

Modèle bayésien hiérarchique pour l'estimation du coefficient de reproduction de la COVID-19 à partir de comptes hebdomadaires^{†,*}

Barbara PASCAL¹ Patrice ABRY²

¹Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France, barbara.pascal@cns.fr

²CNRS, Ens de Lyon, Laboratoire de Physique, Lyon, France, patrice.abry@ens-lyon.fr

Résumé – La surveillance d'une épidémie en temps réel est un enjeu majeur, notamment au long cours comme en témoigne la pandémie de COVID-19 qui reste active plus de cinq ans après son émergence. Un indicateur de l'intensité de la propagation est le coefficient de reproduction, dont l'estimation achoppe sur plusieurs difficultés. Tout d'abord les données collectées en temps réel sont parfois de qualité médiocre les comptes des jours chômés étant souvent sous-estimés ou comptabilisé avec un délai variable. En outre, une fois l'épidémie installée, certaines agences de santé passent d'un report quotidien à un décompte hebdomadaire, rendant caduques les méthodes d'estimation quotidienne du coefficient de reproduction. Pour palier ces écueils, un modèle bayésien hiérarchique adapté à une description à hebdomadaire de la dynamique épidémique est proposé. Un échantillonneur de Gibbs est ensuite construit afin de calculer la moyenne a posteriori et les intervalles de crédibilité du coefficient de reproduction hebdomadaire. La procédure complète est illustrée sur les comptes d'infection à la COVID-19 de plusieurs pays et phases de l'épidémie.

Abstract – Monitoring an epidemic in real time is a major challenge, especially over the long term, e.g., for the COVID-19 pandemic, which remains active more than five years after its emergence. One indicator of the propagation intensity is the reproduction number, the estimation of which might suffer from several issues. First, data collected in real time are often of poor quality, underestimating the counts on daysoff or reporting them with a delay. Furthermore, once an epidemic has taken hold, some health agencies switch from a daily to a weekly basis reporting, impairing the use of standard methods estimating a daily reproduction coefficient. To overcome these pitfalls, a hierarchical Bayesian model adapted to a weekly description of epidemic dynamics is proposed. A Gibbs sampler is then built to compute the a posteriori mean and credibility intervals of the weekly reproduction coefficient. The procedure is illustrated on COVID-19 infection counts from different countries and epidemic phases.

1 Introduction

Contexte. Lors d'une épidémie telle que la pandémie de COVID-19 le suivi en temps réel de la propagation de l'agent pathogène dans la population est un enjeu de santé public majeur [7, 9]. Un indicateur épidémiologique largement utilisé pour quantifier l'intensité d'une épidémie est le coefficient de reproduction, défini au temps t comme le nombre moyen d'infections secondaires générées par un individu infecté typique durant sa période de contagiosité [15, 5, 10]. Or les méthodes d'estimation du coefficient de reproduction de l'état-de-l'art en épidémiologie [5, 11] se heurtent à deux obstacles majeurs. D'une part, les données collectées par les agences de santé souffrent d'erreurs de reports parfois importantes, causées par exemple par un report retardé des cas détectés les jours chômés [6, 4, 8]. D'autre part, dans le cas de la pandémie de COVID-19, après deux ans de reports quotidiens, certaines agences de santé ont opté pour un décompte hebdomadaire du nombre de nouveaux cas. C'est le cas par exemple au Royaume-Uni, dont les reports quotidiens de septembre 2022 à mars 2023 publiés par l'Université Johns Hopkins¹ sont tracés en gris sur la Figure 1 (*haut*). Les données hebdomadaires fluctuent assurément moins sous l'effet des jours chômés, néanmoins leur granularité plus grossière doit conduire à

