Estimation de l'évolution spatio-temporelle de l'intensité de la pandémie de covid-19

P. Abry & équipe Signaux, Systèmes et Physique (SiSyPhe) - Institut rhône-alpin des systèmes complexes
ENS de Lyon

20 juillet 2023

Résumé

Une équipe interdisciplinaire de chercheurs de l'Institut Rhônalpin des Systèmes Complexes à l'ENS de Lyon (équipe Signaux, Système et Physique du Laboratoire de Physique, équipe DANTE du LIP, laboratoire Triangle) a propose un travail original permettant l'estimation de l'évolution spatio-temporelle de l'intensité de la pandémie de Covid-19, mesurée par le taux de reproduction $R(t)$. Cette estimation repose sur la combinaison d'un modèle épidémiologique et de techniques de résolution de problèmes inverses par minimisation convexes non lisses. Cette approche permet de garantir une estimation conjointe (simultanée) du taux de reproduction $R(t)$, qui mesure l'intensité de la pandémie, et du compte corrigé (débruité) de nouvelles infections quotidiennes. Ces estimées sont robustes et réalistes malgré des données covid (nombre de nouvelles infections quotidiennes) de qualité très limité (comptes aberrants, valeurs manquantes, pseudo effets de saisonnalité — week-end/début de semaine —). Elles sont mises à jour automatiquement, quotidiennement pour près de 200 pays, les 101 départements français, les 50 états-unis.)

Table des matières

1 Présentation

1.1 Surveillance de pandémie

La surveillance de l'intensité d'une pandémie repose fondamentalement sur l'observation régulière, par exemple quotidienne, du nombre de nouvelles infections. L'ensemble des pays a donc déployé d'importants moyens (humains, techniques, matériels et financiers) pour la construction de telles données. Les mesures réalisées pendant la pandémie de Covid19 présentent cependant la particularité d'être de qualité très limitée, et ce dans tous les pays du monde. Elles sont, en effet, fortement corrompues par des comptes aberrants, parfois négatifs, des valeurs manquantes (les week-ends ou jours fériés), des comptes reportés les jours suivants. L'organisation des structures de travail a induit un pseudo-rythme hebdomadaire qui néanmoins varie d'un pays à l'autre et même pour un même pays au cours de la pandémie. La France, par exemple, ne produisait de données ni le samedi ni le dimanche au début de la pandémie, avant d'en produire le samedi mais pas le dimanche, pour maintenant fournir une donnée partielle le dimanche. Les jours fériés perturbent cette saisonnalité. Si l'ampleur de la crise sanitaire, le contexte d'urgence qui ont prévalu au début de la pandémie expliquent sans doute cette situation, force est de constater qu'après deux ans de pandémie la qualité des données reste limitée, induisant ainsi une des difficultés majeures pour réaliser une estimation fiable et robuste de l'évolution de l'intensité de la pandémie (cf. figure 1, graphique supérieur).

Quoique la surveillance d'une pandémie puisse être envisagée de multiples façons par différentes sciences [2], l'épidémiologie recourt le plus souvent aux modèles à compartiment [3]. La population est répartie en ç compartiments (sains, malades, guéris, décédés ?) dont les tailles évoluent au cours du temps selon un ensemble de lois ç simples (des équations différentielles non linéaires et couplées). Cependant, pour rendre compte d'une pandémie qui affecte toute la population mais par des canaux divers et avec des intensités différentes, il faut beaucoup de compartiments, ce qui complique significativement l'utilisation pratique et réelle de tels modèles pendant la pandémie et en présence de données de qualité réduite. La surveillance dans un contexte intra-pandémie est donc souvent réalisée à l'aide d'un modèle plus sommaire mais plus robuste concentré sur un seul paramètre : le taux de reproduction de la pandémie R, qui mesure le nombre d'individus qui seront, en moyenne, infecté par une seule et même personne déjà infectée [4, 8]. Ce nombre inquiète donc lorsqu'il est supérieur à 1, la pandémie progresse, et rassure quand il est inférieur à 1, la pandémie régresse. L'idée du modèle épidémiologique développé dans [5] et utilisé ici est de rendre R dépendant du temps, R_t , et d'ainsi pouvoir suivre ses variations : plus R_t s'éloigne de 1 par valeurs supérieures plus la pandémie accélère, plus il se rapproche de 0 plus la pandémie ralentit. C'est à l'estimation (c'est-à-dire l'évaluation de sa

valeur à partir des données observées) de l'évolution en temps et en espace de ce R_t que nous nous sommes intéressés [1, 7, 5].

1.2 Modèle épidémiologique

Le modèle épidémiologique utilisé dans ce travail, celui développé dans [4, 7], préserve le mécanisme fondamental de propagation d'une épidémie : le nombre de nouvelles infections aujourd'hui dépend du taux de reproduction, que l'on cherche à mesurer, multiplié par une moyenne des nombres d'infections de jours précédents, moyenne pondérée par la fonction sérielle d'intervalle. Celle-ci quantifie la probabilité que l'apparition de symptômes aujourd'hui soit due à une infection contractée il y a n jours. Pour les pandémies à coronavirus, cette fonction sérielle est connue, de valeur moyenne 7 jours et prenant des valeurs significatives de 3 à 12 jours. C'est cette fonction sérielle qui fixe l'échelle de temps dynamique de la pandémie et qui implique qu'une mesure prise aujourd'hui ne peut avoir d'effets avant typiquement une dizaine de jours. Techniquement, comme il s'agit de comptes de nouvelles infections, le modèle [4] postule donc que le nombre de nouvelles infections aujourd'hui est, conditionnellement aux observations passées, une variable aléatoire gouvernée par une loi de Poisson dont le paramètre, la moyenne, dépend donc de R_t multipliée par cette moyenne pondérée. A partir de ce modèle, il est aisée d'obtenir une expression théorique de l'estimation de R_t aujourd'hui, à partir d'un principe classique de maximum de vraisemblance. Cette estimation, élémentaire à obtenir, produit, mise en œuvre sur les données réelles de Covid19, une mesure de l'intensité de la pandémie. Cette estimée est cependant inutilisable pour les épidémiologistes car fortement variable et très erratique d'un jour au suivant alors que R_t est, par nature, une quantité qui ne peut varier que faiblement et lentement au cours du temps. Cette piètre estimation résulte de deux causes, l'une structurelle, l'autre conjoncturelle : d'une part, on veut estimer R_t chaque jour à partir d'une seule nouvelle donnée, une situation peu robuste statistiquement ; d'autre part, la faible qualité, les valeurs aberrantes, biaisent l'estimation.

