèles épidémiologiques proposés durant l'épidémie de COVID-19

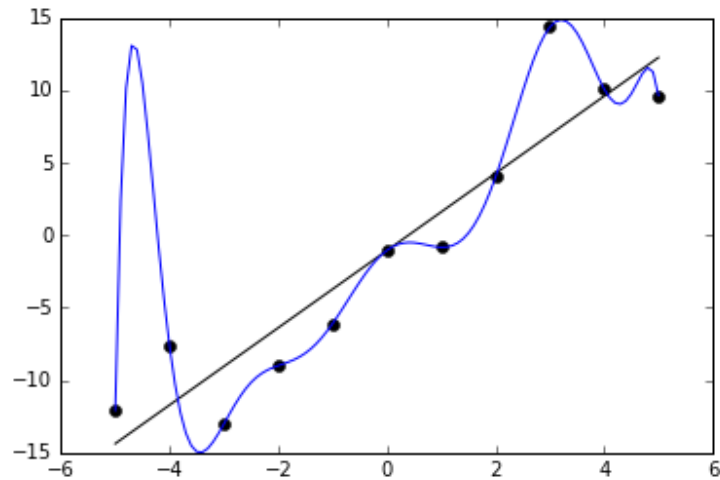


Figure 1: ©<https://en.wikipedia.org/wiki/Overfitting> Des données bruitées (approximativement linéaires) sont approximées par une fonction linéaire et une fonction polynomiale. Bien que la fonction polynomiale soit parfaitement ajustée, on constate que la fonction linéaire généralise surement bien mieux

Un modèle qui ne fonctionne pas (parce que par exemple des R0 sont délirants) peut très bien néanmoins “fitter” les données et donc continuer à prédire correctement les personnes en Réa

30 L'apport de l'informatique :

- des techniques et des outils : algos d'apprentissage
- s'attacher à l'essence du problème : principe de parcimonie

3. Cinq principes issus de l'informatique

(Ici on peut décomposer en 1) définition du problème en épidémiologie puis
35 2) solutions apportées par l'informatique)

3.1. Nécessité

Les modèles d'épidémie vont du simple modèle SIR que l'on peut résoudre sur un coin de feuille à des modèles réunissant des centaines d'agents indépendants,

placés sur des cartes reproduisant des pays entiers (REF nécessaire). Les progrès
40 informatiques rendent possible toutes les constructions auxquelles les épidémiologistes
peuvent penser. Quel doit être le principe qui guide les choix des modélisateurs
?

Un principe central dans la littérature informatique est celui de *nécessité*.
Chaque élément d'un modèle doit être nécessaire, et cette nécessité doit être
45 justifiée.

A détailler : interaction avec validation. 1) Rapport complexité / fitting 2)
Rapport justification / validation (si non validé, alors doit être justifié)

3.2. Calibration

Tout modèle qui vise à estimer un élément de la réalité non accessible (comme
50 le taux de propagation) ou à prédire un élément empirique doit être calibré sur
des éléments réels. Si ce principe n'est pas respecté, il est possible de produire
un modèle qui semble tout à fait répondre aux objectifs fixés mais qui décrit en
fait un tout autre phénomène. (Illustration modèle python).

Il s'agit d'un aspect central dans modélisation informatique et l'apprentissage
55 informatique. Les principes de calibration issus de cette littérature peuvent être
adaptés aux modèles épidémiologiques.

3.3. Approximation

Les possibilités induites par la modélisation donnent souvent l'illusion de la
précision. On lit par exemple dans la presse des estimations très précises après
60 la virgule produites par des modèles épiémiologiques portant sur des éléments
très difficiles à estimer : c'est le cas par exemple des estimations du nombre
d'infectés ou dut aux de propagation, souvent rapportées avec trois chiffres
après la virgule (REF 4,4% presse).

Ces analyses ignorent une question déterminante : le phénomène est-il cap-
65 turable entièrement par des moyens computationnels ? Parfois le mécanisme à
l'oeuvre est une dynamique changeante dont la prévision est hors de notre portée
: c'est le cas des mutations possibles du COVID-19 au cours de l'épidémie,

ou de sa réaction face à l'environnement. C'est particulièrement vrai lorsque l'on s'attaque à des dynamiques sociales, déterminantes dans le calcul du taux
70 de propagation. La difficulté de modéliser les comportements humains et leur aggrégation au niveau de la société est un challenge considérable, qui nourrit une littérature particulièrement active parmi les disciplines informatiques. Il est possible de tirer de nombreux enseignements de cette littérature à destination de l'épidémiologie : le principal est sans doute le recul nécessaire relatif à la
75 précision des estimations et des prévisions des modèles impliquant des décisions humaines (REF article JASSS).