une incertitude majorée sur le coefficient de reproduction. **État-de-l'art.** Si les comptes d'infections à la COVID-19 souffrent d'un bruit administratif très important, cette problématique avait déjà été observée lors d'épidémies passées [6]. Afin de palier cet écueil, des études ont alors proposé d'agréger les données sur une échelle de temps suffisamment grande pour lisser les fluctuations non pertinentes vis-à-vis de la dynamique épidémique [6, 4, 10], induisant typiquement un passage à une granularité hebdomadaire. Une fois éliminé le bruit administratif, une première catégories d'approches consiste à définir des indicateurs à l'échelle de temps plus grossière [6, 4]. Récemment, une interpolation guidée par le modèle épidémiologique [5] a été élaborée afin de revenir, à partir de données agrégées, tels que les comptes tracés en noir à la Figure 1 (*haut*), à des données quotidiennes et pouvoir ainsi appliquer les estimateurs de l'état-de-l'art [5, 11]. Les comptes quotidiens interpolés du Royaume-Uni sont tracés en bleu sur la Figure 1 (*haut*); EpiEstim [5] fournit alors l'estimée du coefficient de reproduction tracée en vert en Figure 1 (*bas*) où les intervalles de crédibilité à 95% associés sont trop petits pour être visibles. L'obstacle majeur à l'utilisation de ces outils pour appuyer des décisions de santé publique à fort impact économique et social est que ces approches ne rendent pas compte, dans le calcul des intervalles de crédibilité du coefficient de reproduction, de la nécessaire majoration de l'incertitude lors du passage à une échelle de temps plus grossière.

Contributions. Un modèle de propagation épidémique strictement hebdomadaire est proposé, rendant compte de la variabilité intrinsèque plus élevée des comptes hebdomadaires en

¹<https://coronavirus.jhu.edu/>

[†] Avec le soutien du financement ANR OptiMoCSI (ANR-23-CE48-0009).
^{*} Les auteurs remercient chaleureusement Gersende Fort et Juliette Chevallier pour les échanges fructueux sur la construction du modèle bayésien hiérarchique et sur l'implémentation de l'échantillonneur de Gibbs.

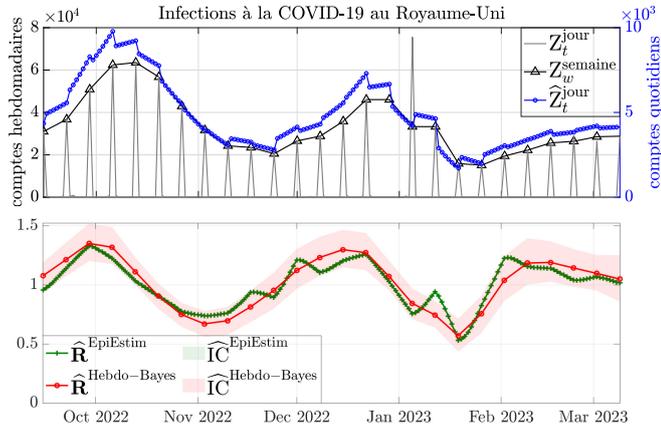


FIGURE 1 : (Haut) Comptes d'infections par la COVID-19 par semaine (en noir) et comptes quotidiens interpolés (en bleu) [10]. (Bas) Estimées du coefficient de reproduction via EpiEstim [5] sur les données interpolées (en vert) vs. via l'estimateur Hebdo-Bayes proposé (en rouge) avec intervalles de crédibilité à 95% associés sous forme de régions colorées.

comparaison des comptes quotidiens. Suivant un formalisme bayésien ce modèle est supplémenté d'une distribution a priori sur le coefficient de reproduction ainsi que d'une distribution a priori sur l'hyperparamètre de ce *prior*. Le modèle bayésien hiérarchique résultant est décrit à la Section 2. Un estimateur bayésien du coefficient de reproduction hebdomadaire, sous forme d'une moyenne a posteriori et d'intervalles de crédibilité calculés via un échantillonneur de Gibbs, est proposé en Section 3. En Section 4, cet estimateur est appliqué à des données réelles de nouvelles infections à la COVID-19 et comparé à un estimateur de l'état-de-l'art qui s'appuie sur une interpolation des données hebdomadaires en comptes quotidiens. Des perspectives sont esquissées en Section 5.

