



## An ensemble model based on early predictors to forecast COVID-19 healthcare demand in France

Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Maylis Layan, Pascal Crepey, Alix Roumagnac, Marc Lavielle, Pierre-Yves Boëlle, Simon Cauchemez

### ► To cite this version:

Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Maylis Layan, Pascal Crepey, et al.. An ensemble model based on early predictors to forecast COVID-19 healthcare demand in France. 2021. pasteur-03690824v1

**HAL Id: pasteur-03690824**

**<https://pasteur.hal.science/pasteur-03690824v1>**

Preprint submitted on 22 Feb 2021 (v1), last revised 8 Jun 2022 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# An ensemble model based on early predictors to forecast COVID-19 healthcare demand in France

Juliette Paireau<sup>1,2†</sup>, Alessio Andronico<sup>1†</sup>, Nathanael Hozé<sup>1</sup>, Maylis Layan<sup>1</sup>, Pascal Crépey<sup>3</sup>, Alix Roumagnac<sup>4</sup>, Marc Lavielle<sup>5,6</sup>, Pierre-Yves Boëlle<sup>7</sup>, Simon Cauchemez<sup>1\*</sup>

## Affiliations

<sup>1</sup>Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, UMR 2000, CNRS, Paris, France.

<sup>2</sup>Direction des Maladies Infectieuses, Santé Publique France, French National Public Health Agency, Saint Maurice, France.

<sup>3</sup>Univ Rennes, EHESP, REPERES « Recherche en Pharmaco-Epidémiologie et Recours aux Soins » – EA 7449, Rennes, France.

<sup>4</sup>PREDICT Services, Castelnau-le-Lez, France.

<sup>5</sup>INRIA, Saclay, France.

<sup>6</sup>CMAP, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, Palaiseau, France.

<sup>7</sup>INSERM, Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris, France.

†These authors contributed equally to this work.

\*Corresponding author. Email: [simon.cauchemez@pasteur.fr](mailto:simon.cauchemez@pasteur.fr)

## Abstract

Short-term forecasting of the COVID-19 pandemic is required to facilitate the planning of COVID-19 healthcare demand in hospitals. Here, we evaluate the performance of 12 individual models and 22 predictors to anticipate French COVID-19 related healthcare needs from September 7th 2020 to January 7th 2021, and build an ensemble model that outperforms all individual models. We find that inclusion of early predictors (epidemiological, mobility and meteorological predictors) can halve the relative error for 14-day ahead forecasts, with epidemiological and mobility predictors contributing the most to the improvement. Our approach facilitates the comparison and benchmarking of competing models through their integration in a coherent analytical framework, ensuring avenues for future improvements can be identified.

## **Main text**

### **Introduction**

Quick increase in hospital and Intensive Care Unit (ICU) admissions have been common since the start of the COVID-19 pandemic. In many instances, this has put the healthcare system at risk of saturation, forced the closure of non-covid-19 wards, cancellation of non essential surgeries, reallocation of staff to COVID-19 wards with dramatic consequences for non COVID-19 patients. In this context, short-term forecasting of the pandemic and its impact on the healthcare system is required to facilitate the planning of COVID-19 and other activities in hospitals (1).

Hospital admission comes late in the history of infection of a patient, so forecasts that only rely on hospital data may miss earlier signs of a change in epidemic dynamics. There have been a lot of discussions about insights we might gain from other types of predictors (e.g. epidemiological predictors such as the number of cases, mobility predictors such as Google data or meteorological predictors) but assessment of the contribution of these predictors have been marred by methodological difficulties. For example, while variations in case counts may constitute an earlier sign of change in epidemic dynamics, these data may be affected by varying testing efforts, making interpretation difficult. Associations between meteorological/mobility variables and SARS-CoV-2 transmission rates have been identified (2–5); but it is yet unknown whether use of these data along with epidemiological predictors may improve forecasts.

Here, we develop a systematic approach to address these challenges. We retrospectively evaluate the performance of 12 individual models and 22 predictors to anticipate French COVID-19 related healthcare needs, from September 7th 2020 to January 7th 2021, and build an ensemble model that outperforms all individual models. Our analysis makes it possible to determine the most promising approaches and predictors to forecast COVID-19 related healthcare demand, indicating for example that inclusion of early predictors (epidemiological, mobility and meteorological predictors) can halve the relative error for 14-day ahead forecasts, with epidemiological and mobility predictors contributing the most to the improvement. Our approach facilitates the comparison and benchmarking of competing models through their integration in a coherent analytical framework, ensuring avenues for future improvements can be identified.

## Results

### *Performance of individual models to forecast hospital admissions*

Twelve individual models are considered to forecast the number of hospital admissions by region with a time horizon of up to two weeks. They use a variety of methods and can rely on epidemiological, mobility and meteorological predictors (see Material and Methods, Supplementary text, Fig. S1, Fig. S2 and Fig. S3). Over the evaluation period, most of these models are able to broadly capture the growth of the epidemic in September and October, the decrease in November, and the plateau in December (Fig. 1). They all over-estimate the peak since the lockdown date could not have been anticipated.

Overall, the performance of the models decreases with the prediction horizon. The mean relative error for hospital admissions ranges from 4-5% at  $t-1$  to 17-61% at  $t+14$  at the national level depending on the model, and from 6-8% to 25-77% at the regional level (Fig. 2). At the national level, two groups of models can be distinguished: six models are below 25% of relative error at 14 days, while the other six are above 30%. Similarly, at the regional level, the first group of models performs better (relative error 25-32% at 14 days) than the second group (relative error >38%). All models in the first group describe the growth rate of hospital admissions, rather than hospital admissions directly, and include several predictors that are described below.

### *Performance of the ensemble model*

To build the ensemble model, we keep the six best performing models (an autoregressive distributed lag model (ARDL), a multiple linear regression model (MLR), a generalized additive model (GAM2), an ARIMA model (ARIMA2), a boosted regression tree (BRT) and a random forest (RF)).

On average, the ensemble model performs better than all individual models at the regional and national level (Fig.2). For hospital admissions, the relative error is 14% at 7 days and 16% at 14 days at the national level, and 19% at 7 days and 25% at 14 days at the regional level. The forecasts are well calibrated for the regional level, with 95% prediction intervals capturing 96% and 97% of observations at 7 and 14 days, respectively (Table S1). The prediction intervals are more conservative at the national level, with coverage of 100% at 7 and 14 days.

For the other targets, the relative errors at 7 days are 12%, 6% and 10% at the national level (21%, 14% and 11% at the regional level) for ICU admissions, ICU beds and general ward beds,

respectively (Table S1). At 14 days, these errors increase to 17%, 11% and 16% at the national level (25%, 20% and 19% at the regional level), respectively. The calibration is good, except for general ward beds at the national level, where the coverage is 87% at 7 days.

For each region, we then rank the individual and ensemble models according to the relative error averaged over all prediction horizons (Fig. 3). The best individual model is not the same in all regions (five models are ranked first in at least one region), but the ensemble model is ranked first on average across all regions.

### *Predictors*

The six best individual models include between 2 and 4 predictors (Table S2). One model has an autoregressive component, i.e. includes lagged values of the growth rate of hospital admissions as covariates. All six models include at least one mobility predictor: the number of visitors to transit stations (in percent change from baseline) is the one that is most often selected, followed by time spent at home (“residential” category). Five models use at least one predictor on confirmed cases: the growth rate of the proportion of positive tests among tests in symptomatic people, and/or the growth rate of the number of positive tests. Three models use one meteorological predictor, either absolute humidity or IPTCC, an index characterizing climatic conditions favorable for the transmission of COVID-19 (6).

Retrospectively, the models are able to reproduce reasonably well the dynamics of the growth rate over time and by region (Fig. S4). Depending on the model, the most important predictors are mobility or epidemiological predictors (Fig. S5). For instance, in the GAM2 model, the change in time spent in residential places contributes to 51% of the explained variance, followed by the growth rate of the positive tests (35%). In the BRT model, the growth rate of the number of positive tests is the most important predictor (relative contribution of 85%) followed by the change in the volume of visits to transit stations (15%). Meteorological factors contribute to 13% in the GAM2 and 5% in the MLR.

We find that an increase in the volume of visits to transit stations, a decrease in the time spent at home, or an increase in the IPTCC, is associated with an increase in the growth rate of hospital admissions 10-12 days later (Fig. 4). Regarding epidemiological predictors, the growth rate of hospital admissions is positively associated with the growth rate of the number of positive tests, with a lag of 4 days, and the growth rate of the proportion of positive tests in symptomatic people, with a lag of 7 days.

## Discussion

We can draw a number of important conclusions from the systematic evaluation and comparison of 12 models and 22 predictors to forecast COVID-19 healthcare demand. First, mathematical models are often calibrated on hospitalization and death data only, as these signals are expected to be more stable than testing data (1, 7, 8). However, we find that such an approach is outperformed by models that also integrate other types of predictors. These include predictors that can detect more quickly a change in the epidemic dynamics (e.g. growth rate in the proportion of positive tests in symptomatic people) or that may be correlated with the intensity of transmission (e.g. mobility data, meteorological data). Inclusion of such predictors can halve the relative error for a time horizon of 14 days.

Second, of the three types of predictors, epidemiological and mobility predictors are those that improve forecasts the most. In models where the lags are estimated, we find that epidemiological predictors precede the growth rate of hospital admissions by 4-7 days, while mobility predictors precede it by 12 days. This is consistent with our understanding of the delays from infection to testing and infection to hospitalization (7). Meteorological variables also improve forecasts, although the reduction in the relative error is more limited. Per se, this result should not be used to draw conclusions on the role of climate on SARS-CoV-2 transmission. Indeed, we are only assessing the predictive power of these variables, not their causal effect, in a situation where the hospitalization dynamics is already well captured by epidemiological predictors. In this context, the additional information brought by meteorological variables is limited and might already be accounted for by epidemiological predictors. Interestingly, despite the diversity of models and retained predictors, estimates of the effect of the different predictors on the growth rate are relatively consistent across models (Fig. 4).

