

UNIVERSITÉ DE GENÈVE

Modèle SIR compartimental appliqué à la pandémie de Covid-19

Auteur : Julien Python

E-mail : julien.python@etu.unige.ch

Travail dirigé par : José Rolim

Master bi-disciplinaire : majeure mathématiques, mineure sciences
informatiques

18 février 2022



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES

Table des matières

1	Introduction	2
2	Système dynamique	2
3	Modèle SIR	3
3.1	Définition	3
3.2	Nombre de reproduction de base	5
3.3	Tentative de simulation de la Covid-19	6
3.4	Qu'est-ce qu'un modèle mathématique ?	8
3.5	Simulation de deux vagues infectieuses	8
3.6	Simplicité du modèle SIR	10
4	Modèle SEIR	11
4.1	Définition	11
4.2	Approximation du nombre de personnes contaminées dans le futur	11
5	Approximation du nombre de personnes contaminées hypothé-	13
	tiquement	
5.1	Première procédure	13
5.1.1	Assignation d'une valeur scalaire à S_0, E_0, I_0 et R_0	13
5.1.2	Estimation des paramètres β, σ et γ	14
5.2	Interprétation des résultats	14
5.3	Deuxième procédure	17
6	Troisième procédure utilisant une estimation du nombre d'in-	19
	dividus infectés	
6.1	Impact sur les décisions politiques	19
6.2	Modèles pour les cas non détectés	20
6.3	Troisième procédure	20
6.4	Interprétation des résultats	21
7	Modèle SEIRS	24
8	Conclusion	24

1 Introduction

Comment fonctionnent les modèles utilisés sur l'épidémie de la Covid-19 ? Nous avons commencé ce travail en février 2020. Et c'est durant cette même période que la Covid-19 se propageait largement au sein des populations partout dans le monde. Lorsque le professeur José Rolim nous a proposé de travailler sur ce sujet après avoir lu les derniers chapitres du livre *Networks, Crowds, and Markets* [1], nous avons immédiatement répondu oui avec enthousiasme, tant le sujet était fascinant et d'actualité.

Début 2020, nous ne savions pas de quelle manière nos vies allaient être tant impactées par la suite. Des scientifiques partout dans le monde étudiaient la Covid-19 afin de mieux la comprendre. Ainsi, les modélisations se sont améliorées avec les connaissances acquises sur la façon dont se comporte le virus.

Au cours des prochains chapitres, nous allons chercher à comprendre le fonctionnement d'un modèle épidémiologique intitulé le modèle SIR. Ce dernier est régulièrement utilisé dans les modélisations liées à la Covid-19.

Comment se comporte-t-il ? Que permet-il de réaliser ? Quelles sont ses limitations ? Nous allons développer chacune de ces questions dans la suite de ce travail.

2 Système dynamique

Afin d'être rigoureux dans notre démarche, nous allons décrire quel type de système nous intéresse pour le modèle SIR.

Un système dynamique est un système qui évolue dans le temps de manière discrète ou continue. Nous notons $s(t)$ l'état du système au temps t . Afin d'explicitier ces deux notions de temps, nous allons énoncer un exemple pour chacune.

Un exemple de système dynamique à temps discret est le nombre de personnes dans un magasin. Ce nombre varie selon les entrées et les sorties des clients. Ces variations sont donc ponctuelles, et notre système, dans ce cas le nombre de clients, peut rester le même durant plusieurs minutes.

Un exemple de système dynamique à temps continu est la quantité de carburant dans un réservoir. Lors de l'utilisation du réservoir, la quantité va continuellement diminuer. C'est donc un système qui va changer sans arrêt au fil du temps. Un système dynamique à temps continu peut généralement être décrit par un système d'équations différentielles.

Nous allons étudier le modèle SIR dans la suite de ce travail. Ce modèle qui peut être utilisé pour modéliser un virus au sein d'une population donnée, est décrit par un système d'équations différentielles et ses états changent constamment au cours du temps jusqu'à atteindre un point fixe. Il s'agit donc d'un système dynamique à temps continu.

3 Modèle SIR

3.1 Définition

Nous avons ainsi formalisé le type de système dont nous nous préoccupons. Il nous reste à définir le cœur de notre travail, le modèle SIR.

Historiquement, il a été cité pour la première fois dans une publication datant de 1927 par Kermack et McKendrick [2]. Le modèle SIR est un modèle épidémiologique simple, qui entreprend de représenter l'évolution d'une maladie infectieuse au cours du temps. Par la suite, différents modèles plus complexes s'en sont inspirés. De nos jours, le modèle SIR est décrit avec le système d'équations différentielles suivant :

$$\begin{cases} \frac{dS}{dt} = -\beta I \frac{S}{N} \\ \frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (1)$$

Notations :

- S pour Susceptible : les individus sains susceptibles de se faire infecter.
- I pour Infectious : les individus actuellement infectés.
- R pour Recovered : les individus qui ne sont plus infectés, comme les morts ou les guéris.
- N pour nombre total d'individus, tel que $N = S + I + R$.
- β : ce paramètre définit l'infection d'individus sains.
- γ : ce paramètre définit le décès ou le rétablissement d'individus infectés.

Remarques :

- Nous supposons que notre population est constante par rapport au temps, c'est à dire que $\frac{dN}{dt} = 0$.
- Nous ne prenons pas en compte les naissances et les morts, c'est à dire les entrées et sorties du système, déterminées par le paramètre μ . Bien qu'il existe des modèles SIR qui les prennent en compte, nous pouvons les négliger pour la Covid-19 car le virus est un processus amplement plus rapide que les naissances et les morts.
- Les paramètres β et γ sont des constantes positives et donc ne dépendent pas du temps t .

Nous pouvons interpréter cette équation différentielle (1) avec le graphe de la figure 1. C'est une forme de système de vases communicants avec trois compartiments distincts mais à sens unique de la gauche vers la droite. Un individu peut aller du compartiment S au compartiment I mais ne peut ni aller du I au S , ni directement du S au R .

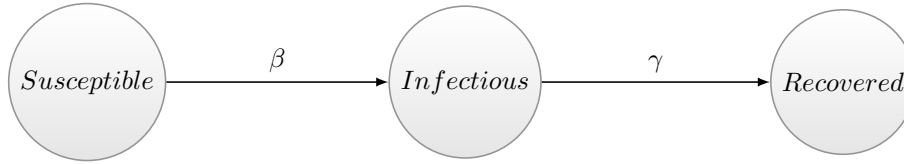


FIGURE 1 – Interprétation graphique du modèle SIR

Lors du début d’une épidémie, il y a un certain nombre de personnes infectées, et le reste de la population est saine. Donc nous commençons avec quelques individus dans I et le reste dans S . La proportion d’individus passant du compartiment S au compartiment I , c’est à dire de sain à infecté, pour une période de temps, dépend du nombre d’individus dans I, S et N ainsi que du paramètre β . Nous remarquons que c’est bien ce que nous décrit la définition du modèle (1). En essayant différents paramètres β, γ et I_0 où I_0 est une notation pour I_{t_0} c’est à dire au temps $t=0$, donc au début de notre système, nous avons pu constater qu’il y a trois schémas de modélisation qui reviennent.

Le premier indique que toute la population a été contaminée au temps final t_f , c’est à dire $S_{t_f} = 0$, $I_{t_f} = 0$ et $R_{t_f} = N$. Ainsi, cela signifie que le nombre d’individus sains et le nombre d’individus infectés sont nuls, et que le nombre d’individus dans le compartiment R est égal à la population totale N .

Le deuxième indique qu’une partie de la population a été infectée, et donc il reste des individus dans le compartiment S , et le nombre d’infectés au temps t_f est égal à 0. Nous avons donc $N_{t_f} = S_{t_f} + R_{t_f}$. Vu qu’il n’y aura plus aucun individu infecté, il n’y aura pas de changement au sein de nos trois compartiments S_{t_k} , I_{t_k} et R_{t_k} pour tout temps $t_k > t_f$.

Le troisième est un cas particulier de deuxième, où nous remarquons que les premiers infectés n’ont pas réussi à transmettre le virus au reste de la population. Donc pour un temps donné $t_f > 0$, nous avons $S_{t_f} \approx N$, $I_{t_f} = 0$ et $R_{t_f} \approx 0$.

