

# Covid-19

**La modélisation a-t-elle conduit à sous-estimer :**

- **la rapidité du développement de l'épidémie en France au printemps 2020 ?**
- **la vigueur de sa reprise à l'automne ?**

**François Xavier MARTIN**

Ingénieur de l'Ecole Polytechnique et de Supélec

[fxm@wanadoo.fr](mailto:fxm@wanadoo.fr)

*L'annexe 4 a été réalisée en commun avec Mathilde VARRET, biologiste à l'INSERM et François VARRET, ancien élève de l'Ecole normale supérieure (Ulm), Professeur émérite de physique à l'Université de Versailles-Saint-Quentin-en-Yvelines.*

# **Covid–19 Pandemic in 2020**

**Did flaws in epidemic modeling lead to underestimate**

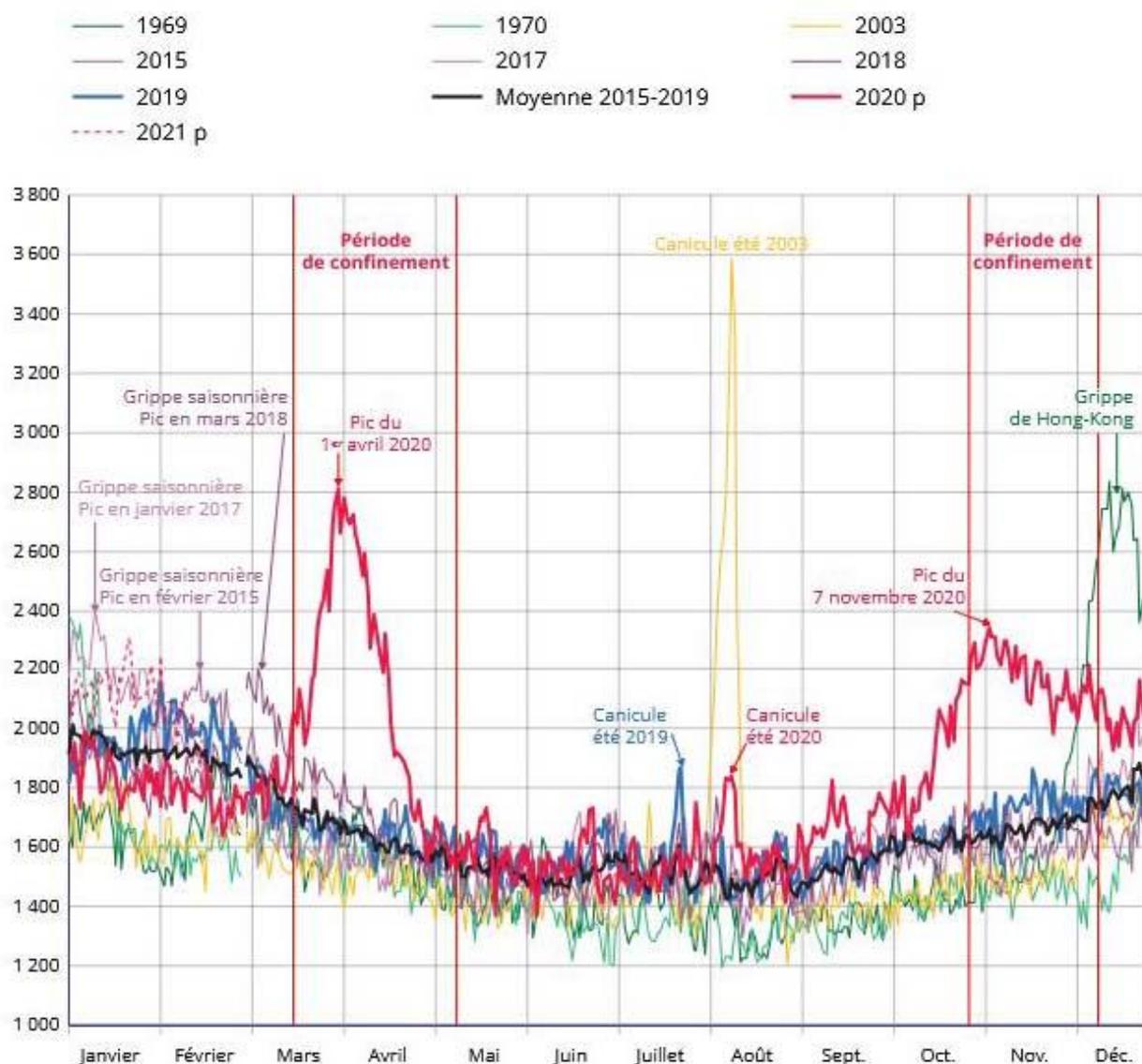
- **the initial spread rate ?**
- **the post lockdowns resurgence intensity ?**

*An Analysis Based on the Evolution of COVID-19 in France*

**François Xavier MARTIN**  
fxm@wanadoo.fr

Appendix 4 was jointly developed with Mathilde VARRET, a biologist at INSERM, and François VARRET, Emeritus Professor of Physics at the University of Versailles Saint-Quentin-en-Yvelines, alumnus of the École Normale Supérieure in Paris.

Figure 1 - Décès quotidiens en 2020 comparés aux cinq dernières années, à la canicule de 2003 et à la grippe de Hong-Kong (1968-1970)



Origine : INSEE

## Résumé

Le point de départ de la modélisation mathématique actuelle des épidémies remonte à une communication d'une vingtaine de pages faite en 1927 par Kermack et McKendrick à la Royal Society de Londres. Cette communication inclut la description théorique d'un modèle appelé SIR ainsi que le jeu d'équations bien connu de tous ceux qui s'intéressent à cette question (voir Annexe 2). **Or ces équations sont uniquement valides dans des périodes où l'épidémie est stabilisée ou évolue très lentement, comme indiqué explicitement par leurs auteurs**, pour qui il s'agit alors d'un :

**« special case » at « constant rates »**

*Ces « rates » incluent en particulier un « recovery rate  $\psi(t)$  » (taux de fin d'infection et de contagiosité - par guérison ou décès - dans une population de personnes contaminées et contagieuses) constant (ou quasi-constant) dans le temps, quelle que soit la date de début d'infection de chacune de ces personnes.*

**Point surprenant, la restriction (« at constant rates ») n'est pratiquement jamais mentionnée dans les cours de base de modélisation qui commencent invariablement par les « équations de 1927 » présentées comme universelles.** Or cette condition (« recovery rate » constant dans le temps) n'est respectée que dans les moments où le coefficient de reproduction  $R_{\text{effectif}}$  est voisin de 1 !

Comme le souligne le Professeur Odo Diekmann, modélisateur de l'Université d'Utrecht mondialement reconnu par les spécialistes de la discipline, elle semble également ignorée de nombreux universitaires qui s'ingénient à « complexifier pour affiner » (pour citer Cédric Villani) par des moyens purement mathématiques le modèle SIR de base en partant des équations de 1927.

Or les autorités sanitaires et plus généralement **les pouvoirs publics ont besoin de modèles tout particulièrement efficaces pendant les périodes de développement dynamique d'une épidémie** (arrivée des premiers malades, début d'un confinement, fin d'un confinement, arrivée d'un nouveau variant, campagne de vaccination, ...) où l'utilisation d'équations uniquement adaptées à des situations statiques conduit à **des prévisions de variations dans les nombres d'infections (à la hausse comme à la baisse) trop lentes et trop faibles par rapport à la réalité.**

Les grands organismes qui disposent d'équipes de modélisation importantes assurent que leurs modèles incluent des correctifs mathématiques qui permettent de porter remède à cette situation. L'observation des nombreuses statistiques relatives à l'épidémie ayant commencé début 2020 permet d'en douter. Le conseil scientifique Covid-19 a lui-même relevé dans certains de ses rapports combien certaines des montées et des descentes du nombre des infections quotidiennes avaient été plus brutales que les prévisions des experts officiels.

Cette situation est d'autant plus regrettable que des méthodes de modélisation dites « discrètes », jusqu'ici peu employées, permettent de porter facilement remède aux erreurs basées sur l'utilisation des équations de 1927 hors de leur domaine de validité. De plus, ces méthodes présentent l'avantage de ne pas demander de connaissances particulières en mathématiques et de permettre une extrême souplesse dans l'introduction des hypothèses les plus variées.

Il paraîtrait donc judicieux de bâtir des formations à cette technique et de formaliser quelques modèles pouvant être utilisés de façon complètement ouverte, avec des calculs vérifiables car facilement reproductibles, dans la perspective d'arrivée de futures épidémies.

## Abstract

The origins of modern mathematical modeling of epidemics date back to a 20-page communication presented in 1927 by Kermack and McKendrick to the Royal Society of London. This communication included the theoretical description of a model called SIR, along with a set of equations familiar to anyone interested in this topic (see Appendix 2). However, these equations are only valid during periods when the epidemic is stabilized or evolving very slowly, as explicitly stated by their authors, for whom they represent a :

### "special case" at "constant rates"

These "rates" include a constant (or quasi-constant) "**recovery rate  $\psi(t)$** " (the rate at which infections and contagiousness end—whether through recovery or death—among a population of infected and contagious individuals), regardless of the infection start date for each individual.

Surprisingly, this limitation ("at constant rates") is almost never mentioned in basic modeling courses, which invariably begin with the "1927 equations" presented as universal. However, this condition (constant "recovery rate" over time) is only met when the effective reproduction coefficient  $R_{\text{eff}}$  is close to 1 !

As emphasized by Professor Odo Diekmann, a recognized modeler from Utrecht University, this limitation also seems overlooked by many academics who, according to Cédric Villani, Field Medal attempt to "complexify for refinement" without the 1927 SIR model equations.

Unfortunately, public health authorities, and more broadly public policymakers, require models that are particularly effective during periods of dynamic epidemic development (arrival of initial cases, start of lockdown, end of lockdown, emergence of new variants, vaccination campaigns, etc.). Using equations suited only to static situations leads to forecasts of infection variations (both increases and decreases) that are too slow and too weak compared to reality.

Major organizations with extensive modeling teams claim that their advanced models incorporate mathematical adjustments to address this issue. However, observations of various statistics related to the epidemic that began in early 2020 cast doubt on this claim. In fact, **the French COVID-19 Scientific Council highlighted in several reports that certain spikes and drops in daily infection numbers were far more abrupt than the projections provided by official experts.**

This situation is particularly regrettable because "discrete" modeling methods, which have been underutilized so far, can easily correct errors resulting from the use of the 1927 equations outside their domain of validity. Furthermore, these methods have the advantage of requiring no specialized mathematical knowledge and allow for extreme flexibility in introducing a wide variety of assumptions.

It would therefore be wise to develop training programs in this technique and formalize a few models that could be used in a completely open manner, with calculations that are verifiable and easily reproducible, in anticipation of future epidemics.

# Sommaire

1 - Rappel historique

2 – Rappel de quelques principes généraux sur la modélisation mathématique

2.1 – Modèles « à éléments finis » (*appelés également* « discrets »)

2.2 - Modèles faisant appel au calcul infinitésimal

2.3 - Avantages et inconvénients comparés des deux méthodes

3 - Situation mondiale début 2020 de la modélisation mathématique des épidémies et de son enseignement

4 – Constatations troublantes faites lors de la récente épidémie de Covid-19

5 – Tentatives d'utilisation de tableurs

6 – L'attitude de Cédric Villani dans sa note dans sa note à l'Office parlementaire d'évaluation des choix scientifiques et technologiques d'avril 2020

7 – Propositions

7.1 – Une réforme de l'enseignement de la modélisation des épidémies afin que ce sujet soit accessible à d'autres que des mathématiciens de haut niveau (en particulier à une partie significative des médecins ainsi qu'à des biologistes)

7.2 – La préparation d'une politique de modélisation permettant la détermination et la justification crédible auprès de l'opinion publique de la conduite à tenir en cas de nouvelle crise sanitaire

## Annexes

**Annexe 1** - Réactions opposées obtenues d'un des principaux modélisateurs français contacté sur le conseil de Cédric Villani ... et de membres du CNRS

**Annexe 2** - Analyse de la « contribution à la théorie mathématique des épidémies » de Kermack et McKendrick » (1927)

**Annexe 3** - Exemple d'utilisation d'un tableur pour le modèle SIR proposé en avril 2020 à des étudiants (licence) en mathématiques.

**Annexe 4** – « A tentative tracking of the SARS-Cov2 pandemic in France, based on a corrected SIR model including vaccination effects »

## Bibliographie

# 1 - Rappel historique

Les premières tentatives de modélisation mathématique des épidémies remontent à la fin du 18<sup>ème</sup> siècle (Bernoulli). Après les avancées du début du 20<sup>ème</sup> siècle (Hamer, Ross) le point de départ réel de la modélisation moderne a été la communication d'une vingtaine de pages de Kermack et McKendrick à la Royal Society de Londres (1927).

Cette communication comprend plusieurs parties :

- Tout d'abord la description d'un modèle « compartimental » dit SIR où la population affectée par une épidémie est divisée **à l'instant t** en 3 catégories : S individus susceptibles d'être contaminés (habituellement au départ l'ensemble de la population), I contaminés et réputés contagieux pendant leur infection, R contaminés puis décédés, ou guéris et de ce fait immunisés (et ne pouvant plus dans ces deux cas transmettre l'infection). Cette description est basée sur une technique mathématique dite « discrète » ou « d'éléments finis » (voir la définition de ces termes au paragraphe 2.1)
- Ensuite des calculs mathématiques retenant le principe des 3 compartiments, mais passant de la technique « discrète » à une utilisation du calcul infinitésimal (dérivées, intégrales, équations différentielles, exponentielles ...). Les calculs mentionnés dans la communication de 1927 sont très difficilement compréhensibles, et certains mathématiciens prétendent même les avoir mis en défaut dans l'exemple d'application qui en est faite à l'épidémie de peste de 1905-1906 à Bombay
- Enfin la proposition de certaines simplifications afin de pouvoir simuler l'évolution générale dans la durée d'une épidémie avec la seule méthode utilisable à une époque où les ordinateurs n'existaient pas : le calcul infinitésimal et les équations différentielles (équations que l'on ne sait résoudre que si on peut les ramener à des cas connus).

Dans l'espoir (qui s'est jusqu'à maintenant révélé vain !) d'apporter au modèle général des simplifications permettant d'obtenir des équations différentielles (relations entre S, I, R et leurs dérivées) pour lesquelles les mathématiciens pourraient trouver des solutions exprimant de façon explicite l'évolution de S, I et R en fonction du temps, Kermack et McKendrick ont envisagé ce qu'ils appellent clairement des « special cases » dont en particulier un à « constant rates » postulant qu'il existe pour l'ensemble ( I ) de la population qui a le statut d'infecté à l'instant t un « recovery rate » (taux de guérison ou de décès) **constant** pendant la durée de déroulement du modèle.

Or pendant les phases critiques telles que le début de l'épidémie, la mise en place d'un confinement et sa levée, l'arrivée d'un nouveau variant, ... les flux journaliers de nouveaux infectés sont éminemment variables ainsi que, par voie de conséquence les flux journaliers de personnes arrivant à la fin d'une infection de durée variant peu d'un malade à l'autre. Pendant ces phases, **la proportion d'infectés passant chaque jour de I à R ( « recovery rate ») n'est donc pas constante dans le temps, car la date de guérison ou de décès de chacun d'entre eux dépend de la date à laquelle il a été contaminé.**

En phase de croissance, l'approche par un « recovery rate » constant fait artificiellement « guérir ou mourir » le jour J une proportion constante du total des individus alors malades, quelle que soit la date à laquelle ils ont été infectés. Dans la réalité, le nombre réel total de guérisons et de décès du jour J sera nettement plus petit, puisque seront concernés ceux qui sont tombés malades autour du jour J - D (D étant la durée moyenne en jours de la maladie), moins nombreux que ceux qui sont

tombés malades les jours suivants. Le nombre de nouveaux infectés étant proportionnel au nombre d'individus susceptibles de les contaminer, l'application des équations de 1927 faisant guérir ou mourir artificiellement trop de malades va retarder et limiter la hausse prévisible du nombre d'infectés par rapport à la réalité.

Inversement, en phase de décroissance de l'épidémie (par exemple après quelques jours de confinement) l'approche par un « recovery rate » constant fait artificiellement « guérir ou mourir » quotidiennement trop peu de malades, ce qui va retarder et limiter la baisse prévisible du nombre d'infectés par rapport à la réalité.

**On peut donc dire que l'application des équations de 1927 pendant des phases dynamiques de l'épidémie ralentit et minimise les variations du nombre de nouvelles infections quotidiennes et donc le nombre total I de personnes entre infection et guérison (ou décès).** Or c'est précisément pendant de telles phases que les autorités sanitaires ont besoin de connaître les prévisions à court terme correspondant à diverses hypothèses de décisions possibles pour la gestion de l'épidémie (début ou arrêt d'un confinement, sévérité de celui-ci, quarantaine, intensité d'une campagne de vaccination, etc.).

Malgré ce qui précède, **les équations correspondant au « special case » à « recovery rate » constant sont passées à la postérité sous le nom d'équations du modèle SIR et la restriction fondamentale selon laquelle elles ne s'appliquent qu'à un cas à « recovery rate » constant est tombée dans l'oubli.** Pratiquement tous les cours de modélisation français (et même mondiaux) commencent par la présentation du « modèle SIR », et aucun de ceux dont nous avons pu prendre connaissance ne mentionne cette restriction, quelle que soit la nationalité du professeur.

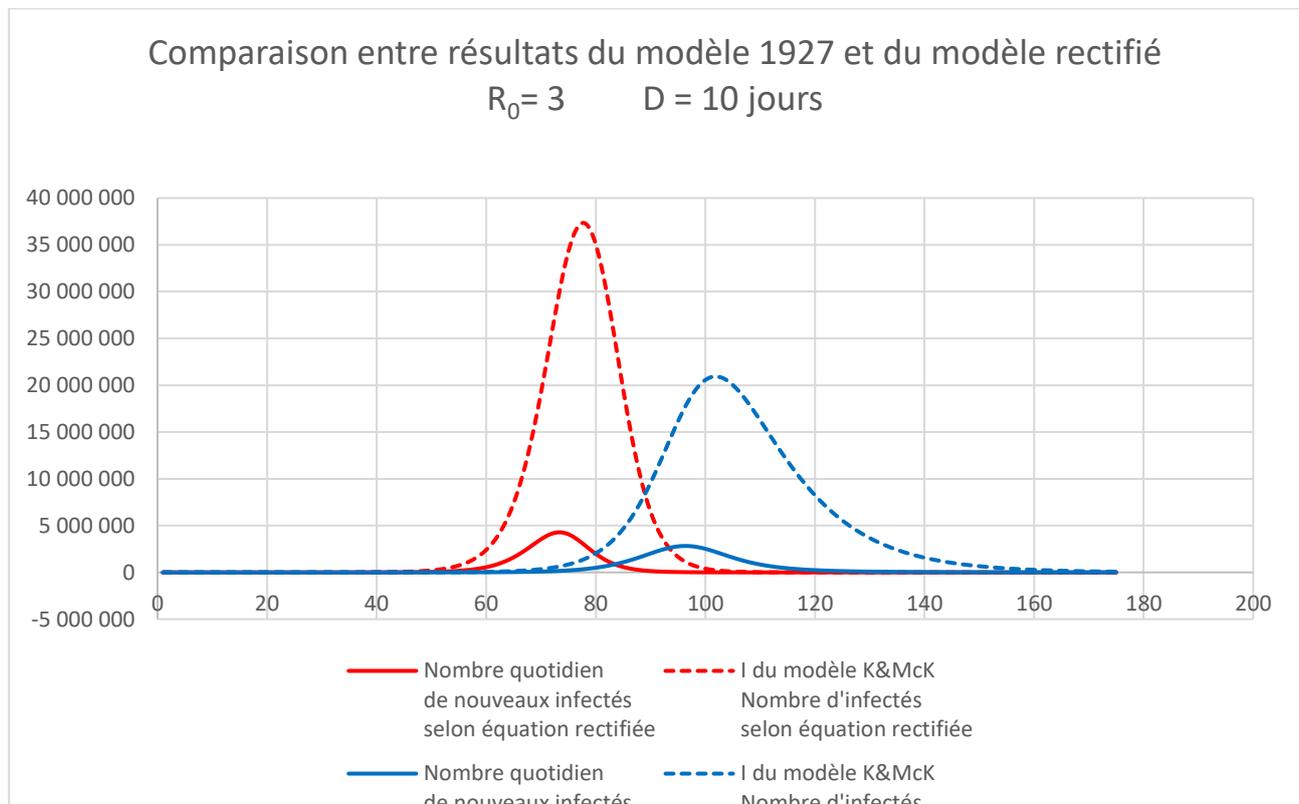
Par ailleurs, en dépit de cette simplification, Kermack et McKendrick ne sont pas arrivés exprimer de façon explicite l'évolution de S, I et R en fonction du temps (sauf en tout début d'épidémie, lorsque S n'a pas encore baissé de façon significative). Depuis cette date, de nombreux mathématiciens ont alors voulu chercher des solutions, mais ils l'ont généralement fait dans l'optique de la résolution d'un pur problème mathématique qu'ils supposaient bien posé par les biologistes (ce qui n'était pas le cas, puisque ces équations ne sont valides qu'en régime statique ou quasi-statique).

