

La modélisation mathématique de propagation et de prédiction de l'épidémie Coronavirus Covid-19 au Maroc

Par Abdeljabar RAFIKI¹

Résumé :

La caractéristique biologique de l'agent pathogène et les données de l'environnement constituent un processus essentiel dans le mode de transmission de virus et joue un rôle crucial dans le choix des modèles épidémiologiques parmi la classe des modèles disponibles. Cet article vise à modéliser la propagation de Coronavirus Covid-19 au Maroc dans la perspective de prédire son jour de pic et son arrêt définitif. A travers des analyses des courbes épidémiques, des simulations du jeu de données de la santé publique par le modèle mathématique S.I.R, et le modèle statistique de la fonction logistique nous sommes arrivés à des conclusions intéressantes. D'abord, le confinement précoce a réduit progressivement le taux de reproduction de base actuellement jusqu'à moins de « 1 ». Ensuite, le mode de transmission est de type « épidémie à propagation » ceci risque de générer une nouvelle vague. Enfin, les deux démarches de modélisation utilisées ont montré que le pic épidémique est atteint entre 20 et 21 avril 2020 et l'arrêt définitif de la crise est prévu à la fin de la 1^{ère} semaine de juin 2020 pour le modèle SIR et à la fin de la dernière semaine du même mois avec un effectif autour de 5000 cas confirmés. La différence des résultats des deux modèles n'est ni surprenante ni trop significative pour qu'elle soit inquiétante, elle est imputée aux bases des calculs différentes et aux hypothèses différentes des deux modèles. La modélisation par S.I.R est plus robuste que celle de la fonction logistique.

Mots-clés : Courbes épidémiques, taux de transmission, taux de reproduction de base et effectif, temps de génération, le pic épidémique, Modèle S.I.R, fonction logistique, arrêt de l'épidémie

¹ Professeur d'économétrie et de statistique à la Faculté des Sciences Juridiques Economiques et Sociales- Université Moulay Ismil- Meknès.

Introduction

Depuis l'apparition de Coronavirus Covid-19 pour la première fois en Chine, les recherches sur la modélisation de propagation et de transmission de virus ne cessent de se développer dans le monde entier sans être convergentes. Malgré la robustesse de ces modèles mathématiques épidémiologiques il n'en demeure pas moins que leurs applications à Covid-19 souffrent des imperfections liées à la nature de l'agent pathogène qui lui-même reste mal connu jusqu'à présent. Il s'agit par exemple des sujets asymptomatiques infectieux non identifiables, de la période d'infection et d'incubation variable d'un sujet à l'autre et du temps de génération mal connu. Ainsi, la méconnaissance de l'estimation précise du taux de reproduction et du taux de transmission de l'infection rendent les résultats de la modélisation mathématique imprécis.

Dans cette modeste contribution, nous allons élaborer une étude propre au contexte marocain dans la perspective de répondre à une question cruciale que se pose tout le monde, à savoir « *Combien de temps pourrait durer l'épidémie causée par Coronavirus ?* ». Pour pouvoir apporter des éléments de réponse cohérents avec la réalité marocaine on devrait choisir le modèle mathématique hypothétique le mieux adapté à la situation marocaine. Pour ce faire on opte pour le modèle S.I.R qui se voit capable de décrire la propagation de virus telle qu'elle est observée dans notre contexte depuis le début de la crise. Ainsi, le modèle mathématique S.I.R qui tient compte à la fois des caractéristiques propre à Covid-19 et des données de l'environnement externe à l'agent pathogène. Dans la construction de cette modélisation nous allons prendre en considération l'hypothèse de confinement précoce qu'avait adopté le Maroc, l'apprentissage par l'expérience des citoyens, la distanciation, l'hygiène et la protection par des masques. Ces procédures ont réduit actuellement la taille de la population des susceptibles d'être infectés de façon considérables et ont fait manquer au virus de ressources pour se propager dans de nombreuses régions du Maroc qui se voient épargnés jusqu'à nos jours.

