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

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

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

Affiliations:

[1]Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Université de Paris, CNRS UMR 2000, Paris, France.

[2]Direction des Maladies Infectieuses, Santé Publique France, French National Public Health Agency, Saint Maurice, France.

[3]Univ Rennes, EHESP, REPERES « Recherche en Pharmaco-Epidémiologie et Recours aux Soins » – EA 7449, Rennes, France.

[4]PREDICT Services, Castelnau-le-Lez, France.

[5]INRIA, Saclay, France.

[6]CMAP, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, Palaiseau, France.

[7]INSERM, Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris, France.

†These authors contributed equally to this work.

*Corresponding author. Email: juliette.paireau@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 19 predictors to anticipate French COVID-19 related healthcare needs from September 7th 2020 to March 6th 2021. We then build an ensemble model by combining the individual forecasts and test this model from March 7th to July 6th 2021. We find that inclusion of early predictors (epidemiological, mobility and meteorological predictors) can halve the root mean square error for 14-day ahead forecasts, with epidemiological and mobility predictors contributing the most to the improvement. On average, the ensemble model is the best or second best model, depending on the evaluation metric. 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.

Significance Statement

The COVID-19 pandemic is inducing important stress on healthcare structures, which can be quickly saturated with negative consequences for patients. As hospitalization comes late in the infection history of a patient, early predictors – such as the number of cases, mobility, climate, and vaccine coverage - could improve forecasts of healthcare demand. Predictive models taken individually have their pros and cons, and it is advantageous to combine the predictions in an ensemble model. Here, we design an ensemble that combines several models to anticipate French COVID-19 healthcare needs up to 14 days ahead. We identify the best early predictors of the growth rate of hospital admissions and propose a promising approach to facilitate the planning of hospital activity.

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 negative 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 19 predictors to anticipate French COVID-19 related healthcare needs, from September 7th 2020 to March 6th 2021. We build an ensemble model by combining the individual forecasts and test this model from March 7th to July 6th 2021. 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 root mean square error (RMSE) 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

Overview of the approach

We first develop a set of individual models to forecast the number of hospital admissions at the national and regional level, up to 14 days ahead. These individual predictions are then combined into a single ensemble forecast (1, 6–10). Finally, we derive three other targets (number of ICU admissions, bed occupancy in general wards and bed occupancy in ICU) from the number of hospital admissions predicted by the ensemble model (see Material and Methods).

We use a two-stage procedure: (i) over the evaluation period (September 7th 2020 to March 6th 2021), we select the predictors and evaluate the performance of the individual models to select the best ones to include in the ensemble model and (ii) over the test period (March 7th 2021 to July 6th 2021), we assess the performance of the ensemble model.

Performance of individual models to forecast hospital admissions over the evaluation period

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 dynamics of hospital admissions from September 2020 to March 2021 (Fig. S4). They all over-estimate the November peak since they were not designed to anticipate the impact of the lockdown before its implementation.

Models are compared using the root mean square error (RMSE) for point forecast error and the weighted interval score (WIS) to assess probabilistic forecast accuracy (see Material and Methods). Overall, the performance of the models decreases with the prediction horizon (Figure 1). Six models outperformed the baseline model (characterized by no change in the number of hospital admissions) at all prediction horizons for both the RMSE and the WIS, at the national and regional levels: 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) model and a random forest (RF) model. All these models describe the growth rate of hospital admissions, rather than hospital admissions directly, and include several predictors that are described below.

Predictors

The six best individual models include between 2 and 4 predictors (Table S1). The best predictors are selected by cross-validation using a forward stepwise selection method (see Material and Methods and Supplementary text). 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: time spent in residential places is the one that is most often selected, followed by the volume of visits to transit stations (in percent change from baseline). Four models use at least one predictor on confirmed cases: the growth rate of the proportion of positive tests (among all tests or among tests in symptomatic people), and/or the growth rate of the number of positive tests. Two models use one meteorological predictor, either absolute humidity or temperature.

In order to determine the importance of the different predictors and explore their effect on the growth rate, we retrospectively fit four individual models (BRT, RF, MLR and GAM2) from June 3rd 2020 to March 6th 2021 (see Material and Methods). Retrospectively, the models are able to reproduce reasonably well the dynamics of the growth rate over time and by region (Fig. S5). Depending on the model, the most important predictors are mobility or epidemiological predictors (Fig. S6). For instance, in the MLR model, the change in time spent in residential places and the growth rate of the positive tests both contributes to 47% of the explained variance. In the BRT model, the growth rate of the number of positive tests is the most important predictor (relative contribution of 89%) followed by the time spent in residential places (6%) and change in the volume of visits to transit stations (5%). Meteorological factors contribute to 37% in the GAM2 model but had no contribution in the three other models.

We find that an increase in the volume of visits to transit stations, a decrease in the time spent in residential places, or a decrease in absolute humidity, is associated with an increase in the growth rate of hospital admissions 10-12 days later (Fig. 2). 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, with a lag of 7 days.

Performance of the ensemble model over the test period

To build the ensemble model, we keep the six best performing models and take the unweighted mean of the individual forecasts. In addition to the previously selected predictors, we also include vaccine coverage and the proportion of variants of concern (VOC) (Supplementary text and Figure

S7) as these two predictors might significantly affect the dynamic of hospitalizations after March 2021. When retrospectively fitting the models from June 3rd 2020 to July 6th 2021, the effects of the other predictors remain relatively stable, compared to the previous fit from June 3rd 2020 to March 6th 2021 (Figure S8).

The ensemble model is evaluated over the test period (March 7th 2021 to July 6th 2021). It is able to capture the growth of hospital admissions in March and the decline in April-July (Figure 3A). The ensemble model performs well in all regions (Figure 3B). On average, the ensemble model is the best at the national level for both the RMSE and the WIS, and at the regional level for the RMSE, and is the second best model at the regional level for the WIS (Fig.S9). For each region or week, we rank the individual and ensemble models according to the RMSE over all prediction horizons. The best individual model is not the same in all regions (Figure 4A) or in all weeks (Figure S10): four models are ranked first in at least one region and eight models are ranked first in at least one week; but the ensemble model is ranked first on average across all regions/weeks.

Finally, to assess the ensemble forecasts of hospital admissions, ICU admissions and bed occupancy in ICU and general wards, we report the mean absolute percentage error (MAPE, mean of the ratio of the absolute error to the observed value) as its interpretation in terms of relative error is straightforward, as well as the 95% prediction interval coverage (proportion of 95% prediction intervals that contain the observed value). For the four targets, the MAPE at 7 days are 11%, 13%, 6% and 5% at the national level (17%, 23%, 8% and 11% at the regional level) for hospital admissions, ICU admissions, ICU beds and general ward beds, respectively (Figure 4B). At 14 days, these errors increase to 20%, 23%, 9% and 10% at the national level (30%, 35%, 13% and 19% at the regional level), respectively. The calibration is good for most of the targets but the 95% prediction interval coverage is lower than 95% for hospital admissions and 14-day ahead forecasts (Table S2).

Discussion

In this study, we evaluated the performance of 19 predictors and 12 models to anticipate French COVID-19 healthcare needs, and built an ensemble model to reduce the average forecast error. We can draw a number of important conclusions from this systematic evaluation. 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, 11, 12). 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 RMSE for a time horizon of 14 days.

Second, of the three types of predictors used over the evaluation period, 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 (11). 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). The effects of the predictors remain relatively stable between the two time periods, although the reduction of the effects of mobility predictors for the BRT model suggests a lower impact of mobility after March 2021 (Figure S8). Other potential predictors could have been considered, such as inter-region mobility or spatial correlations. However, given the 10-day delay between infection and hospitalization, we expect that most patients that will be hospitalized in a given region in the next two weeks will have recently been infected in that region. The benefits of accounting for inter-region mobility therefore appear limited for short-term predictions but might become more important when longer forecast horizons are being considered.

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, 6, 9, 13, 14). Relying on an ensemble model is appealing because it acknowledges that each model has limitations and imperfectly captures the complex reality of this pandemic. Although individual models may perform better in some situations, forecasts that build on an ensemble of models are less likely to

be overly influenced by assumptions of a specific model (15). 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. S11 and S12).

The introduction of vaccines and the emergence of variants such as the Alpha variant, that are more transmissible than historical SARS-CoV-2 viruses (16), opened 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, when vaccines were not widely used. These new factors can modify the association between the different predictors and the epidemic growth rate: it may underestimate the growth rate in a context where a more transmissible variant is also circulating, or over-estimate the growth in vaccinated populations. The flexibility of our approach allows us to adjust the models to this changing epidemiological situation: to account for these new factors, we explicitly integrate the proportion of variants and the vaccine coverage as new predictors of the models over the test period. As expected, we find that vaccine coverage is negatively associated with the growth rate of hospital admissions. For the proportion of VOC, we find a positive association in some models but no association in others (Figure S8). This might be due to the correlation between the rise in vaccine coverage and in the proportion of VOC, and/or to the fact that the effect of VOC is already accounted for by epidemiological predictors. In the meantime, since epidemiological predictors are intermediate factors between external predictors (mobility, climate, vaccine coverage and VOC) and hospital admissions, we also run sensitivity analyses with models that use epidemiological predictors only, with no consideration of external predictors. We find that models with all types of predictors are more performant than purely epidemiological models (Figure S13). Finally, as different predictors might be important at different stages of the epidemic, one could

also update the variable selection at different time points, to continuously revise the best predictors to include in the models.

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. Therefore, in the absence of smoothing or with simple smoothing techniques, forecasts can be biased depending on the day of the week at which the analysis is performed. In order to remove this day-to-day variation and obtain a smooth signal at each date not depending on future data points, we smooth the data using a 2-step approach based on local polynomial regression and the least revision principle (17) (Supplementary text and Figures S11 and S12).

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, 6–10). Finally, we derive the number of ICU admissions, and bed occupancy in general wards and ICU from the predicted number of hospital admissions.

We divide our study period into two periods: (i) over the evaluation period, we select the predictors and evaluate the performance of the individual models in order to select the best ones to include in the ensemble model and (ii) over the test period, we evaluate the performance of the ensemble model on new data.

Modelling hospital admissions

In a first step, 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), 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. We also added a baseline model characterized by no change in the number of hospital admissions.

We evaluate and compare the performance of individual models over a period running from September 7th 2020 to March 6th 2021 ("evaluation period"). As the models are not designed to anticipate the impact of a lockdown before its implementation, we exclude from the evaluation period the forecasts made between October 20th and November 4th (i.e. up to 6 days into the lockdown starting on October 30th) for hospitalizations occuring after November 3th. In other words, between October 20th and November 4th, we consider that the models could not anticipate the impact of the lockdown.

We use a cross-validation approach based on a rolling forecasting origin: for each day *t* of the evaluation period, we make forecasts for the period *t-1* up to day *t+14*, using only past data up to day *t-2* as a training set, and computing evaluation metrics 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*.

Model performance is evaluated using two main metrics: (i) our primary metric for point forecast error, the root mean square error (RMSE), was used to evaluate predictive means (Kolassa 2020),

and (ii) our secondary metric, the weighted interval score (WIS), was used to assess probabilistic forecast accuracy (1, 6). The WIS is a proper score that combines a set of interval scores for probabilistic forecasts that provide quantiles of the predictive forecast distribution. It can be interpreted as a measure of how close the entire distribution is to the observation, in units on the scale of the observed data (6).

