

début de l'épidémie de Covid-19, cette composante était une inconnue majeure. En février 2020, on ignorait, par exemple, la proportion des personnes infectées qui décédaient, une information essentielle pour élaborer le moindre modèle.

La deuxième composante est physique: comment l'épidémie se propage une fois que l'on a pris en compte les contraintes biologiques. C'est la partie la plus « facile », parce que les lois de la physique s'appliquent assez bien au processus de contagion. Mais une troisième composante vient contrecarrer les prédictions: la composante sociale. Les comportements peuvent changer du tout au tout d'un jour à l'autre. Si demain le gouvernement confine le pays, ou si, à l'inverse, il relâche tous les gestes barrières, l'épidémie suivra des trajectoires diamétralement opposées. Prévoir l'état de l'épidémie à plus de deux semaines nécessite donc de prévoir les politiques de santé...

C'est pourquoi on explore différents scénarios. Celui de référence est en général la tendance actuelle. Puis on explore des scénarios qui dévient de cette référence. Nous sommes début septembre: que se passe-t-il si la rentrée a un effet que l'on ne voit pas encore à cause des délais entre l'infection et l'admission en réanimation?

On peut aussi produire des scénarios rétrospectifs. Notre équipe l'a fait dans deux cas: un scénario regardant ce qui se serait passé si l'on n'avait pas eu de vaccin ou si les vaccins avaient uniquement empêché les formes sévères, sans empêcher la transmission, et un scénario où le troisième confinement, qui a eu lieu début avril 2021, avait commencé deux mois plus tôt, suivant les recommandations du Conseil scientifique. Là encore, on ne modélise pas exactement ce qui se serait passé, de même qu'on ne modélise pas le futur. On explore une tendance en fonction d'hypothèses données.

Utilisez-vous les scénarios rétrospectifs pour vérifier vos prévisions?

Oui, notre équipe l'a fait sur nos scénarios prospectifs publiés depuis douze mois afin d'évaluer leur écart avec la dynamique hospitalière observée. Nous avons trouvé qu'en moyenne, notre modèle et nos hypothèses donnaient des tendances robustes sur cinq semaines, ce qui est assez satisfaisant sachant qu'à plus de deux semaines vous êtes à la merci d'un changement de politique sanitaire.

Nous avons aussi identifié pourquoi nos projections étaient trop pessimistes fin avril 2021, lors du troisième déconfinement. Le modèle fonctionnait, mais le problème venait des paramètres d'entrée, car, avec les données de l'époque, on supposait que l'épidémie décroissait moins vite qu'elle ne le faisait en réalité. Pourquoi? On ne peut que faire des

conjectures, car, en France, on ne suit pas d'assez près l'épidémie pour avoir une réponse précise. Une hypothèse est que le troisième confinement a coïncidé avec le début des vacances scolaires et qu'une mesure de contrôle portant sur les écoles met plus de temps à se manifester. En effet, non seulement les chiffres hospitaliers ont un retard de deux semaines sur l'épidémie, mais comme les enfants font très peu de formes sévères, le décalage est encore plus important. D'ailleurs, deux semaines plus tard, le 16 mai, avec les chiffres mis à jour, le scénario a reproduit la tendance observée jusqu'à début juillet.

Quel type de données utilisez-vous dans vos modèles?

