Quantification of the effects of climatic conditions on French hospital admissions and deaths induced by SARS-CoV-2

Hippolyte d'Albis, Dramane Coulibaly, Alix Roumagnac, Eurico de Carvalho Filho, Raphaël Bertrand

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2021 – 63

Quantification of the effects of climatic conditions on French hospital admissions and deaths induced by SARS-CoV-2

Hippolyte d'Albis
Dramane Coulibaly
Alix Roumagnac
Eurico De Carvalho Filho
Raphaël Bertrand

JEL Codes:
Keywords:

Funded by a French government subsidy managed by the ANR under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001

Quantification of the effects of climatic conditions on French hospital admissions and deaths induced by SARS-CoV-2

Hippolyte d'ALBIS[1,*] - Dramane COULIBALY[2] - Alix ROUMAGNAC[3]
Eurico DE CARVALHO FILHO[3] - Raphaël BERTRAND[3]

Abstract

An estimation of the impact of climatic conditions –measured with an index that combines temperature and humidity, the IPTCC – on the hospitalizations and deaths attributed to SARS-CoV-2 is proposed. The present paper uses weekly data from 54 French administrative regions between March 23, 2020 and January 10, 2021. Firstly, a Granger causal analysis is developed and reveals that past values of the IPTCC contain information that allow for a better prediction of hospitalizations or deaths than that obtained without the IPTCC. Finally, a vector autoregressive model is estimated to evaluate the dynamic response of hospitalizations and deaths after an increase in the IPTCC. It is estimated that a 10-point increase in the IPTCC causes hospitalizations to rise by 2.9% (90% CI 0.7-5.0) one week after the increase, and by 4.1% (90% CI 2.1-6.4) and 4.4% (90% CI 2.5-6.3) in the two following weeks. Over ten weeks, the cumulative effect is estimated to reach 20.1%. Two weeks after the increase in the IPTCC, deaths are estimated to rise by 3.7% (90% CI 1.6-5.8). The cumulative effect from the second to the tenth weeks reaches 15.8%. The results are robust to the inclusion of air pollution indicators.

[1] Paris School of Economics, CNRS, 48 boulevard Jourdan, 75014 Paris, France.
* Corresponding author. Email: hdalbis@psemail.eu Phone: +33 1 80 52 16 09
[2] Univ Lyon, Université Lumière Lyon 2, GATE, 93, chemin des Mouilles - B.P.167 69131 - Ecully cedex, France.
[3] PREDICT Services, 20 rue Didier Daurat, 34170 Castelnau-le-Lez, France.

Introduction

SARS-CoV-2 appeared in China in 2019 and has produced a global pandemic –the COVID-19 pandemics– as of March 2020 [1]. To cope with the disease, unprecedented mobility reduction measures such as lockdowns or curfews were implemented in many countries around the world, causing a major financial impact. These measures were justified by the fact that population mobility is a key factor for the virus circulation [2], together with population density, associated with a higher likelihood of infectious contacts between people [3].

Among other causes, several studies have shown the link between the spread of respiratory viruses and climatic conditions [4] [5]. This is the case of the influenza virus, for which the role of absolute humidity on transmission and seasonality has been demonstrated [6]. There are several types of transmission for respiratory viruses, but airborne transmission by aerosols (small particles) is more likely to be impacted by meteorological conditions. However, the mechanisms involved in these processes are still poorly understood, which calls for further investigations [7].

Concerning the SARS-CoV-2 outbreak, a pioneering study [8] conducted at the beginning of the pandemic warned that 90% of infections occurred in areas with temperatures between 3 and 17°C, with an absolute humidity between 4 and 9 g/m^3. Other studies from various countries have linked temperature or humidity to COVID-19. Most notably, it has been shown that for Northern hemisphere countries with a mean regional temperature below 10°C, a variation of 1 g/m^3 of the average absolute humidity can be associated with a variation of 0.15-unit in the basic reproduction number –the R0– and an increase of 1°C can be associated with a 0.16-unit lower R0 [9].

In the present paper, a statistical investigation is proposed using a recently developed indicator created by PREDICT Services, known as the IPTCC, an acronym in French for

PREDICT's Index for COVID-climate transmissivity [10]. This index was notably used by the Pasteur Institute to refine their models and to integrate an environmental factor that aims at better explaining the evolution of epidemiological indexes. Furthermore, the integration of the IPTCC into statistical models (such as the Multiple Linear Regression Model and the Generalized Additive Model) corrected the error between the forecasts and the observations of hospital admissions by 22% and 13% respectively [11].

The virus circulation was spatially and temporally irregular across mainland France in 2020. A first wave affected the country between March and April 2020, particularly the Paris region and the Northeast half of the country. The number of contaminations dropped sharply in late spring and remained considerably low in summer. Then, a second wave homogeneously impacted the country between October and November 2020. Since December 2020, the situation has reached a plateau with a high number of daily cases.

From the meteorological perspective, France is exposed to several climate dynamics. The western shoreline has an oceanic climate characterized by mild temperatures and significant rainfall throughout the year. The northeast of the country has a semi-continental climate with hot summers and cold winters. The Mediterranean climate is characterized by mild winters, warm summers, and less rainfall, which is irregular.

We take advantage of this geographical diversity to conduct a causal analysis of climatic conditions on the hospital admissions and deaths induced by SARS-CoV-2. Our approach is based on the estimation of vector autoregressive (VAR) models, following an established practice in economics since [12], where they are used to quantify an economy's response to an exogenous structural shock; they are also used in the life sciences (see e.g. [13] and [14]), and were found to be useful in particular to analyze the dynamics of SARS-CoV-2 ([15] and

[16]). VAR models can be easily used to quantify the effects of climatic conditions on epidemiological variables, as the former are clearly exogenous to the latter.

Our analysis is conducted at the regional level to exploit the heterogeneity (both in terms of climatic conditions and prevalence of the virus) across regions and to obtain an estimation that is not biased by sanitary measures. Over the period considered, mobility restrictions implemented by the government were indeed generally the same in all regions of mainland France. Specifically, a strict national lockdown was imposed from March 17, 2020 to May 11, 2020 and a mild one (with schools that remained open) from October 30, 2020 to December 15, 2020.

Methods

Data

We considered 54 NUTS3 administrative regions (named *départements* in French) over 42 weeks (from March 23, 2020 to January 10, 2021). We did not consider all existing regions, but only the 54 for which the IPTCC can be computed. In order to evaluate the consequences of the climatic conditions on the pandemic, we used the weekly numbers of hospital admissions and deaths due to SARS-CoV-2 for the same 54 administrative regions considered. The data are official and publicly available at www.data.gouv.fr. The hospitalizations and deaths due to SARS-CoV-2 were consistently measured over the period and are thus more reliable than the series that report the number of infections, as the latter highly depend on the availability of tests and on the population's willingness to get tested. Daily data are available since March 19, 2020, but we chose to use the weekly frequency in order to avoid a seasonal effect induced by lower reporting during weekends.