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. Over the evaluation period, 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,...), 6 mobility predictors (for example the change in volume of visits to workplaces, transit stations, residential places or parks (Google data)) and 4 meteorological predictors (temperature, absolute and relative humidity, and IPTCC, an index characterizing climatic conditions favorable for the transmission of COVID-19 (18). 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 over the evaluation period (not using data from the test period) (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 from June 3rd 2020 to March 6th 2021, on all regions together, whenever possible. We start the fit on June 3rd 2020, when all the predictors are available. The parameters of the ARDL model vary with the prediction horizon and those of the ARIMA2 model vary by region. Therefore, only four models (BRT, RF, MLR and GAM2) are included in this sub-analysis.

In a second step, 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 test the performance of the ensemble model on the period running from March 7th 2021 to July 6th 2021 (“test period”).

To assess the performance of the ensemble model, in addition to the RMSE and the WIS used to compare the models, we also report the mean absolute percentage error (MAPE, mean of the ratio of the absolute error to the observed value) as its interpretation in terms of relative error is

straightforward, as well as the 95% prediction interval coverage (proportion of 95% prediction intervals that contain the observed value).

Over the test period, we use in individual models the previously selected predictors and, in addition, we include the vaccine coverage and the proportion of variants of concern (VOC) (Supplementary text and Figure S7). Indeed, these two predictors, which are negligible over the evaluation period, can significantly affect the dynamic of hospitalizations observed from March 2021.

In order to determine whether the effects of the different predictors might have changed over time, we also retrospectively fit four of the individual models from June 3rd 2020 up to July 6th 2021, and compare the results with those obtained over the previous time period.

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 (11). 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:

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

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

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 bootstrapped smoothed trajectories of hospital admissions, to generate 95% prediction intervals for the three other targets.

To assess the forecasts of ICU admissions and bed occupancy in ICU and general wards, we report the MAPE and the 95% prediction interval coverage.

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 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), the European Union's Horizon 2020 research and innovation program under grants 101003589 (RECOVER) and 874735 (VEO), AXA, Groupama and EMERGEN.

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. Mecenas, 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. 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).

7. R. Polikar, Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine* **6**, 21–45 (2006).

8. E. L. Ray, *et al.*, Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S. *medRxiv*, 2020.08.19.20177493 (2020).

9. 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).

10. 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).

11. H. Salje, *et al.*, Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).

12. 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.

13. 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).

14. C. Viboud, *et al.*, The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics* **22**, 13–21 (2018).

15. R. J. Oidtman, *et al.*, Trade-offs between individual and ensemble forecasts of an emerging infectious disease. *Nat. Commun.* **12**, 5379 (2021).

16. 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).

17. T. Proietti, A. Luati, Real time estimation in local polynomial regression, with application to trend-cycle analysis. *aoas* **2**, 1523–1553 (2008).

18. 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).

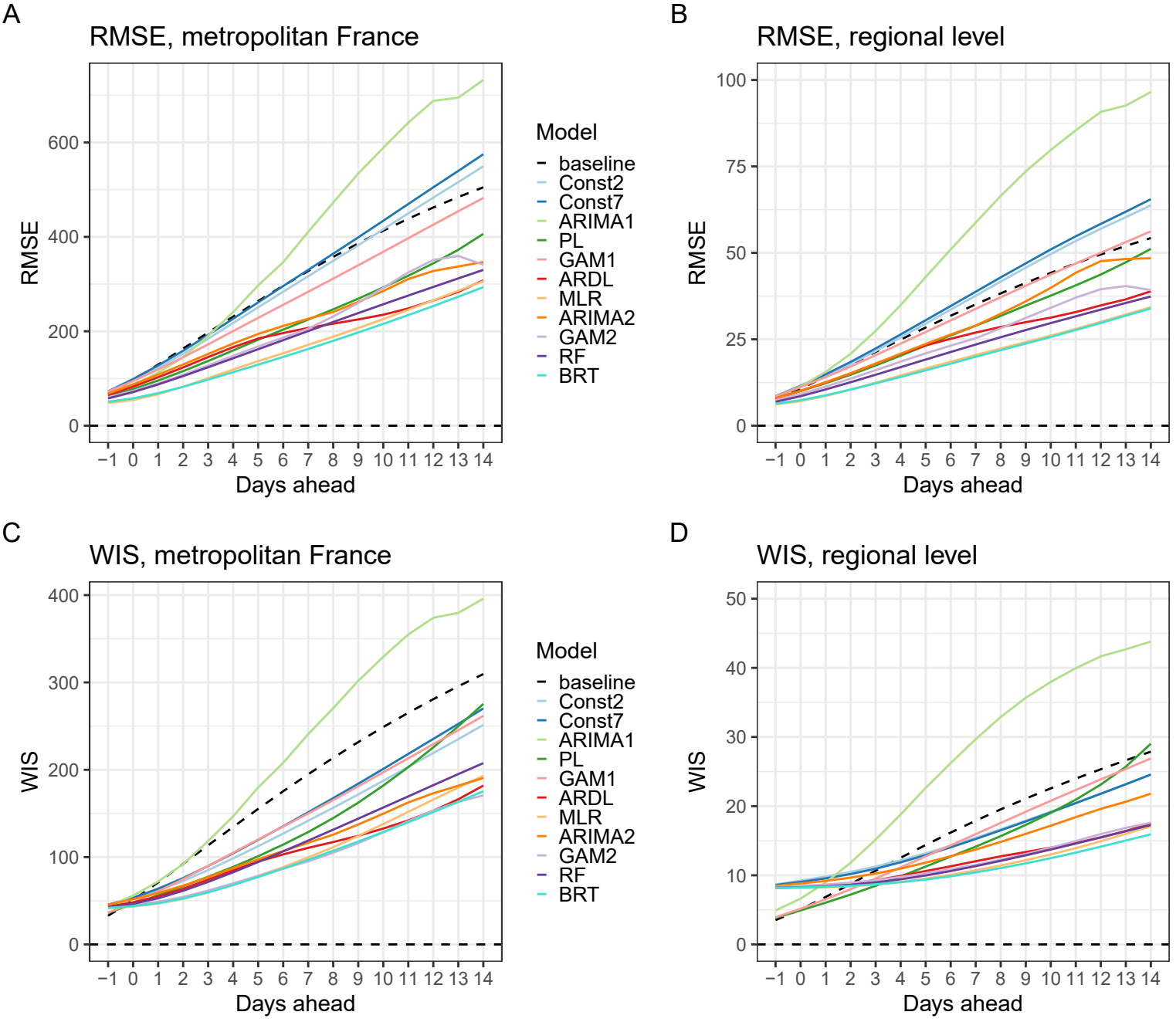
Figure legends

Figure 1: Comparison of the performance of individual models over the evaluation period at the national and regional levels, by prediction horizon, for hospital admissions. A. Root mean squared error (RMSE) in metropolitan France. B. RMSE by region. C. Mean weighted interval score (WIS) in metropolitan France. D. WIS by region.

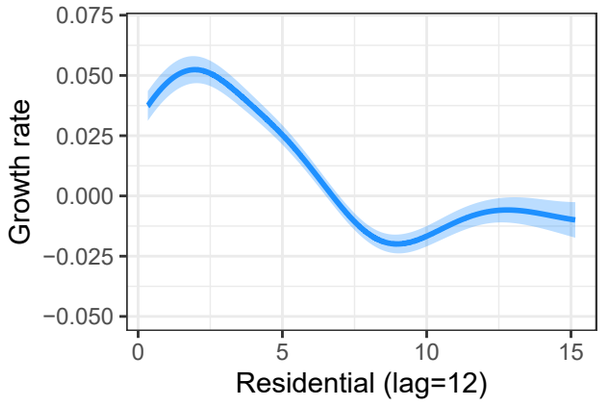
Figure 2: Effects of mobility (blue), epidemiological (green) and meteorological (red) predictors on the growth rate of hospital admissions, for the GAM2, the RF, the BRT and the MLR models. Abbreviations: GR = growth rate. Predictors are described in Supplementary text.

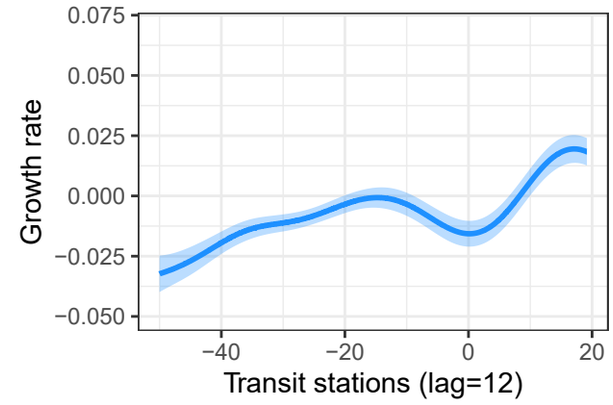
Figure 3: Forecasts of hospital admissions over the test period (March 7th 2021 to July 6th 2021). A. Trajectories predicted by the individual and ensemble models for all prediction horizons in metropolitan France. The black line is the eventually observed data (smoothed), and the colored lines are trajectories predicted on day t, for prediction horizons t-1 up to t+14. B. Forecasts of the ensemble model by region at 3, 7 and 14 days. The black line is the eventually observed data (smoothed). Shaded areas represent 95% prediction intervals. 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).

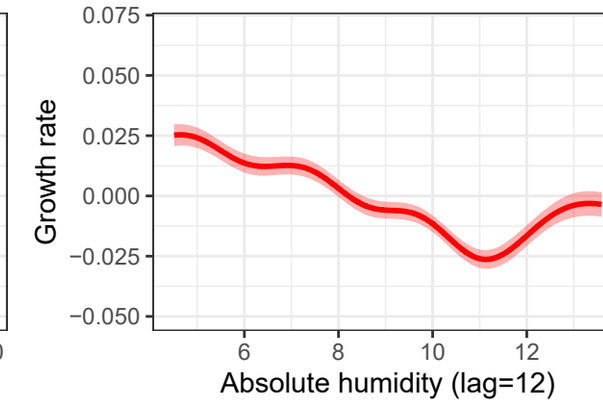
Figure 4: Model ranking by region and performance of the ensemble model for the four targets (hospital admissions, ICU admissions and bed occupancy in general ward and ICU) over the test period. A. Ranks of the models. Models are ranked according to the RMSE over all prediction horizons. Reg. Ave. = regional average (RMSE computed over all regions except metropolitan France). B. Mean absolute percentage error (MAPE) by prediction horizon in metropolitan France. C. MAPE by prediction horizon at the regional level.