Third, rather than using the individual model that performs best, we find that it is better to rely on an ensemble model that averages across the best performing models. This is consistent with results of a number of recent epidemic forecasting challenges (1, 9–12). Relying on an ensemble model is appealing because it acknowledges that each model has limitations and imperfectly captures the complex reality of this pandemic. Forecasts that build on an ensemble of models are less likely to be overly influenced by assumptions of a specific model. The benefits are confirmed in practice, with the ensemble model performing best on average.

Fourth, the systematic evaluation also shed light on important technical lessons for forecasting. We find that the best forecasts are obtained when using the exponential growth rate rather than the absolute value of epidemiological variables. This is true for the dependent variable we aim to forecast (hospital admissions) but also for explanatory variables (e.g. the proportion of positive tests). This finding is not surprising since transmission dynamics are characterized by exponential growth and decline. Using the growth of epidemiological predictors such as the number of positive tests also helps controlling for changes in testing practice that may have occurred over longer time periods. We also find that the approach used to smooth the data is decisive to ensure that forecast quality is not overly dependent on the day of the week (given the existence of important weekend effects); and to find the correct balance between early detection of a change of dynamics and the risk of repeated false alarms (Materials and Methods and Fig. S6).

The emergence of variants such as B.1.1.7, that are more transmissible than historical SARS-CoV-2 viruses (13), opens up new challenges for the forecasting of COVID-19 healthcare demand. Indeed, our models have been calibrated on past data to forecast the epidemic growth rate of the historical virus from a number of predictors. It may underestimate the epidemic growth rate in a context where a more transmissible variant is also circulating. The scale of the problem likely depends on the predictors used. Higher transmissibility of B.1.1.7 should equally impact the growth rate of all epidemiological variables; so that performance of models that only rely on epidemiological predictors may be little affected. The impact may be larger when mobility and meteorological predictors are also being used, since the association between these predictors and the epidemic growth rate might be modified by B.1.1.7 and other variants. Once we have accumulated sufficient data, it will be possible to explicitly integrate the proportion of variants as a new predictor of the models. In the meantime, we can run sensitivity analyses with models that use epidemiological predictors only, with no consideration of mobility and meteorological data. On historical data, this will slightly increase relative errors but may be more robust to a change in dynamics due to the rise of B.1.1.7 and other variants. The vaccination of the population may also modify the association between the different covariates and require a refinement of the models to account for the vaccine coverage, in particular in at-risk individuals.

Through a systematic evaluation, we determined the most promising approaches and predictors to forecast COVID-19 related healthcare demand. Our framework makes it straightforward to compare and benchmark competing models, identify current limitations and avenues for future improvements.

## Material and Methods

### *Hospitalization data*

Hospital data are obtained from the SI-VIC database, the national inpatient surveillance system providing real time data on the COVID-19 patients hospitalized in French public and private hospitals (Supplementary text).

### *Smoothing*

Hospital data follow a weekly pattern, with less admissions during weekends compared to weekdays, and can be noisy at the regional level. In order to remove the noise and prevent forecasts from being biased depending on the day of the week at which the analysis is performed, we remove this weekly pattern and smooth the data using local polynomial regression and the least revision principle (Proietti 2008) (Supplementary text and Figure S6).

### *Exponential growth rate*

We compute the exponential growth rate using a 2-day rolling window, and smooth the resulting time series using local polynomial regression.

### *Overview of the modelling approach*

We build a framework to forecast at the national (metropolitan France) and regional (N=12) levels four targets up to 14 days ahead: the daily numbers of hospital and ICU admissions, and the daily numbers of beds occupied in general wards and ICU.

We first develop a set of individual models to forecast the number of hospital admissions, using a variety of methods. These individual predictions are then combined into a single ensemble forecast, called an ensemble model (1, 10, 12, 14–16). Finally, we derive the number of ICU admissions, and bed occupancy in general wards and ICU from the predicted number of hospital admissions.

### *Modelling hospital admissions*

We evaluate 12 individual models, including exponential growth models with constant or linear growth rates, linear regression models, generalized additive models (GAM), boosted regression trees (BRT), random forests (RF), neural networks auto-regressive models (NNAR), and



autoregressive integrated moving average models (ARIMA) (full description in Supplementary text). Some of the models directly predict the number of hospital admissions, while others predict the growth rate, from which hospital admissions are then derived using an exponential growth model.

Model performance is evaluated using the mean relative error (mean of the ratio of the absolute error to the observed value). We also assess the prediction interval coverage (proportion of prediction intervals that contain the observed value). The evaluation period runs from September 7th 2020 to January 7th 2021, excluding the first two weeks of the second national lockdown (October 30th 2020 to November 13th 2020) as the models are not designed to anticipate the impact of a lockdown before its implementation. We use a cross-validation approach: for each day  $t$  of the evaluation period, we make forecasts for the period  $t-1$  up to day  $t+14$ , using all data up to day  $t-2$  and after day  $t+14$  as a training set, and computing prediction error using the observed data in  $t-1$  to  $t+14$ . We start to make forecasts at  $t-1$  because in real-time, on day  $t$ , values at  $t$  and  $t-1$  are not consolidated yet and the last reliable data point used for forecasts is the value at  $t-2$ .

We keep the 6 best individual models to build the ensemble model. Individual model forecasts are combined into an ensemble forecast by taking the unweighted mean of the point predictions, and the unweighted mean of the 95% confidence intervals.

We include in individual models a set of predictors, chosen for their availability in near real-time and their potential to help to anticipate the trajectory of hospital admissions. Three types of predictors are considered: 9 epidemiological predictors describing the dynamics of the epidemics (for example, growth rate of the number of hospital admissions, of the number of positive tests, of the proportion of positive tests among symptomatic people,...), 9 mobility predictors (for example the change in volume of visits to workplaces, transit stations, residential places or parks (Google data), or the change in volume of requests for itineraries on Apple map App) and 4 meteorological predictors (temperature, absolute and relative humidity, and IPTCC, an index characterizing climatic conditions favorable for the transmission of COVID-19 (6)). All predictors and data sources are described in Supplementary text and Figures S1, S2 and S3. For each individual model, covariates are selected using a forward stepwise selection approach (Supplementary text).

In order to determine the importance of the different predictors and explore their effect on the growth rate, we retrospectively fit the best individual models over the entire time period, on all regions together, whenever possible. The parameters of the ARDL model vary with the prediction horizon and those of the ARIMA2 model vary by region. Therefore, four models (BRT, RF, MLR and GAM2) are included in this sub-analysis.

#### *Modelling ICU admissions, bed occupancy in ICU and general wards*

Predictions for the number of ICU admissions, and bed occupancy in ICU and general wards are derived from predicted numbers of hospital admissions. The expected number  $H_{ICU}(t)$  of ICU admissions at time  $t$  is given by the formula:

$$H_{ICU}(t) = p_{ICU}(t) \sum_{u \leq t} H(u) g_H(t-u)$$

where  $H(u)$  is the number of hospital admissions at time  $u$ ,  $p_{ICU}(t)$  is the probability of being admitted to ICU once in hospital, and  $g_H(t-u)$  is the delay distribution from hospital to ICU admission, i.e. the probability for an individual who entered the hospital on day  $u$  to have a delay of  $d = t-u$  days before ICU admission. We compute  $g_H(t-u)$  assuming that the delay from hospital to ICU admission is exponentially distributed with mean = 1.5 days (7). The probability of ICU admission in time interval  $T = [t_1, t_2]$  can be estimated as follows:

$$p_{ICU}(T) = \frac{\sum_{t \in T} H_{ICU}(t)}{\sum_{t \in T} \sum_{u \leq t} H(u) g_H(t-u)}$$

In practice, we estimate  $p_{ICU}$  on a 10-day rolling window, which makes the estimates relatively stable.

The number of general ward ( $B_H(t)$ ) and ICU ( $B_{ICU}(t)$ ) beds occupied by COVID-19 patients at time  $t$  can be expressed as:

$$B_H(t) = (1 - p_{ICU}(t)) \sum_u H(t-u) s_H(u) + p_{ICU}(t) \sum_u H(t-u) s_{H-ICU}(u)$$

$$B_{ICU}(t) = \sum_u H_{ICU}(t-u) s_{ICU}(u)$$

where  $s_H(u)$  is the probability to stay in hospital for  $u$  days before discharge,  $s_{H-ICU}(u)$  is the probability to spend  $u$  days in the hospital general ward and then move to the ICU, and  $s_{ICU}(u)$  is the probability to spend  $u$  days in ICU. For  $s_H$  and  $s_{ICU}$  we use gamma survival functions with mean to be estimated and coefficient of variation (i.e. standard deviation over mean) fixed at 0.9 for  $s_H$  and 0.8 for  $s_{ICU}$ , while, analogously to  $g_H$ , we take  $s_{H-ICU}(u)$  to be an exponential survival function with a mean of 1.5 days. We minimize the sum of squared errors over the last 5 data points to estimate the free parameters of  $s_H$  and  $s_{ICU}$ . Once all the parameters are estimated and the forecast of  $H$  are available, we use the equations above to forecast  $H_{ICU}$ ,  $B_H$ , and  $B_{ICU}$ , assuming that all parameters' estimates remain constant.

To account for uncertainty in parameter estimates, we use the lower and upper bounds of the smoothed trajectory of hospital admissions at the national level, to generate prediction intervals for the three other targets. For the regional level, this approach gives too narrow prediction intervals. We obtain better coverage by bootstrapping the smoothed trajectories of hospital data.