The meteorological data came from the 63 Météo-France stations. These stations are homogeneously distributed over mainland France and they were chosen in order to better represent the diversity of the country's climate. The dataset provides a daily average for air temperature (measured in °C) and relative humidity (measured in %); these values were calculated as an average of the daily maximum and minimum values of temperature and relative humidity. With them, it is easy to compute the daily absolute humidity for each station, measured in g/m^3, using the Clausius-Clapeyron equation [8].

In order to analyze the potential relation between climate conditions and virus transmission, the IPTCC was created to characterize the potential for virus transmission according to climatic conditions. Following [10], the IPTCC is a function of absolute humidity (AH), relative humidity (RH), and temperature (T), and the formula can be written as follows:

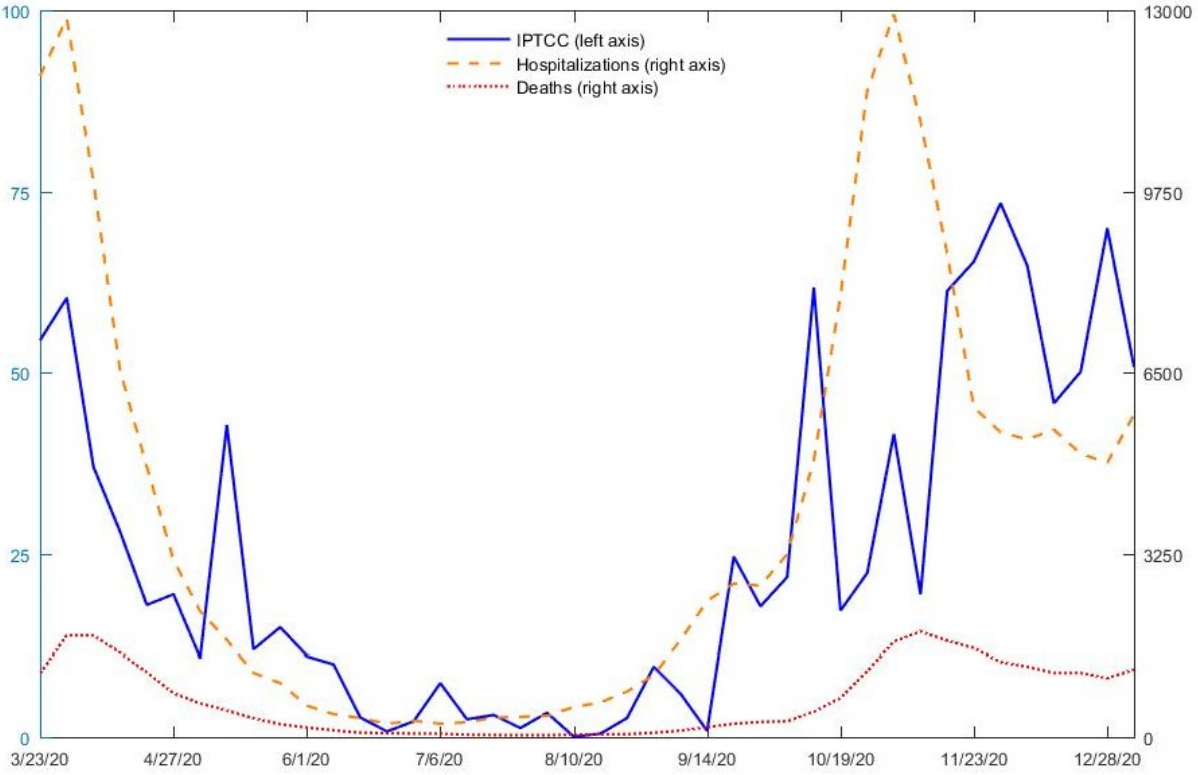
$$IPTCC = 100 * e^{-\frac{1}{2}\left[\frac{(T-7.5)^2}{196}+\frac{(RH-75)^2}{625}+\frac{(AH-6)^2}{2.89}\right]}.$$

The IPTCC is thus maximal when the temperature reaches 7.5°C and the relative humidity 75%. The IPTCC is available on a daily basis since January 1, 2020 for all meteorological stations. Then, for each of the 54 regions considered, we built a weekly indicator computed as the average of the daily IPTCC. Visualizations of the IPTCC through 2D and 3D representations are provided in the Supplementary Materials (Figures A1 and A2).

As a preliminary step, the evolution of the variables over the period considered can be plotted. The figures for each of the 54 regions are reported in the Supplementary Materials (Figure A3). To summarize them, Figure 1 reports weekly hospitalizations and deaths (i.e. the total for the 54 regions) and the weekly IPTCC (the average for the 54 regions). The correlation between the three series is quite remarkable (see Table A2 for the coefficients). Most notably, the peaks in hospitalizations and deaths correspond to periods during which the IPTCC is high on average. Conversely, throughout spring and summer, hospitalizations, deaths, and the

IPTCC are low. We also notice that the IPTCC is more volatile than the epidemiological series. Lastly, the evolution of the IPTCC seems to precede those of hospitalizations and deaths, which is probably due to the time between contamination, the incubation period, and a possible worsening of the disease.

Figure 1. Comparison of the time series of total hospitalizations and deaths and the average IPTCC, March 23, 2020 to January 10, 2021



To go beyond this first graphical analysis and investigate how climatic conditions affect the spread and severity of SARS-CoV-2, we developed a time series analysis. This was done in two steps. Firstly, we demonstrated a causal relationship between the IPTCC and the hospitalization and deaths, and then we evaluated the responses of the latter to an increase in the IPTCC.

Granger causality

We firstly explored whether climate conditions Granger-cause hospitalizations and deaths from SARS-CoV-2, i.e. whether the past values of climatic conditions contain information that are helpful to predict hospitalizations or deaths given that past hospitalization and death information are considered.

To test for Granger causality from climate conditions to hospitalizations and deaths, we considered the following panel dynamic model using the weekly data for 54 regions, from March 23, 2020 to January 10, 2021.

$$y_{it} = \sum_{s=1}^{p} \alpha_s \, y_{it-s} + \sum_{s=1}^{p} \beta_s \, cl_{it-s} + \mu_i + \delta_i . t + \eta_t + \varepsilon_{it}, \qquad i = 1, \dots N \; and \; t = 1, \dots, T \quad (1)$$

where i and t stand respectively for the indices of region and time; $y_{it} \in \{hosp_{it}, death_{it}\}$ with $hosp_{it}$ and $death_{it}$ respectively denote the logarithm of 1 plus the number of hospitalizations and deaths; cl_{it} represents the IPTCC index; μ_i are regional fixed effects, $\delta_i t$ are regional-specific time (linear) trends, and η_t is the common national time (week)-specific effect including seasonal effect.

It is worth noting that taking the logarithm of hospitalizations and deaths allows, through regional fixed effects u_i, to account for regional heterogeneities such as area, population, or density, which are roughly stable over the considered period. Moreover, including the region-specific time trend $\delta_i . t$ with the log specification may capture a potential exponential growth of the disease spreading that could be region specific. Finally, using η_t allows taking into account the national interdependence (including any seasonal effect) of the disease spreading across regions. Preliminary diagnostics (panel unit root tests) rejected the null hypothesis of unit root for the de-trended variables (with a region-specific linear trend). We then considered variables in level or log level while controlling for region heterogeneity (by introducing region-

specific effects and region-specific time trends) and cross-region interdependence (by introducing week-specific effects).

