HAL open science

COMOKIT : un environnement générique et modulaire pour analyser les impacts des politiques d'intervention contre l'épidémie de COVID-19

Kevin Chapuis, Patrick Taillandier, Benoit Gaudou, Arthur Brugière, Alexis Drogoul

To cite this version:

Kevin Chapuis, Patrick Taillandier, Benoit Gaudou, Arthur Brugière, Alexis Drogoul. COMOKIT : un environnement générique et modulaire pour analyser les impacts des politiques d'intervention contre l'épidémie de COVID-19. Journées Francophones sur les Systèmes Multi-Agents (JFSMA) 2021, Jun 2021, Bordeaux, France. hal-03413031

HAL Id: hal-03413031

https://hal.science/hal-03413031

Submitted on 3 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COMOKIT : un environnement générique et modulaire pour analyser les impacts des politiques d'intervention contre l'épidémie de COVID-19

K. Chapuis[a] kevin.chapuis@gmail.com
P. Taillandier[b,c,d] patrick.taillandier@inrae.fr
B. Gaudou[e] benoit.gaudou@ut-capitole.fr
A. Brugière[c,d] arthur.brugière@ird.fr
A. Drogoul[c,d] alexis.drogoul@ird.fr

[a]UMR 228 ESPACE-DEV,
IRD, Université de Montpellier, France

[b]UR 0875 MIAT,
INRAE, France

[c]UMI 209 UMMISCO,
IRD, Sorbonne Université, France

[d]Université Thuyloi, Hanoi, Vietnam

[e]UMR 5505 IRIT,
Université Toulouse 1 Capitole, CNRS, France

Résumé

Depuis le début de la pandémie de la COVID-19, de nombreux travaux ont cherché à modéliser et comparer les impacts des politiques d'intervention non pharmaceutique. Cette volonté s'est cependant heurtée à la difficulté à pouvoir facilement prendre en compte la variété des cultures, des comportements et des contextes dans lesquels se déploient à la fois le virus et les politiques de santé publique. Initié pour soutenir les autorités vietnamiennes dans le déploiement d'interventions à l'échelle locale, l'environnement logiciel COMOKIT, qui repose sur un modèle à base d'agents, a été conçu dès le départ pour être modulaire et facilement déployé sur différents cas d'étude, quelles que soient la disponibilité et la qualité des données démographiques, géographiques et épidémiologiques. Intégrant des outils de construction de l'environnement géographique et de génération de populations synthétiques, COMOKIT permet une modélisation détaillée des individus et de leurs activités à une temporalité fine. Cet article présente son adaptation à la comparaison de l'impact de mesures de confinement dans le centre-ville de Nice.

Mots-clés : *simulation à base d'agent, COVID-19, plate-forme GAMA*

Abstract

Since the beginning of the COVID-19 pandemic, much work has sought to model and compare the impacts of intervention policies. However, this willingness has been hampered by the difficulty of being able to easily consider the variety of cultures, behaviours, and contexts in which both the virus and public health policies are deployed. Initiated to support the Vietnamese authorities in the deployment of interventions on a local scale, the COMOKIT software environment, which is based on an agent-based model, was designed initially to be modular and easily deployed on different case studies, regardless of the availability and quality of demographic, geographic and epidemiological data. COMOKIT integrates tools for constructing the geographic environment and generating synthetic populations, and allows detailed modelling of individuals and their activities at a fine temporality. This paper presents its adaptation to the comparison of the impact of containment measures in the city centre of Nice.

Keywords: *agent-based simulation, COVID-19, GAMA platform*

1 Introduction

L'apparition de la COVID-19 en décembre 2019 a profondément impacté nos sociétés. Face à cette maladie inconnue, les autorités de santé publique ont été contraintes d'expérimenter, en peu de temps, diverses politiques d'intervention, à différentes échelles, sans pouvoir systématiquement anticiper et évaluer leurs impacts. Face à cette double difficulté, la modélisation est très vite apparue comme un outil puissant permet-

tant d'explorer *in silico* une gamme de stratégies d'intervention avant leur éventuelle mise en œuvre sur le terrain. De nombreux modèles ont ainsi été développés et utilisés très vite après l'apparition des premiers cas. Parmi ceux-ci, une majorité repose sur de la modélisation mathématique (modèles SEIR à compartiments comme [2]), et certains ont eu un impact très important sur les choix de politiques d'interventions [1]. Si ces modèles sont particulièrement utiles pour apporter une réponse rapide et globale, ils ne permettent pas de prendre en compte ce qui apparaît aujourd'hui comme un aspect fondamental de l'épidémie, à savoir l'hétérogénéité des individus en termes de relations sociales, de comportements et d'attitudes vis-à-vis de la maladie [9]. Nous affirmons dans cet article que la modélisation à base d'agents offre une alternative capable de prendre en compte cet aspect, et qu'elle permet également de fournir le cadre et les outils appropriés pour construire des modèles qui soient réalistes et généralisables à différents contextes [8].

Ainsi, si [15] propose un modèle à base d'agents qui permet d'étudier l'impact d'interventions, telles que le confinement et la distanciation sociale, sur le nombre de patients hospitalisés, il ne prend pas en compte les aspects sociologiques : la transmission de la maladie ne s'appuie pas sur les foyers, les lieux de travail ou autres lieux de rassemblement, pourtant connus pour faciliter infection et propagation. De plus, la représentation simplifiée de l'espace ne permet pas de modéliser la contamination via l'environnement alors que plusieurs études ont montré que le virus y survit jusqu'à plusieurs heures [6], pouvant ainsi entraîner une contamination et une transmission environnementale, mais aussi des infections nosocomiales [13].