## Acknowledgments

We are grateful to the hospital staff and all the partners involved in the collection and management of SIVIC data. We thank Raphaël Bertrand and Eurico de Carvalho Filho (PREDICT Services) for providing meteorological data. We thank Google and Apple for making their mobility data available online.

## Funding sources

We acknowledge financial support from the Investissement d'Avenir program, the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases program (grant ANR-10-LABX-62- IBEID), Santé Publique France, the INCEPTION project (PIA/ANR16-CONV-0005), and the European Union's Horizon 2020 research and innovation program under grants 101003589 (RECOVER) and 874735 (VEO), AXA.

## Competing interests

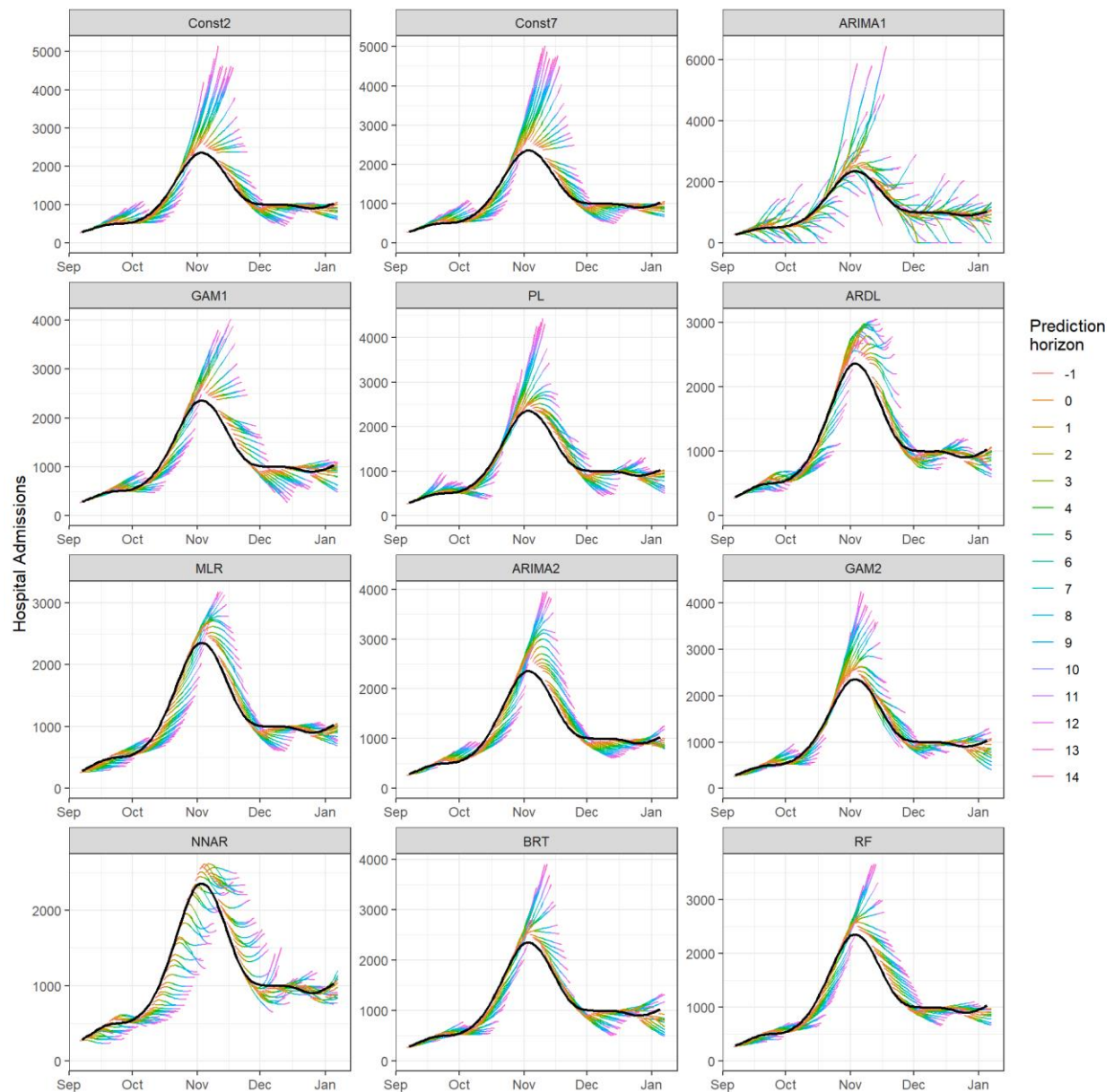
Authors declare that they have no competing interests.

## References

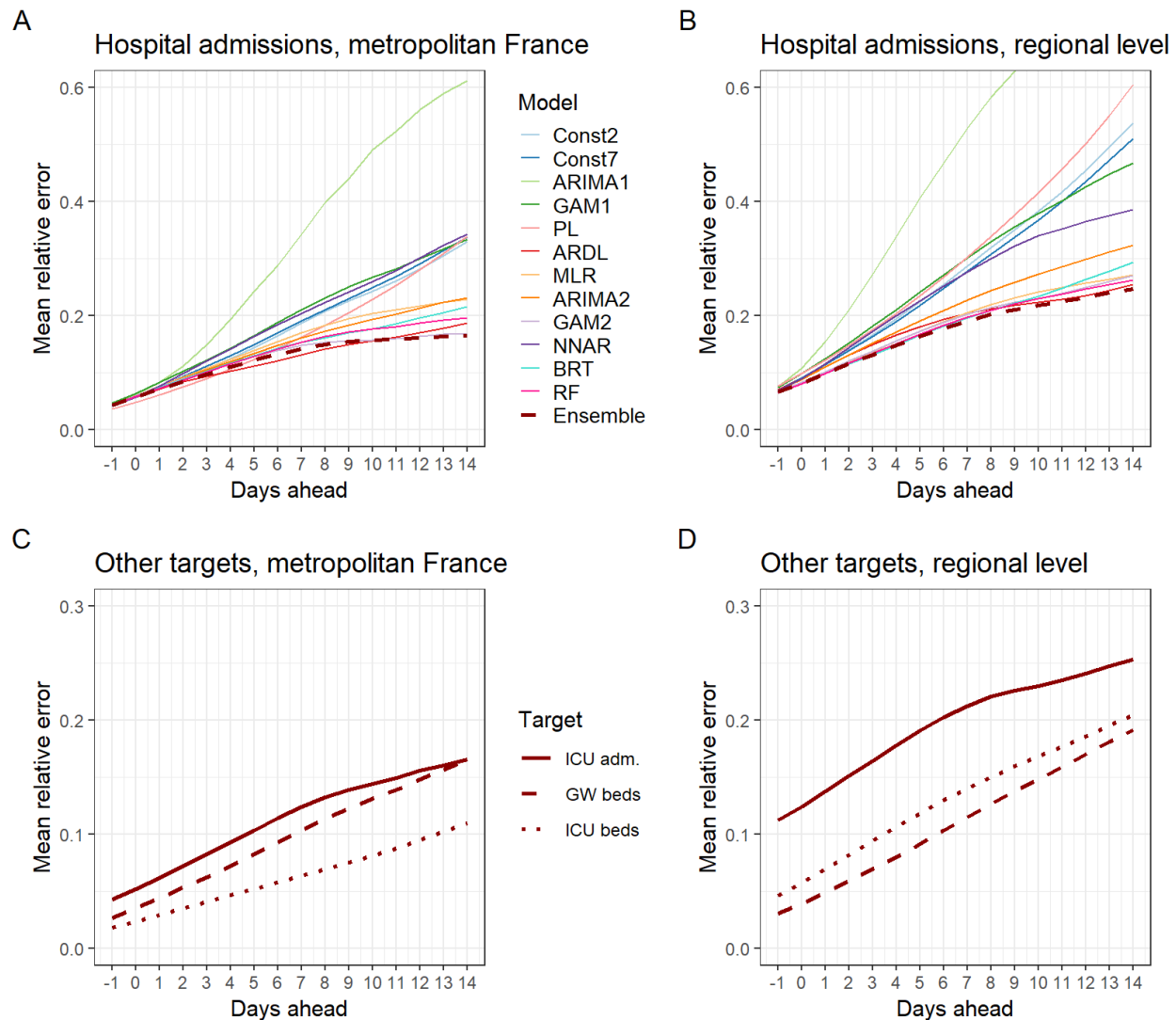
1. S. Funk, *et al.*, Short-term forecasts to inform the response to the Covid-19 epidemic in the UK. *medRxiv*, 2020.11.11.20220962 (2020).
2. P. Mecnas, R. T. da Rosa Moreira Bastos, A. C. R. Vallinoto, D. Normando, Effects of temperature and humidity on the spread of COVID-19: A systematic review. *PLoS One* **15**, e0238339 (2020).
3. Á. Briz-Redón, Á. Serrano-Aroca, The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modelling techniques. *Progress in Physical Geography: Earth and Environment* **44**, 591–604 (2020).
4. M. U. G. Kraemer, *et al.*, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
5. J. Landier, *et al.*, Colder and drier winter conditions are associated with greater SARS-CoV-2 transmission: a regional study of the first epidemic wave in north-west hemisphere countries. *medRxiv*, 2021.01.26.21250475 (2021).
6. A. Roumagnac, E. De Carvalho, R. Bertrand, A.-K. Banchereau, G. Lahache, Étude de l'influence potentielle de l'humidité et de la température dans la propagation de la pandémie COVID-19. *Medecine De Catastrophe, Urgences Collectives* (2021) <https://doi.org/10.1016/j.pxur.2021.01.002> (February 3, 2021).
7. H. Salje, *et al.*, Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
8. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunubá Z, Cuomo-Dannenburg G, Dighe A, Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College COVID-19 Response Team* (March, 16 2020) <https://doi.org/10.25561/77482>.
9. M. A. Johansson, *et al.*, An open challenge to advance probabilistic forecasting for dengue epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24268–24274 (2019).
10. N. G. Reich, *et al.*, Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. *PLoS Comput. Biol.* **15**, e1007486 (2019).
11. C. Viboud, *et al.*, The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics* **22**, 13–21 (2018).
12. E. Y. Cramer, *et al.*, Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *medRxiv*, 2021.02.03.21250974 (2021).
13. E. Volz, *et al.*, Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv*, 2020.12.30.20249034 (2021).
14. R. Polikar, Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine* **6**, 21–45 (2006).
15. E. L. Ray, *et al.*, Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S. *medRxiv*, 2020.08.19.20177493 (2020).

