

# Ce que la Covid-19 nous apprend sur la modélisation des épidémies

Jean-François Mathiot<sup>(1)</sup> (jean-francois.mathiot@clermont.in2p3.fr), Laurent Gerbaud<sup>(2)</sup> et Vincent Breton<sup>(1)</sup>

(1) Laboratoire de Physique de Clermont (UMR6533 CNRS/IN2P3, Université Clermont Auvergne), 4 avenue Blaise Pascal, 63178 Aubière Cedex

(2) Institut Pascal (UMR6602 CNRS, Université Clermont Auvergne, CHU Clermont-Ferrand, SIGMA Clermont), 4 avenue Blaise Pascal, 63178 Aubière Cedex

**L'épidémie de Covid-19 se transmet par les voies respiratoires. Ce mode de transmission dépend essentiellement des relations sociales entre individus, par opposition aux épidémies dont l'élément pathogène est transmis à longue distance, comme par exemple par l'eau.**

**Après un bref résumé des différents types de modèles utilisés jusqu'à présent pour comprendre cette épidémie, nous présentons un nouveau modèle basé sur la théorie de la percolation.**

**Il permet d'intégrer de façon naturelle les trois caractéristiques des relations sociales, pertinentes pour la propagation de l'épidémie : leur densité, leur intensité et leur variété. Quelques résultats sur la modélisation de l'évolution de l'épidémie de Covid-19 en France depuis ses débuts sont présentés.**

La Covid-19 nous a appris à être particulièrement prudents sur les prédictions que l'on pouvait raisonnablement faire sur l'évolution d'une épidémie. Pourtant, sa modélisation mathématique est essentielle à plusieurs points de vue : en premier lieu, bien évidemment, pour comprendre la dynamique propre aux différentes phases de sa propagation. Cette compréhension est indispensable pour anticiper son évolution et pouvoir prendre, si ce n'est les bonnes décisions, tout au moins les moins mauvaises au moment adapté, en évitant de surréagir ou de réagir avec retard.

La propagation d'une épidémie dépend de ses caractéristiques, en particulier sa contagiosité ou la durée de la période pendant laquelle une personne contaminée peut en infecter une autre. Il convient aussi de tenir

compte des conditions physiques de cette propagation. De ce point de vue, nous pouvons identifier deux grandes classes d'épidémies :

- (I) celles qui se propagent de proche en proche par proximité physique lors d'interactions sociales, telles que les épidémies transmises par voies respiratoires ;
- (II) celles qui se propagent à grande échelle par l'intermédiaire d'un élément extérieur, à l'exemple des épidémies transmises par l'eau (choléra...) ou par un vecteur comme le moustique pour le paludisme.

L'étude de la Covid-19 permet de mettre en évidence les caractéristiques essentielles des épidémies de classe I sur plusieurs années. C'est un véritable défi pour tous les modèles utilisés, et une situation tout à fait nouvelle par rapport à la modélisation d'une épidémie

de grippe saisonnière. Cette dernière se caractérise en effet par une propagation simultanée sur tout le territoire, une couverture vaccinale initiale importante, une durée de quelques mois et une contagiosité relativement faible.

L'épidémie de Covid-19 a démarré quant à elle à partir d'un nombre extrêmement restreint de personnes infectées. Différentes publications documentent une circulation précoce du virus en France, plusieurs semaines avant les premiers cas répertoriés officiellement le 24 janvier 2020 [1, 2]. Cette propagation souterraine s'est faite au sein d'une population sans aucune couverture vaccinale et avec un virus bien plus virulent que celui de la grippe. Ces caractéristiques propres induisent des contraintes très fortes sur les modèles existants et ouvrent la voie pour explorer de nouvelles approches.

## Comment modéliser l'épidémie de Covid-19

De manière générale, nous pouvons identifier trois grandes familles de modèles pour décrire la propagation d'une épidémie :

- les modèles « compartimentaux », déployés au milieu des années 1920 pour décrire en particulier l'évolution démographique d'une population animale ;
- les modèles « multi-agents », élaborés dans le milieu des années 1970 avec l'apparition de l'intelligence artificielle distribuée ;
- enfin, les modèles de « percolation », développés en chimie au milieu des années 1950 pour décrire la diffusion d'un liquide dans un milieu poreux.