Un autre exemple de modèle à base d'agents est celui proposé par [3]. Ce modèle, dérivé d'un modèle de transmission de la grippe, permet de tester des interventions telles que la fermeture des écoles, l'interdiction de voyager, la distanciation sociale et l'isolement. Utilisé à l'échelle de l'Australie pour réaliser des simulations sur 24 millions d'individus ayant leurs caractéristiques et leur contexte social, il ne prend cependant pas en compte la capacité limite du système hospitalier ou les activités de loisir qui ont pourtant un impact fort sur la transmission de la maladie [7].

Dans le cas de [12], les auteurs se basent sur un modèle épidémiologique générique à base d'agents, doté d'une représentation explicite de l'espace, afin d'étudier les conditions de sortie de confinement. L'étude au niveau de la France permet de tester différents scénarios et inclut une adhésion variable des agents aux contraintes telles que la distanciation sociale ou le port du masque, sans pour autant représenter la richesse et l'hétérogénéité des pratiques sociales, ni l'impact de la dynamique épidémiologique des pays alentours.

C'est dans ce contexte de foisonnement de modèles et d'urgence à produire des réponses adaptées que nous avons développé l'environnement logiciel COMOKIT (COVID-19 Modeling Kit) [1]. Cet environnement repose sur un modèle à base d'agents conçu pour être générique, évolutif et donc applicable à une variété de contextes sociaux et géographiques : en plus de permettre une représentation très fine des politiques d'intervention, il permet par exemple de représenter les activités de loisir ou d'autres activités sociales, de prendre en compte la transmission environnementale, l'impact des zones environnantes non explicitement modélisées, ainsi que la capacité limite de prise en charge hospitalière [10]. COMOKIT combine quatre sous-modèles qui interagissent de façon dynamique : (1) un modèle de transmission de la maladie de personne à personne et via l'environnement, (2) un modèle d'évolution du statut épidémiologique individuel, (3) un modèle de mobilité individuelle et quotidienne à l'échelle de l'heure et du bâtiment et (4) un modèle permettant de représenter les impacts de la combinaison de plusieurs politiques d'intervention.

Cet article présente une vue d'ensemble de COMOKIT : ses outils de reconstruction des données géographiques et des populations synthétiques (Section 2.1), le modèle à base d'agents qui constitue le coeur de cet environnement (Section 2.2) et les outils d'analyse des résultats dont il est doté (Section 2.3). De plus, nous illustrons, au travers d'une application à un quartier de la ville de Nice (France), comment COMOKIT peut facilement être adapté à un nouveau cas d'étude (Section 3).

2 L'environnement COMOKIT

COMOKIT a été initialement conçu pour répondre au besoin du gouvernement vietnamien d'outils capables d'aider la prise de décision sur les choix de politiques d'intervention (confinement, obligation de port du masque, etc.) à l'échelle d'une petite ville ou d'un quartier

1. https://comokit.org

(d'environ 10 000 habitants). Cette échelle correspond à la taille de la première zone mise en quarantaine au Vietnam au début de la pandémie (commune de Son Loi dans la province de Vinh Phuc, mise en quarantaine le 13 février 2020). Ce contexte initial explique qu'un élément important qui a guidé le développement de COMOKIT a été d'offrir une grande souplesse dans la gestion des données d'entrée : au Vietnam les données géographiques, démographiques ou de mobilités sont généralement inexistantes ou très lacunaires, contrairement à la France elles sont globalement accessibles. Ainsi, COMOKIT a été conçu pour pouvoir être rapidement appliqué à tout cas d'étude en utilisant le minimum de données en entrée, mais en permettant également d'intégrer des données supplémentaires ou plus détaillées lorsqu'elles sont ou deviennent disponibles. De même, un point fort de COMOKIT est sa modularité : comme illustré dans la Section 3, COMOKIT offre la possibilité à tout utilisateur de redéfinir pour son cas d'étude toutes les fonctions proposées, aussi bien concernant le pré-traitement des données que le contenu du modèle lui-même. COMOKIT est ainsi composé de plusieurs modules permettant à la fois de reconstruire toutes les données nécessaires à son application pour un cas d'étude particulier, de lancer des simulations sur un ordinateur ou une grille de calcul et d'analyser les résultats. Cet environnement comprend :

— un ensemble de fichiers modèles en GAML, le langage de programmation de la plateforme GAMA [14], permettant de recueillir les données spatiales nécessaires et générer la population synthétique avec leur agenda.

— un ensemble de fichiers modèles en GAML correspondant au modèle agent simulant le territoire et l'épidémie. Ce module constitue le coeur de COMOKIT.

— une librairie en Python permettant de générer une exploration exhaustive, l'exécuter sur une architecture de calcul haute performance et d'organiser les données brutes de sortie pour en permettre l'analyse.

Les sections suivantes décrivent les 3 composants de COMOKIT.

2.1 Module de construction et de prétraitement des données d'entrées

Appliquer COMOKIT sur un nouveau territoire requiert de disposer de données sur la localisation et la fonction des bâtiments, et de données sur la population, telles que la taille et composition des ménages, ainsi que l'âge, le genre et les activités d'une semaine type pour chaque individu. Dans de nombreux cas, ces données sont soit partiellement disponibles soit proprement absentes. COMOKIT fournit 3 outils spécifiques et indépendants pour pallier l'absence de données sur la zone géographique étudiée, les populations qui l'habitent et leurs activités quotidiennes. Dans les trois paragraphes ci-dessous sont décrits le contenu et fonctionnement de ces modules.

Construction des données géographiques. Cet outil demande à l'utilisateur de fournir l'extension spatiale de son territoire d'étude sous la forme d'un fichier shapefile (fichier de données spatiales vectorielles). A partir de cette extension spatiale, COMOKIT télécharge toutes les données OSM (OpenStreetMap) correspondantes et les prétraite de façon à en extraire les bâtiments avec leur fonction et leur nombre de logements. Si les données OSM sont très complètes pour des pays comme la France, ce n'est pas le cas pour beaucoup de pays du Sud. COMOKIT offre dans ce cadre la possibilité de télécharger automatiquement les données Google map correspondant à l'extension spatiale étudiée, de les vectoriser et de les analyser de façon à compléter les données sur les bâtiments disponibles.