16. T. K. Yamana, S. Kandula, J. Shaman, Individual versus superensemble forecasts of seasonal influenza outbreaks in the United States. *PLoS Comput. Biol.* **13**, e1005801 (2017).

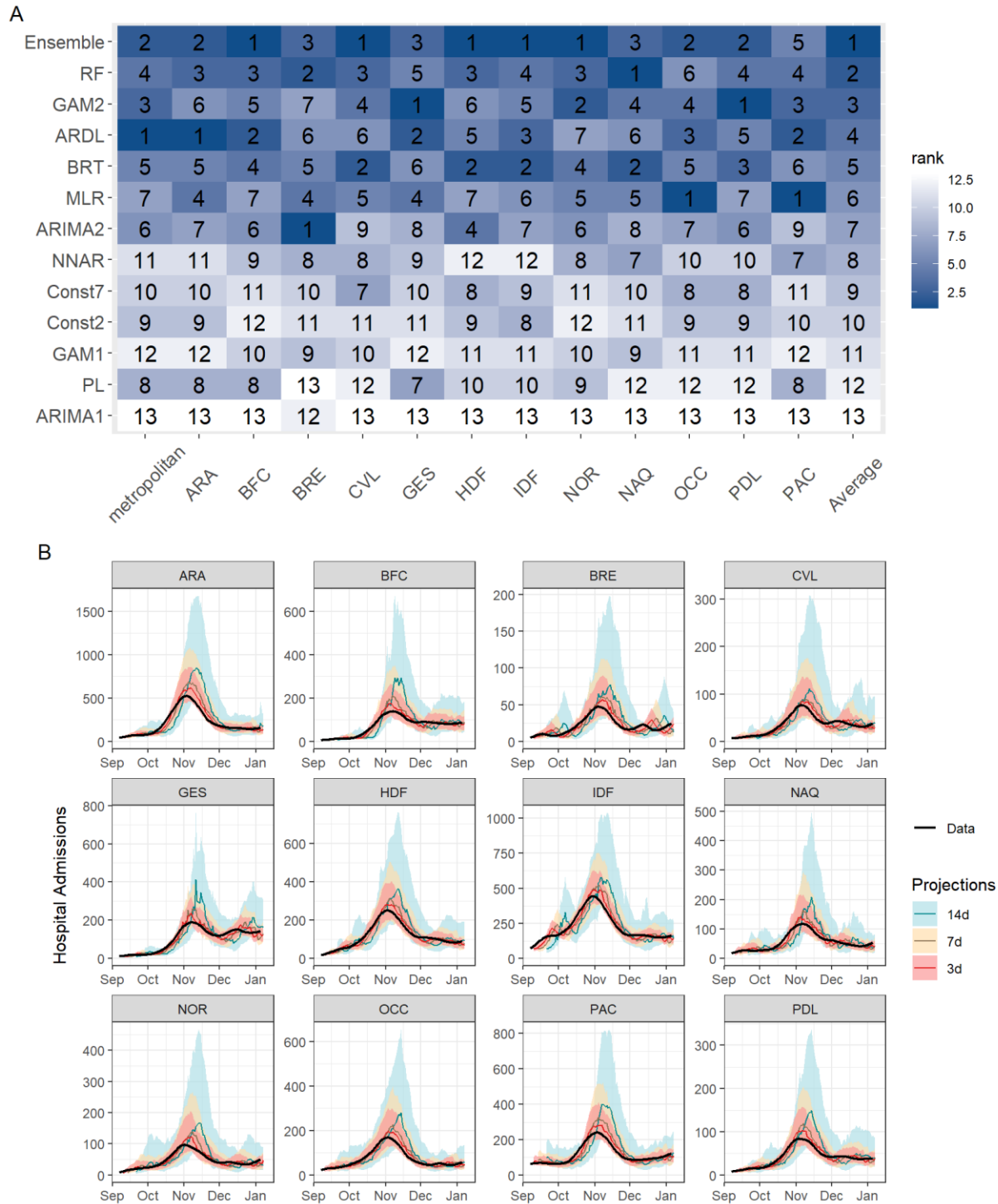
## Figures



**Figure 1: Trajectories of hospital admissions predicted by the 12 individual models in metropolitan France.** The black line is the eventually observed data (smoothed), and the colored lines are trajectories predicted on day  $t$ , for prediction horizons  $t-1$  up to  $t+14$ . The evaluation period runs from September 7th 2020 to January 7th 2021, excluding the first two weeks of the second national lockdown (October 30th 2020 to November 13th 2020), as the models are not designed to anticipate the impact of a lockdown before its implementation.

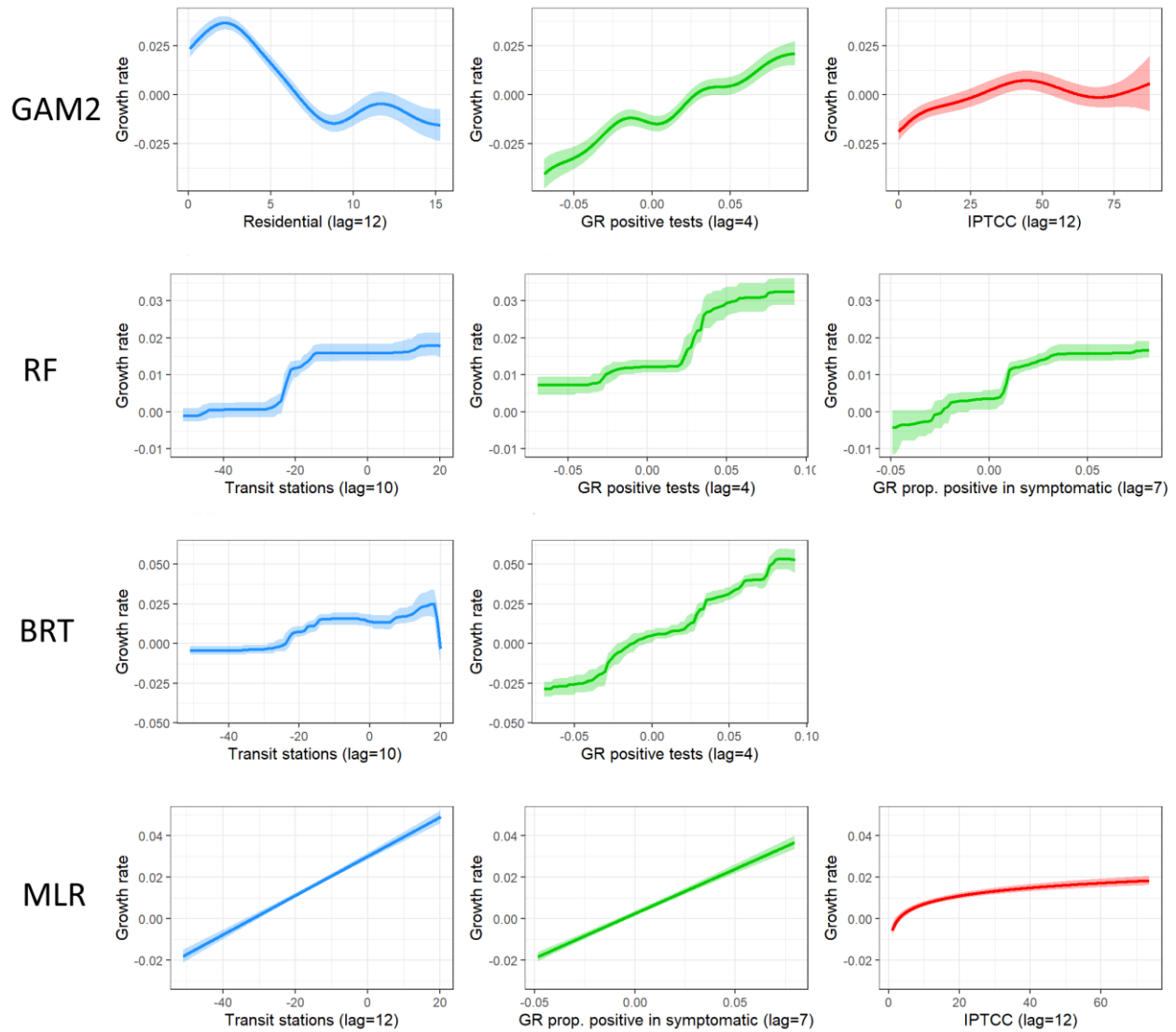


**Figure 2: Mean relative error at the national and regional levels, by prediction horizon, for the four targets.** A. Relative error for hospital admissions for the individual and ensemble models, in metropolitan France. B. Relative error for hospital admissions for the individual and ensemble models, by region. C. Relative error for intensive care unit (ICU) admissions, general wards (GW) beds and ICU beds for the ensemble model, in metropolitan France. D. Relative error for ICU admissions, GW beds and ICU beds for the ensemble model, by region.



