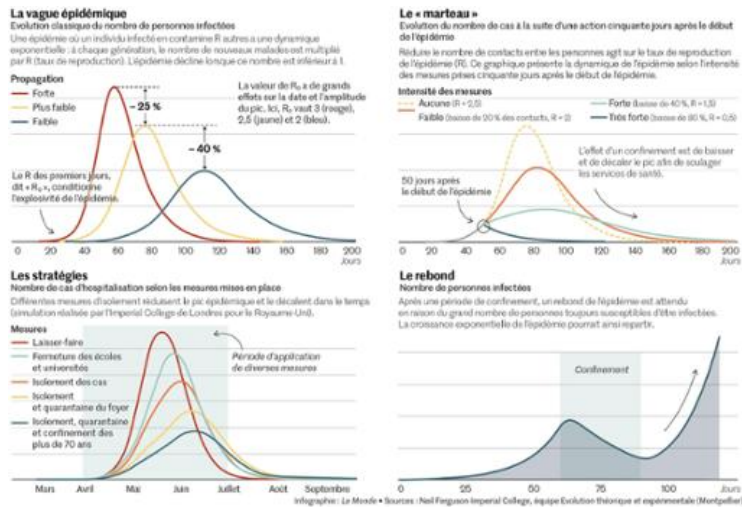


## Modéliser les épidémies, un art difficile



### David Larousserie

Prédire l'évolution des épidémies est délicat, mais si « tous les modèles sont faux, certains sont utiles », disent les statisticiens. Le point sur les outils dont ils disposent pour anticiper la suite de la pandémie due au nouveau coronavirus

**C**ombien de morts du Covid-19 ? Combien d'hospitalisés en réanimation ? Combien de temps le confinement durera-t-il ? Autant de questions brûlantes auxquelles s'attaque la science de la modélisation... avec difficulté.

Le 12 mars, face aux prédictions d'une submersion du système hospitalier avancées par une équipe anglaise de l'Imperial College (Londres), la France s'est engagée dans une stratégie de confinement poussée. « *Les modèles sont un signal parmi d'autres. La situation italienne, mais aussi toutes les connaissances accumulées sur des épidémies précédentes allaient dans le sens de cette décision pour soulager le système de santé* », relativise Simon Cauchemez, chercheur à l'Institut Pasteur, spécialiste en modélisation et membre du conseil scientifique Covid-19 mis en place par le gouvernement.

Il n'y avait en effet pas besoin du modèle anglais pour être inquiet. Un « vieux » modèle des années 1930 et les premières données chinoises sur le nouveau coronavirus fournissent assez vite aux spécialistes des estimations d'au moins 50 % de personnes infectées au pic de l'épidémie, et donc plus de 300 000 morts en France avec une hypothèse de seulement 1 % de létalité due au virus.

### Deux approches

Les scientifiques disposent en fait de deux grandes méthodes pour appréhender la dynamique d'une épidémie, soit une approche dite « compartimentale », soit une approche dite « multi-agent » ou « individu-centrée ». La première remonte aux années 1930 et la seconde au début des années 2000.

En 1927, 1932 et 1933, deux Écossais, Anderson Gray McKendrick et William Ogilvy Kermack, publient trois études dans les comptes rendus de la Royal Society. Ces articles sont si marquants qu'ils furent republiés en 1991 dans *Bulletin of Mathematical Biology*. Le modèle qu'ils y développaient collait très bien aux données d'un épisode de peste, en Inde, en 1906. L'idée est de mettre les individus en trois compartiments : les « sains », S, susceptibles d'être contaminés, les « infectés », I, et les « retirés », R, ceux qui ne sont plus malades, car soit morts, soit immunisés après une infection.

Des taux de passage décrivent les flux entre ces trois catégories, comme le taux de contamination (passage de S à I) et le taux de guérison (ou durée de l'infection, passage de I à S), ce qui fournit des équations faciles à résoudre. Plus tard, ces modèles seront baptisés « SIR ». En conclusion de leur premier article, les deux mathématiciens notaient, message rassurant, qu'« *une épidémie, en général, prend fin avant que la population sensible n'ait été éradiquée* ». Autrement dit, les lois mathématiques protègent une partie de la population : tout le monde n'a pas besoin d'être touché pour qu'on observe un reflux.