Génération de population. COMOKIT fournit deux méthodes de génération de population. La première se base sur des données démographiques (*e.g.* pyramide des âges selon le sexe) mobilisées par le générateur de population Gen* [5, 4]. L'autre méthode est un générateur de population *ad hoc* qui permet de générer une population à partir d'un minimum de données, tel que le nombre d'individus et la proportion d'hommes et de femme.

Génération d'agenda. Cet outil permet, pour chaque individu, en fonction de son âge, de son sexe et de son statut professionnel, de déterminer un agenda d'activités : pour chaque jour d'une semaine type, un individu se voit affecter un ensemble d'activités. Une activité est un tuple comprenant l'heure de début de l'activité, le type d'activité et la liste des personnes partageant l'activité (par exemple, aller au restaurant avec des amis).

2.2 Modèle principal

Échelles spatiales et temporelles. L'unité spatiale de base du modèle est le bâtiment : ces

derniers sont représentés explicitement et constituent les lieux d'activité des agents. Un pas de simulation correspond par défaut à 1 heure. En conséquence, nous avons fait le choix de ne pas simuler les mouvements d'un lieu d'activité à un autre : les individus sont toujours situés dans un lieu d'activité (qui peut être un bâtiment proche ou même un parc extérieur). L'hypothèse sous-jacente est qu'aucune infection ne peut se produire pendant la durée du déplacement.

Entités. L'ensemble des entités constituant le modèle COMOKIT et leurs relations ont été synthétiquement représentées dans le diagramme de classes UML en Figure 1. Par soucis de clarté, les attributs ne sont pas représentés.

Les entités centrales du modèle sont les agents de type *Individual* : ils représentent les habitants de la zone d'étude avec leurs caractéristiques individuelles (âge, sexe, statut professionnel), leur statut épidémiologique (issu d'un modèle classique SEIR), et un ensemble de variables relatives aux interventions politiques (*e.g.* s'ils ont été testés, s'ils doivent se confiner). Ils effectuent leurs activités quotidiennes (par exemple aller au travail, à l'école, faire des courses) en fonction de leur agenda personnel. Ces agents ont également des attributs en lien avec leurs réseaux sociaux : la listes des individus du ménage et de leurs parents proches, leurs amis avec lesquels ils sont susceptibles d'engager des activités partagées, ainsi que leurs collègues de travail (ou selon l'âge, camarades de classe). Ils ont enfin un bâtiment de résidence et selon leurs activités un bâtiment de travail et/ou d'étude.

Les agents *Building* sont les entités spatiales où les agents *Individual* effectuent leurs activités. Deux types de bâtiments spéciaux ont été définis car ils jouent un rôle particulier : l'extérieur (*outside*) et les hôpitaux (*hospital*). L'agent *outside* représente tous les bâtiments situés en dehors de la zone simulée : il est utilisé pour représenter le fait que des personnes peuvent travailler ou effectuer des activités en dehors de la zone considérée. Pour cette zone, une dynamique de contamination particulière est appliquée. Les agents *hospital* sont le lieu où, dans certaines situations, les agents individuels infectés présentant des symptômes critiques peuvent être isolés, traités et le cas échéant guérir.

Le comportement des agents *Individual* est déterminé en grande partie par leur agenda : cet attribut associe à certaines heures de la journée une activité. Chacune des activités possibles, représentée par un agent *Activity*, spécifie l'unité ou les unités spatiales dans lesquelles un agent peut effectuer ce type d'activité. Le choix de l'unité spatiale dépend des préférences de l'agent *Individual* encodées sous la forme de poids et qui sont définies en fonction de son âge et de son sexe : pour une activité de loisir, un enfant peut préférer aller dans un centre de jeux tandis qu'une personne plus âgée peut préférer aller au cinéma. Une fois le type de bâtiment choisi, COMOKIT propose 3 méthodes permettant de déterminer un bâtiment spécifique où effectuer l'activité : choix d'un lieu aléatoire uniforme, choix du lieu le plus proche de l'individu ou méthode gravitaire. Dans ce dernier cas, la probabilité de choisir un lieu est une fonction croissante de la superficie du lieu et inversement proportionnelle de sa distance à l'individu.

Nous avons prédéfini les principaux types d'activités classiques dans un environnement social urbain : visiter un voisin (*visiting_neighbor*), travailler (*working*), rester à la maison *staying_at_home*), étudier (*studying*), rendre visite à un ami (*visiting_friend*). Bien entendu, des activités supplémentaires peuvent également être créées à partir de l'agent générique *Activity*.

Le dernier ensemble d'entités concerne la régulation des comportements sous la forme d'interventions non pharmaceutiques pour lutter contre la propagation du virus. La capacité d'un agents *Individual* à réaliser une activité peut être limitée par l'agent *Authority*. Cet agent gère les différentes politiques qui sont adoptées et représentées par des agents *Policy*. Avant chaque activité, un agent *Individual* demande l'autorisation à l'agent *Authority*. Celle-ci examine toutes les politiques adoptées, pour vérifier qu'elles autorisent cette activité pour cet individu. Ces politiques peuvent se combiner (*CompoundPolicy*) pour former des interdictions limitées (*FowardingPolicy*) dans le temps et l'espace, à un sous groupe d'agents donné ou à un ensemble dynamique (*e.g. PositiveAtHome* : lorsqu'un agent *Individual* a été testé positif, il doit rester chez lui). A noter qu'un agent *Individual* peut choisir de respecter ou non la régulation définie : cette capacité est représentée par une probabilité de respecter la régulation qui est propre à chaque agent *Individual* et qui peut évoluer durant la simulation.