Les figures 2, 3 et 4 possèdent des graphiques où l’axe des ordonnées (l’axe y de la population) est normalisé. Nous avons donc que 1 représente la totalité de nos N individus et inversement, 0 correspond à aucun individu. L’axe des abscisses représente les unités de temps, qui peuvent être des minutes, heures, jours, mois ou années par exemple selon ce que nous souhaitons modéliser. Dans la suite du travail pour la Covid-19, nous avons choisi qu’une unité de temps corresponde à un jour, car nos données collectées recensent les contaminations quotidiennement.

En investiguant, nous avons trouvé ce qui détermine si une maladie infectieuse d’un système est en train de disparaître, il s’agit du paramètre R_0 .

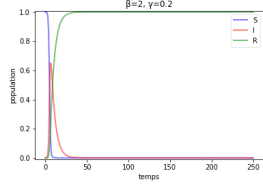


FIGURE 2 – Cas 1

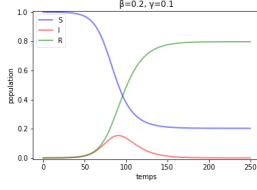


FIGURE 3 – Cas 2

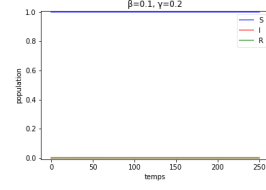


FIGURE 4 – Cas 3

3.2 Nombre de reproduction de base

Une notion importante qui revient régulièrement dans les modèles est le nombre de reproduction de base R_0 . Pour le modèle SIR, $R_0 = \frac{\beta}{\gamma}$. Intuitivement R_0 est le nombre d'individus qu'une personne infectée contamine. Par exemple, pour une population infinie, si $R_0 < 1$ le nombre d'infectés diminuera au cours du temps et inversement si $R_0 > 1$ le nombre d'infectés augmentera.

Pour une population finie, si $R_0 > 1$ l'idée générale est la même que ci-dessus mais il faut prendre en compte l'immunité collective. En effet, les individus du compartiment R ne peuvent plus être infectés, donc ils freinent la propagation du virus comme nous l'avons observé sur la figure 3. Si $R_0 < 1$, le virus arrêtera sa propagation comme nous avons pu le voir sur la figure 4.

Cependant, les paramètres β et γ ne sont pas toujours fixés. Dans notre exemple de la Covid-19, le paramètre β entre S et I peut augmenter ou diminuer lorsque le comportement des populations change. Ainsi il en est de même pour le paramètre R_0 car il est défini par β et γ .

Il est intéressant de constater que la surveillance du paramètre R_0 est primordiale si nous souhaitons contrôler une épidémie. En effet, si le nombre de reproduction de base est toujours strictement plus petit que 1, le virus disparaîtra. Les autorités sanitaires helvétiques ont utilisé cet indice afin d'appliquer des décisions politiques dans la gestion de la Covid-19.

Essayons maintenant d'estimer R_0 . Pendant le début de la propagation d'une maladie infectieuse, nous pouvons [3] approximer le nombre d'infections comme en (2).

$$I(t) \simeq I_0 \cdot e^{(R_0-1) \cdot (\gamma+\mu) \cdot t} \quad (2)$$

De manière informelle, nous pouvons comprendre cette équation par le fait que le virus se propage de manière exponentielle et qu'il faut bien que $R_0 > 1$ afin que l'épidémie continue de vivre, puisque γ, μ et t sont positifs. De plus les individus infectés dépendent bien de μ et γ , et plus le temps avance, plus le nombre d'infectés augmentera. Mais il s'agit bien d'une approximation car seul I_0 est utilisé. En appliquant la fonction logarithme des deux cotés de l'égalité, nous

nous retrouvons avec un terme linéaire, et nous utilisons le fait que $\frac{vertical}{horizontal} \leftrightarrow \frac{I(t)-I_0}{t}$, nous obtenons (3).

$$\begin{aligned} \log(I(t)) &\simeq \log(I_0) + (R_0 - 1) \cdot (\gamma + \mu) \cdot t \\ \log(I(t)) - \log(I_0) &\simeq (R_0 - 1) \cdot (\gamma + \mu) \cdot t \\ \frac{Pente_{regression\ linéaire\ des\ infections}}{\gamma + \mu} + 1 &\simeq R_0 \end{aligned} \quad (3)$$

Or dans le cas de la Covid-19, nous avons que $\mu \ll \gamma$, c'est pourquoi nous négligeons μ afin de simplifier l'équation. La pente de la régression linéaire peut se calculer numériquement et ainsi en isolant R_0 , nous obtenons l'approximation en (4). Ainsi, il nous suffit de savoir la valeur de γ et des données réelles afin d'approximer la valeur de R_0 au début d'épidémie.

$$R_0 \simeq \frac{Pente_{regression\ linéaire\ des\ infections}}{\gamma} + 1 \quad (4)$$

3.3 Tentative de simulation de la Covid-19

Lorsque nous avons commencé notre travail en 2020, nous avons utilisé un fichier mis à disposition sur le site de la confédération suisse contenant toutes les données suisses du virus. Mais il a fallu nettoyer ce fichier, c'est à dire comprendre comment il fonctionnait afin d'en extraire uniquement les données essentielles pour notre cas d'étude. Par exemple, il n'y avait pas le nombre total de contamination par jour qui nous intéressait pour la Suisse mais uniquement le nombre de cas par tranche d'âge par canton et il fallait modifier le format des dates afin que cela corresponde à notre programme utilisé. Nous avons donc écrit un script dans le langage Python qui nous a permis d'obtenir le nombre de cas par jour. Ensuite, nous souhaitions regarder les données pour d'autres pays. Malheureusement, chaque pays adoptait une classification des données différente, et donc il fallait écrire un script différent pour chacun des pays. Nous avons donc cherché une solution afin de pouvoir aisément automatiser notre script pour tous les pays du monde qui publient leurs données.

La solution que nous avons retenue a été d'utiliser les données du site ourworldindata.org, créé et mis à jour par une équipe de recherche de l'université d'Oxford. Sur ce site, nous pouvons télécharger un fichier csv qui contient en particulier le nombre de cas cumulé pour chacun des pays qui publie ses données. Nous allons les utiliser pour la Suisse en premier temps, puis pour d'autres pays en second temps. Dès lors, nous avons écrit un deuxième script pour ce nouveau fichier qui permet d'obtenir le nombre d'individus infectés.

Par exemple, à partir des infections cumulées en Suisse, nous avons essayé de reproduire le graphe SIR respectif. Ce dernier se traduit approximativement par une translation vers la gauche du graphe du compartiment R du modèle SIR.

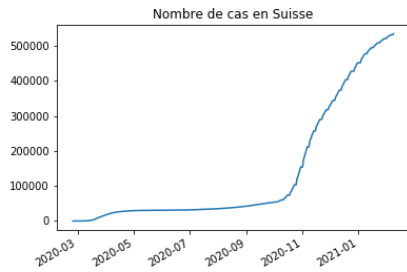


FIGURE 5 – Echelle standard

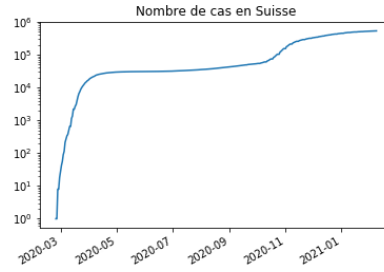


FIGURE 6 – Echelle logarithmique

La translation est approximativement égale au temps durant lequel un individu peut infecter d'autres individus, ce qui est environ de 9 jours [4], même si cette donnée dépend de chaque personne.

Pour ce faire, nous devons choisir des valeurs pour β, γ, S_0 et I_0 , où 0 de S_0 et I_0 est le temps auquel notre système commence. Après plusieurs tentatives, nous ne sommes pas arrivés à créer un graphe issu du modèle SIR semblable aux données suisses.

Cette tentative a été effectuée pendant l'été 2020, et à cette période il y avait environ 1% de la population suisse qui avait été infectée. Or pour notre implémentation utilisée du modèle SIR, il est difficile d'arriver à simuler une vague de moins de 1% de la population manuellement en jouant aléatoirement avec les paramètres. De plus, nous n'arrivions à simuler qu'une seule vague de contamination au sein de la population.

Après cette tentative infructueuse, nous nous sommes demandé si le modèle SIR s'adapterait mieux à d'autres pays dont la population a eu des comportements différents par rapport à la population helvétique ? Pour ce faire, nous nous sommes intéressés, entre autres, aux données de la Chine et des Etats-Unis.