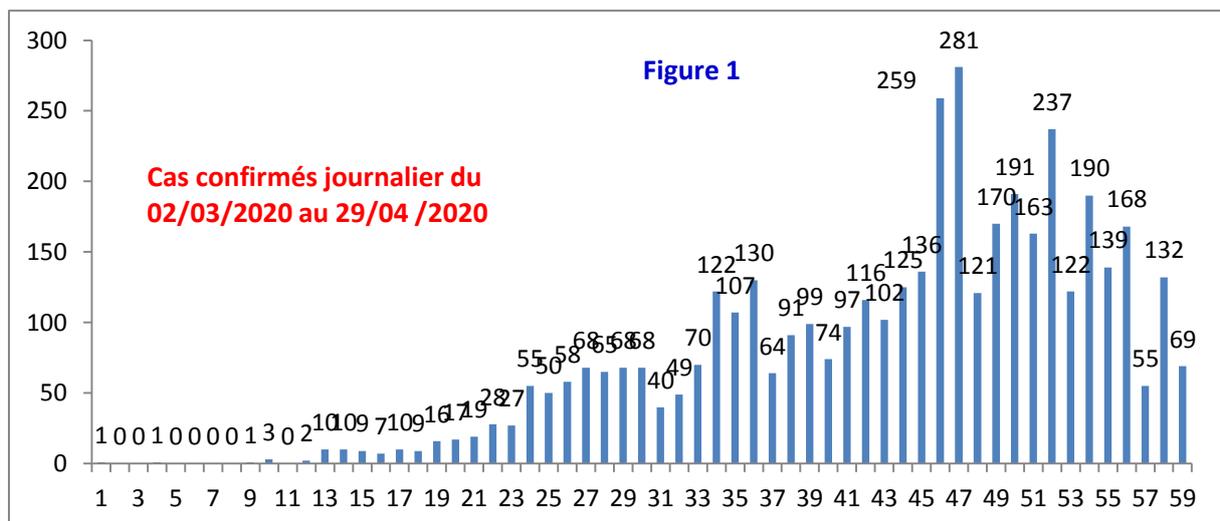
Pour faire face à ses hypothèses restrictives du modèle S.I.R nous avons enrichi notre démarche de modélisation par recours à l'estimation du modèle logistique qui s'ajuste parfaitement à l'évolution type des données épidémiques observées. Par ailleurs, en raison du changement des paramètres des deux modèles au jour le jour, on ne peut en aucun cas admettre l'étalement de cette prédiction à long terme, elle est considérée effectuée à l'instant « T » pour le court terme. Dans les développements qui suivent nous présentons successivement l'analyse des courbes épidémiques, la prédiction de la propagation de virus par le modèle S.I.R et par le modèle de la fonction logistique.

1-Analyse des courbes épidémiques

1.1 Courbes des cas infectés journaliers (incidences)

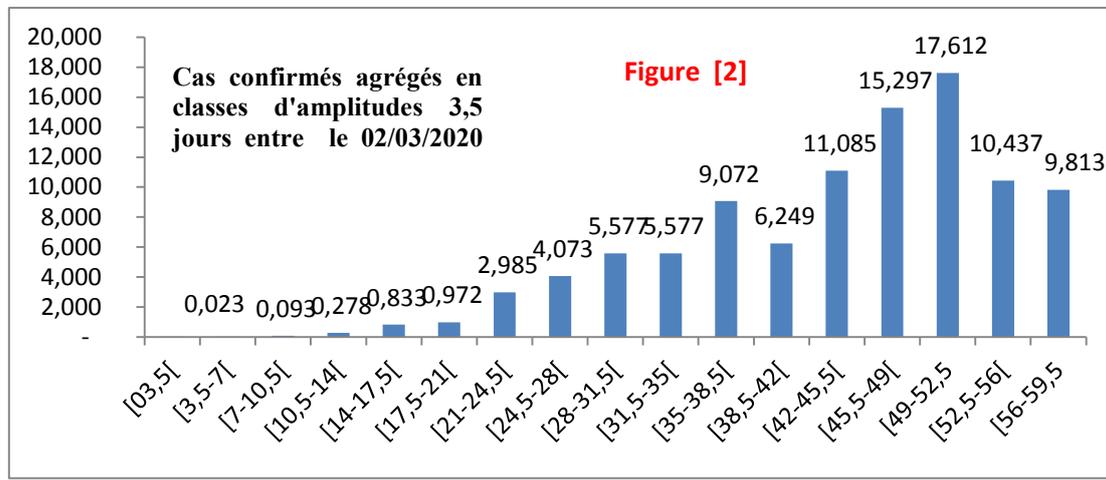
La courbe épidémique de l'évolution des cas confirmés positifs en fonction du temps [figure 1] permet de visualiser plusieurs informations liées à la propagation de l'infection par coronavirus Covid-19 dans notre pays. Nous nous limitons ici aux seules informations d'aide à la prédiction. Il s'agit notamment :

