

Analyse de la transmissibilité du variant B.1.1.7 en France

François Blanquart^{1,2}

¹Centre for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France

²Infection Antimicrobials Modelling Evolution, UMR 1137, INSERM, Université de Paris, Paris, France

J'infère l'avantage de transmissibilité du variant B.1.1.7, appelé « VOC » (Variant of Concern) avec un modèle épidémiologique. L'avantage inféré est +63% en se basant sur les données agrégées au niveau national, et +44% en analysant les données par régions. Il est inféré que le VOC deviendra majoritaire dans les cas détectés au 20 février (analyse au niveau national), ou bien du 22 février au 30 mars (analyse par régions).

Inférence nationale :

En utilisant les **données agrégées au niveau national, l'avantage de transmission inféré du VOC est de +63%** (intervalle de confiance 95% [+56, +70]). L'enquête Flash coordonnée par Bruno Lina et coll. estime la fréquence du VOC dans les cas COVID des 7-8 janvier et 27 janvier 2021. La fréquence estimée du VOC est de 3.3% au 7-8 janvier et de 13.2% (388/11916) au 27 janvier (474/3601). Le chiffre de 13.2% correspond au criblage Thermo Fisher (non-amplification du gène S). Sous un modèle simple où le VOC et le non VOC (appelé ensuite « Wild Type » WT) augmentent exponentiellement (avec un taux de croissance exponentiel qui peut être négatif, positif ou nul), le logit de la fréquence augmente linéairement à un rythme égal à la différence de taux de croissance exponentiel entre le VOC et le WT :

$$\frac{d[\text{logit}(p)]}{dt} = r^{VOC} - r^{WT}$$

Sous l'hypothèse que le taux de croissance du WT est environ $r^{WT} \approx 0$, on a donc :

$$\frac{d[\text{logit}(p)]}{dt} = r^{VOC}$$

Ainsi, le taux de croissance est le coefficient de régression d'une régression logistique de la présence du VOC par rapport à la date : $r^{VOC} = 0.077$ par jour (intervalle de confiance 95% [0.070-0.084]). A un taux de croissance et une distribution du temps de génération, on peut associer le nombre de reproduction sous l'hypothèse que la structure en âge d'infection est stabilisée (1). Avec une distribution du temps de génération gamma avec paramètres 7.8 et 0.833 (moyenne 6.5 jours, déviation standard 2.3 jours), cela donne :

$$R^{VOC} = 1.63 [1.56 - 1.70]$$

L'avantage de transmissibilité du VOC est donc de +63%. Ce chiffre est similaire à celui inféré en Angleterre (2,3). Ce modèle prédit que le variant deviendra majoritaire en France (fréquence 50%) dans les cas détectés le 20 février 2021 (individus infectés aux alentours du 13 février) (Figure 1).