Let $\theta = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p)'$ be the vector of parameters across cross-section units to be estimated. Given the sizes of the cross-region dimension N and the time dimension T in the panel data ($N = 54$ and $T = 42$), in order to deal with short-T dynamic panel data bias or the so-called Nickell bias [17], we used the bias-corrected fixed-effect estimator developed by [18], which is appropriate when $0 < \lim N/T < \infty$, as is the case here (see [19]). This technique consists in removing the asymptotic bias of least square dummy variable (LSDV) or fixed effect estimator (with region-specific time trends and time-specific effects) of θ. The LSDV estimator of θ which is given by the ordinary least square (OLS) regression of $\tilde{y}_{it}$on $\tilde{y}_{it-1}, \dots, \tilde{y}_{it-p}, \widetilde{cl}_{it-1}, \dots, \widetilde{cl}_{it-p}$ where $\tilde{x}_{it}$ is a transformation of x_{it} after removing region specific effects and trends and the national average for each week (this transformation corresponds to include μ_i, $\delta_i . t$, and η_t).

Based on the Bayesian information criterion (BIC) and the Hannan-Quinn information criterion (HQC), we set the lag length p to 3, so that there is no serial correlation in the errors. Taking a lag length higher than 3 does not alter our findings.

In Equation (1), the null hypothesis of no Granger causality from the climate conditions (cl_{it}) to hospitalizations or deaths ($y_{it} = hosp_{it}$ or $death_{it}$) is H_0: $\beta_1 = \beta_2 = \cdots = \beta_p$. The null hypothesis of no Granger causality can be expressed as $R\theta = 0_{p\times 1}$ where R is a known $(p \times 2p)$ matrix with $R = [0: I_p]$. The test statistics, which is a Wald statistics, is given by $W = \hat{\theta}' R' [\hat{\sigma}^2 R(X'X)^{-1} R']^{-1} R\hat{\theta}$ where X is a $(N(T-p) \times 2p)$ matrix of regressors $\tilde{y}_{it-1}, \dots, \tilde{y}_{it-p}, \widetilde{cl}_{it-1}, \dots, \widetilde{cl}_{it-p}$ in column and $\hat{\sigma}^2$ is the estimated variance of residual. Under the null hypothesis W follows a chi-squared distribution of a degree of freedom equal to p, i.e. the number of constraints to be tested that corresponds to the lag length.

The VAR model

To analyze the dynamic responses of the epidemiological variables to a change in the IPTCC, we estimated a panel vector autoregressive (VAR) model where IPTCC (cl) is considered exogenous. It can be written as:

$$Z_{it} = \sum_{s=1}^{p} A_s Z_{it-s} + \sum_{s=0}^{p} b_s cl_{it-s} + u_i + d_i.t + f_t + v_{it}, \quad i = 1, \dots N \text{ and } t = 1, \dots, T \qquad (2)$$

where $Z_{it} = (hosp_{it}, death_{it})'$ is a 2-dimensional vector of endogenous variables including the logarithm of 1 plus the hospitalizations and deaths, where A_s are (2×2) matrices of coefficients associated with Z_{it}, b_s are (2×1) matrices associated with cl_{it-s}, $u_i = (u_i^1, u_i^2)'$ is a vector of regional fixed-effects; $d_i.t = (d_i^1, d_i^2)'.t$ represent region specific-time (linear) trend, $f_t = (f_t^1, f_t^2)$ is a vector of common time (week)-specific effect; $v_{it} = (v_{it}^1, v_{it}^2)'$ is a 2-dimensional vector of errors satisfying $E(v_{it}) = 0$ and $E(v_{it}v_{is}') = \Omega.\mathbb{1}\{t = s\}$ for all t and s. As mentioned above, preliminary diagnostics (panel unit root tests) rejected the null hypothesis of unit root for the de-trended variables (with a region-specific linear trend). Our VAR then considers variables in level or log level while controlling for region heterogeneity (by introducing region-specific effects and region-specific time trends) and cross-region interdependence (by introducing week-specific effects). Moreover, the model was estimated through the bias-corrected fixed-effect estimator developed by [18], and the lag length p was set to 3 based on the Bayesian information criterion (BIC) and the Hannan-Quinn information criterion (HQC). Using a lag length higher than 3 does not change our findings.

After having estimated model (2), we computed the responses of the endogenous variables (hospitalizations and deaths) to an (exogenous) increase in the IPTCC, and the response of

deaths to an increase in hospitalizations. The responses of hospitalizations and deaths (endogenous variables) to climate conditions (exogenous variable) is part of the multiplier analysis; for this reason, there is no need to identify the structural shocks of the endogenous variables (see [20], Section 10.6). However, to identify the response of deaths to hospitalization, it is necessary to identify the structural shocks η_{it} of these two endogenous variables as follows: $\eta_{it} = A_0 v_{it}$ where A_0 is 2×2 matrix such that $E(\eta_{it}\eta_{it}') = I_2$ or $A_0 A_0' = \Omega$. We identified A_0 based on Cholesky decomposition by setting A_0 as the unique lower-triangular Cholesky factor of Ω. This identification relies on the reasonable assumption that hospitalizations can influence deaths contemporaneously, while deaths can potentially influence hospitalizations only with lags.

Results

Causality between the IPTCC and the epidemiological variables

The results of Granger non-causality from climate conditions to hospitalizations and deaths induced by SARS-CoV-2 are reported in Table 1. At the 5% level of significance, we cannot accept the null hypothesis of no Granger causality from climate conditions to either hospitalizations or deaths. In other words, past information on climate conditions is helpful to predict hospitalizations and deaths even when accounting for past hospitalization and death information.

Table 1. Granger causality from the IPTCC to hospitalizations and deaths

Hypothesis	Test statistics	P-value
IPTCC does not Granger-cause hospitalizations	22.044	0.000
IPTCC does not Granger-cause deaths	10.815	0.013

Note: The test statistic is a Wald statistic which follows, under the null hypothesis, a chi-squared distribution of 3 (the number of constraints that corresponds to the lag length).

Although the Granger causality analysis conducted above indicates how the past values of the IPTCC are useful to predict hospitalizations and deaths, it does not provide any evaluation of the response of hospitalizations and deaths to a change in the IPTCC. This is done with the VAR model.