- De plusieurs pic séparés par des périodes d'incubation qui ne sont pas les mêmes pour tous les patients. D'après l'estimation faite par le modèle économétrique d'autorégressif, cette période est en moyenne de 6 jours.
- La présence de certains pics nettement distinctifs par rapport aux autres ne renvoie en aucun cas à la gravité de la chose. Ceci s'explique par deux choses : l'une est liée à la multiplication des laboratoires qui a fait apparaître des cas confirmés avant même l'apparition des symptômes et l'autre revient au fait que la déclaration des cas confirmés est rapportée avec retard aux jours où ils doivent lui être imputés.
- L'instabilité des longueurs des bâtons renvoie à l'infection interhumaine par groupe ami et/ou famille de type épidémie par propagation.
- Le bâton le plus long (47^{ème} jour) ne correspond pas exactement au jour de pic de l'épidémie mais, il a contribué à la formation du pic épidémique théorique que l'on estimera dans les sections qui suivent.



Normalement avant l'arrêt définitif de l'épidémie, le graphique de la figure [1] ne peut en rien se prononcer sur le pic épidémique. Cependant pour obtenir une courbe épidémique plus informative sur cette question nous regroupons la distribution des cas confirmés

journaliers en classe de même amplitude, on prenant comme règle empirique dans les maladies infectieuses une unité de temps d'une durée égale au quart de la période d'incubation, soit une amplitude de 3,5 jours puis nous calculons la fréquence ou la densité de chaque classe. Selon cette règle de manipulation simple nous obtenons l'histogramme de la figure [2] suivante :



Si la situation de confinement, de distanciation, de prudence, de la crainte de virus, de l'hygiène de contact physique etc. est maintenue, il ne reste qu'à attendre l'arrêt de l'épidémie puisque nous sommes déjà vécus son pic. Il correspond à la hauteur du rectangle [figure 2] la plus élevée, soit une fréquence de 17,6125%. Le pic est lu sur l'axe des temps, il se situe entre la 49^{ème} et la 52,5^{ème} journée depuis la déclaration du premier cas positif au début mars 2020, soit le 20 ou le 21 Avril. L'ajustement de la courbe d'incidence cumulée par la fonction logistique et le modèle mathématique de prédiction des épidémies **S.I.R** comme on va le voir ci-après ont montré que le pic épidémique est atteint le 50^{ème} jour, soit le **20/Avril/2020**. Au contraire, on devrait être très prudent à ces conclusions, ***toute désintégration de confinement et de prudence va entrainer une nouvelle vague de de contamination et de propagation puisque déjà la figure [1] montre que nous sommes en présence d'une épidémie de type propagation.***

2.1. La courbe du nombre de cas cumulés

L'épidémie causée par coronavirus covid-19 qui a été déclenchée au Maroc le début du mois de mars va disparaître très prochainement. Les signes, les indices et les calculs mathématiques sont autant d'arguments justificatifs d'une telle conclusion sous l'hypothèse du respect des orientations des autorités et le maintien de confinement et d'hygiènes. A ce point en

particulier, nous sommes reconnaissants de l'effort hautement considérable qu'avait fourni l'Etat marocain pour la lutte contre la propagation de coronavirus Covid-19.

La courbe d'incidence cumulé (Figure [3]) retrace l'évolution quotidienne des cas cumulés depuis le début mars 2020 jusqu'au 29 avril 2020. D'après l'ajustement de cette courbe par une équation exponentielle, elle se démarque par une croissance exponentielle de son début jusqu'au 20 Avril. Dans cette phase, le facteur de croissance ou le nombre de cas d'un jour (j) semble situé entre 1,15 et 2.25 fois le nombre de cas du jour précédent (j-1). **Mais, à partir de cette date jusqu'au 28 Avril, ce coefficient est en décroissance continue, il est asymptotique à la valeur 1 en faisant 1.02 le jour de la rédaction de cet article.** Ce résultat fort intéressant indique qu'on a franchi la frontière exponentielle et on se situe sur la phase linéaire (Figure [3]). Cette zone de passage entre les deux frontières est séparée par un point appelé point d'inflexion, il correspond au pic épidémique.