La force de cette approche est qu'un seul paramètre, que les Ecossais n'avaient pas formellement baptisé, donne une intuition simple de la dynamique épidémique et de ce qui peut se passer. Il s'agit du taux de reproduction, c'est-à-dire le nombre de personnes qu'un infecté peut contaminer. S'il vaut 2, cela signifie qu'un malade contaminera deux personnes, qui ensuite, au total, en contamineront  $2 \times 2 = 4$ , puis  $2 \times 2 \times 2 = 8$ ... Le moteur infernal est enclenché.

En épidémiologie, cette constante macabre, baptisée «  $R_0$  » au début de la flambée, est parmi la première que les statisticiens essaient d'évaluer. Sa valeur exacte joue beaucoup sur les prévisions. Passer de  $R_0 = 3$  à  $R_0 = 2,5$  décale le pic épidémique de trente jours et le fait baisser de presque 25 %. « *Estimer  $R_0$  est simple en théorie. On compte le nombre de cas secondaires, et on divise par le nombre de cas primaires, rappelle Pierre-Yves Boëlle, de l'Institut Pierre-Louis d'épidémiologie et de santé publique. Mais c'est compliqué en pratique !* » En effet, cette méthode suppose d'identifier les cas, de retracer leur historique et de repérer les contaminés secondaires. En pratique, on étudie la courbe épidémique, le nombre de cas ou de morts en fonction du temps, car sa croissance mathématique, exponentielle, est reliée au  $R_0$ . Mais tout dépend de la qualité des données : tous les cas sont-ils détectés ? Les morts attribuables à la maladie sont-ils tous comptés ?

Résultat, par exemple, pour la France, selon Samuel Alizon, chercheur CNRS à Montpellier en évolution des maladies infectieuses, qui utilise des outils développés par le groupe de M. Boëlle pour calculer le  $R_0$ , ce dernier varie entre 2,39 et 2,58 dans le cas du Covid-19.

## Comprendre l'intérêt du confinement

Ce paramètre permet aussi de comprendre simplement l'intérêt d'un confinement. Il dépend en effet notamment du nombre de contacts entre les gens et de la probabilité de contamination. Réduire les contacts diminue le taux de reproduction, et donc la flambée. Au pic, le nombre de nouveaux cas chute, car le taux de reproduction devient inférieur à 1 : une personne contamine moins d'une nouvelle personne. « *Le modèle SIR a beau être simpliste, dans beaucoup de cas, il marche. Surtout, il est souvent implicite dans nos raisonnements, à cause de l'importance du  $R_0$*  », explique M. Alizon.

Mais le vieux rafiot écossais a quand même été rafistolé. Il est ainsi possible d'ajouter des compartiments. On peut avoir une population « exposée », qui incube le virus avant d'être « infectée ». On peut ajouter, notamment dans le cas de l'épidémie due au SARS-CoV-2, des étiquettes « asymptomatiques » pour les gens qui contaminent avec plus ou moins de probabilité leurs semblables, probabilité qui reste encore un mystère pour le nouveau coronavirus. On peut aussi, notamment pour anticiper la surcharge hospitalière, créer des compartiments « malades sévères » ou « malades non sévères »...

Récemment, plusieurs équipes ont ainsi proposé leurs SIR améliorés. A l'université de Bâle, l'équipe de Richard Neher utilise sept compartiments, pour simuler l'évolution de la maladie. Il est même possible, au cours du temps, de changer le taux de reproduction, pour réduire la transmission et « voir » l'effet d'un équivalent d'une situation confinée. La technique étant relativement simple à mettre en œuvre, beaucoup de simulateurs sont proposés en ligne, comme celui de l'équipe suisse ou comme ceux d'« amateurs », tel celui de Gabriel Goh, spécialiste d'apprentissage machine, qui, avec onze « curseurs », permet de se faire une idée de la propagation d'une maladie.