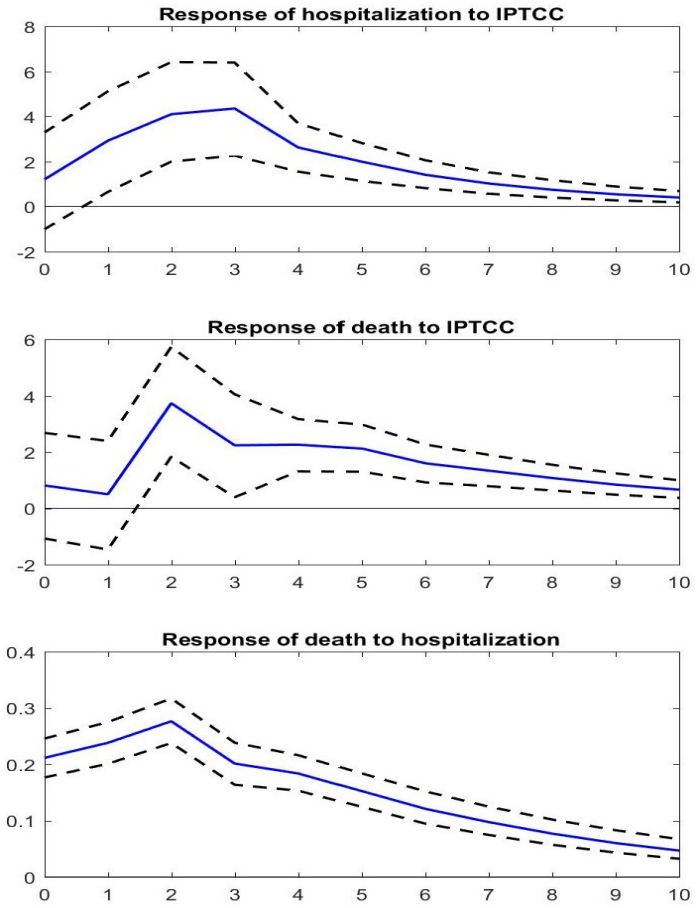
Estimations of the epidemiological responses to a change in the IPTCC

We consider a 10-point increase in the IPTCC, an increase which is understood as relative to the average IPTCC value over the period considered (namely 25.5). One should note that a 10-point increase correspond to rather small changes in temperature and humidity. For instance, the IPTCC moved from 42 to 52 in the Seine-Maritime *département* between April 6 and April 13 of 2020; this change was due to a decrease in temperature from 13°C to 11°C and an increase in relative humidity from 73% to 75%. The largest increase (76.9) was recorded in the Finistère *département* between May 4 and May 11 of 2020 when temperature fall from 16°C to 10°C and relative humidity declined from 83% to 68%. The largest decline (-66.9) was observed in the Hérault *département* between December 7 and December 14 of 2020 when temperature rose from 6°C to 12°C and the relative humidity from 76% to 91%.

The dynamic responses of hospitalizations and deaths to a 10-point increase in the IPTCC are displayed in Figure 2. The upper panel represents the estimated impact of the IPTCC on hospitalizations. Week 0 is the week when the increase occurs. We observe that the response of hospitalizations is delayed but quite persistent: it is significantly positive from one week after the increase until at least 10 weeks later. In terms of magnitude, we estimate that hospitalizations rise by 2.9% (90% CI 0.7-5.0) one week after the increase in the IPTCC, and by

4.1% (90% CI 2.1-6.4) and 4.4% (90% CI 2.5-6.3) in the two following weeks. Over ten weeks, the cumulative effect is estimated to reach 20.1%. Compared to their average value over the period, i.e. 147 individuals per week and region, this represents 30 individuals. Note that to evaluate e.g. the effect of a 50-point increase that would put the IPTCC over 75, our estimates should simply be multiplied by 5. The cumulative effect on hospitalizations is thus expected to reach 100%.

Figure 2. Dynamic responses of hospitalizations and deaths to a shock on the IPTCC



Note: The solid line gives the estimated impulse responses. The dashed lines give the 90% confidence intervals generated by Monte Carlo with 5,000 repetitions. The size of the increase in the IPTCC is set to a 10-point increase. The size of the increase in hospitalizations is set to a one-percent increase. The responses are the percentage change in the number of hospitalizations and deaths.

The middle panel of Figure 2 represents the estimated impact of the same 10-point increase in the IPTCC on deaths. The impact is also positive but is further delayed, as it becomes positive as of the second week after the increase in the IPTCC. That week, we estimate that the deaths increase by 3.7% (90% CI 1.6-5.8). The cumulative effect from the second to the tenth weeks reaches 15.8%. Compared to their average value over the period, i.e. 25 individuals per week and region, this represents 4 individuals. The link between hospitalizations and deaths is further investigated in the bottom panel of Figure 2, which represents the response of the number of deaths to a 1% increase in the number of hospitalizations. We observe an immediate response, of magnitude 0.21%, which grows up to 0.28% the second week, and then decreases. These last evaluations are roughly consistent with the existing evaluations ([21], among others); they are not immediately comparable since we take into account the reaction of hospitalizations to their own change. The existence of persistence in the series explains why our evaluation is slightly higher.

Sensitivity and robustness analysis

The sensitivity and robustness of our findings were assessed by analyzing the dynamics under alternative proxies for climate conditions and by considering the environmental pollution as a possible interfering variable.

Firstly, it seems that alternative indicators for climatic conditions are less effective than the IPTCC to capture the effect of climate on the epidemiological dynamics. We exemplified this with two alternative indicators. Initially, we used a 'false' IPTCC index computed using aberrant target values for temperature and humidity. The 'false' IPTCC reaches 100 when the temperature reaches 30°C and the relative humidity 20%. We obtained that hospitalizations no longer respond significantly to the indicator (see Figure A4 in the Supplementary Materials). Then, we replaced the IPTCC by a normalized temperature index computed as $T_n =$

$100.e^{-(T-7.5)}$, and therefore abstracted from the humidity variables. We obtained that, after three weeks, the response of hospitalizations becomes significant, but the effects are less clear and the magnitude is less important (see Figure A5 in the Supplementary Materials). This confirms that, although partly correlated, temperature and humidity are both useful when combined in a single indicator to evaluate the effects of climate on epidemiological variables. Secondly, it is important to investigate whether our results could be biased by some omitted-variables. In particular, environmental pollution, whose effect on respiratory diseases is well established (see e.g. [22]), has been recently shown to be correlated with the prevalence of COVID-19 ([23], [24], [25]). We have thus estimated extended versions of our models in order to take into account the evolution of the environmental pollution. More specifically, we have collected weekly averages of the concentration of atmospheric particulate matter (PM_{10} and $PM_{2.5}$), of nitrogen dioxide (NO_2) and of ozone (O_3) for a subsample of the regions we considered. We found that, at the 5% level of significance, we still cannot accept the null hypothesis of no Granger causality from climate conditions to either hospitalizations or deaths (Table A3 in the Supplementary Materials). Moreover, the dynamic responses of hospitalizations and deaths to a 10-point increase in the IPTCC are qualitatively similar to those displayed in Figure 2 (Figure A6 in the Supplementary Materials). Our results are thus robust to the inclusion of environmental pollution variables in the models. Interestingly, we also obtained that the null of hypothesis of no Granger causality from air pollution to either hospitalizations or deaths cannot be rejected but that, on the short run, the concentration of atmospheric particulate matter has an impact on hospitalizations and deaths.

Discussion

By testing the statistical impact of the IPTCC on hospital admissions and deaths induced by SARS-CoV-2 in France, this study confirms the potential role of temperature and humidity in

the airborne spread of SARS-CoV-2 by highlighting that there is a particular combination of the two variables that creates a favorable ground for the transmission of the disease: when the distance toward a temperature of 7.5°C and a relative humidity of 75% reduces, hospital admissions and deaths increase. With respect to earlier studies, it provides strong causal evidences and offers a new quantification of the effect of the climate factors on the dynamics of the disease. Most notably, it characterizes the time delay between the climatic conditions and the hospitalizations and deaths and include air pollution indicators as controls.