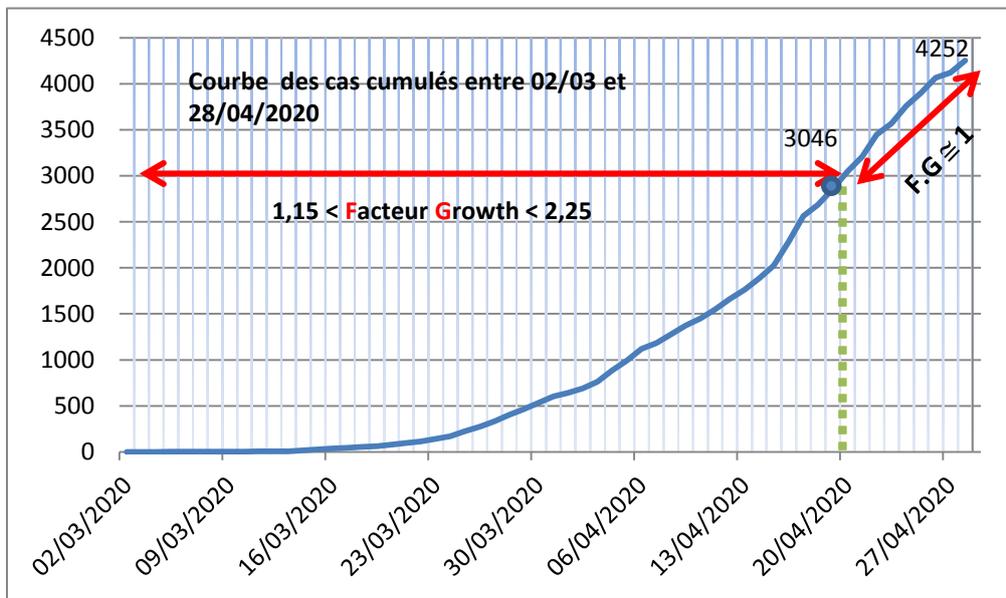
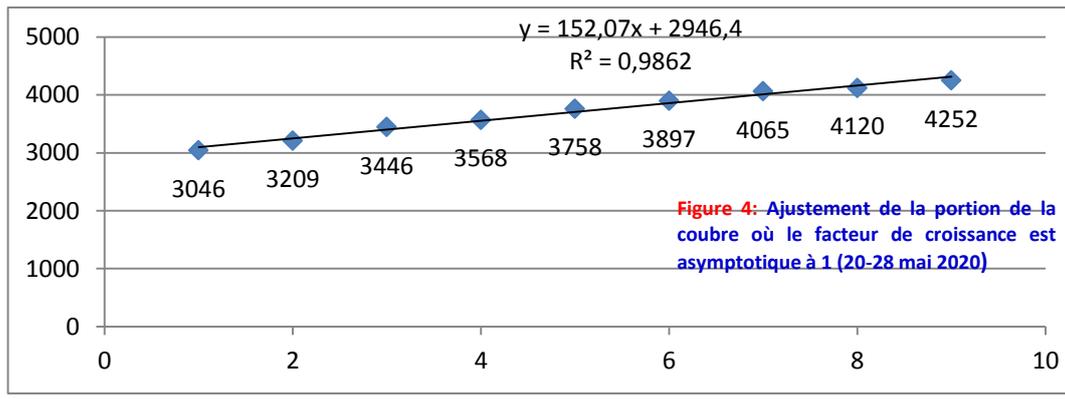


Figure [3]

Lorsque le facteur de croissance est inférieur ou asymptotique à 1, alors on se situe sur la phase linéaire de la courbe. L'ajustement sur le temps de la portion de la courbe d'incidence cumulée au-delà de 20 avril (**figure [3]**) va bien confirmer ce résultat. En effet, le nuage de point sur cette portion est presque aligné donc on peut l'ajuster facilement par l'équation $\bar{C}_m = \hat{\beta}_0 + \hat{\beta}_1 * t$ comme l'illustre la **figure [4]** suivante :



La portion linéaire de la courbe nous indique que le nombre de cas confirmés positifs par jour fluctuera autour de 100 cas jusqu'au 4 ou 5 mai au plus tard. Après cette date le nombre de cas va décroître en-deçà du 100 sans jamais l'atteindre².

2-Prédiction de l'épidémie Covid-19 au Maroc par le modèle mathématique S.I.R

La modèles mathématique de la prédiction de l'épidémie causée par coronavirus doit tenir compte non seulement de ses données historiques des cas confirmés mais aussi des hypothèses propres au virus telles qu'elles sont évoquées par la communauté scientifiques et des hypothèses liées à l'environnement des susceptibles. Trois classes de modèles mathématiques sont proposés dans le cas d'épidémie : SIS, SIR et SEIR. Le choix d'un modèle plus tôt qu'un autre dépend des hypothèses faites sur le coronavirus Covid-19 à savoir :