Dynamiques et processus. La dynamique du modèle peut être résumée en trois processus principaux : la dynamique de l'épidémie, les activités quotidiennes des agents *Individual* et

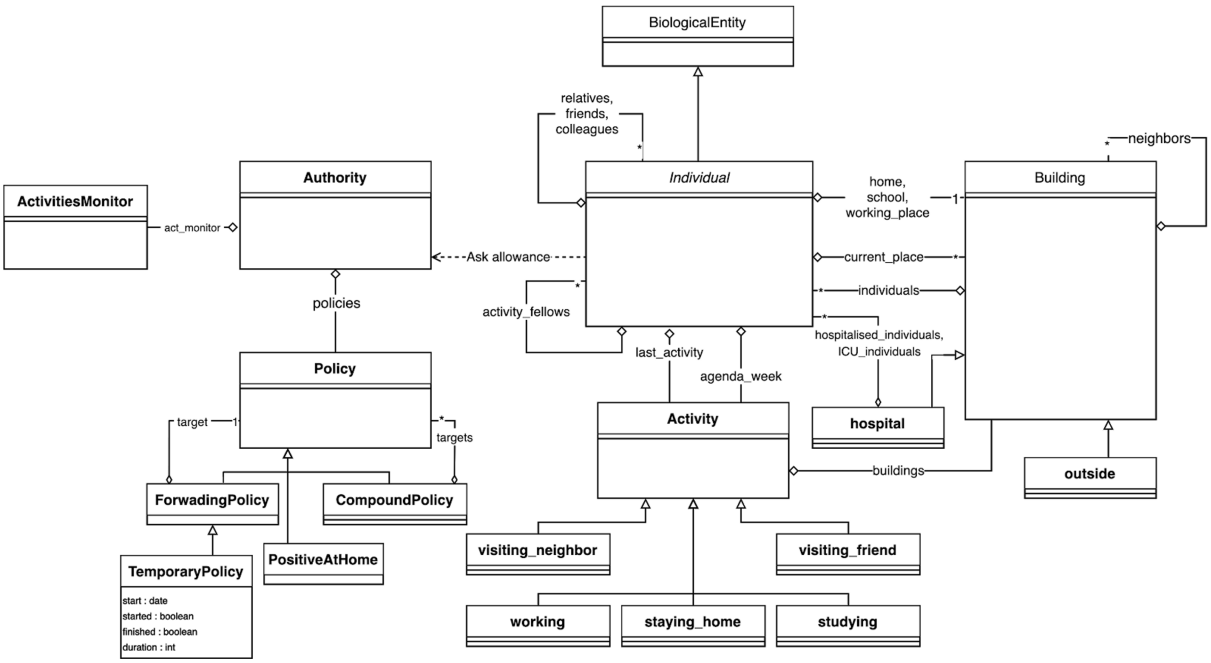


FIGURE 1 – Diagramme de classes UML du modèle COMOKIT

l'adoption et l'application des politiques d'intervention.

Un pas de simulation commence par l'évolution de la charge virale dans les bâtiments (elle diminue au fil du temps, avant de disparaître). Ensuite, les agents *Individual* agissent : ils évaluent d'abord s'ils ont été infectés et, dans le cas où ils sont contagieux, s'ils infectent d'autres agents *Individual* et déposent une charge virale dans le bâtiment où ils se trouvent. Ils mettent ensuite à jour leur statut épidémiologique (voir ci-après) et leur comportement individuel (*e.g.* choix de porter ou non le masque, mise en quarantaine). Enfin, ils exécutent leurs activités quotidiennes : ils trouvent l'activité correspondant à l'heure courante, demandent à l'agent *Authority* s'ils sont autorisés à l'exécuter et agissent en conséquence. Enfin, l'agent *Authority* vérifie sa politique actuelle et essaie de l'appliquer, comme par exemple en exécutant une campagne de test.

L'évolution du statut épidémiologique d'un individu est décrit par un ensemble d'états et de transitions inspiré d'un modèle SEIR : à l'initialisation, nous supposons que la population entière commence la simulation à l'état dit *Susceptible*. En effet, comme il s'agit d'une maladie émergente, personne n'a *a priori* été en contact avec le virus, aucun agent n'est donc immunisé [2]. Le modèle intègre deux voies d'infection différentes pour les agents : soit par transmission d'individu à individu, soit par persistance du virus dans l'environnement. Lorsqu'un agent est en contact avec un autre agent infectieux ou se trouve dans un bâtiment à la charge viral non nulle, il peut être infecté par *Exposition* et passer à l'état *Latent*, en fonction du succès de la transmission. La probabilité qu'un contact avec une surface ou un agent infectieux aboutisse à une exposition effective dépend de la charge virale (*e.g.* la charge virale est *a priori* plus faible chez les agents pre- ou a-symptomatiques), du type de contact (*e.g.* plus "intense" par définition entre deux parents qu'entre deux inconnus) et des précautions éventuelles prises par les agents (*e.g.* port du masque, distanciation sociale). Une fois la période de latence écoulée, un individu passe à l'un des trois états *Infectieux* possibles : asymptomatique, pré-symptomatique ou symptomatique. L'individu reste dans ces états pendant l'intervalle sériel pour les pré-symptomatiques et/ou la période infectieuse pour les symptomatiques et asymptomatiques. A l'issu de la période infectieuse, les agents passent systématiquement à l'état *Retiré* : ils ne sont plus infectieux et tombent dans l'un des deux sous-compartiments, *Rétabli* ou *Mort*. Dans le module épidémiologique de COMOKIT, ce dernier sous-compartiment est une conséquence de l'état clinique grave potentiellement développé par les individus symptoma-

2. L'introduction du virus est un élément de modélisation à part, soit introduite de manière exogène (*i.e.* bâtiment spécifique *outside*) soit en contaminant a priori un ou plusieurs agents.