Ne trouvant pas de solution permettant d'exprimer de façon explicite l'évolution de S, I et R en fonction du temps, ces mathématiciens ont pu cependant déduire des « équations de Kermack et McKendrick » un certain nombre de conclusions sur le développement théorique d'une épidémie, au prix de raisonnements accessibles uniquement à de bons spécialistes du calcul infinitésimal : voir par exemple « [The SIR model and the Foundations of Public Health](#) » d'Howard Weiss (Département de mathématiques du Georgia Institute of Technology à Atlanta).

Dans la seconde moitié du XX<sup>ème</sup> siècle, l'arrivée des ordinateurs a changé la donne. En effet, partant de ces équations de 1927, des programmes informatiques (« solveurs ») permettent maintenant de tracer les courbes représentant l'évolution de S, I et R en fonction du temps. Les modélisateurs se sont alors aperçus sur des cas concrets que les évolutions constatées lors d'épidémies réelles ne correspondaient pas complètement à celles fournies par les solveurs (en particulier, le développement initial prévu par les équations de 1927 de l'épidémie dans un groupe fermé à partir de l'introduction de quelques cas était nettement plus lent que la réalité observée).

Les **courbes bleues** du diagramme qui suit (extrait de l'Annexe 2) donnent par exemple la prévision résultant de l'application de ces équations de 1927 (faites, rappelons-le, pour simuler une situation statique) à la situation dynamique qu'était l'arrivée du Covid-19 début 2020 en France, alors

que la **courbe rouge basée sur une rectification très simple de ces équations pour les rendre valides en régime dynamique** est beaucoup plus proche de la réalité observée.



*Bien noter que ces courbes sont valides à  $R_0$  constant, c'est-à-dire en l'absence toute modification du comportement de la population française (67 millions d'habitants) résultant de la prise de conscience spontanée des dangers de la contagion ou du suivi de directives des autorités sanitaires (confinements, quarantaines, ...).*

Curieusement, les modélisateurs n'ont pas alors donné la priorité à la très facile modification aux équations de 1927 qui aurait permis de rendre ces équations valides, même pendant une phase transitoire. Comme l'indique Cédric Villani dans sa note d'avril 2020 à l'Office parlementaire d'évaluation des choix scientifiques et technologiques, ils ont alors préféré « complexifier pour affiner » ce qui a abouti à une floraison de modèles (dits par exemple « stochastiques » ou « agents ») d'une grande complexité mathématique, impossibles à valider par d'autres que leurs auteurs, d'autant plus que la documentation expliquant leur fonctionnement interne est généralement insuffisante.

En témoigne par exemple l'attitude de l'éminent modélisateur britannique Neil Ferguson en mars 2020 : « Ferguson a admis que son modèle de la maladie Covid-19 de l'Imperial College était basé sur un code informatique vieux de 13 ans, non documenté. Il a refusé de divulguer son code original pour que d'autres scientifiques puissent vérifier ses résultats », déclaration rapportée dans [https://www.liberation.fr/checknews/2020/06/03/les-previsions-de-ferguson-qui-ont-conduit-de-nombreux-pays-a-se-confiner-etaient-elles-fantaisistes\\_1790061/](https://www.liberation.fr/checknews/2020/06/03/les-previsions-de-ferguson-qui-ont-conduit-de-nombreux-pays-a-se-confiner-etaient-elles-fantaisistes_1790061/)

Avant de poursuivre il est nécessaire de préciser un point fondamental sur les approches possibles de la modélisation mathématique des épidémies.

## 2 – Rappel de quelques principes généraux sur la modélisation mathématique

De manière générale, la modélisation de l'évolution dans le temps d'un phénomène physique, biologique, économique ... peut faire appel à deux techniques voisines mais différentes : le calcul infinitésimal et le calcul « discret » ou « à éléments finis »

### 2.1 – Modèles « discrets » ou « à éléments finis »

Cette modélisation est basée sur un calcul pas à pas des variations de grandeurs caractéristiques du phénomène pendant un court espace de temps, compte tenu de lois (ou d'hypothèses de lois) physiques, biologiques, économiques ...

En renouvelant cette opération à partir des valeurs obtenues pour ces grandeurs à la fin d'un premier intervalle élémentaire de temps, on obtient leurs valeurs à la fin d'un second intervalle de temps. La répétition de ce processus permet de calculer de proche en proche l'évolution des grandeurs caractéristiques du phénomène étudié pendant n'importe quelle durée, avec une précision d'autant plus grande que l'intervalle de temps élémentaire utilisé est petit.

Le calcul mentionné devient rapidement fastidieux. De plus, on veut souvent répéter la totalité de ce processus sur une durée importants en testant diverses hypothèses afin d'optimiser la coïncidence du résultat obtenu par le calcul et l'observation du phénomène réel. On arrive alors à des volumes totaux de calculs qui, jusqu'à l'arrivée des ordinateurs, rendaient pratiquement impossible l'emploi de cette méthode pour des phénomènes tels que des épidémies.

### 2.2 - Modèles faisant appel au calcul infinitésimal

Devant une telle impasse, jusqu'au milieu du XX<sup>ème</sup> siècle, les modélisateurs ont tenté d'avoir recours, chaque fois qu'ils le pouvaient, au calcul infinitésimal. Celui-ci consiste à faire tendre vers zéro la durée de l'intervalle de temps mentionné plus haut : la limite du rapport entre la variation d'une grandeur et la durée de cette variation correspond à la notion mathématique de dérivée par rapport au temps de cette grandeur.

On peut alors exprimer des relations entre les grandeurs concernées et leurs dérivées sous forme d'« équations différentielles ». Si on tombe sur un type d'équations connu (ou si on peut s'y ramener par diverses méthodes mathématiques, mais le succès n'est pas garanti ...) il est alors possible d'exprimer l'évolution des grandeurs concernées sur des périodes longues sous forme de fonctions explicites du temps. Un exemple : si la dérivée d'une fonction est toujours proportionnelle à cette fonction, les mathématiciens nous enseignent que cette fonction est la somme d'une exponentielle (dont l'exposant dépend de la proportionnalité entre fonction et dérivée) et d'une constante.

### 2.3 - Avantages et inconvénients comparés des deux méthodes

Si on a la chance de tomber sur une équation différentielle (ou un système d'équations) pour lesquelles on peut trouver une solution exprimable sous forme de fonction explicite du temps, cette méthode est généralement plus précise qu'une méthode discrète. C'est par exemple le cas pour l'étude de l'évolution d'un corps radioactif : si pendant l'unité de temps  $x\%$  des atomes d'un échantillon passent spontanément de l'état radioactif à l'état stable, le nombre total d'atomes radioactif décroît de façon exponentielle et on peut le calculer de façon très précise (à la seconde près si le corps est extrêmement radioactif). Si on utilise une méthode discrète rustique à pas temporel d'un jour, on calcule pendant

chaque journée le nombre de passages de l'état radioactif à l'état stable à partir du nombre d'atomes radioactifs en début de journée. Or en toute rigueur il faudrait faire ce calcul à partir d'un nombre d'atomes radioactif diminuant de façon régulière pendant toute cette journée (ce que fait le calcul infinitésimal).

Cependant, comme l'a signalé le Professeur Odo Diekmann de l'Université d'Utrecht, cet avantage, (réel dans le cas de la modélisation de la radioactivité) n'existe pas dans le cas des épidémies humaines, en raison d'une contagion (donc d'un nombre de nouveaux malades) variant au cours de chaque journée suivant un rythme imposé par l'alternance jour-nuit et l'organisation des activités humaines en fonction de l'heure. Dans le cas de la radioactivité, il peut être intéressant de descendre à un calcul heure par heure (ou même seconde par seconde si le corps est très radioactif). Pour la raison indiquée, ce n'est pas le cas dans l'étude d'une épidémie où un pas temporel de 24 heures apporte une précision qu'aucune autre méthode (pas temporel plus petit ou calcul infinitésimal) ne permet d'améliorer.

La méthode discrète a pour elle un énorme avantage : elle ne demande aucune connaissance mathématique particulière (aucun besoin de savoir ce que sont une dérivée, une équation différentielle, une exponentielle ; connaître les 4 opérations apprises à l'école primaire et savoir utiliser un tableur suffisent).

De plus, comme le souligne Cédric Villani en haut de la page 9 de son rapport, elle apporte une incomparable souplesse pour l'introduction d'hypothèses variées. On peut citer la variation journalière de la contagiosité d'un individu au cours de son infection, le rythme de vaccination de la population et son efficacité, la durée d'acquisition et de conservation de l'immunité apportée par la vaccination ou par l'infection suivie d'une guérison, les variations brutales dans le nombre quotidien de nouvelles infections résultant de directives gouvernementales (confinements, couvre-feux, quarantaines ...), celles, plus lentes, entraînées par des modifications spontanées dans les comportements d'une population motivée par la peur plus ou moins forte de contagion, etc.

### **3 - Situation mondiale début 2020 de la modélisation mathématique des épidémies et de son enseignement**

Face au risque potentiel d'arrivée du Covid-19 en France, les autorités sanitaires françaises pouvaient s'appuyer début 2020 sur deux types d'outils de modélisation :

- De multiples publications émanant de départements de mathématiques universitaires dont les auteurs, purs mathématiciens, sont généralement partis des équations de Kermack et McKendrick de 1927 fournies par leurs collègues biologistes sans les remettre en cause, alors que, comme nous l'avons vu précédemment, ces équations ne sont valides que pour décrire un phénomène statique et conduisent à de graves erreurs dans une réalité foncièrement dynamique. Quant à ceux qui se sont écartés des équations de 1927 en « complexifiant pour affiner » (C. Villani), ils ont généralement continué à utiliser la notion de taux de guérison constants qui, nous l'avons vu, est fautive dans le cas d'une population de malades qui a été contaminée quelques jours plus tôt à un rythme quotidien croissant ou décroissant,
- Un nombre nettement plus restreint de travaux (en particulier ceux provenant de grandes institutions de recherche médicale ou biologique ayant leurs équipes de modélisation) dont les auteurs déclarent (uniquement si on les interroge !) être conscients de la difficulté mentionnée plus haut et avoir utilisé divers procédés de calcul pour en corriger les défauts.

Pour des raisons peu explicables ils n'ont pas voulu remettre en cause officiellement les équations de 1927, et il est très difficile de comprendre si les procédés en question, utilisés quasi clandestinement (et qualifiés parfois d' « astuces » !), sont efficaces, car chaque équipe a sa méthode généralement très mal documentée, rendant impossible toute validation par d'autres que leurs auteurs.

De son côté l'enseignement de la modélisation commence systématiquement par l'étude des équations de 1927, l'utilisation d'un taux de guérison (ou de décès) constant au cours du temps et fait appel aux subtilités du calcul infinitésimal. Il est généralement reconnu que la stricte application du modèle SIR de 1927 conduit à un développement initial de l'épidémie plus lent que celui constaté dans la réalité, mais ce défaut est attribué (sans aucune justification !) au fait que ce modèle ne serait pas assez « stochastique » et ne pourrait pas rendre compte de la propagation de l'épidémie tant que le nombre de malades est faible (alors que la légère modification validant les équations de 1927 pour les régimes dynamiques d'infection suffirait pour obtenir une évolution plus conforme à la réalité).

Une partie importante des cours est consacrée à des variations autour du « modèle SIR » utilisant des équations inspirées par celles de 1927 (donc uniquement valides pendant des phases statiques, sans que cette condition soit jamais rappelée). L'avenir de la modélisation est évoqué sous forme d'une fuite en avant mathématique (modèles « stochastiques », modèles « agents », ...) sans qu'une forme précise de calcul un tant soit peu normalisée soit enseignée. Le résultat est que, persuadés de la nécessité d'avoir un très haut niveau mathématique, la majorité des médecins ou de biologistes fuient cette discipline, persuadés qu'elle ne peut être pratiquée que par des équipes de mathématiciens hyper qualifiés.

#### **4 – Constatations troublantes faites lors de l'épidémie de Covi-19**

Nous avons vu que l'utilisation des équations de 1927 et de la notion de taux de guérison (ou de décès) fixe pendant des phases d'évolution dynamique d'une épidémie donnait un résultat minimisant les variations réelles du nombre d'infections, qu'il s'agisse d'une hausse ou d'une baisse (trop lentes ou trop faibles par « effet tampon »). Les modélisateurs professionnels des grands organismes de recherche affirment que leurs modèles corrigent ce défaut. Or les faits suivants sont troublants :

- Le 1<sup>er</sup> confinement (à partir de mars 2020) a fait descendre le nombre d'infections pendant l'été 2020 à un niveau plus bas que celui prévu par les organismes officiels, à tel point des voix se sont alors élevées pour affirmer que l'épidémie était terminée
- Les organismes officiels avaient bien prévu une « deuxième vague » à l'automne 2020, mais à croissance beaucoup plus lente que la première. En fait, la brutalité de cette vague d'automne a été telle que les pouvoirs publics, qui voulaient éviter une nouvelle décision de confinement pour des raisons politiques, ont été contraints d'en instaurer un fin octobre
- Constatant pendant ce mois d'octobre 2020 la montée de cette deuxième vague, les organismes officiels ont constamment pressé les pouvoirs publics de mettre en place rapidement ce deuxième confinement qui, s'il était trop tardif, risquait de ne pas arriver à temps pour éviter un engorgement du système hospitalier. Pour les raisons politiques rappelées, les pouvoirs publics ont tardé à prendre cette décision, mais fort heureusement la baisse d'infections qui a suivi a ramené plus rapidement que prévu le nombre de malades à un niveau compatible avec la capacité du système hospitalier

Ces faits semblent indiquer qu'en fait les modèles utilisés par les grands organismes de recherche ne corrigent pas complètement les conséquences de l'emploi de la notion de taux de guérison constant pendant des phases transitoires à fortes variations des rythmes quotidiens de nouvelles infections. Le Conseil scientifique Covid-19 lui-même a constaté à plusieurs occasions l'insuffisante réactivité (à la hausse comme à la baisse) des modèles utilisés pour conseiller les pouvoirs publics, par exemple :

- « Cette accélération dont la brutalité est surprenante » (page 3 de sa « Note » du 26/10/2020)
- « Le taux d'incidence a connu une baisse particulièrement rapide depuis mi-mai 2021. Elle a surpris le monde scientifique par son ampleur et sa vitesse » (page 2 de son « Avis » du 6/7/2021)

En juin 2021 est parue « Epidemiology and control of SARS-CoV-2 epidemics in partially vaccinated populations : a modeling study applied to France » co-signée par Paolo Bosetti, Cécile Tran Kiem, Alessio Andronico, Vittoria Colizza, Yazdan Yazdanpanah, Arnaud Fontanet, Daniel Benamouzig et Simon Cauchemez, soit presque toute la fine fleur de la modélisation épidémiologique française. Point inhabituel : dans ses premières versions, la communication contient en page 2 du « Supplément » l'intégralité des équations différentielles utilisées (qui ont disparu des versions ultérieures !). Et là, surprise qui confirme nos soupçons : les fins de maladies sont régies par des équations différentielles de type «1927», dont nous avons vu qu'elles étaient inadaptées aux phases dynamiques d'une épidémie (d'autant plus qu'à cette époque il s'agissait d'un virus Covid-19 particulièrement contagieux, son variant Omicron).

## 5 – Tentatives d'utilisation de tableurs

Un certain nombre de scientifiques jusque là totalement étrangers à la médecine et à la biologie ont découvert à l'occasion de la crise du Covid-19 l'existence des modèles mathématiques du développement d'épidémies. Il a alors semblé à certains d'entre eux que l'outil idéal pour étudier un tel processus était le tableur, qui présente en particulier tous les avantages de souplesse d'utilisation que nous avons évoqués plus haut.

Toutefois, de même qu'un mathématicien chargé de calculer des champs magnétiques ou électriques ne remettra pas en cause les équations de Maxwell fournies par ses collègues physiciens, ces néo-modélisateurs d'épidémies ont généralement pensé qu'il était légitime d'utiliser telles quelles « les équations différentielles de Kermack et McKendrick » enseignées par les spécialistes chevronnés de la discipline qui omettent systématiquement de signaler qu'elles ne sont valides qu'à « recovery rate » constant, la raison étant souvent, selon le Professeur Odo Diekmann déjà cité, parce qu'ils l'ignorent eux-mêmes.

Citons deux exemples de cette utilisation d'un tableur partant d'une discrétisation des équations SIR de 1927 non modifiées, chaque ligne du tableur correspondant à un pas de calcul temporel d'une journée de 24 heures :

- Les exercices proposés en mars 2020 aux étudiants en licence de mathématiques de l'Université d'Aix Marseille (voir l'exemple donné en Annexe 3)
- Jusqu'à sa lettre n°23, le site <https://corona-circule.github.io/lettres/> sur l'épidémie de Covid-19 de François Varret, ancien élève de l'ENS Ulm, Professeur émérite de physique (université de Versailles Saint-Quentin-en-Yvelines) et de Mathilde Varret, biologiste, généticienne (INSERM).

François Varret a pris connaissance en 2020 de nos articles dans la revue des Anciens de l'École polytechnique (voir Bibliographie à la fin de cette note) et a constaté que la correction des équations de 1927 que nous proposons (remplacement du « recovery rate » constant, inadapté aux situations dynamiques, par une durée moyenne de contagiosité constante) permettait d'obtenir des résultats nettement plus conforme à la réalité observée.

Il a alors adopté cette correction sous le nom de  $SIR_{tcc}$  à partir de la lettre n° 24 de janvier 2021 <https://corona-circule.github.io/lettres/let24> et nous avons ensuite publié ensemble « A tentative tracking of the SARS-Cov2 pandemic in France, based on a corrected SIR model including vaccination effects » : communication de Mathilde Varret, François Varret et François Xavier Martin faite en avril 2022 aux 8<sup>th</sup> Complexity-Disorder Days de l'European Physical Journal (EPJ) [https://epjwoc.epj.org/articles/epjconf/abs/2022/07/epjconf\\_complexity-disorder2021\\_01002/epjconf\\_complexity-disorder2021\\_01002.html](https://epjwoc.epj.org/articles/epjconf/abs/2022/07/epjconf_complexity-disorder2021_01002/epjconf_complexity-disorder2021_01002.html)

## **6 – L'attitude de Cédric Villani dans sa note dans sa note à l'Office parlementaire d'évaluation des choix scientifiques et technologiques d'avril 2020**

Pendant l'épidémie, bien qu'il ait obtenu la Médaille Fields en 2010, on ne peut plus considérer Cédric Villani comme un pur scientifique, puisqu'il était député depuis 2017 et que début 2020, il était en pleine campagne pour les élections à la Mairie de Paris.

Il est impensable qu'un mathématicien de son niveau n'ait pas compris immédiatement le problème posé par l'utilisation d'un « recovery rate » (par guérison ou décès) constant ne dépendant pas de l'historique de la contamination de la population malade. Mais, chargé de faire le point auprès de l'Office parlementaire d'évaluation des choix scientifiques et technologiques (dont il était alors premier vice-président), il lui était vraisemblablement difficile de semer en période d'anxiété de la population de graves doutes sur la quasi-totalité des outils sur lesquels les pouvoirs publics pouvaient s'appuyer afin de définir une politique sanitaire destinée à éviter la submersion des capacités du système hospitalier français.

Cédric Villani a donc écrit un rapport très pédagogique présentant les équations de 1927 et expliquant ensuite comment les modélisateurs avaient été amenés à « complexifier pour affiner » leurs modèles. Mais – in cauda venenum ! - il a ajouté un paragraphe exposant certaines possibilités supplémentaires extrêmement intéressantes que pourraient offrir facilement des modèles « discrets », ce qui suggère implicitement qu'on pourrait fort bien aller directement au « tout discret » et s'épargner le passage par le calcul infinitésimal et ses équations différentielles !