This study has several limitations, notably it was performed for continental France, and needs to be replicated in other countries in order to ensure that the benchmark combination for which the IPTCC is maximal (temperature at 7.5°C and relative humidity at 75%) is same. Furthermore, although the model is using regional data, the results are not region specific. They should be interpreted as the response of an average region in France. More data are needed in order to be able to characterize the regional differences. Moreover, the study covers a period that starts in March, 2020 and ends in January, 2021. It thus does not take into account the more recent virus variants nor the vaccination campaign. More generally, it is clear that anthropogenic factors related to the behavior and respect of sanitary recommendations among the citizens, the conditions of hygiene, the access to care, and the quality and resources of the health services have an essential impact on COVID-19 transmission and other indicators. Therefore, the IPTCC should not be taken out of context, and it is not the sole condition or explanation for the crisis the world has been facing. The weather is not a silver bullet.

Nonetheless, incorporating meteorological components into the overall analysis could allow a better surveillance and prediction of the dynamic of the pandemic, as exemplified here by the evolution of the number of hospitalizations. As there is a delay between the IPTCC time

series and hospitalizations, our index could help to forecast the pressure on the health system 2 to 3 weeks beforehand. This can be implemented and continuously updated very easily using the temperature and humidity forecasts provided by Météo-France stations. The evolution of the local or regional weather is a factor that could help the government to predict future epidemic waves and make decisions and communicate to protect population against SARS-CoV-2, or other respiratory diseases spread by airborne transmission.

References

[1] WHO, "WHO Director-General/Speeches," 2020. [Online]. Available: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. [Accessed 02 March 2021].

[2] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, O. C.-1. D. W. Group, L. d. Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus and S. V. Scarpin, "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science,* no. 368, pp. 493-497, 2020.

[3] G. Delnevo, S. Mirri and M. Roccetti, "Particulate Matter and COVID-19 Disease Diffusion in Emilia-Romagna (Italy). Already a Cold Case?," *Computation,* vol. 59, no. 2, 2020.

[4] A. Briz-Redon and A. Serrano-Aroca, "The effect of climate on the spread of the COVID-19 213 pandemic: A review of findings, and statistical and modelling techniques," *Progress in 214 Physical Geography: Earth and Environment,* vol. 44, pp. 591-604, 2020.

[5] D. Marazziti, P. Cianconi, F. Mucci, L. Foresi, I. Chiarantini and A. Della Vecchia, "Climate change, environment pollution, COVID-19 pandemic and mental health," *The Science of the Total Environment,* vol. Jun 15;773:145182, 2021.

[6] J. Shaman and M. Kohn, "Absolute humidity modulates influenza survival, transmission, and seasonality," *Proceedings of the National Academy of Sciences,* vol. 106, no. 9, pp. 3243-3248, 2009.

[7] J. Shaman, V. E. Pitzer, C. Viboud, B. T. Grenfell and M. Lipsitch, "Absolute Humidity and the Seasonal Onset of Influenza," *PLoS Biology,* vol. 8, no. 2, 23 02 2010.

[8] Q. Bukhari and Y. Jameel, "Will Coronavirus Pandemic Diminish by Summer?," *SSRN,* 27 03 2020.

[9] J. P. J Landier, S. Rebaudet, E. Legendre, L. Lehot, A. Fontanet, S. Cauchemez and J. Gaudart, "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,* 26 01 2021.

[10] A. Roumagnac, E. D. C. Filho, R. Bertrand, A.-K. Banchereau and G. Lahache, "Étude de l'influence potentielle de l'humidité et de la température dans la propagation de la pandémie COVID-19," *Médecine de Catastrophe - Urgences Collectives,* 13 01 2021.

[11] J. Paireau, A. Andronico, N. Hozé, M. Layan, P. Crepey, A. Roumagnac, M. Lavielle, P.-Y. Boëlle and S. Cauchemez, "An ensemble model based on early predictors to forecast COVID-19 healthcare demand in France," *Life Sciences [q-bio]/Santé publique et épidémiologie,* 2021.

[12] C. A. Sims, "Macroeconomics and reality.," *Econometrica,* vol. 48, pp. 1-48, 1980.

[13] A. Reeves, S. Basu, M. McKee, D. Stuckler, A. Sandgren and J. Semenza, "Social protection and tuberculosis control in 21 European countries, 1995–2012: a cross-national statistical modelling analysis," *The Lancet Infectious Diseases,* vol. 14, no. 11, pp. 1105-1112, 2014.

[14] A. Duggento, L. Passamonti, G. Valenza, R. Barbieri, M. Guerrisi and N. Toschi, "Multivariate Granger causality unveils directed parietal to prefrontal cortex connectivity during task-free MRI," *Scientific Reports,* vol. 8, no. 5571, 2018.

[15] X. Jiang, L. Chang and Y. Shi, "A retrospective analysis of the dynamic transmission routes of the COVID-19 in mainland China," *Scientific Reports,* vol. 10, no. 14015, 2020.

[16] P. J. Turk, T. P. Tran, G. A. Rose and A. McWilliams, "A predictive internet-based model for COVID-19 hospitalization census," *Scientific Reports,* vol. 11, no. 5106, 2021.

[17] S. J. Nickell, "Biaises in dynamic models with fixed effects," *Econ0metrica,* vol. 49, pp. 1417-1426, 1981.

[18] J. Hahn and G. Kuersteiner, "Asymptotically unbiaised inference for a dynamic panel with fixed effects when both n and T are large," *Econometrica,* vol. 70, pp. 1639-1657, 2002.

[19] H. d'Albis, E. Boubtane and D. Coulibaly, "Macroeconomic evidence suggests that asylum seekers are not a "burden" for Western European countries," *Science Advances,* vol. 4(6), no. eaaq0883, 2018.

[20] H. Hütkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2005.

[21] H. Salje, C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle and S. Cauchemez, "Estimating the burden of SARS-CoV-2 in France," *Science,* vol. 369, no. 6500, pp. 208-211, 2020.

[22] D. Kim, Z. Chen, L.-F. Zhou and S.-X. Huang, "Air pollutants and early origins of respiratory diseases," *Chronic Diseases and Translational Medicine,* vol. 4, no. 2, pp. 75-94, 2018.

[23] Y. Yao, J. Pan, W. Wang, Z. Liu, H. Kan, Y. Qiu, X. Meng et W. Wang, «Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities,» *Science of The Total Environment,* vol. 741, n° %1140396, 2020.

[24] M. Travaglio, Y. Yu, R. Popovic, L. Selley, N. Santos Leal et L. M. Martins, «Links between air pollution and COVID-19 in England,» *Environmental Pollution,* vol. 268 (A), n° %1115859, 2021.

[25] A. F. Skirienė et Z. Stasiskiene, «COVID-19 and Air Pollution: Measuring Pandemic Impact to Air Quality in Five European Countries,» *Atmosphere,* vol. 12 (3), n° %1290, 2021.