Ces différentes modélisations ne sont pas toutes adaptées aux caractéristiques fondamentales de l'épidémie de Covid-19.

### Modèles de type compartimentaux

Ils s'intéressent à l'évolution du nombre d'individus dans certaines catégories [3], comme par exemple les catégories Saine, Infectée ou Rétablie, d'où leur nom générique de « modèles SIR ». Par construction, ces modèles considèrent que n'importe quelle personne contaminée peut infecter une personne saine où qu'elle se trouve dans le territoire étudié, et donc sans qu'il y ait eu nécessairement un contact lors d'une interaction sociale. La propagation de l'épidémie dépend alors d'un paramètre global  $R$ , appelé taux de reproduction. Celui-ci représente le nombre moyen de personnes saines qu'une personne contaminée va infecter pendant toute sa période de contagiosité.

Les modèles compartimentaux sont adaptés à une épidémie se développant à grande échelle, de classe II dans la nomenclature rappelée en introduction, mais le sont peu pour comprendre l'évolution d'une épidémie de type Covid-19 sur le long terme, pour laquelle l'environnement local est essentiel.

Toute la dynamique de l'épidémie est définie dans ces modèles par le taux de reproduction  $R$ . Celui-ci dépend de facteurs de nature très différente : la nature intrinsèque du

virus, le comportement social des personnes, l'historique de l'épidémie ou encore le taux de vaccination de la population concernée. Il est ainsi impossible d'avoir une détermination indépendante de  $R$ , qui doit être estimé typiquement par la mesure des contaminations pendant une quinzaine de jours.

### Modèles de type multi-agents

Ces modèles consistent en une simulation numérique d'un territoire restreint, typiquement une ville ou un quartier, avec tous leurs habitants et leurs différentes interactions, presque heure par heure [4]. Ils nécessitent la connaissance détaillée de ces interactions, qu'elles soient dans le milieu professionnel, associatif, familial, à l'école ou entre amis, ou pour aller faire ses courses. Les échelles caractéristiques d'espace et de temps de ces modèles sont relativement petites ; elles sont entièrement tributaires des capacités de calcul disponibles.

Les modèles multi-agents sont adaptés pour répondre à des questions très précises sur un territoire limité, comme par exemple pour étudier l'importance des contaminations à l'école, dans une entreprise ou dans

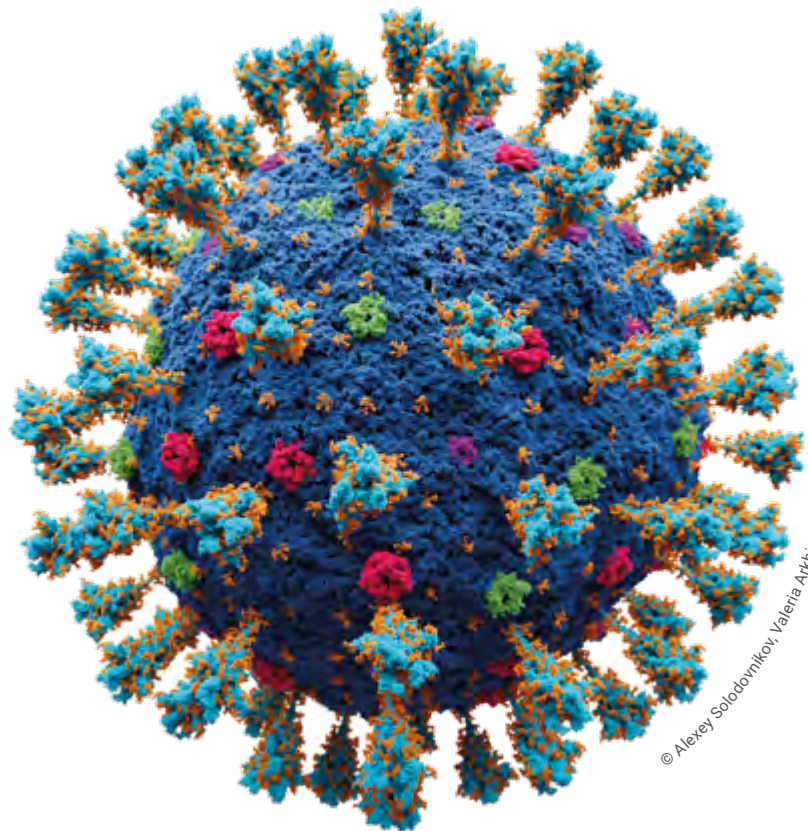
un Ehpad, sur une échelle de temps limitée. Il est donc plus difficile d'extraire une information générique sur un grand territoire, comme une région ou un pays, et de suivre une épidémie sur le long terme.