## **7 – Propositions**

Compte tenu de ce qui précède, il semble souhaitable de mener simultanément :

### **7.1 – Une réforme de l'enseignement de la modélisation des épidémies**

Dès qu'ils dépassent l'application pure et simple des équations de 1927, les cours de modélisation sont d'une extrême complexité mathématique et ne sont suivis que par un effectif très réduit d'étudiants. La plupart des médecins et des biologistes sont persuadés que le domaine requiert un tel niveau mathématique que leur connaissance de la modélisation ne peut aller au-delà de ce qu'ils peuvent retirer d'une brève conférence de vulgarisation.

Il paraîtrait donc intéressant de bâtir, à partir d'un examen aussi exhaustif que possible des formations existantes, un cours montrant que l'on peut arriver à des résultats au moins équivalents avec une méthode de modélisation « discrète » demandant pour tout préalable d'être capable de :

- 1 ) comprendre comment on peut passer:
  - du nombre moyen de contaminations  $R_0$  provoquées par un malade parmi une population (N) où tous peuvent être infectés
  - au nombre total d'infections dans une population où  $[I(t)]$  sont contagieux et où seuls  $[S(t)]$  peuvent être contaminés (ce qui est du niveau des mathématiques enseignées à l'école primaire)
- 2) et savoir utiliser un tableur.

## **8.2 – La préparation d'une politique de modélisation permettant la détermination et la justification permanente auprès de l'opinion publique de la conduite à tenir en cas de nouvelle crise sanitaire**

La concurrence entre modèles au fonctionnement généralement mal documenté et peu compréhensible, l'impossibilité pour d'autres que leurs auteurs d'en reproduire les résultats mettent les autorités politiques dans une situation difficile, d'autant plus qu'elle permet en permanence une contestation des décisions prises, en particulier par des intervenants dans les médias qui déclarent de façon péremptoire : « Le gouvernement aurait dû ... car notre modèle (ou « le modèle de l'équipe X ») montre que ... » sans préciser clairement les hypothèses retenues.

La formalisation dès maintenant de modèles « discrets » ouverts, précisant de façon claire les hypothèses retenues, afin que chacun puisse en reproduire les résultats s'il sait manipuler un tableur (dont l'emploi est connu de nombreux professionnels dans de multiples domaines d'activité) permettrait aux pouvoirs publics, en cas de nouvelle pandémie, de prendre des décisions en meilleure connaissance de cause et de les justifier plus facilement auprès de l'opinion publique.

## Annexe 1

### Réactions opposées obtenues :

#### d'un des principaux modélisateurs français contacté sur le conseil de Cédric Villani ...

*« Vous avez tout à fait raison, dans votre analyse. Les modèles d'équations différentielles ordinaires (EDO) font une hypothèse markovienne, autrement dit que "l'histoire" n'a pas d'effet. Autrement dit, une personne infectée depuis 1 jour a la même probabilité de guérir qu'une personne infectée depuis 14 jours. Ceci est évidemment faux pour les durées d'infection ... Cette approximation a peu de conséquences si on modélise le phénomène sur un temps long (plusieurs années). Elle est, en revanche, très problématique si on se place sur un temps court et si on veut suivre l'épidémie au mieux. »*

#### ... de membres du CNRS

*« Je ne suis pas plus spécialiste du domaine qu'apparemment l'auteur de l'article, mais la naïveté de sa critique des modélisateurs m'étonne. Rejeter les modèles plus sophistiqués sous le prétexte qu'il ne s'agirait que de dizaines de pages d'équations incompréhensibles, et ne garder que le plus simple dans sa forme la plus primitive ... pour ensuite critiquer le manque de réalisme des modélisateurs : tout cela ne me semble pas très sérieux. »*

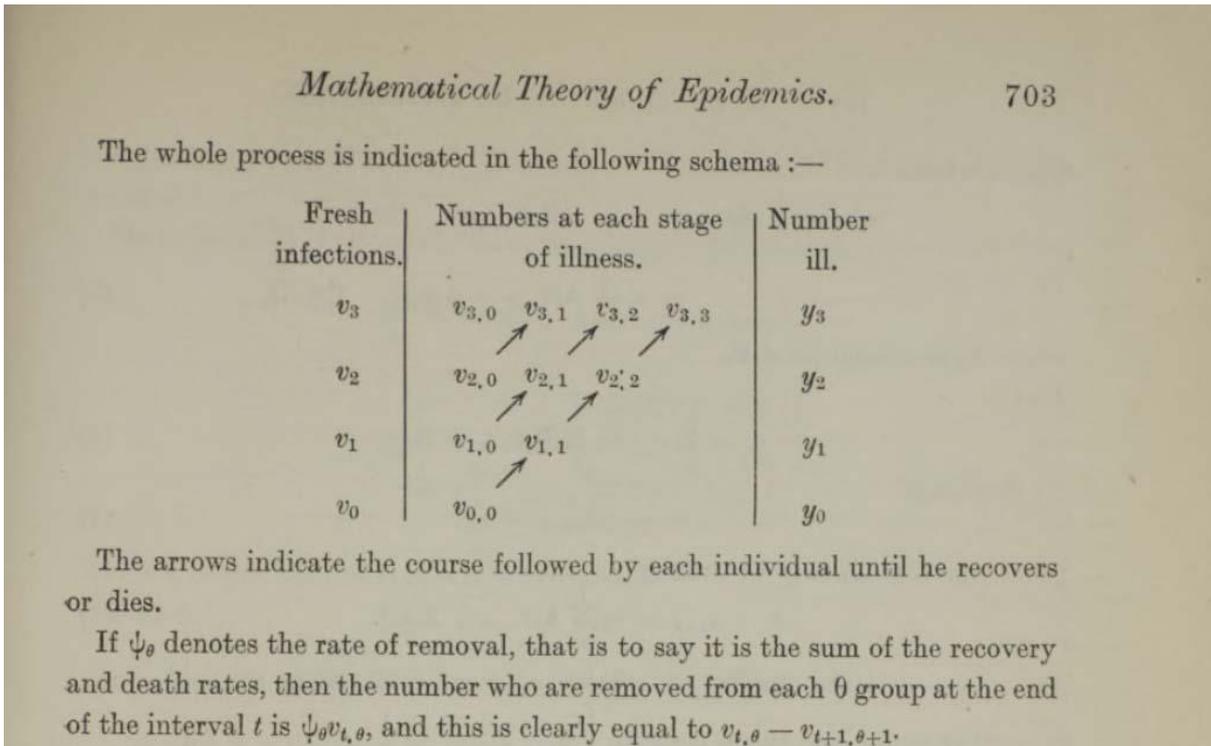
*« Malgré ce que vous semblez penser les modélisateurs professionnels connaissent évidemment ce que vous appelez l'erreur de 1927, et savent quand elle est gênante ou pas, selon ce que l'on veut faire. S'il le faut ils l'éliminent. On peut le faire de différentes façons ... Je ne suis pas l'inventeur de cette astuce numérique, elle fait partie de celles couramment utilisée par les modélisateurs. »*

*« L'épidémiologie mathématique est une discipline scientifique internationale très active, qui a fait de très grands avancées depuis 1927 (notamment pour répondre à des crises de moins grande ampleur et beaucoup moins médiatisées). Les modèles ne "reposent" pas sur une équation erronée mais se nourrissent peu à peu de la meilleure compréhension que l'on peut avoir des épidémies grâce aux précédentes et les hypothèses que ces modèles font ne sont pas un bête copié-collé du travail d'un précurseur datant d'il y a un siècle. Il est très bien d'avoir un esprit critique et il est tout à votre honneur de réfléchir à ces questions, mais ces modèles que vous critiquez un peu vertement ont aussi fait leurs preuves dans d'autres situations. »*

## Annexe 2

### Analyse de la « contribution à la théorie mathématique des épidémies » de Kermack et McKendrick (1927)

Au début de cette « contribution » les auteurs expliquent qu'à un instant donné l'ensemble des individus malades comprend plusieurs sous-ensembles, chacun d'entre eux incluant ceux ont été infectés pendant le même intervalle de temps (et ont donc la même « ancienneté » dans la maladie). Figure alors le diagramme suivant :



qui laisse supposer que le moment où chaque malade a été infecté sera pris en compte dans le modèle pour déterminer le moment où il sortira de son état infectieux par guérison ou décès.

Malheureusement la dizaine de pages de calcul qui suivent, pratiquement incompréhensibles, ne permet pas de voir si les auteurs sont arrivés à une méthode de calcul (à l'époque sans l'aide d'ordinateurs !) permettant de modéliser une épidémie suivant ce principe.

Vraisemblablement dans la recherche du Graal qu'aurait été à cette époque la découverte d'une solution exprimant  $S(t)$ ,  $I(t)$  et  $R(t)$  sous forme de fonctions du temps, Kermack et McKendrick proposent alors ce qu'ils appellent un « special case » à « constant rates » où :

- $\phi(t)$  : « rate of infectivity »
- et
- $\psi(t)$  : « rate of removal » (par guérison ou décès)

ne dépendent pas du temps  $t$ .

from the consideration of the special case in which  $\phi$  and  $\psi$  are constants  $\kappa$  and  $l$  respectively.

In this case the equations are

$$\left. \begin{aligned} \frac{dx}{dt} &= -\kappa xy \\ \frac{dy}{dt} &= \kappa xy - ly \\ \frac{dz}{dt} &= ly \end{aligned} \right\} \quad (29)$$

and as before  $x + y + z = N$ .

Ces équations sont à la base de la plupart des multiples systèmes de simulation d'épidémies proposés depuis cette époque, avec des notations qui sont habituellement :

$$S'(t) = dS / dt = -\beta S(t) I(t)$$

$$I'(t) = dI / dt = \beta S(t) I(t) - \gamma I(t)$$

$$R'(t) = dR / dt = \gamma I(t)$$

$$S(t) + I(t) + R(t) = N$$

Si  $D$  est la durée moyenne d'infection pendant laquelle un malade est réputé contagieux et  $R_0$  le nombre d'individus qu'infecte pendant toute la durée de sa contagiosité un malade plongé dans une population entièrement susceptible d'être infectée :

- ce malade contamine dans ce type de population entièrement contaminable  $R_0/D$  individus par unité de temps (pour le Covid-19 on utilise habituellement la journée de 24 heures)
- si dans une population de  $N$  individus seuls  $S(t)$  sont susceptibles d'être contaminées, la probabilité qu'un malade rencontre une personne contaminable est :  $S(t) / N$ . Le malade ne contaminera donc que  $(R_0/D) \times S(t) / N$  personnes par unité de temps
- $I(t)$  malades contamineront donc  $I(t) \times (R_0/D) \times S(t) / N$  personnes par unité de temps

On a donc  $\beta = R_0 / (D \times N)$

D'autre part si  $I(t)$  personnes sont malades pendant une moyenne de  $D$  unités de temps, on suppose qu'en  $D$  unités de temps toutes seront ou guéries ou décédées, donc que  $I(t) / D$  guériront ou décéderont pendant chaque unité de temps (et donc ne contamineront plus personne).

On pose habituellement  $\gamma = 1 / D$

**Cette approximation qui date de 1927, bien adaptée à des phénomènes statiques ou quasi-statiques, ne rend pas compte correctement de la dynamique réelle d'une épidémie, en particulier dans les moments cruciaux que sont l'arrivée de la maladie et le début ou la fin d'un confinement.**

En effet, pendant une période de croissance du nombre d'individus infectés, au jour  $J$  ceux-ci sont répartis en  $D$  sous-ensembles :

Infectés depuis 1 jour
Infectés depuis 2 jours
..... / .....
Infectés depuis (D – 1) jours
Infectés depuis D jours

**Le nombre d’infectés augmente de jour en jour.**

**La taille de chaque case est proportionnelle au nombre d’individus qu’elle contient.**

Dans la réalité, les malades guérissent ou décèdent en moyenne D jours après avoir été infectés. Le jour J, dans une modélisation correcte, quittent donc le « compartiment I des infectés » (par guérison ou décès) les malades, peu nombreux (voir schéma) qui sont infectés depuis D jours.

L’« approximation de 1927 » de Kermack et McKendrick reprise dans tous les cours d’épidémiologie et la plupart des modèles actuels fait sortir quotidiennement du « compartiment I des infectés » la moyenne entre les D cases du diagramme ci-dessus, c’est-à-dire un nombre très supérieur à la réalité puisque les infectés récents sont arrivés à un rythme quotidien de plus en plus élevé

**Majorant le nombre de sorties du « compartiment I » par rapport au nombre réel, ce type de modèle diminue artificiellement la vitesse de croissance prévue pour une épidémie, par exemple à son début ou pendant une phase de déconfinement.**

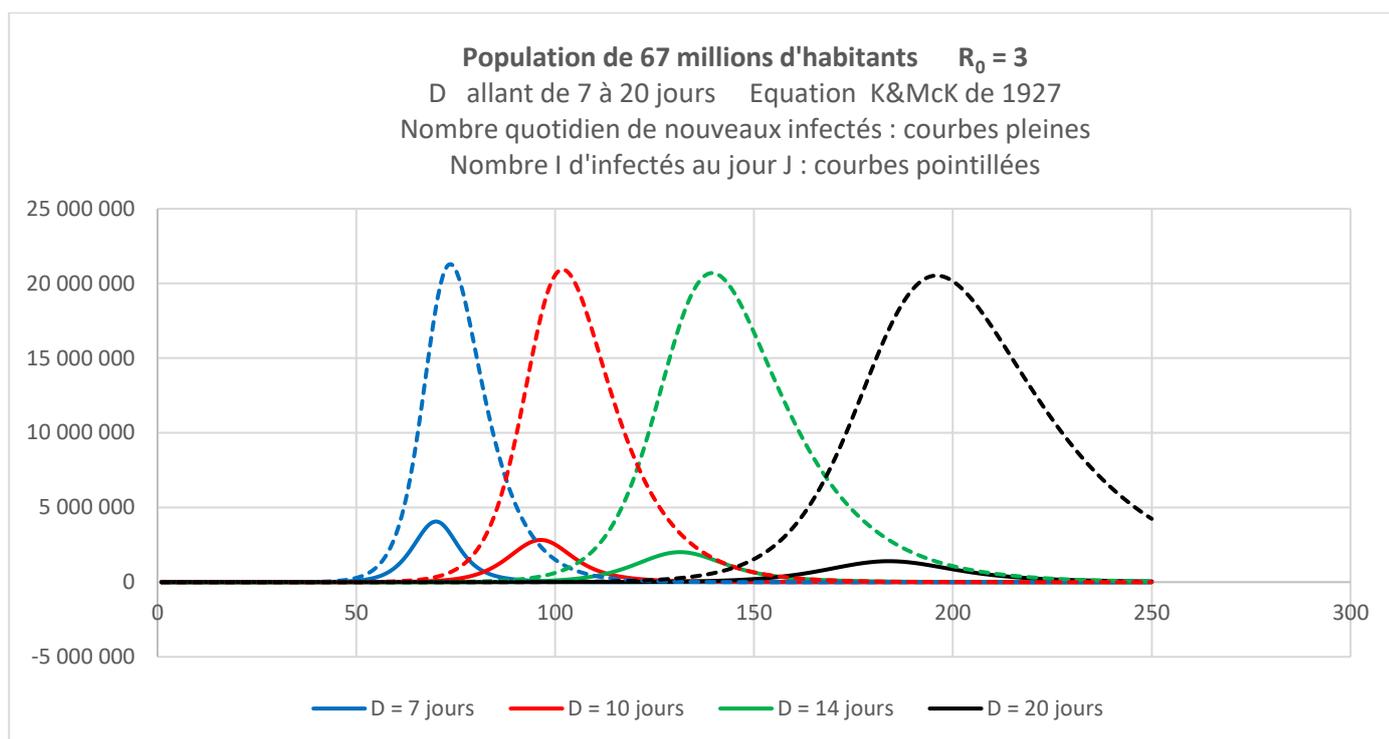
**Dans un modèle digne de ce nom, qui doit simuler une situation en permanence évolutive**, si D est la durée moyenne entre début d’infection et guérison ou décès, chaque jour le nombre de malades qui arrêtent d’être susceptibles de contaminer des personnes saines est sensiblement égal à celui des personnes qui sont tombées malades D jours plus tôt (ce qui peut apparaître comme une lapalissade, mais certains paraissent l’ignorer ... !).

Cette logique est très différente de celle adoptée par Kermack et McKendrick en 1927. Ce qu’on peut appeler « l’approximation de 1927 » aboutit à ce qu’en phase de croissance le modèle SIR classique comptabilise comme sortant de l’infection des individus qui sont tombés malades il y a nettement moins de D jours, ce qui va conduire à la prévision d’une croissance de l’épidémie très inférieure à la réalité par sous-estimation du nombre de personnes restant infectieuses (c’est vraisemblablement ce qui est arrivé en France à l’automne 2020, quand la « 2<sup>ème</sup> vague » Covid-19 a été beaucoup plus rapide et beaucoup plus forte que les prévisions de l’été, pendant lequel le très faible nombre de nouveaux infectés quotidiens avait même conduit certains à annoncer que l’épidémie était terminée). Ce fait est connu de certains modélisateurs, mais ils ne semblent pas en avoir mesuré les conséquences, qu’il est possible d’évaluer par simulation de la même épidémie en utilisant successivement un

programme basé sur « l'approximation de 1927 » et un programme ayant remplacé l'équation correspondante ( $dR / dt = \gamma I$ ) par une équation adaptée à la simulation de phénomènes en régime fortement dynamique ( $dR(t) / dt = - dS(t-D) / dt$ ).

Pour ce faire, il est possible d'utiliser un simple tableur qui permet d'exprimer facilement le nombre de nouveaux infectés, guéris et décédés en fonction du nombre de personnes infectées, guéries et décédées pendant les journées précédentes et de l'évolution de certains paramètres (en particulier  $R_0$ ). L'exemple choisi est un cas théorique souvent présenté dans les cours d'épidémiologie ( $R_0 = 3$  et durées d'infection fixes pouvant aller d'une à trois semaines ; voir en annexe les tableaux permettant de simuler simplement cette épidémie en utilisant les 2 modèles). On l'applique à une population de 67 millions d'habitants.