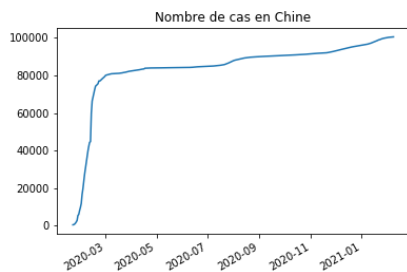


FIGURE 7 – Infections en Chine

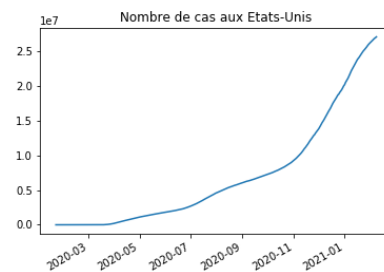


FIGURE 8 – Infections aux Etats-Unis

Puis à partir de ces données, nous avons essayé d'obtenir une courbe des réta-

blissements (*Recovered*) similaire avec le modèle SIR, sans succès.

Pour la Chine et les Etats-Unis, nous remarquons sur les figures 7 et 8 plusieurs vitesses différentes dans l'accélération du nombre de cas, donnée que nous n'arrivions pas à simuler avec le modèle SIR. Mais pourquoi n'arrivons-nous pas à obtenir des résultats moins "lisses" avec le modèle SIR ?

3.4 Qu'est-ce qu'un modèle mathématique ?

Avant de vouloir créer, adapter ou même utiliser un modèle, il faut répondre à la question centrale : quel est l'objectif de notre modélisation ?

Dans notre exemple ci-dessus, nous souhaitions voir si le modèle SIR pouvait simuler un graphe similaire à celui des données réelles collectées.

Il est aussi important de noter qu'un modèle est une simplification de la réalité, il ne sera donc jamais exactement juste. Mais plus nous rajoutons d'informations dans notre modèle, plus nous le complexifions. Ainsi il faut que les informations ajoutées à notre modèle l'améliorent sensiblement.

En résumé, un modèle mathématique doit être simple mais pas trop simple.

Nous pouvons alors nous demander si le modèle SIR est trop simple pour simuler un problème aussi complexe que la Covid-19.

Par exemple, est-ce que le modèle SIR(1) avec les paramètres fixés peut simuler une seconde vague ?

3.5 Simulation de deux vagues infectieuses

Afin de répondre à cette question, nous avons deux cas possibles. Soit nous pouvions trouver un exemple qui simule au moins deux vagues infectieuses, et ainsi cela prouverait que c'est possible. Soit nous pouvions prouver pourquoi cela n'est pas possible. Mais nous ne connaissions pas la réponse à cette problématique, alors nous avons commencé par le cas qui nous semblait le plus simple à prouver, le premier cas. Pour ce dernier, nous avons donc joué avec les paramètres à disposition mais n'arrivions jamais au résultat escompté. La preuve semblait plus complexe à réaliser. Nous avons procédé à plusieurs essais de preuve avant d'arriver à la démonstration qui permet de prouver l'affirmation suivante : le modèle SIR(1) avec les paramètres fixés ne peut pas simuler de seconde vague.

Preuve : nous utilisons le résultat d'un article [5] co-écrit par Ghulam Shabbir, Hina Khan et M. A. Sadiq qui nous fournit une solution exacte au système d'équation du modèle SIR. Les infectés au temps t s'obtiennent donc par l'équa-

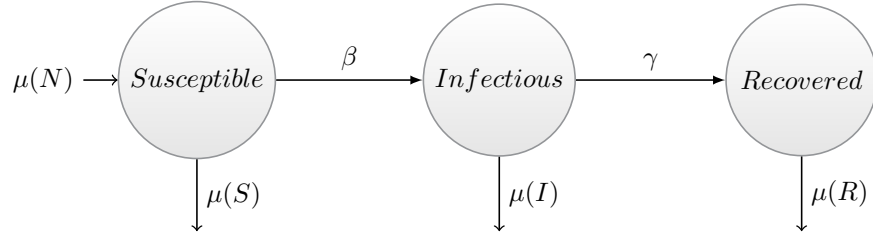


FIGURE 9 – Modèle SIR de l'article utilisé pour notre preuve

tion (5), où $\lambda = -\mu + \beta(S_0 + I_0)$.

$$I(t) = \frac{\lambda}{\beta + \lambda \left(\frac{\lambda - I_0\beta}{\frac{\beta(S_0 + I_0 - 1)}{\mu}} \right)} \cdot e^{-\lambda t + \frac{\beta(S_0 + I_0 - 1)}{\mu}} \quad (5)$$

Toutefois, le modèle SIR de l'article utilisé est légèrement différent de notre modèle SIR(1). En effet, il prend en compte un paramètre additionnel μ qui représente les naissances et morts naturelles comme illustré dans la figure 9. Mais puisqu'il ne modifie pas les hypothèses posées lors de notre définition du modèle, nous pouvons dès lors l'utiliser, et son résultat sera aussi valide pour notre affirmation.

Nous remarquons que dans cette équation, seule la partie $e^{-\lambda t + \frac{\beta(S_0 + I_0 - 1)}{\mu}}$ dépend du temps t .

Ainsi, nous pouvons réécrire cette équation sous la forme

$$\tilde{I}(t) = \frac{a'}{b' + c' \cdot e^{-a' \cdot t + d'}} = \frac{1}{b + c \cdot e^{-a \cdot t + d}} \quad (6)$$

où a', b', c', d', a, b, c et d sont des constantes, et donc elles ne dépendent pas du temps t .

Or, si nous supposons par l'absurde qu'un modèle SIR comporte deux vagues, nous remarquons une contradiction avec la fonction $\tilde{I}(t)$, car sa série ne peut changer de croissante à décroissante (ou inversement) une seule fois. Et il faudrait qu'elle puisse changer au moins deux fois pour avoir deux vagues.

Nous avons donc bien que le modèle SIR avec les paramètres fixés ne peut pas simuler deux vagues.

3.6 Simplicité du modèle SIR

Nous avons dans le cas de la Covid-19 discrétisé notre espace-temps en jours. Mais d'un jour à l'autre, les comportements des populations changent, et donc les paramètres β et γ aussi. Idéalement, il nous faudrait prendre des β_{t_i} où β_{t_i} représente le comportement β de la population au temps t_i . Ainsi, dans l'exemple de la Covid-19, une population qui ne porte pas masque de protection verra sa valeur de β diminuer lorsque le masque est rendu obligatoire. Mais avec le modèle SIR que nous avons défini, les paramètres ne dépendent pas du temps, et sont donc fixés.

Que se passe-t-il si les paramètres du modèle SIR ne sont plus fixés. Par exemple, est-ce une bonne idée de prendre β_t comme la moyenne des β_{t_i} . Mais cette dernière idée n'est pas bonne comme l'illustre l'exemple sur les figures 10 et 11. La première moitié du graphe a été calculée avec $\beta = 0.3$ (respectivement $\beta = 0.5$) puis la seconde moitié du graphe avec $\beta = 0.5$ (respectivement $\beta = 0.3$), tout en gardant les mêmes autres paramètres nécessaires au modèle SIR(1). Bien que la moyenne des β vaille 0.4 dans chaque exemple, le résultat obtenu est sensiblement différent. Nous observons bien deux graphes distincts pour chacune des trois fonctions S, I et R .

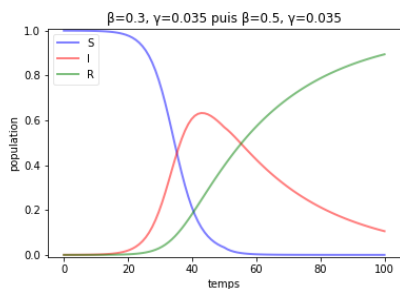


FIGURE 10 – β petit puis grand

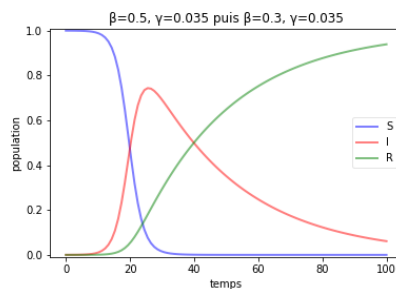


FIGURE 11 – β grand puis petit

Ainsi, même si nous ajoutons à notre modèle SIR des variables β et γ dépendantes du temps, donc non fixées, il n'est pas facile de trouver des valeurs adéquates. Nous pouvons aussi noter que dans ce cas la preuve de la section 3.5 ne fonctionne plus car les hypothèses ont été modifiées. Nous garderons ainsi l'hypothèse que paramètres sont fixés pour la suite du travail.