Un autre aspect relativement méconnu du grand public relatif aux modèles épidémiologiques est la sensibilité aux paramètres initiaux. Un changement relativement faible des conditions initiales peut conduire à des modélisations
80 très différentes. Il s'agit d'un aspect bien connu des disciplines informatiques, étudié notamment par les sciences de la complexité.

Le problème vient du fait que l'approximation n'est pas une fonction "native" des nouveaux outils proposés à l'épidémiologie. La simulation informa-
85 tique d'un modèle tel que les modèles par agents ou les modèles compartimentaux nécessitent des input fixes et donnent des résultats tout aussi fixes : une prédiction telle que "2452 décès" constituera alors un résultat. C'est donc une étape supplémentaire que d'insufler l'approximation à ces modèles là. Soit en transmettant la véritable approximation des entrées, en réalisant
90 plusieurs scénarios avec des valeurs différentes. Soit mathématiquement, en transformant une entrée en distribution de probabilité ce qui permet d'obtenir un résultat aléatoire, que l'on pourra alors caractériser à l'aide d'intervalles de confiance. Soit, si l'approximation provient de mécanismes stochastiques à l'intérieur du modèle, en procédant à un grand nombre de simulations qui
95 permettront d'estimer le caractère aléatoire du résultat, par exemple via une moyenne et un écart-type du nombre de décès prédits. Dans de nombreux cas, la modélisation réclamera une combinaison de ces éléments là.

3.4. Validation

L'épidémie de COVID-19 peut être modélisée de manière très différente, même si l'on s'accordait sur les caractéristiques de base du virus. La question déterminante est donc de savoir si une modélisation est fidèle à la réalité qu'elle prétend simuler, voire prévoir. Il s'agit d'une problématique centrale en modélisation informatique, celle de la *validation*.

La littérature informatique peut nous aider à nous prémunir contre un certain nombre d'écueils, comme celui de l'*overfitting*. (...)

Pour aider à valider un modèle épidémiologique, plusieurs outils issus de la littérature informatique sont disponibles. Un exemple est celui de la *cross-validation*. (...)

Attention : manipulation graphique (aplatissement : ROUX, colonnes : Luo)

Validation : 2 modèles :

1) avec points au hasard 2) prédiction des derniers jours et comparaison avec la réalité (5 jours par exemple)

3.5. Réplicabilité

4. Présence de ces principes dans les modèles actuels

Replicabilité : analyse du coût : 0 : Impossible ✓ : Possible mais coût très élevé pour différentes raisons, avec plusieurs scénarios possibles: - Tout est détaillé et clair mais pas de code - Le code est disponible, tout est disponible, mais pas clair du tout et travail très important à mener

Checkmark : Réalisable, clair, expliqué, code disponible

[ANALYSE UNIQUEMENT POSITIVE]

Approximation :

Mentionner [1] : "The correlogram reporting the ACF and PACF showed that both prevalence and incidence of COVID- 2019 are not influenced by the seasonality."

Article	Nécessité	Calibration	Approximation	Validation	Réplicabilité
Benvenuto et al. [1] (ARIMA)		✓			✓
Zlojutro et al. [2]	~	✓	✓	...	✓
Chinazzi et al. [3] (GLEAM)		✓	✓	...	✓
Scire et al. [4]		✓	✓		✓
Li et al. [5]		✓	✓		✓
Flaxman et al. [6]	✓	✓	✓	✓	✓
Ferguson et al. [7]		✓			
Di Domenico et al. [8]	✓	✓	✓	✓	✓
Wilder et al. [9]	✓	✓	✓	✓	✓
Simha et al. [10]	✓	✓	✓	✓	✓
Roux et al. [11]		✓	✓	✓	
Salje et al. [12]		✓	✓	✓	
Team and Murray [13]		✓	✓		
Luo [14]	✓	✓	✓		
Branas et al. [15]		✓	✓		
Keskinocak et al. [16]		✓		✓	
Wang et al. [17]		✓	✓	✓	✓
Li et al. [18]	✓	✓			✓
Woody et al. [19]		✓	✓	✓	
Platen [20]					
Arenas et al. [21]		✓	✓	✓	✓
Sanche et al. [22]		✓	✓	✓	
Aleta et al. [23]		✓	✓		
REF TUOMISTO	✓	✓		✓	✓