Utilisons tout d'abord l'équation classique de 1927 :  $dR / dt = \gamma I$



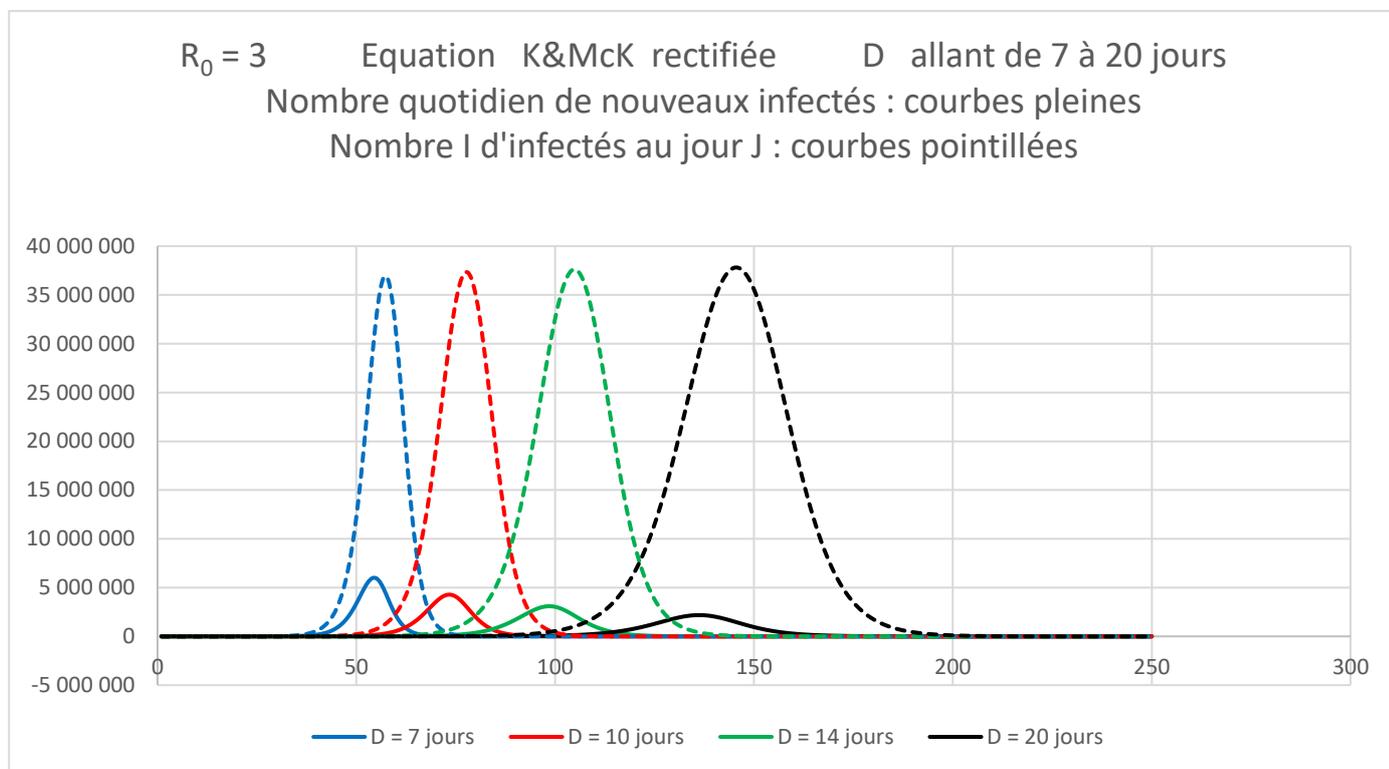
**Abscisses : nombre de jours après l'arrivée d'un premier individu malade au temps 0**  
**Ordonnées : nombre d'individus infectés**

*A noter qu'il s'agit de simulations à  $R_0$  constant uniquement destinées à démontrer la gravissime sous-estimation du modèle de 1927 en phase de croissance d'une pandémie. **Bien évidemment, si la maladie est dangereuse, pouvoirs publics et population feront tout pour que  $R_0$  diminue avant que le nombre réel d'infectés atteigne le maximum théorique de ces courbes.** D'autre part, au cours d'une épidémie réelle,  $R_0$  varie en permanence pour de multiples raisons tenant tout à la fois au comportement de la population et à des données médicales (mutations éventuelles du virus, découverte de traitements changeant la durée de contagiosité des personnes infectées). L'accent quasi exclusif mis dans les cours d'épidémiologie sur des cas théoriques d'épidémies à  $R_0$  constant pouvait se justifier à l'époque où, faute d'ordinateurs, il était important d'essayer de trouver des résultats exprimables sous forme de fonctions analytiques. L'important de nos jours est d'utiliser des équations représentant le mieux possible le phénomène que l'on veut simuler : ensuite, de toute façon, le calcul (par ordinateur) suivra !*

A  $R_0$  constant (égal à 3) le pic **théorique** (voir plus haut) du nombre d'infectés (21 millions sur une population de 67 millions) correspondant à l'atteinte de ce qu'on appelle « l'immunité collective » dépend très peu de la durée d'infection moyenne. En revanche, le nombre quotidien maximal

théorique de nouvelles infections est plus important (4 millions / jour) dans le cas d'une durée d'infectiosité courte (7 jours) que dans le cas d'une infectiosité plus longue (1,5 million / jour pour  $D = 20$  jours).

Utilisons maintenant l'équation  $dR(t) / dt = - dS(t-D) / dt$ ,  $D$  durée de la période d'infection contagieuse étant égale à  $1 / \gamma$



**Abscisses : nombre de jours après la première infection      Ordonnées : nombre d'individus infectés**

**Certaines conclusions sont les mêmes que celles résultant de l'utilisation de l'équation de 1927 :**

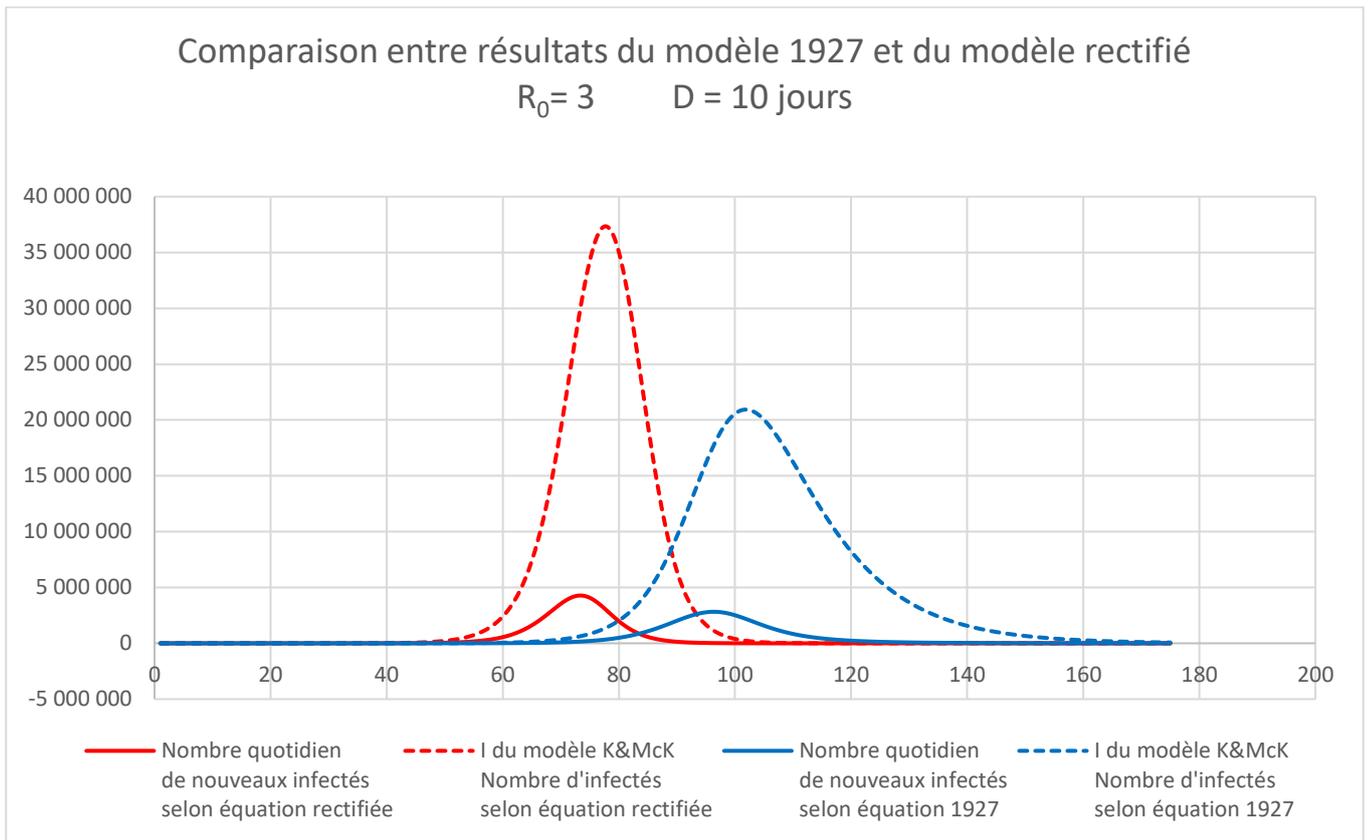
A  $R_0$  constant (égal à 3) le pic théorique du nombre total d'infectés simultanés, s'il est plus important que celui obtenu avec les équations de 1927 (37 millions au lieu de 21 sur une population de 67 millions), dépend très peu de la durée d'infection moyenne. En revanche, le nombre quotidien maximal théorique de nouvelles infections est plus important (6 millions / jour) dans le cas d'une durée d'infectiosité courte (7 jours) que dans le cas d'une infectiosité plus longue (2 millions / jour pour  $D = 20$  jours).

On peut également remarquer que c'est  $R_0$  qui est déterminant pour le calcul du nombre maximum d'infectés correspondant à l'atteinte de l'immunité collective au-delà de laquelle l'épidémie commence à s'éteindre spontanément. A  $R_0$  constant, une faible variation de la durée moyenne d'infection  $D$  ne change pas de façon fondamentale l'évolution de l'épidémie (mais il faut bien sûr remarquer que la valeur de  $D$  est déjà prise en compte dans l'égalité :  $R_0 = \beta D N = \beta N / \gamma$ )

**Mais certaines différences sautent aux yeux :**

- \* le pic théorique du nombre d'infectés maximal théorique donnée par l'équation rectifiée est beaucoup plus élevé que celui donné par l'équation classique de 1927 (37 millions au lieu de 21 millions),
- \* l'équation rectifiée implique un décollage significatif de l'épidémie et une atteinte du nombre maximal de personnes contaminées nettement plus rapides que l'équation de 1927

On peut mettre en évidence ces différences en portant sur le même diagramme les résultats obtenus avec les 2 équations différentes, par exemple pour  $R_0 = 3$  et  $D = 10$  jours :



**Abcisses : nombre de jours après la première infection    Ordonnées : nombre d'individus infectés**

Comme indiqué plus haut, ces courbes ont un caractère théorique, car dans le cas d'une maladie grave,  $R_0$  va généralement diminuer rapidement sous l'influence de changements de comportement de la population spontanés (réduction des contacts avec les malades par peur de la contagion) ou imposés par les pouvoirs publics (confinements, couvre feux, fermeture de magasins et de salles de spectacles ...).

Mais ce qui n'est pas seulement théorique est le fait que l'utilisation de l'équation de 1927, qui traite les guérisons et les décès comme s'il s'agissait d'un phénomène physique tel que la radioactivité, divise pratiquement par deux les variations (à la hausse comme à la baisse) des nombres d'infections calculées par l'équation rectifiée qui, elle, tient compte de la réalité biologique d'une période d'infection variant relativement peu par rapport à une valeur moyenne.

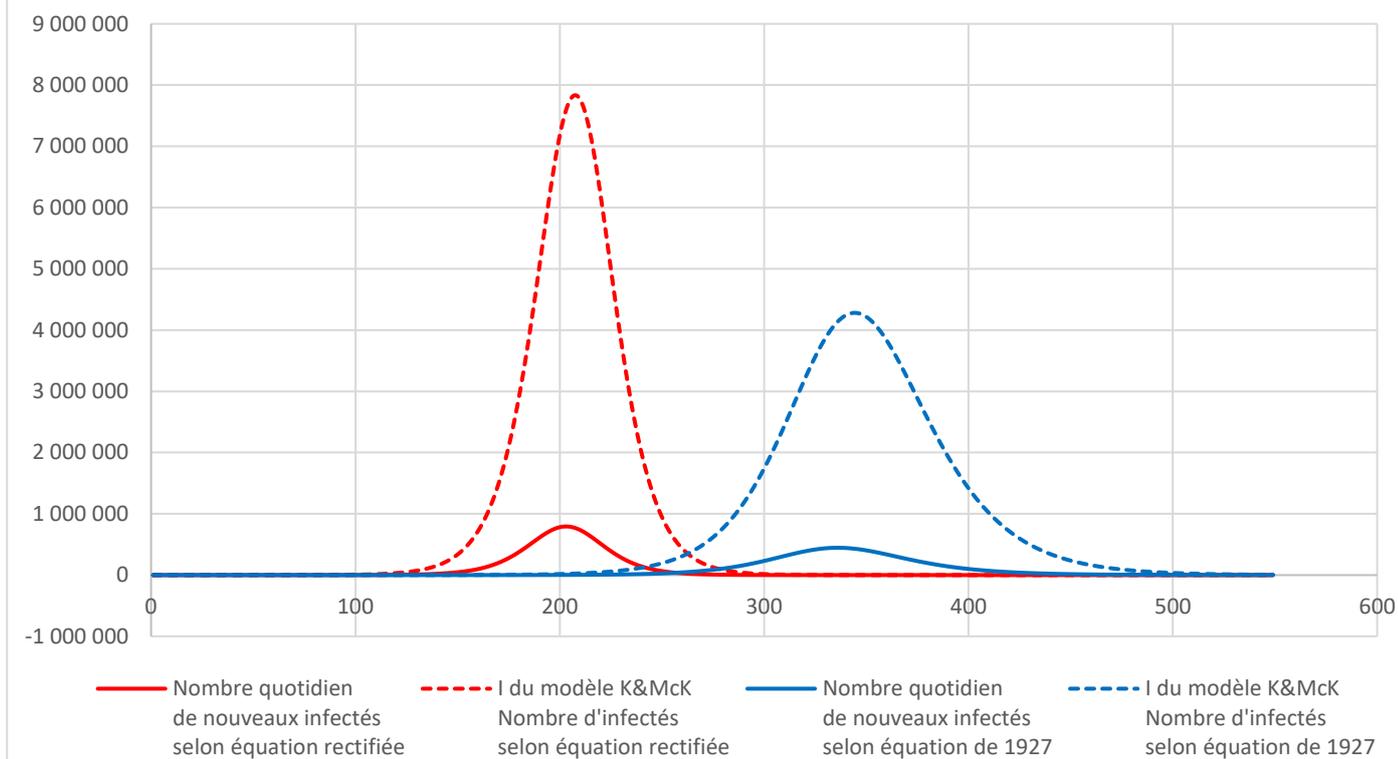
**Sensibilité à  $R_0$**

Il est intéressant de faire la même comparaison entre les résultats obtenus avec d'autres valeurs de  $R_0$ , telles que 1,5 qui est une valeur couramment obtenue pour l'épidémie Covid-19 sans confinement, mais avec des mesures de distanciation physique.

On retrouve les mêmes conclusions que pour  $R_0 = 3$ , avec des valeurs absolues plus petites.

## Comparaison entre résultats du modèle 1927 et du modèle rectifié

$$R_0 = 1,5 \quad D = 10 \text{ jours}$$



**Abscisses : nombre de jours après la première infection      Ordonnées : nombre d'individus infectés**

### Conclusions

**L'équation de 1927, adaptée aux seules situations statiques ou quasi-statiques, génère une solution erronée :**

**- les deux pics (nouveaux infectés quotidiens et total des personnes infectées au jour J) sont égaux à un peu plus de la moitié de ceux donnés par l'équation qui rend compte correctement du phénomène dynamique. De plus, ces pics surviennent avec retard.**

**- en phase ascendante de l'épidémie, le taux de croissance du nombre d'infectés calculé avec l'équation de 1927 est nettement plus faible que celui découlant de l'équation « dynamique »**

Depuis l'arrivée de l'informatique, la contrainte résultant du choix systématique d'équations conduisant aux calculs les plus simples possibles n'existe plus. La bonne méthode dans un tel cas est de concevoir des équations les plus proches possibles des phénomènes que l'on souhaite simuler, et d'utiliser un solveur pour en trouver les solutions. Au passage, ceci permet de simuler facilement des épidémies où  $R_0$  varie, ce qui est le cas dans toutes les épidémies réelles tout à la fois à la suite :

- d'évolutions spontanées (par peur) ou contraintes (confinement) du comportement de la population,  
 - ou d'éventuelles modifications de données biologiques (mutations importantes du virus, découverte de traitements plus efficaces).

Conclusion (qui n'est que la formalisation d'une lapalissade !) :

Pour rendre compte correctement de la dynamique d'une épidémie, les omniprésentes équations de type  $dR / dt = \gamma I$  figurant dans la plupart des modèles en sortie de compartiments où se déroule un phénomène biologique qui peut être estimé comme étant de durée fixe doivent impérativement être remplacées par des équations de type

$$dR(t) / dt = - dS(t-D) / dt$$

D étant la durée du phénomène (inverse de  $\gamma$ )

- dS / dt étant le débit d'individus à l'entrée du compartiment et dR / dt le débit d'individus à la sortie de ce même compartiment.

**III - Simulation de l'épidémie Covid-19 avec un modèle SIR où l'équation  $dR/dt = \gamma I$  est remplacée par une équation adaptée aux phénomènes dynamiques :  $dR(t) / dt = - dS(t-1/\gamma) / dt$**

Compte tenu des nombreux chiffres accessibles sur divers sites Internet, il est possible d'introduire dans un tel modèle des paramètres  $R_0$  et D permettant d'obtenir une bonne cohérence avec l'évolution observée de l'épidémie.

D = 10 jours (soit  $\gamma = 0,1$ )

N = 67 000 000 S : nombre d'individus susceptibles d'être contaminés

$R_0$  évoluant entre 0,7 et 3

$R_{\text{effectif}} = R_0 \times S / N$

$\beta = R_0 / (N \times D)$

1 - jusqu'au 16 mars :  $R_0 = 3$

2 - baisse de 3 à 0,7 en 10 jours (début du 1<sup>er</sup> confinement)

3 -  $R_0 = 0,7$  jusqu'au 10 mai (1<sup>er</sup> confinement)

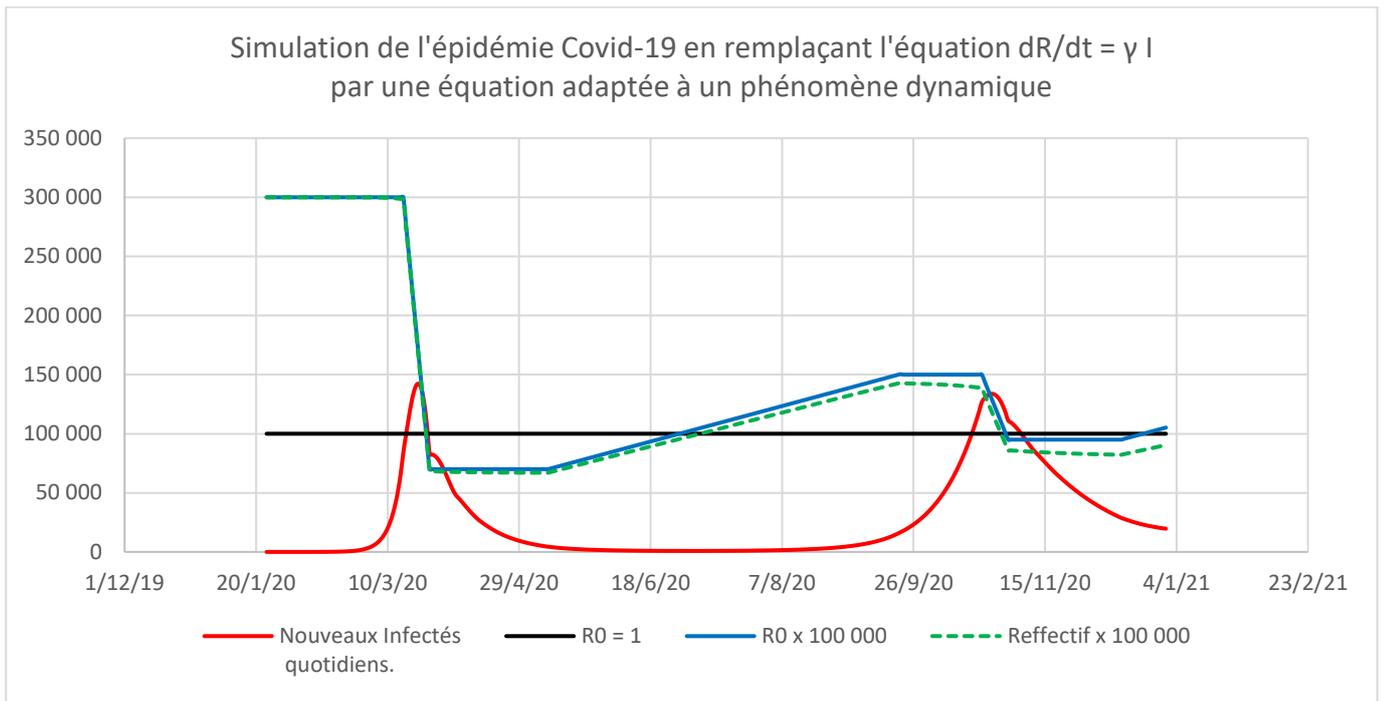
4 - hausse linéaire de 0,7 à 1,5 en 133 jours (1<sup>er</sup> déconfinement)

5 -  $R_0 = 1,5$  du 23 septembre au 22 octobre (fin du 1<sup>er</sup> déconfinement)

6 - baisse de 1,5 à 0,95 en 10 jours (début du couvre-feu suivi du 2<sup>ème</sup> confinement)

7 -  $R_0 = 0,95$  jusqu'au 14 décembre (2<sup>ème</sup> confinement)

8 - à partir du 15/12 hausse linéaire suivant la même pente qu'en 4 (2<sup>ème</sup> déconfinement)

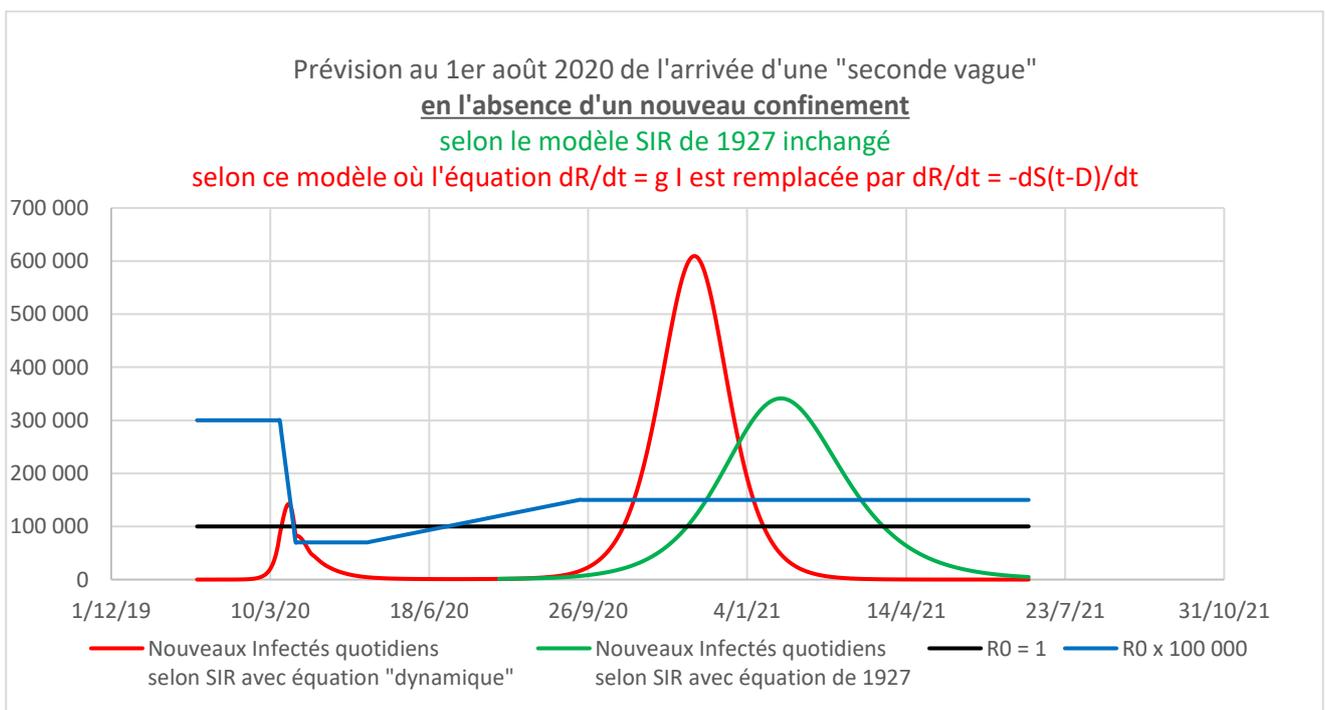


### Comparaison entre les prévisions de « 2<sup>ème</sup> vague » en utilisant les deux types d'équations

On se place au 1<sup>er</sup> août avec les effectifs donnés par le modèle utilisant l'équation rectifiée pour :

- les infectés
- les personnes susceptibles d'être infectées (en excluant donc les personnes immunisées suite à leur guérison ainsi que les personnes décédées).