Notations. L'ensemble des réels strictement positifs est désigné par $\mathbb{R}_{>0}$ et les vecteurs de taille W à composantes toutes strictement positives par $\mathbb{R}_{>0}^W$. La distribution d'une variable aléatoire X est notée $\pi(X)$ et pour Y une seconde variable aléatoire, $\pi(X, Y)$ (resp. $\pi(X|Y)$) désigne la loi jointe de X et Y (resp. la loi conditionnelle de X sachant Y).

2 Modèle bayésien hiérarchique

Afin, d'une part, d'adapter le modèle initial [5] à des données hebdomadaires, et d'autre part de rendre plus fidèlement compte de la variabilité intrinsèque du nombre d'infections au cours du temps, un modèle étendu a été proposé [13, Section 5.1]. Conditionnellement aux comptes *hebdomadaires* passés, Z_1, \dots, Z_{w-1} , le nombre de nouvelles infections à la semaine w , noté Z_w , suit une loi de Poisson remise à l'échelle :

$$Z_w | Z_1, \dots, Z_{w-1}, R_w \sim \gamma \mathcal{P} \left(\frac{\Phi_w^Z R_w}{\gamma} \right). \quad (1)$$

avec R_w le coefficient de reproduction *hebdomadaire*, Φ_w^Z l'infectiosité globale dans la population à la semaine w

$$\Phi_w^Z = \sum_{v=1}^{\tau_\phi} \phi(v) Z_{w-v} \quad (2)$$

obtenue par une somme pondérée des comptes précédents, les poids étant déterminés par la *fonction d'intervalle de série* ϕ

discrétisée à l'échelle de la semaine et $\gamma > 0$ le paramètre d'échelle de la loi de Poisson généralisée. Le coefficient de reproduction R_w quantifie l'intensité de l'épidémie à la semaine w : lorsqu'il est plus grand que un, il signe une augmentation exponentielle du nombre de cas, motivant la mise en place précoce de mesures sanitaires de l'épidémie. Le paramètre d'échelle γ permet d'ajuster la variance de la distribution ; en particulier, $\gamma > 1$ implique une variance plus importante que celle du modèle épidémiologique de Poisson standard [5].

Distributions a priori. Dans la continuité de [2, 8], le coefficient de reproduction hebdomadaire est modélisé par une variable aléatoire de Laplace autorégressive de second ordre

$$\pi(R_w | R_{w-1}, R_{w-2}, \lambda_R) \propto \lambda_R e^{-\frac{\lambda_R}{4} |R_w - 2R_{w-1} + R_{w-2}|} \quad (3)$$

correspondant à une loi de Laplace de paramètre λ_R sur la dérivée seconde de R_w . Puis, guidé par le choix d'un prior conjugué à la loi de Laplace [1], le paramètre λ_R est modélisé par une variable aléatoire suivant une distribution Gamma :

$$\pi(\lambda_R) \propto \lambda_R^{\alpha_R - 1} e^{-\beta_R \lambda_R} \quad (4)$$

où $\alpha_R > 0$ (resp. $\beta_R > 0$) est le paramètre de forme (resp. d'intensité) de la distribution Gamma.²

Loi a posteriori. En appliquant la règle de Bayes à la vraisemblance (1) du modèle de Poisson remis à l'échelle et aux distributions a priori (3) et (4) du coefficient de reproduction $\mathbf{R} = (R_1, \dots, R_W)$ et de l'hyperparamètre λ_R , le logarithme de la distribution jointe s'écrit, à une constante près :

$$\ln \pi(\mathbf{R}, \lambda_R | \mathbf{Z}, R_{-1}, R_0) = - \sum_{w=1}^W \frac{(R_w \Phi_w^Z - Z_w \ln(R_w \Phi_w^Z))}{\gamma} - \lambda_R (\|\mathbf{D}\mathbf{R} + \delta\|_1 + \beta_R) + (W + \alpha_R - 1) \ln \lambda_R, \quad (5)$$

