

Master de deuxième année

GESTION INTEGREE DES MALADIES ANIMALES TROPICALES

Rapport de stage

Modélisation épidémiologique basée sur la théorie du Chaos de la COVID-19 sur le continent africain

Présenté par Nathan THENON

Réalisé de janvier à juin 2021 en France

Sous la direction de Sylvain MANGIAROTTI Chargé de Recherche à l'IRD – UMR CESBIO à Toulouse, France Marisa PEYRE Directrice de Recherche au CIRAD – UMR ASTRE à Montpellier, France

Soutenu le 15 juin 2021

Remerciements

Je tiens tout d'abord à remercier Sylvain MANGIAROTTI, pour son accueil au sein du CESBIO, sa disponibilité, son soutien et sa pédagogie tout au long de mon stage.

Je remercie également Marisa PEYRE, pour son partage d'informations et d'expérience ainsi que ses conseils dans la structuration de mon travail.

J'aimerais aussi remercier Mireille HUC, pour l'accompagnement permanent et l'assistance indispensable qu'elle m'a offert durant ces 6 mois, notamment dans le volet informatique de mon travail.

J'adresse ensuite mes remerciements à Capucine SICARD, en stage sur un sujet complémentaire au mien, avec qui j'ai pu avoir des échanges utiles et enrichissants. De plus, un grand merci pour sa participation dans l'élaboration et la conduite des entretiens auprès d'acteurs locaux en Afrique.

Je tiens particulièrement à remercier MUSE (Montpellier Université d'Excellence) pour avoir financé ce travail grâce au fond d'urgence pour la Recherche sur la COVID-19.

Enfin, je souhaite remercier toutes les personnes qui ont participé, de près ou de loin, à la réalisation de ce stage dans de bonnes conditions, malgré la situation parfois complexe que nous vivons tous.

Table des matières

LISTE DES ABREVIATIONS
INTRODUCTION
PARTIE I : SYNTHESE BIBLIOGRAPHIQUE 10 La COVID-19 : une pandémie prévisible 10 Craintes et réalités de la pandémie sur le continent africain 11 Une évolution complexe de la COVID-19 en Afrique 13
La COVID-19 : une pandémie prévisible
Craintes et réalités de la pandémie sur le continent africain
Une évolution complexe de la COVID-19 en Afrique
La modélisation : un outil pour comprendre les dynamiques épidémiologiques
La modélisation globale : une méthode particulière pour une situation particulière
PARTIE II : TRAVAIL PERSONNEL 19
Objectifs spécifiques de l'étude 19
Matériel et méthodes 20
Collecte des données
Analyse comparative des données épidémiologiques entre pays
Catégorisation et sélection des pays d'étude
Pré-traitement des données des pays d'étude
Recherche de modèles globaux
<i>Etude approfondie de l'espace des phases et de la prédictibilité des modèles</i>
Sélection des modèles et étude de robustesse
Etude de non-stationnarité d'un modèle
Kesultats
Importance des données dans l'analyse comparative de l'impact de la COVID-19 entre les différents pays
au monae
Calegorisation des pays ajricains et selection des pays à etude
Discussion
CONCLUSION 37
RIRLIOGRAPHIE 38
ANNEAE5 40
Annexe 1 : Liste des pays utilisés pour les analyses de pondération
Annexe 2 : Boites à moustaches de la dispersion des facteurs de pondération et des données
Annexe 3 : Nuages de points des cas et décès totaux en fonction des pondérations mises en place, pour une sélection de pays africains et européens
Annexe 4 : Valeurs du nombre de lits d'hôpitaux disponibles par habitant dans les pays africains et ratio d'écart avec la France
Annexe 5 : Sélection de courbes de cas journaliers de la COVID-19 en Afrique
Annexe 6 : Description des modifications ponctuelles apportées aux données épidémiologiques au début de leur pré- traitement et des séries creuses repérées
Annexe 7 : Courbes des cas et décès lissées suite à la préparation des données
Annexe 8 : Liste pour chaque pays des meilleurs modèles trouvés et de leur erreur de prévision ramenée au nombre de cas
(ou de décès) total
Annexe 9 : Description des meilleurs modèles selon leur erreur de prévision
Annexe 10. Fourant de phase et predictionne de 14 modeles selectionnes
Annexe 12 : Variation des prévisions du modèle des cas en Ethiopie selon les valeurs attribuées à ses coefficients
Annexe 13 : Evolution de l'indice de rigueur des mesures gouvernementales en Ethiopie
RESUME
ABSTRACT

Table des illustrations

Figure 1. Nombre de cas et de décès liés à la COVID-19 pondérés par million d'habitants de plus de 60
ans et par lits d'hôpitaux disponibles par million d'habitants, pour une sélection de pays européens et
africains
Figure 2. Catégorisation des pays africains selon la date d'apparition et la durée de leurs différentes
vagues épidémiques entre le début de la pandémie et le 25/04/2021
Figure 3. Séries temporelles et portrait de phase dans sa projection (C1,C2) observé et modélisé par le
modèle à 6 paramètres et par le modèle à 10 paramètres, pour les cas déclarés de COVID-19 en Ethiopie
entre le 22/01/2020 et le 15/03/2021
Figure 4. Evolution de l'erreur absolue $ e(\tau) $ de prévision du nombre journalier de cas de la COVID-19
en Ethiopie estimée par le modèle à 6 paramètres et par le modèle à 10 paramètres, pour un horizon de
prédiction de 0 à 20 jours et pour 200 conditions initiales également réparties sur la période courant du
22 janvier 2020 au 25 février 2021
Figure 5. Erreur absolue de prévision au 15/03/2021, des modèles sélectionnés pour un niveau de
confiance de 95 et 75% et à un horizon de 10 et 20 jours
Figure 6. Etude de non-stationnarité basée sur l'analyse des variations des coefficients des modèles au
cours du temps

Liste des abréviations

CESBIO	Centre d'Etudes Spatiales de la Biosphère
COVID-19	Coronavirus Disease 2019 (Maladie à coronavirus 2019)
INP	Intervention Non-Pharmaceutique
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (Plateforme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques)
JHU	Johns Hopkins University (Université de Johns Hopkins)
MERS-CoV	Middle East Respiratory Syndrome-related Coronavirus (Coronavirus du syndrome respiratoire du Moyen-Orient)
MIE	Maladie Infectieuse Emergente
OMS	Organisation Mondiale de la Santé
SARS-CoV(-2)	Severe Acute Respiratory Syndrome Coronavirus (2) (Coronavirus (2) du syndrome respiratoire aigu sévère)
Sida	Syndrome d'immunodéficience acquise
SRAS	Syndrome Respiratoire Aigu Sévère
VIH	Virus de l'Immunodéficience Humaine

Introduction

L'émergence ou la réémergence de maladies infectieuses est un risque important de nos jours pour les communautés humaines. En effet, les changements environnementaux d'origines anthropiques, comme les mouvements de populations accrus, la modification de l'utilisation des sols et les changements climatiques, sont autant de facteurs induisant une accélération de l'apparition et de la diffusion de nouveaux agents pathogènes. Cet accroissement de la fréquence d'apparition de nouvelles maladies infectieuses émergentes (MIE) au cours du dernier siècle s'accompagne d'une augmentation du risque de pandémies ou de panzooties, aux conséquences potentiellement lourdes, autant en termes sanitaires, sociaux, économiques qu'environnementaux (El-Sayed & Kamel 2020 ; IPBES 2020).

Pour lutter contre les maladies infectieuses et ainsi limiter leurs impacts délétères, il est nécessaire de comprendre leurs dynamiques et d'identifier les facteurs clés de leurs transmissions. Pour cela, la modélisation s'avère être un outil adapté et polyvalent. Selon les méthodes utilisées, les modèles épidémiologiques peuvent avoir pour objectif de représenter, évaluer, analyser et/ou prévoir des dynamiques de transmission (Dubé et al. 2007 ; Djidjou-Demasse et al. 2020). Cependant, dans le cas de MIE, le manque d'information sur la maladie et ses caractéristiques sont souvent problématiques pour obtenir des modélisations utilisables, puisque la plupart de celles-ci nécessitent de fournir un nombre conséquent de données et de connaître la valeur de certains paramètres épidémiologiques, qui sont souvent inaccessibles ou mal connus. Dans ce contexte de données limitées, moins le nombre d'hypothèses sur ces paramètres est élevé, moins les modèles prédictifs obtenus auront de risques d'être erronés (Lloyd-Smith et al. 2015 ; Sridhar & Majumder 2020). De ce fait, les modèles dits « globaux », s'appuyant sur la théorie du Chaos et qui n'émettent pas d'hypothèses fortes sur la maladie étudiée, pourraient être dans ces situations, une alternative intéressante aux modèles plus classiques (Letellier et al. 2009). Au vu de ces constats, l'objectif de ce travail est le développement de modèles épidémiologiques s'appuyant sur la théorie du Chaos et permettant des prévisions à court terme de l'évolution de la COVID-19 pour une sélection de pays africains. Plus globalement, ce travail permet d'explorer le potentiel de cette technique de modélisation dite « globale » afin d'informer la surveillance précoce des MIE.

Partie I : Synthèse Bibliographique

La COVID-19 : une pandémie prévisible

Les MIE, que l'on peut définir comme des maladies dont l'incidence croît rapidement et pour lesquelles l'agent pathogène est nouveau, issu de l'évolution d'un agent pathogène connu, ou nouvellement introduit dans une région jusqu'alors indemne, semblent apparaitre de plus en plus fréquemment ces dernières décennies. Le phénomène est similaire pour les maladies ré-émergentes, c'est-à-dire celles qui avaient disparu ou avaient été oubliées mais qui réapparaissent, voire qui sont déclarées dans de nouvelles zones géographiques (El-Sayed & Kamel 2020). Or, la plupart des pandémies sont provoquées par ces MIE, et plus particulièrement par des zoonoses virales dont les réservoirs sont les mammifères. Le Sida, les grippes, ainsi que le SRAS en sont quelques exemples récents (Morse et al. 2012). Ainsi, d'après les experts de la plateforme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques (IPBES), le monde rentre dans une « ère des pandémies » (IPBES 2020).

Les causes de cette augmentation en fréquence des MIE et par conséquent des pandémies sont multiples mais toutes liées à l'activité anthropique et aux perturbations écologiques dont elle est responsable (El-Sayed & Kamel 2020; IPBES 2020). L'environnement et ses évolutions à travers les changements climatiques (modifications à long terme des propriétés physiques moyennes de l'atmosphère, température, humidité, etc...) ainsi que le changement d'utilisation des terres pour une agriculture intensive, ont de nombreux impacts. Cela provoque la migration ou la modification du comportement de nombreuses espèces animales, voire de populations humaines, un appauvrissement de la biodiversité, ainsi qu'une accentuation des contacts entre faune sauvage, animaux domestiques et humains. Tout cela représente autant de facteurs favorisant la dynamique actuelle des MIE. D'autres facteurs plus anthropogéniques, tels que la mondialisation, l'intensification des transports de biens, d'animaux et d'humains, ou le commerce et la consommation d'espèces sauvages, favorisent eux aussi l'apparition de maladies et la diffusion d'agents pathogènes.

Ces phénomènes font peser un risque pour toutes les régions de la planète, pour plusieurs raisons. Tout d'abord, l'émergence ou la réémergence de maladies peut se faire sur tous les continents, que ce soit dans les régions chaudes, avec notamment les maladies tropicales négligées qui bénéficient de peu de suivi et contre lesquelles la lutte est limitée, ou dans les régions plus froides, avec la fonte des glaces et du pergélisol, pouvant entrainer la libération d'agents pathogènes auparavant piégés en leur sein. De plus, l'hyperconnectivité entre toutes les régions du monde et les flux croissants de réfugiés climatiques induisent un brassage des populations et une diffusion accélérée des agents pathogènes (El-Sayed & Kamel 2020).

Le SARS-CoV-2, un β -coronavirus de la famille des Coronaviridae, est responsable de la COVID-19, une MIE hautement contagieuse pouvant se manifester par de graves pneumonies. Ce virus a été isolé pour la première fois en décembre 2019 dans la province du Hubei, en Chine. En quelques mois, il s'est propagé sur tous les continents, entrainant une pandémie aux dimensions et aux répercussions très importantes (Morse et al. 2012 ; Lone & Ahmad 2020). Du fait de sa proximité phylogénétique avec d'autres β -coronavirus comme le SARS-CoV et le MERS-CoV (eux aussi responsables d'épidémies de pneumonies respectivement en 2002-2004 et en 2012), le SARS-CoV-2 semble être une zoonose probablement transmise par des mammifères sauvages à des mammifères domestiques puis à l'Homme. Actuellement, comme pour le SARS-CoV, le SARS-CoV-2 a réussi à se transmettre d'humains à humains sans avoir à repasser par des réservoirs animaux.

La COVID-19, transmise essentiellement par voie respiratoire avec l'aérosolisation de gouttelettes infectantes, provoque la plupart du temps de la fièvre, des symptômes respiratoires liés à une pneumonie (caractérisés par une douleur thoracique, un essoufflement et de la toux), ainsi que parfois des symptômes digestifs comme des nausées, des vomissements et des douleurs abdominales. Dans un quart des cas, la maladie a tendance à s'aggraver provoquant une pneumonie sévère pouvant entrainer la mort (Li et al. 2021). D'après des revues systématiques de la littérature scientifique et des méta-analyses, la période d'incubation movenne pour cette maladie est autour de 6 jours (intervalle de confiance à 95% de 5,79 - 6,97), avec des amplitudes maximales allant de 2 à 18 jours (Elias et al. 2021). La contagiosité semble commencer moins de 3 jours avant l'apparition de symptômes, et continuer tout le long de la phase clinique pendant près de 20 jours ou jusqu'à la mort de la personne infectée (Xi He et al. 2020). Le nombre de reproduction de base (R0) de cette maladie, sans intervention de contrôle, est quant à lui proche de 3,14 (intervalle de confiance à 95% de 2,69 - 3,59), montrant le caractère hautement contagieux de la COVID-19 (Hussein et al. 2021). Les formes graves de la maladie sont d'autant plus fréquentes que les personnes infectées sont âgées et présentent des facteurs de comorbidité tels que le diabète, l'obésité, l'hypertension ou un état immunodéprimé (Li et al. 2021). Par conséquent, bien que le taux de létalité moyen soit proche de 6%, il existe une forte hétérogénéité selon l'âge, avec un taux de létalité autour de 0,3% pour les moins de 30 ans, 1,1% pour les 40-50 ans, puis augmentant de manière exponentielle et atteignant jusqu'à 29,6% pour les plus de 80 ans (Bonanad et al. 2020). Le R0 et la létalité varient considérablement d'un pays à un autre, en fonction par exemple de la pyramide des âges, de la prévalence des facteurs de comorbidité dans la population, ou des interventions nonpharmaceutiques (INP) comme le confinement ou les mesures de distanciation, menées pour lutter contre le virus (Bonanad et al. 2020 ; Li et al. 2021).

Craintes et réalités de la pandémie sur le continent africain

Cette pandémie, comme d'autres épidémies avant elle, a révélé le manque de préparation de la plupart des états face à cette menace croissante pour nos sociétés que représentent les MIE. Cette vulnérabilité peut se mesurer au travers des conséquences sanitaires (par le nombre de morts causés par la maladie et l'engorgement des systèmes de santé) et socio-économiques (provoquées notamment par les INP) (David & Dévédec 2019 ; IPBES 2020). La majorité des pays occidentaux, dont les systèmes de santé, les économise et les gouvernements sont relativement robustes, ont payé un lourd tribut humain et socio-économique avec cette crise sanitaire. Des craintes fondées se sont donc manifestées concernant les conséquences potentielles pour les pays en voie de développement ayant moins de moyens pour lutter contre la COVID-19, notamment en Afrique où se trouvent 22 des 25 pays du monde les plus à risque d'après l'indice de vulnérabilité aux maladies infectieuses de 2016 (Lone & Ahmad 2020). Des prévisions alarmantes de l'OMS à la mi-avril 2020 prédisaient ainsi entre 29 et 44 millions de cas et jusqu'à 190 000 morts sur le continent africain en une année si les INP s'avéraient inefficaces sur place (Aljazeera News 2020). De même, la Commission Economique des Nations Unies pour l'Afrique a, pour sa part, prédit jusqu'à 300 000 morts causées par la pandémie dans le meilleur des scénarios et plus de 3 millions dans le pire (Ogunleye et al. 2020).

La vulnérabilité de l'Afrique à la COVID-19 s'explique d'abord d'un point de vue sanitaire. De nombreux pays du continent n'ont pas une infrastructure de santé, du personnel et des équipements (de protection, de dépistage, de soin) suffisants pour absorber des épidémies similaires à celles qui ont frappé les pays Occidentaux (Lone & Ahmad 2020). Par exemple, le nombre moyen de lits d'hôpitaux en Afrique est de 1,8 pour 1 000 personnes, soit moins du tiers qu'en France. En termes d'hygiène, plus d'un tiers de la population africaine n'a pas accès aux installations sanitaires de base, ou a un accès limité à l'eau courante, empêchant par exemple un nettoyage des mains régulier (Ogunleye et al. 2020).

Dès lors, une transmission oro-fécale du SARS-CoV-2, en plus de la transmission par voie respiratoire, est une possibilité supplémentaire de contamination à craindre pour certaines populations du continent (Gwenzi 2021). Enfin, de nombreuses maladies infectieuses sont endémiques en Afrique, notamment le Sida avec près 26 millions des personnes infectées d'après l'OMS, la tuberculose avec 2,5 millions de cas, le paludisme avec 213 millions de malades ou encore l'hépatite B ou C avec 71 millions de personnes atteintes (Lone & Ahmad 2020). Ajouté à cela un nombre conséquent de personnes anémiées ou malnutries, en plus de facteurs de comorbidité comme les maladies cardiovasculaires et le diabète, cela amène à penser qu'une proportion importante de la population africaine est à risque en cas d'infection au SARS-CoV-2 (Bell et al. 2020).

De manière indirecte, mais toujours d'un point de vue sanitaire, les INP mises en place pour lutter contre la propagation du virus et la saturation des milieux hospitaliers pourraient avoir des conséquences sur les personnes atteintes d'autres maladies. Par exemple, si les systèmes de santé se concentrent uniquement sur la lutte contre la COVID-19, il pourrait y avoir une réduction du nombre de diagnostics et de traitements réalisés pour le VIH et la tuberculose, voire une interruption des campagnes de lutte contre le paludisme. Le résultat pourrait être une augmentation jusqu'à 36% du nombre de décès pour ces 3 maladies, et ce sur plusieurs années. Ce phénomène a déjà été observé, notamment en 2014 durant l'épidémie d'Ebola en Guinée avec une hausse significative de morts liées au paludisme (Hogan et al. 2020). Dans certains pays, comme l'Ouganda où la population est très jeune, l'impact de la COVID-19 pourrait même être moins important sur le long terme que celui du VIH et du paludisme si des mesures strictes de confinement, à l'image de ce qui a été fait en Europe ou en Chine, étaient appliquées (Bell et al. 2020).

Enfin, d'un point de vue socio-économique, à cause de la dépendance des pays africains au marché international, certains experts ont calculé qu'environ 20 millions d'emplois pourraient être supprimés en Afrique suite à la pandémie, et la Commission économique des Nations Unies pour l'Afrique a elle estimé que la COVID-19 pourrait probablement pousser 29 millions d'africains dans l'extrême pauvreté, de par le ralentissement de l'activité économique mondiale (African Union 2020 ; Ogunleye et al. 2020). L'augmentation du chômage et de la précarité sur le continent provoquerait alors une augmentation de la malnutrition et de l'insécurité alimentaire pouvant générer à son tour une instabilité politique dans les régions les plus pauvres, notamment en Afrique sub-saharienne (African Union 2020 ; Bell et al. 2020). Ainsi, il est difficile de prévoir les conséquences sanitaires ou socio-économiques de la COVID-19 en Afrique, mais il est crucial que la réponse des états africains soit adaptée aux contextes locaux, afin d'arriver aux scénarios les moins néfastes (Bell et al. 2020 ; Hogan et al. 2020).

En pratique, le continent africain a été le dernier touché par la COVID-19. Les premiers cas ont été confirmés en Egypte le 14 février 2020 puis en Algérie le 25 février et au Nigéria le 27 du même mois. 46 autres pays ont ensuite déclaré des cas à partir du mois de mars et 5 autres entre avril et mai 2020 (Dong et al. 2020). La plupart des sources d'introduction du virus sont des voyageurs par avion en provenance d'Europe et des Etats-Unis, la majorité des compagnies aériennes africaines ayant suspendu leurs vols en provenance de la Chine de manière préventive (Lone & Ahmad 2020 ; Nkengasong & Mankoula 2020). Les premiers mois, ce furent les pays avec le plus de vols intercontinentaux qui eurent le plus grand nombre de cas en Afrique (Bamgboye et al. 2020). Cependant, contrairement à la plupart des autres pays où le virus circulait, les dynamiques de transmission de la maladie en Afrique semblent avoir été plus lentes qu'ailleurs, avec une absence de croissance exponentielle des cas, ainsi qu'un faible taux de mortalité. Contrairement aux projections envisagées notamment par l'OMS ou la Commission Economique des Nations Unies pour l'Afrique, le nombre de cas et de morts déclarés se sont avérés inférieurs aux prévisions. Au 14 février 2021, soit un an après le premier cas confirmé en Afrique, 3,75 millions d'africains ont été diagnostiqués positifs et 98 500 en sont morts (Dong et al. 2020).

Malgré le fait que 17,2% des humains vivent en Afrique, ils ne représentent que 3,4% du total des cas et 4,1% de celui des morts de la COVID-19. Ce continent semble donc être celui le moins touché à ce jour par la pandémie (Bamgboye et al. 2020 ; Lone & Ahmad 2020). Il semble aussi le moins meurtri par la COVID-19, avec une létalité du virus inférieure de 4 fois à celle observée en Europe et en Amérique du Nord, et inférieure de 2 fois à celle observée en Asie et en Amérique latine (Mougeni et al. 2020). Enfin, comme ailleurs dans le monde, les pays africains ne sont pas touchés de la même manière ou au même moment par le SARS-CoV-2, et l'Afrique du Sud, l'Egypte, l'Algérie, le Nigéria et le Ghana représentent à eux cinq les trois quarts des cas et morts signalés de la COVID-19 sur le continent (Bamgboye et al. 2020). Nous reviendrons sur ces constats et les questions qui en découlent durant nos analyses.

Une évolution complexe de la COVID-19 en Afrique

De nombreuses pistes ont été envisagées pour expliquer la dynamique actuelle de la COVID-19 en Afrique. Tout d'abord, des critiques ont été émises quant à la capacité de tester les personnes. En effet, l'Afrique ayant moins de moyens à sa disposition, l'indisponibilité, le coût et la mise en place des tests peuvent être des freins à la détection satisfaisante des cas pour certains pays. Ainsi, la proportion de cas non détectés pourrait représenter une part plus importante sur ce continent que pour d'autres régions du monde (Lone & Ahmad 2020). Une certaine résistance à se faire dépister a aussi pu être observée (notamment au Kenya), liée à la peur de se voir placer en quarantaine, ce qui priverait les personnes de travail et de revenus (Rutayisire et al. 2020). Enfin, le nombre de cas pourrait être largement sous-estimé du fait de symptômes bénins, similaires à d'autres maladies, ou même de formes asymptomatiques chez les jeunes, qui sont proportionnellement plus nombreux en Afrique (Pearson et al. 2020). Concernant le décompte des morts de la COVID-19, seuls 8 pays (dont les plus grands sont l'Afrique du Sud, l'Algérie, l'Egypte et la Tunisie) ont des systèmes d'état civil fonctionnels et complets, alors que plus de la moitié des pays d'Afrique sub-saharienne ne conservent que des registres de décès manuscrits (Mbow et al. 2020 ; Shveda & Cardoso 2021). De plus, la pandémie a « paralysé toutes les activités d'état civil » dans un certain nombre de pays où cela était considéré comme non-essentiel, d'après un rapport de l'ONU. Cela doit sûrement entrainer, si ce n'est une sous-déclaration des morts causés par le virus, du moins un retard du décompte de celles-ci. Les rares pays ayant des systèmes d'état civil fonctionnels sont d'ailleurs ceux enregistrant le plus grand nombre de morts dus à la COVID-19, rapporté au nombre d'habitants. De plus, ces registres permettent de constater des excès de mortalité totale non attribués à des personnes diagnostiquées comme positives. Par exemple, en Afrique du Sud entre février 2020 et 2021, 46 200 personnes sont officiellement mortes de la COVID-19. Cependant, la mortalité totale montre un excès de mortalité pour cette année-là de 91 500 personnes supplémentaires à celles officiellement diagnostiqués, qui sont potentiellement des victimes non diagnostiquées de la COVID-19, ou dont le décès est indirectement lié à la crise sanitaire (retard dans le diagnostic ou le traitement d'une maladie, peur de se rendre en milieu hospitalier, etc...). Le phénomène est le même en Egypte (mais aussi dans le reste du monde comme en Europe, dans une moins grande proportion), avec seulement 10% de l'excès de mortalité expliqué par un test positif au SARS-CoV-2 (Shveda & Cardoso 2021).

