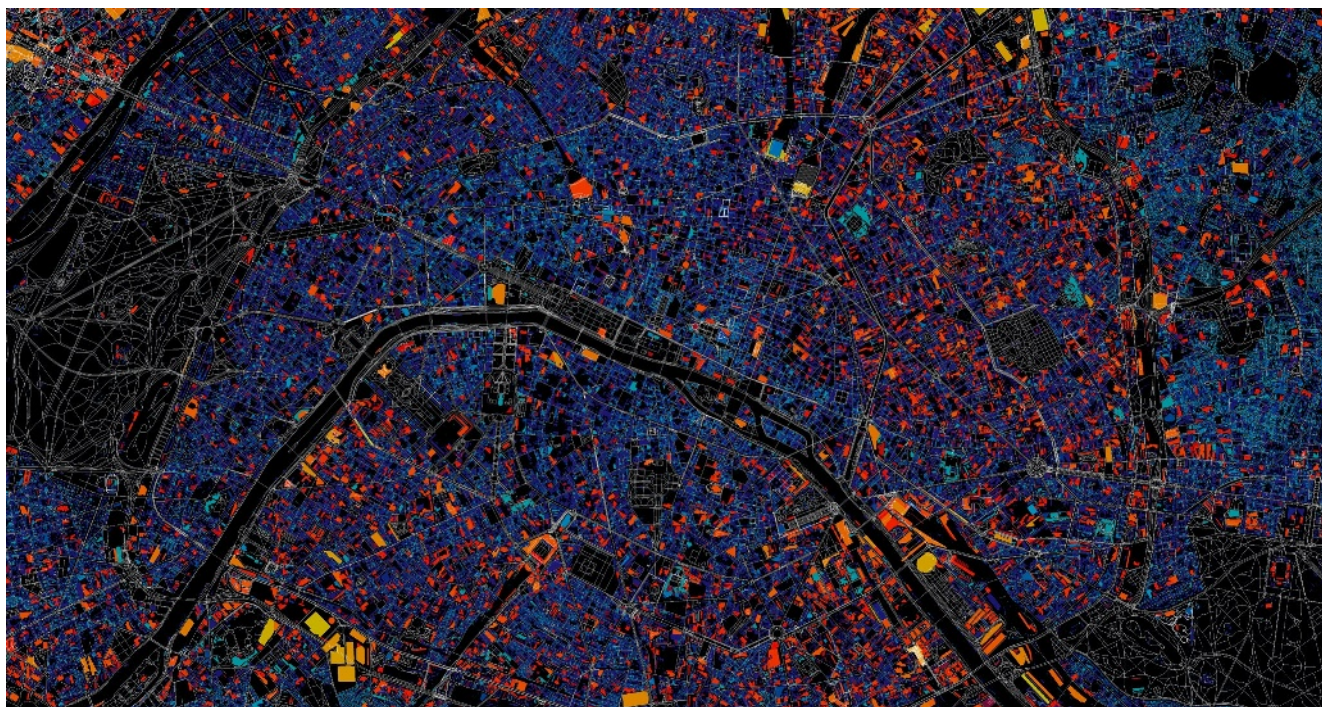




informations liées à la mobilité aide-t-il à anticiper le nombre d'hospitalisations dans une ou dans deux semaines? L'intérêt est que l'initiative Data for Good de Facebook publie des données de colocalisation réelle entre deux personnes (par souci d'anonymat, il s'agit de personnes passant cinq minutes dans la même zone de 600 mètres carrés), alors que les opérateurs de téléphonie ne fournissent que la position de chaque utilisateur. Nous avons donc utilisé ces données afin d'estimer les liens entre tous les départements français. Plus des personnes de deux départements « interagissent », plus les départements sont liés.

Évidemment, l'approche est biaisée: seules sont comptabilisées les personnes qui ont laissé leur téléphone allumé avec Facebook et ont autorisé la collecte de données. On a essayé de s'en affranchir en vérifiant que l'on suivait à peu près le même nombre de personnes chaque mois. Cette approche nous a permis d'affiner considérablement les scénarios d'admission en réanimation aux échelles tant nationale que départementale, sur plus de deux semaines. On voit aussi que les données de colocalisation semblent plus informatives que celles de mobilité fournies par Google, mais aussi que la température ou même les dépistages.