1.3 Estimation de l'intensité d'une pandémie

Pour faire face à cette limitation structurelle (une estimée chaque jour à partir d'une seule nouvelle donnée) , il faut envisager des stratégies de régularisation. On adjoint au modèle épidémiologique des contraintes que l'estimée de R_t doit respecter. On lui impose donc ici naturellement d'être positif ou nul, d'être lentement variable en temps. Techniquement, on impose à R_t de varier linéairement au cours du temps, comme un segment de droite donc, avec une pente qui ne peut changer de valeur et de signe (de sens) que rarement. Le nombre de changements et les dates auxquels ils surviennent ne sont pas choisis par un expert mais ç pilotés par les données elle-même. On

impose également que les R_t mesurés sur des territoires voisins, les départements français partageant une frontière terrestre par exemple, soient peu différents sauf pour de rares frontières, sélectionnées par les données où un changement est estimé significatif et autorisé.

Prendre en compte la limitation conjoncturelle (faible qualité des données observées) se fait classiquement par un prétraitement des données. Dans le contexte temps-réel et évolutif de la pandémie de Covid19, la modélisation de ce prétraitement aurait été fastidieuse et délicate : elle aurait dû dépendre de chaque pays, et, pour un même pays, aurait dû varier au fur et à mesure de l'évolution du suivi de l'épidémie. Nous avons construit une stratégie alternative qui consiste à réaliser en même temps et l'estimation de R_t et le traitement des données aberrantes, sans modélisation préalable de celles-ci. Les temps d'occurrence et les amplitudes et signes de ces valeurs aberrantes sont inconnues, et donc à estimer également, ce qui en apparence, augmente la difficulté du problème. Nous avons alors ç perturbé de façon fine le modèle épidémiologique pour autoriser l'occurrence de valeurs aberrantes, tout en préservant le mécanisme fondamental des contaminations secondaires. Nous avons ainsi évité de modéliser ces valeurs aberrantes au-delà de la simple hypothèse de leur survenue de manière isolées (parcimonie), qui agit comme une contrainte de régularisation supplémentaire. Techniquement, les estimations robustes et conjointes de R_t et des valeurs aberrantes, que nous proposons nécessite la minimisation d'une fonctionnelle. L'originalité de notre approche consiste dans le fait d'avoir construit cette fonctionnelle en combinant le modèle épidémiologique, l'inclusion des valeurs aberrantes, les contraintes pertinentes tout en s'assurant qu'elle ait les ç bonnes propriétés techniques (convexité non lisse) qui assurent qu'il sera possible de construire un algorithme fiable et rapide pour la minimiser.

1.4 Résultats

Ces outils sont mis en oeuvre automatiquement et quotidiennement pour produire une estimée temps réel de la pandémie pour près de 200 pays ou territoires autonomes. Les comptes de nouvelles infections sont extraits du serveur de données de l'université John Hopkins (https ://coronavirus.jhu.edu/), qui a réalisé depuis le début de la pandémie un effort soutenu et remarquable de collecte et d'organisation des données de Covid19 mises à disposition par les Autorités de Santé Publique des différents pays et territoires autonomes du monde.

Ils sont également mis en oeuvre automatiquement et quotidiennement pour mesure l'évolution spatio-temporelle de la pandémie à travers les départements de France métropolitaine, estimée à partir des comptes de nouvelles infections ayant entrainées une hospitalisation, tels que mis en libre accès par Santé-Publique-France.

Avec pour objectif un transfert libre et direct d'information sur l'état de

la pandémie vers la société et les citoyens, les graphiques, mis à jour quotidiennement et représentant l'évolution au cours du temps de la pandémie, sont accessibles à perso.ens-lyon.fr/patrice.abry avec

- pour les départements français :
  - une carte de France interactive et animée,
  - les valeurs des estimées.
- pour presque deux cent pays :
  - une carte du monde interactive et animée,
  - les valeurs des estimées.

Ces estimées spatio-temporelles, graphiques et cartes ont un double intérêt. Rétrospectivement, elles permettent de quantifier l'impact de mesures sanitaires (confinement, couvre-feu, ?) sur l'évolution de la pandémie et de comparer ces impacts sur différents territoires, départements ou pays. Prospectivement, le choix de produire des estimées linéaires par morceaux, a pour conséquence de livrer non seulement une estimée de R aujourd'hui mais aussi de la tendance actuelle, et ainsi de prévoir l'évolution de la pandémie à l'horizon de l'échelle de temps imposée par la fonction sérielle, de quelques semaines seulement donc. L'outil ne permet ainsi pas de prévoir le long terme mais de donner la tendance à court terme.

1.5 Travaux en développements

L'amélioration de ces outils se poursuit.