Parmi les hypothèses du modèle SIR, nous supposons que le système est fermé et homogène avec une population mixée. Or ce n'est jamais vraiment le cas, si nous prenons l'exemple de la Suisse, ses frontières ne sont pas fermées, chaque individu agit différemment, indépendamment et la population n'est pas totalement mélangée.

Et pour la Covid-19, une personne qui est saine ne devient pas directement infectieuse, elle passe d'abord par une période d'incubation. Or le modèle SIR

ne permet pas de modéliser cette période d’incubation. Afin que la modélisation fonctionne au mieux pour la Covid-19, il serait important d’intégrer cette notion à notre modèle.

4 Modèle SEIR

4.1 Définition

Comme certaines maladies infectieuses comprennent une période d’incubation avant d’être infecté, le modèle SEIR ajoute au modèle SIR un état supplémentaire possible pour la population. Il s’agit de du compartiment en période d’incubation (*Exposed*), donc les individus qui ont le virus dans leur corps mais qui ne peuvent pas infecter d’autres individus.

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -\beta I \frac{S}{N} \\ \frac{dE}{dt} = \beta I \frac{S}{N} - \sigma E \\ \frac{dI}{dt} = \sigma E - \gamma I \\ \frac{dR}{dt} = \gamma I \end{array} \right. \quad (7)$$

Par rapport au modèle SIR(1), voici les variables qui changent :

- β : ce paramètre définit l’incubation d’individus sains.
- σ : ce paramètre définit l’infection d’individus en période d’incubation.

Toutefois, ce modèle souffre des mêmes lacunes que le modèle SIR(1), il l’améliore seulement en apportant un nouvel état pour la population, utile pour des modélisations comme la Covid-19 par exemple.

Comme nous l’observons sur la figure 12 du modèle SEIR, il se comporte de manière similaire au modèle SIR.

4.2 Approximation du nombre de personnes contaminées dans le futur

Yougang Gu (YYG) était au milieu d’année 2020 un modèle assez précis pour la prévision du nombre de cas dans le futur [6]. Il s’agit d’une modélisation SEIR (*Susceptible – Exposed – Infectious – Recovery*), combinée avec de l’apprentissage automatique (machine learning). Le modèle s’adapte aux données par rapport aux morts confirmées en utilisant un taux de mortalité par infection (Infection Fatality Ratio IFR) afin de trouver à quels paramètres cela correspond, et donc rétro-calculer une estimation du nombre d’infections.

Parmi les hypothèses, nous admettons qu’il y a un nombre non négligeable de cas non reporté les premières semaines, puis que ce nombre diminue avec le temps

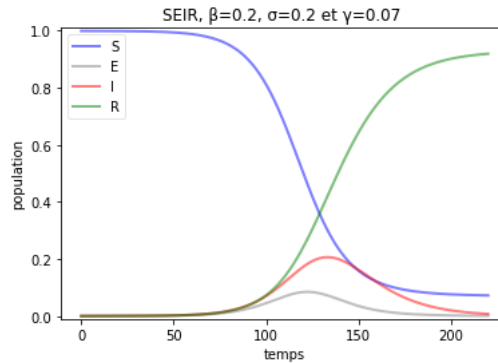


FIGURE 12 – Exemple graphe SEIR pour la Covid-19

(or ce n'est forcément pas le cas). Nous supposons aussi que la réouverture sociale affecte la distanciation sociale.

Afin d'être objectif dans notre critique du modèle, étant donné que la Covid-19 est récente, les modèles sont continuellement mis à jour, et ainsi des problèmes que nous pouvons mettre en avant peuvent être résolus par la suite. L'important pour nous est de voir à un moment donné qu'est-ce qui fonctionnait bien et que fallait-il changer dans un modèle donné.

Un point négatif de ce modèle est qu'il utilise uniquement le nombre de morts. Or, des données comme le nombre de cas et les hospitalisations sont utiles pour améliorer les modèles. Cette décision peut être expliquée par le fait que le nombre d'individus morts de la Covid-19 est l'information la plus exacte que nous pouvons avoir, car le nombre de cas par exemple dépend des tests effectués et des personnes symptomatiques, donc utiliser cette donnée peut entraîner des erreurs dans les projections.

De plus, chaque pays a un système de santé différent, et vu que ce modèle a été conçu par un Américain, il s'adapte peut-être moins bien aux autres pays qui ont des systèmes de santé différents. Utiliser les hospitalisations nous semblerait plus précis, mais aussi plus compliqué à implémenter pour la généralisation à tous les pays.

Et finalement, il ne prend pas en compte les saisons, alors que ceci a une influence sur la transmission du virus pour les projections à long terme.

Un point positif est qu'il est assez simple et s'exécute en 1h sur un ordinateur portable. Ainsi, il peut être mis quotidiennement à jour. De plus, en prenant moins de données il effectue moins d'hypothèses et ainsi minimise les erreurs.

Ce modèle donnait aussi une estimation du nombre de cas dans le futur. Toutefois, il est délicat et compliqué de faire ces projections. En effet, nous avons vu qu'à chaque itération par rapport au temps t , le modèle utilise l'état précédent

$t - 1$. Ainsi, une petite erreur à un moment donné dans les paramètres peut avoir un effet boule de neige et entraîner une erreur importante sur le résultat final. De plus, plus la prédiction sera lointaine, plus l'erreur pourra être grande.

5 Approximation du nombre de personnes contaminées hypothétiquement

5.1 Première procédure

Après avoir étudié les données passées dans le but d'effectuer des prévisions futures, nous allons répondre à la question suivante : si les individus d'un pays n'avaient pas changé leur comportement par rapport à un intervalle de temps donné, c'est à dire que les paramètres β, σ et γ restent constants, comment aurait évolué le virus par la suite ?

Afin de procéder à la réponse de cette question, nous allons décomposer le problème en plusieurs étapes.

1. Choisir un intervalle de temps à étudier, notons-le $T = [t_a, t_b]$ où $t_a < t_b < t_{\text{présent}}$.
2. Extraire les données des contaminations pour un pays donné pour l'intervalle de temps T .
3. Assigner une valeur scalaire à S_0, E_0, I_0 et R_0 .
4. Approximer les paramètres β, σ et γ de ces données pour l'intervalle de temps T .
5. Utiliser le modèle SEIR(7) avec les paramètres $\beta, \sigma, \gamma, S_0, E_0, I_0$ et R_0 obtenus précédemment.

Nous remarquons aisément que le point 1 est facile à implémenter, le point 2 est une extension de ce que nous avons déjà effectué dans ce travail et le point 5 est une application directe du modèle SEIR(7). Il nous reste ainsi à traiter les points 3 et 4.

5.1.1 Assignment d'une valeur scalaire à S_0, E_0, I_0 et R_0

Commençons par les valeurs les plus évidentes à affecter. Si nous modélisons le début de l'épidémie, nous pouvons naturellement supposer que $R_0 = 0$. Sinon, nous pouvons utiliser la somme cumulée des infectés environ 9 jours auparavant, puisque nous avons utilisé l'estimation qu'une personne infectée met environ 9 jours pour être dans le compartiment R [4]. De plus, nous avons les infections recensées chaque jour, donc nous connaissons la valeur de I_0 . Il nous reste à déterminer la valeur E_0 et S_0 . Mais puisqu'elles sont liées par l'équation $S_0 + I_0 + E_0 + R_0 = N$, il nous suffit d'assigner une valeur à une des deux variables, car N est connu. Nous avons effectué la supposition que $E_t = \frac{I_t + I_{t+1}}{2}$ où I_t et I_{t+1} sont connus.

Bien que nous avons dû effectuer des hypothèses supplémentaires, nous avons pu assigner une valeur à chacun des paramètres.