où $\mathbf{Z} = (Z_1, \dots, Z_W)$ est le vecteur des comptes hebdomadaires, $4\delta = (R_{-1} - 2R_0, R_0, 0, \dots, 0)^T \in \mathbb{R}^W$, $\|\cdot\|_1$ désigne la norme ℓ_1 et $\mathbf{D} \in \mathbb{R}^{W \times W}$ est la matrice de l'opérateur de dérivée seconde discrète agissant sur $\mathbf{R} \in \mathbb{R}^W$ par

$$\forall w \in \{2, \dots, W\}, \quad (\mathbf{D}\mathbf{R})_w := (R_w - 2R_{w-1} + R_{w-2})/4,$$

avec les conditions initiales $(\mathbf{D}\mathbf{R})_1 = R_1/4$, $(\mathbf{D}\mathbf{R})_2 = R_2/4 - R_1/2$. Enfin, la distribution a posteriori marginale de \mathbf{R} est obtenue par intégration sur la variable λ_R , $\pi(\mathbf{R} | \mathbf{Z}, R_{-1}, R_0) =$

$$\int_{\mathbb{R}_+} \pi(\mathbf{R} | \mathbf{Z}, R_{-1}, R_0, \lambda_R) \pi_{\alpha_R, \beta_R}(\lambda_R) d\lambda_R. \quad (6)$$

Modélisation de la pandémie de COVID-19. L'intervalle de série de la COVID-19, défini comme la durée entre une infection primaire et une infection secondaire, est modélisée par une variable aléatoire suivant une loi Gamma de moyenne 6,6 jours et d'écart type 3,5 jours [3, 14]. Dans le modèle hebdomadaire étudié, la densité de cette loi Gamma est agrégée de façon hebdomadaire par intégration sur des intervalles d'une semaine [6, 4, 13], fournissant les poids $\{\phi(v), v = 1, \dots, \tau_\phi\}$ utilisés à l'Équation (2) pour le calcul de l'infectiosité hebdomadaire, avec une durée $\tau_\phi = 4$ semaines.³ En s'inspirant de l'heuristique proposée dans [13], le paramètre d'échelle est

²Ayant pour espérance $\mu_R = \alpha_R/\beta_R$ et pour variance $\sigma_R^2 = \alpha_R/\beta_R^2$.

³L'agrégation de la fonction d'intervalle de série s'appuie sur la méthode des rectangles à gauche afin de préserver la causalité.

Algorithme 1 : Échantillonneur de Gibbs ciblant (5)

Entrées : $\mathbf{Z}, \Phi^{\mathbf{Z}}, R_{-1}, R_0, \alpha_R, \beta_R, \varepsilon$ **Résultat :** $\widehat{\mathbf{R}}^{\text{Hebdo-Bayes}}, \widehat{\mathbf{IC}}^{\text{Hebdo-Bayes}}$

- 1 initialisation : $\mathbf{R}^{(0)} \in \mathbb{R}_{>0}^W, \lambda_R^{(0)} \in \mathbb{R}_{>0}$;
 - 2 **pour** $k = 0 \dots k_{\text{max}}$ **faire**
 - 3 $\mathbf{R}^{(k+1)} \sim \text{PGdual}(\mathbf{R}^{(k)} | \lambda_R^{(k)}; \mathbf{Z}, \Phi^{\mathbf{Z}}, R_{-1}, R_0)$;
 - 4 $\lambda_R^{(k+1)} \sim \Gamma(W + \alpha_R, \|\mathbf{DR}^{(k+1)} + \delta\|_1 + \beta_R)$
 - 5 **fin**
 - 6 $\widehat{\mathbf{R}}^{\text{Hebdo-Bayes}} \leftarrow \text{Moyenne}\{\mathbf{R}^{(k)}, k \geq k_{\text{chauffe}}\}$;
 - 7 $\widehat{\mathbf{IC}}^{\text{Hebdo-Bayes}} \leftarrow \text{Quantiles}_{\varepsilon, 1-\varepsilon}\{\mathbf{R}^{(k)}, k \geq k_{\text{chauffe}}\}$;
-