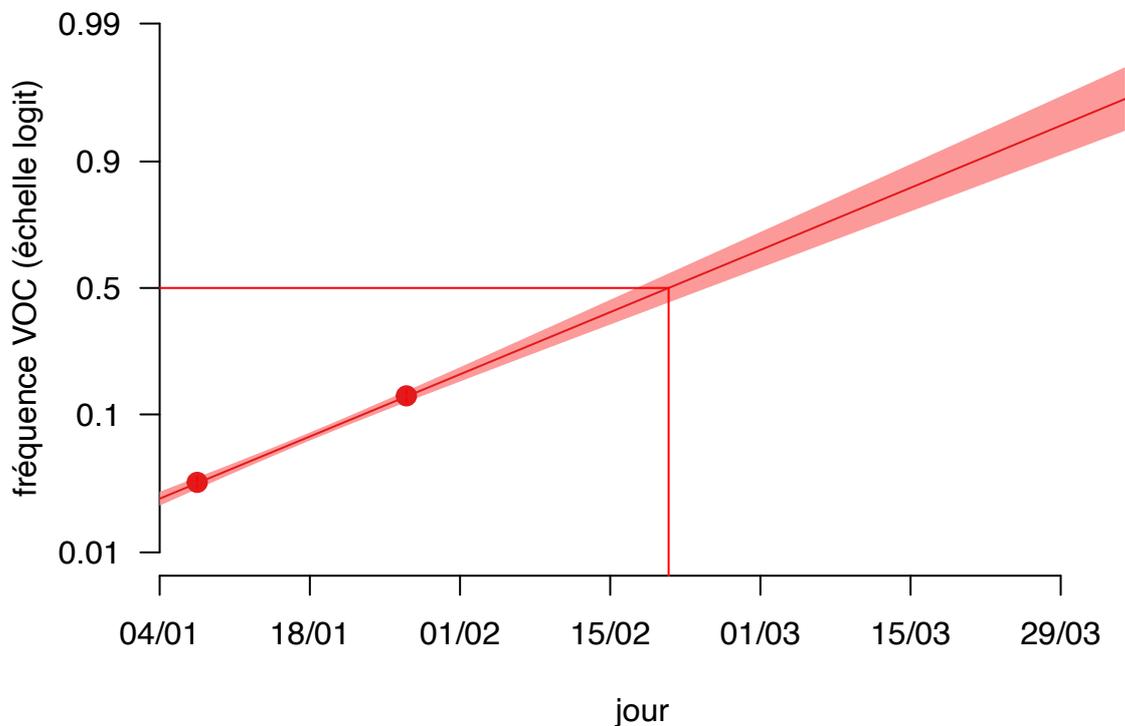


Figure 1 : Fréquence du VOC parmi les cas détectés sur une échelle logit, modèle (droite et intervalle de confiance) et données de l'enquête Flash (points).

Inférence prenant en compte les données par régions :

Dans un deuxième temps, l'inférence se base sur le nombre de cas journaliers par région en France reportés par Santé Publique France, et sur les données de fréquence du VOC de l'enquête Flash ainsi que l'enquête APHP en région parisienne du 11 au 21 janvier (Anne-Geneviève Marcelin et coll.). L'inférence est basée sur un modèle exponentiel simple qui décrit la croissance du nombre de cas VOC et WT. Dans cette analyse qui pondère différemment les différentes régions par rapport à l'analyse des données agrégées au niveau national, **l'avantage de transmissibilité inféré du VOC est +44% [intervalle de crédibilité 95% 39-49%]** sur le nombre de reproduction.

Le nombre de reproduction inféré des non-VOC étant proche de 1 (compris entre 0.92 en Bourgogne Franche Comté et 1.09 en Pays de la Loire), **le VOC augmente en incidence dans toutes les régions** grâce à la transmissibilité accrue (figure 2). Le VOC augmente en fréquence (figure 3), deviendra majoritaire de la fin-février à fin-mars selon les régions (Table). Lorsque le variant sera dominant, le nombre de cas progressera exponentiellement au rythme dicté par le nombre de reproduction du variant (figure 4).