### Modèles de type percolation

Ces modèles s'intéressent à la propagation de proche en proche d'un événement donné dans un milieu constitué d'un très grand nombre d'éléments [5]. Cet événement peut être par exemple un liquide s'écoulant de cavité en cavité dans un milieu poreux, ou un feu se propageant d'arbre en arbre dans une forêt. Dans un réseau de communication ou de transport d'électricité à grande échelle, il correspond à la propagation d'une information ou d'un courant électrique, de relais en relais. Tout naturellement, ces modèles ont été appliqués à la transmission de proche en proche d'une épidémie dans une population donnée, par la formation d'agrégats ("clusters" en anglais).

Les modèles de percolation ont une propriété tout à fait intéressante dans le cas de la propagation d'une épidémie : l'identification du seuil au-delà duquel elle devient incontrôlable sauf

© Alexey Solodovnikov, Valeria Arkhipova (Wikimedia Commons).



Modèle atomique d'un virus de SARS-CoV-2. Chaque petite sphère représente un atome.

>>>

>>>

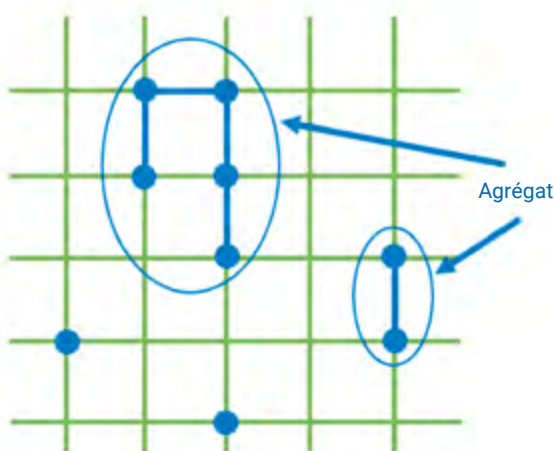
en cas de mesures radicales, confinement ou autres restrictions fortes des relations sociales, par exemple. Ils sont particulièrement flexibles, et sont pertinents dès lors que le nombre de personnes considérées est élevé, comme par exemple pour une ville de 100 000 habitants et plus. Ils sont donc tout à fait adaptés pour décrire une épidémie du type Covid-19 et nous ont inspirés pour construire le modèle PERCOVID [6], en complément des modèles compartimentaux souvent utilisés [7].

## Qu'est-ce que PERCOVID ?

Un modèle de percolation est défini sur un réseau, composé de sites reliés entre eux par des liens. Il est ainsi caractérisé par trois quantités fondamentales :

- la dimension  $D$  du réseau ( $D = 1, 2, 3, 4, \dots$ ), choisie selon la nature du problème étudié ;
- la densité  $p$  de sites dits occupés, comprise entre 0 et 1 ;
- la probabilité  $q$  d'établir un lien entre deux sites occupés voisins, comprise entre 0 et 1.