Authors' contribution

Conceptualization and Supervision: HA, DC, AR
Data collection: HA, EDC, RB
Formal analysis: HA, DC
Interpretation: HA, DC, AR, EDC, RB
Writing: HA, DC, EDC, RB

The authors declare no competing interests.

This research received no specific grant from any funding agency.

Supplementary Materials for

Quantification of the effects of climatic conditions on French hospital admissions and deaths induced by SARS-CoV-2

Hippolyte d'ALBIS - Dramane COULIBALY - Alix ROUMAGNAC - Eurico DE CARVALHO FILHO - Raphaël BERTRAND

This document includes:

1. Geometrical representations of the IPTCC index
2. Plots of the variables of interest for the 54 administrative regions
3. Correlation coefficients between variables
4. Dynamic responses following an increase in the 'false' IPTCC
5. Dynamic responses following an increase in the normalized temperature index
6. Granger causality in a model extended for air pollution indicators
7. Estimations of the epidemiological responses to a change in the IPTCC and in the air pollution indicators

1. Geometrical representations of the IPTCC index

We identified 5 intervals for the IPTCC that correspond to different suitability conditions for the spread of the virus. Following Roumagnac et al. (2021) a color code can be associated to each interval in Table A1. 2D and 3D representation of the IPCC as a function of relative humidity and temperature are then provided in Figures A1 and A2.

Table A1: Color code and intervals on the IPTCC value

IPTCC	**Color code**	**Definition**
[0,20[		Climatic conditions limiting airborne spread of the virus.
[20,75[		Climatic conditions becoming suitable to the airborne spread of the virus.
[75,90[		Climatic conditions suitable to the airborne spread of the virus.
[90,97[		Climatic conditions very suitable to the airborne spread of the virus.
[97,100]		Climatic conditions extremely suitable to the airborne spread of the virus.

Figure A1. IPTCC as a function of temperature and relative humidity, 2D representation

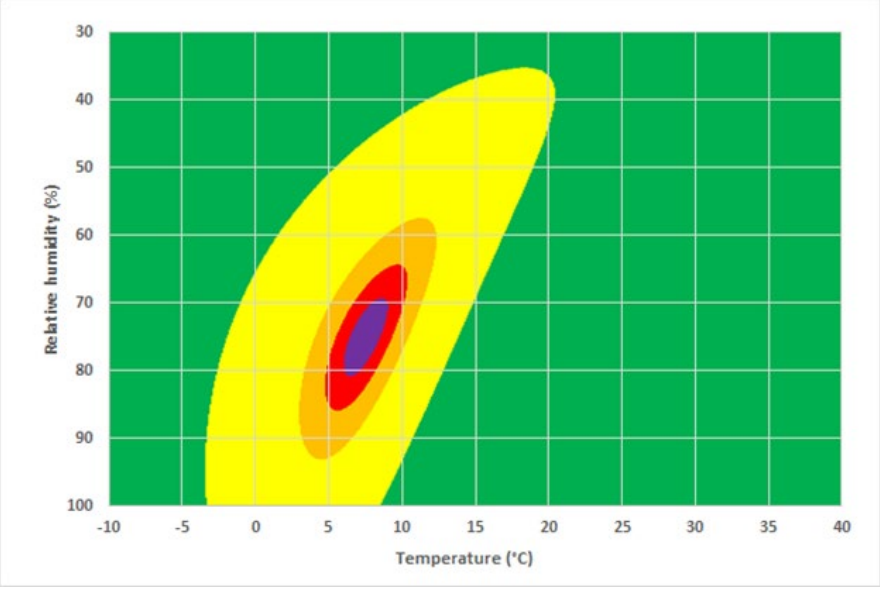
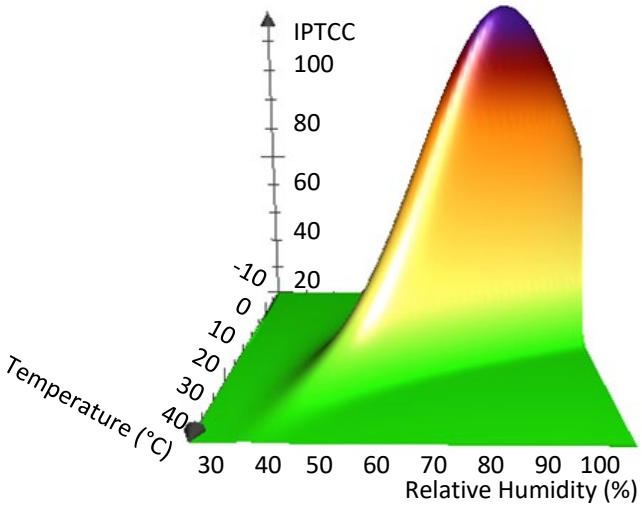
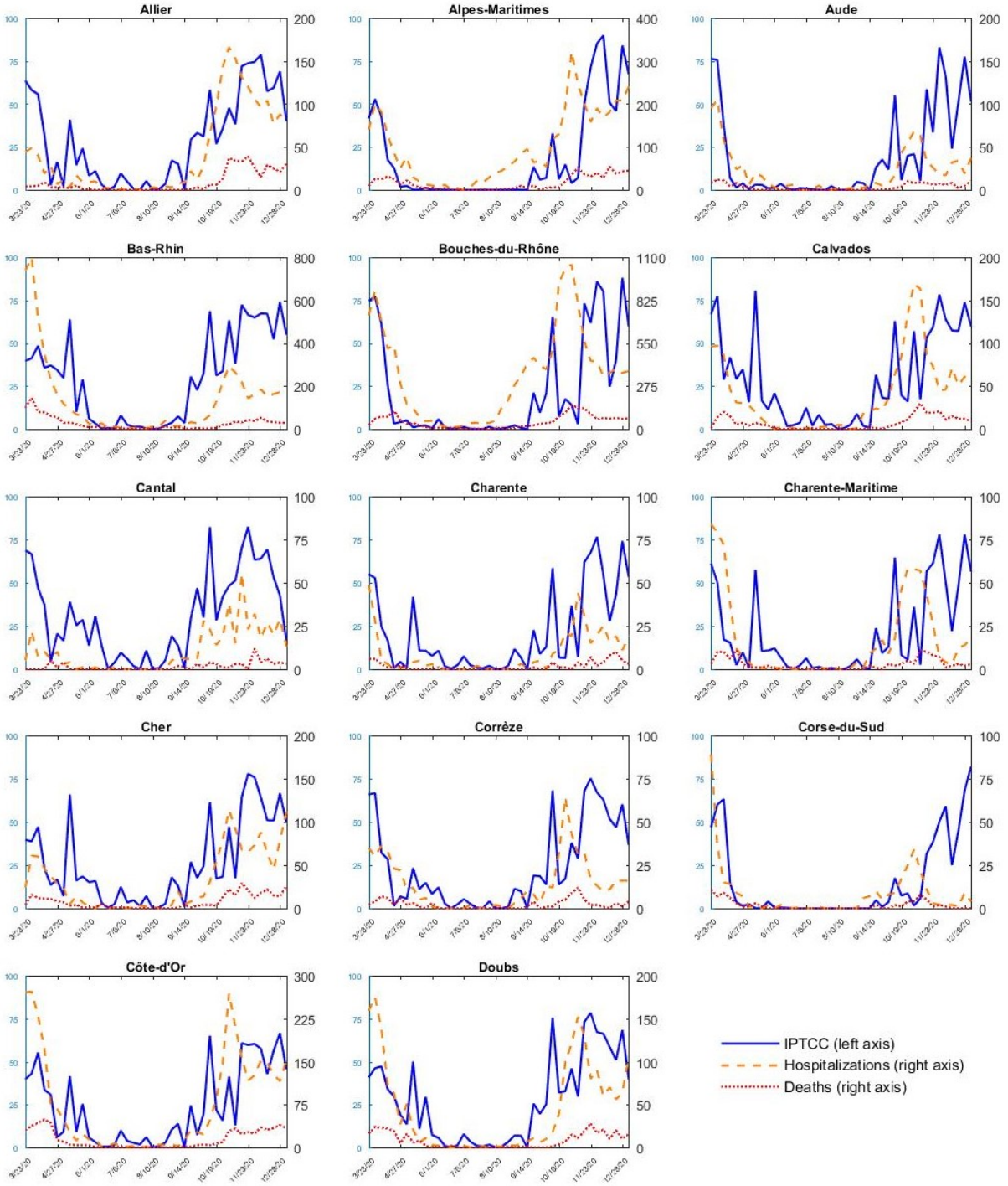