estimé à partir des données via $\gamma = 0,025 \times \text{std}(\mathbf{Z})$. Les coefficients de reproduction initiaux sont calculés à partir des précédents comptes de nouvelles infections de sorte que $R_{-1} = Z_{-1}/\Phi_{-1}^{\mathbf{Z}}$ et $R_0 = Z_0/\Phi_0^{\mathbf{Z}}$. Enfin, les paramètres caractérisant la loi a priori sur l'hyperparamètre λ_R sont fixés à $\alpha_R/\beta_R = 3,5 \times \text{std}(\mathbf{Z})/\gamma$ et $\sqrt{\alpha_R/\beta_R} = 20 \times \alpha_R/\beta_R$ correspondant à un ratio écart-type sur espérance $\sigma_R/\mu_R = 20$ choisi pour que cet *hyperprior* soit quasi-non informatif.

3 Estimateur hebdomadaire bayésien

Dans un formalisme bayésien, plusieurs estimateurs du coefficient de reproduction $\mathbf{R} = (R_1, \dots, R_w)$ sur la période considérée peuvent être envisagés. L'estimateur défini comme la *moyenne* de la distribution *a posteriori* (6), noté $\widehat{\mathbf{R}}^{\text{Hebdo-Bayes}}$ par la suite, est souvent plébiscité pour ses propriétés d'erreur quadratique moyenne minimale. En outre, l'incertitude sur la valeur exacte du coefficient de reproduction à la semaine w peut être quantifiée au moyen de l'intervalle de crédibilité à 95%, noté $\widehat{\mathbf{IC}}_w^{\text{Hebdo-Bayes}}$. La distribution (6) n'étant pas connue de manière explicite, sa moyenne et ses intervalles de crédibilité doivent être calculés à partir d'échantillons générés au moyen d'un algorithme de Monte Carlo par chaîne de Markov. **Échantillonneur de Gibbs.** Parmi les algorithmes de Monte Carlo par chaîne de Markov, les schémas de Gibbs, procédant par échantillonnages conditionnels alternés des variables [8] sont particulièrement adaptés aux distributions jointes de la forme (5). L'échantillonnage de \mathbf{R} conditionnellement à λ_R s'effectue grâce au mécanisme de proposition `PGdual` introduit dans [8], qui tire son efficacité d'une construction ingénieuse inspirée des algorithmes proximaux utilisés pour calculer le maximum a posteriori [12]. Cet échantillonneur de type gradient proximal étend la dynamique de Langevin à des distributions de la même forme que (5), c'est-à-dire non différentiables, par exemple à cause de la présence d'une norme $\|\cdot\|_1$, et impliquant un opérateur linéaire non injectif, tel que l'opérateur de dérivée seconde D . L'hyperparamètre λ_R peut quant à lui être échantillonné conditionnellement à \mathbf{R} de façon explicite grâce au choix d'un a priori conjugué. Le schéma résultant est décrit en détail à l'Algorithme 1. Après une période de chauffe de k_{chauffe} itérations, les échantillons $(\mathbf{R}^{(k)}, \lambda_R^{(k)})$ suivent la distribution cible $\pi(\mathbf{R}, \lambda_R | \mathbf{Z}, R_{-1}, R_0)$ [8]. Les estimées de la moyenne a posteriori et des intervalles de crédibilité à 95% correspondent alors respectivement à la moyenne empirique et aux quantiles d'ordre $\varepsilon = 2,5\%$ et $1 - \varepsilon = 97,5\%$ calculés sur les échantillons correspondant à $k \geq k_{\text{chauffe}}$.

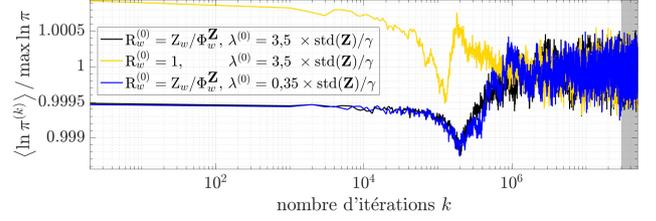


FIGURE 2 : Convergence de l'échantillonneur de l'Algorithme 1 ; $\pi^{(k)} = \pi(\mathbf{R}^{(k)}, \lambda_R^{(k)} | \mathbf{Z}, R_{-1}, R_0)$ est défini par (5) et $\langle \cdot \rangle$ désigne la moyenne sur les dix chaînes indépendantes.