**Figure 3: Model ranking and ensemble forecasts.** A. Ranks of the models. Models are ranked according to the relative error averaged over all prediction horizons. B. Forecasts of the ensemble model for regional hospital admissions at 3, 7 and 14 days. Regions: Auvergne-Rhône-Alpes (ARA), Bourgogne-Franche-Comté (BFC), Bretagne (BRE), Centre-Val de Loire (CVL), Grand Est (GES), Hauts-de-France (HDF), Île-de-France (IDF), Normandie (NOR), Nouvelle-Aquitaine (NAQ), Occitanie (OCC), Pays de la Loire (PDL), Provence-Alpes-Côte d'Azur (PAC).





**Figure 4: Effects of mobility (blue), epidemiological (green) and meteorological (red) predictors on the growth rate of hospital admissions, for the GAM2, the RF, the BRT and the MLR models.** Abbreviations: GR = growth rate. IPTCC = index characterizing climatic conditions favorable for the transmission of COVID-19 (6).

# Supplementary information

## Supplementary text

### Hospitalization data

Hospital data are obtained from the SI-VIC database, the national inpatient surveillance system used during the pandemic. The database is maintained by the ANS (Agence du Numérique en Santé) and provides real time data on the COVID-19 patients hospitalized in French public and private hospitals. Data are sent daily to Santé Publique France, the French national public health agency. All cases are either biologically confirmed or present with a computed tomographic image highly suggestive of SARS-CoV-2 infection. We restrict our analyses to patients newly hospitalized in ICU (“Hospitalisation réanimatoire: réanimation, soins intensifs et unité de surveillance continue”) and general ward beds (“Hospitalisation conventionnelle”). We exclude patients hospitalized in psychiatric care (“Hospitalisation psychiatrique”), long-term care and rehabilitation care (“Soins de suite et réadaptation”) and emergency care patients (“Soins aux urgences”). We consider events (hospitalizations, transfers or discharges) by date of occurrence and correct observed data for reporting delays (1).

### Smoothing

Hospital data follow a weekly pattern, with less admissions during weekends compared to weekdays, and can be noisy at the regional level. In order to remove the noise and prevent forecasts from being biased depending on the day of the week at which the analysis is performed, we remove this weekly pattern and smooth the data using local polynomial regression (LPR). A centered 7-day moving average is often used to remove weekly patterns in a time-series. However, this method leads to the loss of the last three data points, and therefore is not considered here.

We use the biweight kernel in the regression, where the bandwidth  $h$  corresponds to the number of time points below and above the current date used in the local regression. The following algorithm is used to obtain an estimate up to current time  $T$ :

1 – Remove the weekly pattern of the data. We assume that the logged incidence  $y(t) = m(t) + w(d(t)) + \varepsilon(t)$ , where  $m(t)$  is the temporal trend,  $d(t)$  is the day of the week at

date  $t$ , and  $w(d)$  is the weekly pattern. To estimate  $\widehat{w}(d)$ , we first obtain a trend estimate  $m_r(t)$  from the raw data using LPR with a bandwidth of 8 days. We compute the de-trended series by  $y_d(t) = y(t) - m_r(t)$ . We then compute  $\widehat{w}(d)$ , the average of the de-trended residuals by day. The output is a de-patterned series  $\widehat{y}_w(t) = y(t) - \widehat{w}(d(t))$ .

2 – Local polynomial smoothing with adjusted kernels of the de-patterned series  $\widehat{y}_w(t)$ . We assume  $\widehat{y}_w(t) = m_w(t) + \varepsilon_w(t)$  and use local polynomial smoothing to obtain an estimate of the trend and of the derivative of the trend. A major issue in real time smoothing is to account for yet unobserved future data. Several approaches are possible, among which extrapolation of future data, automatic kernel curtailing and the least revision principle. In this last approach, the smoothing kernels are tailored to minimize the error between the real-time predicted value and the eventual estimate. The curtailing approach is theoretically unbiased but is variable. The least revision approach introduces bias but reduces variance. Proietti *et al.* described a framework for implementing the least revision principle using low-order reproducing kernels (2). We use the linear/quadratic approach to compute the approximate kernels and estimate predicted mean, slope and curvature of the current signal. Confidence intervals are computed by bootstrap.

We compare this two-step algorithm with a simple smoothing spline (Fig. S6). The time-series smoothed by a smoothing spline (panel A) is very sensitive to the weekly pattern: the values are systematically under-estimated when the last data point is a Sunday, and over-estimated when the last data point is a Friday. The advantage of our approach (panel B) is that it is insensitive to the day of the week, and more generally, less sensitive to noise. This comes with one drawback: in the case of a sudden change of trajectory, as it occurred at the beginning of November: it can take a few days before the smoothed time-series catches the right trajectory. This loss in reactivity (detecting changes as early as possible) is balanced by the gain in stability (avoiding false alarms).

## Description of individual models

We evaluate 12 individual models to forecast hospital admissions. The first three models directly predict the number of hospital admissions, while the others predict the growth rate, from which hospital admissions are then derived using an exponential growth model.

#### *ARIMA1: Autoregressive integrated moving average model*

We fit a simple ARIMA model of hospital admissions, where the parameters are estimated at each time step using the `auto.arima` function of the R package *forecast*, independently for each region.

#### *GAM1: Generalized additive model*

We fit a GAM model of hospital admissions, with a single smooth term for time, using the R package *mgcv*. The model is calibrated independently for each region.

#### *NNAR: Neural networks auto-regressive model*

We use the function `nnetar` from the R package *forecast* to fit a neural network autoregressive model to the smoothed hospital admissions time series. The function implements a single layer neural network that takes as input  $p$  lagged values of the time series to forecast - and possibly other covariates - and contains  $k$  neurons. The values of  $p$  and  $k$  ( $p = 7$ ,  $k = 5$ ) are set via cross-validation. The model is calibrated independently on each region.

The forecasts are obtained iteratively: for one step ahead forecasts, the algorithm uses just the training set; for two steps ahead forecasts, the algorithm uses the training data along with the one-step forecast obtained in the previous step, and so on.

#### *Const: Exponential growth models with constant growth rate*

We estimate the exponential growth rate  $r$  by fitting a Poisson regression model of the smoothed hospital admissions over a fixed time window. We test windows of 2 (“Const2”) and 7 (“Const7”) days. We project hospital admissions  $H(t)$  by assuming the growth rate will stay constant in the future:

$$H(t) = H_0 \cdot \exp(r \cdot t)$$

#### *PL: Exponential growth model with piecewise linear growth rate*

We consider an extension of the previous model for which the growth rate varies over time:

$$\frac{dH(t)}{dt} = r(t) H(t)$$

and where  $r(t)$  is a continuous piecewise linear function:

$$r(t) = a + bt + \sum_{k=1}^{K-1} h_k \cdot \max(t - \tau_k, 0)$$

Here, the  $\tau_k$  are the instants when the slope changes and  $K$  is the number of segments.

*MLR: Multiple linear regression model*

We fit a multiple linear regression model of the growth rate  $r$ , with covariates selected by forward stepwise selection (see below). The model is fitted on all regions together. The covariates  $X_k$  are introduced in the model as lagged variables with lag  $l_k$ :

$$r(t) = \beta_0 + \sum_k \beta_k \cdot X_k(t - l_k)$$

The best lag  $l_k$  for each covariate is estimated at each time step using Pearson correlation coefficient between the growth rate and the covariate.

The growth rate is then predicted for all prediction horizons by assuming that all covariates will stay constant in the future (equal to their last observed values). We then derive forecasts of hospital admissions recursively:

$$H(t) = H(t-1) \cdot \exp(r(t))$$

*GAM2: Generalized additive model*

We fit a GAM model of the growth rate  $r$ , using the same approach as multiple linear regression, except that the lagged covariates are introduced in the model as smoothed functions  $f_k$  (to relax the linearity assumption):

$$r(t) = \beta_0 + \sum_k f_k(X_k(t - l_k))$$

We use the R package *mgcv*.

### *ARIMA2: Multiple linear regression model with ARIMA error*

We fit a multiple linear regression model of the growth rate with  $k$  lagged covariates and an ARIMA error, to account for autocorrelation in the data:

$$r(t) = \beta_0 + \sum_k \beta_k \cdot X_k(t - l_k) + \eta_t$$

where  $\eta_t$  is an ARIMA process. The model is fitted on each region separately due to the ARIMA structure. We use the R package *forecast*. We select covariates and derive forecasts of hospital admissions using the same approach as for linear regression.

### *ARDL: Autoregressive distributed lag model*

In a distributed lag model, the effect of a covariate on the dependent variable can be distributed over time rather than occur all at once. We use three lags for each covariate. These lags are defined for each prediction horizon, so that we only use the observed values of the covariates, without making any assumption about their future values. For instance, to predict the growth rate five days ahead, we use lags 5, 6 and 7, which correspond to the last three observed values of the covariates. We estimate the lag weights (coefficients of the regression) for each prediction horizon. Therefore, the weights associated to a covariate can be large at short horizons and small at long horizons, or vice versa. We also include lagged values of the growth rate of hospital admissions (autoregressive model). For any prediction horizon  $h$ , the growth rate at  $t+h$  is:

$$r(t+h) = \beta_0 + \sum_k \sum_{i=0}^2 \beta_{k,i} \cdot X_k(t-i) + \sum_{j=0}^2 \gamma_j \cdot r(t-j)$$

We fit all the regions together, and use the same covariate selection procedure as for other models.