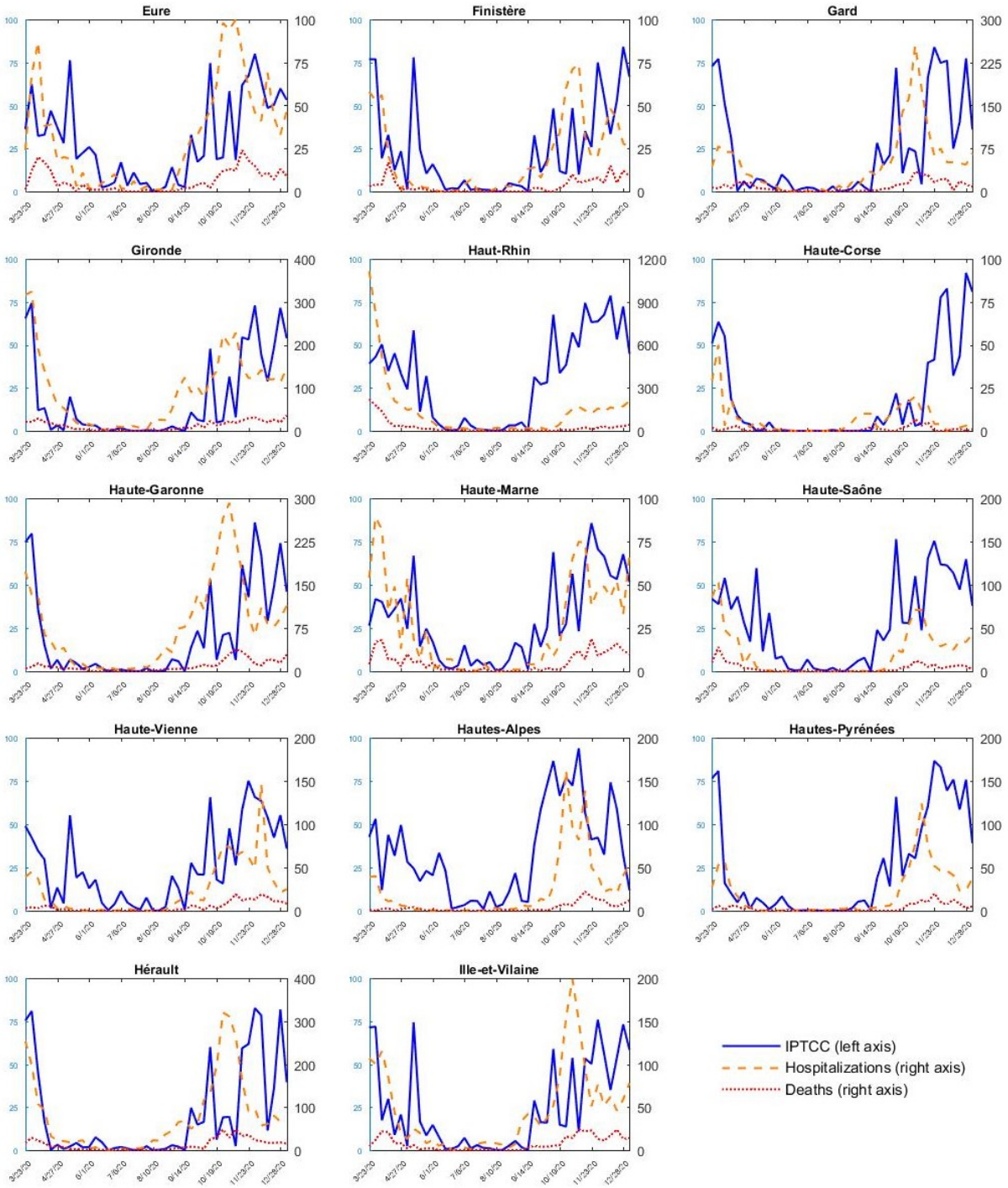
Figure A2. IPTCC as a function of temperature and relative humidity, 3D representation

