PNAS
www.pnas.org

Supplementary information for

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

Juliette Paireau[†*], Alessio Andronico[†], Nathanaël Hozé, Maylis Layan, Pascal Crépey, Alix Roumagnac, Marc Lavielle, Pierre-Yves Boëlle, Simon Cauchemez

†These authors contributed equally to this work.

*Corresponding author. Email: juliette.paireau@pasteur.fr

This PDF file includes:

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 $\widehat{w}(d)$ 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 $\hat{y}_w(t) = y(t) - \widehat{w}_d(t)$.

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. S14). 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 training period is lower with the smoothing algorithm compared to the moving average (Figures S15 and S16). 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 (Figures S15 and S16).

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 training 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 S9) 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. S2):

- 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. S3). 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. S4). 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. S9).

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. S9).

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 training period only. We first include all covariates (N=19) in univariate models and run each univariate model over the training period, using a rolling forecasting origin approach (cross-validation): for each day *t* of the training 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 smoothed 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

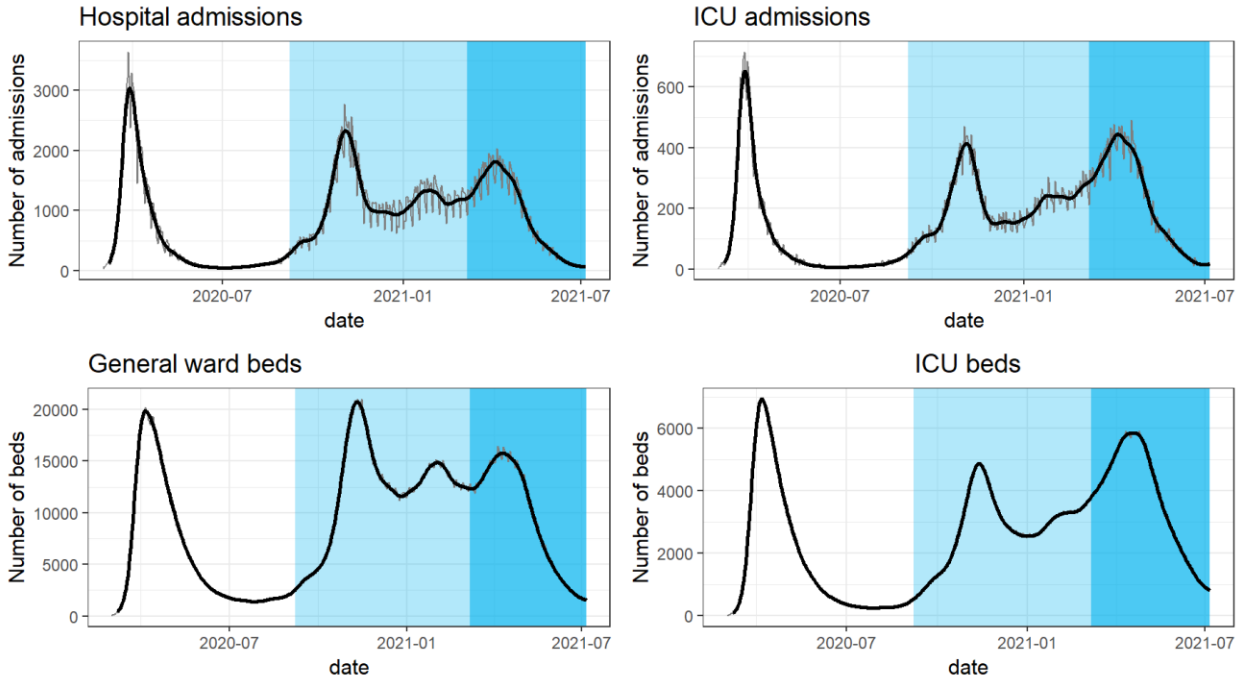

Fig. S1: Time-series of hospital admissions, Intensive Care Unit (ICU) admissions, general ward beds and ICU beds in metropolitan France. The two periods highlighted in blue indicate the training period (September 7th 2020 to March 6th 2021) and the test period (March 7th 2021 to July 6th 2021). The raw series are shown in grey and the smoothed series in black.

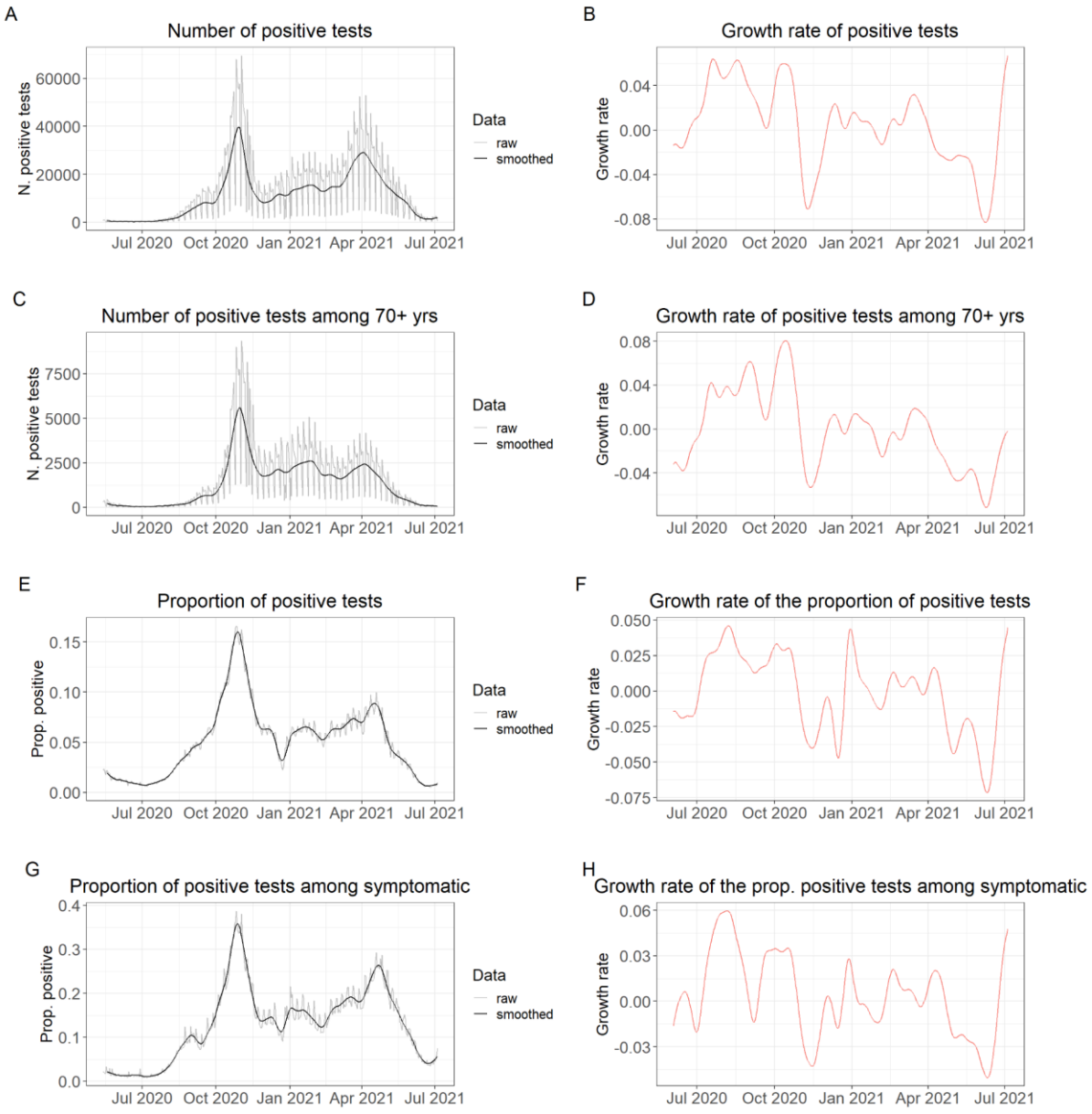

Fig. S2: Epidemiological predictors (see also Supplementary text).