D'autres observations permettent d'être plus confiants dans les données enregistrées en Afrique. En premier lieu, les dynamiques de transmission observées semblent rester représentatives des dynamiques réelles. Une étude a montré qu'il y a au cours du temps une relation linéaire entre le nombre de cas et de morts attribués à la COVID-19. Ainsi, quoique les chiffres soient certainement sous-évalués, la dynamique observée de la pandémie reste quant à elle correcte (Adekunle et al. 2020). Il est aussi important de souligner que la capacité de la majorité des pays africains à tester leurs habitants s'est nettement améliorée depuis avril 2020, grâce à l'envoi de kits de tests et à la formation du personnel sur place par l'OMS et l'African Center for Disease Control and Prevention (Ogunleye et al. 2020). Cela s'est d'ailleurs traduit par une augmentation de 43% des cas détectés sur le continent suite à ces envois (Gaye et al. 2020). Il faut ensuite remarquer qu'en début d'épidémie, de nombreux pays africains ont réalisé plus de tests que d'autres pays à phase épidémique similaire, c'est-à-dire pour un même nombre de cas détectés. Ainsi, bien que dotés de peu de kits de tests au début, une grande partie des pays africains ont mis en œuvre plus précocement des mesures de dépistage que la plupart du reste du monde (Mbow et al. 2020). Un autre indicateur permettant de conforter les chiffres peu élevés de cas et de morts de la COVID-19, notamment en Afrique sub-saharienne, est l'absence d'engorgement dans les hôpitaux de la plupart de ces pays, comme le souligne le Dr John Nkengasong, chef de l'African Center for Disease Control and Prevention (Nguimkeu & Tadadjeu 2021). Pour finir, bien que peu nombreuses, quelques enquêtes de séroprévalence permettent d'avoir un aperçu de la circulation du virus dans certaines régions d'Afrique. Une étude menée au Kenya entre avril et juin 2020 auprès des donneurs de sang a montré une séroprévalence proche de 4,3% (Uyoga et al. 2021). Une autre enquête, auprès du personnel soignant dans les zones urbaines du Malawi entre mai et juin 2020 a conclu à une séroprévalence de près de 12,3% (Chibwana et al. 2020). Une troisième étude, cette fois auprès de la population générale de Brazzaville en République démocratique du Congo entre avril et juillet 2020, montre une séroprévalence autour de 15% (dont 2% en avril et 22% en juillet) (Batchi-Bouyou et al. 2020). Enfin, une dernière enquête, menée auprès des soignants en pédiatrie d'Afrique du Sud entre mai et juillet 2020, affiche une séroprévalence de 7,66% (Goldblatt et al. 2021). Bien que ces séroprévalences soient élevées, montrant une sous-évaluation du nombre de cas de la COVID-19, elles sont pour la plupart issues de personnes habitant dans des milieux urbains, dans lesquels la contamination entre individus se fait plus facilement. De plus, pour la 2ème et 4ème étude, ce sont des personnels médicaux qui sont testés, or le risque accru de transmission nosocomiale est avéré, laissant penser que les personnels soignants sont proportionnellement plus atteints par la maladie que le reste de la population. Enfin, bien que la séroprévalence globale puisse monter jusqu'à 22% dans des foyers épidémiques, cela reste inférieur à des valeurs que l'on peut trouver en Amérique du Nord à la même période, allant jusqu'à 31,5% de séroprévalence (Batchi-Bouyou et al. 2020).

Quelques faits ont aussi été abordés pour potentiellement confirmer l'évolution plus lente de la maladie en Afrique par rapport aux autres continents. Contrairement à des idées reçues largement partagées en Occident, les pays africains ont pour la plupart mis en place précocement des mesures contraignantes pour limiter l'entrée ou la circulation du virus sur leur territoire. Le trafic aérien a été largement réduit, en particulier avec la Chine, de manière préventive dès février 2020 (Nkengasong & Mankoula 2020). D'autres INP ont été entreprises rapidement (limitation des mouvements, couvre-feux, fermeture d'établissements, etc...) de sorte qu'à la fin mars 2020, la quasi-totalité des pays africains avaient considérablement durci leur politique de lutte contre la COVID-19. Ces réponses, intervenant souvent alors que seuls quelques cas étaient détectés, voire même avant les premiers cas, étaient d'autant plus précoces en comparaison avec la lenteur de nombreux pays d'Europe ou d'Amérique, qui n'ont mis en place des INP strictes la plupart du temps qu'à partir du moment où le virus s'était déjà transmis à des centaines de personnes sur le territoire (Gaye et al. 2020). Cette rapidité d'action des pays africains s'explique en partie par l'expérience acquise dans la gestion d'autres épidémies telles que Ebola ou la fièvre de Lassa (Bamgboye et al. 2020; Mbow et al. 2020). Toutes ces INP en Afrique ont permis, à défaut d'empêcher la diffusion du virus, de la limiter et de la ralentir, laissant à ce continent un peu plus de temps pour s'organiser et tirer des leçons des gestions menées sur d'autres continents (Lone & Ahmad 2020). Cependant, les mesures d'endiguement les plus restrictives n'ont pas été pleinement respectées en Afrique, notamment par les personnes les plus pauvres travaillant sur les marchés, dépendant financièrement de leur travail au jour le jour. Aussi, beaucoup de pays ont dû assouplir progressivement les INP mises en place afin de limiter la précarisation de leurs citoyens, entrainant la plupart du temps une hausse des cas de COVID-19 détectés (Bamgboye et al. 2020 ; Mbow et al. 2020).

Un argument justifiant le faible nombre de morts de la COVID-19 observé en Afrique serait la démographie particulière de ce continent. En effet, il s'agit de la région du monde la plus jeune, avec un âge médian situé sous les 20 ans (à comparer par exemple aux 37 ans pour l'Europe). Ainsi, les personnes de plus de 65 ans, les plus durement affectées par le virus et ayant souvent des facteurs de comorbidité, ne représentent que 4% de la population africaine, en comparaison avec l'Europe ou l'Amérique du Nord où elles constituent près de 30% de la population totale, et 37% en Asie de l'Est et du Sud-Est (Bamgboye et al. 2020 ; Lone & Ahmad 2020). Une étude a d'ailleurs montré que la proportion de la population âgée de plus de 65 ans est significativement liée au nombre de cas et de morts observés sur ce contient (Nguimkeu & Tadadjeu 2021). Un autre facteur possible pourrait être le niveau sérique en vitamine D (Bamgboye et al. 2020). Certaines études semblent indiquer que l'exposition au soleil et une production optimale de vitamine D sont autant de facteurs pouvant peut-être limiter la transmissibilité, l'infectiosité et la gravité des signes de la COVID-19. Or, les populations africaines bénéficient en moyenne d'une exposition au soleil plus importante que pour d'autres continents, et les individus jeunes synthétisent mieux la vitamine D.

Sur ce continent, une autre singularité est la densité de population et l'urbanisation qui y sont globalement plus faibles qu'ailleurs. Or, la propagation de la maladie est plus lente et difficile en milieu rural qu'urbain, et c'est d'ailleurs dans les principales mégalopoles d'Afrique (comme Le Caire, Nairobi, Lagos et le Gauteng) que les plus gros foyers sont localisés (Bamgboye et al. 2020 ; Nguimkeu & Tadadjeu 2021). Enfin, en Afrique, les personnes âgées sont moins regroupées au sein d'établissements spécialisés que dans les pays Occidentaux, et vivent plutôt au domicile familial. Cela pourrait être un facteur de protection pour cette population, qui a été durement touchée en Europe et Amérique du Nord suite à la diffusion facilitée du virus dans ces établissements (Bamgboye et al. 2020). L'hypothèse qu'en Afrique des facteurs démographiques, géographiques et culturels particuliers entrainent une meilleure prédisposition de la population contre la transmission du virus est donc envisageable.

Les conditions climatiques en Afrique pourraient avoir un effet sur le ralentissement de la progression du virus. Un certain nombre d'études ont montré une corrélation entre la température, l'humidité, l'exposition aux rayons ultraviolets et la diffusion du SARS-CoV-2. Notamment, des conditions sèches et froides sont apparues idéales pour la transmission et la plus grande virulence du virus. La justification serait que ces conditions météorologiques impactent la survie du virus en milieu extérieur, mais aussi que les comportements des humains changeraient selon le climat (par exemple le fait de rester plus ou moins en milieu confiné) (Gwenzi 2021 ; Nguimkeu & Tadadjeu 2021). Ainsi, les zones tropicales semblent moins propices à la transmission de la COVID-19 que celles plus septentrionales et australes. C'est d'ailleurs dans les régions le plus au Nord de l'Afrique, comme le Maghreb et l'Egypte, tout comme en Afrique du Sud, que le plus grand nombre de cas est recensé (Dong et al. 2020; Ghosh et al. 2020). Cependant, la propagation plus élevée du virus dans d'autres régions tropicales du monde (Amérique latine et Asie), semble indiquer que le facteur météorologique est loin d'être prépondérant dans les conditions de diffusion du SARS-CoV-2 (Ghosh et al. 2020). De plus, ces études se basant sur des corrélations peuvent toujours être empreintes de biais divers, notamment liés à des caractéristiques communes à beaucoup de pays tropicaux (comme un taux de pauvreté plus important) contrairement à d'autres pays.

La question d'une immunité partielle préexistante contre ce virus en Afrique doit être également abordée. Les populations africaines étant globalement plus exposées à certains agents pathogènes, notamment des virus, leur réponse immunitaire pourrait être plus efficace face au SARS-CoV-2 (Mbow et al. 2020). En particulier, la piste d'une immunité croisée liée à une exposition à d'autres *coronavirus* zoonotiques, tels que les coronavirus humains *NL63* et 229*E*, a été étudiée récemment. Ces 2 espèces de *coronavirus*, bien que leurs prévalences soient mal définies sur le continent africain, sont bien présentes et circulent activement. Une étude sur la prévalence de la réactivité sérologique croisée d'échantillons de sang entre le SARS-CoV-2 et d'autres *coronavirus* zoonotiques (tels que *NL63* et 229*E*) a montré qu'elle était significativement plus importante en Afrique Sub-saharienne qu'aux Etats-Unis d'Amérique (Tso et al. 2021). Ainsi, l'exposition accrue à d'autres *coronavirus* en Afrique pourrait aboutir à une protection immunitaire renforcée, au moins partiellement. Dans cette même étude, les personnes atteintes du Sida présentaient aussi une réponse immunitaire croisée existante mais diminuée, probablement liée à l'immunodéficience provoquée par cette maladie. Cela vient modérer l'hypothèse d'une immunité collective plus développée en Afrique, du fait des quelques 26 millions des personnes atteintes par cette maladie sur le continent d'après l'OMS (Bell et al. 2020). Une autre piste avancée, justifiant une sensibilité au virus plus faible en Afrique serait l'effet secondaire bénéfique que produirait la vaccination antituberculeuse par le BCG (Bacille Calmette et Guérin) contre les formes graves de la COVID-19 (Ghosh et al. 2020). Il a été prouvé que ce vaccin limite les réactions cytokiniques, or la principale cause de détresse respiratoire liée à la COVID-19 et donc de décès, serait une production trop intense de cytokines (appelé choc cytokinique). Ainsi, les pays africains utilisant ce vaccin sur leur territoire, pourraient indirectement avoir moins de cas graves et de morts liés au virus grâce à cette vaccination.

Bien qu'un certain nombre d'arguments confortent la véracité d'une diffusion moins rapide du SARS-CoV-2 en Afrique, il ne faut pas pour autant considérer cela comme un fait établi et immuable (Bamgboye et al. 2020 ; Roser et al. 2020). Pour un certain nombre de raisons explorées précédemment, la situation de la COVID-19 sur ce continent peut se dégrader rapidement (comme cela a pu être observé ailleurs dans le monde, notamment en Inde). Pour cela, il suffit de remarquer que le seuil des 1,5 millions de cas en Afrique a été atteint en un peu plus de 7 mois, et celui des 3 millions de cas en seulement 3 mois supplémentaires.

La modélisation : un outil pour comprendre les dynamiques épidémiologiques

La modélisation est de plus en plus utilisée en santé humaine et animale, et son utilité n'est plus à prouver dans le domaine de l'épidémiologie (Dubé et al. 2007 ; Reiner et al. 2021). Un modèle épidémiologique peut être défini comme la représentation de l'évolution au cours du temps, et parfois dans l'espace, de la transmission d'une maladie au sein de populations, au travers d'une formulation mathématique ou algorithmique. En cherchant à simuler et à reproduire des allures de courbes épidémiologiques, de nombreuses informations peuvent être obtenues, que ce soit sur les mécanismes de transmission ou sur la quantification des paramètres épidémiologiques des maladies et des impacts de celles-ci. Les modèles permettent aussi de comprendre voire de prédire l'évolution d'une maladie à plus ou moins long terme, et ainsi de fournir des données aux autorités pour la mise en place et le réajustement de mesures de gestion sanitaire. Au cours de phases actives d'épidémies, ou de manière rétrospective, la modélisation permet aussi de mesurer l'efficacité des mesures prises pour juguler les maladies.

Il existe diverses catégories de modèles épidémiologiques, selon, entre autres, les données et formulations utilisées ou les hypothèses faites. Issus des travaux de Kermack & McKendrick (1927) dans les années 1920, les modèles compartimentaux sont sans doute les plus utilisés en épidémiologie (Dubé et al. 2007 ; Djidjou-Demasse et al. 2020). Ils consistent à séparer la population en différents groupes : les susceptibles (S), qui peuvent contracter la maladie, les infectieux (I) qui peuvent la transmettre, et les guéris/décédés (R pour *recovered* ou *removed* en anglais) qui ne sont plus infectieux. D'autres groupes peuvent exister selon la complexité du modèle formulé, comme les exposés (E), qui ont été en contact avec l'agent pathogène mais ne sont pas encore infectieux. Ces modèles sont assez simplificateurs et faciles d'utilisation, mais reposent sur un certain nombre d'hypothèses, notamment concernant les paramètres épidémiologiques des maladies étudiées. D'autres modèles, appelés modèles basés sur agents (*agent-based models* ou *individual-based models* en anglais), s'appuient sur la simulation du comportement individuel d'une entité (une personne, un groupe d'individus, etc...)

et permettent de prendre en compte des comportements sociaux dans la simulation de la transmission de la maladie. Ces modèles sont souvent complexes et longs à paramétrer ainsi qu'à faire tourner, mais font aussi l'objet de nombreuses simplifications et d'hypothèses concernant le calibrage des paramètres et des comportements. Une troisième catégorie de modèles épidémiologiques concerne les modèles directement extraits des données (*data driven approach* en anglais). Cette technique s'affranchit de la plupart des hypothèses nécessaires dans d'autres types de modélisations, en utilisant directement les données communiquées pour établir les liens entre les variables et simuler les dynamiques épidémiologiques sous-jacentes. Le principal inconvénient de cette approche tient de la nécessité de disposer de très grands jeux de données pour obtenir des résultats satisfaisants. Selon le jeu de données à disposition et l'objectif du modèle, le choix de la technique utilisée peut varier. Cette description des principales catégories de modèles n'est pas exhaustive, beaucoup d'autres types de modélisation existent, ainsi que d'autres manières de les classer.

Depuis le début de la pandémie de COVID-19, de très nombreux modèles ont été formulés. La plupart d'entre eux sont des modèles compartimentaux SIR ou SEIR. Certains de cette catégorie ont été proposés quelques mois à peine après le début de l'épidémie à Wuhan en Chine (Lin et al. 2020 ; Kuniya 2020 ; Lai et al. 2020), d'autres avec des jeux de données beaucoup plus importants, provenant de cas déclarés dans le monde entier (Manchein et al. 2020). L'Institute for Health Metrics and Evaluation aux Etats-Unis d'Amérique a conçu un modèle compartimental SEIR largement utilisé aujourd'hui et qui fait référence en la matière (Reiner et al. 2021). Cette catégorie de modèles a aussi été utilisée en Afrique (Zandvoort et al. 2020). D'autres techniques de modélisations ont été appliquées depuis un an, notamment au Royaume-Uni et aux Etats-Unis d'Amérique, où un modèle basé sur agents développé par l'Imperial College of London est utilisé par les gouvernements pour orienter leurs réponses politiques (Ferguson et al. 2020).

Cependant, la pandémie de COVID-19, comme toutes les MIE, s'avère difficile à modéliser avec ces techniques pour plusieurs raisons. Tout d'abord, cette pandémie progresse sur la base de nombreux facteurs et processus biologiques et sociaux, et ce autant à l'échelle locale qu'internationale. La complexité des dynamiques de transmission et des effets de la COVID-19 limitent ainsi la capacité des modèles à pouvoir prévoir l'évolution de la pandémie (Mangiarotti et al. 2020a ; Reiner et al. 2021). De plus, la COVID-19 est provoquée par un agent pathogène nouveau et dont les caractéristiques biologiques de transmission sont inconnues et difficiles à mesurer, ce qui oblige les modélisateurs à faire un certain nombre d'hypothèses sur la valeur à attribuer aux paramètres épidémiologiques de la maladie, comme la durée de la période d'incubation, de celle de la phase contagieuse, de l'immunité induite par l'infection ou du nombre de reproduction de base (Roda et al. 2020 ; Ferguson et al. 2020 ; Holmdahl & Buckee 2020). Enfin, l'émergence de cette maladie ne permettait pas aux modélisateurs de disposer de bases de données suffisamment fournies sur lesquelles s'appuyer, du moins durant les premiers mois, ce qui réduisait la robustesse de certains types de modèles (Roda et al. 2020 ; Sridhar & Majumder 2020). Du fait de ce manque global d'informations, la plupart des modèles développés la première année de la pandémie ont donc montré une capacité limitée à décrire et encore plus à prévoir les dynamiques épidémiologiques de la COVID-19, et ce d'autant plus que le nombre d'hypothèses simplificatrices formulées pour créer le modèle était important. Cette imprévisibilité de la pandémie s'avère être problématique pour les gouvernements, qui ont besoin de modèles fiables et robustes afin de développer et adapter leurs INP (Cyranoski 2020 ; Kuniya 2020 ; Roda et al. 2020).

La modélisation globale : une méthode particulière pour une situation particulière

La modélisation globale est une technique permettant la formulation de modèles déterministes directement à partir de séries de mesures (Gouesbet & Letellier 1994; Mangiarotti et al. 2012).

En ce sens, elle fait partie des approches *data driven* décrites précédemment, bien qu'elle puisse aussi s'appuyer sur une préstructure du type des modèles compartimentaux. Cette approche est basée sur la théorie des systèmes dynamiques non linéaires (ou théorie du Chaos). Un comportement chaotique est un comportement imprévisible à long terme bien que déterministe, autrement dit : imprévisible quoiqu'entièrement déterminé par ses équations et ses conditions initiales. Cette théorie, très générale, permet de représenter aussi bien les comportements chaotiques que les comportements non chaotiques (périodiques, quasi périodiques, qui en sont des cas particuliers). La dynamique des épidémies présente des caractéristiques chaotiques dans la mesure où les processus de propagation des maladies infectieuses peuvent être représentés par des processus déterministes (les modèles compartimentaux sont généralement basés sur une formulation entièrement déterministe), quoique le plus souvent imprévisibles à long terme. Le fait de recourir à la technique de modélisation globale apparaît donc tout à fait cohérente. Les modèles issus de cette technique cherchent donc à obtenir une formulation déterministe de la dynamique sous la forme d'équations différentielles ordinaires directement à partir des courbes épidémiques ou autres (Mangiarotti 2015).

La modélisation globale peut donc être appliquée dans des situations complexes, où les dynamiques des systèmes semblent chaotiques. De plus, cette technique utilise directement les données fournies pour retrouver des modèles en adéquation avec les dynamiques observées. Elle s'affranchit donc de nombreuses hypothèses et calibrages nécessaires pour d'autres types de modélisation. Ainsi, on peut avoir des modèles globaux robustes sans connaitre au départ les valeurs des paramètres contrôlant la dynamique du système étudié. Un autre atout de la modélisation globale est qu'elle ne nécessite pas obligatoirement pour fonctionner que l'on rentre les valeurs de toutes les variables du système. Une unique variable peut même permettre d'approcher la dynamique sous-jacente du système étudié. La reconstruction de la dynamique d'ensemble à partir d'un nombre réduit de variables observées se fait en substituant les variables manquantes par les dérivées des variables observées ou par décalage de ces variables (théorème de Takens).

En épidémiologie, et encore plus pour les maladies émergentes comme la COVID-19, l'utilisation de la modélisation globale pourrait s'avérer particulièrement intéressante. En effet, les épidémies suivent des dynamiques complexes, instables et non-linéaires, sous l'influence de nombreux paramètres, aussi bien biologiques qu'environnementaux ou sociaux. De plus, en cas de MIE, des données épidémiologiques importantes peuvent faire défaut ou être très incomplètes (Djidjou-Demasse et al. 2020; Reiner et al. 2021), ce qui n'impactera pas les modèles globaux, qui n'en ont pas nécessairement besoin (Mangiarotti et al. 2020a). Enfin, du fait que tous les modèles ont des limites, l'utilisation du plus grand nombre d'approches de modélisation épidémiologique différentes permet de réduire les risques d'erreurs intrinsèquement liés à chaque type de modèle et donc d'avoir une image plus juste de la dynamique réelle de la COVID-19 (Lu et al. 2020; Sridhar & Majumder 2020).

Un autre atout de la modélisation globale par rapport à d'autres techniques apparait lorsque les modèles recherchés s'appuient sur des données issues de pays étrangers (Rhodes et al. 2020 ; Sridhar & Majumder 2020). En Afrique, comme dans de nombreux pays du Tiers-Monde, les modèles sont souvent biaisés par le point de vue du modélisateur, lorsque celui-ci travaille depuis un autre pays. Il a été constaté dans ce genre de situation que des paramètres décrivant le système étudié peuvent être mal renseignés par le créateur du modèle, par ignorance des spécificités locales ou par a priori. Cela entraine alors des diminutions de performance des modèles obtenus. Or, la modélisation globale, parcimonieuse par le nombre limité d'hypothèses à rentrer dans le modèle, notamment concernant les aspects sociologiques, peut donc être considérée comme une alternative utile à appliquer pour des pays dont les modélisateurs auraient des connaissances limitées, bien qu'une connaissance des régions étudiées soit ensuite nécessaire pour interpréter les modèles obtenus.

Dans le domaine de l'épidémiologie, les travaux de modélisation basé sur la théorie du Chaos sont de plus en plus nombreux, principalement depuis ces deux dernières décennies. Ils ont permis, par exemple, de mieux comprendre la dynamique de transmission de la grippe saisonnière en France (Viboud et al. 2003) ou de la rougeole aux Etats-Unis d'Amérique (Dalziel et al. 2016). Des approches de modélisation globale ont quant à elles déjà été utilisées avant la pandémie de COVID-19 pour simuler la peste bubonique de Bombay entre 1896 et 1911 (Mangiarotti 2015), ainsi que pour l'épidémie d'Ebola en Afrique de l'Ouest entre 2013 et 2016 (Mangiarotti et al. 2016). Concernant la COVID-19, un certain nombre d'études tendent à montrer le caractère non-linéaire, imprévisible et déterministe de la propagation de cette maladie (Chen & Yu 2020 ; Fernandes 2020 ; Shaobo He et al. 2020). Or ce sont les critères définissant un système dynamique chaotique, ce qui suggère donc que la pandémie de COVID-19 présente un comportement chaotique (Jones & Strigul 2020). De ce fait, la technique de la modélisation globale a déjà pu être appliquée avec succès au début de la pandémie, en utilisant les données issues de Chine, du Japon, de la Corée du Sud et de l'Italie. Les modèles trouvés ont mis en évidence l'influence des INP sur la dynamique de l'épidémie et ont permis de prévoir l'évolution à court-terme de la propagation du virus dans d'autres pays où l'épidémie débutait à peine (Mangiarotti et al. 2020a). Cependant, aucune modélisation épidémiologique de la COVID-19 en Afrique s'appuyant sur la théorie du Chaos n'a encore été réalisée ou appliquée.