A
RMSE, metropolitan France
RMSE
Days ahead
B
RMSE, regional level
RMSE
Days ahead
C
WIS, metropolitan France
WIS
Days ahead
D
WIS, regional level
WIS
Days ahead
Model
baseline
Const2
Const7
ARIMA1
PL
GAM1
ARDL
MLR
ARIMA2
GAM2
RF
BRT

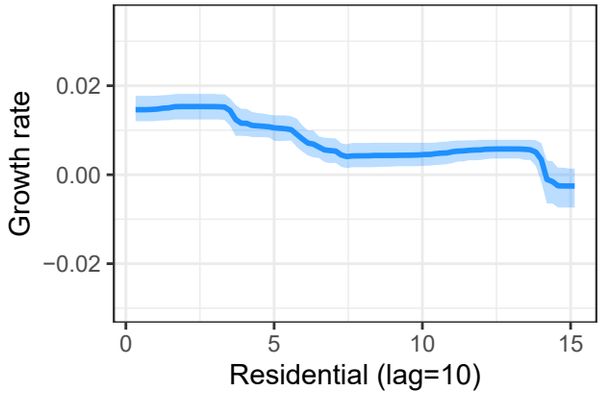
GAM2

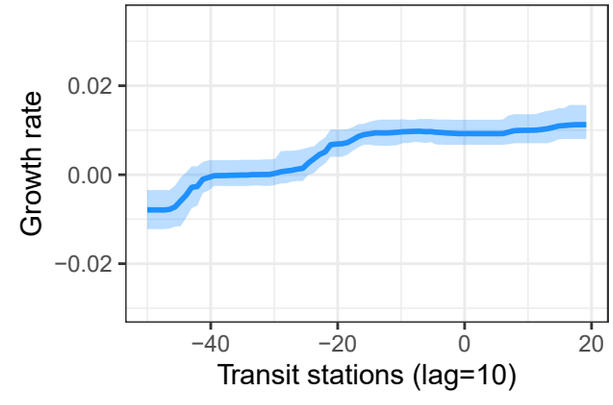


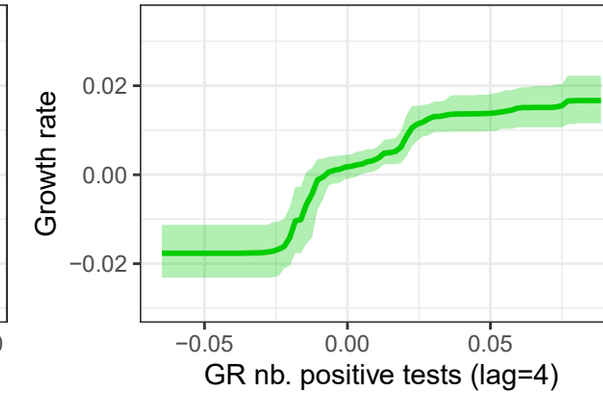




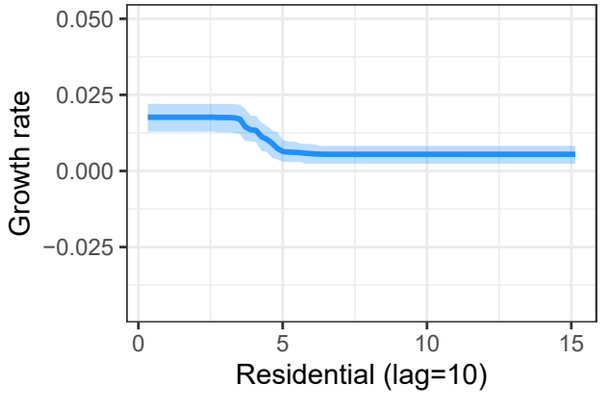
RF

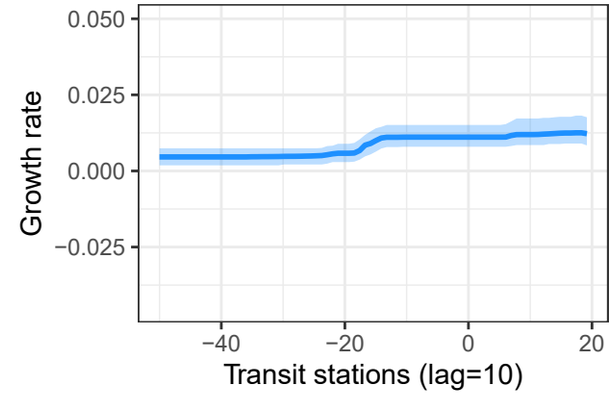


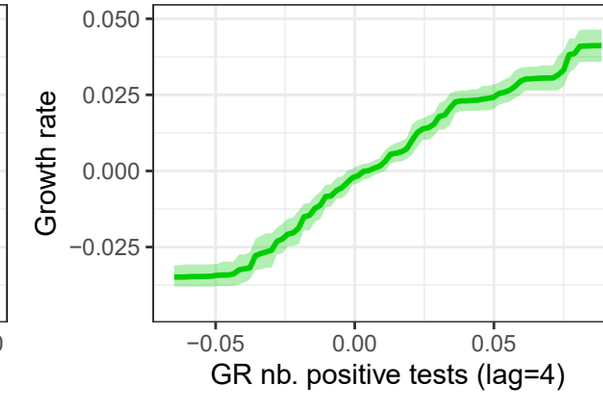




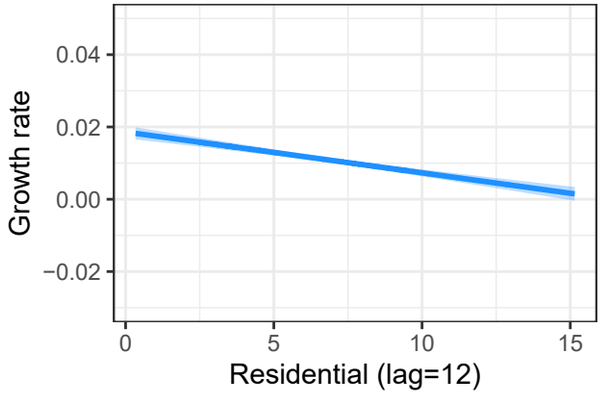
BRT

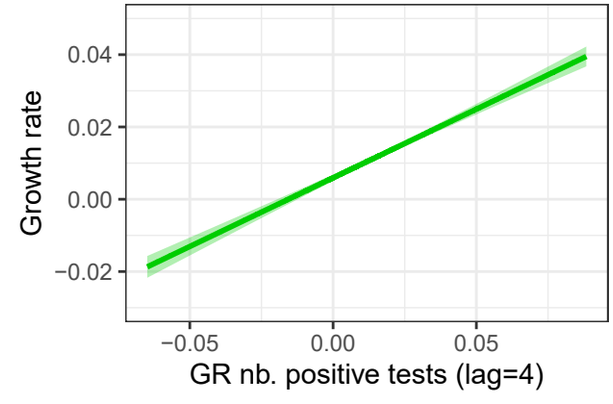


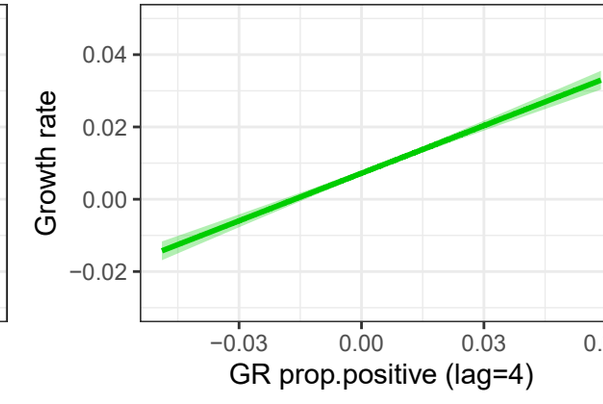


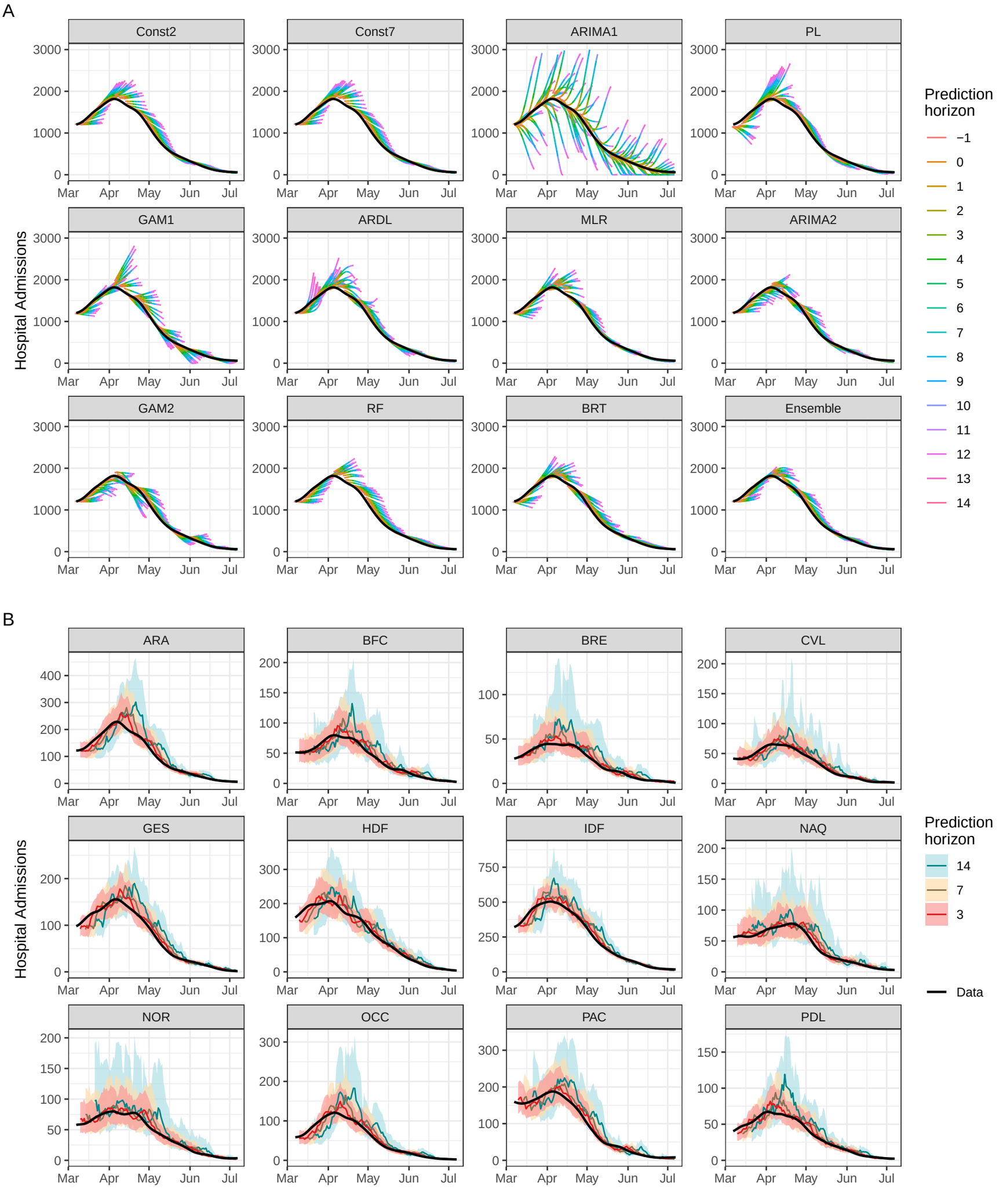


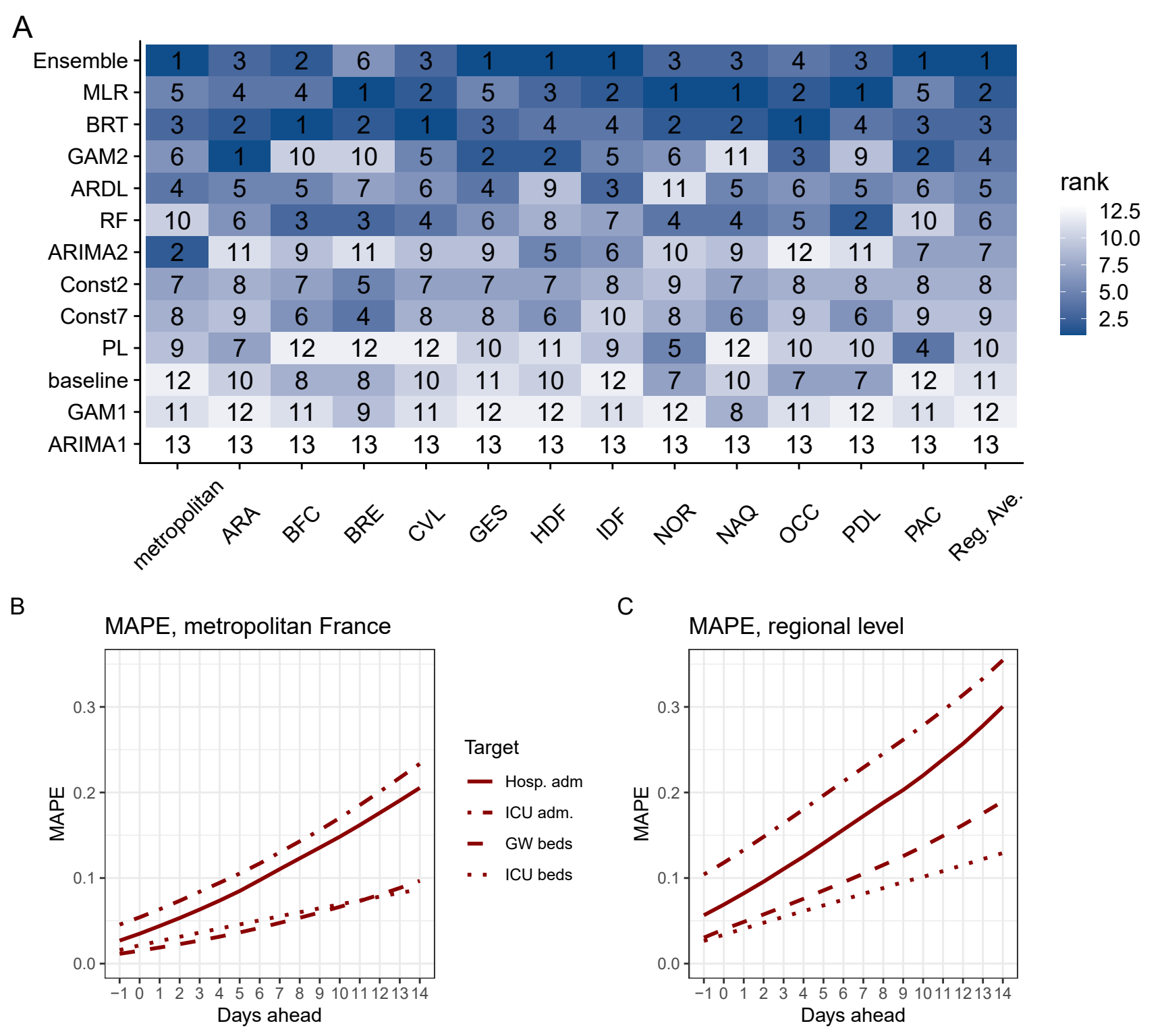
MLR








A
Const2
Const7
ARIMA1
PL
GAM1
ARDL
MLR
ARIMA2
GAM2
RF
BRT
Ensemble
Hospital Admissions
Mar
Apr
May
Jun
Jul
Prediction horizon
−1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
B
ARA
BFC
BRE
CVL
GES
HDF
IDF
NAQ
NOR
OCC
PAC
PDL
Hospital Admissions
Prediction horizon
14
7
3
Data


A
Ensemble 1 3 2 6 3 1 1 1 3 3 4 3 1 1
MLR 5 4 4 1 2 5 3 2 1 1 2 1 5 2
BRT 3 2 1 2 1 3 4 4 2 2 1 4 3 3
GAM2 6 1 10 10 5 2 2 5 6 11 3 9 2 4
ARDL 4 5 5 7 6 4 9 3 11 5 6 5 6 5
RF 10 6 3 3 4 6 8 7 4 4 5 2 10 6
ARIMA2 2 11 9 11 9 9 5 6 10 9 12 11 7 7
Const2 7 8 7 5 7 7 7 8 9 7 8 8 8 8
Const7 8 9 6 4 8 8 6 10 8 6 9 6 9 9
PL 9 7 12 12 12 10 11 9 5 12 10 10 4 10
baseline 12 10 8 8 10 11 10 12 7 10 7 7 12 11
GAM1 11 12 11 9 11 12 12 11 12 8 11 12 11 12
ARIMA1 13 13 13 13 13 13 13 13 13 13 13 13 13 13
metropolitan
ARA
BFC
BRE
CVL
GES
HDF
IDF
NOR
NAQ
OCC
PDL
PAC
Reg. Ave.
rank
12.5
10.0
7.5
5.0
2.5
B
MAPE, metropolitan France
MAPE
0.3
0.2
0.1
0.0
−1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Days ahead
Target
Hosp. adm
ICU adm.
GW beds
ICU beds
C
MAPE, regional level
MAPE
0.3
0.2
0.1
0.0
−1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Days ahead

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 was implemented in March 2020 and is maintained by the ANS (Agence du Numérique en Santé). It 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. Therefore, in the absence of smoothing or with simple smoothing techniques, forecasts can be biased depending on the day of the week at which the analysis is performed. In order to remove day-to-day variation and obtain a smooth signal at each date T not depending on future data points (mimicking the real-time case), we adopt a 2-step approach using state-of-the-art statistical methods and using data only up to date T:

1 – Removal of the day-of-the-week pattern of the data up to current time T. We assume that the logged incidence $y(t)$ can be written as $y(t) = m(t) + w(d(t)) + \epsilon(t)$, where *m(t)* is a smooth temporal trend, *d(t)* is the day of the week at date *t* and *w(d)* is the day of the week effect and $\epsilon(t)$ is noise. We estimate $\widehat{w}(d)$ by fitting a local polynomial regression using the standard biweight kernel with bandwidth *h=8* days (corresponding to the number of time points below and

above used in the smoothing) over the previous 8 weeks of data before date T - not using future data. Local polynomial regression is a state-of-the-art kernel smoother, less biased than simple rolling average because it uses an (optimal) biweight kernel rather than the rectangle kernel and less biased than the classical Nadaraya-Watson estimator even close to the boundaries of the interval of estimation. This regression leads to a trend estimate $m_r(t)$ from the raw data that we compute over an interval in the past [T - 8 weeks +h, T - h], excluding the last *h* days. We then compute the day-of-the-week effect $\hat{w}(d)$ by averaging $y_d(t) = y(t) - m_r(t)$ for each day d of the week. Finally, we output a new series where the day-of-the-week effect has been removed up to time T as $\hat{y}_w(t) = y(t) - \hat{w}_d(t)$.

2 – Smoothing incidence up to time T accounting for real time. The second step allows obtaining a signal that is fit for real time analysis. Indeed, even if local polynomial smoothing is nearly unbiased close to the boundaries of estimation, this comes with increased variance: this means that the estimated trend in the last few days of observation can be misleading. Several approaches are possible to overcome this limitation: automatic kernel curtailing or selecting smoothing according to the least revision principle. Proietti *et al.* described a framework for implementing the least revision principle using low-order reproducing kernels (2). In this approach, the smoothing kernels are tailored to minimize the error between the smooth value predicted in real-time, when data is available only up to time T, and the value that will be obtained as a final estimate once data is present up to time T+h. The method introduces a little bias to reduce the variance of the estimate. As described in Proietti and al, we assume that $\hat{y}_w(t) = m_w(t) + \epsilon_w(t)$ where $m_w(t)$ is a smooth trend estimate and $\epsilon_w(t)$ is noise. The smooth trend is estimated using the linear/quadratic approach of Proietti, whereby a local polynomial of degree 2 is fit on the whole range of the data and is approximated by a local polynomial of degree 1 on the interval [T-h, T] for computation of the real-time smooth estimate of the trend. This last approximation introduces bias and reduces variance, leading to a "least revision" estimate. Confidence intervals are computed by bootstrap.