Un agrégat est constitué par l'ensemble des sites occupés, reliés entre eux par un lien (fig. 1). Selon la valeur des paramètres  $p$  et  $q$ , pour une dimension donnée du réseau, on pourra alors identifier deux configurations bien distinctes : l'une correspond à quelques agrégats de petite taille répartis uniformément sur tout le réseau, l'autre à un très grand agrégat composé typiquement de plus de la



1. Exemple simple de réseau à deux dimensions. Les sites occupés sont représentés par des points bleus, et les liens établis entre deux sites occupés par des traits bleus. Cette situation correspond à  $p = 9/25$  et  $q = 5/6$ .

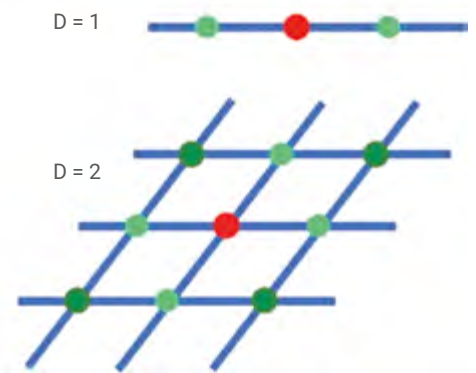
moitié des sites occupés, et entouré de tout petits agrégats. Dans ce dernier cas, on dira que le système se trouve dans la « zone de percolation ». Lorsqu'on augmente  $p$  et/ou  $q$ , la transition entre ces deux configurations est extrêmement brusque dès lors que le nombre de sites considérés est grand. On parle alors de « transition de percolation ».

Les épidémies par voies respiratoires sont entièrement déterminées par les relations sociales. Pour tenir compte de la proximité géographique naturelle au sein de chaque foyer, chaque site du réseau représente un foyer de 1, 2, 3 personnes, ou plus. Les relations sociales faisant basculer l'épidémie d'un foyer vers un autre sont représentées quant à elles par trois de leurs caractéristiques principales, pertinentes pour la propagation d'une épidémie : leur densité, leur intensité et leur variété plus ou moins grande, qui dépendent des pratiques sociologiques de chaque territoire, ainsi que de la réglementation en vigueur.

Ces trois caractéristiques sont directement associées aux trois paramètres fondamentaux d'un modèle de percolation :

- la densité des relations sociales est définie par le nombre moyen de contacts sociaux qu'une personne peut avoir dans une journée, rapporté au nombre maximal de voisins proches ou moins proches sur le réseau. Elle est représentée par la densité  $p$  ;
- l'intensité des relations sociales est définie par la plus ou moins grande facilité de contaminer son voisin dans le réseau social. Cette intensité est représentée par la probabilité  $q$  ;
- la construction d'un réseau à  $D$  dimensions permet de prendre en compte la variété de ces relations sociales.  $D$  pourra être pris égal à 1, 2, 3 ou 4 selon la plus ou moins grande complexité des relations sociales dans le territoire étudié, comme par exemple dans le domaine professionnel, familial ou associatif, ou pour les interactions de la vie courante.

La grande flexibilité des modèles de percolation permet de traiter de façon tout à fait naturelle certaines