Table 2: Présence des cinq principes dans les modèles épidémiologiques proposés durant l'épidémie de COVID-19

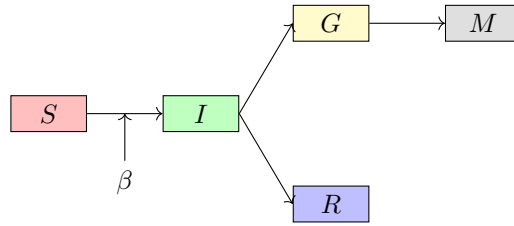


Figure 2: Le modèle épidémiologique par compartiments SIRGM

125 5. Un exemple simple

5.1. *Fitting d'un modèle simple*

Il s'agit d'un modèle SIRGM : une extension du modèle SIR aux deux catégories qui intéressent particulièrement les décideurs en ce qui concerne le COVID-19, à savoir le nombre de personnes en réanimation (R) et le nombre de
130 décès (M). Incidemment, il s'agit des catégories pour lesquelles les données sont les plus fiables. Par opposition, le nombre d'infectés et le nombre d'hospitalisés souffrent d'un biais de sélection important et variable au cours du temps, dus à des changements de politiques publiques.

L'objet du modèle est de trouver les paramètres de ce modèle épidémiologique
135 de sorte qu'il reproduise le plus fidèlement possible les données réelles. D'un point de vue informatique, il s'agit de fitter le plus précisément possible les courbes du nombre de lits occupés en réanimation et du nombre de décès.

Certains paramètres du modèle épidémiologique ont fait l'objet d'études médicales : c'est le cas du temps de contagiosité et du taux de cas grave. Ils
140 sont donc fixés à priori, et nous verrons à la section suivante les conséquences de les laisser libres.

un autre paramètre fixé à priori est celui du nombre d'infecté initial. D'après les données, 3 personnes ont été officiellement depistées au début de la période, néanmoins le nombre d'infecté réel est probablement beaucoup plus élevé. Il
145 n'y a pas de moyen de connaître précisément ce chiffre : nous l'avons fixé arbitrairement à 200. Ce paramètre interagit principalement avec le taux de

propagation de la période initiale. Si le nombre d’infecté initial est moindre, le fitting de la courbe ajustera simplement le taux de propagation initial à la hausse et conservera les autres paramètres à un niveau très similaire. Il s’agit
150 d’un avertissement supplémentaire relatif à une interprétation trop précise des taux de propagation : ils dépendent crucialement d’informations dont nous ne disposons pas.

La propagation du virus est le premier paramètre de ce type de modèle (c’est le bêta en Figure 2, qui peut également être exprimé en termes de R_0).
155 Cependant, ce taux de propagation dépend du comportement humain, et donc des changements de politiques publiques en termes de confinement, de règles de distanciation sociale, etc. AU lieu de fixer un seul taux de propagation, nous en fixons donc quatre, soit un pour chaque grande période en France : la période initiale, une période d’une dizaine de jours précédant l’entrée en
160 vigueur réglementaire du confinement strict, le confinement lui même, puis le déconfinement. La période précédant le confinement a été marquée par des déplacements très importants de la population qui pouvait choisir son lieu de confinement, ce qui explique la nécessité de la prendre en compte.

Les autres paramètres sont les variables d’ajustement laissées libres pour
165 obtenir un modèle épidémiologique cohérent avec les données observées.

Le problème devient dès lors un problème d’optimisation, et plusieurs techniques s’ouvrent à nous pour fitter le modèle. Nous avons utilisé un algorithme génétique basique. Les résultats sont reportés en Table 3 et en Figure 3.

Le modèle fitte raisonnablement bien les données, et ce avec une structure
170 très simple. Les paramètres estimés sont cohérent avec la littérature scientifique portant sur le COVID-19.