- Des techniques bayésiennes d'échantillonnage aléatoire permettront notamment bientôt de proposer des estimations assorties d'intervalles de crédibilité, comme illustré sur la figure 3 [6, 5]. Les premiers résultats pour la France sont mis-à-jour quotidiennes et accessibles à https ://perso.math.univ-toulouse.fr/gfort/ (et en version détaillée à https ://perso.math.univ-toulouse.fr/gfort/project/opsimore/)
- Le réglage de la proportion relative des contraintes de régularisation ajoutées vis-à-vis du modèle épidémiologique sera ajustée au cours du temps par un pilotage par les données, pour ajuster les différences d'intensité entre les vagues successives de la pandémie.
- Enfin, en cohérence avec l'effort d'ouverture de la science vers la société, l'enrichissement en contenu et en ergonomie des cartes interactives et animées se poursuit.

Finalement, ce travail a également été soutenu par un cycle de séminaires La covid-19: regards et questions interdisciplinaires, organisé par l'Institut rhône-alpin des systèmes complexes IXXI.

2 Graphiques et cartes

2.1 France - National

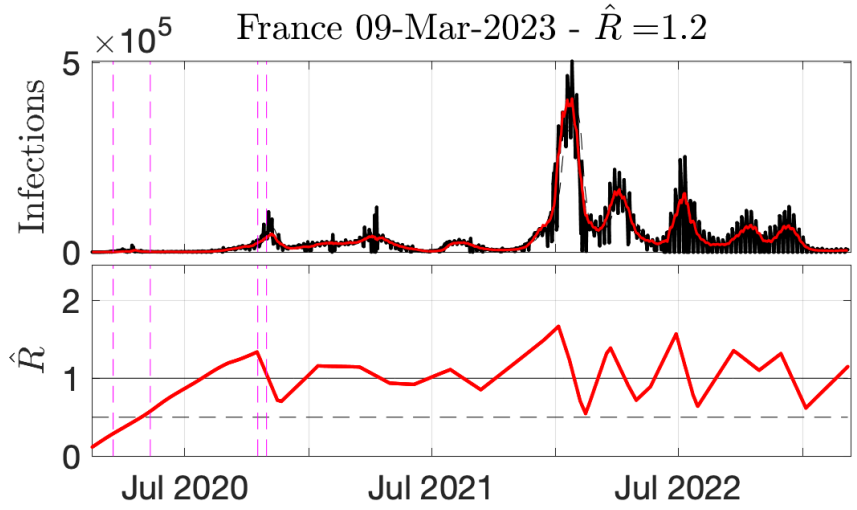

FIGURE 1 – **Estimées quotidiennes de $R(t)$.** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

2.2 France - Estimation par intervalles de crédibilité

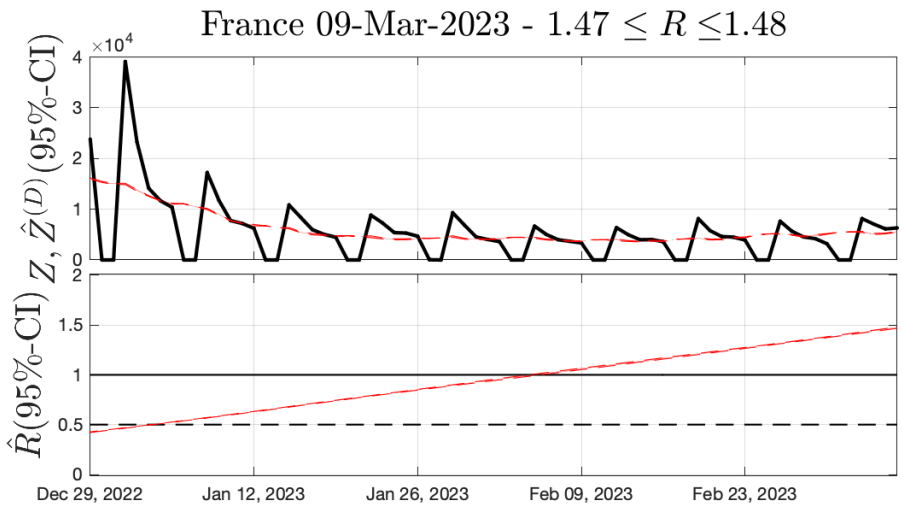

FIGURE 2 – **Estimées quotidiennes de $R(t)$ par intervalles de crédibilité (5 dernières semaines).** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

2.3 France - *Départements*

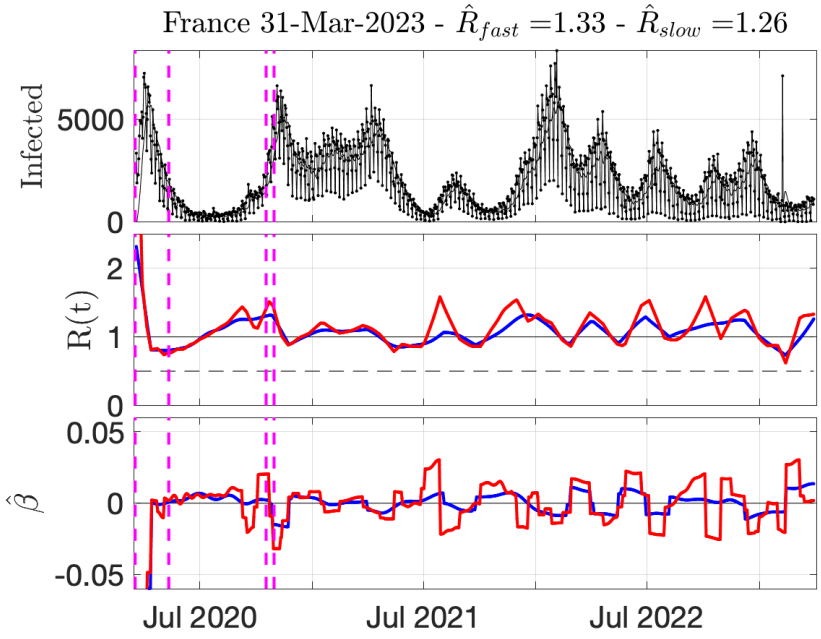

FIGURE 3 – **Estimées de $R(t)$ (à partir des entrées àl'hôpital).** Données : nouvelles infections ayant néceéssitées une entrée à l'hôpital. Source : Santé Publique France.