- La période d'incubation est normalement de 14 jours, mais elle est rapportée à une moyenne de 6 jours ;
- La période d'infection ne correspond pas toujours à la période d'incubation ;
- Une personne guérie devient immunisée et ne peut ni attraper la maladie une deuxième fois ou la transmettre à d'autres individus
- Il y a des sujets symptomatiques et d'autres asymptomatiques mais infectieux ;
- Le virus se transmet d'une personne à une autre par un simple contact ou échange d'air ;

Pour que la modélisation soit fiable, on devrait introduire dans le modèle d'autres hypothèses qui ne sont pas liées aux caractéristiques de l'agent infectieux et qui peuvent modifier la taille des susceptibles, telles que le confinement, les décès pour des raisons autres que le coronavirus et les nouveau-nés au cours de l'épidémie. Dans notre cas de figure on a considéré que les décès égalisent les nouveau-nés et les infectés non identifiés ne sont pas pris

² Ces chiffres sont obtenus par l'ajustement de la courbe cumulée par l'équation logistique

en considération en l'absence des données. En conséquence, on a retenu le modèle S.I.R³ qui a été développé pour la première fois dans l'étude de la propagation de l'épidémie de la peste par Kermack et McKendrick (1927).

En effet, le modèle SIR consiste à diviser la population des citoyens en trois compartiments différents : Les susceptibles (S), les infectées (S) et les rétablies (R). Les susceptibles sont des personnes sains mais peuvent devenir malades à cause de l'infection par contact. Le 2^{ème} compartiment est celui des personnes infectées. Les personnes infectées peuvent quitter ce compartiment une fois guéries ou décès en se plaçant dans le 3^{ème} compartiment appelé rétablies. Le Modèle SIR peut modéliser les trois situations en tenant compte des hypothèses de Coronavirus Covid-19 ci-dessus. Chaque personne susceptible peut devenir malade, soit actuellement infecté et dans ce cas peut infecter d'autre personnes. Une fois guérie elle est immunisée et ne peut ni attraper la maladie ni la transmettre à quelqu'un d'autre. Le schéma suivant illustre les flux d'entrées-sorties unidirectionnels entre les trois compartiments :



Ces flux d'entrées-sorties sont des fonctions du temps, ils peuvent être décrits sous forme d'équations formant ainsi le modèle S.I.R. il est évident de préciser au départ qu'à chaque instant « t » on a : $S(t) + I(t) + R(t) = N$, soit la population totale.

Il est intéressant de définir les conditions initiales du modèle ce qui revient à connaître la valeur de ces fonctions à $t=0$ c'est-à-dire au tout début. L'idée sous-jacente est qu'au début il n'y a aucune personne guérie $R(0)=0$ mais il y probablement un petit nombre d'infectés ou au moins un seul patient⁴ soit, $I(0)=I_0$. En revanche le $S(0)= S_0$ correspond au nombre de personnes susceptibles à $t=0$. Au cours du temps les trois fonctions évoluent suivant à taux d'accroissement soit négatif soit positif. Dès que les personnes saines entre en contact avec les personnes infectées, elles vont créer une certaine interaction en multipliant probablement l'infection. Dans ce cas-là les susceptibles infectés quitteront le compartiment sain et entrent dans le compartiment infecté, en conséquence l'effectif des sains diminue du même nombre d'individus nouvellement infectés. Ainsi, le nombre d'individus infectieux augmente avec

³ (S) Sains, (I) infectés, et (R) Rétablis

⁴ Le patient zéro correspond relativement à un petit nombre ou un à une seule personne infectée à la date $t=0$, noté $I(0)= I_0$

ceux nouvellement infectés et diminue avec ceux retirés en formant le compartiment infecté. Enfin, toutes les individus guéris ou décédés quittent le compartiment infectés et entre dans le compartiment rétablie. La transmission de la maladie s'effectue à un taux (β) et les infectés passe aux rétablis à un taux (ν), soit mathématique l'évolution des fonctions au fil du temps s'écrivent comme suit :

$$\begin{cases} \frac{dS(t)}{dt} = -\beta * S(t) * I(t) \\ \frac{dI(t)}{dt} = \beta * S(t) * I(t) - \nu * I(t) \\ \frac{dR(t)}{dt} = \nu * I(t) \end{cases} \quad \text{Avec :} \quad \begin{cases} S(t) + I(t) + R(t) = N \\ S(0) = S_0; I(0) = I_0; R(0) = 0 \\ \frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} = 0 \end{cases}$$

Pour répondre à la question de présence ou d'absence d'épidémie, nous allons prendre l'équation de la variation des infectés $d(I)/dt$ et on fixe le temps à $t=0$, soit au début de l'infection, soit : $\frac{dI(t)}{d(t)/t=0} = \beta * S_0 * I_0 - \nu I_0$. Le signe de ce rapport renvoie à la présence de

l'épidémie ou de son absence.