5.2. Progression des médecins

Ce type de modèle peut néanmoins ignorer certains phénomènes qui viendraient changer les paramètres au cours du temps. C’est par exemple le cas des taux
175 de mortalité, qui peuvent être affectés par la progression du corps médical sur le sujet. Des entretiens avec des responsables du CHU de Lille nous ont in-

R0 avant	1.77
R0 pré-confinement	3.47
R0 confinement	0.52
R0 déconfinement	0.76
Infectés initiaux	200.0
Taux réanimation	0.005
Taux remis	0.1
Taux réanimation vers grave	0.025
Taux réanimation vers décès	0.064

Table 3: Paramètres du modèle SIGRM.

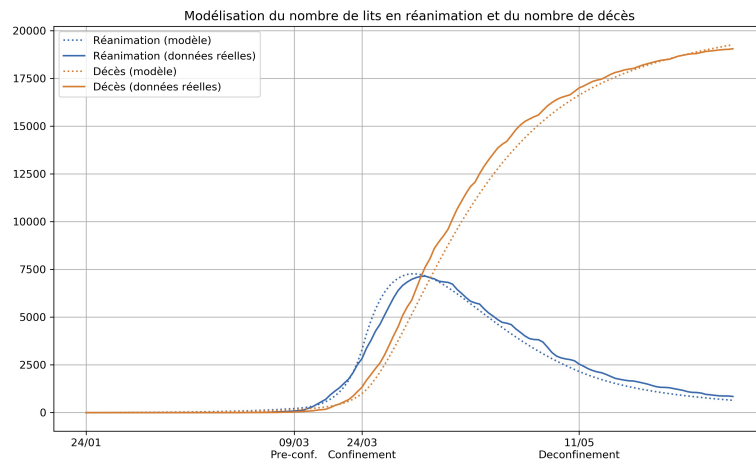


Figure 3: Le modèle SIGRM avec les paramètres donnés en Table 3 correspond aux données.

R0 avant	1.95
R0 pré-confinement	2.76
R0 confinement	0.64
R0 déconfinement	0.76
Infectés initiaux	200.0
Taux réanimation	0.005
Taux remis	0.1
Taux réanimation vers grave	0.025
Taux réanimation vers décès	0.07

Table 4: Paramètres du modèle intégrant la progression du traitement.

formé de l'amélioration du traitement des personnes en réanimation. Ce type de phénomène est compatible avec une modélisation légèrement différente de l'épidémie, comme en témoigne la Figure 4.

180 On observe que le modèle permet un meilleur fitting des entrées en réanimation si la courbe des décès est autorisée à décrocher progressivement de celle prévue par le modèle. Cela témoigne du fait qu'intégrer la progression des médecins améliorerait le fitting des données. Il y a donc un arbitrage à réaliser en termes de perte de simplicité versus l'amélioration de la prédictabilité du modèle.

185 5.3. Des paramètres laissés libres

Le danger de laisser trop de paramètres libres pour l'optimisation est d'obtenir une modélisation plus performante mais qui correspond à un tout autre virus. Prenons l'exemple du temps de contagiosité : on sait désormais que c'est une des caractéristiques singulières du COVID-19, qui présente des temps d'incubation et de contagiosité très importants comparés aux autres virus.

190 Si l'on laisse le temps de contagiosité libre dans la modélisation (via le taux S vers I), on peut obtenir une modélisation telle que décrite dans la Table 5 et la Figure 5.

Il s'agit de la modélisation d'un virus très faiblement mortel, qui présente un temps de contagiosité réduit (entre un et deux jours) et qui se serait répandu

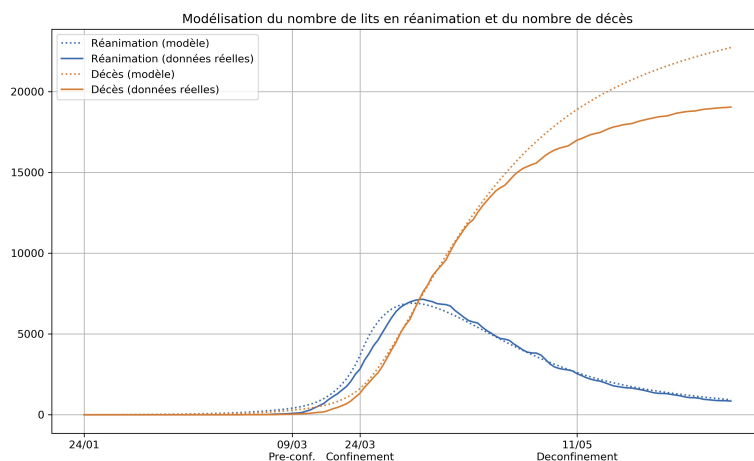


Figure 4: Les données réelles sont compatibles avec une modélisation épidémiologique intégrant la progression du traitement des patients en réanimation.