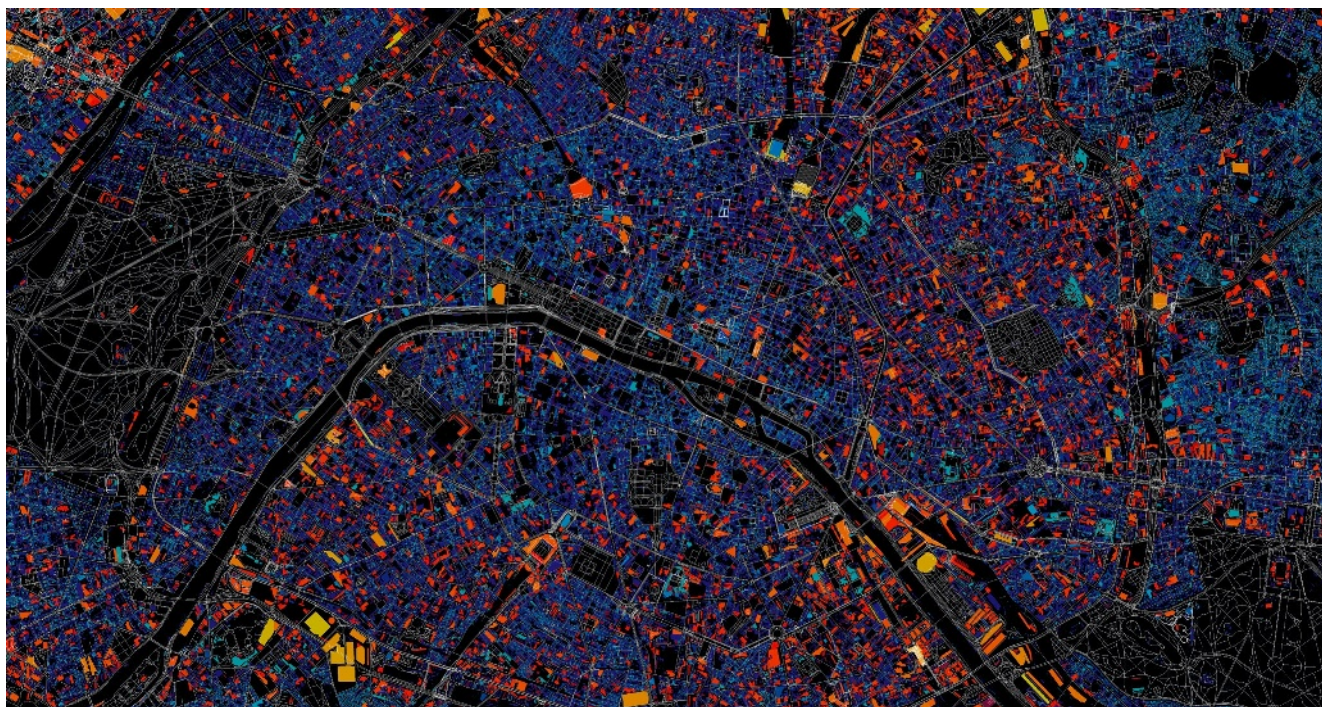
Nous utilisons principalement les données hospitalières, pour deux raisons. D'une part, l'échantillonnage est exhaustif: toutes les personnes hospitalisées à cause du Covid-19 sont comptabilisées, alors qu'on dépiste moins de la moitié des personnes infectées. D'autre part, les admissions en réanimation ont un délai de « seulement » deux semaines entre l'infection et l'observation, alors qu'il est d'un mois avec les décès. Certes, en théorie, les dépistages permettent un suivi au plus près de l'épidémie,



Des informations sur la mobilité aident à anticiper le nombre d'hospitalisations

car les symptômes apparaissent généralement quatre jours après infection. Mais pour s'affranchir des biais, il faut une réflexion statistique. Par exemple, le Royaume-Uni organise des campagnes de dépistage aléatoires où des milliers de personnes, avec ou sans symptômes, sont testées chaque jour, avec un plan d'échantillonnage réfléchi. Ainsi, même avec moins de tests qu'en France, on a une photographie bien plus précise de l'épidémie.

Mais nous avons aussi développé d'autres modèles avec des données plus originales, comme celles de Facebook. L'idée est d'utiliser ce qu'on appelle des « signaux précoces » pour suivre l'épidémie au plus près à l'aide de modèles statistiques: connaissant la dynamique hospitalière passée, avoir des



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