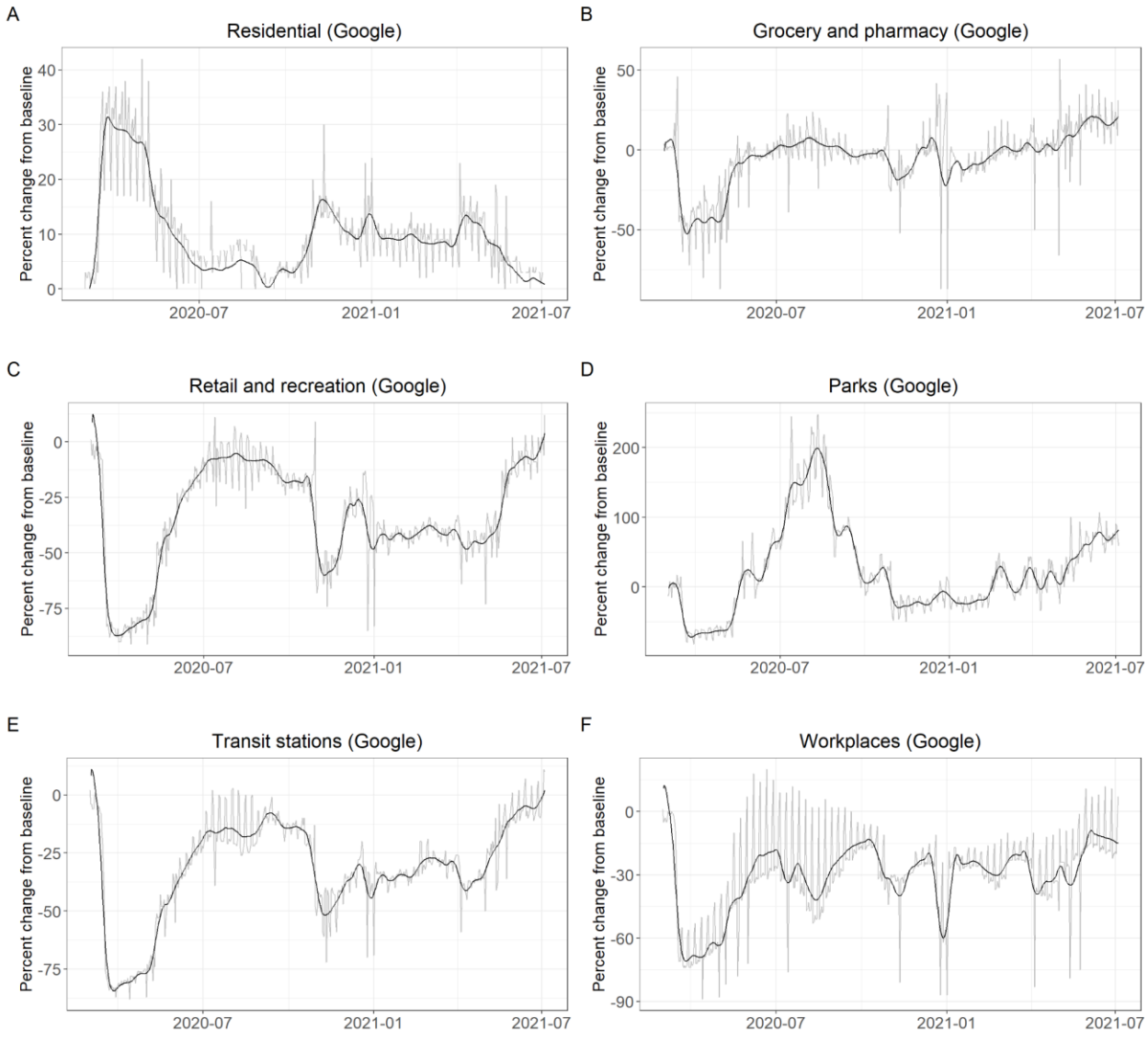

Fig. S3: Mobility predictors (see also Supplementary text).

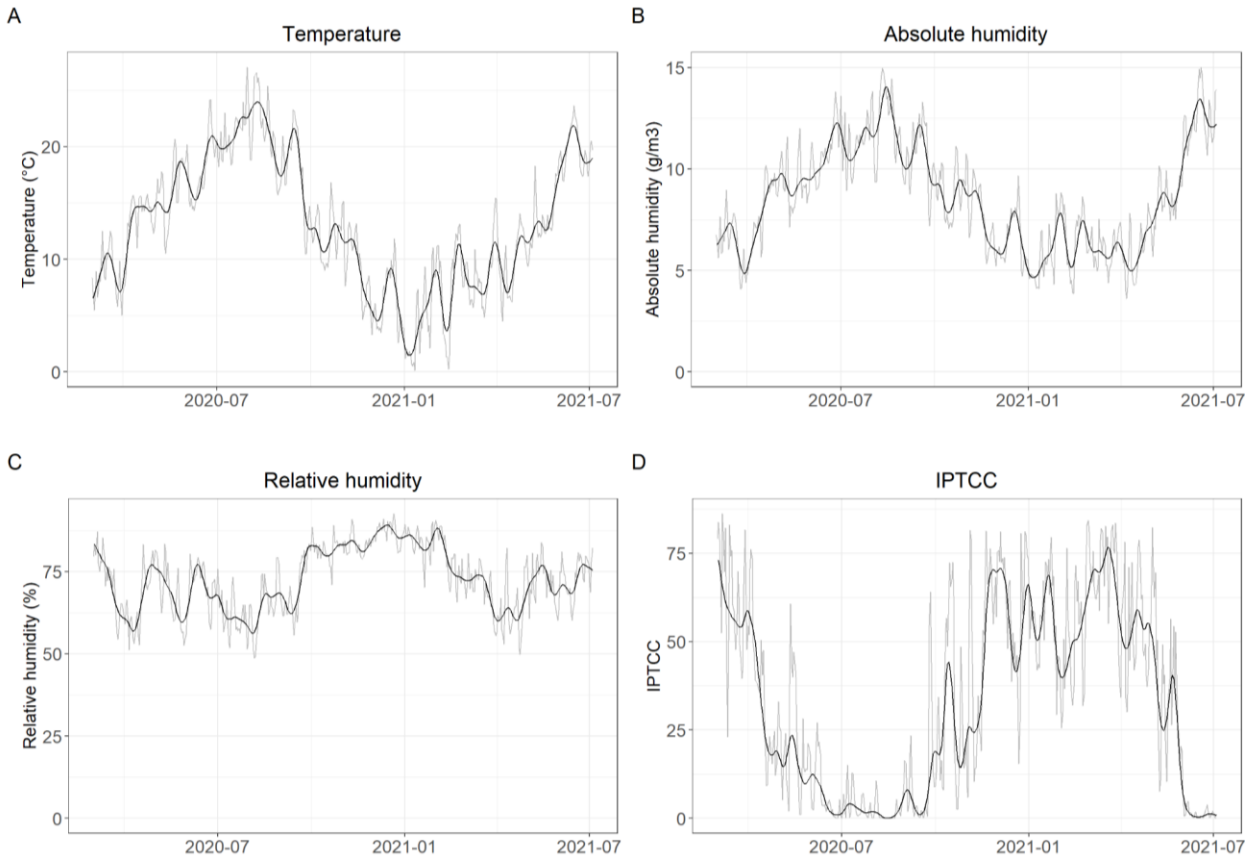

Fig. S4: Meteorological predictors (see also Supplementary text).