### « Matrices de contact »

Malgré ces raffinements, cela ne suffit pas. Anderson McKendrick et William Kermack avaient prévenu : ces modèles ne marchent que pour une population homogène, où tout le monde infecte tout le monde de la même manière. Or, les populations réelles sont hétérogènes, avec des densités différentes sur le territoire et, surtout, n'ont pas le même âge. Un enfant n'a pas la même susceptibilité à une maladie qu'un adulte ou qu'une personne âgée. Il n'a pas non plus le même nombre de contacts quotidiens que ses parents ou grands-parents.

D'où l'idée d'améliorer la description en ajoutant des « compartiments » en classe d'âge, par tranches de cinq ans. Chacun des paramètres précédents aura donc une valeur différente pour chaque tranche. Fort heureusement, de scrupuleux statisticiens ont estimé ces données de contact pour plusieurs pays. En France, par exemple, un enfant de moins de 5 ans a trois fois plus d'interactions quotidiennes avec des enfants de son âge qu'avec des personnes de 30-34 ans. Les adolescents de 15-19 ans sont les plus « tactiles » entre eux, et les Français ont presque deux fois plus de contacts que les Allemands à l'intérieur de cette tranche. « *Ces matrices de contact changent la dynamique d'une épidémie* », insiste Pascal Crépey, de l'Ecole des hautes études en santé publique, à Rennes. Son modèle, en ligne le 20 mars, tient compte de cet effet et prédit l'engorgement hospitalier actuel en fonction de différents  $R_0$ . Dans l'hypothèse où  $R_0 = 3$ , toutes les régions, d'ici au 15 avril, seraient débordées. La moitié résisterait si  $R_0 = 2,25$ . Et cela sous l'hypothèse, désormais fictive, qu'aucune mesure pour réduire les transmissions n'est prise.

Sa consœur Vittoria Colizza, à l'Inserm, a aussi étudié l'effet de la fermeture des écoles sur la propagation. Car, même si une telle mesure n'avait jamais été prise, il existe des données de « contacts » en période scolaire et en période de vacances (proche d'une situation de fermeture). Le modèle à compartiments, mis en ligne le 14 mars, estimait ainsi une réduction de 10 % du pic épidémique en fermant les écoles.

Même philosophie, publiée dans *TheLancet* le 25 mars, par une équipe de la London School of Economics, à propos de la ville chinoise de Wuhan, épicerie de l'épidémie. Ce modèle compartimenté prédit qu'une levée

des mesures de confinement en avril baisserait de plus de 90 % les cas quotidiens par rapport à une levée en mars.

A force d'aller dans les détails, les chercheurs ont voulu pousser jusqu'au bout et zoomer jusqu'à l'individu lui-même. Pourquoi diviser la population en dix-sept tranches quand les ordinateurs peuvent facilement traiter les situations de chaque cas individuel, fussent-ils des millions ? Cette approche, dite « multi-agent » ou « individu-centrée », a été développée depuis une quinzaine d'années et constitue la seconde grande famille de modèles. Comme pour les modèles SIR, les principes sont assez simples. A chaque individu du modèle sont attachés un âge, un lieu d'habitation, des enfants, un foyer, un lieu de travail... Il peut bouger dans l'espace, contraint par les données connues de mobilité, de densité... Arrive une maladie avec ses propriétés propres de temps d'incubation, d'infection, de contagiosité. Ces deux types d'informations sont mélangés, et la simulation est lancée. Elle est forcément aléatoire (on met une personne infectée au hasard, les gens bougent de façon erratique...), et elle est reproduite plusieurs centaines de fois, afin d'en tirer des enseignements moyens sur le nombre de nouveaux cas, d'hospitalisation...

A l'Imperial College, l'équipe de Neil Ferguson est experte en la matière depuis sa simulation, présentée en 2005 dans *Nature*, de la grippe H5N1, en Asie du Sud-Est, par ces méthodes. Ces chercheurs l'ont également appliquée au Covid-19 pour mesurer l'effet de différentes restrictions de circulation et inspirer les gouvernements. D'autres l'ont fait à l'échelle de Singapour dans le même but. Comme ils l'expliquaient dans *TheLancet*, le 23 mars, la stratégie la plus efficace, pour réduire de 99 % le nombre d'infections, est la panoplie complète : fermeture des écoles, quarantaine des malades et éloignement du travail.