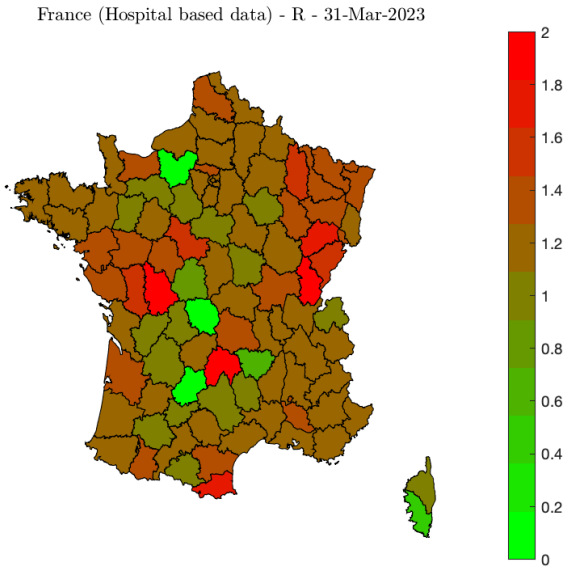

FIGURE 4 – **Estimées de $R(t)$ par départements.** Données : nouvelles infections ayant néceéssitées une entrée à l'hôpital. Source : Santé Publique France.

2.4 Europe

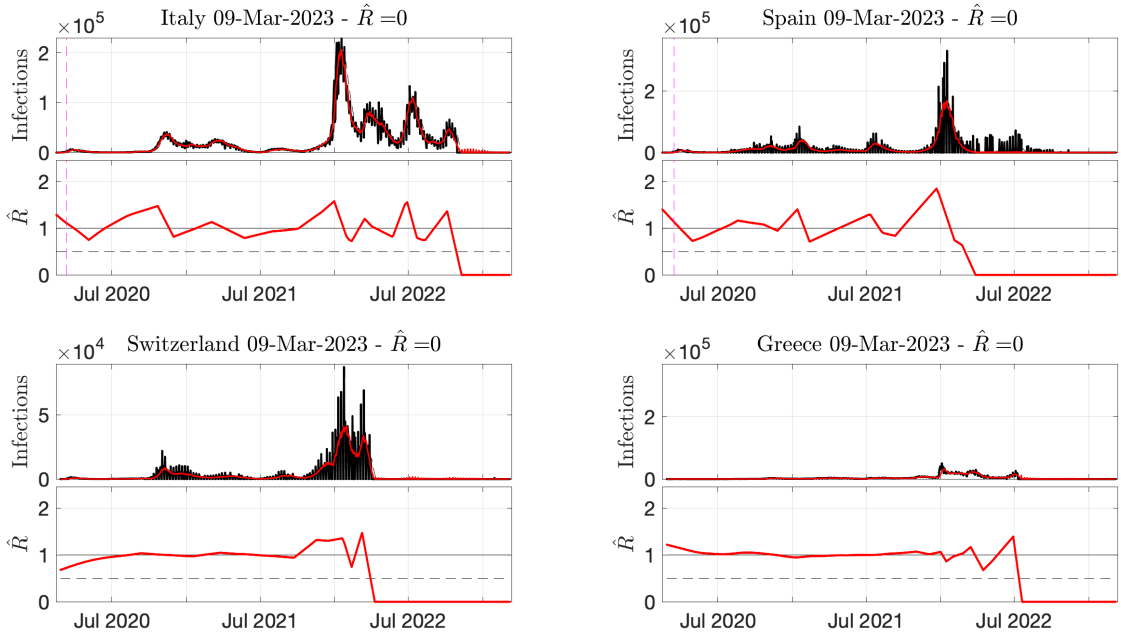

FIGURE 5 – **Estimées de $R(t)$ mises à jour quotidiennement.**

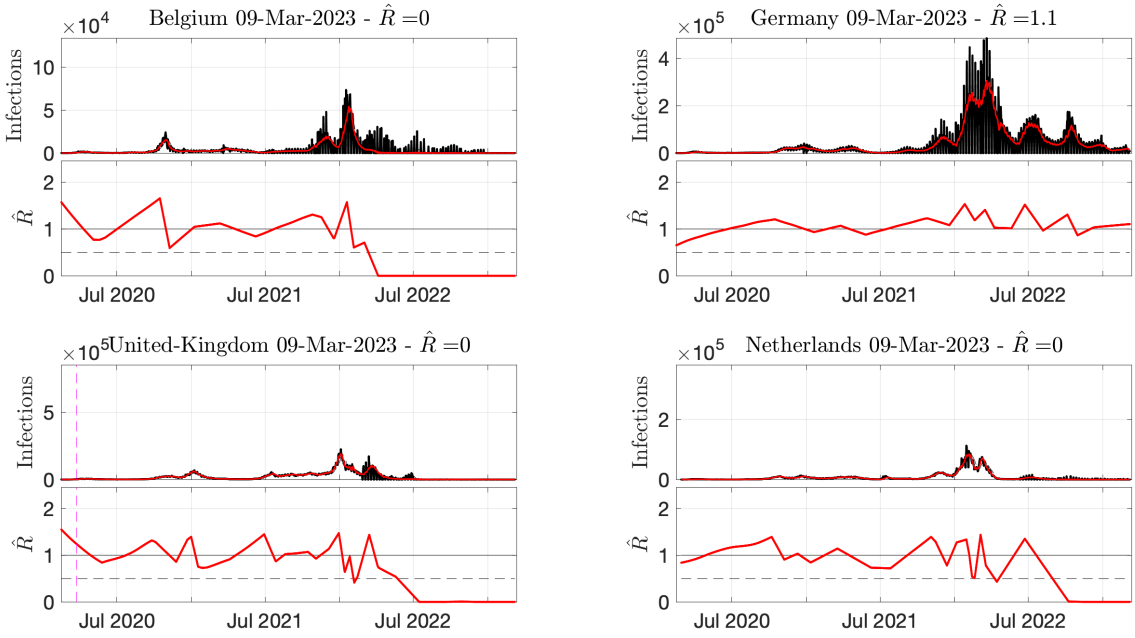

FIGURE 6 – **Estimées de $R(t)$ mises à jour quotidiennement.** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