Fig. S5: Trajectories of hospital admissions predicted by the 12 individual models in metropolitan France over the training 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 training 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.

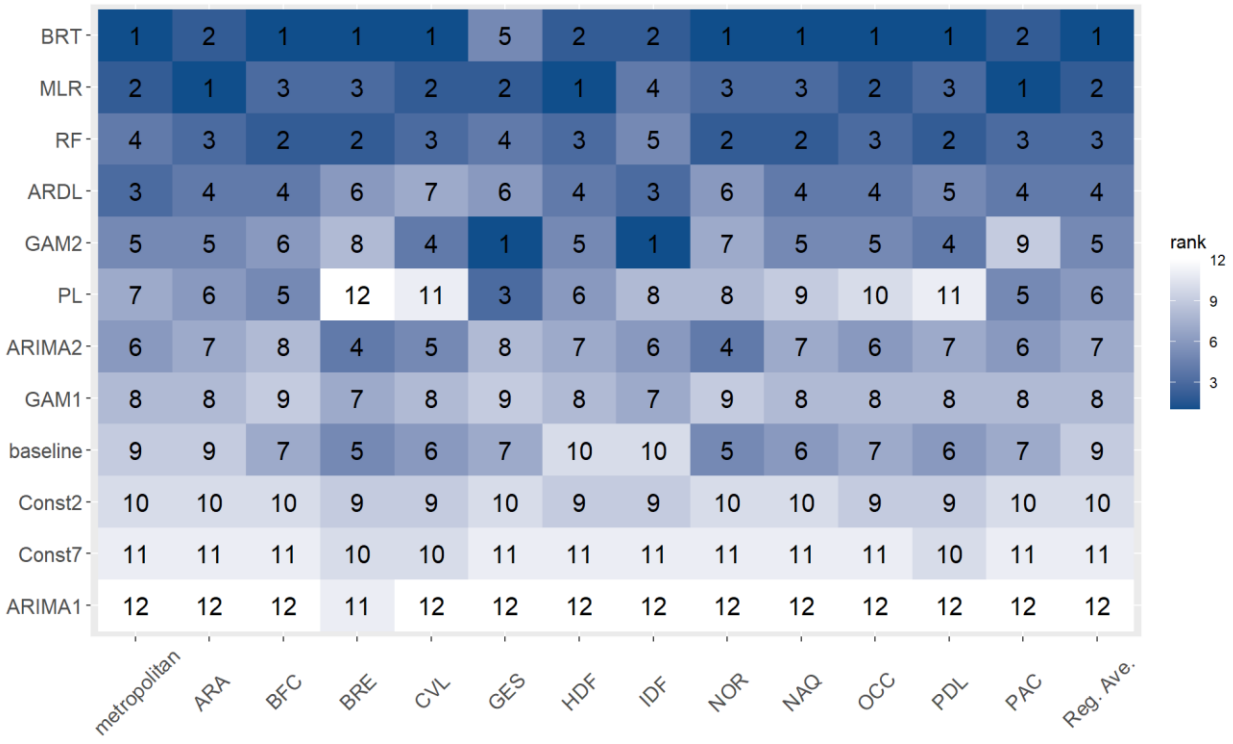

Fig. S6: Model ranking by region over the training period. Models are ranked according to the RMSE over all prediction horizons. Reg. Ave. = regional average (RMSE computed over all regions except metropolitan France).

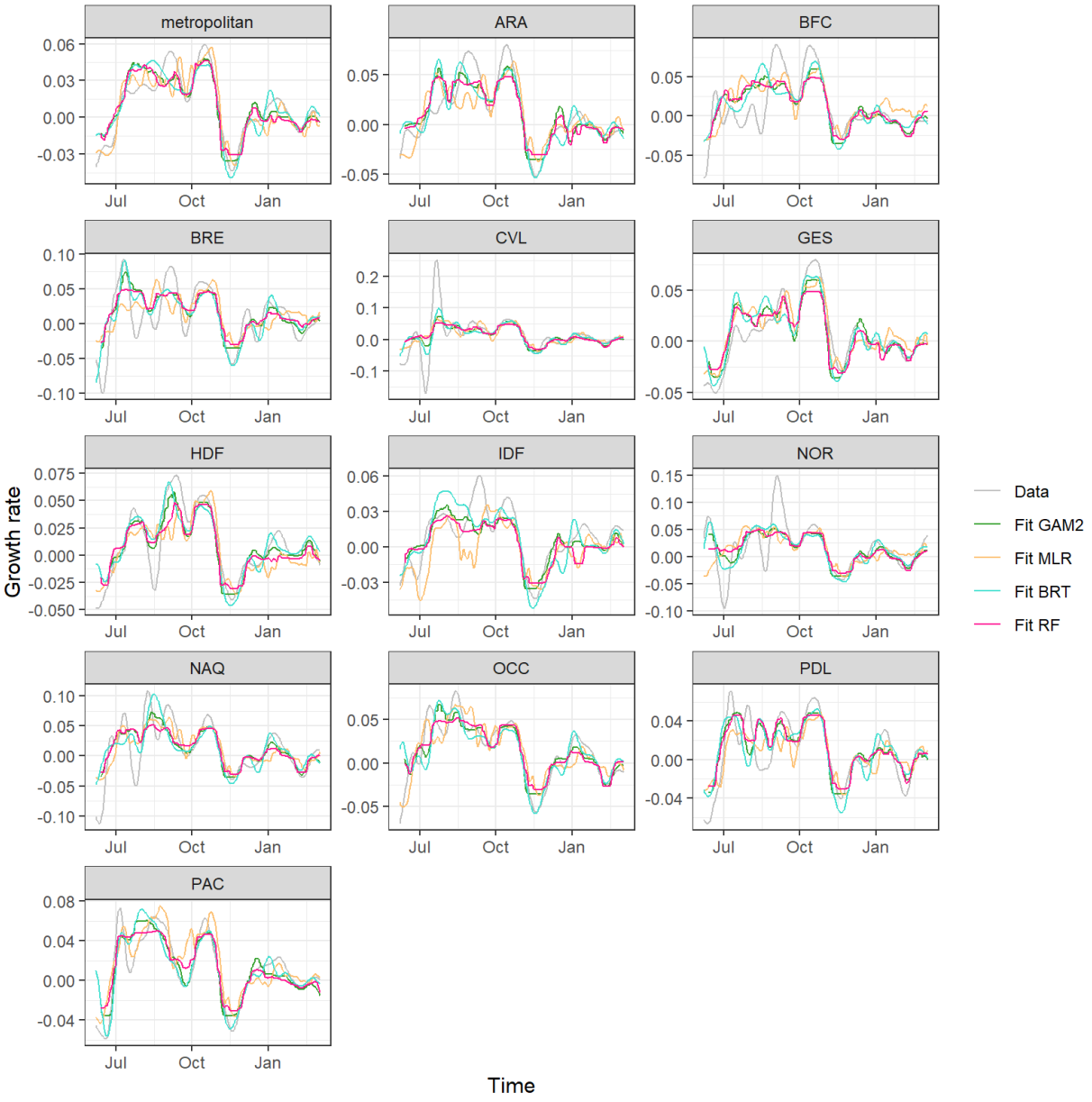

Fig. S7: Model fits for the growth rate of hospital admissions, at the national (metropolitan France) and regional levels, for the GAM2, the MLR, the BRT and the RF models. Each panel shows the observed growth rate (black line) and the predicted growth rates (colored lines) when retrospectively fitting 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).