« Ces simulations sont plus réalistes, mais elles contiennent plus de paramètres et sont difficiles à interpréter, constate M. Cauchemez, qui a travaillé avec M. Ferguson et jongle avec les deux approches. *Les approches multi-agents n'ont pas été utilisées en France, car elles n'étaient pas nécessaires pour répondre aux questions posées.* » M. Ferguson lui-même, dans le dernier rapport de son équipe, datant du 26 mars, reconnaît que le modèle SIR, qu'il a aussi utilisé, donne les mêmes résultats que ses simulations multi-agents, plus complexes.

En fait, à chaque problème, son modèle. « *La règle est qu'on cherche le modèle le plus simple pour répondre à la question posée* », résume Laura Temime, du Conservatoire national des arts et métiers. Mais il faut faire vite, car les questions changent avec l'évolution de l'épidémie. Au début, lorsque le virus n'était qu'en Chine, il s'agissait de savoir quelle était la probabilité qu'il débarque en Europe. La réponse est venue de M<sup>me</sup> Colizza qui a travaillé sans SIR ou multi-agent, « juste » en analysant les flux aériens entre pays et à l'intérieur de la Chine. Puis les statisticiens se sont abîmés les yeux sur les cas chinois pour estimer le fameux  $R_0$ , tandis que leurs collègues, sur le terrain, traquaient les patients atteints pour remonter la chaîne de contamination et identifier deux autres paramètres-clés, le taux de mortalité et le temps de génération (grosso modo le temps entre l'apparition des symptômes chez le premier et le second contaminé).

Ensuite, les modèles à compartiments chauffent pour anticiper la dynamique et les saturations des services de santé. Et très vite, les mêmes ou la version multi-agent tournent pour estimer l'effet de diverses mesures (isolement des cas, quarantaines, fermeture des écoles, arrêt de travail, confinement strict...).

## Barre d'erreurs

« *Et maintenant, on veut savoir quoi faire après le confinement en étudiant diverses options*, note M. Cauchemez, qui pense que le conseil scientifique devrait faire des recommandations cette semaine. *Mais ce confinement est une situation inédite, et nous manquons forcément de données pour alimenter les modèles.* » Les opérateurs téléphoniques pourraient être un recours. Orange a annoncé, le 27 mars, avoir passé un accord avec l'Inserm et l'équipe de M<sup>me</sup> Colizza pour utiliser les données téléphoniques, afin d'évaluer la mobilité en temps normal et en temps confiné.

« *Le problème, ce n'est pas les modèles, mais leurs hypothèses et les données qui les nourrissent* », indique M. Crépey, en écho au célèbre adage des statisticiens. « *Tous les modèles sont faux, mais quelques-uns sont utiles* », ajoute-t-il. Le Covid-19 n'échappe pas à la règle. Quel est le taux de mortalité ? Quelle est la part des asymptomatiques contagieux et avec quelle efficacité transmettent-ils le virus ? Le virus pourrait-il muter ?

Pour ces raisons, chaque simulation est assortie non seulement de la traditionnelle barre d'erreurs, mais aussi d'une analyse de sensibilité des résultats aux fluctuations des divers paramètres utilisés. Quelques dixièmes d'incertitude peuvent changer de plusieurs milliers un nombre de cas. « *Les simulations sont à manier avec précaution en termes de communication* », résume M. Boëlle. En outre, les prédictions initiales, où « on ne fait rien », sont souvent démenties avec le temps, car les gens changent leurs pratiques devant le risque.

Cependant, les spécialistes se risquent à une prévision. La fin d'un confinement causera un rebond de la courbe épidémique, car il n'y aura pas eu assez de « contaminés », donc « d'immunisés ». Comme le bon vieux modèle de 1927 le laisse entendre...