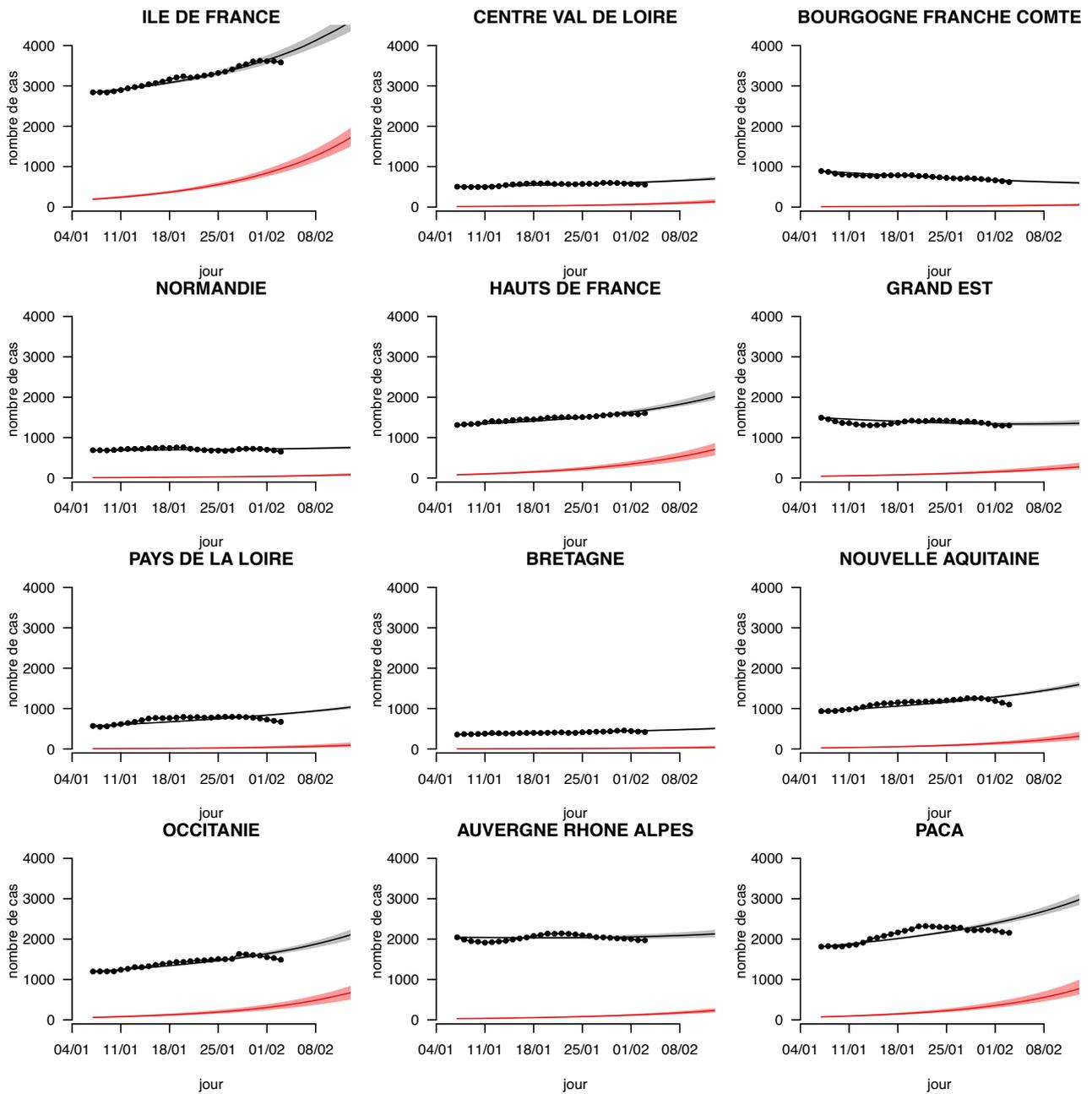


Figure 2 : Nombre de cas journaliers totaux par région (noir ; points = données, ligne et enveloppe = modèle) et cas VOC inférés (rouge, modèle).