We compare this two-step algorithm with a simple smoothing spline (Fig. S11). 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).

We also compare the predictions of the MLR model (taken as an example) using our smoothing algorithm to the predictions made on data smoothed by a centered 7-day moving average (MA). The MA method leads to the loss of the last three data points but is often used to remove weekly patterns in a time-series, due to its simplicity. We show that the RMSE of the predictions over the evaluation period is lower with the smoothing algorithm compared to the moving average (Figure S12).We also show that, with our smoothing algorithm, the RMSE is stable throughout the week, while it varies with the day of the week when using a moving average (Figure S12).

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.

Baseline

The baseline model assumes that the number of hospital admissions stays at its current value indefinitely into the future, with uncertainty levels given by a discretised truncated normal distribution with lower bound 0 and a standard deviation given by past one-day ahead deviations from the value of the metric (3).

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.

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:

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

Here, the τ_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:

$$r(t) = \beta_0 + \sum_k \beta_k . X_k\left(t - l_k\right)$$

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):

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

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:

$$r(t) = \beta_0 + \sum_k \beta_k . X_k\left(t - l_k\right) + \eta_t$$

where η_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:

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

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. Unlike 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 over the evaluation period: 9 epidemiological predictors describing the dynamics of the epidemics, 6 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). In addition, over the test period, we also include vaccine coverage and the proportion of variants of concern (VOC) (Figure S7) as these two covariates can significantly affect the dynamic of hospitalizations from March 2021.

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 SIDEP 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. This surveillance system was implemented from May 13th 2020 and became stable in June 2020.

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/). 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.

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

Vaccine coverage

Vaccine coverage data are obtained from the VAC-SI database, the national information system developed by the French Health Insurance to monitor the deployment of the vaccine campaign. Daily data are made publicly available by Santé publique France (https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-1/). We use the proportion of the population completely vaccinated (i.e. people who received 2 doses in a 2-dose vaccination scheme or 1 dose in a 1-dose scheme) (Fig. S7).

Proportion of variants of concern (VOC)

We obtain data on the proportion of variants of concern (VOC) detected in nasopharyngeal samples, to capture the progressive replacement of the historical strain by more transmissible variants. We use the SIDEP (Système d'Information de Dépistage Populationnel - Information system for population-based testing) database, the national surveillance system describing RT-PCR and antigen tests results for SARS-CoV-2 arising from all 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 postal code of the home address. Aggregated data are made publicly available by Santé publique France (https://www.data.gouv.fr/fr/datasets /donnees-de-laboratoires-pour-le-depistage-indicateurs-sur-les-variants/). VOC were identified among positive PCR or antigen test results, using RT-PCR screening kits. The main VOC circulating during the study period was Alpha variant, followed by Beta and Gamma variants. Data are available from February 15, 2021 to June 9, 2021. In order to impute the proportion of VOC before and after these dates, we fit a logistic regression model, assuming that the proportion of VOC was zero before December 15, 2020 (Fig. S7).