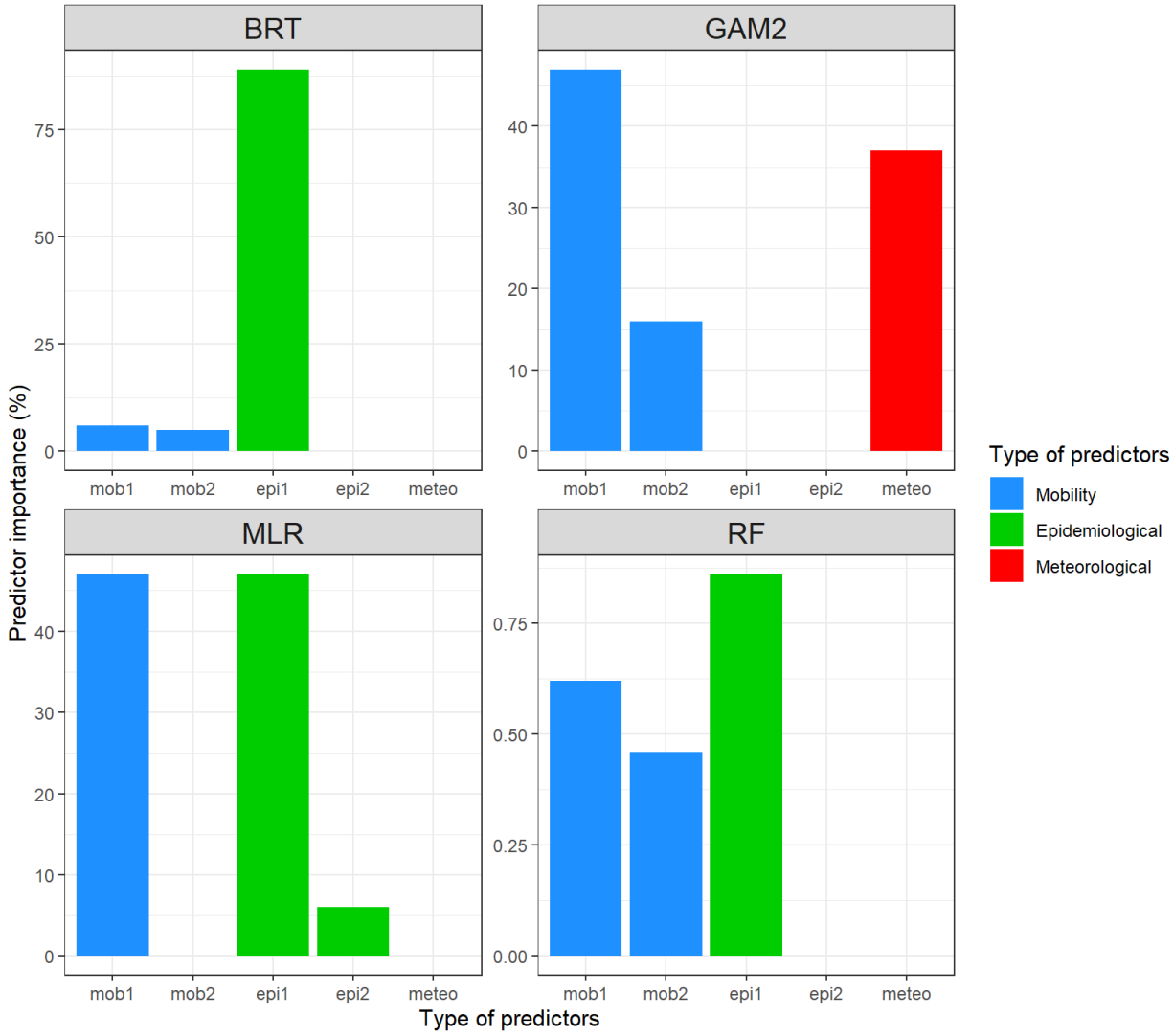

Fig. S8: Importance of predictors, estimated by retrospectively fitting the models from June 3rd 2020 to March 6th 2021. For the BRT model, relative importance is a measure of how each predictor contributes to reducing the loss function (all contributions sum to 100%). For the MLR and the GAM2 models, 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%).

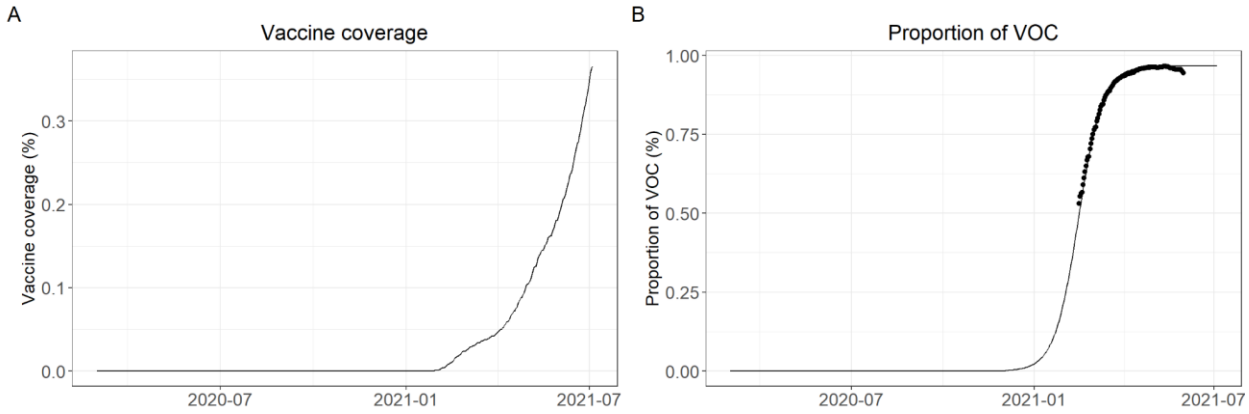

Fig. S9: Additional predictors used for the test period. (A) Vaccine coverage (complete vaccination scheme). (B) Proportion of variants of concern (VOC). The points represent the available data and the line represents the fit of the logistic model (see Supplementary text).