$$\begin{aligned} \text{En effet, si } \beta * S_0 * I_0 - \nu I_0 < 0 &\Rightarrow I_0(\beta * S_0 - \nu) < 0 \\ &\Rightarrow (\beta * S_0 - \nu) < 0 \\ &\Rightarrow \frac{\beta * S_0}{\nu} - \frac{\nu}{\nu} < \frac{0}{\nu} = 0 \\ &\Rightarrow \frac{\beta * S_0}{\nu} < 1. \text{ Dans ce cas il n'y a plus d'épidémie} \end{aligned}$$

On pose $\frac{\beta * S_0}{\nu} = R_0$. Il s'agit du taux de reproduction de base. Il représente le nombre de cas secondaires directement infectés par un sujet contagieux dans une population totalement susceptible en début d'émergence ou en début d'épidémie.

Si le R_0 est inférieur à 1 alors l'épidémie s'éteint sinon elle se propage⁵.

Lorsque $R_0 < 1$, au voisinage de $t=0$, on aura

$$\frac{dI(t)}{d(t)} = \beta * S * I - \nu I \Rightarrow d(I) = I(\beta * S_0 - \nu) dt$$

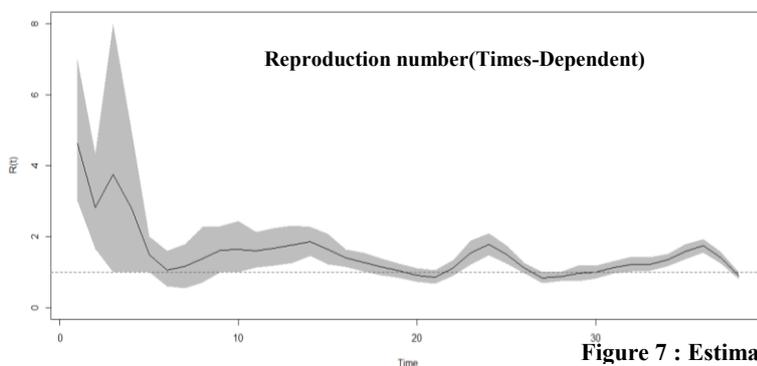
⁵ Pour baisser le R_0 en cas d'épidémie il faut baisser le taux de transmission de la propagation de l'infection (β) au moyen de confinement, l'hygiène, la protection et l'apprentissage par l'expérience. C'est pour cette raison que les pays recourent au confinement.

La solution de cette équation différentielle est $I(t) = \text{Exp}(\beta * S_0 - \nu) * t$. cette équation montre bien qu'en cas de début de l'épidémie la fonction des infectés augmente de façon exponentielle ce qui a nécessité une intervention urgente par le confinement (pour l'illustration, voir la section de la prédiction par la fonction logistique ci-après).

La question qui se pose est en quel jour le pic sera atteint ? Et quand est-ce que l'épidémie s'arrêtera ? Pour répondre à ces questions on devrait chercher la solution du programme S.I.R ci-dessus à chaque instant « t ». Pour ce faire, on va initialiser le processus à l'instant $t=0$ en fixant la population susceptible $S(0)$ le premier individu infecté $I(0)$. La solution elle-même dépend du taux de transmission (β) inconnu qui lui-même dépend de R_0 . De ce fait on estime d'abord R_0 par quatre méthodes en utilisant le logiciel « R » puis on présente les solutions obtenues par le même logiciel.

- **Estimation du taux de reproduction R_0 .**

On estime en même temps le taux de reproduction de base R_0 et le taux de reproduction effectif $R_0(t)$ en utilisant les méthodes : Exponential growth [Wallinga & Lipsitch(2007)], Maximum likelihood estimation [White & Pagano(2008)], Sequential bayesian method (SB) et Estimation of time dependent reproduction numbers [Wallinga & Teunis) (2004)]. Les résultats obtenus sont les mêmes aussi bien pour la méthode **ML($R_0=1,29$)** que la méthode **EG ($R_0=1,28$)**. Mais ils diffèrent pour les autres. Le R_0 le plus significatif reste celui qui tient compte de la phase où l'exponentielle est accentuée au début de l'épidémie. Nous reprenons ici les résultats de la méthode Estimation of time dependent reproduction numbers (TD) puisque celle de séquentielle Bayésienne est sensible aux valeurs nulles constatées au début de l'épidémie.