Partie II : Travail personnel

L'objectif général de ce travail consiste à étudier la nature chaotique de la COVID-19 en Afrique, par la recherche de modèles épidémiologiques globaux, afin d'informer la surveillance précoce des MIE.

Objectifs spécifiques de l'étude

Ce travail de modélisation épidémiologique de la COVID-19 en Afrique s'est construit autour de plusieurs questions :

- Quelles sont les dynamiques de transmission de la COVID-19 en Afrique et peut-on les classer en différents groupes selon l'allure des épidémies, en lien avec la géographie du continent ou de certains facteurs socio-démographiques ou épidémiologiques ? (1)

- Les dynamiques de transmission de la COVID-19 en Afrique ont-elles un comportement chaotique (c'est-à-dire imprévisible et déterministe), et peuvent-elles être approchées par un ou des modèles globaux, capables de reproduire le comportement dynamique dans sa globalité ? (2)

- Peut-on obtenir des modèles globaux permettant de prévoir avec une certaine avance l'évolution des dynamiques épidémiologiques de la COVID-19 dans les pays étudiés ? Si oui, jusqu'à quel horizon, avec quelle précision et avec quel niveau de confiance ? (3)

- Peut-on observer, évaluer ou quantifier à l'aide de nos modèles l'effet des interventions nonpharmaceutiques sur la dynamique de transmission de la COVID-19 ? (4)

- Des modèles globaux de la dynamique de transmission de la COVID-19 dans un pays africain peuvent-ils être appliqués à d'autres pays africains, ou même ailleurs ? (5)

- Les modèles obtenus ici pour la COVID-19, pourraient-ils être utilisés pour étudier l'émergence d'autres maladies ? (6)

La question (1) se concentre sur l'étude des données de chaque pays africain afin de voir si des scénarios d'épidémie se répètent d'un pays à un autre. L'objectif potentiel serait de n'étudier que quelques pays, représentatifs de différents groupes, afin de focaliser le travail sur certaines dynamiques épidémiologiques particulières et suffisamment distantes les unes des autres. Pour la question (2), il s'agit d'étudier si les critères définissant un système chaotique sont remplis d'après les données épidémiologiques de la pandémie en Afrique. Avec cette question, il s'agit également de savoir si l'on peut trouver des modèles globaux pour des pays africains, ainsi que de voir si les modèles obtenus, qui offrent une représentation formelle de l'épidémie, ont des dynamiques chaotiques. La question (3) cherche quant à elle à estimer la robustesse des modèles obtenus et permettre potentiellement de comparer ces modèles entre eux. L'interrogation (4) traite des informations qui pourraient être tirées des modèles globaux, en particulier celles des impacts des interventions non-pharmaceutiques. Les questions (5) et (6) s'intéressent pour leur part à l'application des modèles obtenus à d'autres situations (un autre pays ou une autre maladie). Il est donc question de l'adaptabilité des modèles aux conditions du système.

Matériel et méthodes

Collecte des données

Deux grandes catégories de données ont été collectées : 1) la première concerne la description et la caractérisation de chaque pays africain (54 au total) selon des variables socio-économiques, démographiques et sanitaires; 2) la deuxième concerne les données quotidiennes disponibles liées à la COVID-19 pour chaque pays africain. Les variables utilisées pour chacune de ces catégories de données sont résumées dans le **Tableau 1**.

Il existe une multitude de bases de données recensant le nombre de cas et de morts déclarés positifs à la COVID-19 pour les pays africains. Les bases de données les plus importantes et fréquemment utilisées sont celles de l'Université Johns Hopkins (JHU) aux Etats-Unis d'Amérique, de l'OMS et du site internet WorldOMeters. L'OMS a d'ailleurs créé un tableau de bord comparant les données issues de ces différentes sources, et ce pour tous les pays du monde, facilitant le choix entre les bases de données et permettant de mieux visualiser l'évolution temporelle des variables selon les sources (Epidemic Intelligence from Open Sources initiative 2020). Ainsi, au 19 avril 2021 et pour tout le continent africain, l'OMS décomptait 4 457 000 cas de COVID-19 et 117 930 morts, alors que JHU signalait 4 431 000 cas et 117 898 morts et WorldOMeters 4 421 000 cas et 117 581 morts. Ces écarts, inférieurs à 1% des valeurs totales, sont dus principalement aux différences entre sources utilisées pour faire les décomptes, ainsi qu'à la définition d'un cas et d'un mort de la COVID-19. Dans ce travail, la base de données de JHU a été choisie pour ses valeurs intermédiaires entre les différentes sources possibles et la facilité de la récupération de ses données. Les données exploitées pour élaborer les modèles s'étendent du 22/01/2020 au 15/03/2021, et celles pour tester les modèles obtenus vont jusqu'au 03/05/2021. Les cas de COVID-19 regroupent, pour JHU, les cas probables (signes cliniques et examens d'imagerie médicale évocateurs de la COVID-19) et les cas confirmés (sérologie ou test antigénique confirmant l'infection). De même, les décès dus à la COVID-19 regroupent les décès chez des cas probables ou confirmés de la maladie, sans période de rémission entre le début de la maladie et la mort, et en excluant d'autres causes directement responsables du décès. Du fait de l'hétérogénéité des systèmes de santé entre pays, et du rôle uniquement de collecteur d'informations des organismes tels que JHU ou l'OMS, des limitations sont toutefois à considérer dans l'uniformité de la définition de cas et décès dus à la COVID-19 pour chaque pays.

D'autres variables épidémiologiques ont été recherchées (nombre d'hospitalisations, nombre de cas graves, nombre total de morts au cours du temps), mais les données correspondantes sont rares en Afrique. Seuls quelques pays effectuent un suivi quotidien ou hebdomadaire de ces variables, ce qui ne permet pas de comparer la majorité des pays africains entre eux. Cependant, d'autres variables liées à la COVID-19 sont disponibles pour la plupart des pays, avec notamment les nombres de tests réalisés (et le taux de positivité correspondant), récupérés depuis la plateforme en ligne Our World In Data (Roser et al. 2020). Un indice mesurant la rigueur des mesures gouvernementales au cours du temps, développé par l'Université d'Oxford, a aussi été choisi pour sa capacité à synthétiser en une variable unique l'ampleur des restrictions mises en place pour stopper la propagation du virus au sein d'un pays et au cours du temps. Il a été créé en faisant une agrégation de 9 indicateurs (quantifiés sous forme de scores) représentant des informations sur les INP entreprises à l'échelle nationale, telles que les fermetures d'écoles et les restrictions de mouvement (Hale et al. 2020). Enfin, un indice mesurant la mobilité des populations au cours du temps, dont l'unité est la variation de mouvement par rapport à la période correspondante avant l'arrivée du SARS-CoV-2 (données entre le 3 janvier et le 6 février 2020), est un indicateur direct de l'évolution des mouvements individuels et donc aussi des contacts entre personnes. Cet indicateur, développé par Google en exploitant les données de géolocalisation et de recherche GPS de ses utilisateurs, permet entre autres d'évaluer le respect des mesures gouvernementales de confinement par les populations (Google LLC 2020).

Type de donnée	Nature des données	Sources		
Démographique	Population totale, Densité de population	Base de données de l'ONU (United Nations, Department of Economic and Social Affairs, Population Division 2019)		
9 Ff	Pyramide des âges de la population	Base de données de l'ONU (United Nations, Department of Economic and Social Affairs, Population Division 2020)		
Socio-économique	Taux de pauvreté	Base de données de l'ONU (United Nations, Department of Economic and Social Affairs, Population Division 2019)		
Sanitaire	Nombre d'agents de santé et de médecins par habitant, Nombre de lits d'hôpitaux par habitant	Base de données de la Banque Mondiale (International Bank for Reconstruction and Development, International Development Association 2020)		
	Nombre quotidien cumulé de cas et de morts	Base de données de l'Université Johns Hopkins (Dong et al. 2020)		
Epidémiologique et lié à	Nombre quotidien de tests réalisés et taux de positivité	Base de données <i>Our World In Data</i> (Roser et al. 2020)		
la COVID-19	Indice de rigueur des réponses gouvernementales	Oxford COVID-19 response tracker (Hale et al. 2020) - (Roser et al. 2020)		
	Indice de mobilité globale des populations	Tendances des mobilités par Google (Google LLC 2020) - (Roser et al. 2020)		

Tableau 1. Description des données utilisées

Pour la suite de l'étude, les îles de Mayotte, de la Réunion, de Sainte-Hélène, d'Ascension, de Tristan da Cunha, de Madère et des Canaries n'ont pas été étudiées du fait de leur taille limitée et de leur rattachement à des pays européens. Le Sahara occidental n'a pas été retenu dans l'étude du fait de la complexité administrative locale pour faire remonter les données épidémiologiques. De plus, la Tanzanie a été retirée du jeu de données du fait de la décision gouvernementale de ne plus décompter les cas et les morts de la COVID-19 depuis juin 2020. Enfin, les données issues du Somaliland sont regroupées avec celles de la Somalie. Finalement, 53 pays africains ont donc été retenus dans l'étude.

Analyse comparative des données épidémiologiques entre pays

Une étude des paramètres permettant de réduire au maximum les différences de valeurs de cas et de décès a été réalisée entre pays africains et en comparaison d'autres pays (européens et du monde entier), afin de limiter les biais de comparaison des situations épidémiologiques entre pays. L'ensemble des variables socio-économiques, démographiques et sanitaires décrites dans le **Tableau 1** ont été testées en tant que facteurs de pondération des données.

Catégorisation et sélection des pays d'étude

Afin d'étudier si des pays pourraient avoir des dynamiques épidémiologiques similaires et ainsi être regroupées en catégories, une comparaison qualitative des courbes de cas et de décès quotidiens pour les 53 pays africains a été faite à partir des données comprises entre le début de la pandémie et le 25/04/2021. Cette analyse a consisté à comparer la date de début de chaque vague épidémique, le nombre de vagues observées, ainsi que la différence d'amplitude entre les vagues successives pour chaque pays.

A la suite de cette exploration initiale des données, 17 pays ont été sélectionnés pour l'étude parmi les 53 initialement identifiés. Les pays présentant le plus de données à disposition (absence de trou dans les décomptes quotidiens, nombre de cas/décès significatif) ont été choisis pour l'étude. De plus, quelques pays ont également été inclus dans la sélection du fait de la dynamique singulière de leur épidémie (progression ininterrompue de la maladie sur une longue période, pics épidémiques à des périodes peu communes) et/ou du fait d'une quantité limitée d'informations (afin de tester l'efficacité de la préparation des données à l'étape suivante). L'objectif de cette sélection était de réduire le nombre de pays à étudier, afin d'obtenir une certaine diversité dans les analyses tout en se concentrant sur les pays ayant le plus de chances de générer des modèles robustes.

Pré-traitement des données des pays d'étude

Les traitements de données réalisés par la suite ont été effectués en langage R, version 4.0.3 datant du 10/10/2020 (R Core Team 2020). Le pré-traitement des données épidémiologiques s'est appuyé sur le package 'Covid' en développement au CESBIO de Toulouse (Mangiarotti & Huc 2020), non encore disponible en accès libre. Les modèles globaux ont été générés grâce au package 'GPoM' (Mangiarotti et al. 2020b).

Pour ce travail, nous avons étudié les variations au cours du temps des nouveaux cas et décès quotidiens, ainsi que leurs dérivées premières (représentant la vitesse de variation des nouveaux cas/décès quotidiens) et secondes (représentant cette fois l'accélération). Dans un premier temps, une recherche des valeurs aberrantes dans les données de nouveaux cas/décès quotidiens a été réalisée. Elle a consisté en une identification des pics anormalement élevés par rapports aux valeurs voisines (correspondant généralement à des rattrapages sur les jours précédents), ou de valeurs négatives (correspondant au retrait de cas signalés les jours précédents). Une fois ces valeurs repérées, une explication a été recherchée auprès des rapports quotidiens ou hebdomadaires de l'OMS et des organismes nationaux officiels de comptage des cas. Un réajustement manuel des valeurs aberrantes a ensuite été entrepris, avec une correction des valeurs des jours précédents si nécessaire (les cas de rattrapages ont été re-répartis de façon égale sur la période correspondante).

Après ces quelques modifications ponctuelles, les séries journalières de cumuls ont été souséchantillonnées au pas de temps hebdomadaire, conduisant à un ensemble de sept séries différentes, chacune correspondant à un jour de la semaine. Cette scission en 7 des valeurs quotidiennes a permis de repérer plus facilement les jours de la semaine où les valeurs étaient systématiquement plus basses, du fait d'un arrêt ou d'une réduction la veille des dépistages ou des remontées d'informations. Ces jours, que l'on appelle jours creux, sont généralement les samedis, dimanches et surtout les lundis, résultant des jours de repos de la fin de semaine. Lorsqu'une série creuse a été repérée, elle a été retirée du jeu de donnée, du fait de sa sous-représentativité de la circulation réelle du virus par rapport aux autres jours de la semaine, et de sa perturbation dans la continuité des valeurs. Tous les jours sans nouveaux cas enregistrés ont aussi été retirés du jeu de données, afin de ne garder que les jours ayant des valeurs non nulles.

Chacune des séries hebdomadaires restantes a été ensuite rééchantillonnée par spline cubique, une technique d'interpolation permettant un lissage de trajectoires entre des points par une équation polynomiale de degré 3. Ce rééchantillonnage a été fait au pas de temps horaire, pour à la fois de s'affranchir des perturbations ponctuelles, disposer de séries très lisses à un pas de temps très fin, et disposer d'une estimation de l'erreur associée à notre prétraitement (qui nous est donnée par la dispersion des séries temporelles reconstruites). Des estimations moyennes ont ainsi pu être obtenues par le calcul, à chaque date, de la moyenne des sept séries reconstruites (s'il n'y avait pas de série creuse) et de leur écart-type. Les dérivées par rapport au temps de cette série moyenne de cas ou décès, dont nous aurons besoin par la suite pour appliquer la technique de modélisation globale, ont alors été calculées par l'algorithme de Savitzky-Golai.

Recherche de modèles globaux

Un atout de la modélisation globale est de pouvoir approcher la dynamique épidémiologique d'une maladie grâce à une unique variable et, dans le cas présent, de ses dérivées. Dans cette situation, il s'agit du nombre quotidien de nouveaux cas ou décès. En supposant que la dynamique de la COVID-19 peut être réduite à une dimension 3 (c'est-à-dire que trois variables suffisent à représenter l'évolution de cette dynamique), et qu'il est possible de l'approcher avec des équations différentielles de degré 3, il existerait alors des modèles permettant de reproduire la dynamique épidémiologique de la COVID-19, basés sur une formulation canonique dont les équations sont les suivantes :

$$\begin{cases} \dot{X}_{1} = X_{2} \\ \dot{X}_{2} = X_{3} \\ \dot{X}_{3} = P(X_{1}, X_{2}, X_{3}) + \varepsilon \end{cases}$$
(1)

avec X_1 le nombre de cas ou décès quotidiens, X_2 et X_3 ses dérivées première et seconde, ε les résidus, et P un polynôme de degré choisi et fixé, ici égal à 3. Dans le cas où tous les termes de ce polynôme sont disponibles, celui-ci s'écrit alors:

$$P(X_1, X_2, X_3) = a_0 + a_{100}X_1 + a_{010}X_2 + a_{001}X_3 + a_{110}X_1X_2 + a_{101}X_1X_3 + a_{011}X_2X_3 + a_{200}X_1^2 + a_{020}X_2^2 + a_{002}X_3^2 + a_{111}X_1X_2X_3 + a_{210}X_1^2X_2 + a_{201}X_1^2X_3 + a_{120}X_1X_2^2 + a_{021}X_2^2X_3 + a_{102}X_1X_3^2 + a_{012}X_2X_3^2 + a_{001}X_3^2 + a_{000}X_3^2 + a_{00}X_3^2 + a_{00}X_3^$$

où les a_{ijk} sont les coefficients correspondants aux termes X_1^i, X_2^j, X_3^k , respectivement.

Le package 'GPoM' en langage R permet, grâce à un algorithme de recherche de structures, de retrouver des équations aux dérivées ordinaires de forme très générales, partant d'une ou de plusieurs séries observationnelles. Partant de la formulation donnée par les équations (1), l'algorithme a été utilisé pour obtenir les formulations polynomiales $P(X_1, X_2, X_3)$ et les coefficients a_{ijk} les mieux adaptés afin d'approcher au mieux la dynamique de la série temporelle qui lui est donnée. Pour chacun des 17 pays africains retenus dans l'étude, cette recherche d'équations polynomiales a été réalisée pour les séries temporelles des cas quotidiens et des décès quotidiens préparées précédemment.

Plus précisément, la recherche d'équations polynomiales s'est faite pour l'ensemble de la fenêtre temporelle exploitée (soit 1 an et 2 mois), puis pour des portions de la série temporelle. Par exemple, si la série temporelle comportait 2 vagues épidémiques, l'algorithme a été lancé pour la fenêtre entière, puis pour la fenêtre correspondante à la première vague et enfin pour celle de la deuxième vague.

Etude approfondie de l'espace des phases et de la prédictibilité des modèles

Tous les modèles ne divergeant pas ou ne convergeant pas vers un point fixe et intégrables sur plus de 600 jours (soit 95 modèles), ont été gardés pour une étude plus approfondie de leur espace des phases (ou espace des états). Cet espace nous donne l'évolution au cours du temps de l'épidémie vue au travers d'une variable X_1 (le nombre de cas ou de décès) et de ses deux premières dérivées X_2 et X_3 , et peut être vu sous différentes projections : $(X_1, X_2), (X_2, X_3)$ et (X_1, X_3) . Ces graphes permettent d'avoir une représentation spatiale de la dynamique des modèles obtenus, et de comparer cette dynamique avec la dynamique issue des données réelles. Afin de disposer d'un premier élément quantitatif de validité, la prédictibilité des modèles a été étudiée en s'appuyant sur les données réelles comprises entre le 22/01/2020 et le 15/03/2021. Cette étude s'est faite en mesurant la différence entre les valeurs simulées par les modèles avec les valeurs réelles jusqu'à un horizon de 20 jours et pour des conditions initiales comprises tout au long de la fenêtre temporelle étudiée. Cette analyse ne peut suffire à valider complètement les modèles puisque cette période a servi à obtenir les modèles (les modèles testés ne sont donc pas indépendants des données). Ils offrent toutefois un premier élément de validation, partiel, dans la mesure où les modèles ici testés n'ont pas été ajustés en se basant sur leurs capacités prévisionnelles mais en ajustant le modèle à la dynamique de l'épidémie (c'est-à-dire en ajustant des équations à une trajectoire dans l'espace des phases).

Sélection des modèles et étude de robustesse

Tous les modèles avant une prédictibilité satisfaisante à 5 jours, ne se dégradant pas de manière exponentielle à un horizon de prédiction de 10 et 20 jours, et approchant si possible correctement la dynamique épidémiologique étudiée (cohérence données-modèle dans l'espace des phases), ont été sélectionnés (ce qui représente 37 modèles) afin de comparer plus précisément leur prédictibilité. Pour hiérarchiser les modèles selon leur robustesse, c'est-à-dire leur capacité à prévoir avec le moins d'erreurs possibles l'évolution du nombre de cas ou de décès, un calcul de la valeur de l'erreur absolue avec un intervalle de confiance à 95 et 75% et pour un horizon de 10 et 20 jours a été entrepris pour chaque modèle. Le nombre total de cas ou de décès enregistrés par chaque pays aurait biaisé la comparaison des modèles entre eux si elle avait été faite à partir des valeurs calculées brutes de l'erreur absolue. En effet, un nombre plus important de cas ou de décès dans un pays par rapport à un autre entraîne généralement une erreur plus élevée du nombre de cas/décès estimés par les modèles, même s'ils sont plus robustes, du fait de la simulation de valeurs plus élevées. Pour contourner ce biais, les valeurs calculées des erreurs absolues de chaque modèle ont été divisées par le nombre total de cas (ou de décès) enregistré au 15/03/2021 (dernière date utilisée pour construire les modèles), par le pays dont le modèle est issu. Ce sont ces valeurs d'erreur absolue pondérées par le nombre total de cas ou de décès, qui ont été comparées entre elles, et qui ont permis de classifier les modèles selon leur robustesse et selon les conditions voulues (intervalle de confiance à 95% ou 75% et horizon de prédiction de 10 ou 20 jours). Enfin, les modèles les plus robustes ont été testés pour les données issues après le 15/03/2021 afin d'évaluer leur capacité à prévoir la dynamique épidémiologique à plus long terme et pour des conditions qui ne sont plus celles ayant permis la conception des modèles.

Etude de non-stationnarité d'un modèle

Afin de comprendre le sens des paramètres composant l'équation polynomiale de la formule (1) une étude de la non-stationnarité au cours du temps des coefficients $a_{ijk}(t)$, et du résidu $\varepsilon(t)$ des équations différentielles (1) a été entreprise de manière exploratoire pour un des modèles les plus robustes et ayant peu de termes au sein de son équation polynomiale. Cela a consisté en une analyse de la variation de valeur des coefficients $a_{ijk}(t)$, et du résidu $\varepsilon(t)$ selon la fenêtre temporelle utilisée pour concevoir le modèle, ainsi qu'une exploration de la variation des valeurs prédites par le modèle selon le calibrage des coefficients. L'idée d'une telle reconstruction avait déjà été testée auparavant sur un cas théorique (Mangiarotti et al. 2017).

Résultats

Importance des données dans l'analyse comparative de l'impact de la COVID-19 entre les différents pays du monde

L'analyse des données épidémiologiques, en termes de nombres de cas d'infection et de décès dus à la COVID-19, montre de grandes variations entre pays africains et plus largement entre tous les pays du monde, qui sont potentiellement dues à des biais dans le nombre et la qualité des données disponibles. Afin de réduire les différences de valeurs entre pays, il est nécessaire de considérer la taille des populations de chaque pays. Manier des taux par habitant ne suffit cependant pas à expliquer l'ensemble des différences observées, sachant que l'épidémie ne sera pas homogène au sein de chaque pays, et ce, d'autant plus s'il est grand. Selon la pyramide des âges, selon la définition d'un cas et d'un mort de la COVID-19 et selon la politique de dépistage, les taux de morbidité et de mortalité ne seront pas les mêmes d'un pays à un autre, alors même que les situations épidémiologiques réelles pourraient être identiques (Middelburg & Rosendaal 2020). Seuls 2 des facteurs étudiés ont permis une réduction importante des différences entre pays, en plus de la taille de la population : la part d'habitants de plus de 60 ans et le nombre de lits d'hôpitaux disponibles par habitant. Le **Tableau 2** résume la dispersion maximale des valeurs entre pays (en Afrique, en Europe ou dans le monde en général) pour les cas et les morts sans pondération, ainsi qu'en appliquant un ou plusieurs des 3 facteurs de pondération précédemment énoncés. Les pays utilisés pour cette comparaison et pour les suivantes sont listés en annexe 1.

Tableau 2. Ratio d'écart entre les valeurs minimales et maximales du nombre total de cas ou de décès au 15/03/2021 au sein de 52 pays africains, 13 pays européens et 169 pays du monde entier, pour les données sans pondération (Brut) ou pondérées par le nombre de lits d'hôpitaux par habitant (Lits), par le nombre d'habitants du pays (Hab.), par le nombre d'habitants de plus de 60 ans (Hab.>60) ou par une combinaison de ces facteurs.