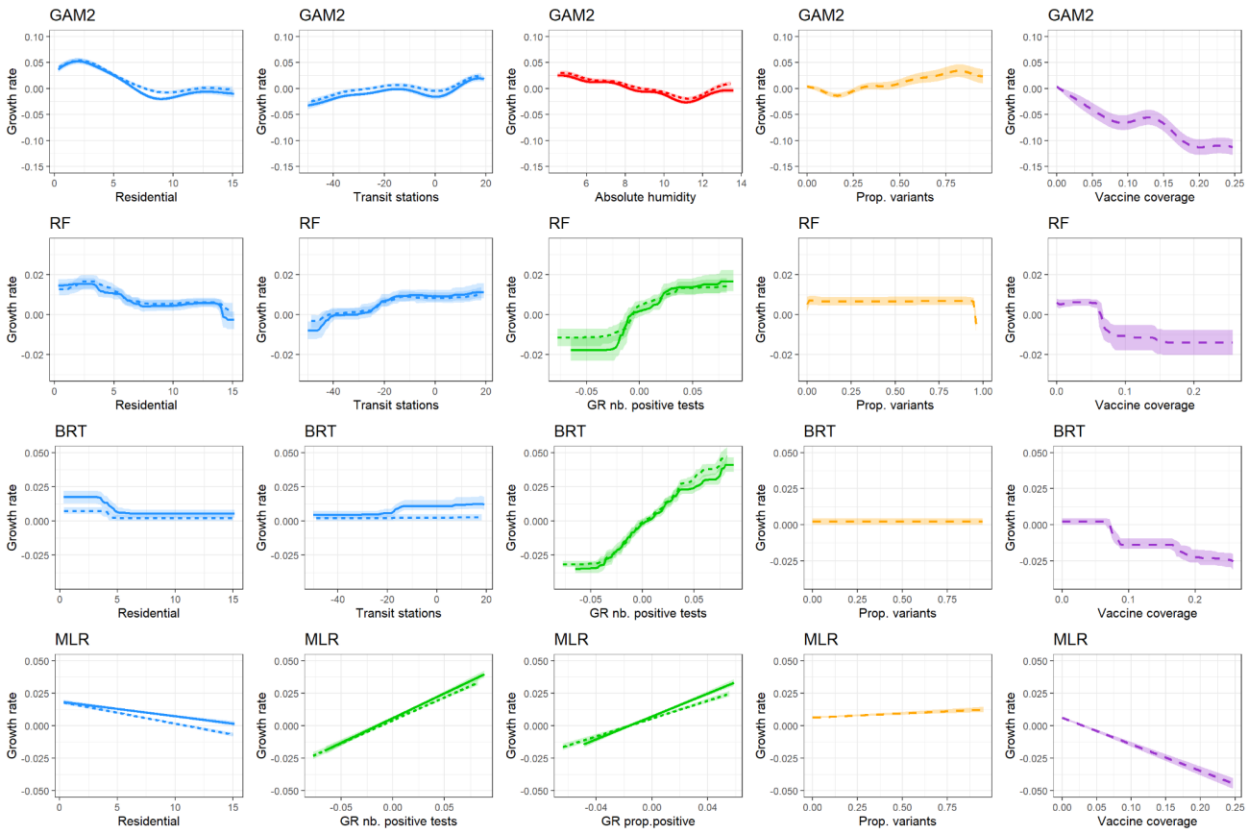

Fig. S10: Effects of mobility (blue), epidemiological (green), meteorological (red), proportion of VOC (orange) and vaccine coverage (purple) predictors on the growth rate of hospital admissions, for the GAM2, the RF, the BRT and the MLR models, by retrospectively fitting the models over two time periods: 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.

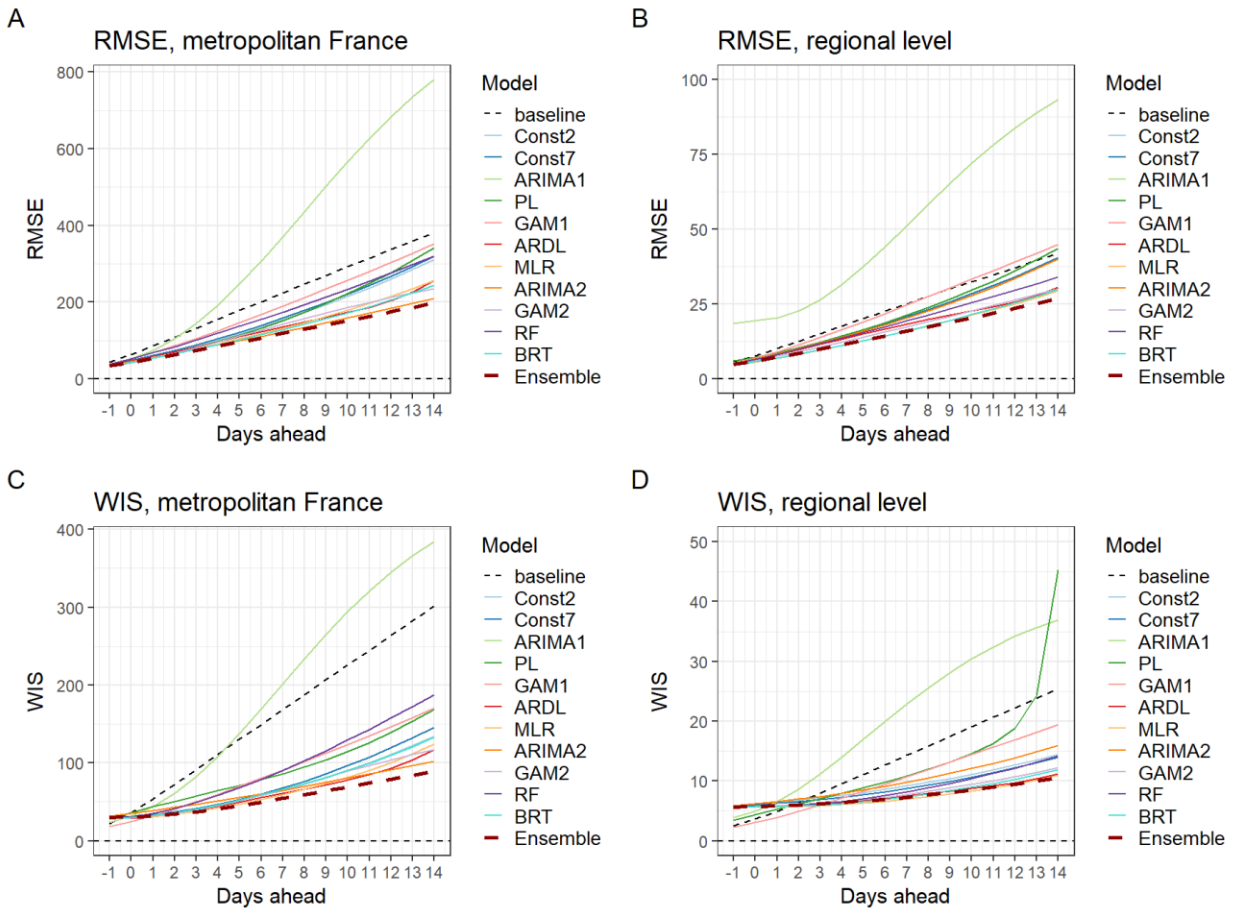

Fig. S11: Comparison of the performance of the individual and ensemble models over the test period at the national and regional levels, by prediction horizon, for hospital admissions. A. Root mean squared error (RMSE) in metropolitan France. B. RMSE by region. C. Mean weighted interval score (WIS) in metropolitan France. D. WIS by region.