D'autres équipes françaises explorent ces nouvelles pistes. Par exemple, celle de Vittoria Colizza, à l'institut Pierre-Louis, à Paris, a un accès privilégié aux données de téléphonie mobile d'Orange. Quant à celle de Simon Cauchemez, à l'institut Pasteur, à Paris, elle inclut les variations de température et d'humidité pour anticiper le nombre d'hospitalisations journalières. L'espoir est qu'en utilisant

Grâce au projet collaboratif de cartographie OpenStreetMap, Olivier Thomine, à Aix-Marseille Université, a construit un modèle qui permet de simuler la transmission du SARS-CoV-2 sur le territoire français à l'échelle des bâtiments. Il répertorie d'abord sur le cadastre les bâtiments habités et leur affluence (ici sur Paris, plus les couleurs sont chaudes plus il y a d'affluence), les écoles et les hôpitaux étant traités à part (en cyan). Les simulations à Paris sont initialisées avec 15 personnes infectées (la simulation pour Paris l'année dernière est disponible ici: <https://bit.ly/3F6HL4I>).

plus de données, on puisse améliorer la qualité des scénarios prospectifs à court terme.

Les progrès informatiques permettent-ils une autre sorte de modélisation ?

À côté des modélisations statistiques et mathématiques, on utilise en effet parfois une autre modélisation plus informatique où l'on crée des simulations avec des agents, un ensemble d'entités (ici des humains) interagissant selon certaines règles, dans un certain environnement. Ces simulations sont presque aussi anciennes que les premiers ordinateurs de bureau, mais elles ont pris une tout autre ampleur avec les progrès de l'informatique, comme récemment le calcul à haute performance. Par exemple, Olivier Thomine, actuellement au Laboratoire d'informatique et systèmes, à Aix-Marseille Université, a développé un modèle de résolution inégalée à l'échelle mondiale: il commence par extraire toutes les données d'OpenStreetMap, un projet collaboratif de cartographie qui a l'avantage, outre d'être libre, de contenir tout le cadastre. Puis il répartit 66 millions d'individus, avec chacun un bâtiment de résidence. À partir de là, il fait des simulations où, chaque jour, un individu visite deux bâtiments autres que le sien. Et s'il y rencontre des personnes infectées, il y a transmission avec une certaine probabilité. La simulation se poursuit sur une année et apporte un réalisme unique au niveau géographique pour un modèle national, puisque l'on travaille à l'échelle des bâtiments.

Attention, comme dans tout modèle, certaines hypothèses sont très simplificatrices. Mais cela nous permet d'explorer la dimension

spatiale des épidémies. On observe par exemple que la proportion de la population infectée dans une commune à la fin d'une vague épidémique tend à dépendre fortement de la densité de population. Plus généralement, ce type de simulation informatique peut être utilisé pour tester des résultats issus de modèles plus simples, comme la théorie des réseaux.

Comment la science des réseaux intervient-elle dans les modélisations ?

On fait l'hypothèse que la propagation d'une épidémie se fait sur un réseau sous-jacent (voir l'article pages 28 à 36). Cela peut correspondre à des contacts entre personnes (qui a serré la main de qui ?) mais aussi entre des villes (combien de personnes transitent chaque jour ?). C'est en quelque sorte un formalisme, une hypothèse supplémentaire du modèle comme lorsqu'on distingue les individus selon des compartiments décrivant leur état clinique – susceptibles, infectés, en réanimation, décédés... L'avantage de cette hypothèse forte de réseau est qu'elle donne accès à des outils et théorèmes mathématiques permettant d'obtenir des résultats plus généraux.

Dans un modèle sans structure de contact explicite, vous supposez que chaque individu peut être en contact avec n'importe quel autre individu. Si cela peut être le cas à l'échelle d'un quartier, on voit bien qu'à une échelle nationale c'est peu crédible. Avec une approche de réseau, vous pouvez faire en sorte que chaque personne soit plutôt en contact avec une dizaine de personnes par semaine en moyenne. Cela génère un réseau social sur lequel vous pouvez simuler votre épidémie et voir en quoi cette structure de contacts modifie les propriétés de l'épidémie.

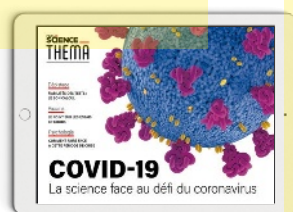
Cela semble très séduisant sur le papier, mais il existe plusieurs obstacles. D'une part, il faut déterminer quels réseaux sont importants, ce qui dépend de la biologie. Dans le cas d'une infection respiratoire comme le Covid-19, à transmission majoritairement aéroportée, les lieux clos visités sont plus importants que le réseau de contacts lui-même. À l'inverse, dans le cas d'une infection sexuellement transmissible, c'est vraiment le réseau de contacts sexuels qui détermine la propagation. D'autre part, même si le réseau qui importe est connu, on a souvent une idée assez vague de sa structure exacte. D'autant que la plupart des réseaux sont très dynamiques : les liens se font et se défont.