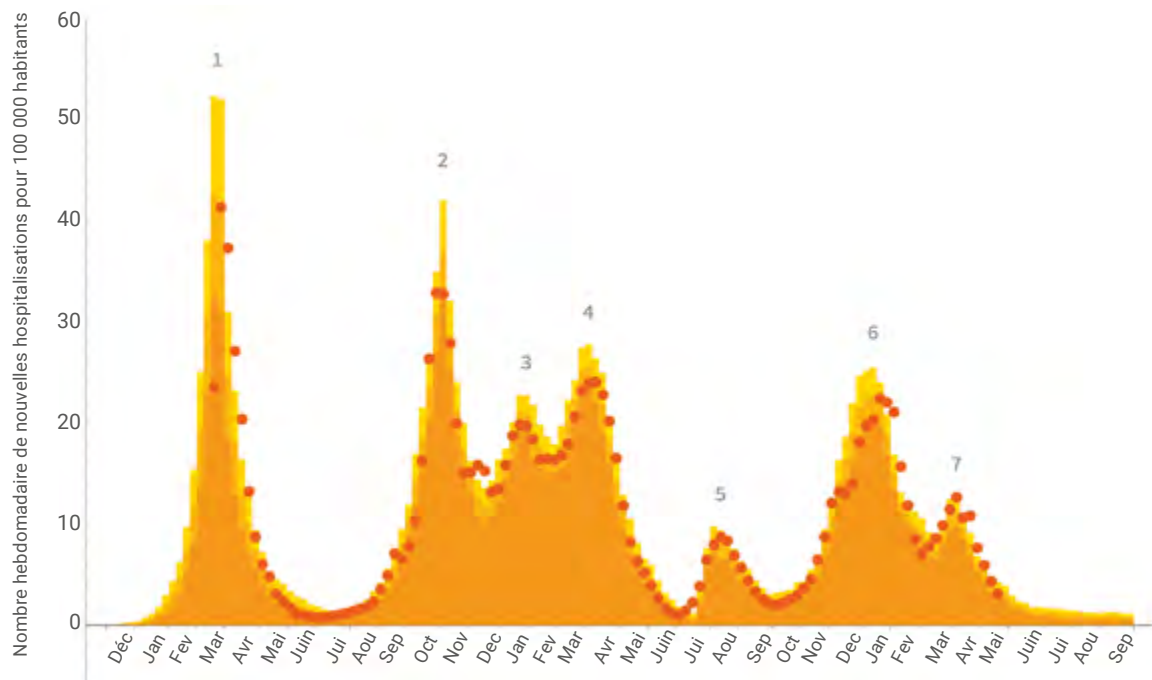
## 2. Réseaux à une ( $D = 1$ ) et deux ( $D = 2$ ) dimensions.

Les voisins proches, ou relations essentielles, d'une personne infectée (en rouge) sont représentés en vert clair, alors que les voisins moins proches, ou relations moins essentielles, sont représentés en vert foncé. À une dimension, il n'existe que des relations essentielles.

spécificités mises en valeur par l'épidémie de Covid-19. La propagation différenciée suivant les relations essentielles ou moins essentielles que chaque individu peut avoir dans son environnement social correspondra par exemple à la propagation différenciée vers les voisins les plus proches ou un peu moins proches (fig. 2).

Chaque individu d'un foyer est dans un état différent : sain, infecté mais non contagieux, contagieux, guéri, hospitalisé, décédé, vacciné avec ou sans rappel. Lorsqu'il est infecté, il pourra l'être avec la souche initiale du virus ou avec chacun de ses variants et être soit symptomatique soit asymptomatique. Tous ces états correspondent à des comportements sociaux différents, et donc à une valeur de la probabilité  $q$  différente. Le taux de vaccination peut être modulé chaque jour, avec une efficacité du vaccin variant en fonction du temps.

Dans le modèle PERCOVID, la propagation de l'épidémie se fait pas à pas, où un pas de temps numérique correspond à une journée. À chaque pas, toute personne contagieuse du réseau va pouvoir infecter un proche voisin dans son premier cercle de relations essentielles, un voisin moins proche dans son second cercle de relations moins essentielles, ou encore une personne de son propre foyer. Elle va le faire avec une probabilité  $q \times r$ , où  $r$  est la contagiosité intrinsèque du virus. Ce paramètre  $r$  est caractéristique de chaque virus et variant, et permet de moduler la propagation du virus selon sa contagiosité. Lorsqu'une



**3. L'épidémie de Covid-19 en France, à partir de décembre 2019.** Le nombre hebdomadaire de nouvelles hospitalisations, pour 100 000 habitants, est indiqué en points rouges (données de Santé Publique France [9]), avec la modélisation en orange. Les différentes teintes d'orange correspondent à l'incertitude statistique associée, pour un calcul moyenné sur une dizaine de simulations. Chaque pic est identifié par un nombre.

personne est infectée, une horloge interne démarre. Cette horloge permet de suivre pas à pas l'évolution de l'infection.