Forward selection of predictors

In order to select the best predictors to include in individual models, we use a forward stepwise selection method (5), using data from the evaluation period only. We first include all covariates (N=19) in univariate models and run each univariate model over the evaluation period, using a rolling forecasting origin approach (cross-validation): for each day *t* of the evaluation period, we make forecasts for the period *t-1* up to day *t+14*, using only past data up to day *t-2* as a training set, and computing evaluation metrics using the observed data in *t-1* to *t+14*. For each univariate model, we compute the RMSE of predictions at *t+7* and *t+14* and we retain the covariate that minimizes the RMSE. We then include the remaining covariates one by one, until no additional covariate can decrease the cross-validated RMSE by more than 1. 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 RMSE among the three multivariate models.

Supplementary figures

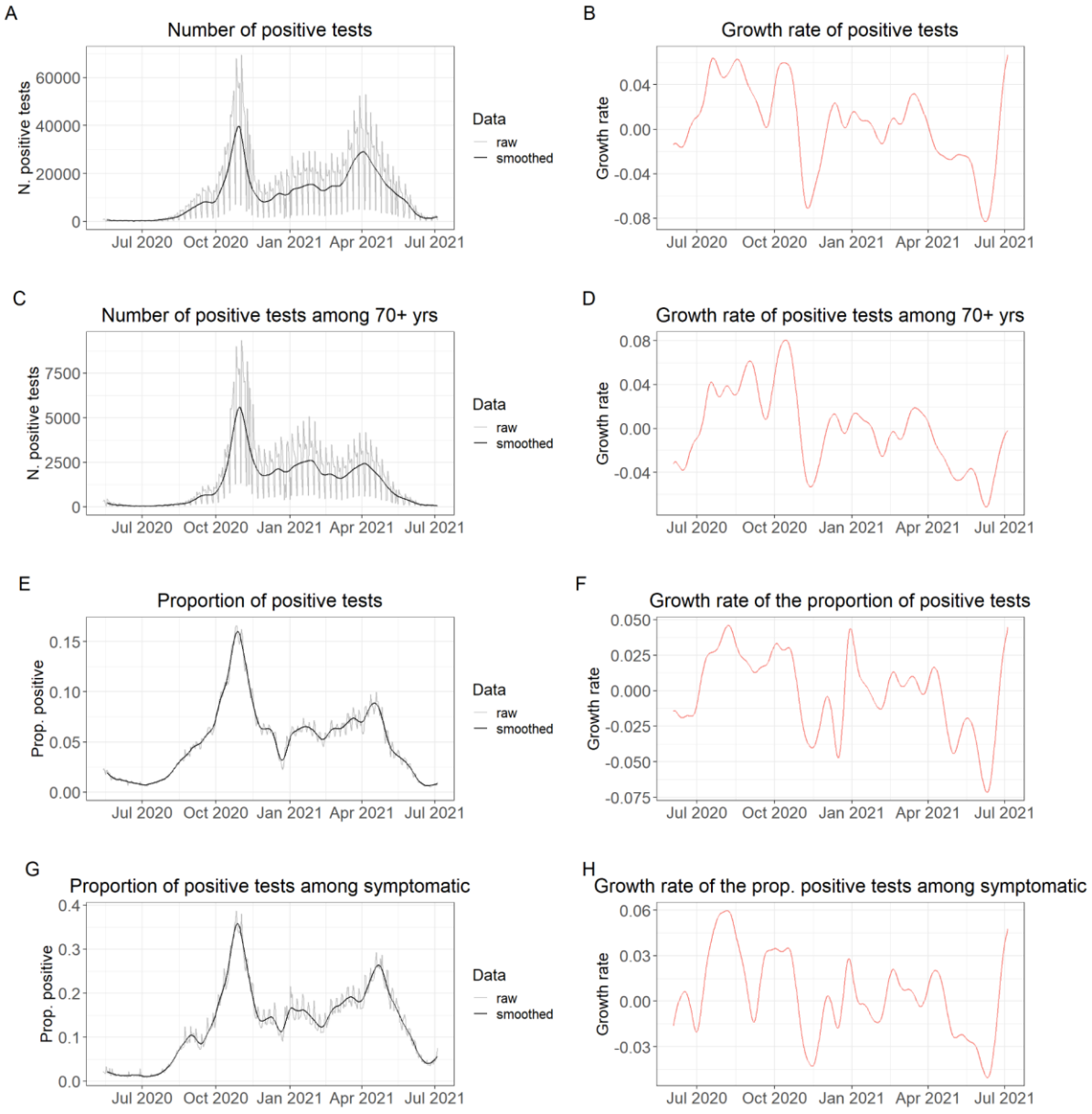


Fig. S1: Epidemiological predictors (see also Supplementary text).

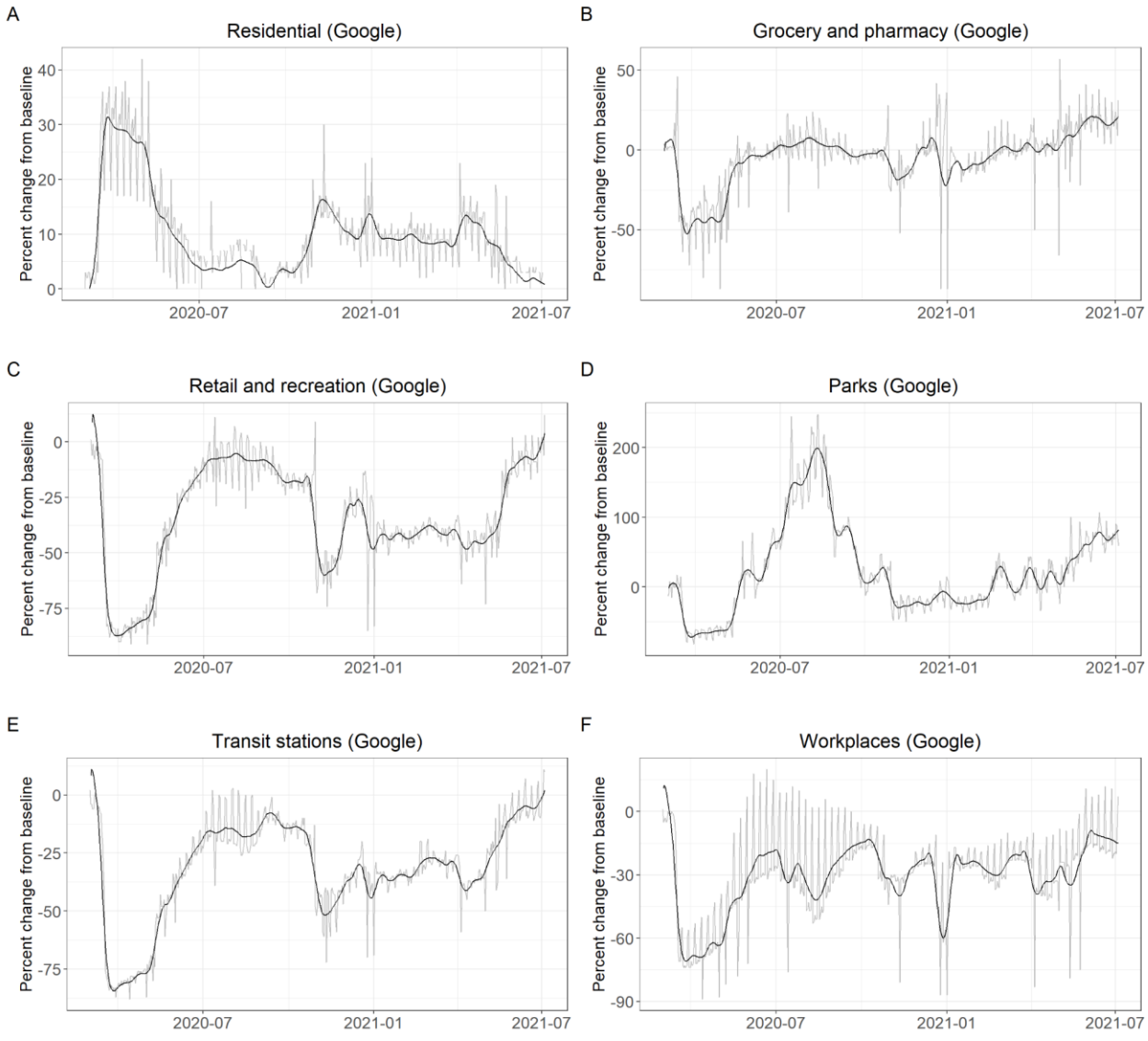


Fig. S2: Mobility predictors (see also Supplementary text).

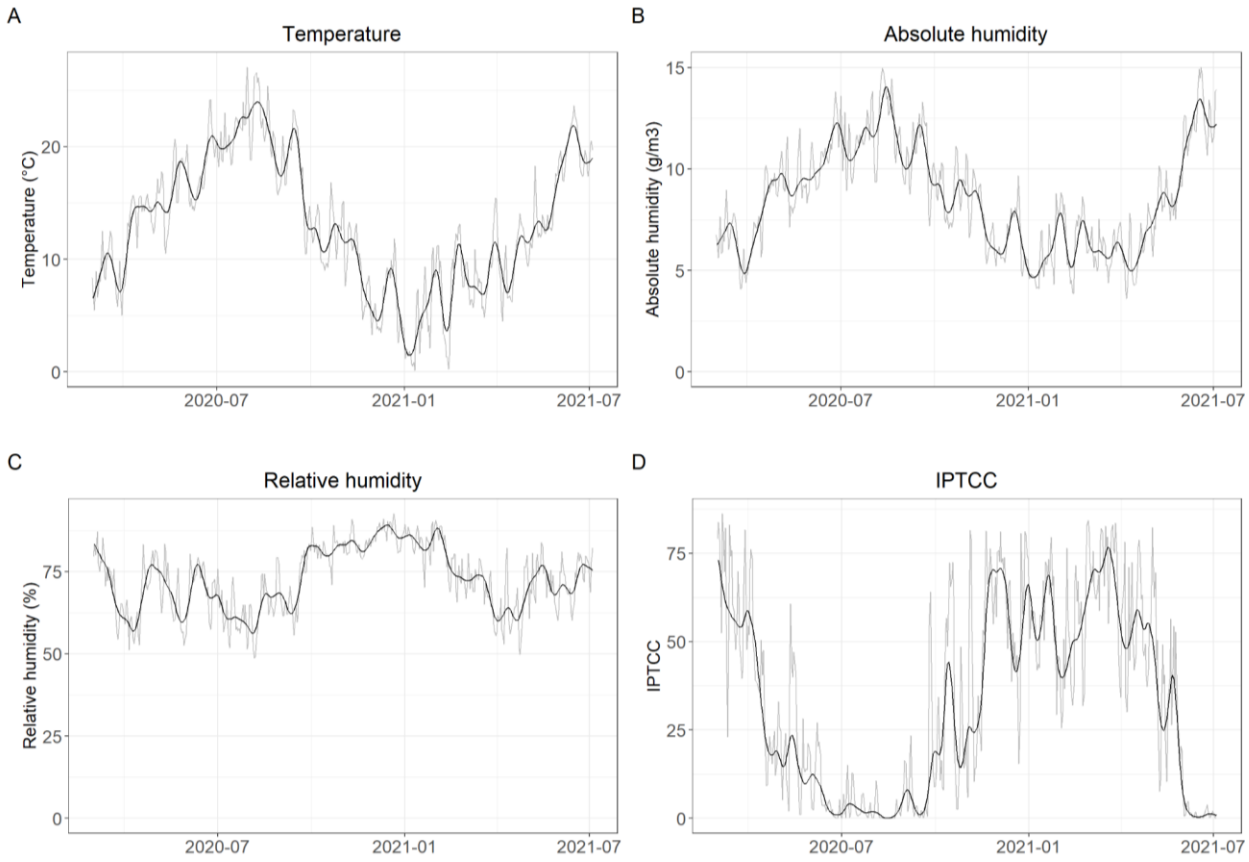


Fig. S3: Meteorological predictors (see also Supplementary text).