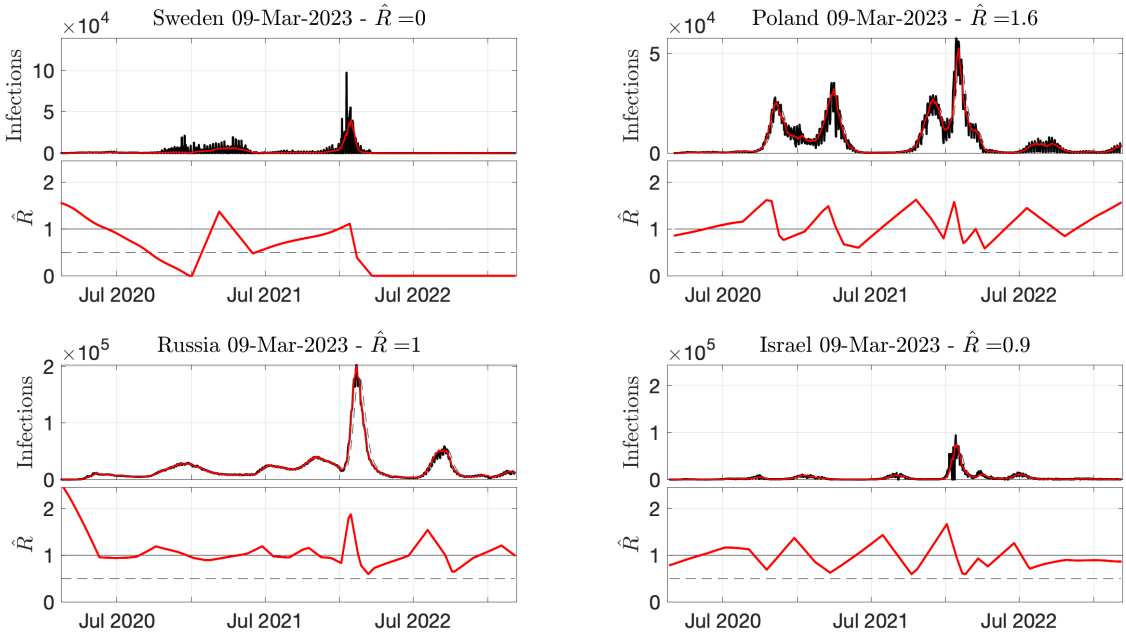

FIGURE 7 – **Estimées de $R(t)$ mises à jour quotidiennement.** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

2.5 America

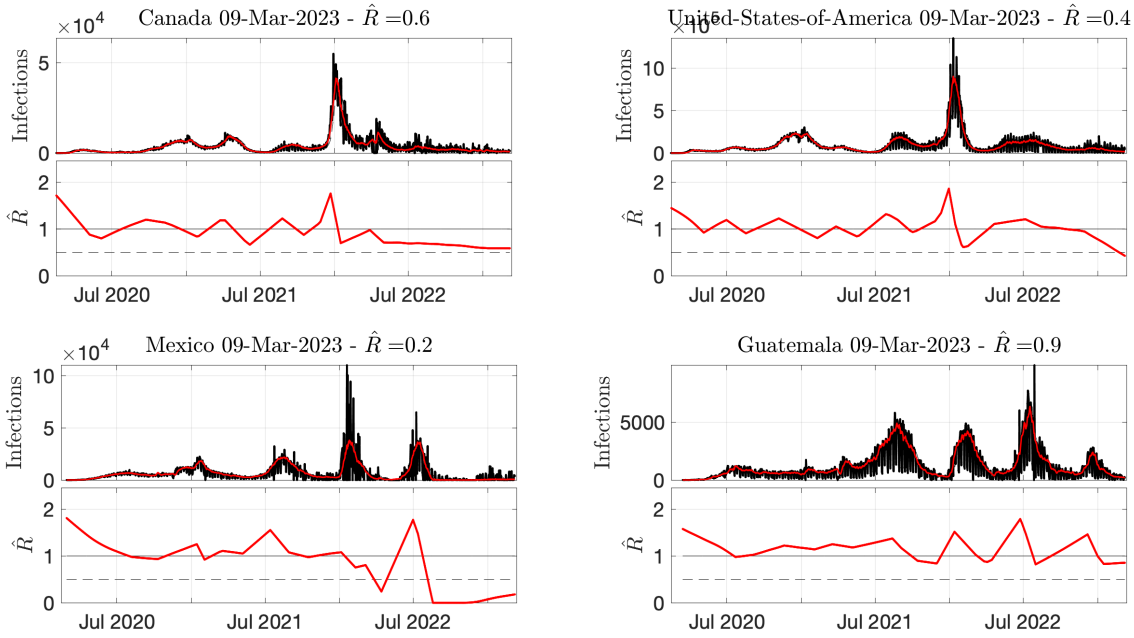

FIGURE 8 – **Estimées de $R(t)$ mises à jour quotidiennement.** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