R0 avant	1.2
R0 pré-confinement	1.16
R0 confinement	1.14
R0 déconfinement	1.34
Infectés initiaux	200.0
Taux réanimation	0.001
Taux remis	0.781
Taux réanimation vers grave	0.006
Taux réanimation vers décès	0.062

Table 5: Paramètres clairement différents de l'épidémie de COVID-19

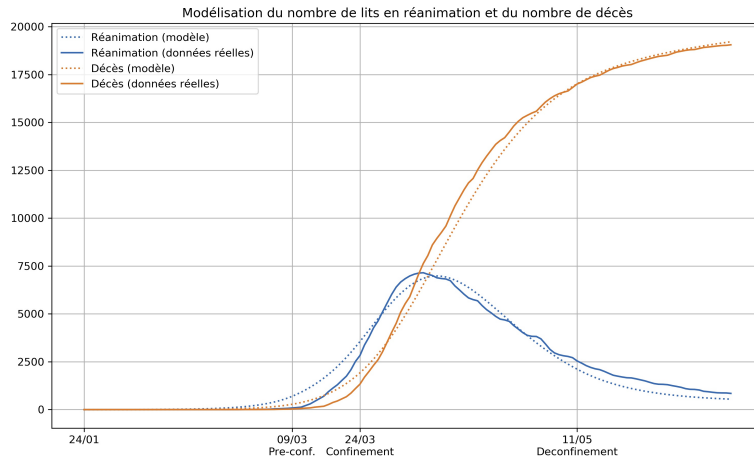


Figure 5: La modélisation issue des paramètres de la Table 5 donne de bons résultats.

dans la population avec un R_0 en permanence supérieur à 1. L'épidémie aurait atteint son pic et faiblirait désormais car le seuil de l'immunité collective serait atteint, et 88% de la population aurait contracté ce virus.

Cette modélisation correspond aux données observées : en l'absence de
 200 spécification volontaire de données médicales observées par ailleurs, il est donc
 tout à fait possible d'obtenir un modèle épidémiologique très différent de l'épidémie
 réelle, mais qui présenterait les mêmes statistiques observables.

References

- [1] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, Appli-
 205 cation of the ARIMA model on the COVID-2019 epidemic dataset, Data
 in Brief 29 (2020) 105340, ISSN 23523409, URL <https://linkinghub.elsevier.com/retrieve/pii/S2352340920302341>.
- [2] A. Zlojutro, D. Rey, L. Gardner, A decision-support framework to opti-
 mize border control for global outbreak mitigation, Scientific Reports 9 (1)

- (2019) 2216, ISSN 2045-2322, URL <http://www.nature.com/articles/s41598-019-38665-w>.
- [3] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y. Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, *Science* 368 (6489) (2020) 395–400, ISSN 0036-8075, 1095-9203, URL <https://science.sciencemag.org/content/368/6489/395>.
- [4] J. Scire, S. Nadeau, T. Vaughan, G. Brupbacher, S. Fuchs, J. Sommer, K. N. Koch, R. Misteli, L. Mundorff, T. Götz, T. Eichenberger, C. Quinto, M. Savic, A. Meienberg, T. Burkard, M. Mayr, C. A. Meier, A. Widmer, R. Kuehl, A. Egli, H. H. Hirsch, S. Bassetti, C. H. Nickel, K. S. Rentsch, W. Kübler, R. Bingisser, M. Battegay, S. Tschudin-Sutter, T. Stadler, Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft, *Swiss Medical Weekly* 150 (1920), URL <https://smw.ch/article/doi/smw.2020.20271>.
- [5] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), *Science* 368 (6490) (2020) 489–493, ISSN 0036-8075, 1095-9203, URL <https://science.sciencemag.org/content/368/6490/489>.
- [6] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan, H. Zhu, T. Berah, J. Eaton, P. Perez Guzman, N. Schmit, L. Cilloni, K. Ainslie, M. Baguelin, I. Blake, A. Boonyasiri, O. Boyd, L. Cattarino, C. Ciavarella, L. Cooper, Z. Cucunuba Perez, G. Cuomo-Dannenburg, A. Dighe, A. Djaafara, I. Dorigatti, S. Van Elsland, R. Fitzjohn, H. Fu, K. Gaythorpe, L. Geidelberg, N. Grassly, W. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, D. Jorgensen, E. Knock, D. Laydon, G. Nedjati Gilani, P. Nouvellet, K. Parag, I. Siveroni, H. Thompson, R. Verity, E. Volz,