5.1.2 Estimation des paramètres β, σ et γ

Cette partie est la plus difficile à effectuer. Intuitivement, les trois paramètres β, σ et γ sont utilisés à chaque itération de temps dans notre modèle SEIR, et ils permettent de calculer les valeurs des compartiments SEIR pour l'itération suivante. Avec un algorithme naïf, nous choisirions aléatoirement trois valeurs pour nos paramètres puis effectuerions les T itérations du modèle SEIR et comparerions notre résultat des contaminations avec les données réelles.

Mais en utilisant des fonctions issues de la librairie Python sous licence libre (open-source) nommé Scipy, nous arrivons à estimer ces paramètres. En effet, `optimize.curve_fit` optimise à l'aide de la méthode des moindres carrés la fonction `fct_odeint`. Cette dernière intègre un système d'équation ordinaire. Ainsi, nous obtenons les paramètres β, σ et γ qui minimisent l'erreur par rapport à nos données réelles.

```
ydata = tableau_des_nouvelles_infections_des_jours_de_T
tdata = tableau_des_jours_de_T

def seir_model(y, t, beta, sigma, gamma):
    dS/dt = -beta * y[0] * y[2] / N
    dE/dt = beta * y[0] * y[2] / N - sigma*y[1]
    dI/dt = sigma*y[1] - (gamma*y[2])
    dR/dt = gamma * y[2]
    return S, E, I, R

def fit_odeint(t, beta, sigma, gamma):
    return integrate.odeint(seir_model, (S0, E0, I0, R0), t, args=(
        beta, sigma, gamma))[:,1]

popt, pcov = optimize.curve_fit(fit_odeint, tdata, ydata)
```

5.2 Interprétation des résultats

Maintenant que nous avons créé et analysé en détail notre procédure, nous pouvons l'utiliser pour des cas concrets.

Tout d'abord, nous avons trouvé intéressant de regarder quels résultats nous obtenions pour la Suisse. Ainsi, nous avons choisi l'intervalle de temps $t_a = 26.02.2020$ et $t_b = 22.03.2020$. Avec ces données, nous obtenons les résultats des figures 13 et 14. Nous trouvons ainsi que, sous les hypothèses de notre modèle et de notre procédure, 80% de la population aurait été contaminé. De plus, nous remarquons que la majorité des personnes seraient passées dans le compartiment R pendant le mois d'avril. Mais il est aussi intéressant de voir que toute la population n'aurait pas été contaminée. En effet, 20% est toujours dans le compartiment sain grâce à l'immunité collective.

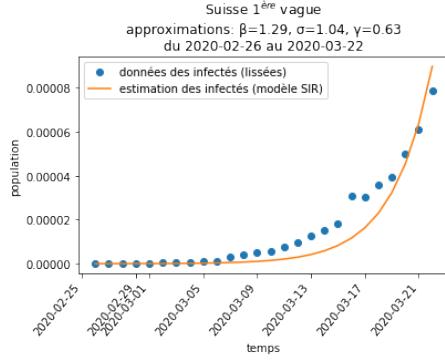


FIGURE 13 – point 4.

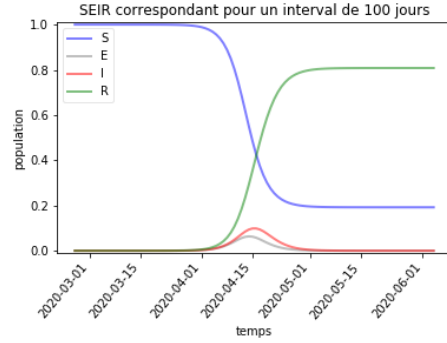


FIGURE 14 – point 5.

Nous avons bien précisé que ces résultats dépendaient des hypothèses émises. Afin de démontrer que les hypothèses d'origine jouent un rôle important dans les résultats de notre modélisation, nous allons effectuer la même procédure avec les mêmes variables, en augmentant uniquement I_0 de 2.5%. Nous obtenons les figures 15 et 16. Nous obtenons alors des valeurs différentes pour nos trois paramètres, ce qui donne un résultat différent pour l'évolution des contaminations. Nous remarquons ainsi que presque toute la population aurait été contaminée.

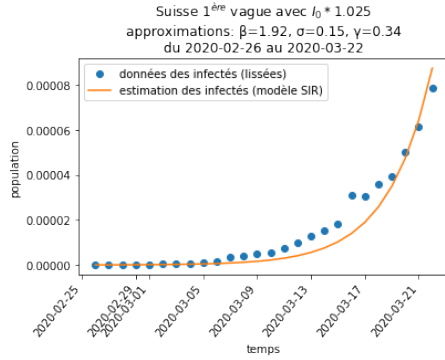


FIGURE 15 – Point 4.

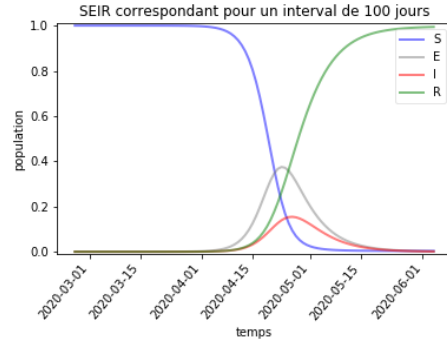


FIGURE 16 – Point 5.

De même, regardons ce qu'il se passe lorsque nous effectuons la même procédure mais en diminuant uniquement I_0 de 2.5%. Nous observons sur les figures 17 et 18 un résultat avec des paramètres proches aux paramètres de la figure 15. Toutefois, ici 98% de la population finirait contaminée, et 2% échapperait au virus.

De manière similaire, si nous choisissons un autre intervalle de temps, nous obtenons un résultat différent sur les figures 19 et 20. En avançant $t_a = 27.02.2020$, il y aurait 65% de la population qui serait contaminé.

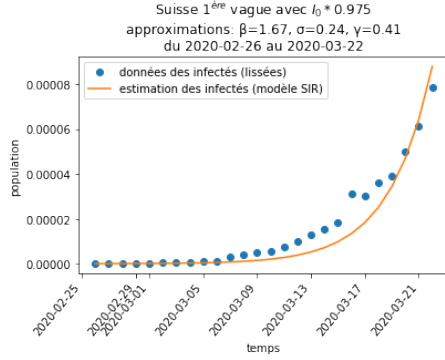


FIGURE 17 – Point 4.

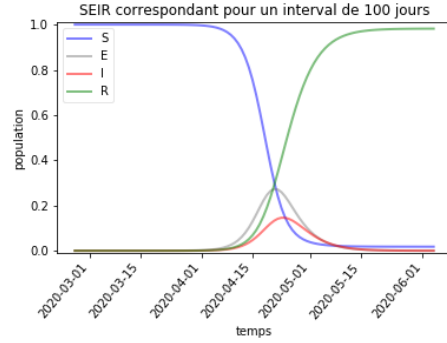


FIGURE 18 – Point 5.

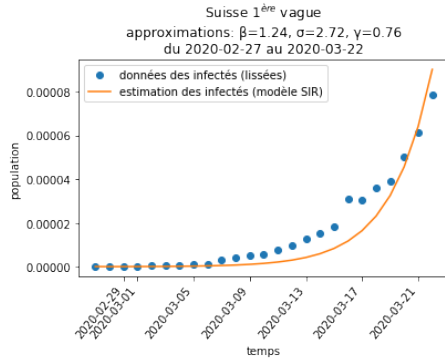


FIGURE 19 – Point 4.

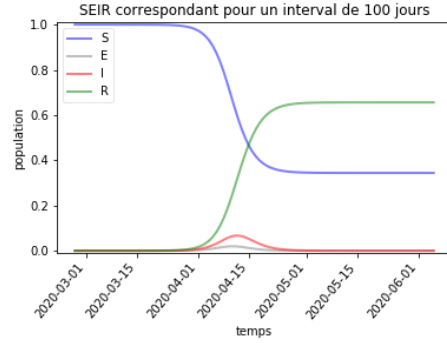


FIGURE 20 – Point 5.

Nous avons ainsi réussi à répondre à la problématique posée en début de chapitre. Mais il faut rester attentif quant à l'intervalle de temps choisi ainsi qu'aux conditions initiales utilisées pour résoudre notre problème.

Afin de tester notre algorithme, nous avons voulu voir s'il fonctionnait toujours ou bien s'il y avait des cas où il nous affichait un résultat erroné. Par exemple, que se passe-t-il si le compartiment $I_0 > E_0$?