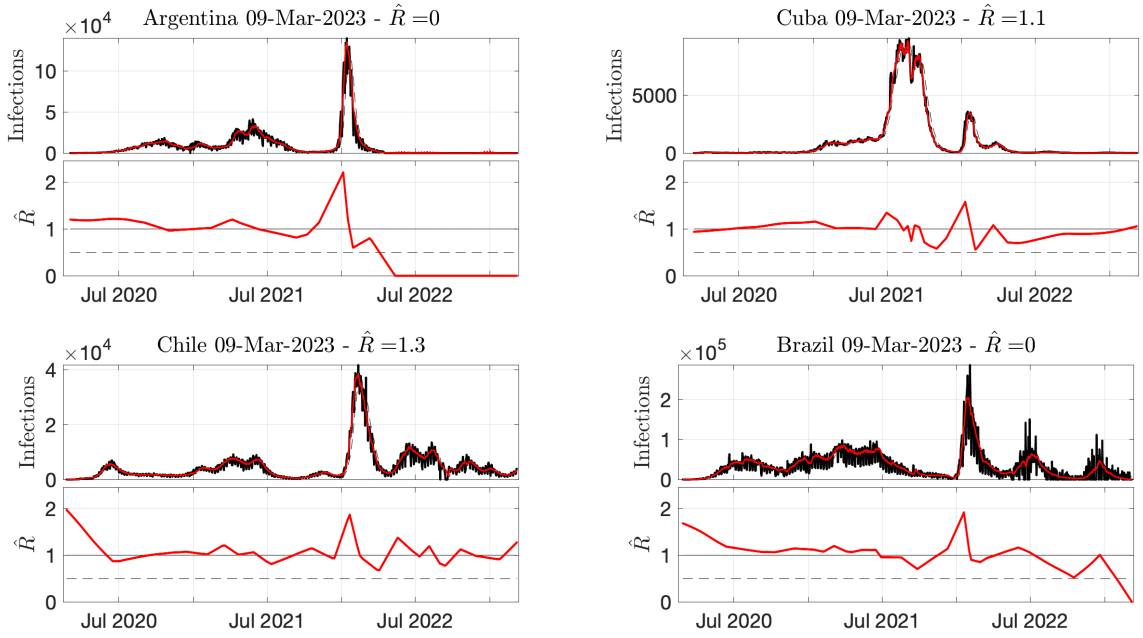

FIGURE 9 – **Estimées de $R(t)$ mises à jour quotidiennement.** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

2.6 Asia

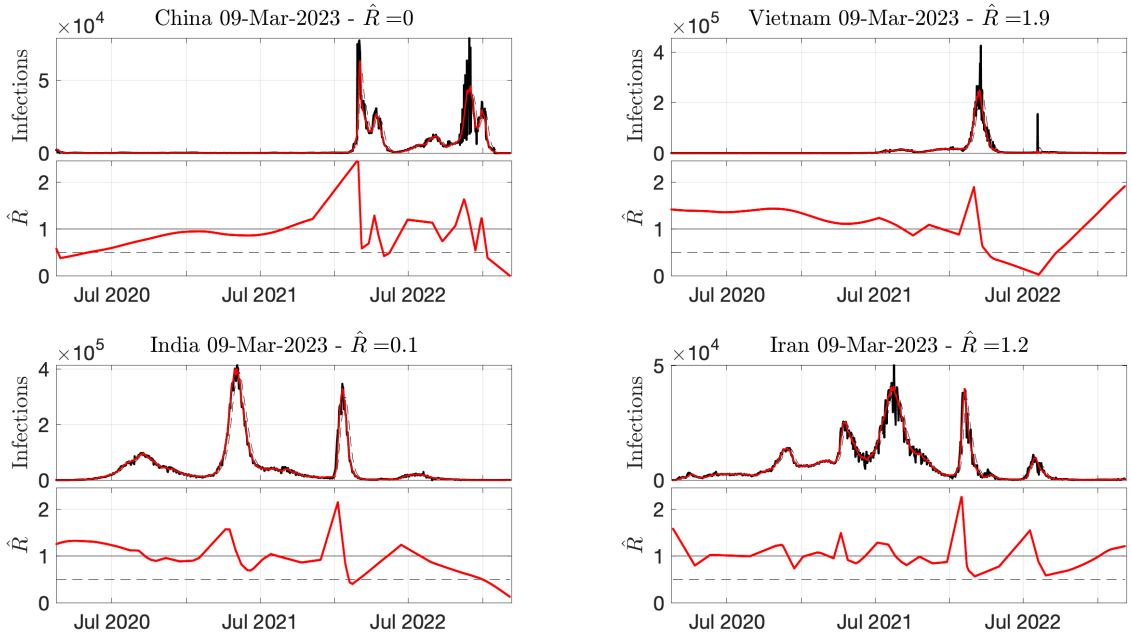

FIGURE 10 – **Estimées de $R(t)$ mises à jour quotidiennement.** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