Puis on demande aux tableurs utilisant les deux types d'équations de donner l'évolution des infections jusqu'à la fin de juin 2021.



On constate immédiatement que le modèle utilisant l'équation « dynamique » prévoit à partir de septembre une montée beaucoup plus rapide du nombre de nouveaux infectés quotidiens que le modèle utilisant l'équation de 1927, avec un pic théorique, en l'absence de nouveau confinement, de plus de 600 000 infections quotidiennes atteint dès décembre 2020. Le modèle fondé sur « l'approximation de 1927 » prévoit une montée nettement plus lente avec un pic théorique de plus de 300 000 infections quotidiennes atteint en février 2021.

Là encore, l'équation « dynamique » est nettement plus conforme à la réalité observée à l'automne 2020 que ce que prévoient les modèles basés sur l'équation traditionnelle de 1927.

### Cas d'une décroissance du nombre d'infectés

Il convient de remarquer que le caractère « dynamique » de l'équation proposée fonctionne dans les deux sens : il rend compte, en cas de confinement amenant  $R$  en dessous de 1, d'une diminution réelle du nombre d'infections nettement plus forte que ce que laissent prévoir les modèles basés sur l'équation de 1927.

En effet, en cas de décroissance du nombre d'infections, si on considère l'ensemble des personnes infectées au jour  $J$ , elles sont divisées en  $D$  sous-ensembles :

Infectés depuis 1 jour
Infectés depuis 2 jours
..... / .....
Infectés depuis $(D - 1)$ jours
Infectés depuis $D$ jours

**Le nombre d'infectés diminue de jour en jour.**

**La taille de chaque case est proportionnelle au nombre d'individus qu'elle contient.**

L'« approximation de 1927 » de Kermack et McKendrick fait sortir quotidiennement du « compartiment I des infectés » la moyenne entre les  $D$  cases du diagramme ci-dessus, c'est-à-dire un nombre très inférieur à la réalité puisque les infectés qui sortent réellement du compartiment I sont ceux, nombreux, qui sont arrivés  $D$  jours auparavant.

**Minorant le nombre de sorties du « compartiment I », ce type de modèle diminue artificiellement la prévision de la vitesse à laquelle une épidémie décroît, par exemple lorsqu'un**

### **confinement fait passer $R_{\text{effectif}}$ en-dessous de 1.**

Ceci donne une possible explication, au moins partielle, de la différence entre la déclaration du Président Macron indiquant le 28 octobre 2020 que, "quoi qu'on fasse", il allait y avoir 9 000 patients Covid-19 en réanimation le 15 novembre et le chiffre réel observé à cette date, nettement plus faible. Il se peut que l'effet du couvre-feu (23 octobre) puis de 2ème confinement (29 octobre) ait été plus rapide que ce que prévoient des modèles basés, au moins partiellement, sur l'équation de 1927.

## **IV - Conclusion**

**L'équation proposée par Kermack et McKendrick en 1927 pour le calcul du nombre de malades sortant d'infection par guérison ou décès n'est valable que si leur nombre est stable, ce qui dans la réalité d'une épidémie n'est jamais le cas.** Si on remplace l'équation de 1927 par une autre tenant compte du caractère dynamique du flux de malades on constate que ceci induit une forte modification des prévisions du modèle : par exemple à  $R_0$  constant, l'équation de 1927 donne une pointe du nombre maximum d'infectés deux fois plus faible que l'équation « dynamique » par laquelle elle devrait être remplacée pour être conforme à la réalité, ainsi qu'un taux de croissance du nombre de personnes infectées très inférieur.

**Globalement, l'équation de 1927 conduit à minimiser les futures variations réelles du nombre d'infections. Elle est dangereuse pendant les phases où l'épidémie se développe, car elle peut conduire à sous-estimer les besoins futurs en équipement et en personnel médical. Mais inversement cette caractéristique joue dans l'autre sens pendant les phases de décélération, car finalement les actions des pouvoirs publics destinées à réduire le nombre de malades seront plus efficaces, dès que  $R_{\text{effectif}}$  passe en-dessous de 1, que ce qu'indique un modèle utilisant l'équation de 1927.**

## Annexe 3

### Exemple d'utilisation d'un tableur pour le modèle SIR proposé en avril 2020 à des étudiants (licence) en mathématiques.

*Le Professeur d'Université (mathématiques), consulté, nous a indiqué ne pas avoir réalisé que les équations de 1927 n'étaient pas adaptées aux régimes dynamiques. Toutefois l'objet de l'exercice n'était pas de chercher à modéliser au mieux l'évolution d'une épidémie, mais de montrer un exemple d'utilisation possible d'Excel*

Université d'Aix-Marseille  
Licence de maths PES

2019–2020  
Avril 2020

#### DM 5 UN MODÈLE SIR À TEMPS DISCRET

On étudie ici un modèle très classique : le modèle SIR.

**1.1. Base du modèle.** On part d'une population de  $N$  personnes. Certaines de ces personnes sont atteintes d'une maladie dont elles guérissent avec une certaine probabilité : chaque jour, une proportion  $\gamma$  guérissent (ou meurent), et on les appellent les personnes retirées. Ces personnes ne peuvent plus, ni tomber malade, ni contaminer les autres (c'est pourquoi, du point de vue de cette épidémie, les personnes mortes ou guéries jouent le même rôle).

On note  $S_n$  le nombre de personnes saines le jour  $n$  (on dit aussi susceptibles, ce qui veut dire susceptibles de tomber malade),  $I_n$  le nombre de personnes infectées (ou malades), et  $R_n$  le nombre de personnes retirées (d'où le nom du modèle : SIR, qui date des années 1920). On a toujours  $S_n + I_n + R_n = N$ , où  $N$  est la population totale.

On va supposer que le nombre de personnes qui tombent malade est proportionnel au nombre de personnes infectées et aussi au nombre de personnes saines (quand il n'y a presque plus personne de sain, il y a forcément très peu de gens qui tombent malade). Donc le nombre de personnes qui tombent malades à l'étape  $n$  est

$$\alpha S_n I_n / N$$

d'autre part, le nombre de personnes qui guérissent est égal, comme on l'a vu plus haut, à  $\gamma I_n$  : plus il y a de malades, plus il y a de personnes qui guérissent.

**Remarque 1.1.** Il est plus commode pour les calculs de normaliser la première formule en divisant la constante par  $N$ , on le verra plus tard.

1.1.1. La formule de base. On peut donc, si on connaît  $S_n$  et  $I_n$ , calculer  $S_{n+1}$ ,  $I_{n+1}$  et  $R_{n+1}$ , puisqu'on sait combien de gens tombent malade et guérissent.

Donner la formule qui permet de calculer la situation à l'étape  $n + 1$  en fonction de la situation à l'étape  $n$ .

1.1.2. Combien de variables indépendantes ? Notre présentation laisse croire qu'il y a

trois variables qui décrivent le système.

Montrer qu'en réalité, si on connaît les suites  $S_n$  et  $I_n$ , on connaît aussi la suite  $R_n$ , donner la formule qui permet de trouver  $R_n$  en fonction de  $S_n$  et  $I_n$ , et donner la formule qui permet de trouver  $S_{n+1}$  et  $I_{n+1}$  en fonction de  $S_n$  et  $I_n$ .

1.1.3. Paramètres et conditions initiales. Quels sont les paramètres du système, et la condition initiale ?

**1.2. Formalisation informatique.** On a donc une constantes  $\alpha$  (en pratique, comprise entre 0 et 10), une constante  $\gamma$  pour la guérison, et une constante  $N$ , taille de la population. On a trois tableaux  $S$ ,  $I$  et  $R$ , où  $S[n]$  est le nombre de personnes saines au jour  $n$ ,  $I[n]$  le nombre de personnes malades au jour  $n$ , et  $R[n]$  le nombre de personnes retirées.  $S$  pour Saine ou Susceptible,  $I$  pour Infecté,  $R$  pour retiré.

On fixe les conditions initiales, valeurs de  $S[0]$ ,  $I[0]$  et  $R[0]$  dont la somme fait  $N$ . Il reste à calculer le jour suivant. Mais  $S[n + 1]$ , c'est le nombre de ceux qui sont sains au jour  $n + 1$ ,  $S[n]$ , et ceux qui étaient sains le jour  $n$ , d'où :

$$S[n + 1] = S[n] + \alpha \times S[n]I[n] / N$$

On calcule de même  $I[n + 1]$  et  $R[n + 1]$ , on laisse ce soin au lecteur.

**1.3. Mise en place sur tableur.** On recommence comme pour le modèle précédent : On met le nombre de jours (pour se repérer) dans la colonne A : écrire  $n$  en A1, 0 en A2,  $=A2+1$  en A3, et recopier cette dernière cellule dans le reste de la colonne. On écrit  $\alpha$  en E1 et la valeur de  $\alpha$  en E2 (on met 1 pour commencer) et de même on écrit  $\gamma$  en F1 et la valeur  $\gamma$  en F2, et  $N$  en G1 et la valeur de  $N$  en G2.. On écrit  $S[n]$ ,  $I[n]$ ,  $R[n]$  en B1, C1, D1 et on met les valeurs initiale  $S[0]$ ,  $I[0]$ ,  $R[0]$  (par exemple  $N - 1$ , 1, le patient 0, et 0) en B2, C2, D2. en B3, C3, D3, on met les formules de calcul trouvées ci-dessus, et on recopie ces formules dans le reste des 3 colonnes. Ceci permet de calculer le développement de l'épidémie, si on admet les règles du Modèle

**1.4. Etude pratique.** Etudier le développement de l'épidémie, suivant les valeurs des paramètres  $\alpha$ ,  $\gamma$  et  $N$  (population : il est plus intéressant de prendre un grand nombre, 1000 voire un million), et suivant la valeur initiale  $I[0]$  (nombre de zombies au départ). On tracera le graphique donnant l'évolution de  $S_n$  et  $I_n$ , en faisant attention de prendre suffisamment d'étapes pour voir ce qui est significatif (cela dépend bien sur des paramètres  $\alpha$  et  $N$ , et de la condition initiale  $I_0$ ; dans quel cas faut-il beaucoup d'étapes pour voir l'évolution ? Que peut-on en dire ? Que peut-on dire du début de l'évolution si  $N$  est très grand, et  $I_0 = 1$  ? On peut dans ce cas calculer  $\log(I_n)$  (mettre en colonne G  $=\log(Cn)$ , et représenter graphiquement cette suite en insérant un graphique ; on pourrait aussi, plus rapidement, insérer un graphique montrant la colonne C, et choisir une échelle logarithmique pour l'axe vertical. Que peut-on dire pour la suite  $I[n]$  ? est-ce que son type de croissance dépend des conditions initiales ?

## 2. Les solutions possibles

**2.1. Décroissance de  $S_n$ .** Montrer que, si  $0 < S_n < N$ , on a  $S_{n+1} < S_n$ .

En déduire que, si  $0 < S_0 < N$ , la suite  $I_n$  est croissante, et la suite  $S_n$  est décroissante.

Indication : on connaît la valeur de  $I_{n+1} - I_n$ , qui est aussi l'opposé de  $S_{n+1} - S_n$ .

Montrer que la suite  $R_n$  est croissante.

Que peut-on dire de la suite  $I_n$  ?

**2.2. Limite de  $S_n$ ,  $I_n$  et  $R_n$ .** Que peut-on dire des limites de ces trois suites ?

**2.3. Début de l'évolution.** On suppose que  $N$  est très grand, et  $I_0$  très petit ; par exemple,  $N$  vaut un milliard, et  $I_0$  vaut 1. Vérifier sur des exemples que dans ce cas, au début de l'évolution,  $I_n$  reste petit par rapport à  $N$ , et que  $S_n$  est approximativement égal à  $N$ .

Montrer que dans ces conditions, on peut simplifier l'équation qui donne  $I_{n+1}$  en fonction de  $I_n$  ; écrire cette équation, et donner sa solution. Expliquer ce que l'on a constaté à la question 1.4. Quelle est la condition nécessaire et suffisante pour avoir une épidémie ?

**2.4. Stratégies face à une épidémie.** Expliquer, en terme d'épidémie, ce qu'on appelle l'immunité de groupe, et l'effet d'une stratégie de confinement et de mitigation.

## Annexe 4

### A tentative tracking of the SARS-Cov2 pandemic in France, based on a corrected SIR model including vaccination effects

Mathilde Varret<sup>1</sup>, François Xavier Martin<sup>2</sup> and François Varret<sup>3\*</sup>

<sup>1</sup> Université de Paris, INSERM, Laboratory for Vascular Translational Science (LVTS), F-75018 Paris, France

<sup>2</sup> Member of the Editorial Committee of the Ecole Polytechnique alumni review, 12 rue de Poitiers, 75007 Paris, France

<sup>3</sup> Emeritus Professor, Université de Versailles Saint-Quentin, 45 avenue des Etats-Unis, 78035 Versailles cedex, France

\* corresponding author : varret\_francois@yahoo.fr

**Abstract.** We developed successive extensions of the SIR model in order to track the dynamics of the SARS-Cov2 disease. The analysis of health system available data is published in a chronicle accessible on the net: <https://corona-circule.github.io/lettres/>. This chronicle was initiated on late march 2020 and up to now contains 50 issues. A constant concern was the reliability of the data: for instance, we very soon evidenced that the number of confirmed cases, because of the asymptomatic carriers and the erratic testing policy, was hugely underestimated. By the end of 2020 we made a basic change in the model which consisted in accounting for a constant contagiousness time (SIR-tcc) instead of the probabilistic evolution of the end of the infection assumed so far. Recently we completed this SIR-tcc model for the vaccination effects in order to properly track the evolution of the group immunity threshold. Calculations were performed using the Excel facility (Microsoft), allowing a manual fitting of the model parameters. The results have dealt with a large number of countries, but we focus here on the data regarding France. Further pieces of information are also presented, in order to help elucidating some the factors responsible for the complex history of the pandemic dynamics. (submitted dec 14<sup>th</sup> 2021)

# 1 The epidemiological context

## 1.1 The SARS-Cov2 virus

The 21st century saw the emergence of three coronaviruses responsible for deadly severe acute respiratory syndromes in humans (SARS): SARS-CoV [1,2], MERS-CoV [3], and SARS-CoV-2 [4,5] which is causing the COVID-19 disease. As for the first two, the reservoir host for SARS-CoV-2 is the bat [6]. Whereas palm civets [7] and dromedary camels [8] are recognized for the transmission of, respectively, SARS-CoV and MERS-CoV from bats to humans, the SARS-CoV-2 intermediate host is still not known. Like other respiratory pathogens, including influenza and rhinovirus, coronavirus transmission occurs through airborne droplets produced by sneezing, coughing, and breathing [9].

Coronaviruses are ribonucleic acid (RNA) viruses, their genetic material is a single-stranded RNA molecule, carrying four main structural proteins - Spike, envelope, membrane, nucleocapsid - and several non-structural proteins [10]. The spike surface protein of the coronaviruses mediates their entry into host cells through a variety of receptors and entry mechanisms [11]. Like SARS-CoV, SARS-CoV-2 uses the angiotensin-converting enzyme 2 (ACE2) as a receptor to enter cells [12,13]. Inside the host cell, viral RNA is translated into non-structural proteins that suppress the host genes expression in favour of the virus ones [14,15]. Then, viral RNA is replicated, translated and new virus particles are formed that will strike other host cells [16]. Infected cell fuses to neighbouring cells expressing ACE2, giving rise to massive individual respiratory cells that prosper for long and produce more and more new virus particles [17]. Beside the rapid virus replication, the innate immune cells response takes place with the secretion of pro-inflammatory cytokines and proteases [18]. This excessive inflammation, called “cytokine storm”, leads to acute respiratory distress syndrome and pulmonary oedema [6]. However, since the virus suppresses the host genes expression, including those encoding proteins that alert the immune system, it escapes the immune response of the host.

The consequences of Sars-CoV-2 infection vary greatly from one person to another. While 90% of the infected individuals are not very symptomatic or asymptomatic, 10% develop severe, or even critical, forms with pneumonia requiring an intensive care stay. The viral load is obviously a major risk factor. Age is also a major risk factor for the development of severe forms, with individuals over 65 years having the greatest risk of requiring intensive care [19], and men are more susceptible than women [20]. Nearly 25% of severe forms are due to a deficiency in the type 1 interferon (INF1) pathway, a cytokine usually produced rapidly by the immune system in response to a viral infection and whose main effect is to inhibit virus replication in infected cells). With 3-4% of *INF1* gene alterations [21] and 15-20% of presence of autoantibodies neutralizing INF1 and blocking its antiviral action [22-23]. The presence of INF1 neutralizing autoantibodies is very rare before age 65 (0.2 to 0.5%) and then increases exponentially with aging to reach 4% between 70 and 79 years, and 7% between 80 and 85 years. The causes and mechanisms of this increase in the general population remain to be elucidated, but it partly explains why age is a major risk factor in the development of severe forms of Covid-19. Defects in the gene encoding the toll-like receptor 7 (TLR7), an RNA intracellular receptor playing a major role in INF1 production, are found in 1.8% of men with severe forms under age 65 [24]. The TLR7 gene being on the X chromosome explains why men are more frequently affected by severe Covid-19 than women.

## 1.2 SARS-CoV-2 variants

The SARS-CoV-2 virus is constantly mutating and numerous SARS-CoV-2 variants have emerged, naturally selected for their higher transmissibility, lethality or capacity to escape the host’s immune response in comparison with the original virus (Table 1). Transmissibility is related to permeability and represents the ability of the virus to enter host cells. Lethality is the proportion of infected individuals who die from the infection in a given place, at a given time. Lethality is to be differentiated from virulence which corresponds to the propensity of the virus to harm its host. Virulence is quantified in the absence of specific care. The same variant will have a different lethality from one country to another depending, for example, on the quality of the hospital system, but its virulence will be unchanged.