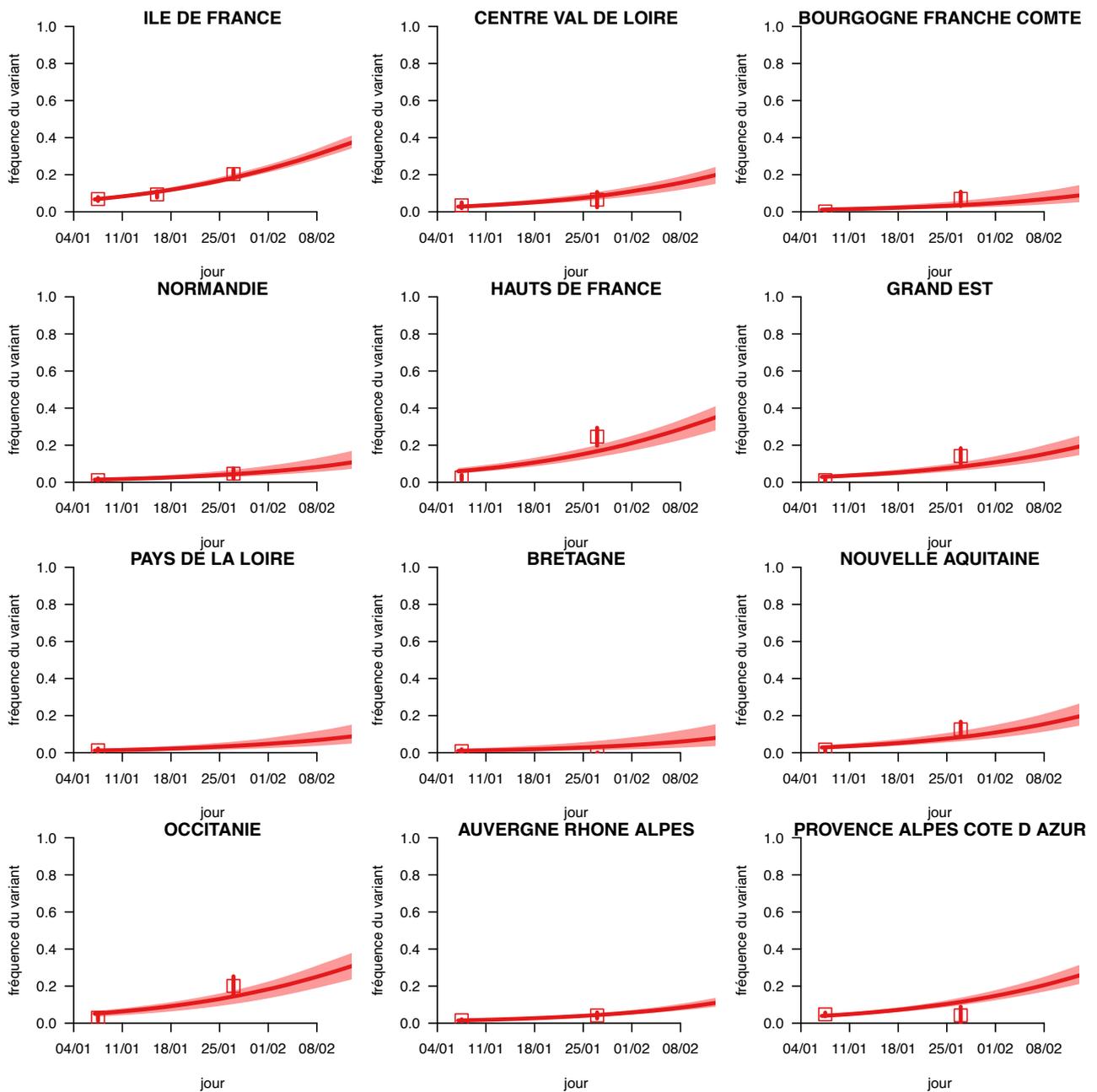


Figure 3 : Fréquence du VOC parmi les cas inférée (ligne) avec les points de données (enquêtes Flash et APHP) (carrés avec intervalle de confiance binomial 95%).