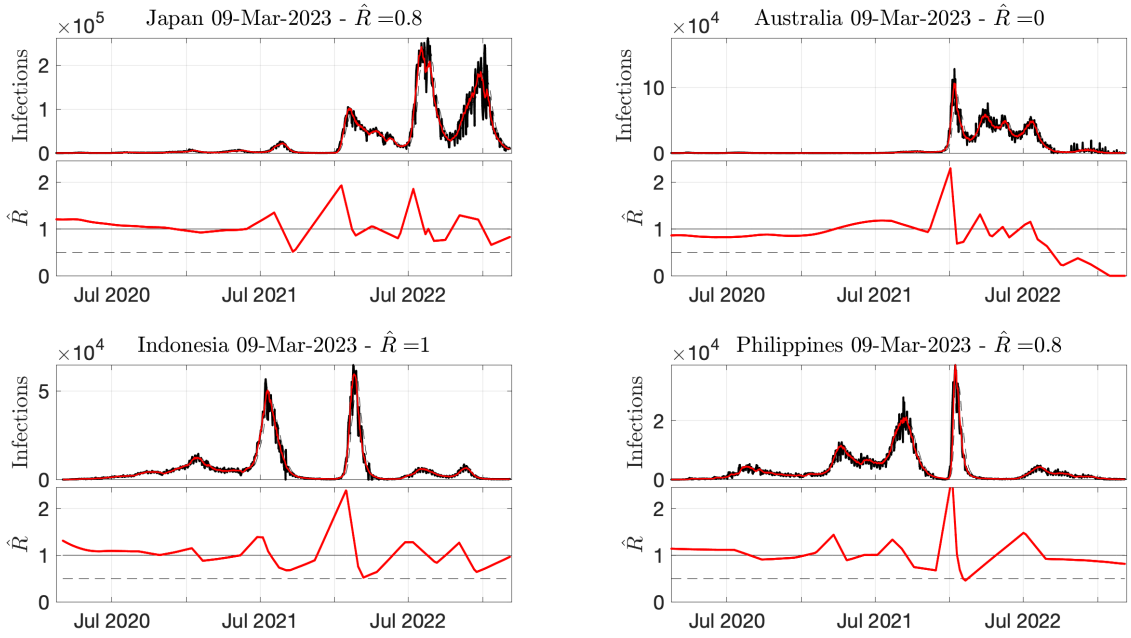

FIGURE 11 – **Estimées de $R(t)$ mises à jour quotidiennement.**Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

2.7 Africa

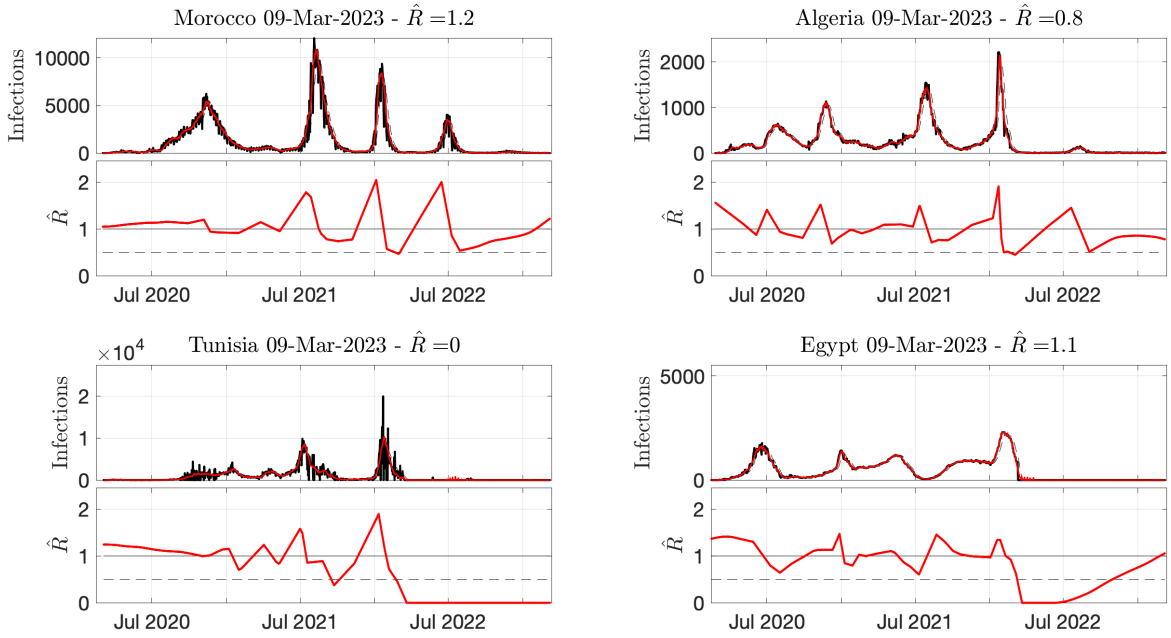

FIGURE 12 – **Estimées de $R(t)$ mises à jour quotidiennement.** Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

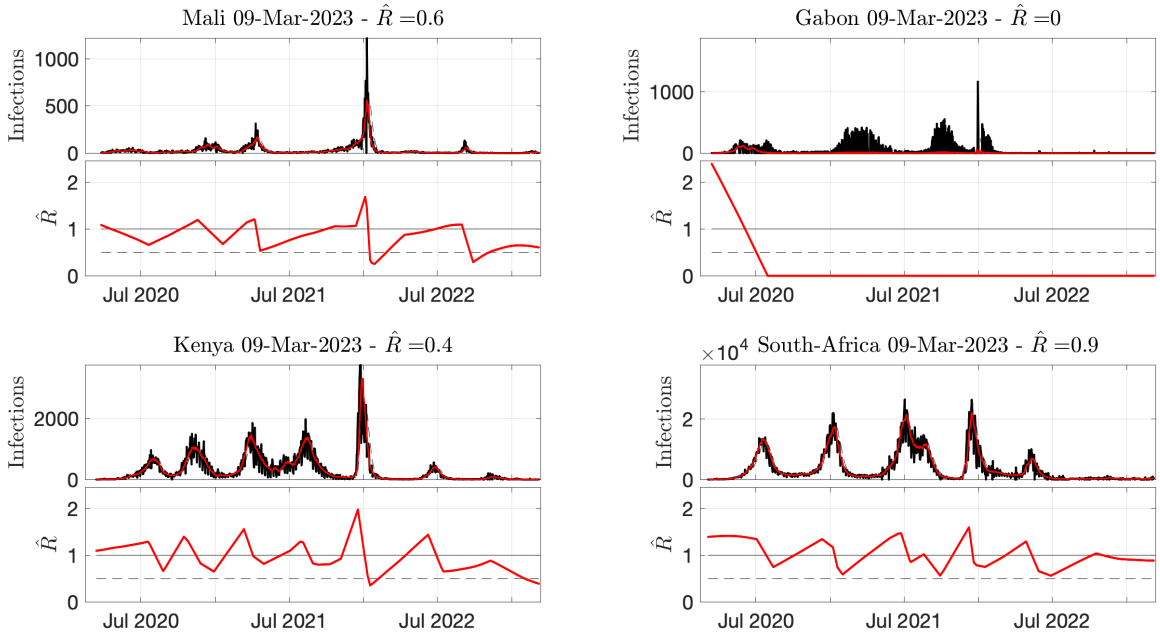

FIGURE 13 – **Estimées de $R(t)$ mises à jour quotidiennement.**Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