Fig. S4: Trajectories of hospital admissions predicted by the 12 individual models in metropolitan France over the evaluation period. 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 March 6th 2021. We exclude the forecasts made between October 20th and November 4th (i.e. up to 6 days into the lockdown starting on October 30th) for hospitalizations occuring after November 3rd, as the models were not designed to anticipate the impact of a lockdown before its implementation. The excluded forecasts are shown with transparent lines and a grey shaded area.

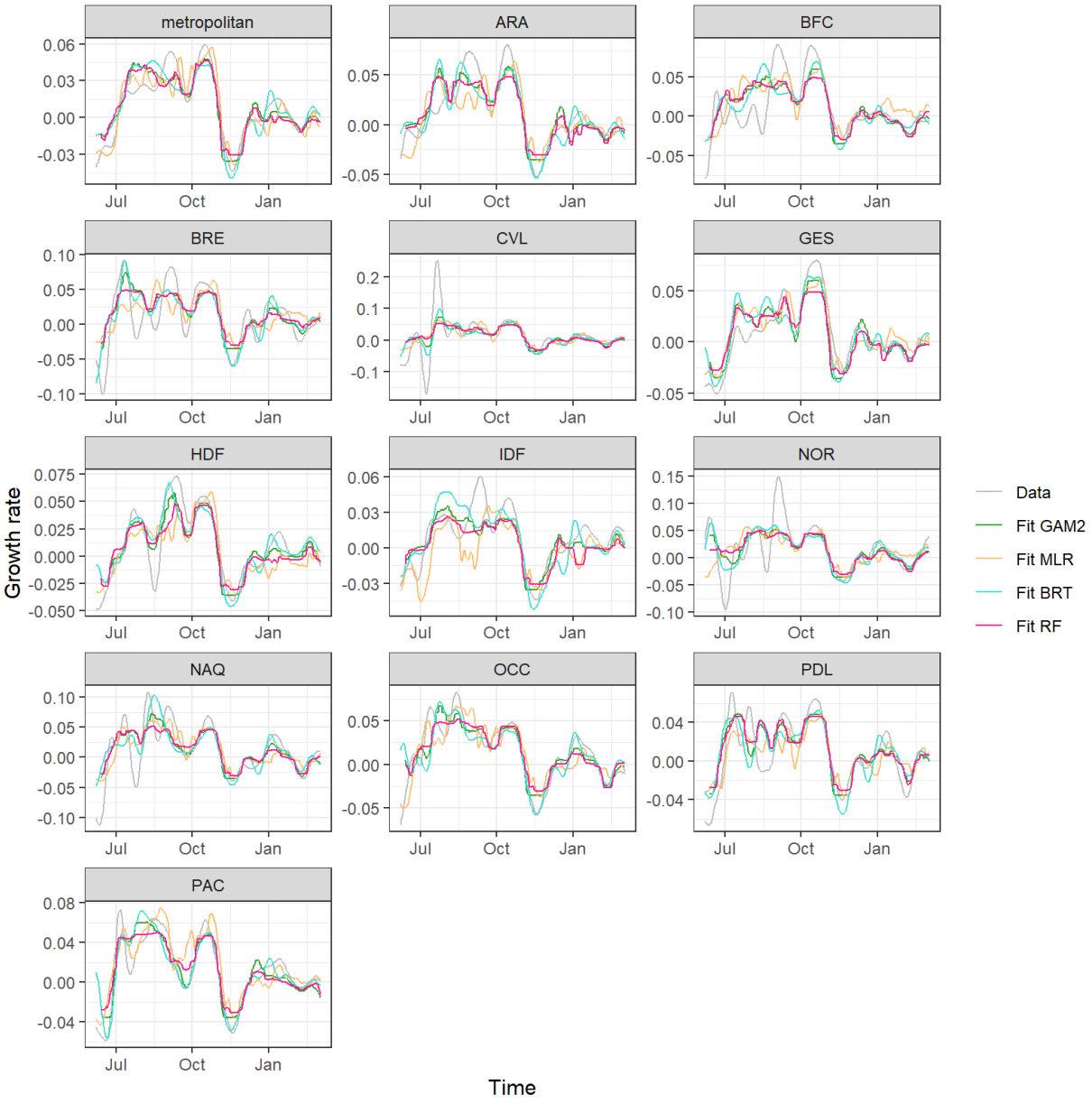


Fig. S5: Model fits for the growth rate of hospital admissions, at the national (metropolitan France) and regional levels, for the GAM2, the MLR, the BRT and the RF models. Each panel shows the observed growth rate (black line) and the predicted growth rates (colored lines) when retrospectively fitting each model from June 3rd 2020 to March 6th 2021, 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).

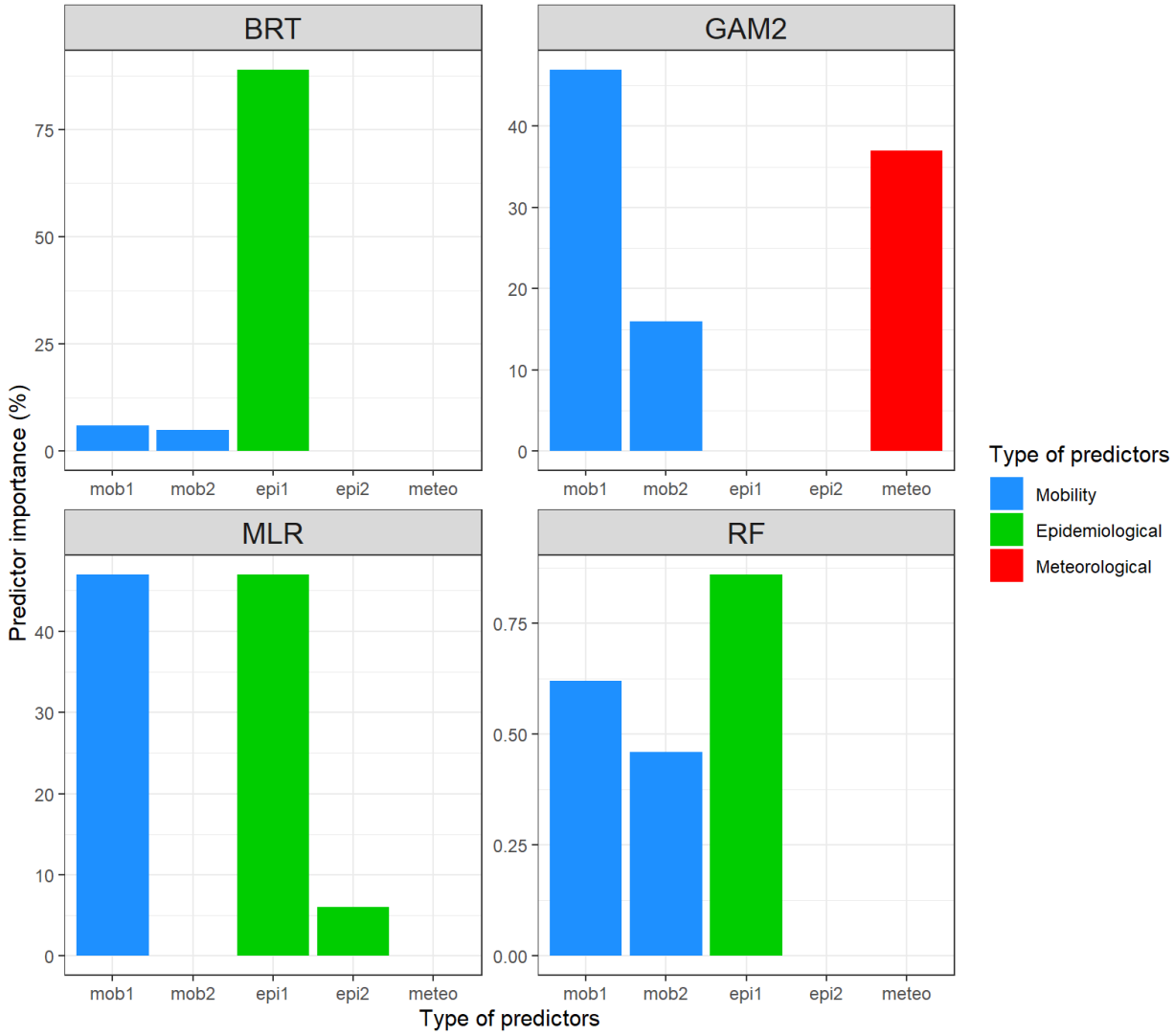


Fig. S6: Importance of predictors, estimated by retrospectively fitting the models from June 3rd 2020 to March 6th 2021**.** For the BRT model, relative importance is a measure of how each predictor contributes to reducing the loss function (all contributions sum to 100%). For the MLR and the GAM2 models, relative importance is a measure of how each predictor contributes to the total explained variance (all contributions sum to 100%). For the RF model, 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%).

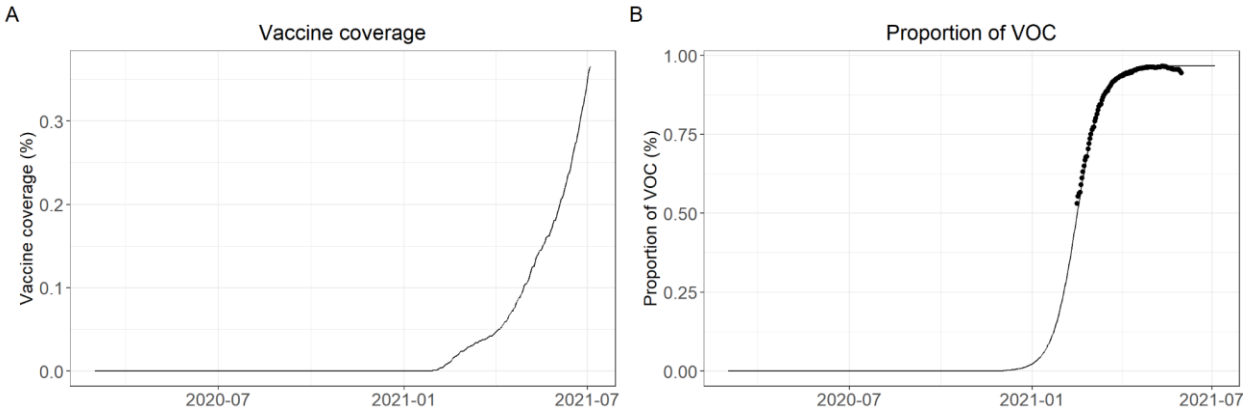


Fig. S7: Additional predictors used for the test period. (A) Vaccine coverage (complete vaccination scheme). (B) Proportion of variants of concern (VOC). The points represent the available data and the line represents the fit of the logistic model (see Supplementary text).