D'après ce modèle, la variant atteindra une fréquence de 50% des cas à ces dates :

Région	Médiane	Intervalle 2.5% - 97.5%	
ILE DE FRANCE	22/02/2021	19/02/2021	26/02/2021
CENTRE VAL DE LOIRE	10/03/2021	04/03/2021	20/03/2021
BOURGOGNE FRANCHE COMTE	26/03/2021	16/03/2021	09/04/2021
NORMANDIE	24/03/2021	15/03/2021	04/04/2021
HAUTS DE FRANCE	24/02/2021	20/02/2021	04/03/2021
GRAND EST	10/03/2021	01/03/2021	18/03/2021
PAYS DE LA LOIRE	25/03/2021	11/03/2021	09/04/2021
BRETAGNE	30/03/2021	14/03/2021	15/04/2021
NOUVELLE AQUITAINE	10/03/2021	01/03/2021	19/03/2021
OCCITANIE	27/02/2021	22/02/2021	06/03/2021
AUVERGNE RHONE ALPES	21/03/2021	15/03/2021	29/03/2021
PROVENCE ALPES COTE D'AZUR	05/03/2021	27/02/2021	11/03/2021

Cette analyse simple a un certain nombre de limites :

- Les données de l'enquête Flash ne sont pas encore consolidées. Les inférences et projections pour certaines régions peuvent être fragiles. En particulier, la fréquence du VOC dans les cas au 27 janvier se base sur la non-amplification du gène S dans la PCR Thermo Fisher. Tous ces cas ne sont pas forcément des VOC.
- Le nombre de reproduction est considéré constant sur la période du 7 janvier à début février. Ce nombre de reproduction change en réalité. Par exemple, en région PACA, on observe une période de hausse des cas suivie par une stabilisation qui est mal capturée par le modèle. Si le nombre de reproduction diminue, le variant progressera également moins rapidement en incidence.
- La modélisation des données de fréquence du variant comme une distribution binomiale peut sous-estimer la variabilité de ces données et donc sous-estimer les intervalles de crédibilité.