**Table 1.** Effects of the main SARS-CoV-2 variants, in comparison with the original virus, on their transmissibility, lethality, and ability to escape the host immune response [25-28].

Variant	Transmissibility	Lethality	Immune escape
Alpha B.1.1.7	43-90% higher	No change	Slightly increased
Beta B.1.351	1.5x higher	Possibility increased fatality rate	Higher
Kappa & Delta	2.6 x higher	Elevated in Brazil (may be due only	Higher

B.1.617.1 &2		to healthcare system overload)	
Gamma P.1	1.7-2.4 x higher	Higher	Higher
Epsilon B.1.427&429	19-24% higher	Yet not known	Slightly increased

The most recent variant which is in the headlines at the end of 2021, Omicron (B.1.1.529), carries many mutations present in other variants. Of the 32 Omicron mutations in the spike protein, 16 are in the Delta variant. These known mutations seem worrying since they are linked to a greater infectivity and a better ability to escape the immune response. And numerous new mutations seem able to contribute to a further escape ability, and a more efficient virus replication [29-31]. Nevertheless, there is still a lot to understand about the Omicron variant at the biological, clinical, vaccination and epidemiological level.

### 1.3 Vaccines

In just one year, 322 candidate vaccines, based on the Wuhan virus, had been proposed and 9 were approved for use to adults, and in some cases to adolescents through various region-specific and regulatory agency approval procedures (Table 2).

**Table 2.** Characteristics of COVID-19 vaccines approved to date [32-33].

Vaccine manufacturer (vaccine name)	Platform	Efficacy (%)	
		From trials	For severity*
Sinovac Biotech (CoronaVac)	Inactivated virus	51	51 SC, 100 SD, H
Sinopharm (BBIBP-CorV)	Inactivated virus	78	100 SD, H
AstraZeneca–University of Oxford (AZD1222)	Viral vector	81	100 H
Johnson & Johnson (Ad26.COV2-S)	Viral vector	66	85 SD
Gamaleya (Sputnik V)	Viral vector	92	-
Bharat Biotech (Covaxi)	Viral vector	78	100 H
Pfizer–BioNTech (BNT162b2)	mRNA	95	100 SD
Moderna (mRNA-1273)	mRNA	94	100 SD
VECTOR (EpiVacCorona, NCT04780035)	Protein subunit	-	-

\* SD: severe disease, H: hospitalization, SC: symptomatic COVID-19 disease.

Three major vaccine technology platforms were exploited. In its most traditional form, the principle of a vaccine is to inoculate the virus against which we intend to prepare the immune system. To prevent the virus from causing disease, it will first be either inactivated (like the flu and polio vaccines) or attenuated (like the pertussis and tuberculosis vaccines). Viral vector and mRNA vaccines are of the so-called gene type. These technologies consist of inoculating the whole or parts of the virus genetic material, DNA or mRNA, so that the vaccinated organism himself will produce the immunogenic proteins. The carrier, or vector, is generally a benign virus, adenovirus, which will stimulate the immune response while remaining harmless. The protein subunit vaccine technology uses three peptides of the spike protein planted on a chimeric protein constituted of two parts (viral nucleocapsid protein and bacterial maltose-binding protein). The vaccinated organism will respond by making antibodies against the Spike protein. This technology has already been used, for the hepatitis-B vaccine by example.

Efficacy against symptomatic COVID-19 disease is calculated from end points defined in the clinical trial, severe disease, hospitalization, and/or symptomatic COVID-19 disease, which differ between trials, even for the symptom's definition (Table 2). It is an important value for the approval of vaccines, but does not reflect the real-world impact, especially when trials have been performed on specific population (younger, healthy adults, individuals at risk of severe disease), and in the absence of the more recently reported SARS-CoV-2 variants. It is thus essential to know the extent and duration of protection across all age groups and populations. Especially given the higher risk of severe form in the elderly (age > 70). The efficacy may be affected by various factors including the population, the vaccine, the vaccination schedules and the handling/administration of the vaccine.

Most of these vaccines were developed against the Wuhan virus and showed an efficacy between 50% and 95% (Table 3). Since numerous SARS-CoV-2 variants have emerged and continue to emerge, characterizing their impact on vaccines efficacy quickly became necessary. Their efficacy against the dominant variant today, Delta, remains very good (Table 3).

**Table 3.** COVID-19 vaccines efficacy according to SARS-CoV-2 variants [32-34].

Vaccine manufacturer (vaccine name)	Efficacy (%) against variant				
	Wuhan strain	Alpha	Beta	Gamma	Delta
Sinovac Biotech (CoronaVac)	50	-	-	-	59
	70				
Sinopharm (BBIBP-CorV)	78	-	-	-	-
AstraZeneca- University of Oxford (AZD1222)	55	75	10	78	92
	81				
Johnson & Johnson (Ad26.COV2-S)	66	70	57 72	68	-
Gamaleya (Sputnik V)	-	-	-	-	90
Bharat Biotech (Covaxi)	-	-	-	-	65
Pfizer-BioNTech (BNT162b2)	95	94	86	51	88
Moderna (mRNA-1273)	94	100	96	-	92

While vaccines are reported to reduce the number of symptomatic COVID-19 cases, direct evidence for their effect on virus transmission is limited. The BNT162b2 (Pfizer-BioNTech) vaccine has been shown to reduce the viral load [35]. And since a reduced viral load is associated with decreased transmission of the virus, these data together suggest that vaccination may reduce transmission.

#### 1.4 Data and websites

Basic data for the number of confirmed cases, deceased persons and vaccine doses can be found on the various websites of National Health Systems [36]. Some sites update worldwide compilations [37, 38], while others built-up useful figures or maps which provide synthetic views of the evolution of the pandemic [39, 40]. Among them, the CovidTracker site [39] provides very easy-to-read figures of crucial effects such as the age effect which was the constant background of our investigation.

Prior to any analysis it may be noted that all reliable indicators (some are not!) of the pandemic are not necessarily sensible. For example, the number of removed people is not known in real time. Consequently, the number of infectious people at a given date is not known. Only the total number of infected persons, that is, which have been contaminated in the past, is provided by the health systems. Infected persons are usually termed "cases", and the recognition of a new case is based on the positive result of a PCR test. However, the number of cases is obviously underestimated due to asymptomatic carriers. This gives rise to the famous Trump theorem [40] "*more covid19 testing creates more cases*". On the contrary the number of reported deaths assigned to the COVID-19 appears to be (for most countries) reliable.

On the example of France, we could check a good compliance between the number of deaths quoted during the first pandemic and the seasonal excess mortality. [CC 4]

A model is used here for deriving correlations between the available sensible data, and eventually deriving the non-available ones, in order to elucidate some of the basic factors of the pandemic. The reverse way consists in using models for predicting the effects of possible factors. Such models may be extremely complex, they mostly deal with

inhomogeneous effects and require the efforts of large groups or researchers [42-46]. Their conclusions were often used as guidelines for the present approach which modestly remained in the frame of the mean-field approximation.

The analysis of health system available data is published in a chronicle accessible on the net: <https://corona-circule.github.io/lettres/>. For convenience the chronicle issues are abbreviated here “CC n...” where n is the issue number

## 2 The SIR model

### 2.1 The uncorrected SIR model

The historical SIR model, where S, I, R respectively stand for susceptible, infectious and removed people, was created by mathematicians almost one century ago [47]. It was widely used till now through its mean-field approach, which actually was the “special case” for which an analytical solution could be given, see [48] for a recent description (restricted to the special case). For convenience we call this special case “the uncorrected SIR model”. The population is considered as a 3-state system which irreversibly evolves from the unstable initial state S to the stable state R through the intermediate state I.

#### 2.1.1 Description of the uncorrected SIR model

In the mean-field approach, S(t), I(t), R(t) are the time dependences of the population fractions associated with these 3 states. The probabilistic approach is based on transition rates  $\beta$  and  $\gamma$  for the successive transitions  $S \rightarrow I$  and  $I \rightarrow R$  respectively, see figure 1. The peculiarity of the pandemic mechanism is that the flux  $S(t) \rightarrow I(t)$  is proportional to the infectious fraction. This gives rise to a self-accelerated dynamics at the beginning of the pandemic, a well-known phenomenon in the field of population dynamics.

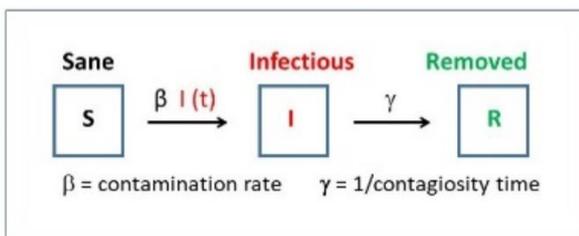


Fig. 1. A schematic view of the uncorrected SIR model.

$\beta$  the contamination rate is defined as the probability per unit time for a sane person to be contaminated due to meeting a contagious one. This crucial parameter depends on the intensity/frequency of contacts between people as well upon the intrinsic contagiousness of the virus. It is strongly impacted by the irruption of virus mutants (the so-called variants) and by the level of social constraints (lock-in, curfew, travel or activity restrictions, masks...). For these reasons it is essentially time dependent. The adjustment of  $\beta(t)$  to the sanitary data (deceases, hospitalization, contamination cases) is a must for any sensible tracking of the pandemic.

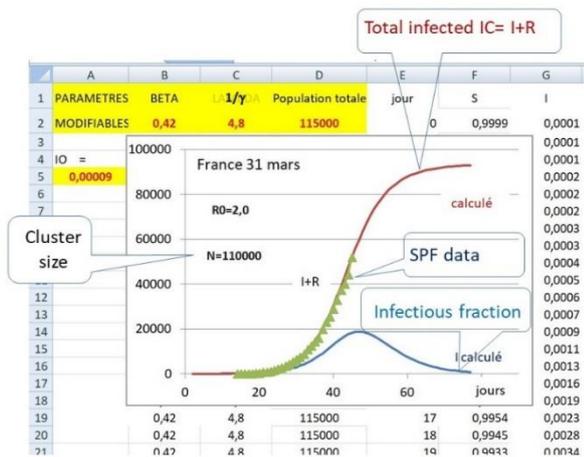
In the traditional approach  $\gamma$  is the probability per time unit for a given infectious person to reach the removed state, dead or recovered. It is the reverse of the average duration of the contagious state (or contagiousness time).

The master equation of the SIR system follows:

$$dS/dt = -\beta S I; \quad dI/dt = \beta S I - \gamma I; \quad dR/dt = \gamma I \quad (1)$$

such as  $dS/dt + dI/dt + dR/dt = 0$   
because  $S(t) + I(t) + R(t) = 1$

The set of coupled differential equations (1) is easily solved by using an Excel spreadsheet (Microsoft Office), where time t is increased by steps of one day. The adjustment is performed manually thanks to the real-time display of the calculated curves. Our first calculation, dated 2020/03/31, see Figure 2, led to claim that the top of the first wave was nearly reached. The peak of daily cases actually occurred 1 – 2 weeks later.



**Fig. 2.** An early adjustment using the uncorrected SIR model based on the total case data for France (adapted from [CC 2]). IC is the fraction associated with the total number of cases

In this preliminary calculation, all parameters were kept constant. The cluster size,  $\beta$  and  $\gamma$  were both adjusted. Later on,  $\beta$  was taken time-dependent, and  $\gamma$  was given a better value 1/14 days suggested by health system websites.

A crucial problem was to properly assign the date of the beginning of the pandemics. In principle the use of continuous functions is not consistent with the existence of a “patient zero”. In addition, the date of the first contamination was unknown – and till now remains so. In practice, this problem is solved by arbitrarily fixing the first day of the pandemics, and adjusting the initial (non-null) value of the infectious fraction. Here, day 1 was arbitrarily fixed March 1, 2020.

**2.1.2 Initial dynamics**

The initial dynamics of the pandemic is exponential. Indeed for small t:

$$S(t) \sim 1 \text{ and } I(t) \sim 0 \Rightarrow dI/dt \sim (\beta - \gamma) I$$

$$\Rightarrow I(t) \sim I(0) \exp \{ t / (\beta - \gamma) \}$$

The exponential is growing for  $\beta > \gamma$  and decreasing in the opposite case.

**2.1.3 Reproduction rates**

The initial reproduction rate  $R_0$  is the average number of persons infected by each infectious person at the beginning of the pandemic.  $R_0 = \beta / \gamma$ , that is, the contamination rate x the contagiousness time. For  $R_0 > 1$  the exponential speeds up, for  $R_0 < 1$  it slows down, in agreement with sec. 2.1.2.

The value of  $R_0$  depends on both the virus contagiousness (which depends on the variant) and on the frequency of contacts between persons. Social restrictions tend to decrease the reproduction rate, while the irruption of each more contagious variant tends to increase it. The balance between these opposite effects is crucial and has to be accurately followed.

However, the most popular rate is the effective reproduction rate  $R_{eff}(t) = R_0(t) \times S(t)$ , which through eqn. (1) governs the dynamics of the pandemic at time t. For  $R_{eff}(t) > 1$  the pandemic grows, for  $R_{eff}(t) < 1$  it shrinks. The tracking of  $R_{eff}(t)$  is a central problem (see [49] for a review of the methods). Santé Publique France quotes the Pasteur Institute results obtained through the Cori method [50]; a very simple - approximate - method was recently introduced by the CovidTracker site [39].

In our approach  $R_{eff}(t)$  is an output of the data analysis, it is derived from the fitted sequence of the  $\beta(t)$  values.

**2.1.4 Group immunity**

Group immunity is a state expected when the effective rate  $R_{eff}(t) = R_0(t) \times S(t)$  becomes smaller than the threshold value 1 due to the irreversible decrease of  $S(t)$ . For a constant  $R_0$  the behavior is clear and once the threshold value  $S_{thres} = 1 / R_0$  is reached, the pandemic irreversibly slows down and the infected fraction tends to saturate significantly below the ultimate limit 1. For a time-dependent  $R_0(t)$  the behavior is far from clear and group immunity should be better understood as a condition which is (temporarily!) fulfilled when  $S(t)$  is smaller than the threshold value  $1 / R_0(t)$ . In absence of vaccination effects this is strictly equivalent to  $R_{eff}(t) < 1$ . The additional immunity created by large scale vaccination will make it different (see Sec. 2.4).

## 2.2 Adaptation of the uncorrected SIR model

### 2.2.1 Hidden part and visibility ratio

A first adaptation was required by to the large number of asymptomatic carriers, which was first recognized by the Pasteur Institute in April 2020 [51], and led us an analysis exclusively based on the decease data.

We therefore completed the master equation (1), with the deceased fraction,  $D(t) = \ell R(t)$ , (2)

where  $\ell$  is the lethality ratio, that is, the probability for a contaminated person - non-vaccinated - to decease. This lethality ratio is strongly age-dependent, it drastically increases as a function of age.

Meanwhile, we observed that the adjusted size of the cluster was continuously increased at each new analysis; we were also puzzled by the incredible diversity of the lethality rates reported for various countries. We thought that this diversity was due to the existence of asymptomatic carriers and introduced the idea of the “**hidden part**” of the pandemics [CC 4],

Due to the presence of this hidden part, we changed the basic assumptions of the analysis, as follows: (i) the case data were corrected by an adjustable constant factor which is the **visibility ratio** of the pandemics quoted in the previous section; (ii) the lethality ratio was initially fixed:  $\ell=0.53\%$  in agreement with a detailed report of the Pasteur Institute [52], and later on  $1.06\%$  [CC 23] according to a more extended study [53]; (iii) the total population (France = 67 million inhabitants) was accounted for; (iv) last, but not least, the contamination rate  $\beta(t)$  was no longer a constant. To avoid handling too many free parameters, we used a stepped  $\beta(t)$  function with 10-day steps (5-day, later on).

We show in Figure 3 an example of such an analysis when the first wave was just over. The value of the visibility ratio was adjusted so as to match both the case and death data.

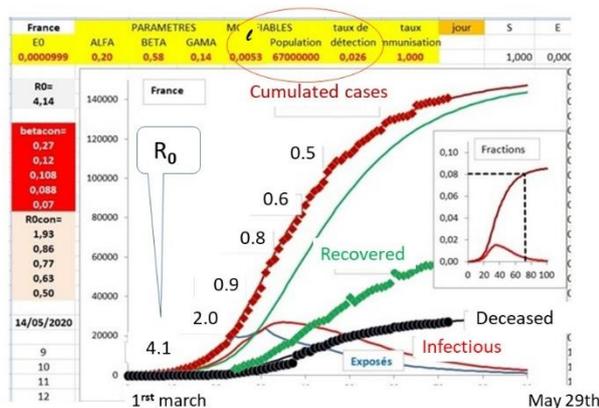


Fig. 3. The first wave in France, dated late May 2020, adapted from [CC 4]. Complete lock-in was effective since mid-march.

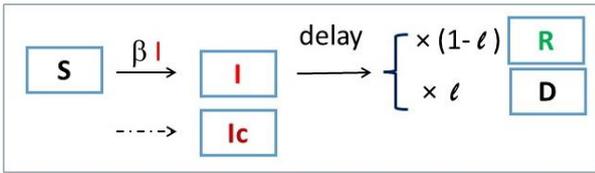
The main result shown in Figure 3 is the decrease of the reproduction rate, associated with the decay of the first wave. This decay is due to the complete lock-in from mid-March to mid-May. The visibility ratio, taken here as a constant, is extremely small, a few % only, and was confirmed by the detailed report of the Pasteur Institute in May 2020 [52].

When studying the following phase of the pandemic, we found that such a simultaneous fit of the death and case data was no longer possible. Indeed, the value of the visibility ratio irreversibly increased for each new wave. We then decided to fit a single set of data; to start with, the decease data, because they were consistent with the seasonal mortality data (Sec. 1.4). Then the visibility ratio became another output of the analysis, time dependent, defined as the total number of confirmed cases divided by the calculated total number of infected persons. This ratio presently reaches around 66 % (detailed data in [CC 43]).

## 2.3 The corrected SIR-tcc model

By the end of the year 2020 our attention was drawn to a contribution to the French Ecole Polytechnique alumni review [54] which pointed out the lack of relevance of the uncorrected SIR model, where the recovery from (or death after) the contamination of a given person is a random process governed by a transition rate. The main drawback of this assumption is that the recovering probability of a given person is maximum on the very day of contamination and smoothly decreases as a function of time. This is totally against the idea of an average recovery time, which is commonly admitted in the medicine field (for instance, around 14 days for the covid-19 disease).

It is very surprising that the obvious deterministic character of  $I \rightarrow R$  process was clearly accounted for so lately (the model is 90-years old). For our part, we turned to the duly corrected version and termed it SIR-tcc (for constant contagiousness time).



**Fig. 4.** The corrected model (SIR-tcc). Ic stands for the total infected fraction

The change in the spreadsheet organization was straightforward [CC 23] by defining delayed output fluxes from  $I(t)$  to  $R(t)$  and  $D(t)$ , which merely are the input flux at time  $(t - \ell)$ , multiplied by the respective  $(1 - \ell)$  and  $\ell$  factors. The delay value was taken in continuity of the previous model, that is,  $1/\gamma$ , same for all and constant in time (15 days).

In practice, the output flux was taken as a 7-day average value in order to smooth eventual discontinuities generated by the model [CC 24].

### 2.3.1 Comparison to the previous model

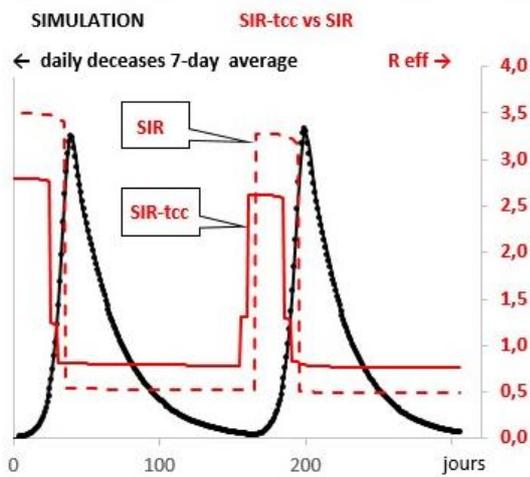
As evidenced by one of us (FXM) [55], the uncorrected SIR model provides a too weak and delayed response to the variations of the  $\beta$ -parameter.