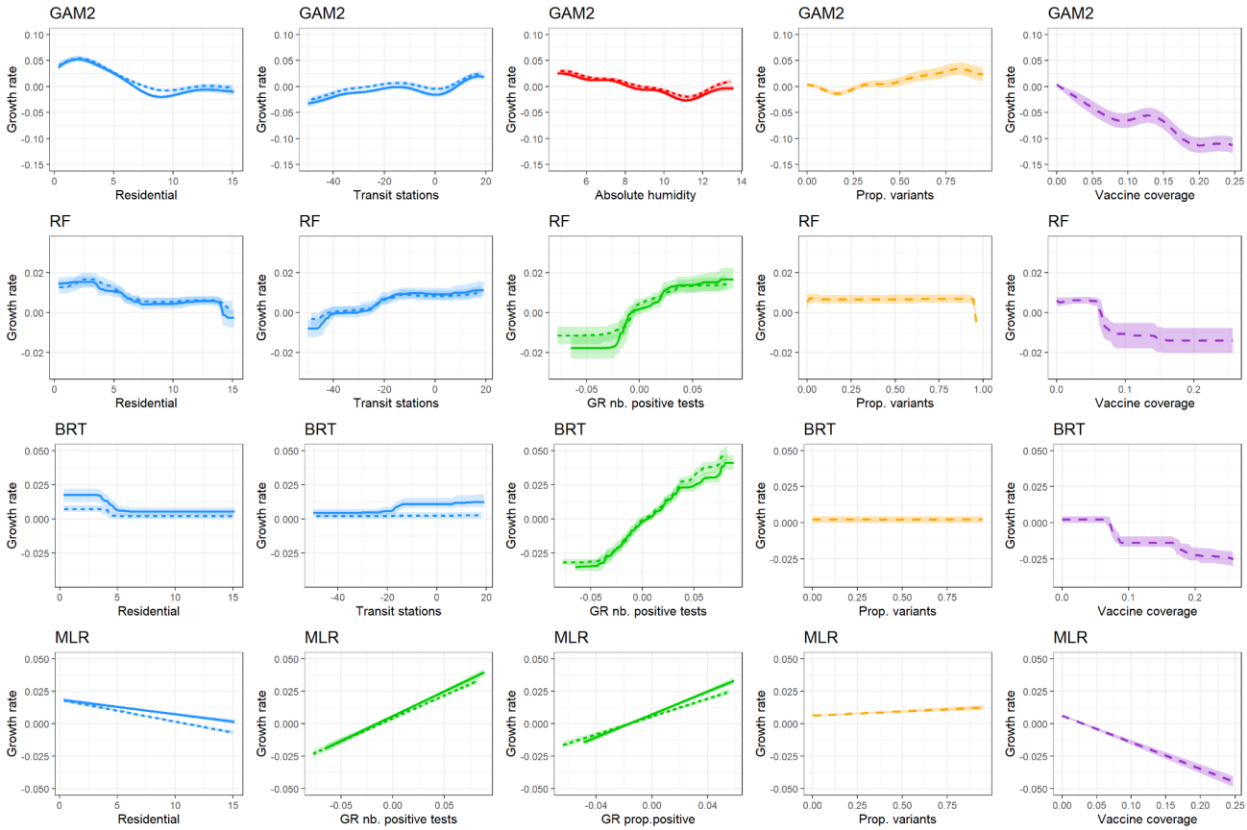


Figure S8: Effects of mobility (blue), epidemiological (green), meteorological (red), proportion of VOC (orange) and vaccine coverage (purple) predictors on the growth rate of hospital admissions, for the GAM2, the RF, the BRT and the MLR models, by retrospectively fitting the models over two time periods: from June 3rd 2020 to March 6th 2021 (solid lines) or from June 3rd 2020 to July 7th 2021 (dashed lines). Abbreviations: GR = growth rate. Predictors are described in Supplementary text.

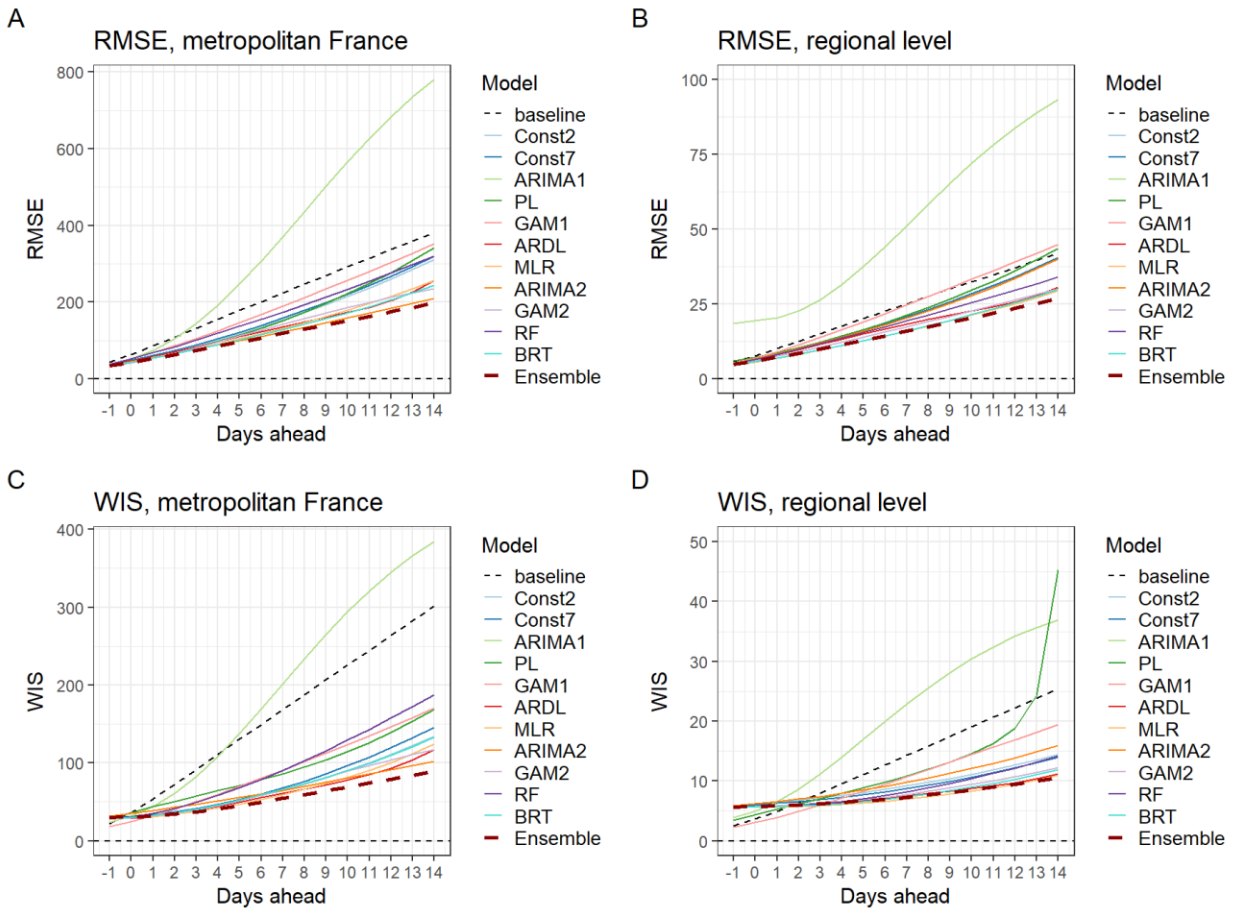


Figure S9: Comparison of the performance of the individual and ensemble models over the test period at the national and regional levels, by prediction horizon, for hospital admissions. A. Root mean squared error (RMSE) in metropolitan France. B. RMSE by region. C. Mean weighted interval score (WIS) in metropolitan France. D. WIS by region.

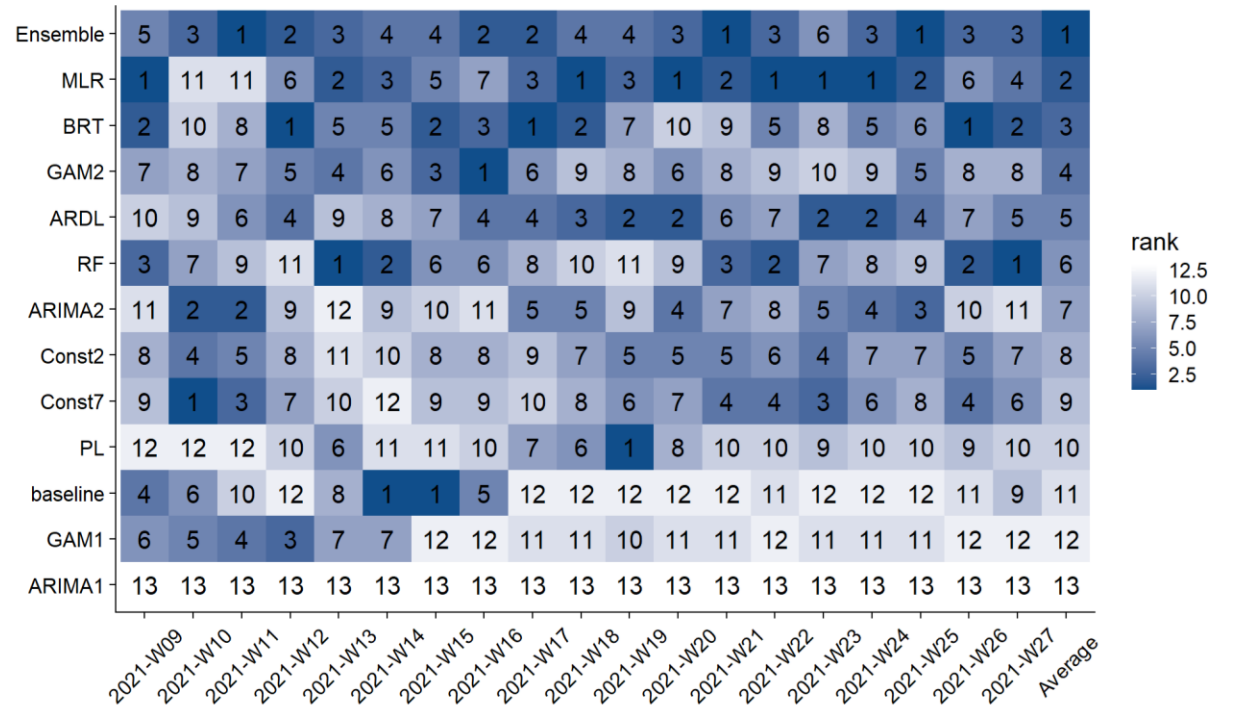


Figure S10: Model ranking by week. Models are ranked according to the RMSE over all prediction horizons and all regions.