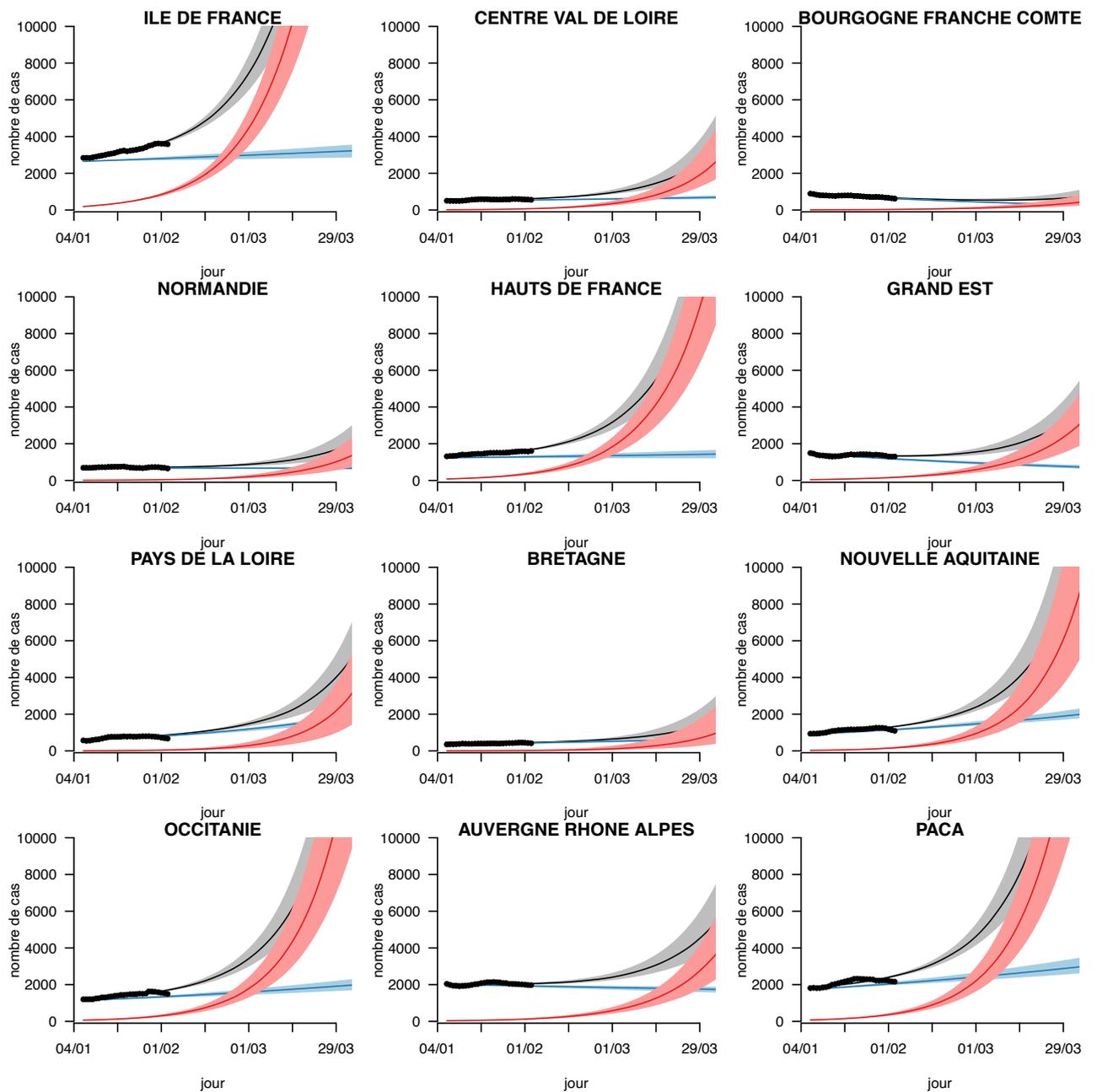


Figure 4 : Nombre de cas journaliers totaux par région extrapolés à plus long terme (noir ; points = données, ligne et enveloppe = modèle) et cas VOC (rouge, modèle) et non-VOC (bleu, modèle).

Détails du modèle et de la procédure d'inférence pour l'analyse par régions :

Le modèle suppose que le nombre de reproduction effectif des souches originales dans la région i est R_i^{WT} (WT = Wild Type) et celui du VOC est $R_i^{VOC} = R_i^{WT} (1 + \alpha)$ où α est l'avantage de transmissibilité. Par ailleurs, la fréquence du VOC dans la région i au 7 janvier 2021 est $p_{i,0}$. On infère les paramètres R_i^{WT} , α et $p_{i,0}$.

Le nombre total de cas à la date t est noté $N_{i,t}$, qui se décompose en $N_{i,t}^{WT}$ (wild type) et $N_{i,t}^{VOC}$ (VOC).

Aux nombres de reproduction R et une distribution de temps de génération $g(a)$ sont associés des taux de croissance exponentiels, r , solution de cette équation :

$$R \int_0^{\infty} e^{-ra} g(a) da = 1$$

La distribution du temps de génération $g(a)$ est gamma avec paramètres 7.8 et 0.833 (moyenne 6.5 jours, déviation standard 2.3 jours).

Ainsi, les nombre de cas générés par le WT et le VOC au temps t sont donnés par :

$$\begin{aligned} N_{i,t}^{WT} &= N_{i,0}^{WT} e^{r_i^{WT} t} = (1 - p_{i,0}) N_{i,t} e^{r_i^{WT} t} \\ N_{i,t}^{VOC} &= N_{i,0}^{VOC} e^{r_i^{VOC} t} = p_{i,0} N_{i,t} e^{r_i^{VOC} t} \end{aligned}$$

Ce modèle est ajusté par une méthode Bayésienne. La vraisemblance se base sur deux composantes :

- La probabilité du nombre de cas quotidiens observés par région, modélisé par une distribution négative binomiale avec pour moyenne le nombre de cas prédit $N_{i,t}$. Nous utilisons le nombre de cas moyenné sur sept jours glissants.
- La probabilité des fréquences de variants observés dans les enquêtes Flash, modélisées par une distribution binomiale (cette hypothèse pourrait être relâchée pour prendre en compte la surdispersion liée à l'hétérogénéité spatiale ou une variabilité technique, qui pourrait être modélisée par une loi beta-binomiale).