tiques. En effet l'état *Infectieux* se décompose en sous-compartiment spécifique aux agent développant des symptômes : ils sont a priori placés dans un sous compartiment ne nécessitant pas d'hospitalisation, puis suivant une probabilité d'aggravation sont susceptibles de passer dans le sous compartiment nécessitant une hospitalisation (I_H), puis dans un troisième temps dans le sous compartiment nécessitant une unité de soins intensifs (I_{ICU}). Cette probabilité dépend de la gravité (donnée) de la maladie pour la catégorie d'âge de l'agent et des soins qui lui ont été prodigués ; en effet, si la capacité hospitalière (défini dans les bâtiments spéciaux *Hospital*) permet de traiter les agents du sous compartiment I_H, ils auront relativement au facteur déterminant de l'âge une moindre probabilité de nécessiter une prise en charge en sois intensifs (*i.e.* transition vers l'état I_{ICU}). A l'inverse si la prise en charge est impossible, la probabilité de transition entre ces deux états s'en trouve accrue. Comme souligné plus haut, dans le cas où par manque de capacité de traitement hospitalier, l'agent ne peut pas recevoir les soins intensifs requis par son état, il mourra systématiquement. A l'inverse, tout agent symptomatique qui ne requiert pas de soins intensifs sera systématiquement rétabli à l'issue de la phase infectieuse de la maladie.

Les temporalités ainsi que les conditions de transition entre état sont toutes paramétrables. Elles ont fait par ailleurs l'objet de recherche bibliographique pour en établir les valeurs par défaut en fonction des connaissances disponibles. Pour de plus ample détails sur le modèle, les dynamiques et le choix des valeurs de paramètre, voir [10].

2.3 Module d'analyse des résultats

Exploration du modèle. COMOKIT fournit un ensemble d'outils de visualisation et d'expérimentation permettant de sauvegarder les sorties d'intérêts, tel que l'incidence cumulée, le nombre de patients traités à l'hôpital et en soins intensifs, le nombre de cas réels et connus par voie de test, ou encore le nombre d'individu qu'une personne contaminée va infecter (base désagrégée du taux de reproduction R0). Ces fichiers de sorties proposent par défaut la valeur de l'ensemble des indicateurs pour chaque pas de simulation, par tranche d'âge de 5 ans et pour chaque type d'activité, totalisant plusieurs dizaines de fichiers au format csv par simulation pouvant avoisiner plusieurs gigaoctets de données brutes par exploration.

Pour explorer le modèle, l'environnement COMOKIT inclut également un ensemble de scripts pour la définition d'exploration exhaustive et leur déploiement sur un supercalculateur. L'ensemble des paramètres du modèle peuvent ainsi être explorés conjointement, en spécifiant pour chacun d'eux l'ensemble des valeurs possibles. La condition d'arrêt des simulations peut également être spécifiée.

Dans [10], nous avons ainsi pu lancer plusieurs centaines de simulations pour déterminer *a priori* un seuil de stabilité des résultats afin de limiter le nombre de réplications : suite à l'analyse de la dispersion des indicateurs d'intérêts en fonction du nombre de réplication de simulation, il a été déterminé qu'un seuil de 50 constituait le meilleur compromis entre temps de calcul et qualité des résultats.

Post-traitement. Le traitement des données massives (125Mo [3] de données au format CSV par combinaison de paramètres) de sortie de l'exploration se fait à l'aide de scripts Python exécutés directement sur le supercalculateur pour limiter les transferts de données. Le post-traitement consiste en une agrégation de toutes les données brutes, puis la compilation et traitement de celles-ci pour produire un fichier CSV synthétique contenant divers indicateurs paramétrables (comme le calcul des quartiles, la valeur moyenne, etc.) et des images de visualisations des résultats (cf. graphiques Section 3.2).

3 Cas d'étude de Nice

3.1 Adaptation du modèle

Dans le cadre de cet article, nous présentons une application pour la ville de Nice en France où l'enjeu est d'utiliser COMOKIT pour mieux étudier l'impact spatiale de politiques de confinement sur la propagation du COVID-19. Cette exemple est une bonne illustration d'adaptation de COMOKIT à un cas d'étude. En effet, par rapport au cas d'étude initial au Vietnam, diverses adaptations de COMOKIT ont été nécessaires : disponibilité de données supplémentaires sur les bâtiments (données IGN), les populations et leurs mobilités (enquête ménage déplacement), meilleure prise en compte de la dynamique épidémiologique à l'extérieur de la zone et nouvelle politique de confinement basée sur

3. 115Ko CSV * 21 CSV * 50 réplications = 125Mo

la distance des lieux d'activités au logement des ménages.

Une première adaptation concerne l'intégration de données supplémentaires dans la génération de l'environnement. COMOKIT peut récupérer les données sur un cas d'étude à partir des données OSM (qui sont généralement en France de bonne qualité). Néanmoins, nous disposions ici des données IGN (BD Topo) sur la zone d'étude (qui sont plus complètes [11]), nous avons donc modifié la méthode de création des bâtiments sur la zone d'étude de façon à fusionner ces deux types de données, en particulier pour enrichir les données sur le type de bâtiments et le nombre de logements à l'intérieur.

En ce qui concerne la génération de la population, nous avons croisé, dans le cadre de l'application à l'aire urbaine de Nice, la disponibilité de données agrégées sur la composition des ménages, la distribution des attributs individuels et les densités de population (individus et ménages) à l'échelle de la cellule du carroyage INSEE (carreaux de 200m). En mobilisant un algorithme hiérarchique basé sur le principe de reconstruction synthétique [4] disponible dans la librairie Gen*, nous avons pu générer les caractéristiques individuelles, agencer les agents synthétiques en ménage tout en respectant au mieux les marges statistiques concernant leur structure, et localiser les lieux d'habitations sous contraintes du respect du nombre d'individus et de ménages par cellule (échelle spatiale de référence).