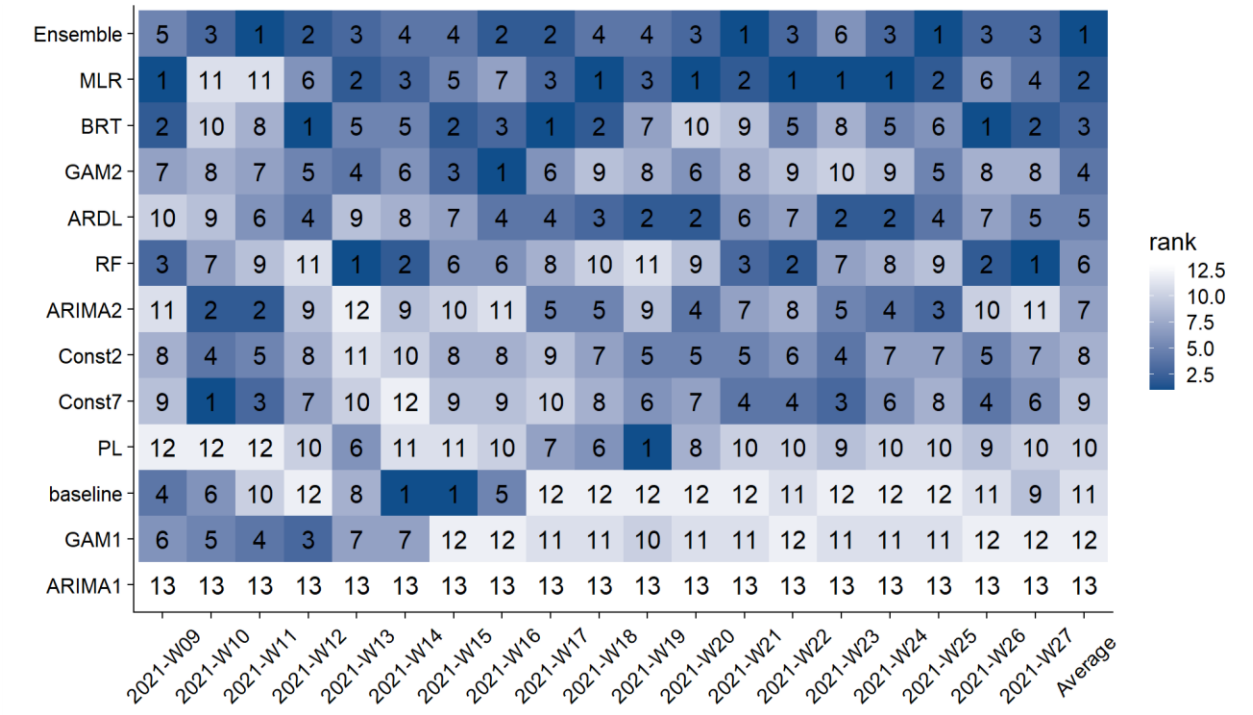

Fig. S12: Model ranking by week. Models are ranked according to the RMSE over all prediction horizons and all regions.

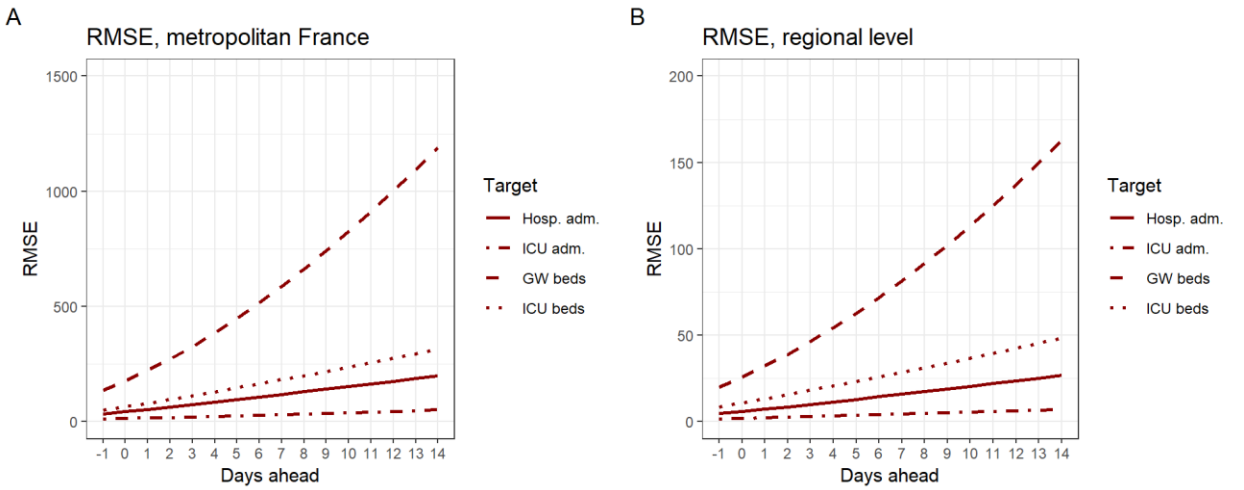

Fig. S13: Root mean squared error (RMSE) of the ensemble model for the four targets (hospital admissions, ICU admissions and bed occupancy in general ward (GW) and ICU) over the test period. A. RMSE of the ensemble model in metropolitan France. B. RMSE of the ensemble model at the regional level.

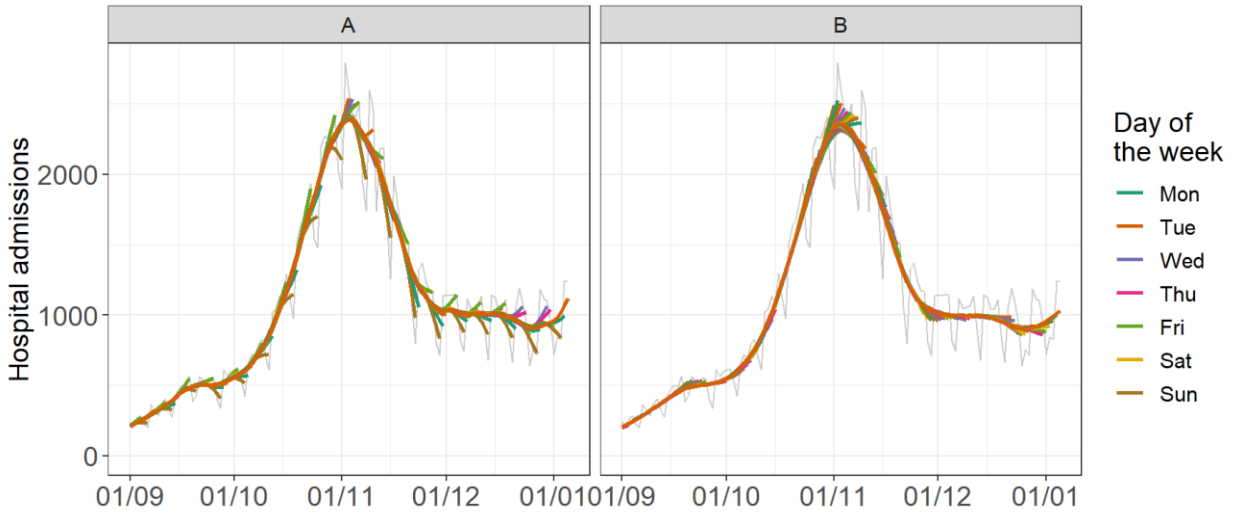

Fig. S14: Comparison of smoothing methods. (A) Smoothing spline. (B) Our two-step algorithm. The grey line shows the raw data of hospital admissions in metropolitan France from September 2020 to January 2021.The colored lines are time-series smoothed in real time (i.e. knowing only the past values), with different colors indicating the day of the last data point.