2.8 Worldwide COVID-19

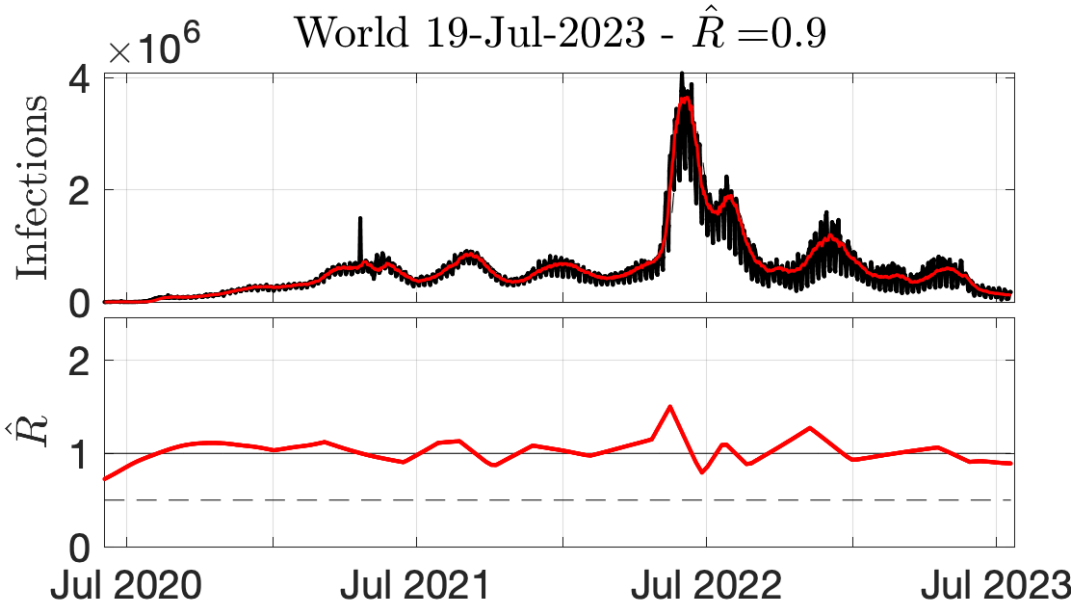

FIGURE 14 – **Estimées de $R(t)$ pour le monde.**Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

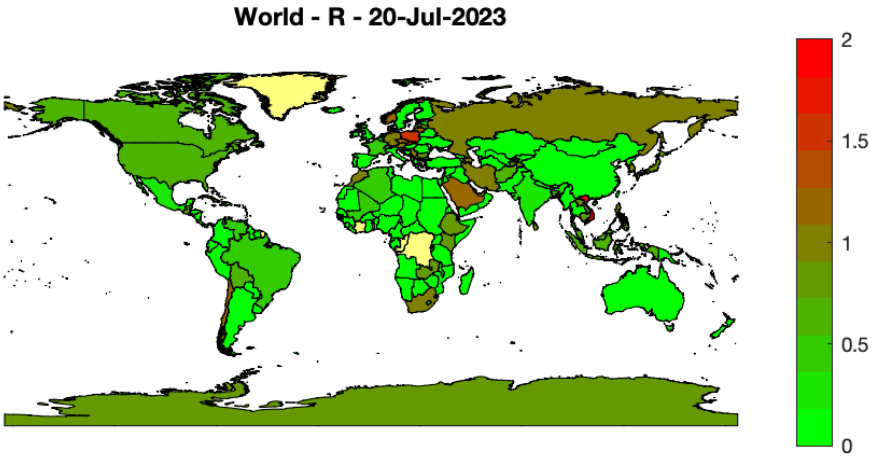

FIGURE 15 – **Carte du monde des estimées de $R(t)$.**Donnés : Nouvelles infections quotidiennes. Source : Université Johns Hopkins. Graphique supérieur : nouvelles infections quotidiennes brutes (noir) et corrigées (rouge). Graphique inférieur : Estimation de $R(t)$ (rouge).

Références

[1] P. Abry et al. Spatial and temporal regularization to estimate COVID-19 reproduction number $R(t)$: Promoting piecewise smoothness via convex optimization. *PLOS One*, 15, 2020. e0237901.

[2] J. Arino. Describing, modelling and forecasting the spatial and temporal spread of COVID-19–A short review. Technical report, arXiv :2102.02457, 2021.

[3] F. Brauer, C. Castillo-Chavez, and Z. Feng. *Mathematical models in epidemiology*. Springer, New York, 2019.

[4] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.*, 178 :1505–1512, 2013.

[5] G. Fort, B. Pascal, P. Abry, and N. Pustelnik. Covid19 Reproduction Number : Credibility Intervals by Blockwise Proximal Monte Carlo Samplers. Technical Report hal-03611079v1, 2022.

[6] G. Fort, B. Pascal, P. Abry, and N. Pustelnik. Covid19 reproduction number : Credibility intervals by blockwise proximal monte carlo samplers. *submitted*, 2022.

[7] B. Pascal, P. Abry, N. Pustelnik, S. Roux, R. Gribonval, and P. Flandrin. Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data. Technical report, arXiv 2109.09595, 2022.

[8] R.N. Thompson et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29 :100356, 2019.