### Quelques résultats

La simulation présentée ici est effectuée sur un réseau à trois dimensions comportant environ trois millions de foyers (un dixième de la population française). Elle est initialisée au 1<sup>er</sup> décembre 2019, avec une quinzaine de personnes infectées. Cela correspond au nombre minimal de personnes infectées pour que l'épidémie puisse se propager sur le réseau. Chaque simulation dure environ trente minutes sur un ordinateur portable.

Les paramètres de la modélisation sont de trois types : épidémiologiques, ou liés aux comportements sociaux des populations concernées, ou liés à la mobilité des individus en dehors du premier ou du second cercle de leurs relations sociales. Les paramètres épidémiologiques ne sont pas toujours connus avec une grande précision, en particulier au démarrage de l'épidémie comme ce fut le cas pour la Covid-19. La modélisation de l'épidémie consiste alors à étudier plusieurs scénarios possibles suivant la valeur de ces paramètres, en les affinant au fur et à

mesure de la publication de nouvelles études.

Certains paramètres liés aux comportements sociaux peuvent être déduits d'études sociologiques, comme pour la densité des relations sociales en France [8]. L'estimation de l'intensité de ces relations sociales, et de sa variation en fonction du temps, est plus subjective, et ne peut être faite que par ajustement de la simulation avec des données expérimentales. Le suivi de l'épidémie de Covid-19 est de ce point de vue particulièrement instructif, avec pratiquement toutes les situations possibles de relations sociales : sans aucune restriction au tout début de l'épidémie avec un comportement social typique de chaque population considérée, restrictions plus ou moins fortes pendant les périodes de confinement, restriction de l'accès aux relations moins essentielles et restrictions dues aux couvre-feux, influence des gestes barrières, relâchement des restrictions et des comportements lors des vacances d'été.

La mobilité des personnes en dehors de leurs relations essentielles ou moins essentielles a été très peu étudiée. Cette mobilité joue un rôle important dans les phases de démarrage de l'épidémie ou d'apparition de

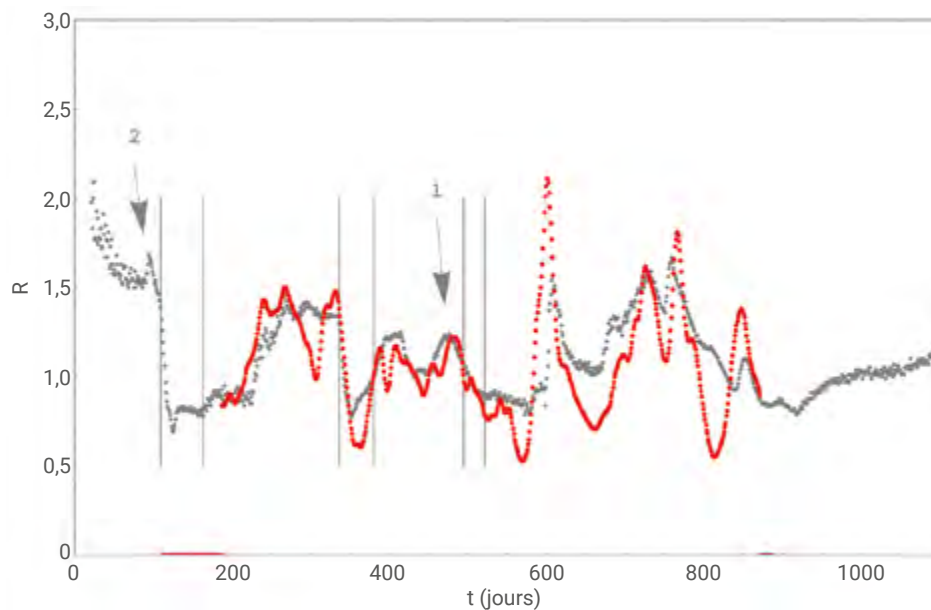
nouveaux variants, car elle permet une dissémination du virus dans des territoires nouveaux. Une fois l'épidémie implantée, cette mobilité n'a pas d'effet dominant dans notre simulation.