### *RF: Random forests*

Regression trees approaches consist in recursively partitioning the data using binary splits and to build a set of decision rules on the predictors. RF combine decision trees with bagging - bootstrap aggregation of multiple trees run in parallel. We use RF for regression of the growth rate of hospital admissions at time  $t+h$  using the covariates at time  $t$ . The R package *randomForest* is used. In practice we use  $h=10$  days for mobility and climate predictors, and  $h=4$  or  $h=7$  days

for epidemiological predictors, and the minimum size of the nodes is set to 1,000 to reduce overfitting. Importance of variables is assessed with the increase in node impurity, computed as the total decrease in residual sum of squares obtained after each splitting on the variable and averaged over all trees. The dependency between the growth rate and a covariate is visualized using partial dependence plots, where we determine the marginal effect of the covariate while setting the other covariates to their median value.

#### *BRT: Boosted regression trees*

BRT combine decision trees with boosting. At the difference of RF, trees are added sequentially and not in parallel. At each step, the tree that best reduces a loss function is added. We use the R package *gbm* and choose the default parameters offered by the package: fits are made on 100 trees; a Gaussian loss function is used; interaction depth =1; the shrinkage (learning rate) is set to 0.1. We use the same lags as in the RF model. Relative importance of the covariates is a measure of how each variable contributes to reducing the loss function. Similarly to the RF, we visualize the dependency between the growth rate and the covariates using partial dependence plots.

### **Description of predictors**

We include in individual models a set of predictors, chosen for their availability in near real-time and their potential to help to anticipate the trajectory of hospital admissions. Three types of predictors are considered: 9 epidemiological predictors describing the dynamics of the epidemics, 9 mobility predictors and 4 meteorological predictors. All predictors are available at the region and day levels. Most of them follow a strong weekly pattern. Data are smoothed using the methodology used for hospitalization data, in order to remove the weekly pattern and reduce edge effects (see “Smoothing” above).

#### *Epidemiological predictors*

In addition to the growth rate of hospitalizations, we include predictors on confirmed cases, given that cases are expected to be reported a few days before hospitalizations. Case data are obtained from the SIDEPA database (Système d’Information de Dépistage Populationnel - Information system for population-based testing), the national surveillance system describing RT-PCR and antigenic tests results for SARS-CoV-2 arising from private and public French laboratories.

Anonymized data are transmitted daily to Santé publique France through a secured platform. Test results are reported by date of nasopharyngeal swab and include patient information such as age, delay since symptoms onset and postal code of the home address.

We explore 8 potential predictors (Fig. S1):

- the number of positive tests, and their growth rate
- the number of positive tests, in people aged >70 years, and their growth rate
- the proportion of positive tests among all tests, and their growth rate
- the proportion of positive tests among tests in symptomatic people, and their growth rate.

The exponential growth rate is computed using a 2-day rolling window, and the resulting time series is smoothed using local polynomial regression. Due to reporting delays, case data can be used up to 2 days before the date of analysis.

#### *Mobility predictors*

Mobility data are obtained from Google (<https://www.google.com/covid19/mobility/>) and Apple (<https://covid19.apple.com/mobility>) mobility reports.

Google mobility data describe how visitors to (or time spent in) categorized places change compared to a baseline (the 5-week period Jan 3 – Feb 6, 2020). The 6 categorized places are: residential (time spent at home), workplaces, grocery and pharmacy, retail and recreation, parks, and transit stations (Fig. S2). Reports are updated every other day and contain data up to 2 days prior to the day the dataset is generated. They are uploaded 2 days after the day the dataset is generated. Therefore, the maximum delay for data availability is 5 days.

Apple mobility data describe the relative changes in the volume of requests on Apple map App for itineraries, compared to January 13 2020. Itineraries are divided in 3 categories: walking, transit, and driving (Fig. S2). Reports are updated every day and contain data up to one or two days before.

#### *Meteorological predictors*

Climate data are obtained from Météo France/PREDICT Services, and include temperature, absolute humidity and relative humidity, for each weather station in France (N=63) (Fig. S3). We also include the IPTCC index (*Index PREDICT de transmissivité climatique de la COVID-19*), an index characterizing favorable climatic conditions for the transmission of COVID-19 (3). We take

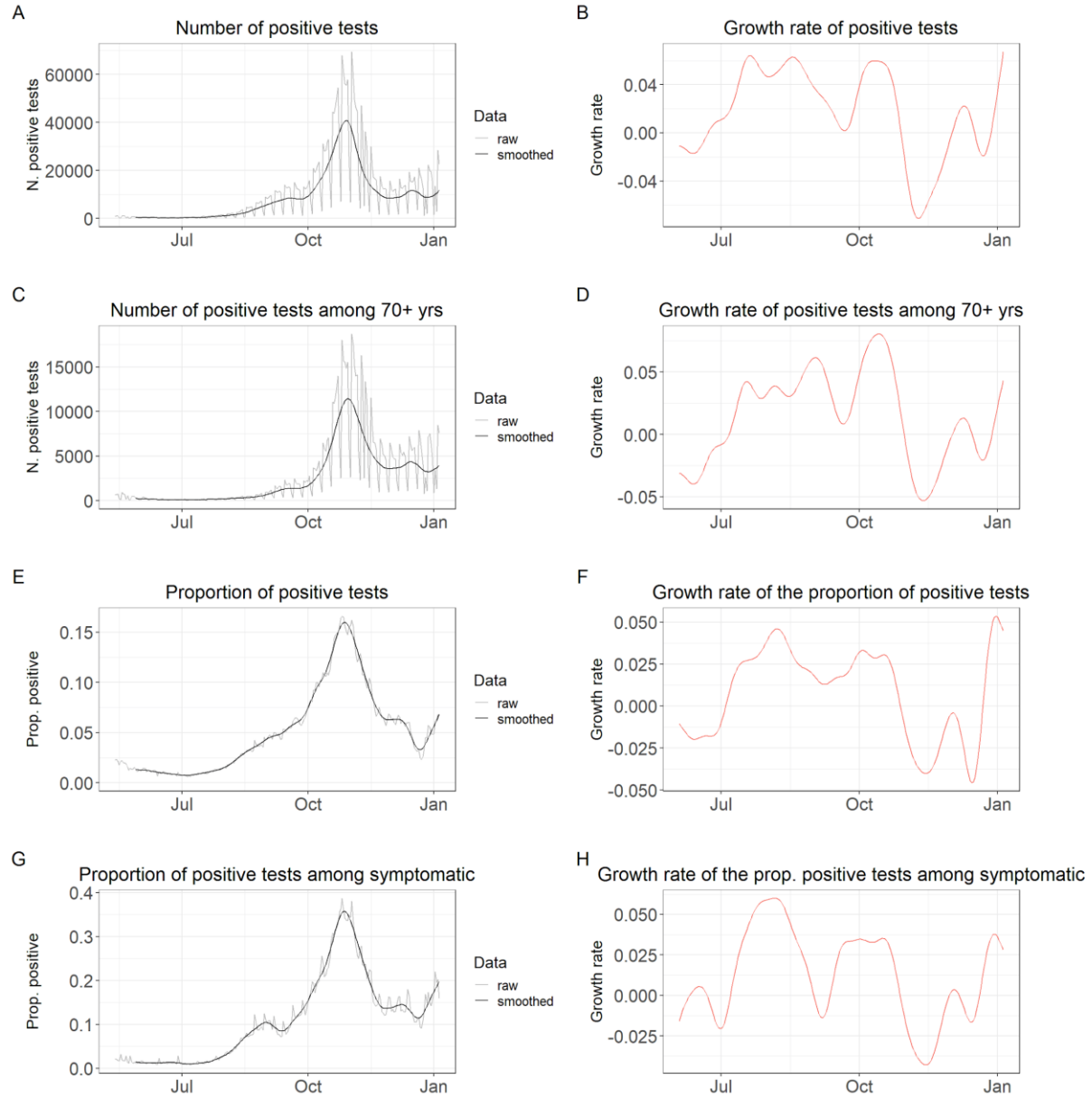


the median of the four variables in each region. In linear models, IPTCC is also tested in its logarithmic form.