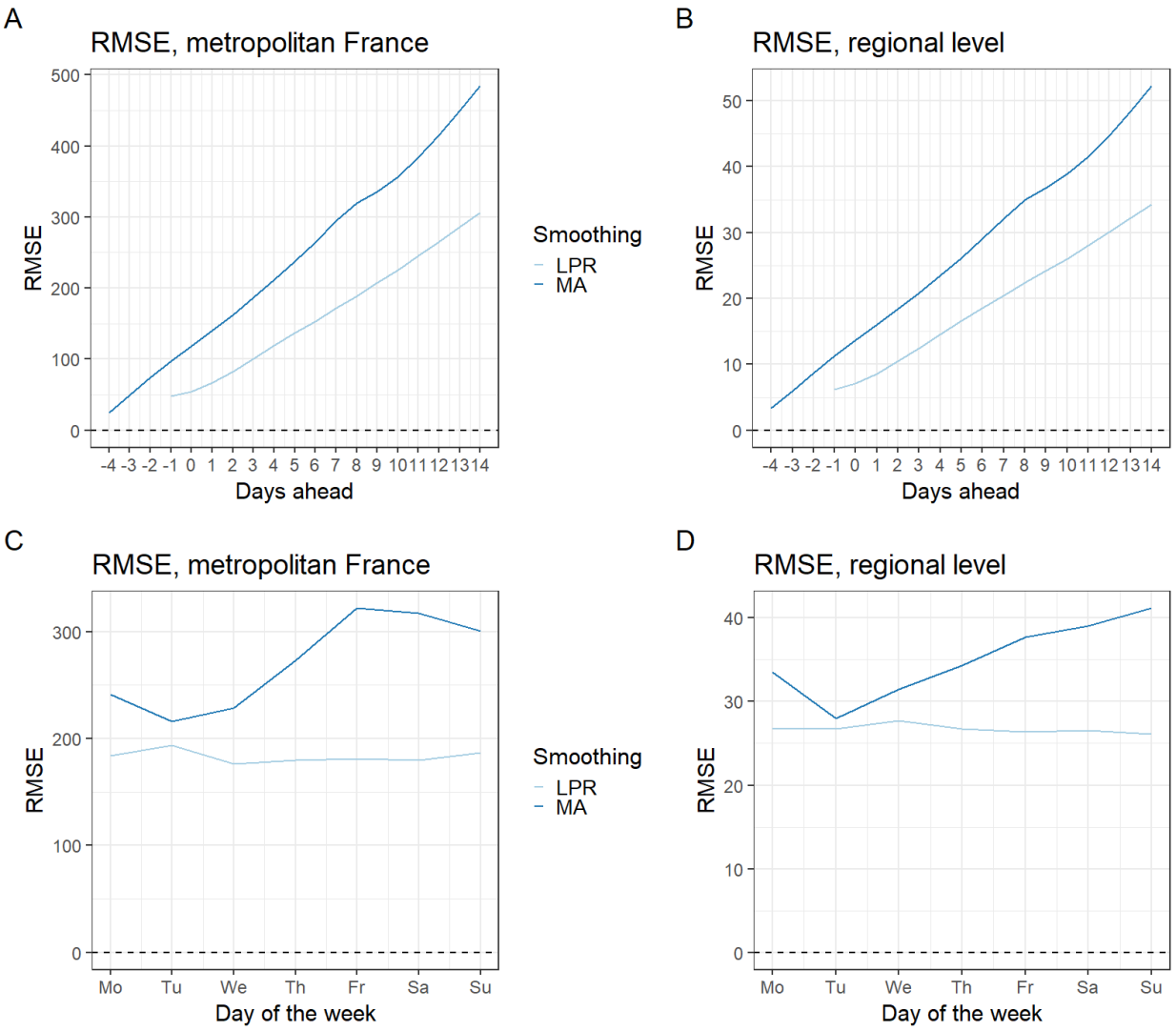

Fig. S15: Comparison of RMSE for hospital admissions over the training period, at the national and regional level, for the multiple linear regression (MLR) model, according to the smoothing method (LPR = local polynomial regression, MA = 7-day moving average). A. RMSE by prediction horizon in metropolitan 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. Here, we computed the RMSE of the LPR method against the LPR smoothed time series, and we computed the RMSE of the MA method against the MA smoothed time series.

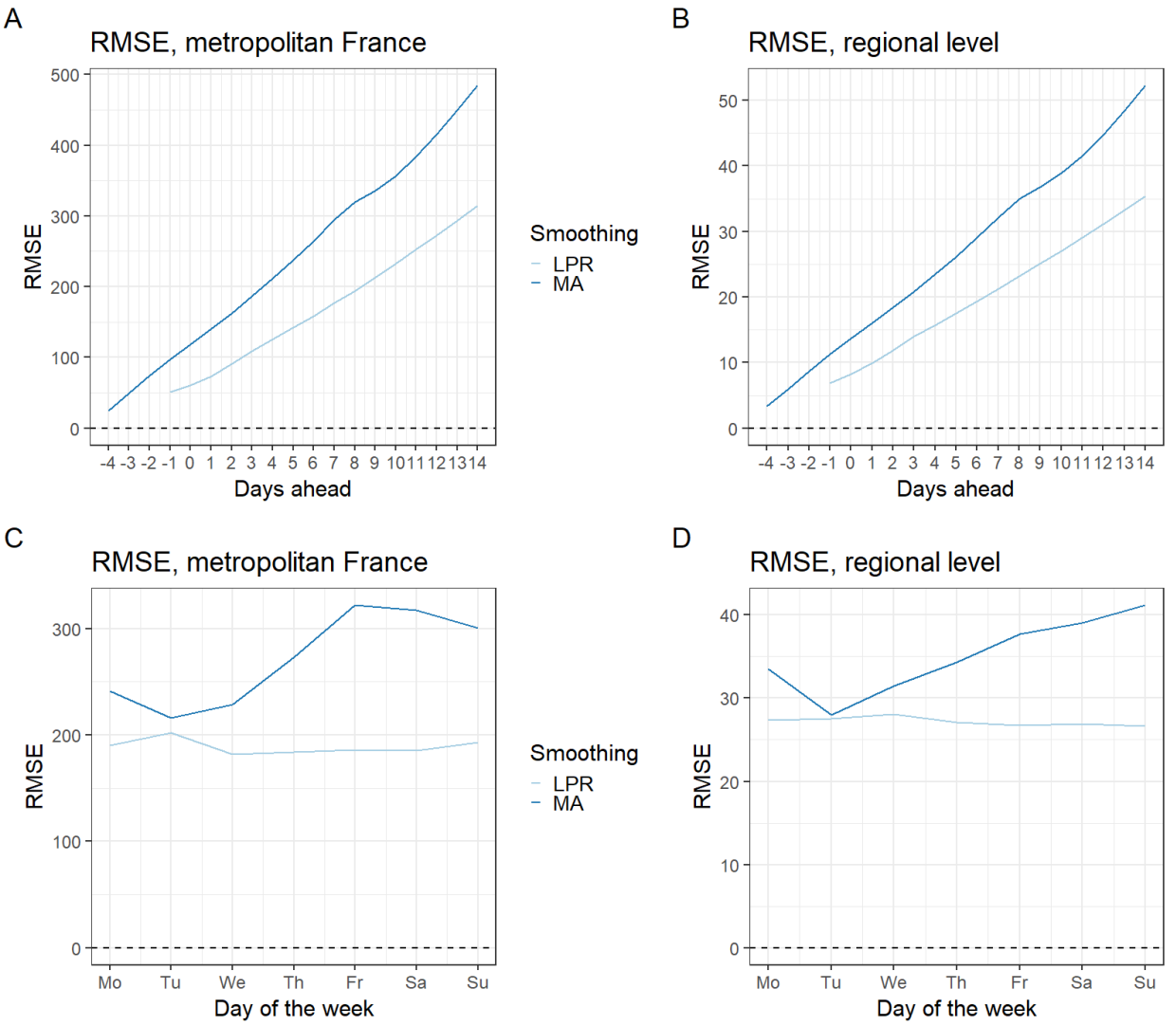

Fig. S16: Comparison of RMSE for hospital admissions over the training period, at the national and regional level, for the multiple linear regression (MLR) model, according to the smoothing method (LPR = local polynomial regression, MA = 7-day moving average), using MA smoothed series as "truth". A. 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. This figure differs from Figure S15 as RMSE of both methods was computed against the same times series (the MA smoothed series), showing that the conclusions of the comparison do not depend on the time series used as "truth".