Variables	Pays étudiés	Brut	Lits	Hab.	Hab.>60	Lits*Hab.	Lits*Hab.>60
	Afrique	2048	2584	164	121	81	201
Cas totaux	Europe	52	73	5	4	7	8
	Monde	446900	602461	15273	8015	4852	16577
	Afrique	17140	6244	3377	1628	952	603
Décès totaux	Europe	196	275	16	15	22	21
	Monde	535628	890557	36059	17095	13807	9489

On observe ainsi qu'en ne pondérant les données qu'avec le nombre de lits d'hôpitaux par habitant, les dispersions de valeurs sont accentuées, quel que soit le groupe de pays considéré. La seule exception observée concerne les décès en Afrique, où ce paramètre permet de réduire la dispersion maximale d'un facteur 2.7 (de 17140 à 6244). En tenant compte uniquement du nombre d'habitants par pays, les différences entre pays sont fortement réduites pour tous les groupes de pays, ce qui apparait logique. En tenant compte cette fois uniquement du nombre d'habitant de plus 60 ans, cette réduction des différences est encore accentuée, et plus pour les pays africains et du monde entier que pour les pays européens. En Afrique, cette réduction des différences par rapport aux données pondérées par le nombre total d'habitants va d'un facteur 1.36 pour les cas à 2 pour les décès. Pour l'Afrique et l'ensemble des pays du monde, la réduction de la dispersion maximale varie selon la variable épidémiologique étudiée : pour les cas d'infection, il s'agit de la pondération par le nombre de lits par habitant et par le nombre total d'habitants (réduction d'un facteur 2 en Afrique par rapport aux données pondérées par le nombre total d'habitants), tandis que pour les décès, il s'agit de la pondération par le nombre de lits par habitant et par le nombre d'habitants de plus de 60 ans (réduction d'un facteur 3.5 en Afrique par rapport aux données pondérées par le nombre total d'habitants). Ainsi, pris individuellement, le facteur nombre de lits d'hôpitaux par habitant ne réduit pas nécessairement les différences, mais utilisé avec d'autres facteurs, il peut les réduire significativement. Pour la sélection de pays européens par contre, la prise en compte simultanée du nombre de lits par habitant et du nombre d'habitants (total ou de plus de 60 ans) ne permet pas de mieux expliquer les différences observées.

Lorsque les facteurs de pondération sont appliqués pour prendre en compte des biais liés à l'âge des populations et à l'infrastructure médicale de chaque pays, on observe une réduction nette des différences d'impact de la COVID-19 entre les pays et une homogénéisation des valeurs entre pays africains et européens notamment (**Figure 1** et **annexes 2** et **3**). Enfin, l'**annexe 4** fournit une comparaison des valeurs de lits d'hôpitaux disponibles par habitant en France et pour les pays africains, afin de pouvoir mieux estimer les variations de valeurs de ce facteur entre pays, et ainsi les biais potentiels qu'il pourrait refléter.



Afrique OEurope

Figure 1. Nombres de cas et de décès liés à la COVID-19 pondérés par million d'habitants de plus de 60 ans et par lits d'hôpitaux disponibles par million d'habitants, pour une sélection de pays européens et africains.

Catégorisation des pays africains et sélection des pays d'étude

Les pays africains ont recensé leur premier cas de COVID-19 entre la mi-février et la mi-mai 2020 (3 en février, 46 en mars, 4 en avril et 1 en mai). Ils ont présenté un nombre variable de vagues épidémiques, à des périodes de l'année elles aussi assez variables. Cependant, certains rapprochements entre pays ont pu être faits, et son présentés dans la **Figure 2**. Les allures de quelques courbes épidémiologiques sont données en exemple dans l'**annexe 5**.



Figure 2. Catégorisation des pays africains selon la date d'apparition et la durée de leurs différentes vagues épidémiques entre le début de la pandémie et le 25/04/2021. Carte générée depuis le site internet Mapchart (*Mapchart 2021*).

17 pays ont été retenus dans l'étude sur une base de 53 pays potentiels. L'Afrique du Sud, l'Algérie, l'Egypte, l'Ethiopie, le Ghana, le Kenya, la Libye, le Maroc, le Nigeria, la Tunisie et la Zambie ont été sélectionnés pour le grand nombre de données disponibles. Le Cameroun, la Côte d'Ivoire, la Namibie, le Sénégal et le Zimbabwe ont été sélectionnés pour diversifier les dynamiques épidémiologiques étudiées tout en préférant les pays avec le plus de données, l'Ouganda et le Mozambique ont été écartés. Le Cameroun présente des courbes de cas et de décès en dent de scie. Il a ainsi été sélectionné pour tester l'efficacité de la préparation des données à l'étape suivante, dont l'un des objectifs est précisément de s'affranchir de ce type de biais. Enfin, le Togo, un pays ayant beaucoup moins de données, a été ajouté à la sélection afin de vérifier les capacités de l'approche à modéliser des dynamiques à partir d'une quantité limitée d'informations. Ce pays est aussi intéressant à étudier du fait de sa dynamique épidémiologique singulière, consistant en une croissance des cas et des décès tout au long de l'année 2020, avec une décroissance de l'épidémie à partir d'avril 2021.

Traitement des données et étude des modèles obtenus

Pour les 17 pays, le traitement numérique des données épidémiologiques préalablement préparées (**annexes 6** et **7**) a conduit à la sélection de 95 modèles potentiels, qui sont intégrables et qui ne divergent pas ou ne convergent pas en un point fixe après 600 jours de modélisation. Cependant, certains d'entre eux ont un portrait de phase éloigné de celui des données réelles, et/ou une prédictibilité générale assez mauvaise. Le modèle à 6 paramètres simulant les cas en Ethiopie à partir des données

entre la mi-mars 2020 et la mi-février 2021 en est un exemple (panels supérieurs des Figures 3 et 4). A l'inverse, d'autres modèles ont un portrait de phase beaucoup plus proche de la dynamique réelle et une prédictibilité plus robuste, comme celui à 10 paramètres simulant lui aussi les cas en Ethiopie mais cette fois à partir de données allant de la mi-mars 2020 à la mi-janvier 2021. Les panels inférieurs des Figures 3 et 4, représentent le portrait de phase et la croissance de l'erreur de prévision de ce modèle, et peuvent servir de comparaison avec le modèle décrit précédemment. On observe notamment que le portrait de phase du modèle à 6 paramètres a des amplitudes de variation plus élevées que les données réelles, et converge sur un attracteur chaotique dont le nombre de cas oscille autour de 80-100 cas par jour (en abscisse) sans permettre d'atteindre les valeurs observées les plus fortes. A l'inverse, le portrait de phase du modèle à 10 paramètres est en bon accord avec les données tout au long de l'intégration (Figure 3). Pour ces modèles, on observe en revanche que les portraits de phase offrent une mauvaise relation en termes de séries temporelles. Cela vient du fait qu'il n'est possible de faire de bonnes prévisions qu'à des échéances de temps relativement réduites, (ce qu'on observe par la suite avec la Figure 4). Or, on part pour ces simulations du tout début de l'épidémie, on ne peut donc pas espérer une bonne prévision de son déroulement (le début de croissance est assez bien prévu, mais de façon très lisse, mais les variations de grandes amplitudes qui suivent sont prévues beaucoup plus tard par le modèle). Le modèle voit donc assez bien la dynamique transitoire du début de croissance (mais en plus lisse), il voit également assez bien le type de variation qui peuvent avoir lieu lors d'une flambée et ses irrégularités. Par contre, il ne permet pas de prévoir leur succession (car c'est du long terme et que ces successions dépendent fortement des conditions initiales et des perturbations).



Figure 3. Séries temporelles (côté gauche) et portrait de phase dans sa projection (C1,C2) (côté droit) observé (en noir) et modélisé (en couleur) par le modèle à 6 paramètres (panels du haut) et par le modèle à 10 paramètres (panels du bas), pour les cas déclarés de COVID-19 en Ethiopie entre le 22/01/2020 et le 15/03/2021. Les valeurs observées en trait plein sont celles qui ont été utilisées pour identifier le modèle. Les valeurs modélisées sont en couleur. Les couleurs correspondent à des simulations issues de conditions initiales légèrement différentes, choisies à partir des données observées en tout début d'épidémie le 17/03/2020 à 7h (violet) et à 19h (rouge) et le 18/03/2020 à 7h (orange). Les cas journaliers sont normalisés par million d'habitants de plus de 60 ans.

Concernant l'erreur de prévision, on observe visuellement avec les graphes bidimensionnels que l'erreur croit plus vite, dès 5 jours d'horizon, pour le modèle à 6 paramètres que pour celui à 10 paramètres (**Figure 4**). Cela se confirme lorsque l'on regarde, toujours pour les mêmes figures l'erreur maximale à 20 jours avec un niveau de confiance de 75% : le modèle à 6 paramètres se situe autour de 45 cas par million d'habitants de plus de 60 ans, pour 39 cas avec le modèle à 10 paramètres. Cependant, cet écart se resserre pour des niveaux de confiance plus élevés (autour de 77 cas pour les 2 modèles avec un niveau de 90% et 106 cas avec 95%). Cela pourrait s'interpréter comme le fait qu'au-delà d'un certain seuil, l'erreur ne croit plus : à 20 jours, ce seuil est déjà atteint depuis plus longtemps par

le modèle à 6 paramètres que par celui à 10 paramètres. C'est grâce à une analyse de ce type que les meilleurs modèles pour chaque pays et chaque variable épidémiologique ont pu être sélectionnés, réduisant la liste à 37 modèles conservés pour la suite du travail. Cette liste, accompagnée pour chacun des modèles de l'erreur absolue de prévision divisée par le nombre total de cas (ou de décès), est fournie en **annexe 8**. Aucun modèle suffisamment fiable n'a été trouvé pour la Zambie, que ce soit pour les cas ou décès.



Figure 4. Evolution de l'erreur absolue $|e(\tau)|$ de prévision du nombre journalier de cas de la COVID-19 en Ethiopie estimée par le modèle à 6 paramètres (panels du haut) et par le modèle à 10 paramètres (panels du bas), pour un horizon de prédiction de 0 à 20 jours et pour 200 conditions initiales également réparties sur la période courant du 22 janvier 2020 au 25 février 2021. 4 graphes sont présentés : l'évolution de l'erreur de prévision sous forme de courbes (côté gauche) avec pour chaque courbe verte la variation de l'erreur sur 20 jours pour une condition initiale tirée des données observées. Les courbes en noir correspondent à l'erreur associée à un niveau de confiance de 75% (en tirets), 90% (pointillés et tirets) et 95% (pointillés). L'autre représentation, sous forme bidimensionnelle (côté droit), avec en abscisse le nombre de jours depuis le 1^{er} janvier 2020 et en ordonnée l'horizon de prédiction, lui aussi en jours. L'erreur de prévision y est représentée en couleur, avec des tons plus foncés pour une surestimation du nombre de cas et des tons plus clairs pour une sous-estimation. Des courbes de niveau pour des erreurs de -100, -25, -1, 1, 25 et 100 ont été représentées sur ces graphes. Les cas journaliers sont normalisés par million d'habitants de plus de 60 ans.

Pour les 37 modèles conservés, les modèles présentent une grande variabilité de qualités prédictives, comme le montre la **Figure 5**. Les niveaux d'erreur varient dans l'ensemble d'un facteur 1 à 3, quel que soit l'horizon et le niveau de confiance, même si des niveaux d'erreur beaucoup plus grands sont observés pour certains pays (l'Afrique du Sud, la Namibie, le Togo et le Zimbabwe). En retirant de la sélection les modèles de ces pays, il n'en reste que 28, présentés dans l'**annexe 9** de la même manière que pour la **Figure 5**, avec en plus une attribution pour chaque point du nom du modèle. Il est aussi à noter que la croissance de l'erreur absolue est plutôt linéaire entre les horizons de prédiction (une erreur 3 fois plus grande en passant de 10 à 20 jours) et les niveaux de confiance étudiés ici (une erreur entre 2 et 3 fois plus grande en passant de 75 à 95%). Cela montre le caractère linéaire de la relation entre erreur de prévision et niveaux de confiance dans les 4 conditions étudiées ici (horizon de prédiction à 10 et 20 jours et niveaux de confiance de 75% et 95%). Dans l'ensemble, les prévisions des cas d'infection présentent généralement une erreur plus faible que ceux des décès (qui ne représentent qu'un tiers des modèles illustrés sur la **Figure 5**), bien que ce ne soit pas systématique.



Figure 5. Erreur absolue de prévision divisée par le nombre total de cas (ou de morts) au 15/03/2021, des modèles sélectionnés pour un niveau de confiance de 95 et 75% et à un horizon de 10 jours (haut) ou 20 jours (bas). Les points correspondent aux modèles simulant des cas d'infection et les triangles à ceux simulant des décès.

Les 37 modèles initialement conservés ont été utilisés pour prévoir à 20 jours la dynamique épidémiologique comprise entre le 22/01/2020 et le 03/05/2021, afin d'étudier leurs capacités prévisionnelles sur une période de validation (non utilisée au cours du processus de modélisation) du 15/03/2021 au 03/05/2021. Les portraits de phases et graphiques de prédictibilité de 14 de ces modèles sont présentés en **annexe 10**. Pour 13 de ces modèles, une comparaison de la variation de l'erreur absolue de prévision divisée par le nombre total de cas (ou de décès), est fournie en annexe 11. On constate que la variation de la prédictibilité des modèles pour les derniers mois est très dépendante de la dynamique épidémiologique étudiée durant cette période. Il y a globalement moins d'erreurs dans les prévisions lorsque la maladie circule à bas bruit plutôt que lorsqu'elle est en phase épidémique. La prise en compte de ce biais par la division de l'erreur absolue par le nombre de cas d'infection ou de décès total au 03/05/2021 permet de ne pas observer de tendance particulière concernant les capacités prévisionnelles des modèles sur la période de validation et pour des horizons de prédiction de 10 et 20 jours. La prédictibilité globale des modèles reste plutôt stable et ne se détériore pas, du moins pas à partir d'un horizon de prédiction de 10 jours et pour 1 mois et demi de données supplémentaires. Il a aussi été constaté que les modèles générés à partir des données les plus anciennes permettent également de prévoir l'évolution de la maladie durant les derniers mois, bien qu'une dégradation de l'erreur absolue de prévision ait été observée pour certains d'entre eux. Enfin, parmi les modèles conservés pour cette étude, ceux générés avec plus de données n'ont pas particulièrement tendance à être plus robustes que ceux conçus avec moins de données.

Le modèle prédisant les cas en Ethiopie a été choisi pour l'étude de la stationnarité de ses paramètres, car il possède des équations différentielles relativement courtes (10 paramètres) et se place parmi les meilleurs modèles en termes de prédictibilité. Sa formulation canonique est la suivante :

$$\begin{cases} \frac{dX_1}{dt} = X_2 \\ \frac{dX_2}{dt} = X_3 \\ \frac{dX_3}{dt} = 0.267966X_3^2 - 0.039073X_3^3 - 0.059365X_2 + 0.004392X_1 - 0.001147X_1X_3^2 + 1.18e^{-5}X_1X_2^2 - 7.57e^{-5}X_1^2 + 2e^{-7}X_1^3 + e^{-7}X_1^3 + e^{-7}X$$

Pour étudier l'évolution de l'épidémie et pour tenter de détecter l'influence des mesures nonpharmaceutiques, une analyse de non-stationnarité est proposée ici. Cela consiste à reconstruire l'évolution temporelle des paramètres de modèle au cours de l'épidémie. Le potentiel de l'approche a pu être démontré sur un cas théorique (Mangiarotti et al. 2017), mais son application à des données réelles n'a jamais été testée. Comme l'illustre la Figure 6, les coefficients des paramètres de l'équation polynomiale du modèle varient beaucoup au cours du temps (avec une alternance de valeurs positives/négatives pour la plupart des coefficients). Ce comportement n'est pas incohérent avec la situation étudiée dans la mesure où un changement brutal des paramètres d'un comportement dynamique (telle l'application soudaine d'un confinement ou autres INP dans le cas réel) peuvent justement conduire à des variations extrêmes des coefficients, sur des fenêtres réduites. Cette situation est d'autant plus difficile à étudier car elle part de données épidémiologiques estimées à partir de conditions réelles. L'analyse des courbes obtenues montre que certaines parties de ces courbes reconstruites apparaissent relativement stables, généralement pour des périodes qui sont les mêmes pour la plupart des paramètres. Deux de ces portions ont été sélectionnées et correspondent à la période milieu de croissance et décroissance de la première vague jusqu'à sa stabilisation pour la première (jours 210 à 260), et au début de croissance de la deuxième vague pour la deuxième (jours 390 à 410). Outre les problèmes qui viennent d'être évoqués, une des difficultés de l'approche tient au fait que la signification des coefficients ici considérés nous est a priori inconnue. L'analyse comparative des valeurs des coefficients estimées sur ces deux fenêtres est présentée dans le Tableau 3. Cette analyse montre des comportements dynamiques de ces paramètres assez nettement symétriques entre ces deux périodes qui sont en cohérence avec les phases épidémiologiques de chacune des deux périodes. Pour mieux comprendre l'influence des coefficients des paramètres sur le modèle, des changements de valeurs de ceux-ci ont été réalisés afin de voir comment les prévisions variaient, mais les conclusions de ce travail se sont avérées très limitées. Quelques résultats sont présentés en annexe 12. Une comparaison a été faite entre l'évolution des coefficients des paramètres et celle d'une variable connue pour ce pays, l'indice de rigueur des mesures gouvernementales (annexe 13 pour la courbe de cette variable). Cependant, aucune corrélation franche n'a été trouvée pour ce modèle.

L'analyse des résidus (**Figure 6**), montre que le modèle obtient les meilleurs ajustements sur le début de la fenêtre temporelle (au 100ème jour), soit avant et au début de la première vague. L'ajustement se dégrade progressivement et de manière continue puis se stabilise à partir du 180ème jour (régressant même un peu autour du 350ème jour). Ces variations d'ajustement semblent assez directement liées à l'amplitude du signal. On observe aussi en tout début et toute fin d'intervalle temporel un bond net des valeurs, peut-être imputable aux données tronquées aux extrémités.



Figure 6. Etude de non-stationnarité basée sur l'analyse des variations des coefficients des modèles au cours du temps. Partant de la courbe du nombre de cas journaliers en Ethiopie pondéré par million d'habitants de plus de 60 ans (premier graphe en haut à gauche), les variations des paramètres $a_{ijk}(t)$ ont été estimées par une succession d'ajustements estimés sur une fenêtre glissante de 30 jours (le terme correspondant à chaque coefficient est fourni en titre de chaque graphe). La dernière courbe correspond aux résidus $\varepsilon(t)$ de ces ajustements. Les points en noir étant ceux retenus pour l'étude en raison de leur relative stabilité (les variations rapides, en gris, ont été exclues). La ligne verte représente la valeur des paramètres du modèle issu de la fenêtre temporelle complète. Le jour 1 pris comme référence correspond au 01/01/2020, et les dernières valeurs temporelles correspondent au 15/03/2021. Les deux traits verticaux sur la courbe du nombre de cas journaliers permettent d'illustrer la largeur de la fenêtre d'analyse (30 jours) qui a été déplacée le long de la courbe (heure par heure) pour étudier les variations des paramètres.

Coefficients	Accélération de l'épidémie	Ralentissement de l'épidémie
X3 ²	+	-
X_3^3	-	+
X ₂	<-0.15	> -0.15
\mathbf{X}_1	- ou لا	⊅ ou +
$X_1 X_3^2$	-	+
$X_1 X_2^2$	+ ou -	0
X1 ²	<i>∧</i> ou +	- ou ک
X1 ³	- ou ک	⊅ ou +

Tableau 3. Comparaison du signe, de la valeur et de l'évolution des coefficients du modèle des cas en Ethiopie selon la dynamique du nombre de cas journaliers durant les 2 phases retenues pour l'étude de la stationnarité

Discussion

Les pays du monde entier et en particulier les pays africains présentent de très grandes différences de nombres de cas et de décès de la COVID-19 par habitant avec des ratios pouvant aller de 1 à 3376 pour les pays africains, et de 1 à 8408 entre pays africains et européens. La recherche de facteurs permettant d'expliquer ces différences ont permis de faire ressortir deux facteurs principaux : (1) la proportion de personnes âgées (>60ans) et (2) le nombre de lits d'hôpital par habitant. Ces facteurs ont l'intérêt d'être disponibles pour la majorité des pays, et sont apparus comme très efficaces pour réduire les écarts entre pays, autant en Afrique que sur d'autres continents. La proportion de personnes âgées est un facteur qui permet d'expliquer une part importante des variations observées, notamment en Afrique (permettant une réduction d'un facteur 1,3 à 2 des dispersions maximales par rapport aux nombres de cas et de décès par habitant) ce qui rend vraisemblablement compte d'une situation réelle : les personnes les plus symptomatiques et les plus à risque de mourir de cette maladie sont les personnes âgées, avec une croissance du risque plus élevée à partir de 60 ans (Bonanad et al. 2020). Les pyramides des âges jouent donc un rôle important dans le nombre de morts enregistrés, ainsi que très probablement dans le nombre de personnes dépistées, car symptomatiques. A l'inverse, le second facteur relève d'un biais, et le révèle : moins il y a de lits d'hôpitaux et plus on sous-estime le nombre de cas. Il s'agit donc d'un biais et les estimations pourraient être sous-estimées en Afrique d'un facteur allant de 1,8 à 64,8 (en comparaison du nombre moyen de lit d'hôpitaux par habitant en France). En considérant en plus de la population totale le nombre de lits disponibles, on réduit d'ailleurs les différences sur ce continent d'un facteur 2 à 3,5. On observe également que la capacité de réduction des différences par ces facteurs est variable, pour l'Afrique tout comme plus globalement pour l'ensemble des pays du monde, selon l'indicateur épidémiologique considéré : pour les cas, c'est la combinaison du nombre de lits disponibles et de la population totale qui permet au mieux de réduire la variabilité des valeurs, tandis que pour les morts, c'est la combinaison du nombre de lits disponibles et de la population de plus de 60 ans qui est la plus efficace. Cela pourrait s'interpréter comme le fait que la proportion de personnes âgées n'a pas particulièrement d'influence sur le nombre de cas d'infection détectés. Enfin, d'autres facteurs ont été testés (taux de pauvreté, PIB, densité de population, nombre d'agents de santé et de médecins par habitant) mais aucun de ceux-ci n'a permis de mettre en évidence des réductions aussi significatives des différences entre pays. Cette étude mériterait d'être approfondie car la littérature permettant d'expliquer ces différences reste à ce stade encore très limitée.

Une des questions de ce travail portait sur la classification potentielle des pays africains selon leurs courbes épidémiologiques. Cinq grands groupes ont été proposés, en fonction du nombre, de la durée et de la date de commencement des vagues épidémiques. Cependant, de nombreux pays sont à la limite entre plusieurs groupes, et au fur et à mesure de l'avancée dans le temps, les classifications évoluent et se complexifient. De plus, au début de la pandémie, les capacités de dépistage étant moins importantes et l'épidémiosurveillance moins organisée, les malades et décès de la COVID-19 ont vraisemblablement été sous-estimés, voire confondus avec d'autres maladies. Il est donc difficile de trouver des critères permettant de distinguer les pays entre eux de manière claire et précise, autant par le grand nombre de pays et de situations différentes que par les données utilisées, parfois inexactes. En revanche, des tendances ont été observées. Tout d'abord, la majorité des pays étudiés ont présenté 2 vagues épidémiques ou plus de grande ampleur, avec une période d'accalmie entre la mi-août et la minovembre. Il est donc à noter que les pays africains ont une dynamique épidémique assez proche de celle d'autres pays dans le monde, et notamment d'Europe. Les 4 pays n'ayant eu qu'une vague épidémique tardive à partir de novembre ont tous la particularité d'être de petits pays et d'avoir été parmi les derniers à signaler leur premier cas de COVID-19 (entre le 21 mars et le 13 mai 2020), ce qui pourrait expliquer en partie l'absence d'épidémie avant l'automne. Ensuite, d'un point de vue géographique, les pays tropicaux et sahéliens semblent être ceux ayant eu les premières et deuxièmes vagues épidémiques les plus précoces, tandis que les pays les plus septentrionaux et australs ont plutôt tendance à avoir eu des vagues plus tardives dans l'année ou plus prolongées dans le temps. Pour finir, cette difficulté de classer les pays africains entre eux, du fait de la grande diversité de scénarios observés, pourrait être un indicateur d'une dynamique épidémiologique imprévisible et déterministe, définissant en partie un système chaotique. Cette hypothèse est donc intéressante en soit, permettant de justifier l'usage de modèles basés sur la théorie du Chaos pour simuler la COVID-19.