Nous avons donc expérimenté ce cas de figure en prenant $E_0 = I_0 \cdot 0.95$, comme l'illustrent les figures 21 et 22. Nous voyons clairement que l'estimation des infectés ne correspond pas aux données collectées. De plus, il en résulte un paramètre incohérent avec la valeur négative de σ . Nous ne devrions pas effectuer le graphe car les données initiales sont incorrectes. Toutefois, nous l'avons quand même affiché afin d'illustrer qu'avec des hypothèses erronées, ce que nous avons dit précédemment ne s'applique pas forcément.

Ensuite, est-ce que les valeurs des paramètres peuvent correspondre à la réalité ? Pour cela, nous allons utiliser des approximations énoncées dans un article [7]

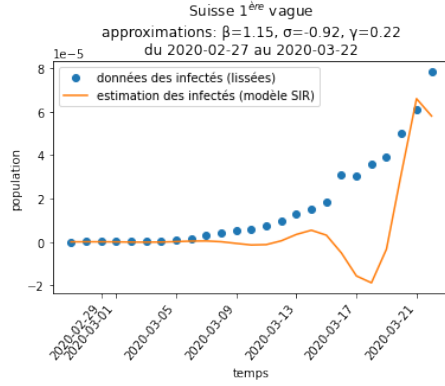


FIGURE 21 – Point 4.

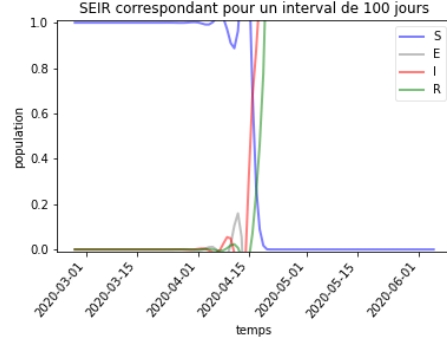


FIGURE 22 – Point 5.

qui indique les valeurs approximatives (8) de la Covid-19 pour le modèle SEIR.

$$\begin{aligned} 0.1 &\leq \beta \leq 0.3 \\ 0.07 &\leq \sigma \leq 0.5 \\ 0.05 &\leq \gamma \leq 0.1 \end{aligned} \quad (8)$$

Dans les exemples énoncés dans ce chapitre, nous n'avons jamais eu les trois paramètres qui étaient proches de ces trois critères simultanément. Toutefois, nous avons seulement étudié le début de l'épidémie, et il y avait les gestes barrières qui n'étaient pas encore présents au début de l'épidémie. Lors de la deuxième vague, observe-t-on des paramètres qui correspondent mieux aux approximations énoncées en 8 ? Nous remarquons sur la figure 23 que ce n'est pas forcément le cas. Ainsi, notre algorithme nous retourne des paramètres mathématiquement et théoriquement corrects. C'est à dire le résultat le plus proche mais il peut exister d'autres paramètres qui sont plus proches de la réalité mais plus éloigné au sens de la distance par rapport au résultat obtenu. Malheureusement, nous ne pouvons pas définir un intervalle borné pour chaque paramètre avec la méthode que nous avons proposé.

5.3 Deuxième procédure

Nous avons précédemment créé une méthode qui, à partir d'hypothèses fournies par rapport aux paramètres initiaux, nous retourne une estimation des paramètres β, σ et γ . Donc les deux problèmes auxquels nous sommes confrontés sont le fait que nous avons émis des hypothèses supplémentaires et que nous ne pouvons pas restreindre d'intervalle pour les paramètres. Afin de remédier à ceci, nous avons cherché une autre méthode qui ne contiendrait pas ces deux contraintes.

Ainsi, le professeur Rolim m'a orienté vers un article [3] qui pourrait nous être

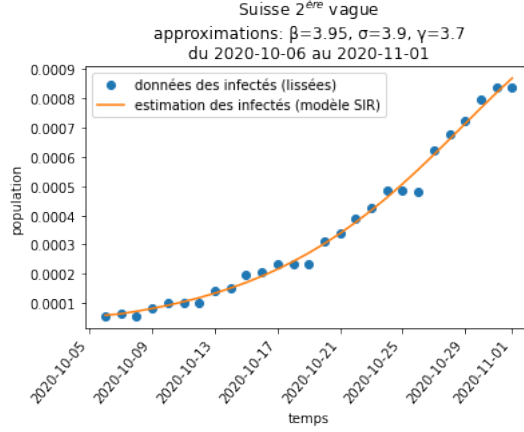


FIGURE 23 – Deuxième vague en Suisse

utile. En effet, nous avons utilisé des propositions de cet article afin de créer une deuxième procédure. Cette dernière s'utilise de la manière suivante :

1. Choisir un intervalle de temps à étudier, notons-le $T = [t_a, t_b]$ où $t_a < t_b < t_{\text{présent}}$.
2. Extraire les données des contaminations pour un pays donné pour l'intervalle de temps T .
3. Attribuer un interval, borné ou non, à chacun des paramètres (optionnel), et donner une estimation des conditions initiales des paramètres.
4. Assigner une valeur aux paramètres N et I_0 connus.
5. Approximer les paramètres $\beta, \sigma, \gamma, S_0$ et E_0 de ces données pour l'intervalle de temps T .
6. Utiliser le modèle SEIR(7) avec les paramètres $\beta, \sigma, \gamma, S_0, E_0, I_0$ et R_0 obtenus précédemment.

L'avantage de cette méthode est qu'elle n'a pas besoin d'hypothèse forte quant aux conditions initiales. Les valeurs initiales soumises aident à la convergence de la résolution du problème et sont souvent différentes de celles obtenues dans le résultat. De plus, nous pouvons à présent borner nos paramètres. Ceci est très utile car nous n'aurons ainsi plus de comportement étrange comme avec la précédente méthode où nous pouvions avoir des valeurs de paramètre négatives. Et puisque nous avons une approximation de ceux-ci pour la Covid-19 (8), nous avons pris par exemple ces paramètres avec une marge de 20% par rapport au minimum et maximum. En procédant ainsi, nous avons obtenu le résultat de la figure 24.

Il est intéressant de la comparer avec la figure 25 car il s'agit du cas identique, avec les mêmes valeurs de départ utilisées mais avec la première procédure. Ainsi,

nous remarquons un comportement global assez semblable. Nous pouvons voir toutefois un pic d'infectés plus fort avec la deuxième méthode, et les courbes S et R sont moins rigides avec la deuxième méthode. Mais nous remarquons que les dates des pics d'infections ne correspondent pas entre les deux graphes. Ceci peut être dû au fait que les conditions initiales soient plus agressives dans la première méthode ce qui induit une infection globale plus rapide.

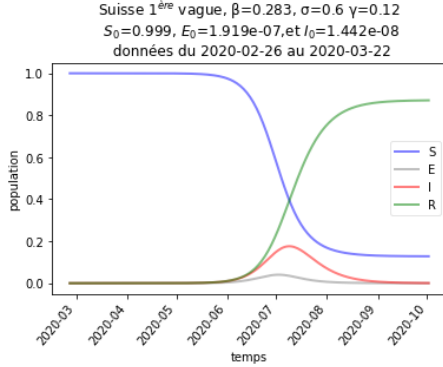


FIGURE 24 – Deuxième méthode

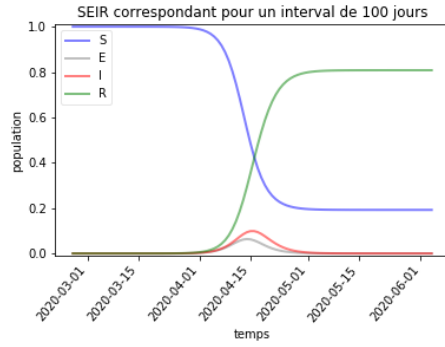


FIGURE 25 – Première méthode (figure 14)

Après de multiples essais, lorsque nous utilisons des bornes inférieures et supérieures (par exemple (8)) pour les paramètres, les résultats obtenus ne sont pas toujours cohérents. Mais si nous indiquons uniquement la borne inférieure (0.001 par exemple), les résultats sont meilleurs. Toutefois, les valeurs des paramètres n'appartiennent pas à l'intervalle souhaité (8). Les deux procédures que nous avons décrites pour estimer les paramètres utilisent les infections recensées, pas forcément proches de la réalité. Pourrions-nous faire différemment ?