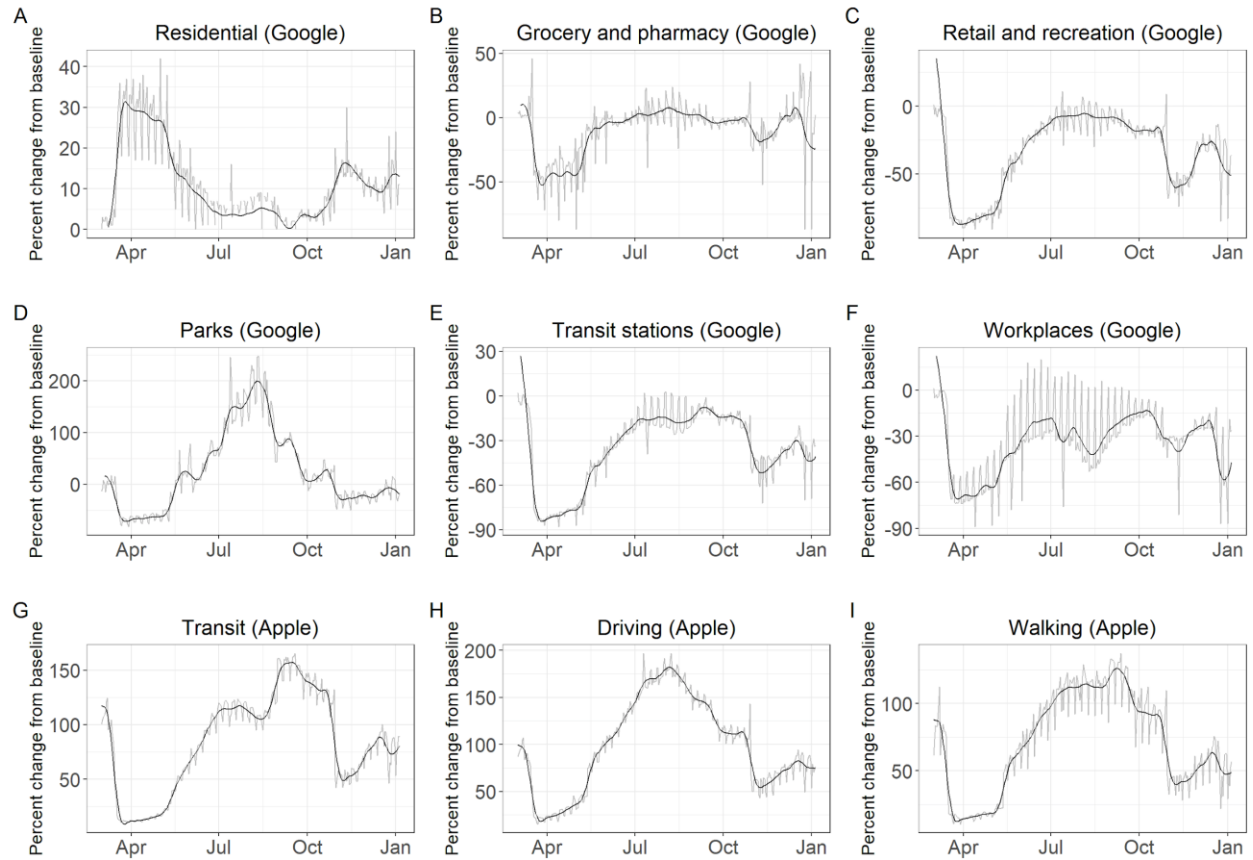
### **Forward selection of predictors**

In order to select the best predictors to include in individual models, we use a forward stepwise selection method. We first include all covariates ( $N=22$ ) in univariate models and run each univariate model over the evaluation period. We retain the covariate that minimizes the average relative error of predictions at  $t+7$  and  $t+14$ . We then include remaining covariates one by one, until no additional covariate can decrease the relative error by more than 0.5%. We also consider two alternative models starting from the second or the third best covariate in univariate analysis. In the end, we retain the model with the lowest relative error among the three multivariate models.

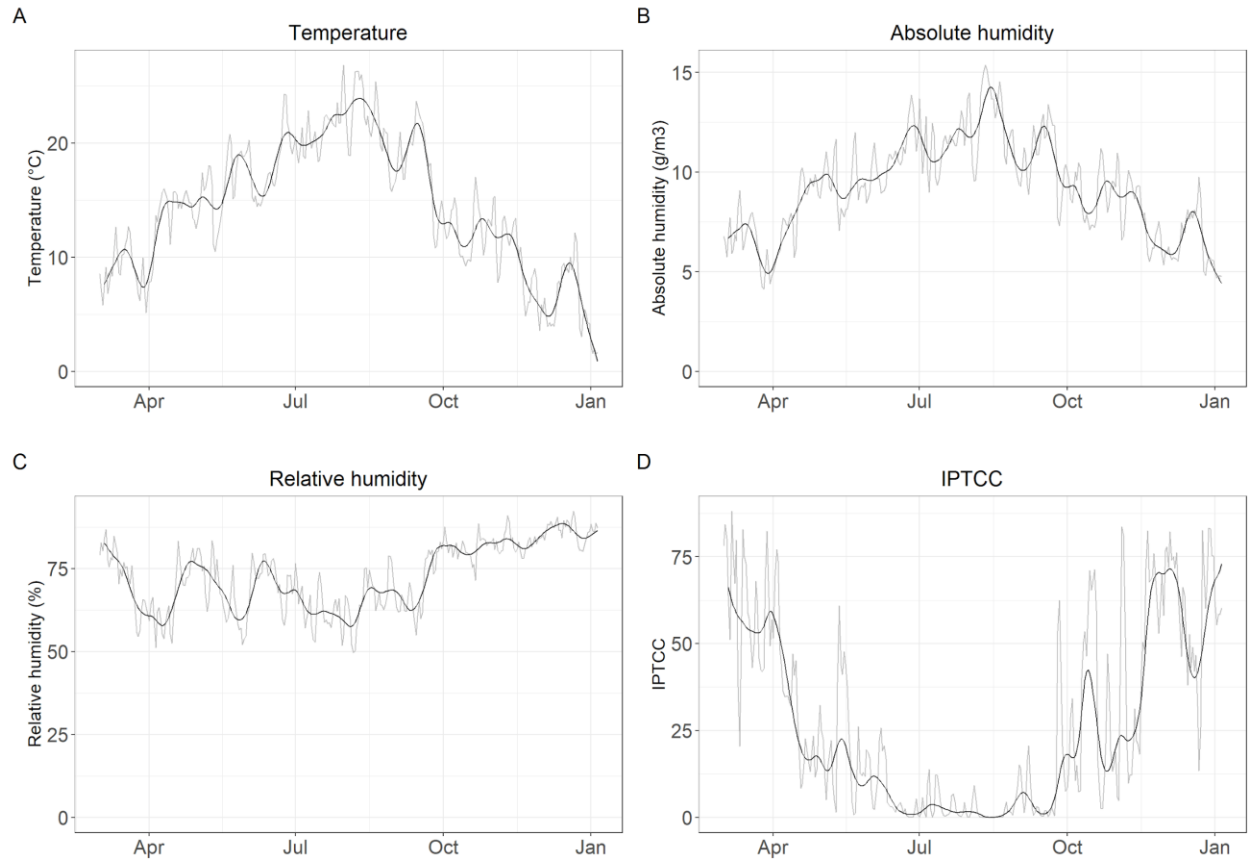
## Supplementary figures



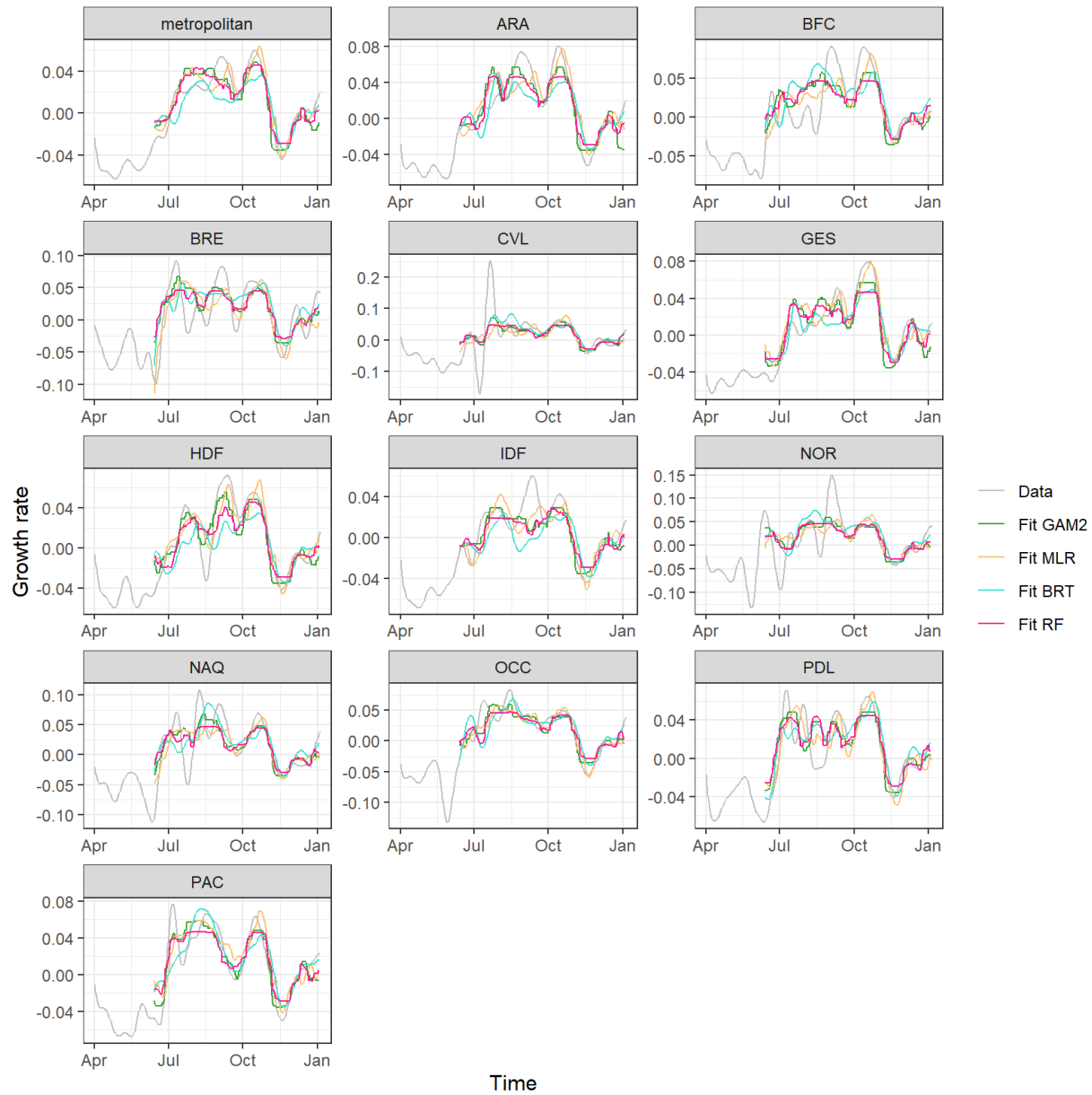
**Fig. S1: Epidemiological predictors (see also Supplementary text).**