Si les données de nouveaux cas/décès quotidiens utilisées pour la modélisation présentent de fortes variations d'un jour à l'autre (souvent dues à des retards de déclaration de cas, à des rattrapages et à des corrections rétrospectives de valeurs), les dérivées successives auront alors des variations de valeurs et de signes d'autant plus élevées. Or, la modélisation globale est une technique s'appuyant justement sur ces variations de données. Par conséquent, cette technique générera des modèles de mauvaise qualité ou en inadéquation avec la réalité si les variables fournies sont empreintes de bruits et d'irrégularités. Il faut donc limiter au maximum les allures en dent de scie des courbes épidémiologiques, et procéder à un pré-traitement des données adapté afin de s'affranchir des fortes irrégularités journalières. Le prétraitement décrit en début de manuscrit et appliqué à l'ensemble des séries temporelles analysées dans cette étude s'est avéré efficace, à la fois pour lisser les courbes, disposer d'une information robuste à l'échelle de temps hebdomadaire, mais mise à jour de façon journalière, tout en disposant d'une estimation rigoureuse de l'erreur associée. En effet, même les données issues du Cameroun et du Togo, pourtant très irrégulières, et souvent très éparses pour le Togo ont réussies à être préparées de manière à avoir des courbes lissées tout en restant très cohérentes avec les autres jeux de données. La modélisation globale a donc pu être appliquée à de nombreux pays, donnant des modèles assez robustes pour le Togo et parmi les meilleurs pour le Cameroun, alors que cela aurait été impossible à obtenir avec les données brutes. Une limitation apparait cependant en fin de période avec ce type de pré-traitement des données. Lorsque beaucoup de données journalières sont sans valeur durant les dernières semaines, les courbes rééchantillonnées peuvent évoluer vers des valeurs positives ou négatives infinies, du fait de l'absence de point par où passer à la fin de la fenêtre temporelle. Pour remédier à ce problème, lorsqu'il manquait des données pour le même jour de la semaine et sur une ou plusieurs des dernières semaines, le rééchantillonnage par splines n'était plus appliqué, remplacé par un rééchantillonnage linéaire vers la dernière donnée disponible de la série temporelle. Cela a eu pour conséquence de diminuer la qualité du lissage des courbes de dérivées vers les dernières valeurs. C'est un effet de bord très courant et fréquemment rencontré en traitement de signal, et cela n'a pas été très gênant dans l'analyse présente car nous n'avons pas appliqué de prévisions à partir des toutes dernières valeurs de la courbe. Cependant, ce problème est en cours de correction pour de futures analyses.

La technique de modélisation globale, s'appuyant sur la théorie du Chaos, a permis de trouver un grand nombre de modèles potentiels permettant de simuler les cas et les décès pour les 17 pays étudiés. Tous ne se valent pas, et plus de la moitié de la centaine de modèles étudiés ont une prédictibilité qui se dégrade rapidement dès 5 jours d'horizon de prédiction. Cependant, 28 modèles ont une capacité de prévision assez bonne, avec pour le meilleur modèle (simulant les cas au Cameroun), moins de 33 cas d'erreur à 10 jours d'horizon et pour un niveau de confiance de 95%. Généralement, les modèles ont eu le plus de difficultés à prévoir la dynamique de la COVID-19 au niveau des zones d'inflexion de la courbe, c'est-à-dire au tout début de l'accroissement exponentiel du nombre de cas/décès, ainsi qu'au pic de l'épidémie. Cette limitation tient essentiellement au fait de s'appuyer sur une variable unique, comme cela a pu être constaté en appliquant l'approche à une thématique environnementale tout à fait différente, Mangiarotti et al. 2019). Un autre fait notable est la plus grande difficulté à obtenir des modèles pour l'évolution du nombre de décès par rapport au nombre de cas d'infection : parmi les 28 meilleurs modèles obtenus, il n'y en a que 8 pour les décès, soit moins du tiers. Cela pourrait trouver son explication, au moins partiellement, par le nombre de déclarations de décès moins important que celui des cas, ce qui rend leurs courbes moins lisses, plus en dent de scie et qui complique le travail de modélisation globale, très dépendant du jeu de données qu'on lui fournit en entrée. Un autre élément d'explication pourrait relever d'un problème d'observabilité : les cas d'infection nous offrent une information plus directe sur la dynamique de l'épidémie que les décès, qui sont la conséquence d'une infection, puis d'une évolution défavorable de cette infection sur le plan clinique. Enfin, l'étude de la prédictibilité des modèles pour des données plus récentes n'a pas montré de détérioration particulière de la capacité prévisionnelle de ceux-ci lorsqu'ont été rajoutées 1 mois et demi de données supplémentaires (non utilisées au cours du processus de modélisation). Cependant, l'erreur absolue de prévision n'a été étudiée que pour des horizons de prédiction de 10 et 20 jours. Il est donc possible qu'une saturation de l'erreur de prévision ait été atteinte pour certains modèles avant 10 jours d'horizon de prédiction, et que la capacité prévisionnelle ait été altérée pour des horizons de prédiction plus petits. La période utilisée de validation des modèles n'était ici que de 1 mois et demi de données supplémentaires. Pour apprécier avec plus de confiance la robustesse des modèles, il serait préférable d'utiliser des périodes de validation plus longues, de l'ordre de plusieurs mois, dès qu'assez de données seront disponibles. Parmi les modèles conservés pour cette étude, les plus robustes n'ont pas été nécessairement ceux ayant été générés à partir du plus grand nombre de données et/ou des données les plus récentes. Il est important d'actualiser voire de recréer des modèles avec les nouvelles données au fur et à mesure que le temps passe, afin de conserver une bonne capacité prévisionnelle. Cependant, ce travail semble montrer que des modèles générés avec de données parfois anciennes et sur seulement quelques mois peuvent tout de même être assez robustes, et conserver leur capacité prévisionnelle au moins pendant quelques mois. Enfin, une dégradation de l'erreur absolue de prévision a été observée pour quelques modèles, généralement conçus à partir de données épidémiologiques relativement anciennes. Cela pourrait être un indicateur d'une dynamique évolutive de la maladie (apparition de variants du virus), de la facon d'appréhender l'épidémie, ou des mesures sanitaires (vaccinations, mesures de contrôles plus ou moins efficaces).

Un autre point important de ce travail est que la plupart des modèles qui ont été retenus lors de leur sélection pour leur bonne adéquation de leur portrait de phase avec celui observé et pour leur bonne prédictibilité, présentent un comportement chaotique. Les modèles chaotiques de dimension trois (car basés sur trois variables) sont une bonne approximation de la dynamique originale. Leur obtention apporte un bon élément de preuve (1) du déterminisme sous-jacent (les équations du modèle sont déterministes), (2) du petit nombre de variables en jeu (en première approximation), (3) de la forte sensibilité aux conditions initiales de la dynamique épidémiologique de la COVID-19 (une petite différence initiale, tel une contamination localisée, peut conduire à un développement exponentiel). On observe une certaine diversité de dynamiques épidémiologiques entre les modèles et les pays (comme l'illustrent les portraits de phases) qui peuvent résulter des spécificités associées aux contextes différents pour chaque pays (densité de population, facilité d'accès à l'eau, etc.) mais aussi de la non-stationnarité de la situation épidémiologique (évolution des mesures non-pharmaceutiques, migrations, etc...).

Une étude exploratoire de la stationnarité des paramètres d'un des modèles a été réalisée. Le but était d'essayer de comprendre le sens de certains paramètres et de relier la valeur que prenaient certains coefficients au cours du temps à des variables connues telles que l'indice de rigueur des mesures gouvernementales. L'analyse des 8 paramètres composant l'équation polynomiale du modèle des cas en Ethiopie a permis de mettre en évidence un comportement très différent entre deux périodes correspondant l'une à une phase de milieu de flambée épidémique (pendant la première vague), l'autre à un début de flambée (début de la deuxième vague). Cependant, cette analyse n'a toutefois pas permis de révéler de corrélation significative avec l'indice de rigueur des mesures gouvernementales, indice qui s'avère d'ailleurs être entièrement construit, et basé au moins partiellement sur des critères subjectifs et par conséquent assez difficile à valider. Une autre étude de stationnarité, non décrite dans ce rapport, a aussi été faite avec cette fois un modèle décrivant les cas en Libye. La prédictibilité de ce modèle était moindre en comparaison de celle de l'Ethiopie mais le modèle était plus concis (seulement 7 paramètres). On disposait de plus pour ce pays d'informations sur la mobilité des populations ainsi que sur le nombre de tests effectués chaque jour, en plus de l'indice de rigueur des mesures gouvernementales. Cependant, là non plus, aucune corrélation significative n'a pu être relevée entre ces variables et les coefficients des paramètres. Ces études exploratoires de l'évolution des coefficients sont difficiles, car les paramètres des modèles peuvent être liés entre eux et évoluer simultanément dans le temps. Ainsi, une variation de valeur d'un coefficient devrait être considérée non pas uniquement selon l'évolution des courbes épidémiques, mais aussi en regard de tous les autres paramètres. Les conclusions de ces recherches sur la stationnarité ne sont donc que partielles, et un travail plus poussé pourrait peut-être permettre de mieux comprendre le sens des paramètres composant les modèles étudiés dans ce rapport.

Un des objectifs énoncés au début de ce travail était de savoir s'il était possible de transposer un modèle trouvé pour un pays à un autre pays. En l'état, cela s'avère compromis du fait de la grande variabilité des dynamiques épidémiologiques observée entre pays. De plus, des modèles ont été trouvés pour chaque pays étudié. Même si les modèles de certains pays, comme la Zambie, sont moyennement bons en termes de prévision, utiliser un modèle issu d'un autre pays amènerait à des erreurs de prévision plus importantes, du fait de dynamiques moins en adéquation avec celle propre au pays considéré. En plus de dynamiques différentes, les échelles de valeurs de cas et décès varient entre pays, même après normalisation. Un facteur de correction serait donc nécessaire en cas d'utilisation d'un modèle étranger pour améliorer la correspondance des valeurs prédites à celles du pays étudié. Un autre objectif de l'étude était de voir la capacité d'utilisation des modèles trouvés pour simuler l'émergence de futures maladies. Là aussi, il est compliqué de pouvoir statuer, car les modèles trouvés sont totalement issus des données fournies. Or, chaque maladie évolue différemment, même au sein d'un même pays. Ces modèles pourraient donc être utilisés pour le même pays et pour une MIE différente, mais leur prédictibilité s'en retrouverait possiblement très diminuée. L'intérêt serait surtout au tout début de l'épidémie, afin de générer des prévisions le temps que d'autres modèles (globaux ou basés sur d'autres techniques), plus robustes, soit créés.

Pour finir, on peut citer un biais lié aux données utilisées, et ayant eu probablement un impact sur les modèles générés. Pour chacun des pays sélectionnés, une recherche plus approfondie de la répartition temporelle et géographique des cas (et de la population) a été menée, afin d'avoir une vision régionale et non plus uniquement nationale de la dynamique épidémiologique de la COVID-19. Ce travail a mis en lumière certaines limites à la description d'une épidémie à partir de données uniquement nationales : la répartition spatiale des cas et des morts de la COVID-19 n'était pas nécessairement homogène en comparaison de celle de la population. Pour ne citer qu'un exemple, au Nigéria, presque la moitié des cas enregistrés étaient concentrés dans la région de Lagos, la plus grosse mégalopole du pays, qui représente 10% de la population totale. Ce biais est d'autant plus important que le pays étudié est grand. Cela a possiblement réduit la robustesse des modèles développés, qui se sont basés sur une dynamique épidémiologique moyenne nationale cachant parfois une importante hétérogénéité régionale.
Conclusion

Ce travail de modélisation globale de la dynamique de la COVID-19 a contribué à prouver le potentiel encore rarement exploité en épidémiologie de la théorie du Chaos. Les modèles obtenus à partir d'une seule variable, les cas ou décès journaliers déclarés, parfois sur une fenêtre temporelle assez courte, de l'ordre de quelques mois, ont montré pour une partie d'entre eux une capacité à prévoir à court terme l'évolution de la maladie avec une erreur limitée. L'atout de cette technique consiste en l'absence d'hypothèses fortes formulées pour concevoir le modèle, celui-ci se basant uniquement sur la dynamique dans l'espace des phases de la variable étudiée. Cette technique est donc intéressante à utiliser dans le cas de MIE pour lesquelles de nombreux paramètres épidémiologiques sont encore inconnus, ou pour des maladies dont la dynamique semble chaotique (ce qui est vraisemblablement le cas pour un grand nombre d'entre elles), tel que démontré ici pour la COVID-19 en Afrique.

Après la préparation des données, afin de lisser les courbes épidémiologiques, des modèles globaux satisfaisants ont pu être conçus pour la quasi-totalité des pays étudiés, alors que les variables utilisées s'avéraient parfois pauvres en données ou assez bruitées. Cette technique de modélisation apparait donc tout à fait utilisable de façon pratique et en conditions réelles, à partir des données de qualité très variable, à condition d'un bon pré-traitement préalable.

Les modèles globaux obtenus pour ces pays africains peuvent être utilisés de différentes manières, comme illustré dans ce travail. Lorsque leur prédictibilité est bonne, ils peuvent permettre de simuler à court terme l'évolution de la maladie. Ils peuvent, à plus long terme, mettre en lumière l'évolution de la dynamique de la maladie si celle-ci change en cours d'épidémie, en raison par exemple de l'apparition de variants du virus, d'interventions non-pharmaceutiques d'efficacité différente, ou par l'effet de la vaccination des populations. Une troisième utilisation possible de ces modèles serait pour comprendre, de manière rétrospective, la dynamique de la maladie. En effet, si les modèles arrivent à bien approcher la dynamique épidémiologique de la COVID-19, une étude des termes des équations différentielles de ces modèles peut amener à mieux comprendre les facteurs impliqués dans l'évolution de la maladie.

Ce type de modélisation pourrait être appliquée à plus de pays, ainsi qu'à d'autres maladies. Il est aussi envisageable d'utiliser les modèles conçus pour la COVID-19 afin de prévoir l'évolution de nouvelles maladies émergentes, durant le stade précoce de celles-ci, quand il manque à la fois des données et des connaissances sur la maladie. Pour conclure, cette technique de modélisation doit être vue comme une alternative à d'autres approches. Chaque façon de modéliser les maladies possède ses propres qualités, défauts et limites, et c'est en approchant les dynamiques épidémiologiques par différents types de modèles, complémentaires les uns des autres, qu'une meilleure compréhension des épidémies sera possible.

Un dernier résultat important de ce travail a consisté en la mise en lumière de deux facteurs permettant de réduire les grandes différences de valeurs observées entre pays pour les cas d'infection et de décès dus à la COVID-19: la proportion de personnes âgées de plus de 60 ans (rendant compte d'un effet réel) et le nombre de lits d'hôpital par habitant (relatif à une sous-estimation). Ces facteurs ont l'intérêt d'être disponibles pour la majorité des pays, et sont apparus comme très efficace pour réduire les écarts entre pays, autant en Afrique (permettant une réduction d'un facteur 2 à 5,6 de la dispersion maximale par rapport au nombre de cas et de décès par habitant) que dans le monde dans sa globalité.

Bibliographie

ADEKUNLE, Ibrahim Ayoade, ONANUGA, Abayomi Toyin, AKINOLA, Olanrewaju Olugbenga et OGUNBANJO, Olakitan Wahab, 2020. Modelling spatial variations of coronavirus disease (COVID-19) in Africa. In: *Science of The Total Environment*. 10 août 2020. Vol. 729, pp. 138998. DOI 10.1016/j.scitotenv.2020.138998.

AFRICAN UNION, 2020. *Impact of the coronavirus (COVID-19) on the African economy* [en ligne]. 2020. S.l.: s.n. Disponible à l'adresse : https://www.tralac.org/documents/resources/covid-19/3218-impact-of-the-coronavirus-covid-19-on-the-african-economy-african-union-report-april-2020/file.html.

ALJAZEERA NEWS, 2020. Africa coronavirus cases could hit 10 million in six months: WHO COVID-19 cases on the continent may shoot up from thousands to millions if models are accurate, UN health agency says. In : [en ligne]. 17 avril 2020. [Consulté le 2 mars 2021]. Disponible à l'adresse : https://www.aljazeera.com/news/2020/4/17/africa-coronavirus-cases-could-hit-10-million-in-sixmonths-who.

BAMGBOYE, Ebun L., OMIYE, Jesutofunmi A., AFOLARANMI, Oluwasegun J., DAVIDS, Mogamat Razeen, TANNOR, Elliot Koranteng, WADEE, Shoyab, NIANG, Abdou, WERE, Anthony et NAICKER, Saraladevi, 2020. COVID-19 Pandemic: Is Africa Different? In : *Journal of the National Medical Association* [en ligne]. 3 novembre 2020. [Consulté le 15 janvier 2021]. DOI 10.1016/j.jnma.2020.10.001. Disponible à l'adresse : http://www.sciencedirect.com/science/article/pii/S002796842030345X.

BATCHI-BOUYOU, Armel Landry, LOBALOBA, Line, NDOUNGA, Matthieu, VOUVOUNGUI, Jeannhey Christevy, MFOUTOU, Chastel Mapanguy, BOUMPOUTOU, Kamal Rauchelvy et NTOUMI, Francine, 2020. High SARS-COV2 IgG/IGM seroprevalence in asymptomatic Congolese in Brazzaville, the Republic of Congo. In : *International Journal of Infectious Diseases* [en ligne]. 25 décembre 2020. [Consulté le 18 janvier 2021]. DOI 10.1016/j.ijid.2020.12.065. Disponible à l'adresse : http://www.sciencedirect.com/science/article/pii/S1201971220325893.

BELL, David, HANSEN, Kristian Schultz, KIRAGGA, Agnes N., KAMBUGU, Andrew, KISSA, John et MBONYE, Anthony K., 2020. Predicting the Impact of COVID-19 and the Potential Impact of the Public Health Response on Disease Burden in Uganda. In : *The American Journal of Tropical Medicine and Hygiene*. septembre 2020. Vol. 103, n° 3, pp. 1191-1197. DOI 10.4269/ajtmh.20-0546.

BONANAD, Clara, GARCÍA-BLAS, Sergio, TARAZONA-SANTABALBINA, Francisco, SANCHIS, Juan, BERTOMEU-GONZÁLEZ, Vicente, FÁCILA, Lorenzo, ARIZA, Albert, NÚÑEZ, Julio et CORDERO, Alberto, 2020. The Effect of Age on Mortality in Patients With COVID-19: A Meta-Analysis With 611,583 Subjects. In : *Journal of the American Medical Directors Association*. 1 juillet 2020. Vol. 21, n° 7, pp. 915-918. DOI 10.1016/j.jamda.2020.05.045.

CHEN, Xinguang et YU, Bin, 2020. First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model. In : *Global Health Research and Policy*. 2 mars 2020. Vol. 5, n° 1, pp. 7. DOI 10.1186/s41256-020-00137-4.

CHIBWANA, Marah Grace, JERE, Khuzwayo Chidiwa, KAMNG'ONA, Raphael, MANDOLO, Jonathan, KATUNGA-PHIRI, Vincent, TEMBO, Dumizulu, MITOLE, Ndaona, MUSASA, Samantha, SICHONE, Simon, LAKUDZALA, Agness, SIBALE, Lusako, MATAMBO, Prisca, KADWALA, Innocent, BYRNE, Rachel Louise, MBEWE, Alice, HENRION, Marc Y. R., MORTON, Ben, PHIRI, Chimota, MALLEWA, Jane, MWANDUMBA, Henry C., ADAMS, Emily R., GORDON, Stephen B. et JAMBO, Kondwani Charles, 2020. High SARS-CoV-2 seroprevalence in Health Care Workers but

relatively low numbers of deaths in urban Malawi. In : *medRxiv*. 5 août 2020. pp. 2020.07.30.20164970. DOI 10.1101/2020.07.30.20164970.

CYRANOSKI, David, 2020. When will the coronavirus outbreak peak? In : *Nature* [en ligne]. 18 février 2020. [Consulté le 14 janvier 2021]. DOI 10.1038/d41586-020-00361-5. Disponible à l'adresse : https://www.nature.com/articles/d41586-020-00361-5.

DALZIEL, Benjamin D., BJØRNSTAD, Ottar N., PANHUIS, Willem G. van, BURKE, Donald S., METCALF, C. Jessica E. et GRENFELL, Bryan T., 2016. Persistent Chaos of Measles Epidemics in the Prevaccination United States Caused by a Small Change in Seasonal Transmission Patterns. In : *PLoS Computational Biology* [en ligne]. 1 février 2016. [Consulté le 14 janvier 2021]. Disponible à l'adresse : https://doaj.org.

DAVID, Pierre-Marie et DÉVÉDEC, Nicolas Le, 2019. Preparedness for the next epidemic: health and political issues of an emerging paradigm. In : *Critical Public Health*. 27 mai 2019. Vol. 29, n° 3, pp. 363-369. DOI 10.1080/09581596.2018.1447646.

DJIDJOU-DEMASSE, Ramsès, SELINGER, Christian et SOFONEA, Mircea T., 2020. Épidémiologie mathématique et modélisation de la pandémie de Covid-19 : enjeux et diversité. In : *Revue Francophone des Laboratoires*. 1 novembre 2020. Vol. 2020, n° 526, pp. 63-69. DOI 10.1016/S1773-035X(20)30315-4.

DONG, Ensheng, DU, Hongru et GARDNER, Lauren, 2020. An interactive web-based dashboard to track COVID-19 in real time. In: *The Lancet Infectious Diseases*. 1 mai 2020. Vol. 20, n° 5, pp. 533-534. DOI 10.1016/S1473-3099(20)30120-1.

DUBÉ, C., GARNER, G., STEVENSON, M., SANSON, R., ESTRADA, C. et WILLEBERG, P., 2007. *L'utilisation de modèles épidémiologiques pour la gestion des maladies animales* [en ligne]. 2007. S.1. : s.n. Disponible à l'adresse : https://www.oie.int/doc/ged/D4534.PDF. Conf. OIE 2007, 1-11

ELIAS, Christelle, SEKRI, Abel, LEBLANC, Pierre, CUCHERAT, Michel et VANHEMS, Philippe, 2021. The incubation period of COVID-19: A meta-analysis. In : *International Journal of Infectious Diseases*. 1 mars 2021. Vol. 104, pp. 708-710. DOI 10.1016/j.ijid.2021.01.069.

EL-SAYED, Amr et KAMEL, Mohamed, 2020. Climatic changes and their role in emergence and reemergence of diseases. In : *Environmental Science and Pollution Research*. 1 juin 2020. Vol. 27, n° 18, pp. 22336-22352. DOI 10.1007/s11356-020-08896-w.

EPIDEMIC INTELLIGENCE FROM OPEN SOURCES INITIATIVE, 2020. EIOS COVID-19 Case Count Dashboard. In : [en ligne]. 2020. [Consulté le 20 avril 2021]. Disponible à l'adresse : https://portal.who.int/report/eios-covid19-counts/.

FERGUSON, N, LAYDON, D, NEDJATI GILANI, G, IMAI, N, AINSLIE, K, BAGUELIN, M, BHATIA, S, BOONYASIRI, A, CUCUNUBA PEREZ, ZULMA, CUOMO-DANNENBURG, G, DIGHE, A, DORIGATTI, I, FU, H, GAYTHORPE, K, GREEN, W, HAMLET, A, HINSLEY, W, OKELL, L, VAN ELSLAND, S, THOMPSON, H, VERITY, R, VOLZ, E, WANG, H, WANG, Y, WALKER, P, WINSKILL, P, WHITTAKER, C, DONNELLY, C, RILEY, S et GHANI, A, 2020. *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand* [en ligne]. S.I. Imperial College London. [Consulté le 12 janvier 2021]. Disponible à l'adresse : http://spiral.imperial.ac.uk/handle/10044/1/77482.

FERNANDES, T., 2020. Chaotic model for COVID-19 growth factor. In : *Research on Biomedical Engineering* [en ligne]. 26 juillet 2020. [Consulté le 11 janvier 2021]. DOI 10.1007/s42600-020-00077-5. Disponible à l'adresse : https://doi.org/10.1007/s42600-020-00077-5.

GAYE, Bamba, KHOURY, Stéphanie, CENE, Crystal W., KINGUE, Samuel, N'GUETTA, Roland, LASSALE, Camille, BALDÉ, Dadhi, DIOP, Ibrahima Bara, DOWD, Jennifer Beam, MILLS, Melinda C. et JOUVEN, Xavier, 2020. Socio-demographic and epidemiological consideration of Africa's COVID-19 response: what is the possible pandemic course? In : *Nature Medicine*. juillet 2020. Vol. 26, n° 7, pp. 996-999. DOI 10.1038/s41591-020-0960-y.