En l'absence d'une stratégie de tests aléatoires systématiques sur tout le territoire, la façon la plus objective de suivre la propagation de l'épidémie est de considérer le nombre hebdomadaire de nouvelles hospitalisations. Cette donnée est indiquée sur la figure 3 en points rouges, avec la modélisation en orange. Elle est caractérisée par des périodes de brusque intensification pendant lesquelles le nombre de cas de Covid-19 évolue rapidement vers un pic épidémique, et se traduit par une évolution en forme de vagues successives. Chacun des pics observés peut être identifié dans notre simulation à un événement particulier.

- Le pic (1), en mars 2020, correspond à un développement exponentiel de l'épidémie dans une population non vaccinée, en l'absence de restrictions fortes et avec un virus relativement contagieux. Avec un confinement très strict, la décade du nombre d'hospitalisations est extrêmement brutale.
- Le pic (2), en octobre 2020, correspond à la reprise épidémique de l'automne

>>>





**4. Taux de reproduction  $R$  en fonction du temps.** Les données de Santé Publique France [9] sont indiquées en rouge, et les prédictions du modèle PERCOVID en gris. Le temps  $t = 0$  correspond au 1<sup>er</sup> décembre 2019. Les paires de traits verticaux encadrent les trois périodes de confinement en France.

>>>

suite à un relâchement des pratiques sociales et des gestes barrières, et après une dissémination de l'épidémie sur tout le territoire pendant les vacances d'été. La décline du nombre d'hospitalisations est ici aussi brutale, suite à la deuxième période de confinement.

- Le pic (3), en janvier 2021, correspond juste à un relâchement des comportements sociaux après le deuxième confinement et pendant et après les fêtes de fin d'année, sans explosion de l'épidémie.
- Le pic (4), en mars/avril 2021, est entièrement dû à l'apparition du variant *alpha*, bien plus contagieux que la souche initiale du virus. Il correspond à un pic classique de propagation d'une épidémie.
- De la même façon, le pic (5) de l'été 2021 est dû à la propagation du variant *delta*. La contagiosité de ce variant est encore bien plus forte que celle du variant *alpha*. Cela lui permet de se propager de façon significative, même en été, sans cependant beaucoup d'intensité compte tenu de l'immunité naturelle acquise par les personnes guéries et du taux de vaccination croissant de la population française.
- Le pic (6) en janvier 2022 est la conséquence de deux effets bien différents : tout d'abord, en novembre/décembre, une remontée des infections par le variant *delta* due à la

perte progressive de l'immunité vaccinale et naturelle, renforcée à partir de décembre 2021 par la propagation du variant *omicron* bien plus contagieux. Cette très forte virulence induit un pic de contamination élevé, puis une décroissance très rapide en janvier 2022 une fois ce pic franchi, grâce à l'immunité naturelle acquise par les personnes guéries dans l'environnement social de chaque individu.

- Le pic (7), observé à partir de la fin février 2022, trouve son origine dans la propagation de la mutation BA.2 du variant *omicron*, avec une contagiosité encore plus forte. Cette propagation est favorisée par le relâchement des comportements sociaux au retour des vacances scolaires.

La propagation de l'épidémie pendant l'été 2022 devrait rester faible mais continue, comme le montre la valeur du taux de reproduction  $R$  proche de 1 (fig. 4). Dans notre simulation, cela est dû à la perte progressive de l'immunité. Cette propagation pourra devenir importante seulement en présence d'un nouveau variant échappant de façon significative à la protection vaccinale ou naturelle.