Toujours concernant la génération de la population, contrairement au cas d'étude vietnamien, il existe pour la France des données sur les mobilités quotidiennes des personnes : les enquêtes mobilités EMC^2 [4]. Il existe ainsi pour la ville de Nice des données de 2017 sur les déplacements quotidiens d'un petit échantillon de ménages. Nous avons utilisé ces données pour régénérer des agendas pour nos agents *Individual*. Néanmoins, ces informations étant lacunaire (pas de données sur les activités du week-end), nous avons dû proposer une méthode hybride de reconstruction d'agenda utilisant quand cela était possible nos données et dans le cas contraire le générateur de COMOKIT.

Concernant les aspects dynamiques, nous avons modifié la dynamique d'infection en dehors de la zone géographique modélisée. En effet, dans COMOKIT, des agents *Individual* peuvent effectuer toute ou partie de leurs activités à l'extérieur de la zone d'étude (*i.e.* dans le bâtiment spécifique *outside*) ou au contraire habiter à l'extérieur de la zone d'étude et venir faire une ou plusieurs activités dans l'espace explicitement modélisé. Par défaut, COMOKIT propose une dynamique épidémiologique très simple pour l'extérieur : chaque individu a, à chaque heure passée à l'extérieur, une probabilité fixe d'être infecté. L'idée pour notre application de Nice était de tirer partie des données disponibles sur la propagation de la maladie en France. Ainsi, le modèle ajuste la probabilité de contamination sur les valeurs connues de cas issues des bases de données nationales [5]. La probabilité pour un agent d'être contaminée à l'extérieur correspond pour un jour donnée à l'incidence connue sur la zone la plus précise possible (*i.e.* Département des Alpes-Maritimes, Région PACA, France et monde) par rapport à la population totale considérée.

Une dernière modification a concerné la modélisation des politiques d'intervention non pharmaceutique. Nous souhaitions en effet tester l'effet de politiques spatiales, notamment pour n'autoriser que les activités proches du domicile des individus, dans un radiant fixe ou sur une zone urbaine pré-définie.

3.2 Expérimentions réalisées

De façon à évaluer l'impact de différentes politique de confinement, nous avons simulé la propagation de la maladie dans un quartier de la ville de Nice du 24 janvier au 20 octobre 2020 en prenant en compte le contexte epidémiologique connu sur cette période. Parmi les conditions limites de notre expérience, nous avons choisi de ne pas contraindre le nombre de lits dans les hôpitaux, y compris les unités de soins intensifs, et de considérer que les agents *Individual* ont une attitude passive vis-à-vis de leur propre statut épidémiologique (le fait d'être infecté ne modifie pas leur comportement).

Nous avons testé trois scénarios, dont deux impliquant une intervention : le scénario de base peut être décrit par l'absence de politique d'intervention, les deux seconds se caractérisant d'une part par une politique réaliste inspirée par les actions gouvernementales menées en France et d'autre part par une politique hypothétique de fermeture de zones urbaines. Chacun des deux scénarios suit le développement temporel des mesures prisent en France lors du 1er confinement entre le 17 mars et le 11 mai 2020.

4. anciennement enquête ménages-déplacements ou EMD

5. donnée issue du dépôt officiel à l'adresse suivante : `https://github.com/opencovid19-fr/data`

Le confinement "réaliste" implique, à l'image des mesures prises par le gouvernement Français en Mars 2020, une suspension de la plupart des activités : les agents *Individual* ne peuvent pas quitter leur domicile pour effectuer une activité dans un autre bâtiment sauf pour les activités essentielles, à savoir les achats de première nécessité, soins, activités administratives, travail ne pouvant être réalisé à domicile. Concernant cette dernière limitation, nous avons fait le choix de définir un ensemble d'agents *Individual* autorisés à travailler : dans notre cas, nous exprimons cela comme une proportion de personnes qui peuvent agir librement (25%) - l'enquête INSEE Focus n°207 estime que durant le 1er confinement, 45% des français ne travaillaient pas ou peu, alors que 47% étaient en télétravail. Les autres activités ne peuvent être effectuées que dans un rayon donné de 1 km autour du lieu de résidence et pour une durée maximale d'une heure.

La dernière politique testée est un confinement spatialisé. A l'image de la limitation d'un km autour du domicile pour les activités non essentielles dans le cas de la politique d'intervention française, l'idée est de produire des zones de clôture, que l'on ne peut pas quitter mais où toutes les activités sont autorisées, en s'appuyant sur la structure urbaine plutôt que sur une distance arbitraire par rapport au domicile. En effet, cette distance unique ne prend pas en compte le contexte sociale et urbain qui constitue le cadre de vie limite autorisé par les interventions. En proposant un découpage associé à la morphologie et au fonctions urbaines, il est *a priori* possible de délimiter des zones d'activités autorisées susceptibles de fournir l'ensemble des services urbains essentiels (*e.g.* parcs, commerce de proximité).

La zone d'étude choisie est le centre-ville de Nice, qui comprend 39 000 habitants, et qui voit venir chaque semaine près de 58 000 personnes uniques pour travailler ou faire des activités dans cette zone. De plus, plus de 60% des habitants de cette zone urbaine travaillent à l'extérieur.

Pour chaque simulation, 50 réplications ont été exécutées sur le super-ordinateur GAIA d'EDF R&D. Les résultats sur l'évolution du nombre de personnes susceptibles, pré-symptomatiques, asymptomatiques, symptomatiques, hospitalisées, en réanimations et mortes sont respectivement montrées Figure 2.a (pas de confinement), Figure 2.b (confinement réaliste) et Figure 2.c (confinement spatiale). Outre les valeurs de quartiles (ligne) encadrées par les valeurs min/max (couleur) sur chaque indicateurs pour chaque jour simulé, les figures montrent le temps d'application des politiques d'intervention (barre grisée) entre le 17 mars et le 11 mai 2020.