GHOSH, Debajyoti, BERNSTEIN, Jonathan A et MERSHA, Tesfaye B, 2020. COVID-19 pandemic: The African paradox. In: *Journal of Global Health* [en ligne]. 2020. Vol. 10, n° 2. [Consulté le 15 janvier 2021]. DOI 10.7189/jogh.10.020348. Disponible à l'adresse : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7506193/.

GOLDBLATT, David, JOHNSON, Marina, FALUP-PECURARIU, Oana, IVASKEVICIENE, Inga, SPOULOU, Vana, TAMM, Eda, WAGNER, Michael, ZAR, Heather J., BLEOTU, Laura, IVASKEVICIUS, Rimvydas, PAPADATOU, Ioanna, JÕGI, Piia, LISCHKA, Julia, FRANCKLING-SMITH, Zoe, ISAROVA, Dana, GRANDJEAN, Louis et ZAVADSKA, Dace, 2021. Cross Sectional Prevalence of SARS-CoV-2 antibodies in Health Care Workers in Paediatric Facilities in Eight Countries. In : *Journal of Hospital Infection* [en ligne]. 7 janvier 2021. [Consulté le 18 janvier 2021]. DOI 10.1016/j.jhin.2020.12.019. Disponible à l'adresse : http://www.sciencedirect.com/science/article/pii/S0195670121000049.

GOOGLE LLC, 2020. Rapports sur la mobilité de la communauté - COVID-19. In : *COVID-19 Community Mobility Report* [en ligne]. 2020. [Consulté le 15 février 2021]. Disponible à l'adresse : https://www.google.com/covid19/mobility/.

GOUESBET, G. et LETELLIER, C., 1994. Global vector-field reconstruction by using a multivariate polynomial L2 approximation on nets. In : *Physical Review E*. 1 juin 1994. Vol. 49, n° 6, pp. 4955-4972. DOI 10.1103/PhysRevE.49.4955.

GWENZI, Willis, 2021. Dangerous liaisons? As the COVID-19 wave hits Africa with potential for novel transmission dynamics: a perspective. In: *Journal of Public Health* [en ligne]. 2 janvier 2021. [Consulté le 18 janvier 2021]. DOI 10.1007/s10389-020-01467-w. Disponible à l'adresse : https://doi.org/10.1007/s10389-020-01467-w.

HALE, T., ANGRIST, N., CAMERON-BLAKE, E., HALLAS, L., KIRA, B., MAJUMDAR, S., PETHERICK, A., PHILLIPS, T., TATLOW, H. et WEBSTER, S., 2020. *Oxford COVID-19 Government Response Tracker* [en ligne]. 2020. S.I.: Blavatnik School of Government. [Consulté le 15 février 2021]. Disponible à l'adresse : https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.

HE, Shaobo, PENG, Yuexi et SUN, Kehui, 2020. SEIR modeling of the COVID-19 and its dynamics. In : *Nonlinear Dynamics*. 1 août 2020. Vol. 101, n° 3, pp. 1667-1680. DOI 10.1007/s11071-020-05743-y.

HE, Xi, LAU, Eric H. Y., WU, Peng, DENG, Xilong, WANG, Jian, HAO, Xinxin, LAU, Yiu Chung, WONG, Jessica Y., GUAN, Yujuan, TAN, Xinghua, MO, Xiaoneng, CHEN, Yanqing, LIAO, Baolin, CHEN, Weilie, HU, Fengyu, ZHANG, Qing, ZHONG, Mingqiu, WU, Yanrong, ZHAO, Lingzhai, ZHANG, Fuchun, COWLING, Benjamin J., LI, Fang et LEUNG, Gabriel M., 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. In : *Nature Medicine*. mai 2020. Vol. 26, n° 5, pp. 672-675. DOI 10.1038/s41591-020-0869-5.

HOGAN, Alexandra B, JEWELL, Britta L, SHERRARD-SMITH, Ellie, VESGA, Juan F, WATSON, Oliver J, WHITTAKER, Charles, HAMLET, Arran, SMITH, Jennifer A, WINSKILL, Peter, VERITY, Robert, BAGUELIN, Marc, LEES, John A, WHITTLES, Lilith K, AINSLIE, Kylie E C, BHATT, Samir, BOONYASIRI, Adhiratha, BRAZEAU, Nicholas F, CATTARINO, Lorenzo, COOPER, Laura V, COUPLAND, Helen, CUOMO-DANNENBURG, Gina, DIGHE, Amy, DJAAFARA, Bimandra A,

DONNELLY, Christl A, EATON, Jeff W, VAN ELSLAND, Sabine L, FITZJOHN, Richard G, FU, Han, GAYTHORPE, Katy A M, GREEN, William, HAW, David J, HAYES, Sarah, HINSLEY, Wes, IMAI, Natsuko, LAYDON, Daniel J, MANGAL, Tara D, MELLAN, Thomas A, MISHRA, Swapnil, NEDJATI-GILANI, Gemma, PARAG, Kris V, THOMPSON, Hayley A, UNWIN, H Juliette T, VOLLMER, Michaela A C, WALTERS, Caroline E, WANG, Haowei, WANG, Yuanrong, XI, Xiaoyue, FERGUSON, Neil M, OKELL, Lucy C, CHURCHER, Thomas S, ARINAMINPATHY, Nimalan, GHANI, Azra C, WALKER, Patrick G T et HALLETT, Timothy B, 2020. Potential impact of the COVID-19 pandemic on HIV, tuberculosis, and malaria in low-income and middle-income countries: a modelling study. In : *The Lancet Global Health*. 1 septembre 2020. Vol. 8, n° 9, pp. e1132-e1141. DOI 10.1016/S2214-109X(20)30288-6.

HOLMDAHL, Inga et BUCKEE, Caroline, 2020. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. In : *New England Journal of Medicine*. 23 juillet 2020. Vol. 383, n° 4, pp. 303-305. DOI 10.1056/NEJMp2016822.

HUSSEIN, Mohammad, TORAIH, Eman, ELSHAZLI, Rami, FAWZY, Manal, HOUGHTON, August, TATUM, Danielle, KILLACKEY, Mary, KANDIL, Emad et DUCHESNE, Juan, 2021. Meta-analysis on Serial Intervals and Reproductive Rates for SARS-CoV-2. In : *Annals of Surgery*. mars 2021. Vol. 273, n° 3, pp. 416-423. DOI 10.1097/SLA.00000000004400.

INTERNATIONALBANKFORRECONSTRUCTIONANDDEVELOPMENTetINTERNATIONAL DEVELOPMENT ASSOCIATION, 2020. The World BankGroup DataBank. In :[en ligne].2020.[Consulté le 19 avril 2021].Disponibleàl'adresse :https://databank.worldbank.org/home.aspx.

IPBES, 2020. Workshop Report on Biodiversity and Pandemics of the Intergovernmental Platform on Biodiversity and Ecosystem Services. In : [en ligne]. octobre 2020. [Consulté le 26 février 2021]. DOI 10.5281/zenodo.4147317. Disponible à l'adresse : http://ipbes.net/pandemics.

JONES, Andrew et STRIGUL, Nikolay, 2020. Is spread of COVID-19 a chaotic epidemic? In : *Chaos, Solitons & Fractals*. octobre 2020. pp. 110376. DOI 10.1016/j.chaos.2020.110376.

KERMACK, William Ogilvy et MCKENDRICK, A. G., 1927. A contribution to the mathematical theory of epidemics. In : *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1 août 1927. Vol. 115, n° 772, pp. 700-721. DOI 10.1098/rspa.1927.0118.

KUNIYA, Toshikazu, 2020. Prediction of the Epidemic Peak of Coronavirus Disease in Japan, 2020. In : *Journal of Clinical Medicine*. mars 2020. Vol. 9, n° 3, pp. 789. DOI 10.3390/jcm9030789.

LAI, Shengjie, RUKTANONCHAI, Nick W., ZHOU, Liangcai, PROSPER, Olivia, LUO, Wei, FLOYD, Jessica R., WESOLOWSKI, Amy, SANTILLANA, Mauricio, ZHANG, Chi, DU, Xiangjun, YU, Hongjie et TATEM, Andrew J., 2020. Effect of non-pharmaceutical interventions to contain COVID-19 in China. In: *Nature*. septembre 2020. Vol. 585, n° 7825, pp. 410-413. DOI 10.1038/s41586-020-2293-x.

LETELLIER, Christophe, AGUIRRE, Luis A. et FREITAS, U. S., 2009. Frequently asked questions about global modeling. In: *Chaos (Woodbury, N.Y.).* juin 2009. Vol. 19, n° 2, pp. 023103. DOI 10.1063/1.3125705.

LI, Jie, HUANG, Daniel Q., ZOU, Biyao, YANG, Hongli, HUI, Wan Zi, RUI, Fajuan, YEE, Natasha Tang Sook, LIU, Chuanli, NERURKAR, Sanjna Nilesh, KAI, Justin Chua Ying, TENG, Margaret Li Peng, LI, Xiaohe, ZENG, Hua, BORGHI, John A., HENRY, Linda, CHEUNG, Ramsey et NGUYEN, Mindie H., 2021. Epidemiology of COVID-19: A systematic review and meta-analysis of clinical

characteristics, risk factors, and outcomes. In : *Journal of Medical Virology*. 2021. Vol. 93, n° 3, pp. 1449-1458. DOI https://doi.org/10.1002/jmv.26424.

LIN, Qianying, ZHAO, Shi, GAO, Daozhou, LOU, Yijun, YANG, Shu, MUSA, Salihu S., WANG, Maggie H., CAI, Yongli, WANG, Weiming, YANG, Lin et HE, Daihai, 2020. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. In: *International Journal of Infectious Diseases*. 1 avril 2020. Vol. 93, pp. 211-216. DOI 10.1016/j.ijid.2020.02.058.

LLOYD-SMITH, James O., FUNK, Sebastian, MCLEAN, Angela R., RILEY, Steven et WOOD, James L. N., 2015. Nine challenges in modelling the emergence of novel pathogens. In : *Epidemics*. 1 mars 2015. Vol. 10, pp. 35-39. DOI 10.1016/j.epidem.2014.09.002.

LONE, Shabir Ahmad et AHMAD, Aijaz, 2020. COVID-19 pandemic – an African perspective. In : *Emerging Microbes & Infections*. 22 mai 2020. Vol. 9, n° 1, pp. 1300-1308. DOI 10.1080/22221751.2020.1775132.

LU, Fred S., NGUYEN, Andre T., LINK, Nick et SANTILLANA, Mauricio, 2020. Estimating the Prevalence of COVID-19 in the United States: Three Complementary Approaches. In : *medRxiv*. 23 avril 2020. pp. 2020.04.18.20070821. DOI 10.1101/2020.04.18.20070821.

MANCHEIN, Cesar, BRUGNAGO, Eduardo L., DA SILVA, Rafael M., MENDES, Carlos F. O. et BEIMS, Marcus W., 2020. Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies. In : *Chaos: An Interdisciplinary Journal of Nonlinear Science*. avril 2020. Vol. 30, n° 4, pp. 041102. DOI 10.1063/5.0009454.

MANGIAROTTI, S., COUDRET, R., DRAPEAU, L. et JARLAN, L., 2012. Polynomial search and global modeling: Two algorithms for modeling chaos. In : *Physical Review E*. 8 octobre 2012. Vol. 86, n° 4, pp. 046205. DOI 10.1103/PhysRevE.86.046205.

MANGIAROTTI, S. et HUC, M., 2020. Covid: Covid article codes. R. Toulouse, France : s.n.

MANGIAROTTI, S., PEYRE, M., ZHANG, Y., HUC, M., ROGER, F. et KERR, Y., 2020a. Chaos theory applied to the outbreak of COVID-19: an ancillary approach to decision making in pandemic context. In : *Epidemiology & Infection* [en ligne]. ed 2020. Vol. 148. [Consulté le 14 janvier 2021]. DOI 10.1017/S0950268820000990. Disponible à l'adresse : https://www.cambridge.org/core/journals/epidemiology-and-infection/article/chaos-theory-applied-to-the-outbreak-of-covid19-an-ancillary-approach-to-decision-making-in-pandemic-context/07A53305F996B2B9498F25606DB83B0B#.

MANGIAROTTI, S., HUC, M., LE JEAN, F., CHASSAN, M. et DRAPEAU, L., 2020b. *GPoM*: *Generalized Polynomial Modelling* [en ligne]. R. Toulouse, France : s.n. Disponible à l'adresse : https://CRAN.R-project.org/package=GPoM.

MANGIAROTTI, S., SAEZ ANDREU, M., PEYRE, M., LEBLANC, P., MUDDU, S., MOLERO, B., KERR, Y., ANAND, S., KARBOU, F., RICHAUME, P., GASCOIN, S., LEROUX, D. et HUC, M., 2017. Modélisation globale multi-échelle & lent-rapide de dynamiques environnementales - applications à l'épidémiologie, à l'hydrologie karstique et à l'hydrologie nivale. In : *Journée de restitution du défi Interfaces Interdiscipliaires Numérique et Théorique (InFiNiTi)*. Poster. Institut Henri Poincaré, Paris. 3 novembre 2017.

MANGIAROTTI, S., ZHANG, Y. et LEBLANC, M., 2019. Chaos theory applied to the modelling of karst springs: first results from univariate time series. In : *Hydrogeology Journal*. 1 septembre 2019. Vol. 27, n° 6, pp. 2027-2043. DOI 10.1007/s10040-019-01971-8.

MANGIAROTTI, Sylvain, 2015. Low dimensional chaotic models for the plague epidemic in Bombay (1896–1911). In: *Chaos, Solitons & Fractals.* 1 décembre 2015. Vol. 81, pp. 184-196. DOI 10.1016/j.chaos.2015.09.014.

MANGIAROTTI, Sylvain et HUC, Mireille, 2019. Can the original equations of a dynamical system be retrieved from observational time series? In : *Chaos: An Interdisciplinary Journal of Nonlinear Science*. février 2019. Vol. 29, n° 2, pp. 023133. DOI 10.1063/1.5081448.

MANGIAROTTI, Sylvain, PEYRE, Marisa et HUC, Mireille, 2016. A chaotic model for the epidemic of Ebola virus disease in West Africa (2013-2016). In : *Chaos (Woodbury, N.Y.)*. novembre 2016. Vol. 26, n° 11, pp. 113112. DOI 10.1063/1.4967730.

MAPCHART, 2021. Create your own Custom Map. In: *MapChart* [en ligne]. 2021. [Consulté le 25 avril 2021]. Disponible à l'adresse : https://mapchart.net/index.html.

MBOW, Moustapha, LELL, Bertrand, JOCHEMS, Simon P., CISSE, Badara, MBOUP, Souleymane, DEWALS, Benjamin G., JAYE, Assan, DIEYE, Alioune et YAZDANBAKHSH, Maria, 2020. COVID-19 in Africa: Dampening the storm? In: *Science*. 7 août 2020. Vol. 369, n° 6504, pp. 624-626. DOI 10.1126/science.abd3902.

MIDDELBURG, Rutger A. et ROSENDAAL, Frits R., 2020. COVID-19: How to make betweencountry comparisons. In: *International Journal of Infectious Diseases*. 1 juillet 2020. Vol. 96, pp. 477-481. DOI 10.1016/j.ijid.2020.05.066.

MORSE, Stephen S., MAZET, Jonna A. K., WOOLHOUSE, Mark, PARRISH, Colin R., CARROLL, Dennis, KARESH, William B., ZAMBRANA-TORRELIO, Carlos, LIPKIN, W. Ian et DASZAK, Peter, 2012. Prediction and prevention of the next pandemic zoonosis. In : *Lancet (London, England)*. 1 décembre 2012. Vol. 380, n° 9857, pp. 1956-1965. DOI 10.1016/S0140-6736(12)61684-5.

MOUGENI, Fabrice, MANGABOULA, Ance et LELL, Bertrand, 2020. The potential effect of the African population age structure on COVID-19 mortality. In: *medRxiv*. 21 mai 2020. pp. 2020.05.19.20106914. DOI 10.1101/2020.05.19.20106914.

NGUIMKEU, Pierre et TADADJEU, Sosson, 2021. Why is the number of COVID-19 cases lower than expected in Sub-Saharan Africa? A cross-sectional analysis of the role of demographic and geographic factors. In: *World Development*. 1 février 2021. Vol. 138, pp. 105251. DOI 10.1016/j.worlddev.2020.105251.

NKENGASONG, John N. et MANKOULA, Wessam, 2020. Looming threat of COVID-19 infection in Africa: act collectively, and fast. In : *The Lancet*. 14 mars 2020. Vol. 395, n° 10227, pp. 841-842. DOI 10.1016/S0140-6736(20)30464-5.

OGUNLEYE, Olayinka O., BASU, Debashis, MUELLER, Debjani, SNEDDON, Jacqueline, SEATON, R. Andrew, YINKA-OGUNLEYE, Adesola F., WAMBOGA, Joshua, MILJKOVIĆ, Nenad, MWITA, Julius C., RWEGERERA, Godfrey Mutashambara, MASSELE, Amos, PATRICK, Okwen, NIBA, Loveline Lum, NSAIKILA, Melaine, RASHED, Wafaa M., HUSSEIN, Mohamed Ali, HEGAZY, Rehab, AMU, Adefolarin A., BOAHEN-BOATEN, Baffour Boaten, MATSEBULA, Zinhle, GWEBU, Prudence, CHIRIGO, Bongani, MKHABELA, Nongabisa, DLAMINI, Tenelisiwe, SITHOLE, Siphiwe, MALAZA, Sandile, DLAMINI, Sikhumbuzo, AFRIYIE, Daniel, ASARE, George Awuku, AMPONSAH, Seth Kwabena, SEFAH, Israel, OLUKA, Margaret, GUANTAI, Anastasia N., OPANGA, Sylvia A., SARELE, Tebello Violet, MAFISA, Refeletse Keabetsoe, CHIKOWE, Ibrahim, KHULUZA, Felix, KIBUULE, Dan, KALEMEERA, Francis, MUBITA, Mwangana, FADARE, Joseph, SIBOMANA, Laurien, RAMOKGOPA, Gwendoline Malegwale, WHYTE, Carmen, MAIMELA, Tshegofatso, HUGO, Johannes, MEYER, Johanna C., SCHELLACK, Natalie, RAMPAMBA, Enos M., VISSER, Adel, ALFADL, Abubakr, MALIK, Elfatih M., MALANDE, Oliver Ombeva, KALUNGIA, Aubrey C., MWILA, Chiluba, ZARANYIKA, Trust, CHAIBVA, Blessmore Vimbai, OLARU, Ioana D., MASUKA, Nyasha, WALE, Janney, HWENDA, Lenias, KAMOGA, Regina, HILL, Ruaraidh, BARBUI, Corrado, BOCHENEK, Tomasz, KURDI, Amanj, CAMPBELL, Stephen, MARTIN, Antony P., PHUONG, Thuy Nguyen Thi, THANH, Binh Nguyen et GODMAN, Brian, 2020. Response to the Novel Corona Virus (COVID-19) Pandemic Across Africa: Successes, Challenges, and Implications for the Future. In : *Frontiers in Pharmacology* [en ligne]. 2020. Vol. 11. [Consulté le 15 janvier 2021]. DOI 10.3389/fphar.2020.01205. Disponible à l'adresse : https://www.frontiersin.org/articles/10.3389/fphar.2020.01205/full?&utm_source=Email_to_authors_ &utm_medium=Email&utm_content=T1_11.5e1_author&utm_campaign=Email_publication&field= &journalName=Frontiers_in_Pharmacology&id=568873.

PEARSON, Carl AB, SCHALKWYK, Cari Van, FOSS, Anna M., O'REILLY, Kathleen M., TEAM, SACEMA Modelling and Analysis Response, GROUP, CMMID COVID-19 working et PULLIAM, Juliet RC, 2020. Projected early spread of COVID-19 in Africa through 1 June 2020. In : *Eurosurveillance*. 7 mai 2020. Vol. 25, n° 18, pp. 2000543. DOI 10.2807/1560-7917.ES.2020.25.18.2000543.

R CORE TEAM, 2020. *R* : *A language and environment for statistical computing* [en ligne]. Vienna, Austria : R Foundation for Statistical Computing. Disponible à l'adresse : https://www.R-project.org/.

REINER, Robert C., BARBER, Ryan M., COLLINS, James K., ZHENG, Peng, ADOLPH, Christopher, ALBRIGHT, James, ANTONY, Catherine M., ARAVKIN, Aleksandr Y., BACHMEIER, Steven D., BANG-JENSEN, Bree, BANNICK, Marlena S., BLOOM, Sabina, CARTER, Austin, CASTRO, Emma, CAUSEY, Kate, CHAKRABARTI, Suman, CHARLSON, Fiona J., COGEN, Rebecca M., COMBS, Emily, DAI, Xiaochen, DANGEL, William James, EARL, Lucas, EWALD, Samuel B., EZALARAB, Maha, FERRARI, Alize J., FLAXMAN, Abraham, FROSTAD, Joseph Jon, FULLMAN, Nancy, GAKIDOU, Emmanuela, GALLAGHER, John, GLENN, Scott D., GOOSMANN, Erik A., HE, Jiawei, HENRY, Nathaniel J., HULLAND, Erin N., HURST, Benjamin, JOHANNS, Casey, KENDRICK, Parkes J., KHEMANI, Apurva, LARSON, Samantha Leigh, LAZZAR-ATWOOD, Alice, LEGRAND, Kate E., LESCINSKY, Haley, LINDSTROM, Akiaja, LINEBARGER, Emily, LOZANO, Rafael, MA, Rui, MÅNSSON, Johan, MAGISTRO, Beatrice, HERRERA, Ana M. Mantilla, MARCZAK, Laurie B., MILLER-PETRIE, Molly K., MOKDAD, Ali H., MORGAN, Julia Deryn, NAIK, Paulami, ODELL, Christopher M., O'HALLORAN, James K., OSGOOD-ZIMMERMAN, Aaron E., OSTROFF, Samuel M., PASOVIC, Maja, PENBERTHY, Louise, PHIPPS, Geoffrey, PIGOTT, David M., POLLOCK, Ian, RAMSHAW, Rebecca E., REDFORD, Sofia Boston, REINKE, Grace, ROLFE, Sam, SANTOMAURO, Damian Francesco, SHACKLETON, John R., SHAW, David H., SHEENA, Brittney S., SHOLOKHOV, Aleksei, SORENSEN, Reed J. D., SPARKS, Gianna, SPURLOCK, Emma Elizabeth, SUBART, Michelle L., SYAILENDRAWATI, Ruri, TORRE, Anna E., TROEGER, Christopher E., VOS, Theo, WATSON, Alexandrea, WATSON, Stefanie, WIENS, Kirsten E., WOYCZYNSKI, Lauren, XU, Liming, ZHANG, Jize, HAY, Simon I., LIM, Stephen S., MURRAY, Christopher J. L., et IHME COVID-19 FORECASTING TEAM, 2021. Modeling COVID-19 scenarios for the United States. In: Nature Medicine. janvier 2021. Vol. 27, n° 1, pp. 94-105. DOI 10.1038/s41591-020-1132-9.

RHODES, Tim, LANCASTER, Kari, LEES, Shelley et PARKER, Melissa, 2020. Modelling the pandemic: attuning models to their contexts. In : *BMJ Global Health*. 1 juin 2020. Vol. 5, n° 6, pp. e002914. DOI 10.1136/bmjgh-2020-002914.

RODA, Weston C., VARUGHESE, Marie B., HAN, Donglin et LI, Michael Y., 2020. Why is it difficult to accurately predict the COVID-19 epidemic? In : *Infectious Disease Modelling*. 1 janvier 2020. Vol. 5, pp. 271-281. DOI 10.1016/j.idm.2020.03.001.

ROSER, M., RITCHIE, H., ORTIZ-OSPINA, E. et HASELL, J., 2020. Coronavirus Pandemic (COVID-19). In : *Our World in Data* [en ligne]. 2020. [Consulté le 1 mars 2021]. Disponible à l'adresse : https://ourworldindata.org/excess-mortality-covid.

RUTAYISIRE, Erigene, NKUNDIMANA, Gerard, MITONGA, Honore K., BOYE, Alex et NIKWIGIZE, Solange, 2020. What works and what does not work in response to COVID-19 prevention and control in Africa. In: *International Journal of Infectious Diseases*. 1 août 2020. Vol. 97, pp. 267-269. DOI 10.1016/j.ijid.2020.06.024.