If  $\beta$  increases, the uncorrected SIR model equations lead to overestimate the number of recoveries and deaths, as they do not take in account the delay between the  $S \rightarrow I$  and  $I \rightarrow R$  processes for an given individual. As a consequence, the SIR computed number of infectious is too small.

For the same reason, if  $\beta$  decreases, the uncorrected SIR model equations lead to underestimate the number of recoveries and deaths: the computed number of infectious is too high.

The uncorrected SIR model is not reactive enough to  $\beta$  variations. No doubt the SIR-tcc model definitely has to be preferred.

Accordingly, the analysis of the data using the SIR-tcc model will obviously provide results different from those of the uncorrected model. This is illustrated in Figure 5.



**Fig. 5.** Comparison of the  $R_{\text{eff}}(t)$  curves derived from obtained with the uncorrected SIR and SIR-tcc models, for the same set of (simulated) data.

For convenience the data shown in Figure 5 were generated using the uncorrected model, with a schematic squared-shaped  $\beta(t)$  involving two contamination waves. The asymmetric shape of the resulting peaks is worth noting and contrasts with the symmetrical shape obtained with the corrected model (not shown here).

The responses of the SIR and SIR-tcc analyses of these simulated data, as shown in Figure 5, are sizably different: the uncorrected model analysis overestimates the fitted  $\beta(t)$  curve – with respect to the corrected model analysis - so as to compensate its weaker response to  $\beta$ -variations. The uncorrected model analysis delays the fitted  $\beta(t)$  curve, a far from intuitive consequence of its delayed response to  $\beta(t)$  variations. The latter effect is larger on the decreasing side of the peaks.

In addition, we found that the manual adjustment procedure was much easier with the SIR-tcc model. This might be a consequence of the endless character of the recovery process, which induces strong mutual influences between the values of the adjustable parameters during the fitting procedure.

### 2.3.2 Results for France

The results for France, recently updated, are shown below:

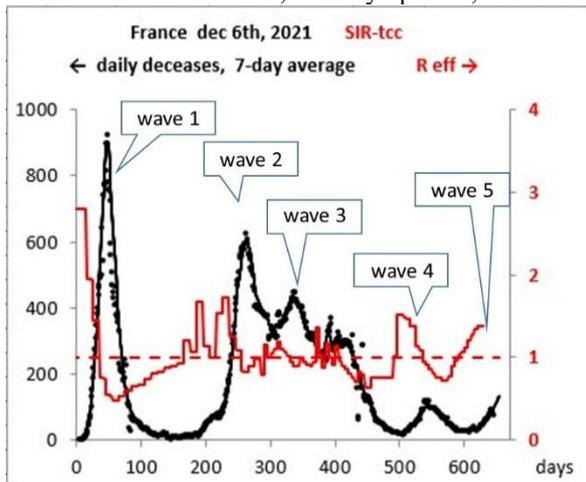


Fig. 6. The successive 5 waves of the pandemic in France, analysed through the SIR-tcc model, adapted from [CC 43].

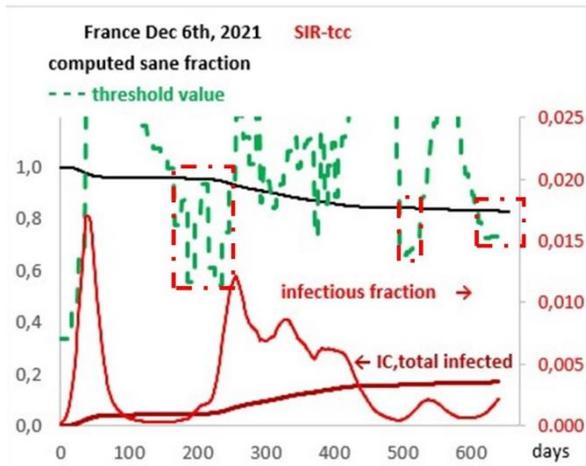
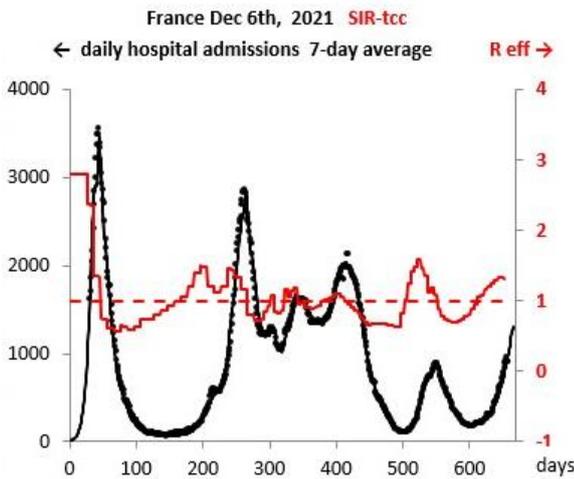


Fig. 7. The compared evolutions of the computed fraction  $S(t)$  (solid black line) and of its group immunity threshold (broken green line), for France. The situation of group immunity has been reached several times, when the solid black line was below the slashed green one; the situations where group immunity was not reached are been pointed out by red dashed frames. Adapted from [CC 43].

### 2.3.3 Alternative indicators

Alternative indicators such as the daily number of confirmed cases, hospitalisations or admissions to the intense care sections may be used. These data presumably are almost as reliable as the decease ones. We have treated the hospitalization data in France, in the same way we did for the decease data. The lethality ratio merely is substituted by a **severity ratio** assumed to be time-independent and the same for all. The value given to this severity ratio (4 %) was determined in order to yield visibility ratio values close to those obtained by the decease-based analysis. [CC 30]

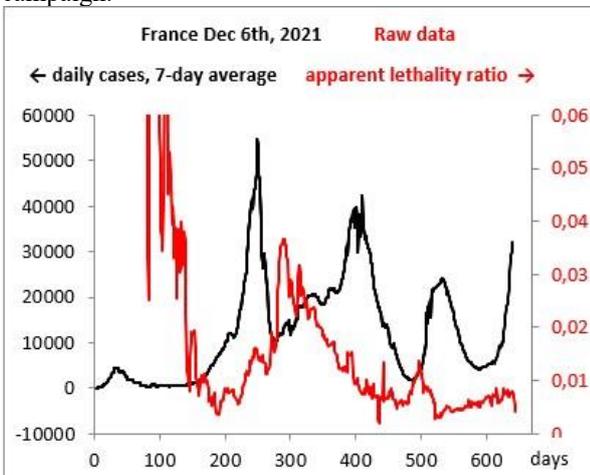


**Fig. 8.** An independent analysis based on the daily hospitalization data of France. Note that the  $R_{\text{eff}}(t)$  curve is similar to that obtained from the decease data (figure 6). Updated from [CC 30].

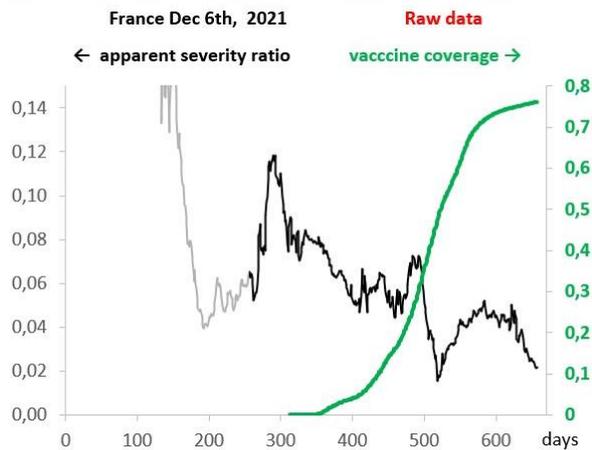
The  $R_{\text{eff}}(t)$  curves displayed on figures 6 and 8 are similar. A complete comparison of the curves associated with the possible 3 indicators (cases, hospitalizations, deaths) will be presented in Sec. 3, Figure 16.

### 2.3.4 Apparent lethality and severity ratios

Daily hospitalization and death data may be directly compared to the case data, provided that a convenient delay is introduced. We denoted these ratios “apparent” because they refer to an indicator – the number of cases – which is subject to the absence of asymptomatic cases. The apparent lethality ratio is the number of deaths on day  $d$ , divided by the number of cases confirmed on day  $d - 14$  (elsewhere denoted *the case fatality rate*). The apparent severity ratio is the number of persons hospitalized on day  $d$ , divided by the number of cases confirmed on day  $d - 2$  (to start with: confirmed on day  $d - 14$ , but this excessive delay value was corrected later on [CC 43]). Both ratios are impacted by the intensity of the testing campaign.



**Fig. 9.** The evolution of the apparent lethality ratio in France. The large values prior to day 200 are attributed to the sufficient number of tests. The increase in the ratio after day 200 is assigned to the onset of variants (Beta, and later on Delta). The sizeable decrease after day 300 reflects the efficiency of a massive vaccination campaign targeted to the most fragile or eldest persons. Updated from [CC 30].



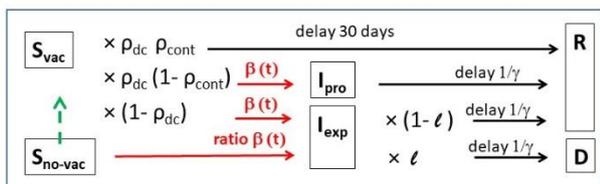
**Fig. 10.** The evolution of the apparent severity ratio in France. The global decrease after day 300 confirms the efficiency of vaccination reported in the previous figure. Additional features around days 300 and 480 might be assigned to concomitant weaknesses of the testing activity. Adapted and updated from [CC 30].

## 2.4 The SIR-tcc-vaccination extension

Worldwide vaccination campaigns started on the very end of 2020. The campaign was rather massive in France and reached a high coverage value (number of vaccinated persons / total population) close to 75 %, see Figure 10 (in previous section).

It appeared that this campaign has significantly hindered the larger contagiousness (by 50 %) of the so-called British variant with respect to the initial virus. In terms of protection against severe forms of the disease leading to intensive care sections and ultimately, to death, efficiency ratios up to 90% or more were reported, see Sec. 1.3. But, for a long time, little was known about the protection against contamination. Thanks to a detailed study of the correlation between the vaccinal coverage as a function of age and the relative intensity of the fourth wave [CC 42] we reached the conclusion that the protection against contamination was lesser, say 60-70 % smaller. It seems that the reduction of the viral load due to vaccination is the mechanism which leads to the large decrease in the probability of severe forms and death.

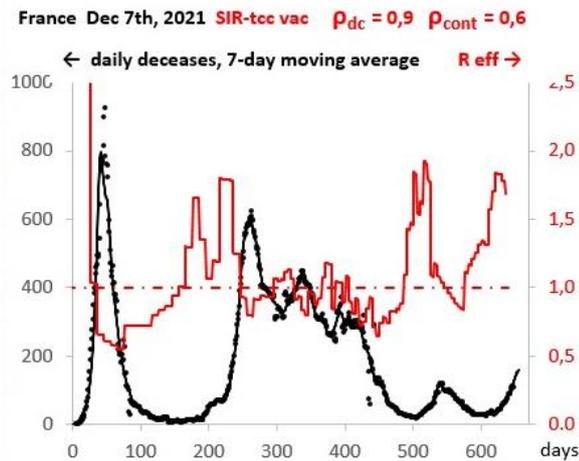
The model extension basically consisted in creating a second sane fraction for the vaccinated persons, and splitting the infectious fraction according to their protection or not against severe forms and death, see figure 11.



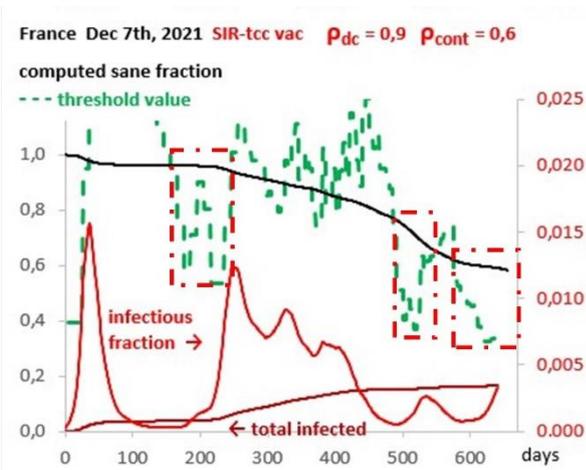
**Fig. 11.** Scheme of SIR-tcc-vaccination model.  $S_{vac}$ ,  $S_{no-vac}$ , respectively stand for the vaccinated or non-vaccinated sane fractions,  $I_{pro}$ ,  $I_{exp}$  for the protected or exposed infectious fractions, and  $\rho_{dc}$ ,  $\rho_{cont}$  for the protection ratios against decease and against contamination.

$$\text{The flux } S_{no-vac} \rightarrow S_{vac} = S_{no\_vac}(t) \times d \text{ Couv}(t)/dt \quad (3)$$

where  $Couv(t)$  is the vaccinal coverage (2-doses here). We assume for simplicity the same contamination rate  $\beta(t)$  for both  $I_{pro}$ ,  $I_{exp}$ . The equations follow straightforwardly and can be found in [CC 43].



**Fig. 12.** The France deceased data, analysed through the SIR-tcc-vaccination model, adapted from [CC 43]. To be compared to Figure 6.

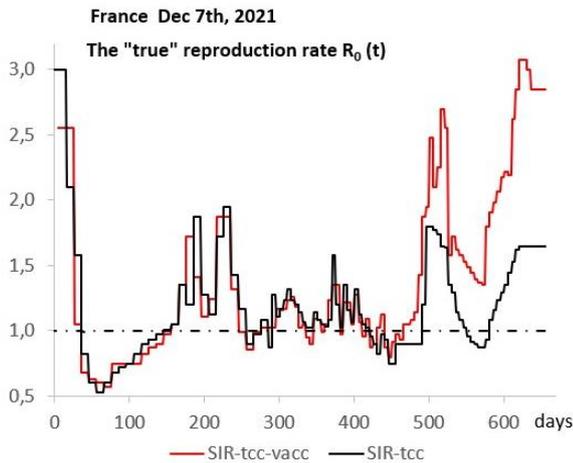


**Fig. 13.** The sane fraction  $S(t)$  (solid black line) and of its group immunity threshold (broken green line), obtained through the SIR-tcc-vaccination model. Adapted from [CC 43]. To be compared to Figure 7. It is worth noting that the infected fraction  $I(t)$  presently is in large excess of the peak of the previous wave.

An obvious change provided by the SIR-tcc-vac model is the lowering of both the  $S(t)$  and threshold value curves (compare Figures 7, 13). The combination of both effects finally leads to hinder the condition of group immunity. In other words, the analysis discarding vaccination effect sizably **overestimates** the possibility of reaching a future group immunity.

The  $R_{eff}(t)$  curve reported in Figure 12 only shows minor differences with respect to Figure 6. These differences presumably are due to the imperfect character of the manual adjustment procedure. An automatic fitting procedure might be helpful here.

On the contrary, the  $R_0(t)$  values are visibly increased by the account of the vaccination effect, see Figure 14.



**Fig. 14.** The “true” reproduction rate,  $R_0(t)$  obtained through the SIR-tcc and SIR-tcc-vaccination models, for France, adapted from [CC 43].

$R_0(t)$ , corrected for the vaccination effect, is presently larger than it was at the beginning of the pandemics. This statement means that the present social restrictions, however quite severe, hardly balance the contagiousness excess of the present variant ( $\delta$ ).

### 3 Discussion

Numerous research works are presently carried out by Universities or Institutes. They are mainly based on sophisticated models which aim to reveal all various factors which may govern the dynamics of the pandemic. Understanding these factors is needed to make sensible predictions of the long-term evolution of the pandemic.

Our approach is less ambitious, and only allows short-term extrapolations. However, the analysis of the past evolution, like as any historic research work, provides *a posteriori* measurements of the effect of social decisions such as lock-in, curfew, vaccination, sanitary pass... but this is a sociologic matter which is not in the very scope of the present work.

#### 3.1.1 The CovidTracker website approach

Meanwhile, the huge amount of data was processed in several websites [39, 40], so as to provide an easier access to the pandemic evolution through maps, graphs and figures. The CovidTracker site [39] reports an alternative determination of the effective reproduction rate, see Figure 15:

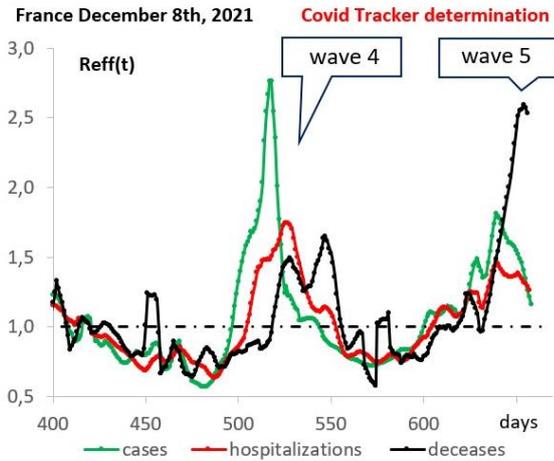


**Fig. 15.** The effective reproduction rate, derived from the relative variations of indicators (case, hospitalisations), after [39] dated 2021/12/08.

Here,  $R_{eff}(t)$  is assumed to be the relative variation of any indicator over 7 days (of course using 7-day moving averages). The relevance of this model-free method was a posteriori justified by the close compliance with the values quoted by Santé Publique France [36], that is, determined according to the Cori method [50] by the Pasteur Institute simulation group, see for example [56]. These results are also in good compliance with those of the present analysis. This simple approach required, of course, the tuning of the time interval over which the relative variation is calculated. We noticed that this 7-day interval exactly matches half the contagiousness time of the SIR-tcc model (14 days). A sensible

explanation is that a given contagious person may have been contaminated by a previous one all along its contagiousness time, that is, after 7 days in average [CC 40].

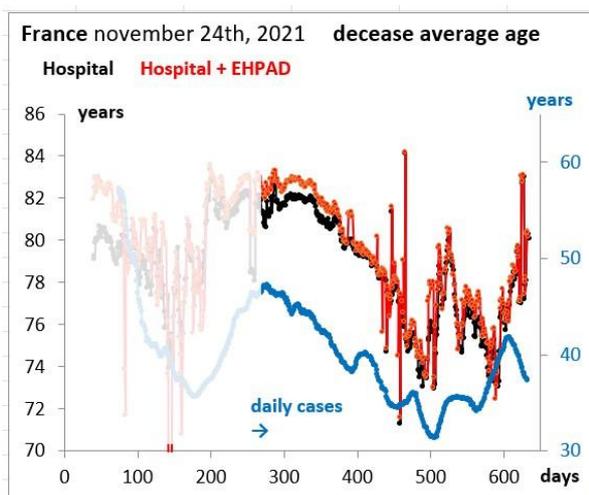
Such a model-free determination of  $R_{eff}(t)$  avoids the hazardous step of manual fitting. We used it to illustrate the various time shifts associated with the possible indicators, see Figure 16.



**Fig. 16.** An example of  $R_{eff}(t)$  variations determined from the CovidTracker method. The hospitalization peak is clearly delayed with respect to the case peak, by roughly 6-7 days. The decess peak - split - occurs obviously later on. Adapted from [CC 41]

### 3.1.2 The impact of age

Since the very beginning, it was observed that most of decesses were those of aged persons. The younger persons were thought to be almost free from severe forms, and even from contamination. This is no longer true, and presently the number of cases drastically increases in the young population. We have followed the evolution of the average age for the various steps of the disease, see figures 17, 18, adapted from [CC 39].



**Fig. 17.** Evolution of the average ages for contamination and decess. EHPAD stands for Hospitals for Aged or Dependent Person.

The decrease in the average decess age, between days 300-500 is assigned to the impact of the vaccination campaign targeted to the eldest population. The final increase may reveal a sizable weakening of the vaccinal protection, which the reason why a third vaccine dose was decided in most countries. Note that that the case curve also exhibited a strong increase (weakening of the vaccinal protection against contamination), before a final decrease which may reveal a recent change in the virus circulation, which now is extremely active in the younger population.

France november 24th, 2021 average ages hosp. data

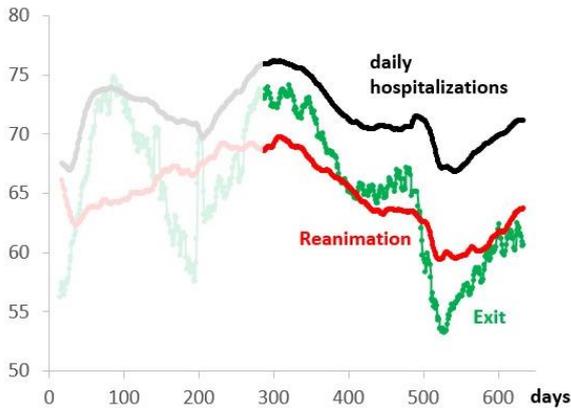


Fig. 18. Evolution of the average ages for hospitalization, admission in reanimation care service, and exit.