On observe sur les figures b et c des politiques de confinement une diminution relative, comparativement au scénario de base, du nombre de personnes nécessitant une hospitalisation, d'entrée en service de réanimation et de morts, ce qui va dans le sens de l'efficacité des politiques de confinement. La diminution s'explique en grande partie par la présence de trajectoire de simulation menant à une quasi absence d'épidémie dans le cas du confinement spatialisé (Q1 avec moins de 5 hospitalisations en sois intensifs et 1 mort) et une épidémie très faible dans le cas du confinement réaliste (Q1 avec moins de 20 hospitalisation en sois intensifs et 6 morts), alors que toutes les simulations dans le cas sans intervention mène à un nombre similaire de mort et d'occupation des soins intensifs.

Cette différence d'amplitude est d'autant plus marquée dans le cas du confinement "spatialisé" (la médiane du nombre de morts est 2 fois moins importante par rapport au confinement réaliste) qu'elle s'accompagne d'un délai plus important dans le déclenchement du pic épidémique. Sur ce point, on peut constater que le scénario sans intervention exhibe également un décalage de plusieurs mois avant d'observer une augmentation de cas significative : l'explication est à chercher du côté du choix fait de ne pas introduire à l'initialisation un cas infecté arbitrairement, mais de s'appuyer sur les données de cas d'infection avérés dans la région ou le département. Ainsi, cette information "contient" en quelque sorte les effets du confinement, puisque la courbe de cas, et donc la probabilité de voir un agent infecté en dehors de la zone y entrer, reste particulièrement basse dans les premiers mois simulés.

L'autre point important qui différencie le confinement spatial et détermine son relatif succès par rapport aux deux autres scénarios, est sa capacité à bloquer le mécanisme d'introduction exogène d'agents infectieux en interdisant durant l'intervention l'entrée d'agents externes. Dans le cas de la politique réaliste, un quart des agents externes, étant associé à des métiers essentiels, sont encore susceptibles d'infecter des habitants de la zone. La mise en place de politiques d'intervention cibler sur l'introduction exogène de nouveau cas (*e.g.* tracer et isoler), loin de contenir entièrement l'épidémie, permet néanmoins d'envisager d'infléchir les courbes de contamination : dans

un cas sur quatre, le rebond épidémique n'intervient qu'à la fin des simulations, *i.e.* plus de 4 mois après la fin de la politique de restriction des activités.

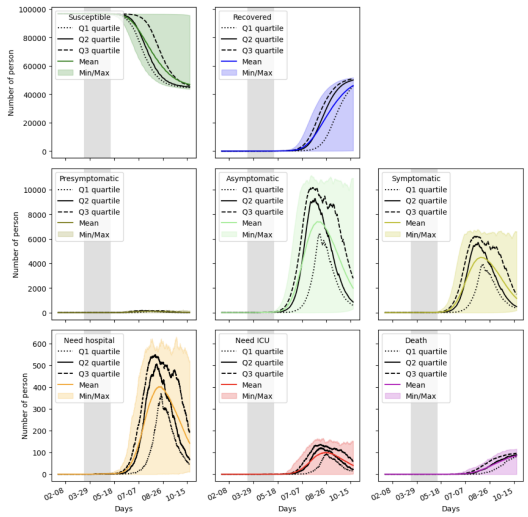
Néanmoins, on observe qu'aucune des interventions modélisées ne permet d'éviter un rebond épidémique, souvent au coeur des débats lors du déconfinement. Cela confirme la nécessité d'introduire dans notre modèle des interventions pharmaceutiques, notamment l'introduction d'une politique de vaccination, afin d'étudier les possibilités de sortie de crise.

4 Conclusion

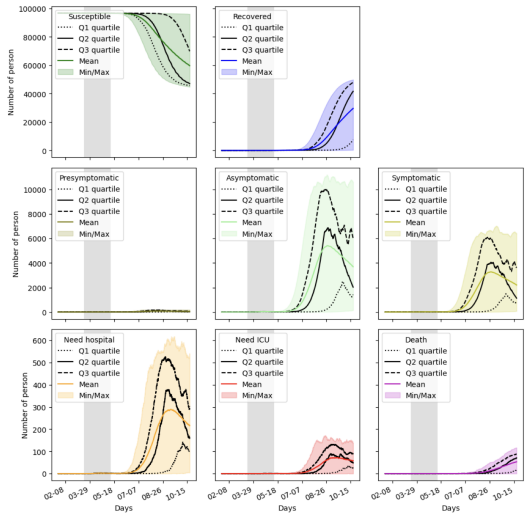
Nous avons présenté dans cet article l'environnement de modélisation COMOKIT, qui permet de tester l'impact de politiques d'intervention locales face à la pandémie de COVID-19. Une application au centre-ville de Nice, qui permet d'illustrer la versatilité de COMOKIT et le type d'adaptation pouvant être réalisée, a également été présentée. COMOKIT est aujourd'hui utilisé pour des projets de modélisation dans des contextes radicalement différents, ce qui démontre l'intérêt de son architecture et des choix de conception initiaux qui ont été réalisés ; néanmoins, des travaux sont toujours en cours afin d'améliorer sa généralité.

Par exemple, l'une des hypothèses sous-jacente à COMOKIT est que les infections se produisent en des lieux donnés (foyer, entreprises, écoles, etc.) mais pas pendant le déplacement entre ces lieux. Cette hypothèse, valide pour des territoires où la majorité des déplacements sont effectués dans des véhicules individuels (e.g. Vietnam), pose néanmoins question pour des terrains d'étude sans ce type de restrictions ou habitudes. Une extension future de COMOKIT sera de prendre en compte les transports collectifs, de même qu'autoriser la modulation du pas de temps horaire aujourd'hui appliqué par défaut, afin de représenter de façon plus fine des contextes possibles d'infection.