SHVEDA, K. et CARDOSO, K., 2021. Measuring Africa's Data Gap: The cost of not counting the dead. In : *BBC News* [en ligne]. 22 février 2021. [Consulté le 1 mars 2021]. Disponible à l'adresse : https://www.bbc.com/news/world-africa-55674139.

SRIDHAR, Devi et MAJUMDER, Maimuna S., 2020. Modelling the pandemic. In : *BMJ*. 21 avril 2020. Vol. 369, pp. m1567. DOI 10.1136/bmj.m1567.

TSO, For Yue, LIDENGE, Salum J., PEÑA, Phoebe B., CLEGG, Ashley A., NGOWI, John R., MWAISELAGE, Julius, NGALAMIKA, Owen, JULIUS, Peter, WEST, John T. et WOOD, Charles, 2021. High prevalence of pre-existing serological cross-reactivity against severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) in sub-Saharan Africa. In : *International Journal of Infectious Diseases*. 1 janvier 2021. Vol. 102, pp. 577-583. DOI 10.1016/j.ijid.2020.10.104.

UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, POPULATION DIVISION, 2019. *World Population Prospects* [en ligne]. 2019. S.l.: s.n. Disponible à l'adresse : https://population.un.org/wpp/Download/Standard/Population/.

UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, POPULATIONDIVISION, 2020. ST/ESA/SER.A/451: World Population Ageing 2020 Highlights [en ligne]. S.I.[Consulté le 15 janvier 2021].Disponibleàl'adresse :https://www.un.org/development/desa/pd/news/world-population-ageing-2020-highlights.

UYOGA, Sophie, ADETIFA, Ifedayo M. O., KARANJA, Henry K., NYAGWANGE, James, TUJU, James, WANJIKU, Perpetual, AMAN, Rashid, MWANGANGI, Mercy, AMOTH, Patrick, KASERA, Kadondi, NG'ANG'A, Wangari, ROMBO, Charles, YEGON, Christine, KITHI, Khamisi, ODHIAMBO, Elizabeth, ROTICH, Thomas, ORGUT, Irene, KIHARA, Sammy, OTIENDE, Mark, BOTTOMLEY, Christian, MUPE, Zonia N., KAGUCIA, Eunice W., GALLAGHER, Katherine E., ETYANG, Anthony, VOLLER, Shirine, GITONGA, John N., MUGO, Daisy, AGOTI, Charles N., OTIENO, Edward, NDWIGA, Leonard, LAMBE, Teresa, WRIGHT, Daniel, BARASA, Edwine, TSOFA, Benjamin, BEJON, Philip, OCHOLA-OYIER, Lynette I., AGWEYU, Ambrose, SCOTT, J. Anthony G. et WARIMWE, George M., 2021. Seroprevalence of anti–SARS-CoV-2 IgG antibodies in Kenyan blood donors. In : *Science*. 1 janvier 2021. Vol. 371, n° 6524, pp. 79-82. DOI 10.1126/science.abe1916.

VIBOUD, Cécile, BOËLLE, Pierre-Yves, CARRAT, Fabrice, VALLERON, Alain-Jacques et FLAHAULT, Antoine, 2003. Prediction of the Spread of Influenza Epidemics by the Method of Analogues. In : *American Journal of Epidemiology*. 15 novembre 2003. Vol. 158, n° 10, pp. 996-1006. DOI 10.1093/aje/kwg239.

ZANDVOORT, Kevin van, JARVIS, Christopher I., PEARSON, Carl A. B., DAVIES, Nicholas G., GROUP, CMMID COVID-19 working, RUSSELL, Timothy W., KUCHARSKI, Adam J., JIT, Mark, FLASCHE, Stefan, EGGO, Rosalind M. et CHECCHI, Francesco, 2020. Response strategies for COVID-19 epidemics in African settings: a mathematical modelling study. In : *medRxiv*. 3 mai 2020. pp. 2020.04.27.20081711. DOI 10.1101/2020.04.27.20081711.

Annexes

Afrique	Europe	Monde			
Algeria	Belgium	Afghanistan	Gambia	Oman	
Angola	Denmark	Albania	Georgia	Pakistan	
Benin	France	Algeria	Germany	Panama	
Botswana	Germany	Angola	Ghana	Paraguay	
Burkina Faso	Ireland	Antigua and Barbuda	Greece	Peru	
Burundi	Italy	Argentina	Grenada	Philippines	
Cabo Varda	Natharlands	Armonia	Guatamala	Poland	
Cabo verde	Norman	Australia	Guinea	Portugal	
Control African Danublia	Dortugal	Austria	Cuinca Dissou	Ostar	
Chad	Foltugal	Ausula	Guillea-Dissau	Qatai	
Chau	Span	Azerbaijan		Romania	
Comoros	Sweden	Banamas	Halu	Russia	
Congo	Switzerland	Bangladesh	Honduras	Rwanda	
Congo DR	United	Barbados	Hungary	Saint Lucia	
Côte d'Ivoire	Kingdom	Belarus	Iceland	Saint Vincent and the	
Djibouti		Belgium	India	Grenadines	
Egypt		Belize	Indonesia	Sao Tome and Principe	
Equatorial Guinea		Benin	Iran	Saudi Arabia	
Eritrea		Bhutan	Iraq	Senegal	
Eswatini		Bolivia	Ireland	Serbia	
Ethiopia		Bosnia and Herzegovina	Israel	Seychelles	
Gabon		Botswana	Italy	Sierra Leone	
Gambia		Brazil	Jamaica	Singapore	
Ghana		Brunei	Japan	Slovakia	
Guinea		Bulgaria	Jordan	Slovenia	
Guinea-Bissau		Burkina Faso	Kazakhstan	Somalia	
Kenva		Burundi	Kenva	South Africa	
Lesotho		Cabo Verde	Korea, South	Spain	
Liberia		Cambodia	Kuwait	Sri Lanka	
Libva		Cameroon	Kyrøyzstan	Sudan	
Madagascar		Canada	I atvia	Suriname	
Malawi		Central African Republic	Lehanon	Sweden	
Mali		Chad	Lesotho	Switzerland	
Mauritania		Chile	Liberia	Swrig	
Mouritius		China	Libua	Tajikistan	
Moroago		Colombia	Lithuania	Tanzania	
Morembique		Comoros	Luuramhaura	Theiland	
Nomihio		Control	Luxembourg		
Niner		Congo Conce DD	Malagascar	Togo	
Niger		Congo DR	Malawi Malawi	Trinidad and Tobago	
Nigena			Malaysia		
Rwanda		Cote d'Ivoire	Maldives	Turkey	
Sao Tome and Principe		Croatia	Mali	Uganda	
Senegal		Cuba	Malta	Ukraine	
Seychelles		Cyprus	Mauritania	United Arab Emirates	
Sierra Leone		Czech Republic	Mauritius	United Kingdom	
Somalia		Denmark	Mexico	United States of	
South Africa		Djibouti	Moldova	America	
Sudan		Dominican Republic	Mongolia	Uruguay	
Togo		Ecuador	Montenegro	Uzbekistan	
Tunisia		Egypt	Morocco	Venezuela	
Uganda		El Salvador	Mozambique	Viet Nam	
Zambia		Equatorial Guinea	Namibia	Yemen	
Zimbabwe		Eritrea	Nepal	Zambia	
		Estonia	Netherlands	Zimbabwe	
		Eswatini	New Zealand		
		Ethiopia	Nicaragua		
		Fiji	Niger		
		Finland	Nigeria		
		France	North Macedonia		
		Gabon	Norway		

Annexe 1 : Liste des pays utilisés pour les analyses de pondération

Annexe 2 : Boites à moustaches de la dispersion, selon les regroupements géographiques choisis, des facteurs de pondération et des données épidémiologiques avec ou sans pondération

Chaque point correspond à un pays, le rectangle coloré pour chaque regroupement de pays correspond aux limites des quartiles inférieur (Q1) et supérieur (Q3) et la barre centrale au sein de ce rectangle correspond à la médiane. La barre à l'extrémité inférieure, sous le rectangle, correspond au maximum entre la valeur minimale du regroupement de pays et la valeur de Q1-(1.5*(Q3-Q1)) équivalant à la distance sous Q1 de l'intervalle interquartile multipliée par 1.5. La barre à l'extrémité supérieure, au-dessus du rectangle, correspond au minimum entre la valeur maximale du regroupement de pays et la valeur maximale du regroupement de pays et la valeur maximale du regroupement de pays et la valeur de Q3+(1.5*(Q3-Q1))) équivalant à la distance au-dessus de Q3 de l'intervalle interquartile multipliée par 1.5.





Europe

80

60 40 20







Nombre total de cas (en millions) par million d'habitants de plus de 60 ans













Annexe 3 : Nuages de points des cas et décès totaux en fonction des pondérations mises en place, pour une sélection de pays africains et européens

Dans cette annexe, les pays européens étudiés sont les mêmes que précédemment (voir **annexe 1** pour leurs noms), tandis que les pays africains sélectionnés sont ceux pris pour la recherche de modèles (soit Afrique du Sud, Algérie, Cameroun, Côte d'Ivoire, Egypte, Ethiopie, Ghana, Kenya, Lybie, Maroc, Namibie, Nigeria, Sénégal, Togo, Tunisie, Zambie et Zimbabwe), afin d'avoir un nombre limité de points dans les graphiques. Les données utilisées datent du 15/03/2021 et les cercles de couleurs sur les nuages de points représentent l'amplitude des valeurs que prennent chaque sélection de pays.













Annexe 4 : Valeurs du nombre de lits d'hôpitaux disponibles par habitant dans les pays africains et ratio d'écart avec la France

Pays	Nombre de lits d'hôpitaux disponibles pour 10 000 habitants	Ratio d'écart avec la France	
France	64.8		
Algeria	19	3.4	
Angola	8	8.1	
Benin	5	13.0	
Botswana	18	3.6	
Burkina Faso	4	16.2	
Burundi	7.9	8.2	
Cabo Verde	21	3.1	
Cameroon	13	5.0	
Central African Republic	10	6.5	
Chad	4	16.2	
Comoros	21.6	3.0	
Congo	16	4 1	
Congo DR	8	8 1	
Côte d'Ivoire	1	16.2	
Diibouti		16.2	
Fgynt	15.6	4.0	
Egypt Equatorial Guinea	21	3.1	
Fritrea	7	93	
Eswatini	21	3.1	
Eswatiin	3.1	20.9	
Cabon	12	5.0	
Gaboli	13	5.0	
Chana	0	3.9	
Guinea	2	7.2	
Guinea	3	21.0	
Guillea-Dissau	10	0.5	
Kenya	14	4.6	
Lesotno	13	5.0	
	8	8.1	
	37	1.8	
Madagascar	2	52.4	
Malawi	13	5.0	
Mali	1	64.8	
Mauritania	4	16.2	
Mauritius	34	1.9	
Morocco	11	5.9	
Mozambique	7	9.3	
Namibia	27	2.4	
Niger	2.8	23.1	
Nigeria	5	13.0	
Rwanda	16	4.1	
Sao Tome and Principe	29	2.2	
Senegal	3	21.6	
Seychelles	36	1.8	
Sierra Leone	4	16.2	
Somalia	8.7	7.4	
South Africa	28	2.3	
Sudan	8.2	7.9	
Togo	7	9.3	
Tunisia	22.9	2.8	

Pays	Nombre de lits d'hôpitaux disponibles pour 10 000 habitants	Ratio d'écart avec la France
Uganda	5	13.0
Zambia	20	3.2
Zimbabwe	17	3.8

Annexe 5 : Sélection de courbes de cas journaliers de la COVID-19 en Afrique

Le premier graphe illustre l'évolution de 3 pays classés comme ayant une première vague débutant avant la mi-mai et une deuxième vague après la mi-octobre parfois suivie d'une troisième vague après la mi-février.



Le second graphe illustre l'évolution de 3 pays classés comme ayant une première vague commençant entre la mi-mai et la mi-juillet, et une deuxième vague après la mi-novembre.



Le troisième graphe illustre l'évolution de 2 pays classés comme ayant une première vague commençant au printemps, et une deuxième vague commençant entre la mi-septembre et la mi-octobre, avec ici aussi une troisième vague après janvier 2021.



Le quatrième graphe illustre l'évolution de 3 pays classés comme ayant une première vague commençant à partir de la mi-mars et prolongée dans le temps pendant au moins 7 mois, avec parfois une deuxième vague moins de 1 mois après la fin de la première vague, voire avant la fin de cette première vague.



Le cinquième graphe illustre l'évolution de 3 pays classés comme ayant une seule vague épidémique commençant entre la mi-novembre et la mi-décembre. A noter quelques cas signalés entre mai et août 2020, dus à une vague épidémie sous-estimée par les états ou à un début de vague épidémique correctement stoppée avant sa croissance exponentielle.



Annexe 6 : Description des modifications ponctuelles apportées aux données épidémiologiques au début de leur pré-traitement et des séries creuses repérées

Variable	Pays	Jour	Valeur	Valeur	Commentaire
	-		initiale	modifiée	
Cas	Cameroun	06/07/2020	0	1445	Réattribution de valeurs regroupées en 1 jour
		07/07/2020	2324	487	
		08/07/2020	0	392	
	Côte	15/12/2020	70	36	Correction de valeurs négatives
	d'Ivoire	16/12/2020	-34	0	
	Ethiopie	02/07/2020	0	202	Réattribution des valeurs regroupées en un
	Lunopie	03/07/2020	0	115	pic de rattrapage
		04/07/2020	0	54	
		05/07/2020	0	169	
		07/07/2020	0	280	
		08/07/2020	928	108	
	Ghana	31/05/2020	302	113	Réattribution de valeurs regroupées en 1 jour
		01/06/2020	0	189	
		14/06/2020	846	304	
		15/06/2020	0	542	
		03/07/2020	1254	758	
		04/07/2020	0	496	
	Namibie	11/07/2020	0	45	Réattribution de valeurs regroupées en 1 jour
		12/07/2020	117	72	
		08/08/2020	0	50	
		09/08/2020	147	97	
	Sénégal	15/09/2020	223	15	Réattribution proportionnelle de 208 valeurs
		22/01/2020	Х	X +	d'un pic de rattrapage aux jours précédents
		au		X*(208/14306)	(14306 = cas cumulés au 14/09/2020)
		14/09/2020		200	
	Zambie	13/07/2020	0	388	Réattribution de valeurs regroupées en 1 jour
	A : 1 1	1//0//2020	915	527	
	Zimbabwe	30/04/2020	8	2	Correction de valeurs negatives
		02/05/2020	-0	0	
		27/05/2020	/0 V	0	d'un nie de rettrenage aux jours précédents
		22/01/2020	Λ	$\Lambda +$ V*(76/56)	(56 - cas cumulés au 26/05/2020)
		26/05/2020		A (70/50)	(50 – eus cumules du 20/05/2020)
		18/06/2020	62	0	Réattribution proportionnelle des valeurs
		22/01/2020	Х	X +	d'un pic de rattrapage aux jours précédents
		au		X*(62/401)	(401 = cas cumulés au 17/06/2020)
		17/06/2020			
		09/07/2020	0	41	Réattribution de valeurs regroupées en 1 jour
		10/07/2020	57	16	
		16/07/2020	273	42	Réattribution proportionnelle de 231 valeurs
		09/07/2020	Х	X +	d'un pic de rattrapage aux 7 jours précédents
		au		X*(231/204)	(204 = cas cumulés entre le 09/07/2020 et le
		15/07/2020			15/07/2020)
		05/08/2020	0	118	Correction de valeurs mal attribuées
		06/08/2020	118	56	
		07/08/2020	112	56	Díathaileachta an Ian (11)
		05/09/2020	0	140	Realification de valeurs regroupées en 1 jour
		00/09/2020	161	139	
		20/11/2020	212	102	
		30/11/2020	0	8/	
		02/12/2020	0	114	
		03/12/2020	295	181	
		03/12/2020	275	101	

Les valeurs des données dans ce tableau sont des cas et décès journaliers.

Variable	Pays	Jour	Valeur	Valeur	Commentaire
			initiale	modifiée	
Décès	Cameroun	08/06/2020	0	3	Correction de valeurs mal attribuées
		13/06/2020	0	55	
		15/06/2020	64	5	
		16/06/2020	0	1	
		06/07/2020	0	17	
		07/07/2020	46	12	
		08/07/2020	0	17	
	Ethiopie	08/07/2020	17	0	Retrait des valeurs d'un pic de rattrapage
	1				inexpliqué
	Libye	29/07/2020	9	6	Correction de valeurs négatives
	5	30/07/2020	-3	0	
	Nigeria	08/11/2020	-1	0	Retrait d'une valeur négative inexpliquée
	Zambie	14/07/2020	0	40	Réattribution de valeurs regroupées en 1 jour

Les pays présentant des jours creux (ou réduits en nombre) par rapport aux autres jours sont le Kenya (série des cas du lundi), la Libye (séries des cas et des décès du samedi) et le Maroc (série des cas du lundi). Ces séries creuses ont été retirées des données.

Annexe 7 : Courbes des cas et décès lissées suite à la préparation des données

Les différentes étapes de la préparation des données au cours du pré-traitement ont été illustrées pour le Cameroun ci-dessous. Les 4 premiers graphes représentent les courbes des cas (côté gauche) et des décès (côté droit) de la COVID-19 par million d'habitants de plus de 60 ans, sous forme cumulée (panels du haut) ou journalière (panels du bas) pour le Cameroun entre le 22/01/2020 et le 15/03/2021. En noir les courbes après modification ponctuelles de quelques valeurs aberrantes et en rouge avant modification. Les traits verticaux en gris représentent le premier jour de chaque mois, le trait plus épais celui du 1er janvier 2021.



Les 4 graphes suivants sont les mêmes que ceux ci-dessus, mais sont reconstruits au pas de temps horaire à partir des sous-échantillonnages hebdomadaires, partant de chacun des jours de la semaine (une couleur par jour). Les points représentent les valeurs conservées pour le rééchantillonnage (dont certaines cumulent parfois plusieurs jours de retard).



Les 4 graphes suivants sont les mêmes que ceux ci-dessus, mais résultent de la moyenne des courbes reconstruites. En noir, les courbes des données avant préparation des données (mais après modification ponctuelles des quelques valeurs aberrantes) sont également présentées. Les aires colorées délimitées par des bords rouges fournissent l'écart-type $(\pm 1\sigma)$ à la courbe moyenne (en bleue).













0

C par million d'hab.>60 ans











Annexe 8 : Liste pour chaque pays des meilleurs modèles trouvés et de leur erreur de prévision ramenée au nombre de cas (ou de décès) total

Ce tableau donne le nom du pays et de la variable (Cas ou Morts) de chaque modèle conservé. Les modèles sont désignés par l'intervalle de temps sur lesquels ils ont été conçus (l'unité est l'heure et 1 correspond au 23 janvier 2020 à 5 heures) ainsi que par le nombre de paramètres composant leurs équations différentielles (5p correspond à 5 paramètres). E.A. signifie Erreur Absolue de prévision, 95% et 75% donnent le niveau de confiance choisi et 10j ou 20j désignent l'horizon de prédiction étudié. Toutes les valeurs d'erreur absolue ont été divisées par le nombre total de cas (ou de morts selon la variable étudié) au 15/03/2021 (dernière date utilisée pour construire les modèles), afin d'obtenir l'erreur absolue par unité épidémiologique (1 cas ou 1 mort selon la variable étudiée), et permettant la comparaison des modèles entre eux (facilitée ici par des dégradés de couleur pour chaque colonne, le plus vert étant le modèle le plus juste et le plus rouge celui qui l'est le moins).

Pays	Variable	Modèle	E.A 95% 10j	E.A 95% 20j	E.A 75% 10j	E.A 75% 20j
Afrique	М	[7500:9850] 9p	0.00327	0.00686	0.00033	0.00163
du Sud	С	[1300:5900] 10p	0.00196	0.00915	0.00131	0.00294
		[1:9984] 22p	0.00292	0.00680	0.00097	0.00389
Algérie	C	[1200:9980] 19p	0.00168	0.00491	0.00070	0.00140
	М	[3420:9800] 11p	0.00210	0.00491	0.00070	0.00140
		[3420:9800] 8p	0.00092	0.00443	0.00037	0.00148
Cameroun	C	[1:6400] 22p	0.00082	0.00219	0.00027	0.00110
	М	[1800:6400] 13p	0.00148	0.00648	0.00046	0.00185
Côte	М	[1500:6300] 21p	0.00141	0.00311	0.00057	0.00141
d'Ivoire	C	[1300:9750] 17p	0.00238	0.00635	0.00079	0.00191
Egypte	C	[6500:9800] 19p	0.00172	0.00431	0.00043	0.00108
	М	[1200:9850] 22p	0.00146	0.00437	0.00036	0.00146
Ethiopie	C	[1300:8700] 10p	0.00135	0.00338	0.00034	0.00135
Ghana	C	[1300:6400] 17p	0.00183	0.00548	0.00055	0.00146
		[1300:6400] 19p	0.00183	0.00548	0.00055	0.00183
		[1300:6400] 16p	0.00183	0.00439	0.00073	0.00183
		[1300:9600] 18p	0.00183	0.00439	0.00073	0.00183
		[1300:9600] 16p	0.00229	0.00457	0.00073	0.00183
		[1300:6400] 14p	0.00274	0.00731	0.00091	0.00219
Kenya	C	[5750:8900] 17p	0.00154	0.00326	0.00058	0.00144
		[5750:8900] 13p	0.00173	0.00384	0.00077	0.00173
Libye	C	[1500:8250] 7p	0.00122	0.00257	0.00064	0.00128
	М	[3000:9000] 15p	0.00195	0.00585	0.00078	0.00195
Maroc	C	[3000:9900] 11p	0.00106	0.00398	0.00044	0.00194
Namibie	М	[4000:6500] 12p	0.00253	0.00675	0.00068	0.00169
	C	[3300:6800] 16p	0.00481	0.00542	0.00120	0.00301
Nigeria	C	[1200:9850] 19p	0.00225	0.00451	0.00056	0.00197
Sénégal	C	[1200:9800] 18p	0.00170	0.00403	0.00064	0.00159
		[1200:9800] 19p	0.00106	0.00318	0.00053	0.00170
		[7200:9800] 11p	0.00106	0.00318	0.00053	0.00170
		[1000:6750] 7p	0.00159	0.00339	0.00064	0.00170

Pays	Variable	Modèle	E.A 95% 10j	E.A 95% 20j	E.A 75% 10j	E.A 75% 20j
Togo	М	[1:9984] 20p	0.00395	0.00593	0.00079	0.00316
Tunisie	М	[5200:9750] 8p	0.00150	0.00374	0.00056	0.00168
		[1200:2900] 6p	0.00262	0.00449	0.00093	0.00187
Zimbabwe	С	[2700:6700] 13p	0.00706	0.01115	0.00074	0.00130
	М	[3500:9800] 21p	0.00226	0.01263	0.00090	0.00226
		[6800:9900] 21p	0.00271	0.01443	0.00090	0.00226

Annexe 9 : Description des meilleurs modèles selon leur erreur de prévision ramenée au nombre de cas (ou de décès) total

Ces nuages de points représentent l'erreur absolue de prévision ramenée au nombre de cas (ou de décès) total au 15/03/2021 des modèles sélectionnés, pour un niveau de confiance de 95 et 75% et à un horizon de 10 jours (haut) ou 20 jours (bas). Les triangles signifient que les modèles simulent des cas d'infection (20 modèles en tout), tandis que les ronds sont associés aux simulations de morts (8 modèles en tout).