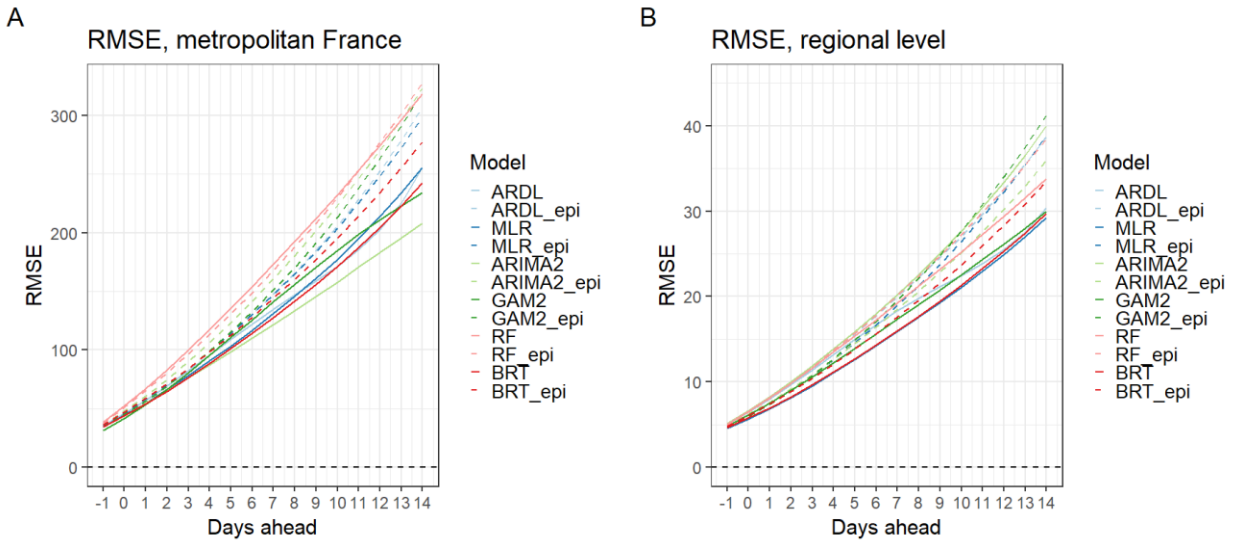

Fig. S17: Comparison of models with all types of predictors (solid lines) *vs* models with epidemiological predictors only (denoted "_epi", dashed lines), over the test period. In general, the models with all types of predictors perform better than the models with epidemiological predictors only.

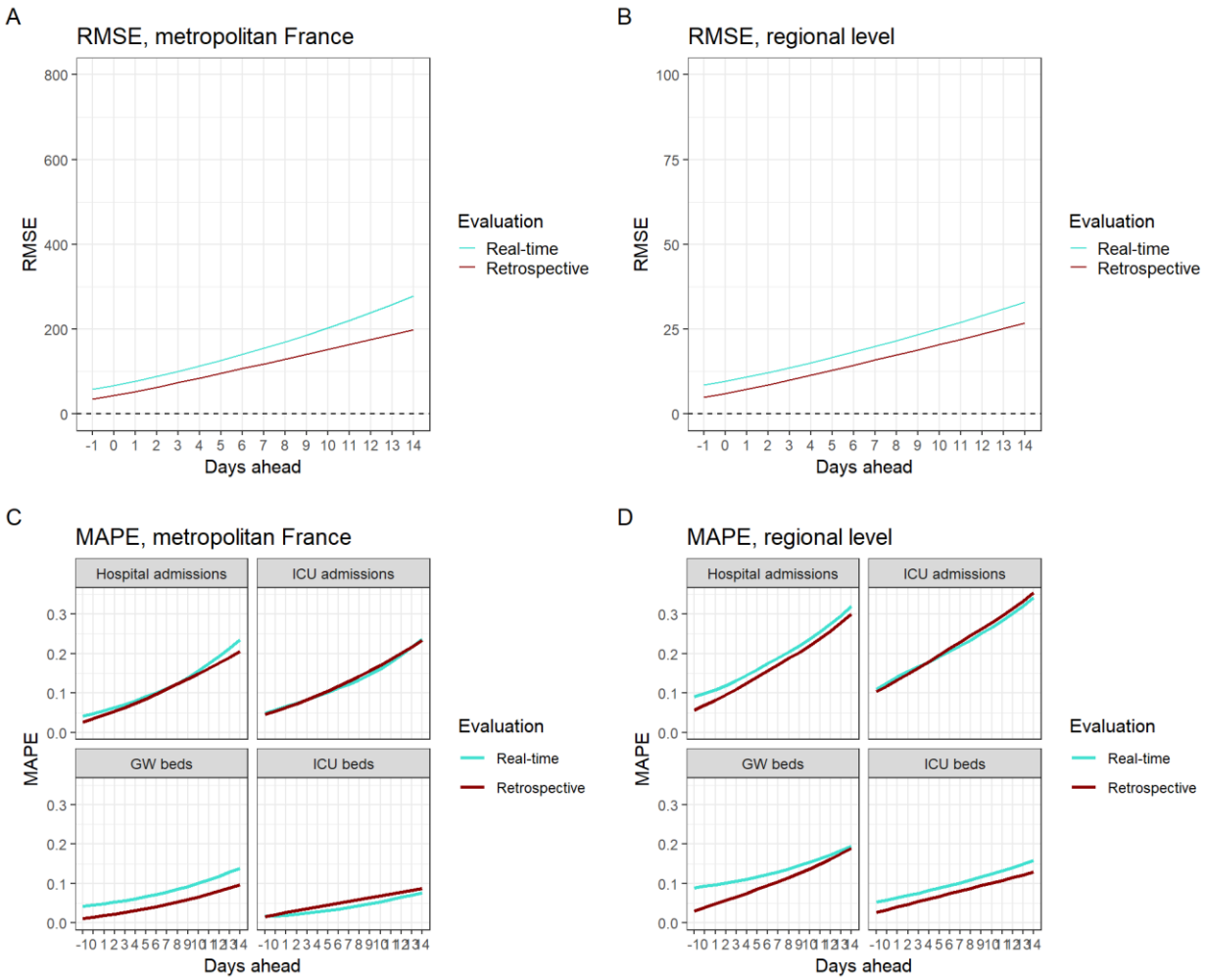

Fig. S18: Comparison of retrospective forecasts with real-time forecasts over the test period (March 7th 2021 to July 6th 2021). (A) RMSE for the ensemble forecasts of hospital admissions in metropolitan France. (B) RMSE for the ensemble forecasts of hospital admissions at the regional level. (C) MAPE for the ensemble forecast of the four targets in metropolitan France. (D) MAPE for the ensemble forecasts of the four targets at the regional level. Real-time forecasts were produced approximately every week to support public health decision-making. The models used to produce the real-time forecasts relied on an ensemble approach similar to the one presented in the retrospective study but they differed slightly (e.g. different set of predictors) and evolved over time.

Supplementary tables

Table S1: Best predictors selected for each individual model using the forward stepwise selection procedure over the training 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).