L'étude du nombre hebdomadaire de nouvelles hospitalisations ne permet cependant pas d'anticiper sur la dynamique propre de l'épidémie : l'hospitalisation éventuelle d'une

personne infectée a lieu environ une quinzaine de jours après son infection. Cette dynamique est entièrement déterminée par le taux de reproduction  $R$ . Grâce à l'horloge interne déclenchée dans notre modélisation lorsqu'une personne est infectée, nous pouvons calculer quotidiennement la valeur exacte de ce paramètre et donc anticiper sur l'évolution de l'épidémie. Ce taux est ici une prédiction, alors qu'il est un paramètre nécessaire pour initialiser les modèles de type compartimentaux.

La figure 4 présente une comparaison entre la valeur de  $R$  prédite par le modèle PERCOVID et celle estimée par Santé Publique France à partir des données de taux d'infection journalier [9]. Typiquement, un pic dans le comportement de  $R$  apparaît environ une quinzaine de jours avant un pic des contaminations. Le pic noté par une flèche (1) sur cette figure est ainsi antérieur d'une quinzaine de jours sur le pic des contaminations noté (4) sur la figure 3, correspondant à la troisième période de confinement. Celle-ci n'a eu ainsi que peu d'influence sur la valeur de  $R$  car la dynamique de l'épidémie était déjà dans une phase décroissante, contrairement aux deux premières périodes de confinement qui ont induit une chute brutale de cette valeur. Le pic (2) correspond, quant à lui, au franchissement des seuils épidémiques dans le Haut-Rhin [2] au tout début de l'épidémie.

## Conclusion

Le modèle PERCOVID, par sa capacité à intégrer les caractéristiques pertinentes des relations sociales dans la propagation des épidémies des voies respiratoires, est un outil original pour l'analyse des maladies émergentes. Il est un bel exemple d'étude pluridisciplinaire à l'interface entre épidémiologie, mathématique, physique et sociologie.

Cette étude représente une nouvelle manière de concevoir les caractéristiques d'une épidémie des voies respiratoires, en séparant clairement le comportement intrinsèque du virus, sa contagiosité en particulier, des effets dus aux comportements sociaux des populations au sein desquelles il se propage. Cela peut se faire par exemple en comparant la dynamique

de plusieurs épidémies — comme la grippe, la Covid-19, Ebola — dans un même territoire ou une même épidémie dans des territoires aux pratiques sociologiques différentes — comme différents pays en Europe, Afrique ou en Asie. Le modèle PERCOVID permet alors d'accéder aux propriétés du virus, indépendamment du contexte social dans lequel ce dernier se propage. ■



- 1• F. Carrat *et al.*, "Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based "CONSTANCES" cohort", *Eur. J. Epidemiol.* **36** (2021), 219-222 ; <https://doi.org/10.1007/s10654-020-00716-2>.
- 2• L. Gerbaud *et al.*, *Int. J. Environ. Res. Public Health*, **17** (2020) 7175 ; <https://doi.org/10.3390/ijerph17197175>.
- 3• Modélisation d'une épidémie : <https://cutt.ly/Modelisation-d-une-epidemie>
- 4• B. Gaudou *et al.*, "COMOKIT: a modeling kit to understand, analyze and compare the impacts of mitigation policies against the COVID-19 epidemic at the scale of a city", *Front. Public Health* **8** (2020) 563247 ; <https://cutt.ly/doi-10-3389> ; <https://comokit.org>.
- 5• Modèles de percolation : <https://cutt.ly/La-percolation-jeu-de-pavages-aleatoires>
- 6• J.-F. Mathiot, L. Gerbaud et V. Breton, "Highlighting the impact of social relationships in the propagation of respiratory viruses using percolation theory", *Scientific Reports* **11** (2021) 24326 ; [www.nature.com/articles/s41598-021-03812-9](http://www.nature.com/articles/s41598-021-03812-9).
- 7• H. Salje *et al.*, "Estimating the burden of SARS-CoV-2 in France", *Science* **369** (2020) 208-211 ; <https://doi.org/10.1126/science.abc3517>
- 8• G. Béraud *et al.*, "The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases" *PLoS One* **10(7)** e0133203 ; <https://doi.org/10.1371/journal.pone.0133203> (2015).
- 9• Santé Publique France : <https://geodes.santepubliquefrance.fr>.