Annexe 10 : Portrait de phase et prédictibilité pour des données allant jusqu'au 03/05/2021 de 14 modèles sélectionnés

Les modèles présentés ici correspondent à une sélection parmi ceux ayant la meilleure prédictibilité pour les données allant du 22/01/2020 au 15/03/2021. Pour chaque modèle, 6 figures sont présentées : l'évolution du nombre de cas journaliers au cours du temps (a), les projections du portrait de phase de cette dynamique, selon la variable des cas journaliers C1 et sa dérivée première C2 (b), selon C2 et C3 (la dérivée seconde de C1) (c), et selon C1 et C3 (d), ainsi que l'évolution de l'erreur de prévision sous forme de courbes (e) et sous forme bidimensionnelle (f). Les données utilisées vont du 22/01/2020 au 03/05/2021. Pour les 4 premières figures, les valeurs observées sont en noir, celles utilisées pour générer le modèle en trait plein, et celles modélisées en rouge. Les modèles ont été intégrés sur 625 jours à partir des conditions initiales du premier jour de la fenêtre temporelle utilisée pour générer le modèle. Pour les graphiques e) et f), l'horizon de prédiction est de 20 jours et 250 conditions initiales ont été utilisées, tirées de données observées, réparties également entre le 22/01/2020 et le 13/04/2021. Pour le graphique e), chaque courbe verte correspond à la variation de l'erreur sur 20 jours pour une des conditions initiales, les courbes en noir correspondant à l'erreur associée à un niveau de confiance de 75% (en tirets), 90% (pointillés et tirets) et 95% (pointillés). Pour le graphe f), l'abscisse correspond au nombre de jours depuis le 1er janvier 2020 et l'ordonnée à l'horizon de prédiction, lui aussi en jour. L'erreur de prévision est en couleur, avec des tons plus foncés pour une sur-estimation du nombre de cas et des tons plus clairs pour une sous-estimation. Les cas journaliers sont normalisés par million d'habitants de plus de 60 ans. Les modèles sont désignés par l'intervalle de temps sur lesquels ils ont été conçus (l'unité est l'heure et 1 correspond au 23 janvier 2020 à 5 heures)

Les 6 graphes ci-dessous correspondent au modèle à 10 paramètres simulant les cas en Afrique du Sud, généré à partir des données [1300:5900] et dont la formule polynomiale est la suivante :

 $\label{eq:dX3/dt} dX3/dt = -0.126958415 \ X2 \ -1.4488e \ -05 \ X2^3 \ + \ 0.003290612 \ X1 \ + \ 0.000298827 \ X1 \ X2 \ + \ 3.93e \ -07 \ X1 \ X2^2 \ - \ 8.236e \ -06 \ X1^2 \ - \ 1.03e \ -07 \ X1^2 \ X2 \ + \ 3e \ -09 \ X1^3$



Les 6 graphes ci-dessous correspondent au modèle à 19 paramètres simulant les cas en Algérie, généré à partir des données [1200 : 9980] et dont la formule polynomiale est la suivante :

 $dX3/dt = 0.0796686 + 0.0662343 X3^2 - 0.0796214 X2 + 0.0387395 X2 X3 + 0.0043666 X2^2 - 0.0022654 X2^2 X3 - 0.0004227 X2^3 - 0.0036161 X1 + 0.0011516 X1 X3 - 0.000429 X1 X3^2 + 0.0009288 X1 X2 - 0.0002453 X1 X2 X3 - 2.05e-05 X1 X2^2 + 3.93e-05 X1^2 - 1.04e-05 X1^2 X3 - 2.2e-06 X1^2 X2 - 1e-07 X1^3$



Les graphes ci-dessous correspondent au modèle à 22 paramètres simulant les cas au Cameroun, généré à partir des données [1 : 6400] et dont la formule polynomiale est la suivante :

 $dX3/dt = 0.01450574 + 0.17768333 X3 + 0.22661589 X3^2 - 0.10327662 X3^3 - 0.11632778 X2 + 0.07260197 X2 X3 - 0.01174499 X2 X3^2 + 0.00747561 X2^2 - 0.0028243 X2^2 X3 - 0.00112245 X2^3 - 0.0011693 X1 - 0.00281063 X1 X3 - 0.00160972 X1 X3^2 + 0.00188238 X1 X2 - 0.00035329 X1 X2 X3 - 4.802e - 05 X1 X2^2 + 1.56e - 05 X1^2 + 1.051e - 05 X1^2 X3 - 5.8e - 06 X1^2 X2 - 4e - 08 X1^3$



Voici un zoom sur les données utilisées pour générer le modèle (du fait d'un changement d'échelle avec les toutes dernières données) :



Pour la prédictibilité, suite à 3 erreurs dans la fonction permettant d'obtenir les graphes, il est impossible de représenter correctement les données. Le problème vient du fait que suite à un pré-traitement des données ne permettant pas de lisser au mieux la courbe des cas en fin de courbe, une prévision est impossible à faire pour certaines conditions initiales. Voici cependant une représentation au cours du temps de e) ainsi qu'une représentation de f) (les lignes blanches correspondent aux prévisions impossibles à réaliser) :



Les graphes ci-dessous correspondent au modèle à 21 paramètres simulant les morts en Côte d'Ivoire, généré à partir des données [1500:6300] et dont la formule polynomiale est la suivante :

 $dX3/dt = -0.002 + 0.153 X3 - 9.097 X3^2 - 0.209 X2 - 2.502 X2 X3 - 278.886 X2 X3^2 + 5.502 X2^2 - 167.401 X2^2 X3 - 31.164 X2^3 + 0.01 X1 - 0.748 X1 X3 + 10.657 X1 X3^2 + 0.641 X1 X2 - 0.999 X1 X2 X3 - 4.931 X1 X2^2 - 0.025 X1^2 + 0.686 X1^2 X3 - 0.434 X1^2 X2 + 0.014 X1^3$



Voici un zoom sur les données utilisées pour générer le modèle, pour les 4 premiers graphes (du fait d'un changement d'échelle avec les toutes dernières données) :



Les 6 graphes ci-dessous correspondent au modèle à 22 paramètres simulant les morts en Egypte, généré à partir des données [1200:9850] et dont la formule polynomiale est la suivante :

 $dX3/dt = 0.0036 - 0.19649 X3 + 2.63564 X3^2 - 1.17859 X3^3 - 0.1939 X2 + 1.54975 X2 X3 - 4.1405 X2 X3^2 - 0.13161 X2^2 + 0.52573 X2^2 X3 - 0.12677 X2^3 - 0.00122 X1 + 0.10633 X1 X3 - 0.32858 X1 X3^2 + 0.06097 X1 X2 - 0.26089 X1 X2 X3 + 0.02645 X1 X2^2 + 0.00012 X1^2 - 0.01164 X1^2 X3 - 0.00531 X1^2 X2 - 1e-05 X1^3$



Les 6 graphes ci-dessous correspondent au modèle à 10 paramètres simulant les cas en Ethiopie, généré à partir des données [1300 : 8700] et dont la formule polynomiale est la suivante :

 $dX3/dt = 0.2679677 \ X3^2 \ -0.0390729 \ X3^3 \ -0.0593653 \ X2 \ + \ 0.0043915 \ X1 \ -0.0011471 \ X1 \ X3^2 \ + \ 1.18e-05 \ X1 \ X2^2 \ -7.57e-05 \ X1^2 \ + \ 2e-07 \ X1^3$



Les 6 graphes ci-dessous correspondent au modèle à 18 paramètres simulant les cas au Ghana, généré à partir des données [1300 : 9600] et dont la formule polynomiale est la suivante :

$$\label{eq:dX3/dt} \begin{split} & = -0.31087501 \; X3 - 0.02375191 \; X3^2 \; -0.1306806 \; X2 + 0.03231115 \; X2 \; X3 - 0.00186872 \; X2 \\ & X3^2 \; -9.482e - 05 \; X2^3 + 0.00373462 \; X1 + 0.00371468 \; X1 \; X3 + 5.488e - 05 \; X1 \; X3^2 + 0.00125104 \; X1 \\ & X2 \; -0.00010522 \; X1 \; X2 \; X3 + 5.57e - 06 \; X1 \; X2^2 \; -3.705e - 05 \; X1^2 \; -7.35e - 06 \; X1^2 \; X3 \; -2.63e - 06 \; X1^2 \; X2 + 6e - 08 \; X1^3 \end{split}$$



Les 6 graphes ci-dessous correspondent au modèle à 17 paramètres simulant les cas au Kenya, généré à partir des données [5750:8900] et dont la formule polynomiale est la suivante :

 $dX3/dt = -0.51889772 - 0.00695195 X3^2 - 0.26669634 X2 + 0.03579 X2 X3 - 0.00538123 X2 X3^2 - 0.00502265 X2^2 - 0.0004693 X2^3 + 0.01112785 X1 + 8.23e-06 X1 X3^2 + 0.00283422 X1 X2 - 0.00012847 X1 X2 X3 + 1.351e-05 X1 X2^2 - 4.266e-05 X1^2 - 4.81e-06 X1^2 X2 + 4e-08 X1^3$



Voici un zoom sur les données utilisées pour générer le modèle, pour les 4 premiers graphes (du fait d'un changement d'échelle avec les toutes dernières données) :



Les 6 graphes ci-dessous correspondent au modèle à 7 paramètres simulant les cas en Libye, généré à partir des données [1500:8250] et dont la formule polynomiale est la suivante :



dX3/dt = -0.02959998 X2 + 0.004514481 X1 -5.376e-06 X1^2 -2.1e-08 X1^2 X2 + 2e-09 X1^3

Les 6 graphes ci-dessous correspondent au modèle à 12 paramètres simulant les morts en Namibie, généré à partir des données [4000:6500] et dont la formule polynomiale est la suivante :

 $dX3/dt = 0.55873 X3 - 1.27773 X3^3 + 0.05899 X2^2 - 0.03201 X2^3 - 0.11966 X1 X3 - 0.00608 X1 X2^2 - 0.00037 X1^2 + 0.00567 X1^2 X3 - 0.00016 X1^2 X2 + 4e - 05 X1^3$



Les 6 graphes ci-dessous correspondent au modèle à 19 paramètres simulant les cas au Nigeria, généré à partir des données [1200:9850] et dont la formule polynomiale est la suivante :

$$\label{eq:dX3/dt} \begin{split} & = 0.0343334 \ \text{-}0.0756397 \ \text{X3}^2 \ \text{-}0.11032 \ \text{X2} + 0.0339922 \ \text{X2} \ \text{X3} + 0.0101453 \ \text{X2}^2 \ \text{-}0.0036629 \\ & \text{X2}^2 \ \text{X3} \ \text{-}0.0009342 \ \text{X2}^3 \ \text{-}0.0023776 \ \text{X1} + 0.0011545 \ \text{X1} \ \text{X3} + 0.0009709 \ \text{X1} \ \text{X3}^2 + 0.0022362 \ \text{X1} \\ & \text{X2} \ \text{-}0.0002411 \ \text{X1} \ \text{X2} \ \text{X3} \ \text{-}8.8\text{e}\text{-}05 \ \text{X1} \ \text{X2}^2 + 3.49\text{e}\text{-}05 \ \text{X1}^2 \ \text{-}5.2\text{e}\text{-}06 \ \text{X1}^2 \ \text{X3} \ \text{-}1.17\text{e}\text{-}05 \ \text{X1}^2 \ \text{X2} \ \text{-}1\text{e}\text{-}07 \ \text{X1}^3 \end{split}$$



Les 6 graphes ci-dessous correspondent au modèle à 19 paramètres simulant les cas au Sénégal, généré à partir des données [1200:9800] et dont la formule polynomiale est la suivante :

$$\label{eq:dX3/dt} \begin{split} &dX3/dt = 0.07409904 + 0.05432847 \; X3^2 \; -0.00585201 \; X3^3 \; -0.06526278 \; X2 + 0.00524753 \; X2 \; X3 \; -0.00400956 \; X2 \; X3^2 + 0.00312921 \; X2^2 \; -0.00046723 \; X2^3 \; -0.00319631 \; X1 \; + \; 0.00067275 \; X1 \; X3 \; -0.00022793 \; X1 \; X3^2 \; + \; 0.00030833 \; X1 \; X2 \; -6.03e\text{--}06 \; X1 \; X2^2 \; + \; 2.564e\text{--}05 \; X1^2 \; -2.48e\text{--}06 \; X1^2 \; X3 \; -3e\text{--}08 \; X1^2 \; X2 \; -5e\text{--}08 \; X1^3 \; -0.00312921 \; X2^2 \; -0.00046723 \; X2^2 \; + \; 2.564e\text{--}05 \; X1^2 \; -2.48e\text{--}06 \; X1^2 \; X3 \; -3e\text{--}08 \; X1^2 \; X2 \; -5e\text{--}08 \; X1^3 \; -3e\text{--}08 \; X1^2 \; X3 \; -3e\text{--}08 \; X1^2 \; X3 \; -3e\text{--}08 \; X1^3 \; -3e\text{--}08 \;$$



Les 6 graphes ci-dessous correspondent au modèle à 11 paramètres simulant les cas au Sénégal, généré à partir des données [7200:9800] et dont la formule polynomiale est la suivante :

 $dX3/dt = -0.10874414 X2 + 0.00914253 X2^2 - 0.0006102 X2^3 - 0.00023449 X1 + 0.00119765 X1 X2 - 2.792e - 05 X1 X2^2 + 4.53e - 06 X1^2 - 2.26e - 06 X1^2 X2 - 1e - 08 X1^3$



Les 6 graphes ci-dessous correspondent au modèle à 8 paramètres simulant les morts en Tunisie, généré à partir des données [5200:9750] et dont la formule polynomiale est la suivante :

dX3/dt = -0.0605287 X2 + 0.0013054 X1 + 0.00541 X1 X2 -6.76e-05 X1^2 -0.0001309 X1^2 X2 + 9e-07 X1^3



Annexe 11 : Evolution pour une sélection de modèles de leur erreur de prévision selon la fenêtre temporelle considérés

Ce tableau donne le nom du pays et de la variable (Cas ou Morts) de chaque modèle conservé. Les modèles sont désignés par l'intervalle de temps sur lesquels ils ont été conçus (l'unité est l'heure et 1 correspond au 23 janvier 2020 à 5 heures) ainsi que par le nombre de paramètres composant leurs équations différentielles (5p correspond à 5 paramètres). E.A. signifie Erreur Absolue de prévision, 95% et 75% donnent le niveau de confiance choisi et 10j ou 20j désignent l'horizon de prédiction étudié. Les erreurs absolues correspondent, pour chaque modèle, à celles obtenues sur la fenêtre temporelle du 22/01/2020 au 15/03/2021 (ligne supérieure sur fond blanc), et à celles obtenues sur la fenêtre temporelle du 22/01/2020 au 03/05/2021 (ligne du milieu sur fond gris). La ligne inférieure pour chaque modèle correspond au ratio de l'erreur absolue pour les données allant du 22/01/2020 au 15/03/2021 par celle pour les données allant du 22/01/2020 au 03/05/2021, et permet d'étudier l'évolution de la capacité prévisionnelle des modèles dans le temps. Elle est colorée en vert si le ratio est inférieur à 0.95, en jaune s'il est compris entre 0.95 et 1.05 et en rouge s'il est supérieur à 1.05. Toutes les valeurs d'erreur absolue ont été divisées par le nombre total de cas (ou de morts selon la variable étudié) à la dernière date de la fenêtre temporelle considérée (15/03/2021 ou 03/05/2021), afin d'obtenir l'erreur absolue par unité épidémiologique (1 cas ou 1 mort selon la variable étudiée), permettant la comparaison de l'erreur absolue entre modèles et au sein de chaque modèle pour les différentes fenêtres temporelles.

Pays	Variable	Modèle	E.A. 95%	E.A. 95%	E.A. 75%	E.A. 75%
			10j.	20j.	10j.	20j.
			et ratio	et ratio	et ratio	et ratio
Afrique du Sud	С	[1300:5900] 10p	0.00327	0.00686	0.00033	0.00163
			0.00266	0.00673	0.00028	0.00133
			0.815	0.981	0.859	0.815
Algérie	С	[1200:9980] 19p	0.00092	0.00443	0.00037	0.00148
			0.00076	0.00382	0.00035	0.00104
			0.828	0.862	0.940	0.705
Côte d'Ivoire	М	[1500:6300] 21p	0.00141	0.00311	0.00057	0.00141
			0.00256	0.00390	0.00050	0.00137
			1.811	1.255	0.876	0.971
Egypte	М	[1200:9850] 22p	0.00146	0.00437	0.00036	0.00146
			0.00122	0.00342	0.00031	0.00110
			0.838	0.782	0.838	0.754
Ethiopie	С	[1300:8700] 10p	0.00135	0.00338	0.00034	0.00135
			0.00177	0.00492	0.00050	0.00123
			1.309	1.456	1.481	0.909
Ghana	С	[1300:9600] 18p	0.00183	0.00548	0.00055	0.00146
			0.00225	0.00536	0.00070	0.00163
			1.233	0.978	1.275	1.115
Kenya	С	[5750:8900] 17p	0.00154	0.00326	0.00058	0.00144
			0.00159	0.00341	0.00045	0.00131
			1.035	1.044	0.788	0.907
Libye	C	[1500:8250]	0.00122	0.00257	0.00064	0.00128
		7p	0.00126	0.00235	0.00058	0.00119
			1.035	0.914	0.909	0.928

Pays	Variable	Modèle	E.A. 95%	E.A. 95%	E.A. 75%	E.A. 75%
			10j.	20j.	10j. et	20j.
			et ratio	et ratio	ratio	et ratio
Namibie	М	[4000:6500] 12p	0.00481	0.00542	0.00120	0.00301
			0.00385	0.00435	0.00102	0.00239
			0.800	0.804	0.844	0.795
Nigeria	C	[1200:9850] 19p	0.00225	0.00451	0.00056	0.00197
			0.00165	0.00490	0.00066	0.00212
			0.730	1.087	1.169	1.077
Sénégal	С	[1200:9800] 19p	0.00106	0.00318	0.00053	0.00170
			0.00116	0.00361	0.00058	0.00175
			1.097	1.134	1.097	1.029
		[7200:9800] 11p	0.00159	0.00339	0.00064	0.00170
			0.00144	0.00388	0.00072	0.00242
			0.902	1.143	1.128	1.429
Tunisie	М	[5200:9750] 8p	0.00150	0.00374	0.00056	0.00168
			0.00115	0.00331	0.00040	0.00144
			0.770	0.885	0.719	0.855

Annexe 12 : Variation des prévisions du modèle des cas en Ethiopie selon les valeurs attribuées à ses coefficients

Les courbes suivantes du nombre de cas journaliers sont simulées sur 130 jours à partir de conditions initiales du 17/08/2020, lors de la phase ascendante de la première vague. La valeur de base de chaque coefficient correspond à celle du modèle construit sur l'ensemble de sa fenêtre temporelle.

Pour analyser l'influence du coefficient du paramètre X_3^2 sur le modèle, nous modifions sa valeur, en prenant au lieu de 0.268 (courbe verte), 0.257 (courbe bleue) et 0.277 (courbe rouge). On observe que comme attendu dans le **Tableau 3**, une valeur inférieure à celle du modèle se traduit par une réduction globale du nombre de cas au cours du temps, et une valeur légèrement plus positive fait augmenter le nombre de cas prédit.



Pour analyser l'influence du coefficient du paramètre X_3^3 sur le modèle, nous modifions sa valeur, en prenant au lieu de -0.039 (courbe verte), -0.049 (courbe bleue) et -0.029 (courbe rouge). On observe cette fois que contrairement à l'hypothèse émise dans le **Tableau 3**, une valeur plus négative n'accentue pas l'épidémie, elle réduit même l'amplitude des oscillations. De plus, en prenant des valeurs supérieures, on voit que le modèle diverge rapidement vers $+\infty$, au lieu d'avoir des valeurs inférieures au modèle initial (on observe d'ailleurs un résultat similaire lorsque le coefficient est très faiblement négatif ou positif).



Pour analyser l'influence du coefficient du paramètre X_2 sur le modèle, nous modifions sa valeur, en prenant au lieu de -0.059 (courbe verte), -0.069 (courbe bleue) et -0.049 (courbe rouge). On observe qu'à l'inverse de ce qui était attendu dans le **Tableau 3**, une valeur inférieure au modèle n'accélère pas l'épidémie, et une valeur supérieure ne la freine pas, au contraire.



Pour analyser l'influence du coefficient du paramètre X_1 sur le modèle, nous modifions sa valeur, en prenant au lieu de 0.004 (courbe verte), 0.003 (courbe bleue) et 0.005 (courbe rouge). On observe comme ce qui était attendu dans le **Tableau 3**, qu'une valeur inférieure au modèle accentue le nombre de cas, mais de manière périodique et pas au tout début. Finalement, l'épidémie fait des oscillations avec plus d'amplitude. Aussi, une valeur positive plus élevée que dans le modèle initial fait régresser l'épidémie mais seulement après une première vague plus intense. Ainsi, selon la durée que l'on étudie, le changement de valeur du paramètre peut avoir un effet différent au cours du temps.



Pour le coefficient du paramètre $X_1X_3^2$, les résultats sont similaires à X_3^3 et donc inverse à ce qui était attendu (une valeur inférieure accentue l'épidémie, et une valeur supérieure la réduit).

Pour le coefficient des paramètres $X_1X_2^2$ et X_1^2 , les résultats sont similaires à X_3^3 et $X_1X_3^2$. Mais cette fois donc, comme ce qui était attendu dans le **Tableau 3**, une valeur inférieure au modèle diminue le nombre de cas et une valeur supérieure au modèle initial accentue le nombre de cas. Cependant, une valeur inférieure réduit le nombre de cas à court terme, sans nécessairement le réduire à long terme (il réduit surtout l'amplitude de l'oscillation).

Pour analyser l'influence du coefficient du paramètre X_1^3 sur le modèle, nous modifions sa valeur, en prenant au lieu de 2^e-7 (courbe verte), 1^e-7 (courbe bleue) et 3^e-7 (courbe rouge). A court terme, les hypothèses s'avèrent inexactes, une augmentation de la valeur du paramètre n'entraine pas une diminution du nombre de cas prédits mais une augmentation, et inversement. Cependant, à l'horizon du 30eme jour, c'est l'opposé qu'on observe avec une diminution brutale du nombre de cas pour une valeur plus élevée et une augmentation momentanée pour des valeurs inférieures.



Annexe 13 : Evolution de l'indice de rigueur des mesures gouvernementales en Ethiopie

Pour mieux comparer les données, la **Figure 6**, correspondant à l'évolution des coefficients des paramètres du modèle étudié, a été rajoutée en premier, ci-dessous. Les données s'arrêtent au 15/03/2021.



Indices de rigueur des mesures gouvernementales en Ethiopie



Résumé

Modélisation épidémiologique basée sur la théorie du Chaos de la COVID-19 sur le continent africain.

La pandémie de COVID-19 s'est montrée difficile à modéliser de par son caractère imprévisible ainsi que par le manque d'informations sur les paramètres épidémiologiques de celle-ci au début de son émergence. La technique de la modélisation globale, permettant de générer des modèles déterministes à partir d'une seule variable (ici les cas ou décès journaliers), et sans faire d'hypothèses fortes, a été appliquée à 17 pays africains. Pour la quasi-intégralité de ces pays, des modèles chaotiques avec une prédictibilité à court terme intéressante ont pu être générés, avec certains modèles approchant parfois très bien la dynamique de la COVID-19 dans l'espace des phases. De plus, l'étude des paramètres composant les équations différentielles des modèles obtenus pourrait fournir des informations sur les facteurs impliqués dans l'évolution épidémiologique de la maladie. Ce travail vient confirmer l'intérêt et le potentiel de ce type d'approches dans le cas de maladies infectieuses émergentes, ainsi que la nature chaotique de la dynamique de cette pandémie. Enfin, cette étude a mis en lumière 2 facteurs permettant de réduire significativement les différences de valeurs observées entre pays (autant en Afrique que pour la plupart des pays du monde) pour les cas d'infection et de décès dus à la COVID-19: la proportion de personnes âgées de plus de 60 ans et le nombre de lits d'hôpitaux par habitant.

<u>Mots Clés:</u> Modélisation – COVID-19 – Théorie du Chaos – Afrique – Epidémiologie – Emergence de maladie – Zoonose

Abstract

Epidemiological modelling based on the Chaos theory applied to the outbreak of COVID-19 in Africa.

The COVID-19 pandemic has been difficult to model due to its unpredictable nature as well as the lack of information on its epidemiological parameters at the start of its emergence. The technique of global modelling, which can generate deterministic models from a single variable (here the daily number of cases or deaths), and without making strong assumptions, has been applied to 17 African countries. For almost all these countries, chaotic models with interesting short-term predictability were generated, and some models approach very well the dynamics of COVID-19 in the phase space. Moreover, the study of the parameters composing the differential equations of these models could provide information on the factors involved in the epidemiological course of the disease. This work confirms the interest and the potential of this type of approach in the case of emerging infectious diseases, as well as the chaotic nature of the dynamics of this pandemic. Finally, this study highlighted 2 factors that significantly reduce the differences in values observed between countries (both in Africa and in most countries of the world) for cases of infection and death due to COVID-19: the proportion of people over 60 and the number of hospital beds per capita.

<u>Keywords:</u> Modelling – COVID-19 – Chaos Theory – Africa – Epidemiology – Disease emergence – Zoonosis