Le graphique montre l'évolution de R_0 depuis le 2 mars sur un temps de génération distribué suivant Gamma de moyenne 2,6 et de dispersion 1. Au départ de l'épidémie le R_0 était autour de 4 mais il est au-dessous de 1 actuellement. Cette valeur signifie que coronavirus Covid-19 en voie de disparition.

Figure 7 : Estimation de R_0 et de $R_0(t)$

Puisque la méthode du maximum de vraisemblance et celle de la croissance exponentielle ont donné la presque la même valeur de R_0 (1.28) alors on a retenu cette dernière pour calculer le taux de transmission (β) et la résolution des équations différentielles du modèle S.I.R. En

supposant une taille des susceptibles réduite au faite de la diminution du taux de reproduction effectif $R_0(t)$ au cours du temps, le confinement, l'hygiène et l'apprentissage par expérience. La solution par le logiciel R-Project des équations différentielles du modèle S.I.R à ces résultats qui schématisent le mode de propagation du virus Covid-19 au Maroc.

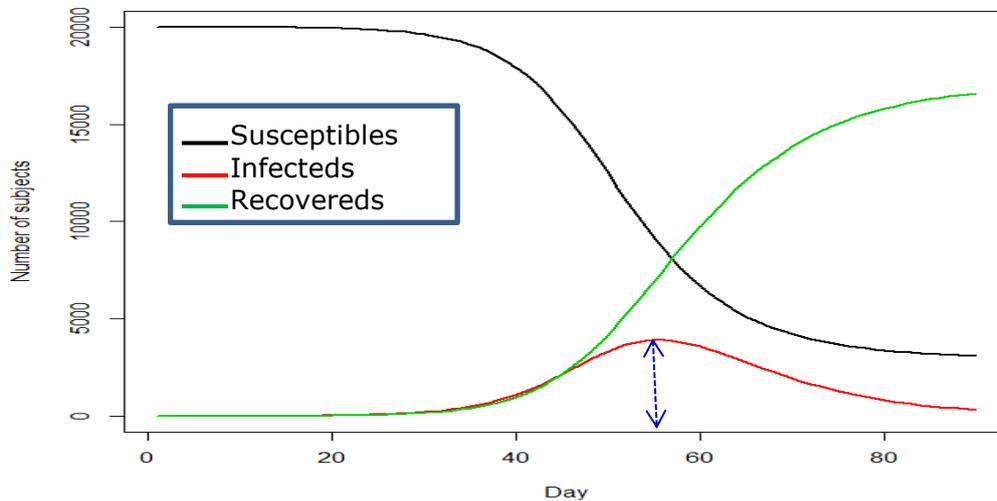


Figure 8 : Solution d'équilibre du modèle SIR

A la lumière de ces résultats, le modèle S.I.R (figure [8]) montre que le pic épidémique ne se diffère pas beaucoup de celui obtenu par la méthode d'estimation de la fonction logistique. On constate sur la figure qu'il est atteint le 22 Avril 2020, Certainement, cette épidémie va régresser après le pic et finira par disparaître très prochainement, mais le risque de rebond est toujours probable tant qu'il n'y pas de vaccin. Théoriquement, une épidémie dure environ deux fois la période qui sépare son début de la date du jour de pic épidémique. La modélisation SIR sous examen nous indique (**figure 8**) qu'elle va durer jusqu'à la fin de la 1^{ème} semaine de juin avant de disparaître complètement.

3. Prédiction de la propagation de l'épidémie par la fonction logistique

La courbe épidémique contient trois types d'évolution : une croissance exponentielle au début, une croissance ralentie sous forme linéaire ensuite et plus tard, l'arrêt de la croissance complète. Cette évolution typique de **la population** est modélisable par l'équation logistique du belge Pierre-François Verhulst [1888]⁶ à savoir $f(t) = \frac{L}{1 + e^{-k*t}}$ où k est positif et « L » la limite de la fonction au voisinage de l'infini. La courbe de cette fonction prend la

⁶ Dans les études épidémiologiques, la fonction logistique est fréquemment utilisable dans la modélisation de prédiction et de propagation des épidémies.

forme de la lettre « S »⁷ c'est-à-dire elle évolue exponentiellement au début puis effectue un point d'inflexion et change de direction en gardant une forme linéaire avant de s'arrêter définitivement. L'ajustement de la courbe épidémique du Maroc (**Figure [5]**) par cette

fonction logistique a donné l'équation estimée suivante⁸: $\hat{C}_m = \frac{5822,81}{(1+Exp(+5,89-0,1196*time))}$

sur cette figure sont configurés les données observées des cas actifs (en bleu), le modèle prédit (en noire) et les prévisions (en vert). Sur la figure [6], il est reporté le pic théorique (estimé) et les cas confirmés par jour prédit avant et après le pic. La lecture des résultats d'estimation par la fonction logistique révèle :