Les distributions *a priori* des paramètres sont :

- Distribution beta avec moyenne 0.033 et déviation standard 0.02 pour les fréquences régionales du VOC au 7 janvier (0.033 correspond à la moyenne nationale à cette date) (paramètre $p_{i,0}$).
- Distribution normale avec moyenne 1 et déviation standard 0.05 pour le nombre de reproduction du *wild type*.
- Distribution normale avec moyenne 0.5 et déviation standard 0.2 pour l'avantage sélectif du VOC (paramètre α).
- Distribution uniforme entre 0.1 et 1000 pour le paramètre de dispersion de la distribution négative binomiale.

La distribution postérieure est échantillonnée par MCMC (15000 itérations, les itérations 1500 à 15000 sont retenues).

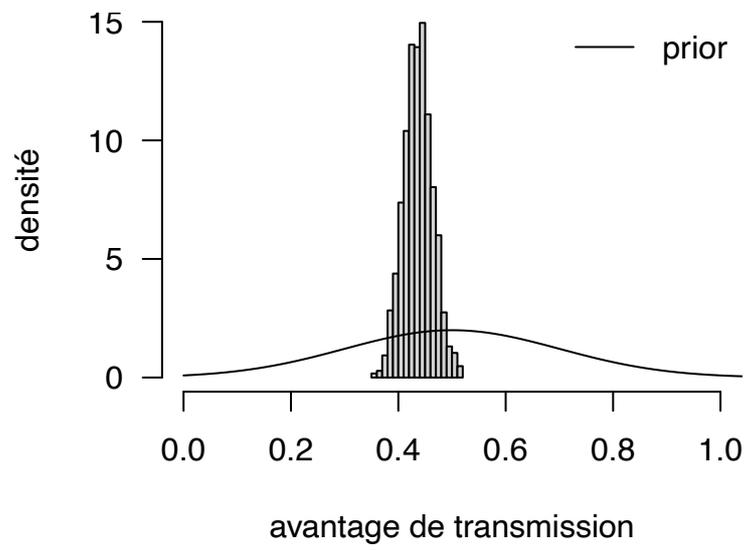


Figure 5 : Distribution postérieure de l'avantage de transmissibilité (histogramme) et prior (courbe)

Sources de données :

Cas journaliers Santé Publique France : <https://www.data.gouv.fr/fr/datasets/taux-dincidence-de-lepidemie-de-covid-19/>

Données des enquêtes Flash 1 et 2 : points épidémiologiques hebdomadaires Santé Publique France du 14 janvier, 28 janvier (résultats consolidés), 4 février et 11 février:

<https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-14-janvier-2021>

<https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-28-janvier-2021>

<https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-4-fevrier-2021>

<https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-11-fevrier-2021>

Données de l'enquête APHP : communiquées par exemple ici :

<https://www.nouvelobs.com/coronavirus-de-wuhan/20210126.OBS39397/le-variant-anglais-represente-desormais-environ-10-des-cas-en-ile-de-france-s-inquiete-l-ap-hp.html>

Sites web visités le 10 février 2021.

Remerciements : je remercie les personnes ayant collecté et mis à disposition les données utilisées ici, ainsi que Samuel Alizon et Florence Débarre pour leurs commentaires.

Référence

1. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc B Biol Sci. 2007 Feb 22;274(1609):599–604.
2. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Transmission of SARS-CoV-2 Lineage B. 1.1. 7 in England: Insights from linking epidemiological and genetic data. medRxiv. 2021;2020–12.
3. Davies NG, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, Pearson CAB, et al. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. medRxiv. 2020 Dec 26;2020.12.24.20248822.