Convergence vers la distribution a posteriori. Afin d'illustrer la convergence de la chaîne de Markov générée par l'Algorithme 1 vers la loi cible (5) *quel que soit le point initial*, dix chaînes indépendantes de $5 \cdot 10^7$ itérations sont calculées à partir des comptes d'infections hebdomadaires au Royaume-Uni tracés en noir à la Figure 1 (*haut*) pour trois initialisations différentes. Le coefficient de reproduction est initialisé soit à $R_w^{(0)} = Z_w/\Phi_w^{\mathbf{Z}}$, maximisant la vraisemblance associée au modèle (1), soit à la valeur non informative $R_w^{(0)} = 1$ pour tout $w = 1, \dots, W$; quant à l'hyperparamètre il est initialisé soit à la valeur historique utilisée dans [12, 2, 8] divisée par γ afin de s'adapter au modèle hebdomadaire remis à l'échelle (1), soit à un dixième de cette valeur. Le logarithme de la distribution a posteriori (5) au cours des itérations, moyenné sur les dix chaînes et normalisé par sa valeur maximale, est tracé en Figure 2. Pour les trois initialisations, le logarithme de la distribution cible converge vers la même valeur moyenne et fluctue de la même façon autour de cette valeur. La zone colorée en gris correspond aux dernières $2 \cdot 10^7$ itérations : après une phase de chauffe de $k_{\text{chauffe}} = 3 \cdot 10^7$ itérations la chaîne de Markov s'est effectivement stabilisée en loi, fournissant des échantillons à partir desquels estimer le coefficient de reproduction et les intervalles de crédibilité.

4 Application au suivi de la COVID-19

Données. Dans un premier temps, des données quotidiennes, reportées par l'Université Johns Hopkins,¹ tracées en gris Figure 3 (*haut*), sont agrégées en comptes d'infections hebdomadaires, tracés en noir, auxquels est appliqué l'estimateur Hebdo-Bayes. Afin d'explorer des dynamiques épidémiques et des niveaux de bruit administratif variés, plusieurs phases de la pandémie et différentes régions du monde sont considérées : les États-Unis, entre les semaines du 24 septembre 2020 et du 18 mars 2021 en Figure 3a, la France du 4 novembre 2021 au 28 avril 2022 en Figure 3b, l'Allemagne du 21 juillet 2022 au 12 janvier 2022 en Figure 3c, et Israël du 15 septembre 2022 au 9 mars 2023 en Figure 3d. Enfin, les comptes d'infections reportés de façon hebdomadaire au Royaume-Uni entre la semaine du 15 septembre 2022 et celle du 9 mars 2023, tracés en Figure 1 (*haut*), permettent d'illustrer en conditions réelles l'estimateur du coefficient de reproduction proposé.

Paramètres algorithmiques. Les chaînes de Markov sont toutes initialisées à $R_w^{(0)} = Z_w/\Phi_w^{\mathbf{Z}}$ pour $w = 1, \dots, W$ et $\lambda^{(0)} = 3,5 \times \text{std}(\mathbf{Z})/\gamma$. Elles sont simulées par l'Algorithme 1 avec $k_{\text{max}} = 5 \cdot 10^7$ itérations ; les estimées sont calculées en retirant une période de chauffe de $k_{\text{chauffe}} = 3 \cdot 10^7$ itérations. Enfin, le calcul des intervalles de crédibilité est effectué à partir des quantiles empiriques d'ordre $\varepsilon = 2,5\%$ et $1 - \varepsilon = 97,5\%$.