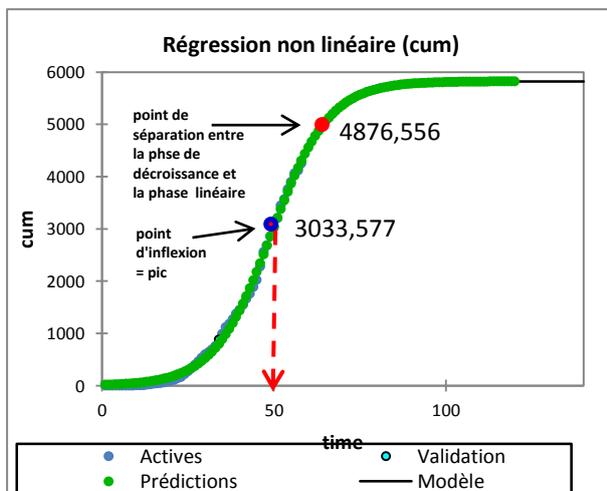


Figure [5]

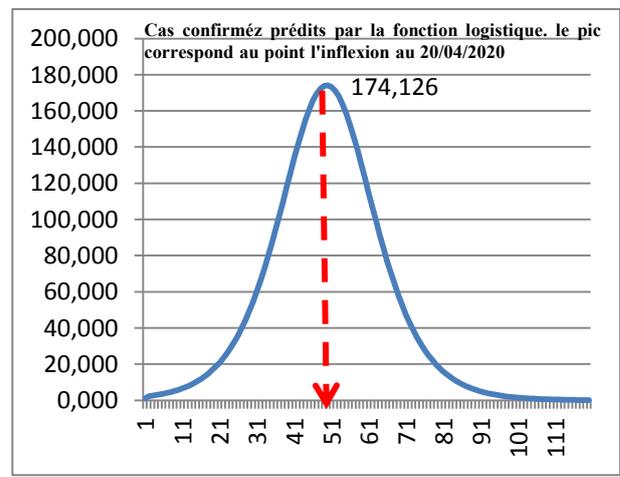


Figure [6]

- Le passage de la phase exponentielle à la forme linéaire a commencé le 20-Avril 2020, ce qui laisse penser au pic épidémique qui se situe autour de cette date. Il ne faut pas s'attendre à un pic très remarquable comme celui de la chine. D'après les estimations par la fonction logistique, le pic théorique (estimé) est atteint pour un effectif théorique de 174,12 cas (**figure [6]**) auquel correspond un pic observé de 191.
- Au Maroc, vu le nombre de cas faible d'un jour à l'autre, le pic peut être inaperçu ;
- Les cas confirmés par jour restent généralement au tour de 100 durant toute la portion linéaire (entre le 20 et 05 Avril au plus tard);
- Dès le franchissement de la phase linéaire de la courbe en vert le 05 mai 2020 (point rouge sur la figure [5]), le nombre de cas par jour va décroître en se situant autour de 17 cas le 20 Mai. Après cette date les cas présentant des périodes d'incubation longues apparaîtront avec retard mais ils ne sont pas nombreux.

⁷ La fonction logistique est connue par sa forme de dessin tel un « S »

⁸ Tous les résultats d'estimation sont établis sur la période de 02 au 30 Avril 2020 sur des données des cas cumulés

- Le nombre total des infectés se situent entre de 5600 et 5800.
- L'arrêt définitif de l'épidémie causée par coronavirus est prévu au plus tard la semaine du 14 juin 2020.

Conclusion

La lecture graphique reste une preuve de la crédibilité des prédictions par les calculs mathématique, mais elle ne peut en aucun cas remplacer les outils de prévision. Par ailleurs, il faut noter que chacune des méthodes de modélisation qu'on a présenté ci-dessus est dotée d'une base de calcul propre et des hypothèses propres, et pourtant elles se sont convergées presque vers le même résultat.

Les ces conclusions auxquelles nous sommes parvenus sont le fruit de l'exploitation des outils de modélisation dans le domaine épidémique d'aspect universel, il se peut qu'elles ne soient traduites effectivement dans la réalité.

Enfin j'espère bien du fond du cœur que ces prévisions se produisent et voient le jour tel qu'il est approuvé par les modèles théorique E.I.R et logistique.