- 240 C. Walters, H. Wang, Y. Wang, O. Watson, P. Winskill, X. Xi, C. Whittaker, P. Walker, A. Ghani, C. Donnelly, S. Riley, L. Okell, M. Vollmer, N. Ferguson, S. Bhatt, Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, Report, URL <http://spiral.imperial.ac.uk/handle/10044/1/77731>, 2020.
- 245 [7] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. Okell, S. Van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. Walker, P. Winskill, C. Whittaker, 250 C. Donnelly, S. Riley, A. Ghani, Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand, Tech. Rep., Imperial College London, URL <http://spiral.imperial.ac.uk/handle/10044/1/77482>, 2020.
- [8] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, V. Colizza, 255 Expected impact of lockdown in Île-de-France and possible exit strategies, preprint, Infectious Diseases (except HIV/AIDS), URL <http://medrxiv.org/lookup/doi/10.1101/2020.04.13.20063933>, 2020.
- [9] B. Wilder, M. Charpignon, J. Killian, H.-C. Ou, A. Mate, S. Jabbari, A. Perrault, A. Desai, M. Tambe, M. Majumder, The Role of Age Dis- 260 tribution and Family Structure on COVID-19 Dynamics: A Preliminary Modeling Assessment for Hubei and Lombardy, SSRN Electronic Journal ISSN 1556-5068, URL <https://www.ssrn.com/abstract=3564800>.
- [10] A. Simha, R. V. Prasad, S. Narayana, A simple Stochastic SIR model for COVID 19 Infection Dynamics for Karnataka: Learning from Europe, 265 arXiv:2003.11920 [math, q-bio] URL <http://arxiv.org/abs/2003.11920>, arXiv: 2003.11920.

- [11] J. Roux, C. Massonnaud, P. Crépey, COVID-19: One-month impact of the French lockdown on the epidemic burden, preprint, Epidemiology, URL <http://medrxiv.org/lookup/doi/10.1101/2020.04.22.20075705>, 2020.
- [12] H. Salje, C. T. Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hoze, J. Richet, C.-L. Dubost, Y. L. Strat, J. Lessler, D. Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boëlle, S. Cauchemez, Estimating the burden of SARS-CoV-2 in France (2020) 44.
- [13] I. C.-. h. s. u. f. Team, C. J. Murray, Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries, medRxiv (2020) 2020.04.21.20074732 URL <https://www.medrxiv.org/content/10.1101/2020.04.21.20074732v1>.
- [14] J. Luo, Predictive Monitoring of COVID-19 (2020) 12.
- [15] C. C. Branas, A. Rundle, S. Pei, W. Yang, B. G. Carr, S. Sims, A. Zebrowski, R. Doorley, N. Schluger, J. W. Quinn, J. Shaman, Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-CoV2) cases in US counties, preprint, Epidemiology, URL <http://medrxiv.org/lookup/doi/10.1101/2020.04.01.20049759>, 2020.
- [16] P. Keskinocak, B. E. O. Aglar, A. Baxter, J. Asplund, N. Serban, The Impact of Social Distancing on COVID19 Spread: State of Georgia Case Study, medRxiv (2020) 2020.04.29.20084764 URL <https://www.medrxiv.org/content/10.1101/2020.04.29.20084764v1>.
- [17] L. Wang, G. Wang, L. Gao, X. Li, S. Yu, M. Kim, Y. Wang, Z. Gu, Spatiotemporal Dynamics, Nowcasting and Forecasting of COVID-19 in the United States, arXiv:2004.14103 [stat] URL <http://arxiv.org/abs/2004.14103>, arXiv: 2004.14103.