6 Troisième procédure utilisant une estimation du nombre d'individus infectés

6.1 Impact sur les décisions politiques

Tout d'abord, l'Institute for Health Metrics and Evaluation (IHME) utilise le modèle SEIR. Il essaie d'adapter les données à l'aide d'un modèle qui utilise notamment les cas confirmés, les morts confirmés ainsi que les données des tests effectués. De plus, il utilise un taux de mortalité par infection (IFR) estimé, et dépend aussi de plusieurs hypothèses, ainsi que de leur pondération.

Il peut être utile aux politiques afin d'estimer quel impact aura une décision destinée à limiter la propagation du virus, afin de choisir par exemple des restrictions ayant un fort effet sur le virus et un faible effet économique. Il permet aussi d'estimer le nombre de cas total d'individus infectés.

Un problème de ce modèle est qu’au début il n’utilisait pas le modèle SEIR. Il utilisait uniquement les données passées où la Covid-19 avait fortement touché les villes/pays, puis adaptait ces données aux données américaines. Il a été vivement critiqué car c’était le modèle le plus optimiste, et les politiciens américains l’utilisaient. Après avoir remarqué que leurs projections ne correspondaient pas à la réalité, ils ont intégré le modèle SEIR afin d’améliorer leurs projections.

6.2 Modèles pour les cas non détectés

Combien y a-t-il réellement de cas ? C’est une question que nous nous sommes tous posée avec la Covid-19. Dans l’idéal, il aurait fallu augmenter le nombre de test effectués mais nous connaissons les difficultés auxquelles les Etats se sont confrontés avec les tests PCR. Ainsi, est-ce que nous pourrions estimer le nombre de cas total ?

Par exemple, la London School of Hygiene and Tropical Medicine (LSHTM) a créé un algorithme qui est utilisé pour mieux comprendre le niveau de progression indétectée de la Covid-19 parmi la population.

Nous avons vu précédemment que les modèles YGG et IHME utilisaient le nombre de morts afin d’estimer le nombre d’infectés.

Il y a aussi le modèle Imperial College London (ICL) qui permet, entre autres, de fournir une estimation du nombre de cas d’infectés.

6.3 Troisième procédure

Jusqu’à présent, nous avons essayé de trouver comment le modèle s’adaptait aux données collectées par les autorités avec deux méthodes distinctes. Toutefois, les paramètres du modèle SEIR ne correspondent pas à ce que nous devrions obtenir pour la Covid-19.

De plus, nous venons de parler d’algorithmes qui utilisent les morts afin, entre autres, d’estimer le nombre d’infections total, c’est à dire les infections recensées par les gouvernements ainsi que les infection non recensées.

Est-ce que nous arriverions mieux à approximer les paramètres du modèle SEIR en utilisant une approximation du nombre de cas réel qui prend en compte, entre autres, le nombre de décès parmi les individus.

Ainsi, nous allons procéder de la façon suivante :

1. Choisir un pays et un intervalle de temps à étudier, notons-le $T = [t_a, t_b]$ où $t_a < t_b < t_{\text{présent}}$.
2. Sélectionner une des quatre méthodes qui utilise les morts recensées, la compiler, puis enregistrer la moyenne de ses résultats (infections) retournés.
3. Utiliser la procédure présentée dans la sous-section 5.3.

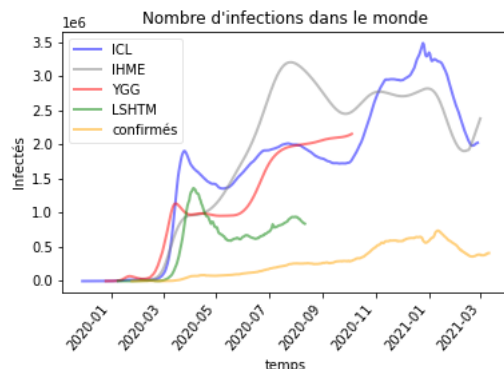


FIGURE 26 – Infections mondiales

Comme nous pouvons l’observer sur le graphe (26), chacune des méthodes estime un nombre différent d’infections. Donc nous ne pouvons malheureusement pas affirmer avec certitude le nombre total d’individus infectés. Toutefois, ces quatre méthodes sont d’accord avec le fait que le vrai nombre d’infectés est supérieur au nombre de cas recensé.

Notons tout de même que chacune de ces méthodes ne nous retourne pas un unique scalaire mais plusieurs valeurs avec un intervalle de confiance. Et nous avons décidé d’utiliser la médiane pour LSHTM et la valeur moyenne pour les autres afin d’en faire une interprétation. LSHTM et YGG ont arrêté de fournir des données respectivement au mois d’août et d’octobre 2020. Pour YGG, cet arrêt s’est expliqué par le fait que le projet prenait trop de temps à son créateur indépendant car il nécessitait d’être amélioré et mis à jour régulièrement.

Après avoir essayé ces quatre méthodes distinctes avec la procédure ci-dessus, appliqué à plusieurs pays, avec des dates différentes, tous les paramètres n’appartiennent pas aux bornes énoncées en (8), rappelons-les : $0.1 \leq \beta \leq 0.3$, $0.07 \leq \sigma \leq 0.5$ et $0.05 \leq \gamma \leq 0.1$. Nous allons toutefois parler de certains résultats intéressants.

6.4 Interprétation des résultats

Comment confronter nos deux procédures, c’est-à-dire la comparaison entre l’ancienne du chapitre 5.3 qui utilise les infections recensées (notée P2) et la nouvelle du chapitre 6.3 qui utilise des infections simulées à l’aide d’un des modèles (notée P3).

Une idée serait de calculer soit l’erreur potentielle maximale, soit l’espérance de l’erreur pour un grand nombre d’échantillons d’exemples afin de pouvoir affirmer de manière rigoureuse qu’une procédure est meilleure qu’une autre. Dans notre cas, l’erreur serait la différence entre les paramètres estimés et les paramètres réels, nous pourrions appliquer la méthode des moindres carrés par exemple.

Mais le premier problème est que nous ne connaissons pas exactement la valeur des paramètres réels c'est pourquoi cette première idée n'est pas envisageable. Le second problème est que nous choisissons manuellement les dates, ce qui ne convient pas pour un grand échantionnage. Nous avons alors choisi de prendre une dizaine d'exemples en choisissant des pays sur divers continents et en sélectionnant des intervalles de dates pertinents et de les recenser dans le tableau 1. La colonne nommée "Inf. expo" signifie infections exponentielles, et représente donc si dans l'intervalle de temps donné pour le pays sélectionné, les nouveaux cas apparaissent de manière exponentielle. Et pour chacun des exemples, nous effectuons P2 (infections recensées, chapitre 5.3) et P3 (infections simulées, chapitre 6.3), et regardons les valeurs des paramètres. Nous avons utilisé aléatoirement une des méthodes disponible. Si nous changeons de méthode, le résultat est bien entendu différent car les données d'entrée des infections sont différentes. Mais la tendance générale que nous verrons par la suite ne change pas.

Pays	Inf. expo	Dates	Infections	β	σ	γ
Suisse	oui	20-07-20 au 20-10-01	recensées	0.1843	0.2624	0.0887
			simulées	1.2527	0.001	0.001
Suisse	oui	20-10-02 au 20-10-28	recensées	1.0044	1.8118	0.9167
			simulées	0.2402	0.2160	0.1440
Suisse	non	20-11-01 au 21-01-01	recensées	3.3345	0.7095	3.2691
			simulées	2.4834	0.5353	2.4334
France	oui	20-07-20 au 20-10-01	recensées	0.0191	0.7357	0.001
			simulées	0.0513	0.4184	0.0343
France	oui	20-07-22 au 20-09-30	recensées	0.7293	5.0540	0.7091
			simulées	0.3930	0.4809	0.3570
France	non	20-05-01 au 20-07-01	recensées	5.4171	1.0679	5.4696
			simulées	5.9903	0.9935	6.1030
Etats-Unis	oui	20-02-09 au 20-02-29	recensées	0.0447	3.6875	0.001
			simulées	0.4166	0.3129	0.08
Afrique du Sud	oui	20-04-20 au 20-05-17	recensées	1.1588	0.0020	0.0010
			simulées	1.1895	0.0034	0.0224
Afrique du Sud	non	20-09-20 au 20-10-25	recensées	1.1333	0.0058	0.1440
			simulées	3.2155	0.7947	3.1326
Inde	oui	20-03-20 au 20-05-17	recensées	0.0292	11982	0.001
			simulées	0.7057	0.0055	0.0444
Inde	oui	20-04-20 au 20-06-08	recensées	0.4920	0.4749	0.4460
			simulées	0.4161	0.2785	0.3744
Inde	non	20-09-01 au 20-11-01	recensées	6.6674	2.0345	6.6383
			simulées	2.3339	0.4924	2.2808

TABLE 1 – Tableau comparatif

Pour une raison technique au code de notre algorithme, nous avons dû choisir une borne minimale valant 0.001 pour P2 et P3. Lorsque ce nombre apparaît dans

la valeur d'un paramètre, l'algorithme est probablement resté bloqué dans un minimum local. Donc il faudrait relancer l'algorithme avec des données d'entrée différentes afin d'arriver à une meilleure solution. C'est ce que nous avons fait dans le tableau 1, sauf pour les Etats-Unis, nous n'avons pas réussi à trouver un interval de dates durant le mois de février 2020 qui ne reste pas bloqué avec P2, alors que les résultats de P3 sont cohérents avec bornes souhaitées.