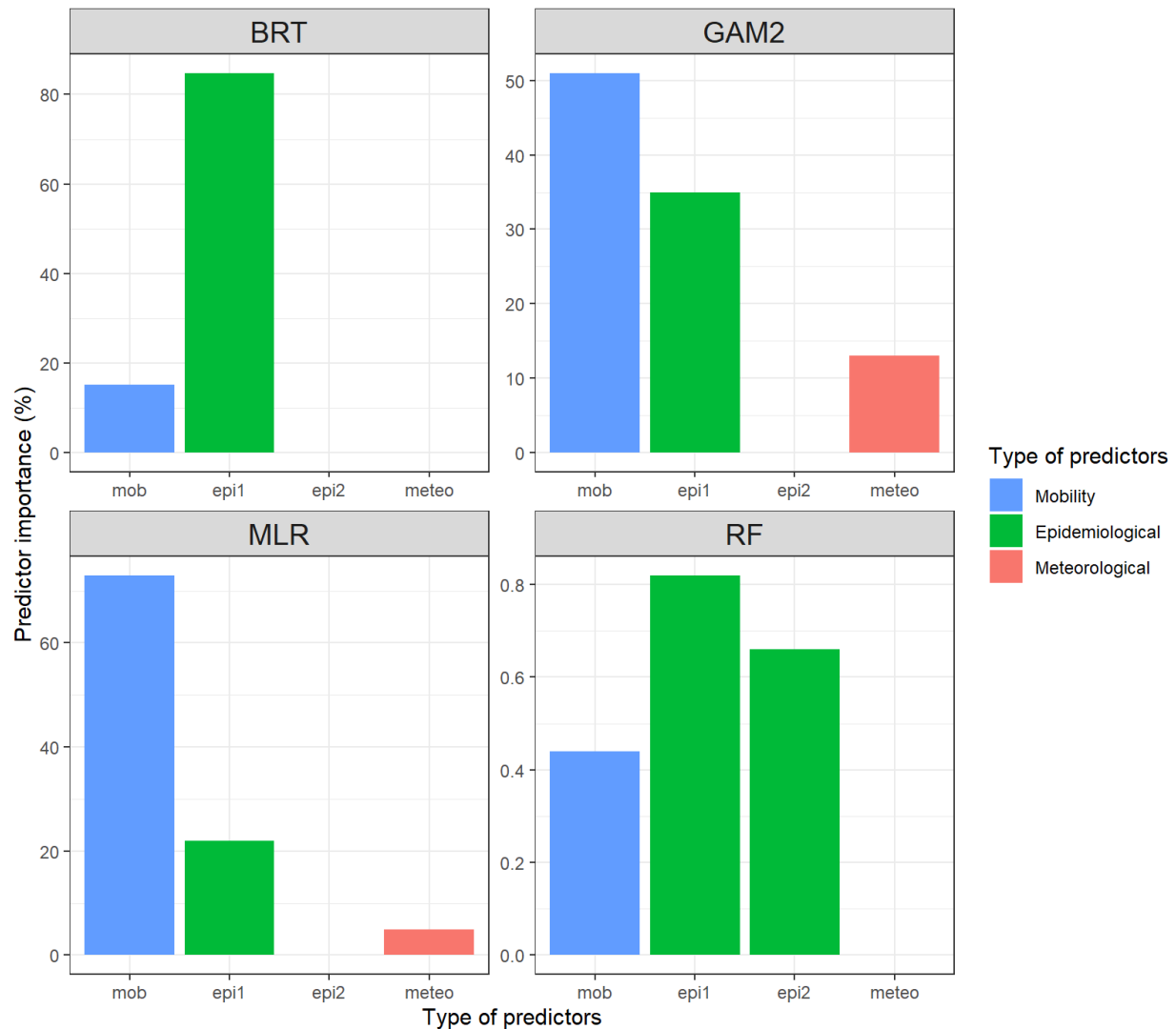
**Fig. S2: Mobility predictors (see also Supplementary text).**



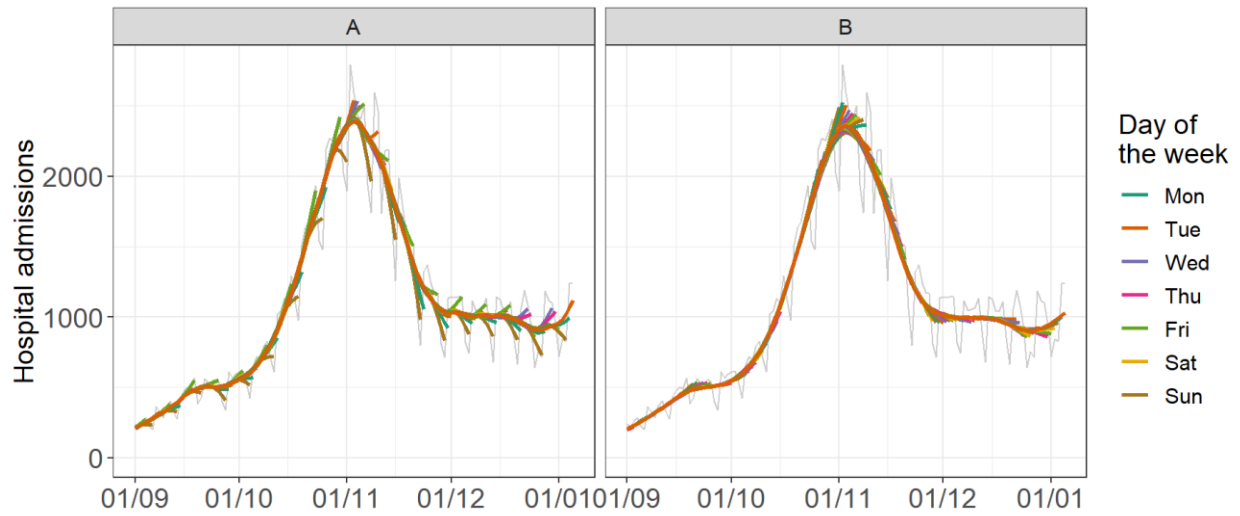
**Fig. S3: Meteorological predictors (see also Supplementary text).**



**Fig. S4: Model fits for the growth rate of hospital admissions, at the national (metropolitan France) and regional levels, for the GAM2, the MLR, the BRT and the RF.** Each panel shows the observed growth rate (black line) and the predicted growth rates (colored lines) when retrospectively fitting each model over the entire time period, on all regions together. Abbreviations for regions: Auvergne-Rhône-Alpes (ARA), Bourgogne-Franche-Comté (BFC), Bretagne (BRE), Centre-Val de Loire (CVL), Grand Est (GES), Hauts-de-France (HDF), Île-de-France (IDF), Normandie (NOR), Nouvelle-Aquitaine (NAQ), Occitanie (OCC), Pays de la Loire (PDL), Provence-Alpes-Côte d’Azur (PAC).



**Fig. S5: Importance of predictors.** For the BRT, relative importance is a measure of how each predictor contributes to reducing the loss function (all contributions sum to 100%). For the MLR and the GAM2 models, relative importance is a measure of how each predictor contributes to the total explained variance (all contributions sum to 100%). For the RF, predictor importance is assessed with the increase in node impurity, computed as the total decrease in residual sum of squares obtained after each splitting on the variable and averaged over all trees (importance measures do not sum to 100%).



**Fig. S6: Comparison of smoothing methods.** (A) Smoothing spline. (B) Our two-step algorithm. The grey line shows the raw data of hospital admissions in metropolitan France from September 2020 to January 2021. The colored lines are time-series smoothed in real time (i.e. knowing only the past values), with different colors indicating the day of the last data point.

## Supplementary tables

**Table S1: Prediction interval coverage of the ensemble model, at the national and regional level, for 7- and 14-day ahead forecasts.**

Level	Target	7-day ahead	14-day ahead
National	Hospital admissions	1	1
	ICU admissions	0.90	1
	General wards beds	0.87	0.94
	ICU beds	0.92	0.96
Regional	Hospital admissions	0.96	0.97
	ICU admissions	0.99	1
	General wards beds	0.98	0.97
	ICU beds	0.99	1



**Table S2: Best predictors selected for each individual model using the forward stepwise selection procedure.** The first six models were included in the ensemble model.

<b>Model</b>	<b>Dependent variable</b>	<b>Are lagged values of the dependent variable used as covariate?</b>	<b>Epidemiological predictors</b>	<b>Mobility predictors</b>	<b>Meteoro-logical predictors</b>
ARDL	Growth rate of hospital admissions	Yes	Growth rate of the proportion of positive tests among tests in symptomatic people	Transit stations (Google)	absolute humidity
MLR	Growth rate of hospital admissions	No	Growth rate of the proportion of positive tests among tests in symptomatic people	Transit stations (Google)	IPTCC (log)
GAM	Growth rate of hospital admissions	No	Growth rate of the number of positive tests	Residential (Google)	IPTCC
ARIMA2	Growth rate of hospital admissions	No		Transit stations (Google) and residential (Google)	
BRT	Growth rate of hospital admissions	No	Growth rate of the number of positive tests	Transit stations (Google)	
RF	Growth rate of hospital admissions	No	Growth rate of the number of positive tests Growth rate of the proportion of positive tests among tests in symptomatic people	Transit stations (Google)	
NNAR	Number of hospital admissions	Yes	Number of positive tests in people aged >70 years	Grocery and pharmacy (Google)	Absolute humidity

## SI References

1. H. Salje, *et al.*, Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
2. T. Proietti, A. Luati, Real time estimation in local polynomial regression, with application to trend-cycle analysis. *Ann. Appl. Stat.* **2**, 1523–1553 (2008).
3. A. Roumagnac, E. De Carvalho, R. Bertrand, A.-K. Banchereau, G. Lahache, Étude de l'influence potentielle de l'humidité et de la température dans la propagation de la pandémie COVID-19. *Medecine De Catastrophe, Urgences Collectives* (2021) <https://doi.org/10.1016/j.pxur.2021.01.002> (February 3, 2021).