The recent increase, over the last two months, shows that the present (5<sup>th</sup>) wave again induces severe forms of the covid disease in the eldest population.

The different behaviours of the various age ranges are conveniently illustrated in the CovidTracker site, see Figure 19.

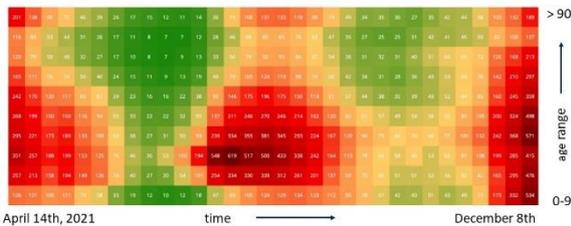
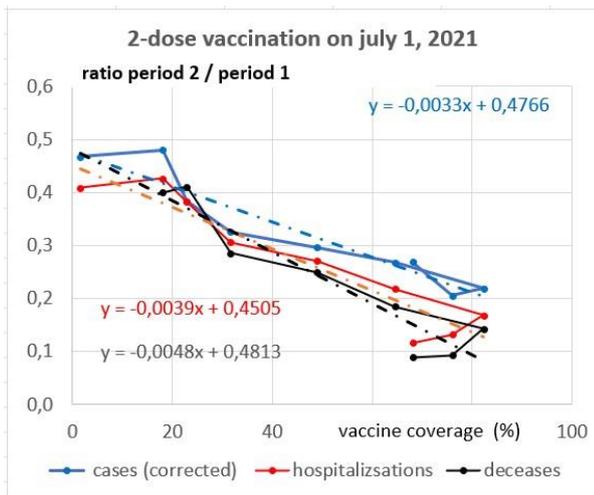


Fig. 19. Evolution of the incidence rate (number of new cases over the last 7 days per 100 thousand persons), as a function of time (horizontal scale), for the various age ranges (vertical scale). The colour scale spans from green to dark red, on increasing values of the incidence rate. Adapted from [39]

Figure 9 covers waves 3, 4 and 5, which are featured by dark areas. Remarkably wave 4 is peaked on the 20-29 range, while the other waves have a wider extension. The peculiar character of the 4<sup>th</sup> wave is interpreted as a consequence of the unachieved character of the vaccination campaign when wave 4 began. Details are given below. On the contrary wave 5 is widespread over all ages. The intense contamination of the 0–9-range is a true concern because this range is not yet allowed for vaccination in our country. It is worth mentioning that several countries already allowed it.

We have more precisely documented [CC 42] the effect of the vaccination coverage at the beginning of wave 4 by comparing the data of cases, hospitalizations and deceases, cumulated over two periods: period 1, before the beginning of the vaccination campaign, that is, before day 300, and period 2 associated with wave 4, that is, from day 488 to day 579. The ratios of the data period 2 / period 1 are plotted in figure 20 for each age range, as a function of the vaccine coverage on day 300. As the testing activity was extremely different in the various age ranges, we corrected the case ratio, by dividing it by the ratio of the total test numbers in each age range.



**Fig. 20.** The ratio of cumulated data over periods 1 and 2 (see Figure 19), plotted versus the vaccinal coverage of the various age ranges on the beginning of period 2, with the fit of their linear correlations. Adapted from [CC 42]

A clear dependence upon the age is obtained for all indicators, which suggests a linear correlation, except for the oldest two age ranges. A possible reason for the peculiar behaviour of the upper age tranches might be a reinforcement of the sanitary protection, with respect to wave 1; indeed, on the beginning of the pandemic, mortality was extremely high in the medical nursing homes (EHPAD), presumably because of an obvious lack of knowledge and experience in front of an unknown disease.

Whatsoever, the linear fits of the curves of Figure 20 leads to correlation slopes which shows that vaccination provides different degrees of protection against contamination (case data), severe forms (hospitalization data) and decease. In particular, the vaccination efficiency against contamination reveals to be approximately 25 % smaller than against severe forms and decease. Since efficacy against decease and severe forms is close to 90 % (sec. 1.3); against contamination it should be in the range 60-70 %. This value remains to be confirmed by real life data.

### 3.1.3 Summary of the main results

Important features were revealed by the present analysis which combined model improvements and a critical analysis of “experimental” data. The SIR model was corrected for a more sensible approach of the  $I \rightarrow R$  process. We recently completed it in order to account for vaccinal immunity against severe forms and decease, and independently against contagiousness. The application to the French data showed the reality of this immunity, at different protection degrees. We also showed that presently - on the manuscript submission date - the group immunity condition is far from being fulfilled.

We stress on the need for a critical analysis of the data (this holds any kind of experimental data). The most useful result of our investigation may be the recognition of the hidden part of the pandemic, leading to bias any analysis exclusively based on the case data (in other words, based on the incidence ratio). We preferred a decease-based analysis, casually completed by analyses based on hospitalisation data.

The main outputs of our calculations were the reproduction rates: the effective rate  $R_{\text{eff}}(t)$  characterizes the evolution of the dynamics while the “true”  $R_0(t)$  reveals the combined evolutions of the (average) individual behaviour and of the virulence of the prominent variant.

We pointed out the interest of additional information derived from raw data, for instance in the terms of “apparent severity” and “apparent lethality ratios”, we introduced.

A less intuitive comparison of different periods of the pandemic allowed drawing correlations between the impact of the pandemic and the degree of vaccinal coverage, on the occasion of wave 4 which occurred at a time when the coverage degree was strongly age-dependent.

We have provided figures which illustrate the successive waves of the pandemic. We did not pretend to explain their detailed features, obviously impacted by the various regulations (lock-in, curfew, travel restrictions, homework, closure of non-essential activities, sanitary pass ...). Such explanations require a detailed modelling which is only tractable by large institutions.

## 4 Conclusion

As reported in the previous section (3.1.3) the simple SIR mean-field model, corrected and completed for vaccinal immunity, has provided interesting results upon the average dynamics of the pandemic at the national level. However, we have used an extremely small part of the available data, which are worldwide and declined according multiple criteria.

We have often analysed the data of foreign countries and twice those of the age ranges in France [CC 22-23, CC 33]. In all cases each country or age range was treated independently from the others.

A generalized approach involving several countries or age ranges would require introducing the possible interactions between them. Obviously, the resolution of such a problem with input parameters in the form of a  $\beta_{ij}(t)$  matrix is not tractable by the means of Excel spreadsheets. *This consequently is another story.*

As for us, modesty is required. We intend to continue the current chronicle, accessible as usual, as long as the COVID-19 disease is not mastered. Presently, the growth rhythm of wave 5 seems to slow down. However, the Omicron variant transmissibility, lethality, and vaccines response are not well characterised.

Acknowledgements are due to our friends: Henri Lorain and Professor Didier Astruc, for their stimulating discussions and advices.

## References

1. C Drosten, S Günther, W Preiser, et al. *N Engl J Med.* **348**, 1967–1976 (2003).
2. TG Ksiazek, D Erdman, CS Goldsmith, et al. *N Engl J Med.* **348**, 1953–1966 (2003).
3. AM Zaki, S van Boheemen, TM Bestebroer, et al. *N Engl J Med.* **367**, 1814–1820 (2012).
4. N Zhu, D Zhang, W Wang, Li X., et al. *N Engl J Med.* **382**, 727–733 (2020).
5. C Huang, Y Wang, MD, X Li, et al. *Lancet* **395** (10223), 497–506 (2020).
6. P Zhou, XL Yang, XG Wang, et al. *Nature* **579** (7798), 270–273 (2020).
7. B Kan, M Wang, H Jing, et al. *J. Virol.* **79**, 11892–11900 (2005).
8. BL Haagmans, SH Al Dhahiry, CB Reusken, et al. *Lancet Infect. Dis.* **14**, 140–145 (2014).
9. CC Wang, KA Prather, J Sznitman, et al. *Science* **373** (6558), eabd9149 (2021).
10. JFW Chan, KH Kok, Z Zhu, et al. *Emerg Microbes Infect.* **9**, 221–236 (2020).
11. MA Tortorici & D Veeler. *Adv. Virus Res.* **105**, 93–116 (2019).
12. R Yan, Y Zhang, Y Li, et al. *Science* **367**, 1444–1448 (2020).
13. AC Walls, YJ Park, MA Tortorici, et al. *Cell* **180**, 181–292 (2020).
14. Y Finkel, A Gluck, A Nachshon, et al. *Nature* **594** (7862), 240–245 (2021).
15. K Zang, L Miorin, T Makio, et al. *Sci Adv.* **7** (6), eabe7386 (2021).
16. M Scudellari. *Nature* **595**(7869):640–644. (2021).
17. L Braga, H Ali, I Secco, et al. *Nature* **594** (7861), 88–93 (2021).
18. MZ Tay, CM Poh, L Rénia, et al. *Nat Rev Immunol.* **20**, 363–374 (2020).
19. M O'Driscoll, GRD Santos, L Wang, et al. *Nature* **590**, 140–145 (2021).
20. P. Brodin. *Nat. Med.* **27**, 28–33 (2021).
21. Q Zhang, P Bastard, Z Liu, et al. *Science* **370** (6515), eabd4570. (2020)
22. P Bastard, LB Rosent, Q Zhang, et al. *Science* **370** (6515), eabd4585. (2020)
23. P Bastard, A Gervais, T Le Voyer, et al. *Sci Immunol.* **6** (62), eabl4340. (2021)
24. T Asano, B Boisson, F Onodi, et al. *Sci Immunol.* **6**(62): eabl4348. (2021)
25. E Boehm, I Kronig, RA Neher, et al. *Clin Microbiol Infect.* **27** (8), 1109–1117 (2021).
26. SM Soh, Y Kim, C Kim, et al. *J Microbiol.* **59**, (9), 807–818 (2021).
27. E Janik, M Niemcewicz, M Podogrocki, et al. *Pathogens* **210** (6), 633 (2021).
28. A Kumar, R Parashar, S Kumar, et al. *J Med Virol.* Epub ahead of print (2021). doi: 10.1002/jmv.27467.
29. J Chen, R Wang, NB Gilby, et al. *ArXiv.* 2112.01318v1. Preprint (2021).
30. SJ Gao, H Guo, and G Luo. *J Med Virol.* Online ahead of print (2021). doi: 10.1002/jmv.27491
31. E Callaway. *Nature.* **600** (7887), 21 (2021).
32. M Levine-Tiefenbrun, I Yelin, R Katz. et al. *Nat. Med.* **27**, 198–199 (2021).
33. JS Tregoning, KE Flight, SL Higham, et al. *Nat Rev Immunol.* **21** (10), 626–636 (2021).
34. A Puranik, PJ Lenehan, E Silvert et al. medRxiv. Preprint (2021). doi:10.1101/2021.08.06.21261707
35. HH Ibkel. *Rev Med Virol* e2313. Online ahead of print (2021). doi: 10.1002/rmv.2313.
36. <https://www.data.gouv.fr/fr/organizations/sante-publique-france/>
37. <https://www.eficiens.com/coronavirus-statistiques/>

38. <https://ourworldindata.org/coronavirus>
39. <https://covidtracker.fr/covidtracker-france>
40. [Visualisation Covid-19 \(germain-forestier.info\)](#)
41. <https://www.statnews.com/2020/07/20/trump-said-more-covid19-testing-creates-more-cases-we-did-the-math/>
42. French platform [MODCOV19 \(cnrs.fr\)](#)
43. Montpellier University « [Avec le Covid-19, la modélisation des épidémies s'ouvre aux données massives](#) » | [Pour la Science; Modélisation de l'épidémie de COVID-19 \(ouvaton.org\)](#)
44. Pasteur Institute : [Modélisation COVID-19 - Modélisation COVID-19 \(pasteur.fr\)](#)
45. « [Avec le Covid-19, la modélisation des épidémies s'ouvre aux données massives](#) » | [Pour 373, 889–895 \(2021\)](#)
46. N. Hozé, J. Paireau, N. Lapidus, et al, [Lancet Public Health 6 e408-15 \(2021\)](#)
47. W.O. Kermack, A. McKendrick, G.T. Walker, [Proc. R. Soc. A 115, 700–721 \(1927\)](#)
48. <https://images.math.cnrs.fr/Modelisation-d-une-epidemie.html>
49. K. M. Gostic, L. McGough, E. B. Baskerville, et al, [PLOS Comput. Biol. 16 \(12\), e1008409. \(2020\)](#)
50. A. Cori\*, N. M. Ferguson, C. Fraser, et al, [Am. J. Epidemiol 178 1505–1512 \(2013\)](#)
51. [Press Review, Pasteur Institute, April 2020](#)
52. H. Salje, C. Tran Kiem, N. Lefrancq, et al, [Science 369, 6500 \(2021\)](#)
53. M. O'Driscoll, G. R. Dos Santos, Lin Wang, et al, [Nature 590, 140–145 \(2021\)](#)
54. [Covid-19 l'erreur de 1927 | CORONA-CIRCULE](#)
55. <https://www.lajauneetlarouge.com/covid-19-interrogations-sur-le-modele-epidemiologique-prise-en-compte-de-la-vaccination-et-du-variant-anglais/>
56. P. Bosetti, C. Tran Kiem, A. Andronico, et al, preprint (2021) <https://hal-pasteur.archives-ouvertes.fr/pasteur-03272638/document>

# Bibliographie

## Documents de base

1 - « Épidémie de COVID-19 – Point sur la modélisation épidémiologique pour estimer l’ampleur et le devenir de l’épidémie de COVID-19 », note présentée par Cédric Villani le 30 avril 2020 à l’Office parlementaire d’évaluation des choix scientifiques et technologiques. <https://www2.assemblee-nationale.fr/15/les-delegations-comite-et-office-parlementaire/office-parlementaire-d-evaluation-des-choix-scientifiques-et-technologiques/secretariat/a-la-une/modelisation-epidemiologique-pour-estimer-l-ampleur-et-le-devenir-de-l-epidemie-de-covid-19>

2 – « A Contribution to the Mathematical Theory of Epidemics », note de W. O. Kermack et A. G. McKendrick, pages 700 à 721 d’un document transmis en mai 1927 à la Royal Society de Londres, contenant les équations enseignées dans pratiquement tous les cours d’épidémiologie du monde sans mentionner le fait que les auteurs précisent bien dans leur note que ces équations simplificatrices ne s’appliquent qu’à un « special case » correspondant à des périodes (peu intéressantes ...) de l’épidémie où le taux quotidien des fins d’infections chez les malades (par guérison ou décès) est constant (« constant rates ») et ne dépend donc pas, comme dans les phases dynamiques, de l’ancienneté de leur infection.  
<https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118>

## Cours de modélisation

3 – « Modèle mathématique des épidémies » cours d’Etienne Pradoux, Professeur de mathématiques (université d’Aix-Marseille)  
<https://www.youtube.com/watch?v=Z27HG2dtgck>

4 – « *The SIR model and the Foundations of Public Health* »  
Howard Weiss (Département de mathématiques, Georgia Institute of Technology, Atlanta)

5 - « Modèles épidémiologiques » par Suzanne TOUZEAU (cours de l’Agro ParisTech)  
<https://acoustique.ec-lyon.fr/chaos/INRA-Touzeau10.pdf>

## « Complexifier pour affiner » ?

(cf note de C. Villani à l’Office parlementaire d’évaluation des choix scientifiques et technologiques)

6 - [https://www.yvespeyson.fr/coronavirus/memoire\\_Jego\\_Falconet.pdf](https://www.yvespeyson.fr/coronavirus/memoire_Jego_Falconet.pdf)

ou

[https://scholar.google.com/citations?view\\_op=view\\_citation&hl=en&user=jGbzMV0AAAAJ&citation\\_for\\_view=jGbzMV0AAAAJ:Y0pCki6q\\_DkC](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=jGbzMV0AAAAJ&citation_for_view=jGbzMV0AAAAJ:Y0pCki6q_DkC)

Modéliser la progression d’une épidémie

Hugo Falconet et Antoine Jego sous la direction d’Amandine Veber et Vincent Calvez

**Bel exemple de « complexification » absurde :** des pages et des pages exposant une utilisation des

chaînes de Markov dont le principe de base général est que l'évolution d'un processus à partir d'un instant T ne dépend que de la situation à l'instant T et absolument pas de ce qui s'est passé avant cet instant T. Or pendant une épidémie le moment de la guérison ou du décès d'un individu dépendent de la date à laquelle il a été infecté ...

## Articles de la revue des anciens de l'Ecole polytechnique

7 – Articles de François Xavier Martin parus dans « La Jaune et la Rouge » revue mensuelle de l'Association des anciens élèves et diplômés de l'Ecole polytechnique, dont en particulier :

<https://www.lajauneetlarouge.com/covid-19-une-modelisation-simple-utilisant-excel-accessible-aux-non-mathematically-et-pleine-denseignements/> (octobre 2020)

[https://www.lajauneetlarouge.com/deuxieme-vague-covid-19-perspectives-2020-2021\\_model\\_sir/](https://www.lajauneetlarouge.com/deuxieme-vague-covid-19-perspectives-2020-2021_model_sir/) (novembre 2020)

<https://www.lajauneetlarouge.com/covid-19-interrogations-sur-le-modele-epidemiologique-prise-en-compte-de-la-vaccination-et-du-variant-anglais/> (février 2021)

<https://www.lajauneetlarouge.com/covid-19-enseignements-et-credibilite-des-modeles-epidemiologiques/> (avril 2021)

<https://www.lajauneetlarouge.com/la-modelisation-des-epidemies-covid-19/> (mai 2021)  
CONFERENCE SUR YOUTUBE

<https://www.lajauneetlarouge.com/covid-19-perspectives-dinfection-par-le-variant-omicron/>  
(janvier 2022, écrit en commun avec François Varret)

## Communication collective de presque toute la fine fleur de la modélisation épidémiologique française (juin 2021)

8 - « Epidemiology and control of SARS-CoV-2 epidemics in partially vaccinated populations: a modeling study applied to France »

<https://hal-pasteur.archives-ouvertes.fr/pasteur-03272638/document>

Paolo Bosetti, Cécile Tran Kiem, Alessio Andronico, Vittoria Colizza, Yazdan Yazdanpanah, Arnaud Fontanet, Daniel Benamouzig et Simon Cauchemez

**Où l'on découvre page 2 du «Supplément» de l'édition du 28 juin 2021 l'utilisation d'équations de type « 1927» pour simuler les fins d'infections. Or nous avons montré qu'elle minimisait fortement les variations du flux quotidien de nouveaux infectés**

**On notera que cette note n'est pas co-signée par l'équipe de l'Université de Montpellier (Samuel Alizon – Mircea Sofonea) qui emploie un modèle discret et qui a très aimablement accepté de répondre positivement et de façon détaillée à nos remarques sur l'utilisation abusive des équations différentielles de 1927 par un grand nombre de modélisateurs.**

## **Site « Corona circule » de Mathilde et François Varret**

8 – Site sur l'épidémie de Covid 19 de François Varret, Ancien élève de l'ENS Ulm, Professeur émérite de physique (université de Versailles Saint-Quentin) et de Mathilde Varret, biologiste, généticienne (INSERM)

<https://corona-circule.github.io/lettres/>

## **Communication commune de F. Varret, M. Varret et F.X. Martin**

9 - « A tentative tracking of the SARS-Cov2 pandemic in France, based on a corrected SIR model including vaccination effects ». Communication de Mathilde Varret, François Varret et François Xavier Martin faite en avril 2022 aux 8<sup>th</sup> Complexity-Disorder Days de l'European Physical Journal (EPJ)

[https://epjwoc.epj.org/articles/epjconf/abs/2022/07/epjconf\\_complexity-disorder2021\\_01002/epjconf\\_complexity-disorder2021\\_01002.html](https://epjwoc.epj.org/articles/epjconf/abs/2022/07/epjconf_complexity-disorder2021_01002/epjconf_complexity-disorder2021_01002.html)

## **Article de synthèse du « Berliner Zeitung »**

10 - « The great corona modelling controversy. One of Germany's most respected mathematicians has discovered a systemic bias in pandemic forecasts ».

Article très exhaustif de Tobias Haberkorn dans le « Berliner Zeitung »

<https://www.berliner-zeitung.de/en/the-great-corona-modelling-controversy-were-contact-restrictions-overrated-li.171039>