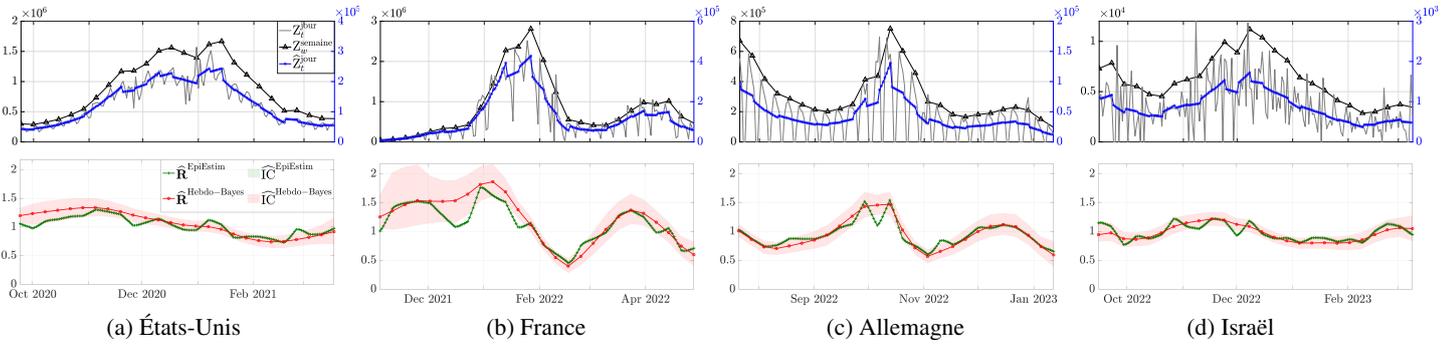


FIGURE 3 : (Haut) Comptes d’infections par la COVID-19 quotidiens (en gris), agrégés à l’échelle de la semaine (en noir), puis ré-interpolés en données quotidiennes (en bleu) [10]. (Bas) Estimées du coefficient de reproduction quotidien via EpiEstim [5] (en vert) vs. hebdomadaire via l’estimateur Hebdo-Bayes proposé (en rouge) avec intervalles de crédibilité à 95% (régions colorées).

Comparaison des estimées et intervalles de crédibilité. Les estimées du coefficient de reproduction et des intervalles de crédibilité quotidiens, via la méthode EpiEstim [5, 10], et hebdomadaires, via la stratégie Hebdo-Bayes proposée, sont présentées en Figure 3 (bas). L’estimateur de moyenne a posteriori quotidien EpiEstim (en traits pleins verts) et l’estimateur hebdomadaire Hebdo-Bayes (en traits pleins rouges) sont en accord pour chacun des quatre exemples de la Figure 3. Il est intéressant de noter que les estimées $\hat{R}^{\text{Hebdo-Bayes}}$ évoluent régulièrement au cours du temps, comportement attendu pour un indicateur épidémiologique en population générale. Au contraire, les estimées $\hat{R}^{\text{EpiEstim}}$ présentent des à-coups ne reflétant pas la dynamique épidémique mais provenant directement des artefacts visibles sur les comptes interpolés par la méthode proposée dans [10] tracés en bleu en Figure 3 (haut). Ainsi, les changements de dynamique épidémique, tels que les passages du coefficient de reproduction au-dessus et en dessous de un ainsi que les pics de $w \mapsto R_w$ sont plus nettement lisibles sur l’estimée $\hat{R}^{\text{Hebdo-Bayes}}$, permettant un suivi de l’épidémie plus précis. En outre, les intervalles de crédibilité $\hat{\text{IC}}^{\text{EpiEstim}}$, en vert en Figure 3 (bas) sont si étroitement concentrés autour de la moyenne a posteriori qu’ils ne sont pas visibles : ils ne rendent pas compte de manière réaliste de l’incertitude majorée engendrée par l’utilisation de comptes interpolés. Par contraste, les intervalles de crédibilité $\hat{\text{IC}}^{\text{Hebdo-Bayes}}$, en rouge en Figure 3 (bas), beaucoup plus larges, témoignent plus fidèlement de l’incertitude sur l’estimation du coefficient de reproduction majorée par l’utilisation de comptes reportés à une échelle de temps plus grossière. La taille variable des intervalles de crédibilité Hebdo-Bayes pour les différents pays et phases de la pandémie montre la capacité de l’estimateur à capter précisément non seulement la dynamique épidémique mais aussi le niveau d’incertitude. Ces conclusions s’appliquent également à l’estimation du coefficient de reproduction sur des données collectées directement à l’échelle de la semaine, comme l’illustre l’exemple du Royaume-Uni en Figure 1 (bas).