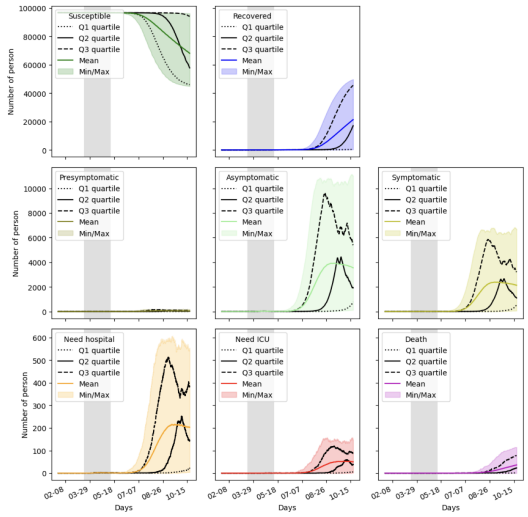
Enfin, l'une des contributions de COMOKIT a été de montrer à quel point l'approche de modélisation à base d'agents était adaptée à la représentation de phénomènes socio-environnementaux complexes. Les extensions prévues, ainsi que celles déjà réalisées par les autres utilisateurs de COMOKIT, en mettant en valeur la versatilité de cette approche, permettent en effet d'enrichir le domaine plus général de la simulation sociale, de plus en plus sollicitée pour répondre aux questionnements des décideurs et aux interrogations



(a) Politique : "pas de confinement"



(b) Politique : "confinement réaliste"



(c) Politique : "confinement spatial"

FIGURE 2 – Résultats des simulations pour les 3 politiques

des différentes parties prenantes dans les crises sociétales ou environnementales que l'humanité traverse (adaptation aux changements environnementaux, résilience face aux catastrophes...).

Remerciements

Le travail sur le noyau de COMOKIT est financé par l'ANRS (projet COV15, AAP COVID19-Sud Flash 2020). Les expérimentations à large échelle ont été rendues possibles par l'établissement d'une convention entre EDF et l'IRD pour l'utilisation du super-calculateur GAIA [6].

Références

[1] How epidemiology has shaped the COVID pandemic. *Nature*, 589(7843) :491–492, Jan. 2021.

[2] D. W. Berger, K. F. Herkenhoff, and S. Mongey. An seir infectious disease model with testing and conditional quarantine. Technical report, National Bureau of Economic Research, 2020.

[3] S. L. Chang, N. Harding, C. Zachreson, O. M. Cliff, and M. Prokopenko. Modelling transmission and control of the covid-19 pandemic in australia. *Nature communications*, 11(1) :1–13, 2020.

[4] K. Chapuis, P. Taillandier, B. Gaudou, F. Amblard, and S. Thiriot. Gen* : an integrated tool for realistic agent population synthesis. In *Social Simulation Conference*, 2019.

[5] K. Chapuis, P. Taillandier, M. Renaud, and A. Drogoul. Gen* : a generic toolkit to generate spatially explicit synthetic populations. *Int. Journal of Geographical Information Science*, 32(6) :1194–1210, 2018.

[6] A. Chin, J. Chu, M. Perera, K. Hui, H.-L. Yen, M. Chan, M. Peiris, and L. Poon. Stability of sars-cov-2 in different environmental conditions. *MedRxiv*, 2020.

[7] C. L. Correa-Martínez, S. Kampmeier, P. Kümpers, V. Schwierzeck, M. Hennies, W. Hafezi, J. Kühn, H. Pavenstädt, S. Ludwig, and A. Mellmann. A pandemic in times of global tourism : Superspreading and exportation of covid-19 cases from a ski area in austria. *Journal of clinical microbiology*, 2020.

[8] A. Drogoul, P. Taillandier, B. Gaudou, M. Choisy, K. Chapuis, Q. N. Huynh, N. D. Nguyen, D. Philippon, A. Brugière, and P. Larmande. Designing social simulation to (seriously) support decision-making : COMOKIT, an agent-based modelling toolkit to analyse and compare the impacts of public health interventions against COVID-19. *Review of Artificial Societies and Social Simulation*, 2020.

[9] N. Ferguson. Capturing human behaviour. *Nature*, 446(7137) :733–733, 2007.

[10] B. Gaudou, Q. N. Huynh, D. Philippon, A. Brugière, K. Chapuis, P. Taillandier, P. Larmande, and A. Drogoul. 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. *Frontiers in Public Health*, 8 :563247–563247, 2020.

[11] J.-F. Girres and G. Touya. Quality assessment of the french openstreetmap dataset. *Transactions in GIS*, 14(4) :435–459, 2010.

[12] N. Hoertel, M. Blachier, C. Blanco, M. Olfson, M. Massetti, M. S. Rico, F. Limosin, and H. Leleu. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine*, 26(9) :1417–1421, Sept. 2020.

[13] S. W. X. Ong, Y. K. Tan, P. Y. Chia, T. H. Lee, O. T. Ng, M. S. Y. Wong, and K. Marimuthu. Air, surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (sars-cov-2) from a symptomatic patient. *JAMA*, 323(16) :1610–1612, 2020.

[14] P. Taillandier, B. Gaudou, A. Grignard, Q.-N. Huynh, N. Marilleau, P. Caillou, D. Philippon, and A. Drogoul. Building, composing and experimenting complex spatial models with the gama platform. *GeoInformatica*, 23(2) :299–322, 2019.

[15] Q. Wang, N. Shi, J. Huang, T. Cui, L. Yang, J. Ai, H. Ji, K. Xu, T. Ahmad, C. Bao, et al. Effectiveness and cost-effectiveness of public health measures to control covid-19 : a modelling study. *medRxiv*, 2020.

6. https ://www.edf.fr/groupe-edf/qui-sommes-nous/activites/recherche-et-developpement/toutes-les-actualites-de-la-r-d/la-r-d-d-edf-et-des-equipes-de-l-ird-s-allient-dans-la-lutte-contre-la-covid-19