- 295 [18] M. L. Li, H. T. Bouardi, O. S. Lami, N. Trichakis, M. F. Zarandi, D. Bertsimas, Overview of DELPHI Model V2.8 - COVIDAnalytics (2020) 6.
- [19] S. Woody, M. G. Tec, M. Dahan, K. Gaither, M. Lachmann, S. Fox, L. A. Meyers, J. G. Scott, Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile
300 phones, medRxiv (2020) 2020.04.16.20068163URL <https://www.medrxiv.org/content/10.1101/2020.04.16.20068163v2>.
- [20] E. Platen, Stochastic Modelling of the COVID-19 Epidemic, SSRN Scholarly Paper ID 3586208, Social Science Research Network, Rochester, NY, URL <https://papers.ssrn.com/abstract=3586208>, 2020.
- 305 [21] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Panos, B. Steinegger, A mathematical model for the spatiotemporal epidemic spreading of COVID19, medRxiv (2020) 2020.03.21.20040022URL <https://www.medrxiv.org/content/10.1101/2020.03.21.20040022v1>.
- 310 [22] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, R. Ke, High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2, Emerging Infectious Diseases journal 26 (7), URL https://wwwnc.cdc.gov/eid/article/26/7/20-0282_article.
- [23] A. Aleta, D. Martin-Corral, A. P. y. Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini, S. Merler,
315 A. Pentland, A. Vespignani, E. Moro, Y. Moreno, Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic, medRxiv (2020) 2020.05.06.20092841URL <https://www.medrxiv.org/content/10.1101/2020.05.06.20092841v1>.
320
- [24] R. M. May, R. M. Anderson, Population biology of infectious diseases: Part II, Nature 280 (5722) (1979) 455–461.

- [25] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, et al., Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees, *BMC medicine* 9 (1) (2011) 87.
- [26] W. Duan, Z. Fan, P. Zhang, G. Guo, X. Qiu, Mathematical and computational approaches to epidemic modeling: a comprehensive review, *Frontiers of Computer Science* 9 (5) (2015) 806–826.
- [27] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115 (772) (1927) 700–721.
- [28] M. J. Keeling, P. Rohani, *Modeling infectious diseases in humans and animals*, Princeton University Press, 2011.
- [29] C. Kamp, M. Moslonka-Lefebvre, S. Alizon, Epidemic spread on weighted networks, *PLoS computational biology* 9 (12).
- [30] B. Wilder, M. Charpignon, J. A. Killian, H.-C. Ou, A. Mate, S. Jabbari, A. Perrault, A. Desai, M. Tambe, M. S. Majumder, The role of age distribution and family structure on COVID-19 dynamics: A preliminary modeling assessment for hubei and lombardy, Available at SSRN 3564800 .
- [31] B. M. Ndiaye, L. Tendeng, D. Seck, Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting, *arXiv preprint arXiv:2004.01574* .
- [32] J.-Q. Liao, X.-B. Hu, M. Wang, M. S. Leeson, Epidemic modelling by ripple-spreading network and genetic algorithm, *Mathematical Problems in Engineering* 2013.
- [33] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, Application of the ARIMA model on the COVID-2019 epidemic dataset, *Data in Brief* 29 (2020) 105340, ISSN 2352-3409.

- [34] N. Pearce, F. Merletti, Complexity, simplicity, and epidemiology, 2006.
- [35] S. A. Levin, T. G. Hallam, L. J. Gross, Applied mathematical ecology, vol. 18, Springer Science & Business Media, 2012.
- [36] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai,
 355 G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu, et al., Estimates of the severity of coronavirus disease 2019: a model-based analysis, The Lancet Infectious Diseases .
- [37] D. Chicco, Ten quick tips for machine learning in computational biology, BioData mining 10 (1) (2017) 35.
- [38] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai,
 360 G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang,
 365 X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: a model-based analysis, The Lancet Infectious Diseases (2020) S1473309920302437ISSN 14733099, URL <https://linkinghub.elsevier.com/retrieve/pii/S1473309920302437>.
- [39] G. Pullano, E. Valdano, N. Scarpa, S. Rubrichi, V. Colizza, Population
 370 mobility reductions during COVID-19 epidemic in France under lockdown (2020) 22.
- [40] S. Pei, J. Shaman, Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US, preprint, Epidemiology, URL <http://medrxiv.org/lookup/doi/10.1101/2020.03.21.20040303>, 2020.