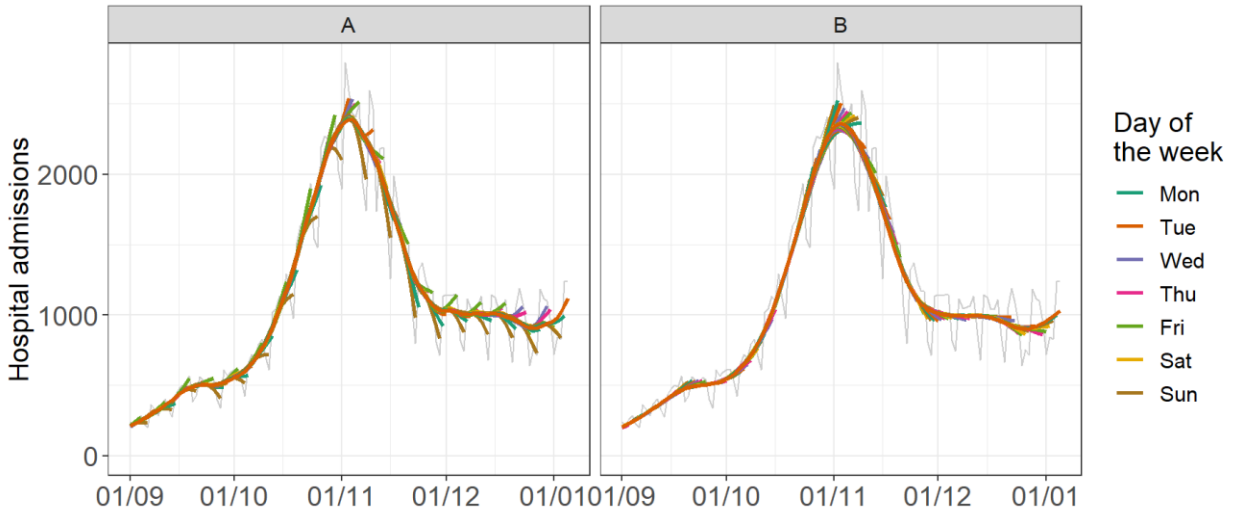


Fig. S11: Comparison of smoothing methods. (A) Smoothing spline. (B) Our two-step algorithm. The grey line shows the raw data of hospital admissions in metropolitan France from September 2020 to January 2021.The colored lines are time-series smoothed in real time (i.e. knowing only the past values), with different colors indicating the day of the last data point.

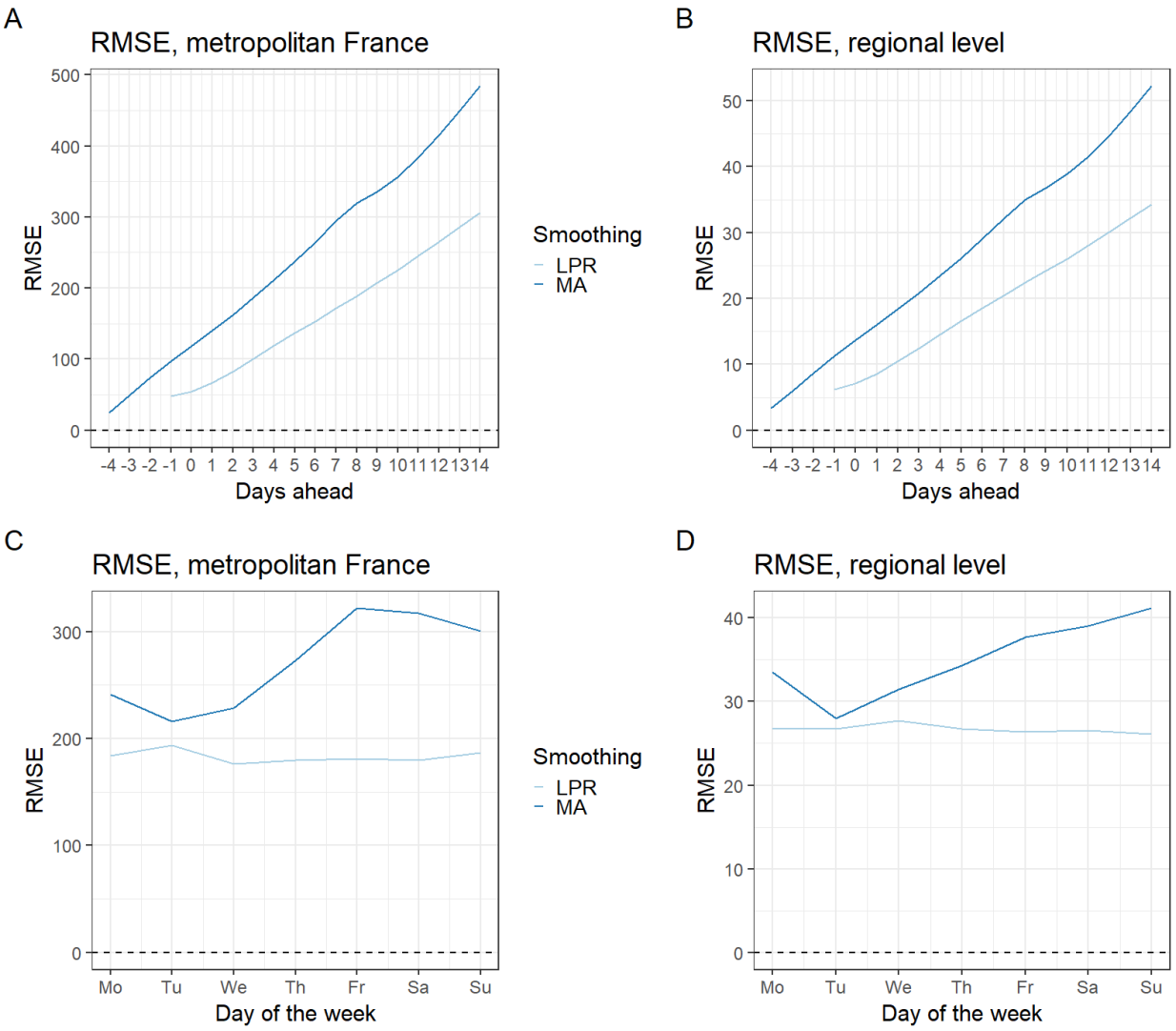


Fig. S12: Comparison of RMSE for hospital admissions over the evaluation period, at the national and regional level, for the multiple linear regression (MLR) model, according to the smoothing method (LPR = local polynomial regression, MA = 7-day moving average). A. RMSE by prediction horizon in metropolitan France. B. RMSE by prediction horizon at the regional level. C. RMSE according to the day of the week at which the predictions were made, in metropolitan France. D. RMSE according to the day of the week at which the predictions were made, at the regional level.

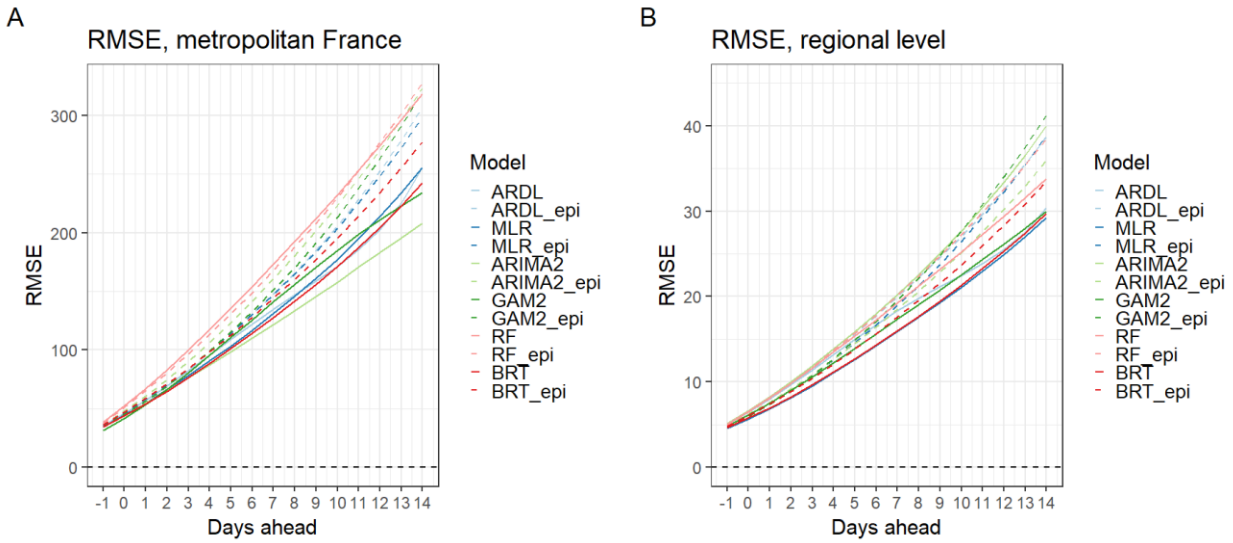


Figure S13: Comparison of models with all types of predictors (solid lines) *vs* models with epidemiological predictors only (denoted "_epi", dashed lines), over the test period. In general, the models with all types of predictors perform better than the models with epidemiological predictors only.

Supplementary tables

Table S1: Best predictors selected for each individual model using the forward stepwise selection procedure over the evaluation period. The six models were included in the ensemble model.

Model	Dependent variable	Are lagged values of the dependent variable used as covariate?	Epidemiological predictors	Mobility predictors	Meteoro-logical predictors
ARDL	Growth rate of hospital admissions	Yes	Growth rate of the number of positive tests Growth rate of the proportion of positive tests among tests in symptomatic people	Residential	Temperature
MLR	Growth rate of hospital admissions	No	Growth rate of the number of positive tests Growth rate of the proportion of positive tests	Residential	
GAM	Growth rate of hospital admissions	No		Residential and transit stations	Absolute humidity
ARIMA2	Growth rate of hospital admissions	No		Transit stations and residential	
BRT	Growth rate of hospital admissions	No	Growth rate of the number of positive tests	Transit stations and residential	
RF	Growth rate of hospital admissions	No	Growth rate of the number of positive tests	Transit stations and residential	

Table S2: 95% prediction interval coverage of the ensemble model, at the national and regional level, for 7- and 14-day ahead forecasts, over the test period.

Level	Target	7-day ahead	14-day ahead
National	Hospital admissions	0.76	0.69
	ICU admissions	0.96	0.81
	General wards beds	0.90	0.84
	ICU beds	0.79	0.83
Regional	Hospital admissions	0.89	0.80
	ICU admissions	0.95	0.91
	General wards beds	0.90	0.90
	ICU beds	0.93	0.96

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. S. Funk, *et al.*, Short-term forecasts to inform the response to the Covid-19 epidemic in the UK. *bioRxiv* (2020) https:/doi.org/10.1101/2020.11.11.20220962.

4. 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).

5. R. J. Hyndman, G. Athanasopoulos, *Forecasting: principles and practice* (OTexts: Melbourne, Australia, 2021).