Néanmoins, les réseaux peuvent se révéler utiles dans le cadre du Covid-19. Dans notre étude avec les données de Facebook, nous avons essayé d'expliquer la propagation de l'épidémie en utilisant ces données pour paramétrer un réseau de contacts entre

94 départements français. Pour décrire ce réseau, nous avons construit une matrice – un tableau de 94×94 cellules, chacune contenant un nombre réel compris entre 0 et 1, qui représentait le pourcentage de contacts entre les deux départements concernés et qui évoluait chaque jour. Ce réseau pondéré nous a permis de voir comment, au cours du temps, les contacts entre les départements changeaient.

La théorie des réseaux est extrêmement développée en physique. En biologie, l'engouement a été fort dans les années 1990 et au début des années 2000, quand la théorie a atteint l'épidémiologie. Beaucoup d'articles sont parus sur les réseaux « invariants d'échelle », par exemple, des réseaux très hétérogènes dont les propriétés décrivent bien le réseau de contacts sexuels : la majorité des individus ont peu de contacts et quelques individus en ont énormément. Cette composante physique des épidémies a donné de très beaux résultats théoriques, en montrant notamment que l'hétérogénéité des réseaux modifie la vitesse de propagation des épidémies. Mais depuis la fin des années 2000, on étudie surtout comment les composantes biologique et sociale des épidémies mettent à mal certaines prédictions généralistes, et les limites de la transcription des résultats de la physique à l'épidémiologie.

à lire



THEMA Covid-19

La science face au défi du coronavirus

Comment fonctionnent les vaccins contre le Covid ? Les masques sont-ils efficaces ? Les tests de dépistage sont-ils fiables ? Que signifie le « taux de reproduction » R ? Le point sur l'épidémie dans ce n° 21 de la collection « Thema ».

À télécharger gratuitement

sur https://www.pourlascience.fr/sd/epidemiologie/thema-n021-20_430.php

Le Covid-19 a-t-il accéléré cette prise de conscience ?

L'épidémie d'Ebola en Afrique de l'Ouest avait déjà mis la question sur le devant de la scène en 2014. Une équipe de l'université Northeastern, à Boston, aux États-Unis, avait annoncé, en se fondant sur un modèle de réseau qu'elle avait développé, que l'épidémie qui avait ressurgi quelques mois plus tôt risquait de se propager à l'ensemble de la planète et de causer des millions de morts. Cela ne s'est pas du tout réalisé. Le modèle négligeait le cycle de vie de l'infection ainsi que l'efficacité des mesures de prévention. Mais le Covid-19 a aussi fait ressortir l'importance d'inclure les composantes sociale et biologique en introduisant un net changement d'échelle. Des milliers d'articles ont été publiés sur le sujet. Tout le monde s'est pris de passion pour l'épidémiologie, ce qui est très bien, mais souvent avec beaucoup d'applications directes de résultats mathématiques ou physiques, au détriment des composantes sociale et biologique.

Avez-vous observé d'autres changements de la modélisation des épidémies, déclenchés par l'étude de la pandémie actuelle ?

La collecte de l'information a fait un bond. La modélisation dépend beaucoup des données. Sans données, on peut juste savoir à quelle vitesse double l'épidémie et produire

un modèle de croissance exponentielle, comme début mars 2020. Plus on a de données, plus on peut affiner les modèles. Et, dans tous les pays, on a observé un développement considérable de l'acquisition des données de suivi de l'épidémie.

De plus, jamais auparavant la modélisation en épidémiologie n'avait eu un tel impact politique et médiatique. Dans la plupart des pays européens, les premiers confinements nationaux ont été décidés sur la base de modèles mathématiques ou informatiques. Certes, on était dans une situation inédite où le poids du biologique s'imposait de manière historiquement forte pour les pays occidentaux. Il n'en reste pas moins que ce constat provenait de modèles, ce qui était complètement nouveau.

Par ailleurs, la modélisation en épidémiologie est une science en constante évolution. Par exemple, certains outils couramment utilisés sont assez récents. Notamment, la technique dite «d'EpiEstim» pour estimer au mieux le R_0 , le nombre moyen de personnes infectées par une personne porteuse de l'infection, a été développée par Anne Cori, chercheuse à l'Imperial College de Londres, lors de l'épidémie d'Ebola en 2014. Les modèles, même ceux purement analytiques, changent aussi.