Analysons maintenant les données récoltées dans le tableau (1). Nous remarquons que chacune des deux procédures est potentiellement restée bloquée dans un minimum local, nous avons surligné chacune de ces occurrences en bleu afin de les distinguer. Dans le tableau, P2 comporte la valeur 0.001 trois fois alors que P3 la présente une fois. En effet, au cours d'expérimentations afin d'obtenir des exemples, nous avons remarqué que P2 nous rend plus fréquemment cette valeur. Ceci peut s'expliquer par le fait que P3 utilise des infections qui sont plus lisses que P2 comme nous pouvons l'observer sur le graphe 28, donc P3 a probablement moins de chance de rester bloqué dans un minimum local lors du calcul des paramètres.

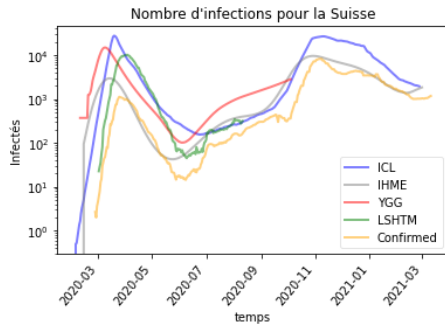


FIGURE 27 – Echelle logarithmique

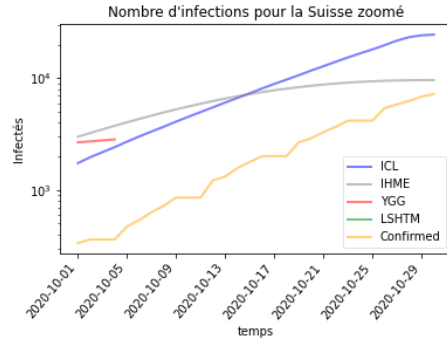


FIGURE 28 – Deuxième vague

Lorsque ce n'est pas le cas, la première observation évidente est pour les lignes où les infections ne sont pas exponentielles. Puisque le modèle SEIR modélise les infections de façon exponentielle, si nous lui demandons d'approximer les paramètres pour des données non exponentielles ou des données avec plusieurs vagues d'infection, les valeurs des paramètres obtenues seront incohérentes pour P3. Intuitivement, il est clair que si les données ne sont pas de la forme attendue alors le résultat ne doit pas être cohérent. Même si nous ne prouvons rien, nous pouvons penser qu'il s'agit d'une propriété de notre troisième procédure.

Puis, analysons nos données restantes, c'est-à-dire où les infections sont exponentielles et la valeur des paramètres est plus grande que 0.001. Pour la Suisse, P3 est très proche des bornes alors que P2 est éloignée. Pour la France, P2 nous rend un résultat incohérent avec $\sigma > 5$ alors que P3 est sensiblement plus proche des bornes. Pour l'Afrique du Sud, β n'est pas proche du résultat escompté pour les deux procédures. Pour l'Inde, P3 est plus proche des bornes que P2 mais

assez éloignée pour σ .

Ainsi, au vu des résultats analysés ci-dessus, il semblerait que P3 soit plus précise que P2, mais n'est malheureusement pas optimale. Cette tendance peut s'expliquer par le fait que certains de ces modèles qui simulent le nombre de cas utilisent un modèle SEIR dans leur implémentation, ce qui favorise donc le calcul des paramètres. Nous pouvons aussi évoquer le fait que puisque nous travaillons à partir des morts recensées, il est possible que l'approximation des contaminations soit meilleure et ainsi que les paramètres recherchés obtenus soient plus précis.

7 Modèle SEIRS

Nous avons vu que parmi les modèles utilisés plusieurs utilisent à un moment donné le modèle SEIR. Est-ce la meilleure solution ? Est-ce que le modèle SEIRS (*Susceptible – Exposed – Infectious – Recovery – Susceptible*) ne serait pas mieux adapté puisqu'il a été prouvé qu'une personne peut attraper deux fois le virus ? En effet, avec le modèle SEIRS, les individus contaminés dans le passé peuvent être contaminés à nouveau.

Etant donné que ce cas n'est pas représentatif de l'ensemble de la population actuellement, nous y négligeons. Cependant suivant l'évolution du virus dans la population, il se peut que ces algorithmes se mettent à utiliser le modèle SEIRS si les experts virologues estiment qu'une partie non négligeable de la population peut être infectée une seconde fois. De plus, avec l'apparition de variants de la Covid-19, il n'est pas exclu qu'un variant mute suffisamment pour se propager à nouveau au sein d'une population qui était auparavant dans le compartiment R .

8 Conclusion

Lors de ce travail, nous nous sommes d'abord intéressés à l'étude d'un système dynamique à temps continu, le modèle SIR compartimental. Après avoir énoncé sa définition, répertorié ses différents comportements possibles et abordé le nombre de reproduction de base R_0 , nous avons réussi à démontrer que nous ne pouvons pas simuler deux vagues d'infections successives au sein d'une population donnée avec le modèle SIR(1).

Ensuite, puisqu'il fonctionne relativement bien pour certaines maladies infectieuses, nous avons regardé le modèle SEIR qui ajoute un compartiment pour l'incubation du virus afin de perfectionner sa modélisation. Mais ces modèles sont caractérisés par des paramètres qui nous sont inconnus. C'est pourquoi une grosse partie de notre travail a été de chercher à approximer ces paramètres, pour un intervalle de temps fermé précis, à partir des données récoltées. Nous avons commencé avec une première procédure (chapitre 5.1), nous devons effectuer plusieurs hypothèses de départ ce qui fragilisait notre résultat. Ensuite

nous avons utilisé un nouvel algorithme (chapitre 5.3) qui nous permettait de ne plus émettre d’hypothèse supplémentaire.

Ces deux premières techniques utilisent comme données d’entrée les infections recensées par les gouvernements. Afin d’améliorer l’approximation des paramètres de notre système, nous avons finalement programmé une troisième procédure (chapitre 6.3) qui utilise une approximation du nombre d’individus infectés en fonction de différentes données. Cette dernière tentative, bien qu’imparfaite, est celle qui est la plus stable et semble nous rendre les meilleurs résultats pour la Covid-19 parmi nos trois procédures testées.

Références

- [1] David Easley et Jon Kleinberg. *Networks, Crowds, and Markets*. Cambridge University Press, 2010.
- [2] Ottar N. Bjørnstad. *Epidemics Models and Data using R*. Springer, 2018.
- [3] Aaron A. King. *Introduction to inference : parameter estimation*.
<https://kingaa.github.io/clim-dis/parest/parest.html>
- [4] Harvard Medical School. *If you’ve been exposed to the coronavirus*.
<https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus>
- [5] Ghulam Shabbir, Hina Khan et M. A. Sadiq. *A note on Exact solution of SIR and SIS epidemic models*.
www.researchgate.net/publication/48173364_A_note_on_Exact_solution_of_SIR_and_SIS_epidemic_models
- [6] Youyang Gu. *A Simple Machine Learning Model to Forecast COVID-19 Deaths*.
<https://www.youtube.com/watch?v=f88bYf1JYEo>
- [7] Anca Rădulescu, Cassandra Williams et Kieran Cavanagh. *Management strategies in a SEIR-type model of COVID 19 community spread*.
<https://www.nature.com/articles/s41598-020-77628-4>