5 Conclusion et perspectives

Un modèle bayésien hiérarchique pour la propagation d’une épidémie virale a été proposé. Dans ce modèle, les comptes de nouvelles infections sont rapportés de façon hebdomadaire et le coefficient de reproduction estimé est donc lui-même hebdomadaire. Plusieurs paramètres du modèle sont fixés a priori, c’est le cas de la moyenne et de l’écart-type de la fonc-

tion d’intervalle de série, ainsi que du paramètre d’échelle de la loi de Poisson. Une estimation pilotée par les données de ces paramètres, tirant parti des données COVID-19 mises à disposition par l’Université Johns Hopkins couvrant trois ans de pandémie dans plus de deux-cent pays, est à l’étude.

Références

- [1] P. ABRYS, J. CHEVALLIER, G. FORT et B. PASCAL : Hierarchical Bayesian Estimation of COVID-19 Reproduction Number. *In Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Hyderabad, India, 2025.
- [2] P. ABRYS, G. FORT, B. PASCAL et N. PUSTELNIK : Estimation et intervalles de crédibilité pour le taux de reproduction de la Covid19 par échantillonnage Monte Carlo Langevin proximal. Nancy, France, Sept. 6 - 9 2022.
- [3] D. CEREDA *et al.* : The early phase of the COVID-19 outbreak in Lombardy, Italy. *Preprint arXiv :2003.09320*, 2020.
- [4] K. CHARNIGA *et al.* : Spatial and temporal invasion dynamics of the 2014–2017 Zika and chikungunya epidemics in Colombia. *PLoS Comput. Biol.*, 17(7):e1009174, 2021.
- [5] A. CORI, N. M. FERGUSON, C. FRASER et S. CAUCHEMEZ : A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.*, 178(9):1505–1512, 2013.
- [6] N. M. FERGUSON *et al.* : Countering the Zika epidemic in Latin America. *Science*, 353(6297):353–354, 2016.
- [7] A. FLAHAULT : Covid-19 cacophony : is there any orchestra conductor? *The Lancet*, 395(10229):1037, 2020.
- [8] G. FORT, B. PASCAL, P. ABRYS et N. PUSTELNIK : Covid19 reproduction number : Credibility intervals by blockwise proximal monte carlo samplers. *IEEE Trans. Signal Process.*, 71:888–900, 2023.
- [9] E. KRYMOVA *et al.* : Trend estimation and short-term forecasting of COVID-19 cases and deaths worldwide. *Proc Natl Acad Sci U S A*, 119(32):e2112656119, 2022.
- [10] R. K. NASH, S. BHATT, A. CORI et P. NOUVELLET : Estimating the epidemic reproduction number from temporally aggregated incidence data : A statistical modelling approach and software tool. *PLoS Comput. Biol.*, 19(8):e1011439, 2023.
- [11] R. K. NASH, P. NOUVELLET et A. CORI : Real-time estimation of the epidemic reproduction number : Scoping review of the applications and challenges. *PLOS Digital Health*, 1(6):e0000052, 2022.
- [12] B. PASCAL *et al.* : Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data. *IEEE Trans. Signal Process.*, 70:2859–2868, 2022.
- [13] B. PASCAL et S. VAITER : Risk Estimate under a Nonstationary Autoregressive Model for Data-Driven Reproduction Number Estimation. *Preprint arXiv :2409.14937*, 2024.
- [14] F. RICCARDO *et al.* : Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. *Euro Surveill.*, 2020.
- [15] J. WALLINGA et P. TEUNIS : Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *Am. J. of Epidemiol.*, 160(6):509–516, 09 2004.