On a aussi utilisé des outils théoriques récents qui avaient principalement été testés rétrospectivement, sur des épidémies passées. C'est le cas pour la phylodynamique, qui consiste à analyser les séquences génétiques des virus afin de comprendre l'histoire des épidémies. Car la manière dont les microbes se propagent laisse des traces dans leurs génomes. C'est un champ très jeune, dont les outils les plus populaires ont moins de dix ans. Le terme lui-même ne date que de 2004. La phylodynamique existe à peine en France, mais nous montons justement un groupement de recherche avec le CNRS et l'Inrae pour la développer. Les Britanniques, en revanche, sont bien avancés dans le domaine, ce qui explique leur avance considérable en termes de séquençage et de génomique.

Quels sont les grands enjeux de la modélisation des épidémies, à présent ?

Aujourd'hui, on comprend bien la dynamique hospitalière. On sait assez bien la modéliser pour peu qu'il n'y ait pas de changement brusque de politique de santé publique. Les principaux enjeux sont les inconnues biologiques que sont l'évolution virale et l'efficacité de la réponse immunitaire. Dans l'équipe, en ce moment, nous essayons de développer des modèles mathématiques avec des formalismes dits «d'équation aux dérivées partielles» qui, justement, capturent mieux les variations d'immunité, car pour l'instant les hypothèses à ce sujet sont un peu simplistes: on suppose soit

que la personne est vaccinée soit qu'elle a été infectée, a guéri et est immunisée.

Un autre défi qui intéresse beaucoup la communauté est l'intégration des données dans les modèles statistiques: pour l'instant, en France, les modèles de description de l'épidémie s'appuient sur des données classiques (incidence de la maladie à l'hôpital, dépistage...). Le pays s'est mis doucement au séquençage des virus chez les personnes infectées par le SARS-CoV-2 afin de détecter les mutations associées aux variants, mais on n'injecte pas encore ces données dans les modèles, contrairement à d'autres pays comme le Royaume-Uni. On pourrait rajouter à cela les données de mobilité et les données environnementales, sachant que l'hétérogénéité géographique est encore difficile à suivre.

Mais la principale inconnue est probablement sociale. On ne se pose pas les mêmes questions de modélisation si on se lance massivement dans la prévention par des mesures sanitaires, si au contraire on restructure les capacités hospitalières pour augmenter les admissions en réanimation par an, ou si on table sur une troisième dose vaccinale pour augmenter l'immunité.

Et concernant les autres maladies ?

Pour le moment, les quelques modélisateurs en France commencent tout juste à se remettre de l'effort collectif produit en 2020 et 2021. Mais il y aura des ponts, c'est certain. Par exemple, l'approche de simulation sur OpenStreetMap, d'Olivier Thomine, pourra très clairement être transposée à d'autres maladies respiratoires, comme la grippe, voire à la transmission de maladies par des vecteurs comme le moustique-tigre. En fait, la modélisation des épidémies se nourrit en permanence des nouvelles approches. Là encore, les Britanniques ont de l'avance. Ils ont déjà commencé, *via* des numéros spéciaux de revues scientifiques, une rétrospective des approches des différentes équipes pour stimuler le dialogue et tirer les bilans sur l'épidémie actuelle. En France, la nouvelle agence ANRS-MIE tente d'impulser cela pour les maladies humaines en regroupant une trentaine de modélisateurs en épidémiologie, mais les moyens alloués sont dérisoires par rapport aux Anglo-Saxons.

Cela dit, un enseignement est déjà clair. L'interdisciplinarité est un élément crucial en épidémiologie: il est indispensable de maîtriser biologie, physique et enjeux sociaux. Cela illustre, en somme, ce que devrait être la santé publique: une institution profondément interdisciplinaire prenant en compte les différentes composantes de la santé, de l'individu aux populations.

**Propos recueillis par
Marie-Neige Cordonnier**

BIBLIOGRAPHIE

C. Selinger *et al.*, **Predicting Covid-19 incidence in French hospitals using human contact network analytics**, *Int. J. Infect. Dis.*, vol. 111, pp. 100-107, 2021.

M. T. Sofonea *et al.*, **Memory is key in capturing Covid-19 epidemiological dynamics**, *Epidemics*, vol. 35, article 100 459, 2021.

M. T. Sofonea et S. Alizon, **Anticipating COVID-19 intensive care unit capacity strain : A look back at epidemiological projections in France**, *Anaesth. Crit. Care Pain Med.*, vol. 40(4), article 100 943, 2021.

G. Danesh *et al.*, **Quantifying transmission dynamics of acute hepatitis C virus infections in a heterogeneous population using sequence data**, *Plos Pathog.*, vol. 17(9), article e100 9916, 2021.