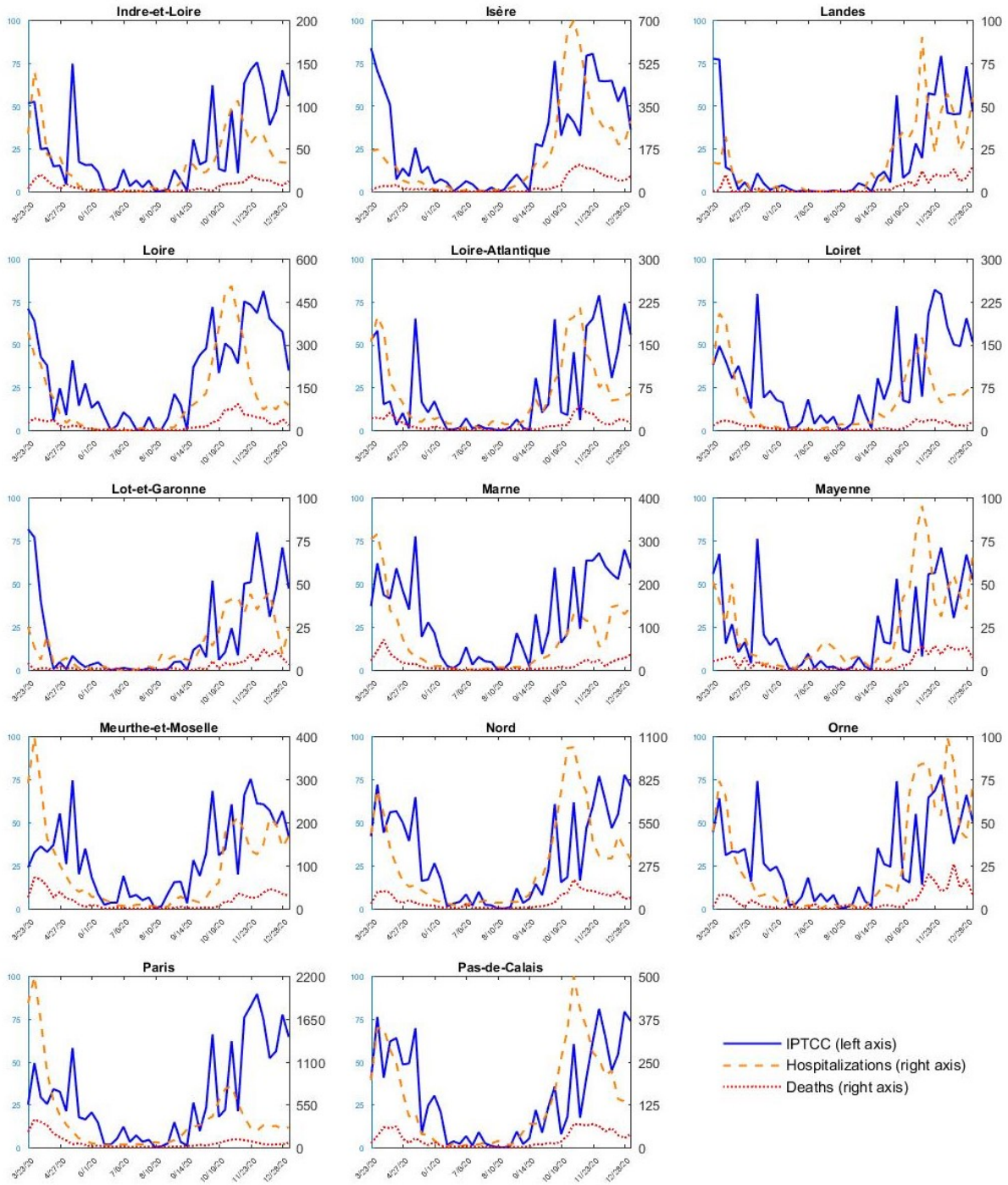


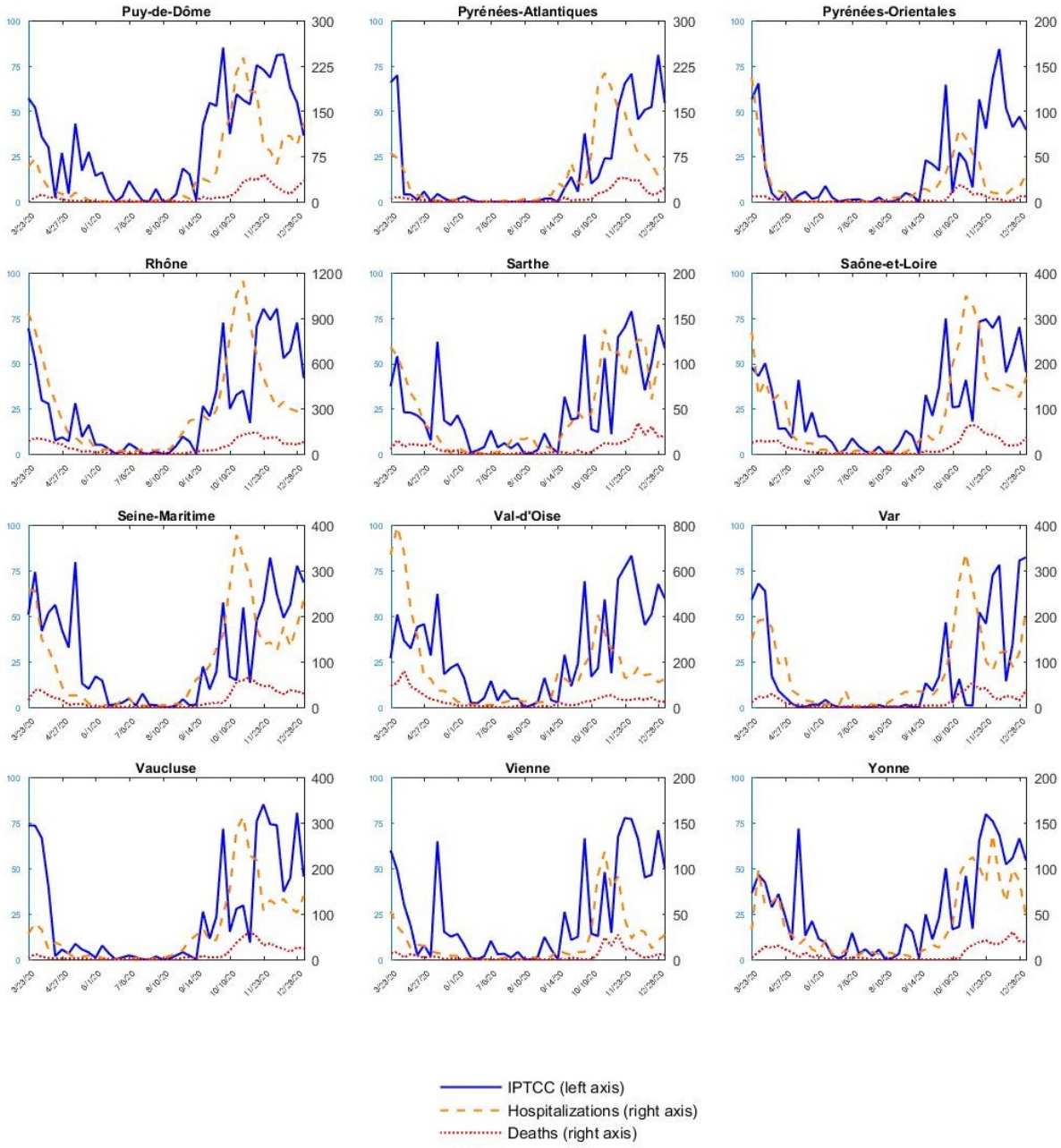
2. Plots of the variables of interest for the 54 administrative regions

Figure A3. Comparison of the time series of hospitalizations, deaths and IPTCC for each administrative region, from March 23, 2020 to January 10, 2021




Eure
Finistère
Gard
Gironde
Haut-Rhin
Haute-Corse
Haute-Garonne
Haute-Marne
Haute-Saône
Haute-Vienne
Hautes-Alpes
Hautes-Pyrénées
Hérault
Ille-et-Vilaine
IPTCC (left axis)
Hospitalizations (right axis)
Deaths (right axis)


Indre-et-Loire
Isère
Landes
Loire
Loire-Atlantique
Loiret
Lot-et-Garonne
Marne
Mayenne
Meurthe-et-Moselle
Nord
Orne
Paris
Pas-de-Calais
IPTCC (left axis)
Hospitalizations (right axis)
Deaths (right axis)


Puy-de-Dôme
Pyrénées-Atlantiques
Pyrénées-Orientales
Rhône
Sarthe
Saône-et-Loire
Seine-Maritime
Val-d'Oise
Var
Vaucluse
Vienne
Yonne
IPTCC (left axis)
Hospitalizations (right axis)
Deaths (right axis)

3. Correlation coefficients between variables

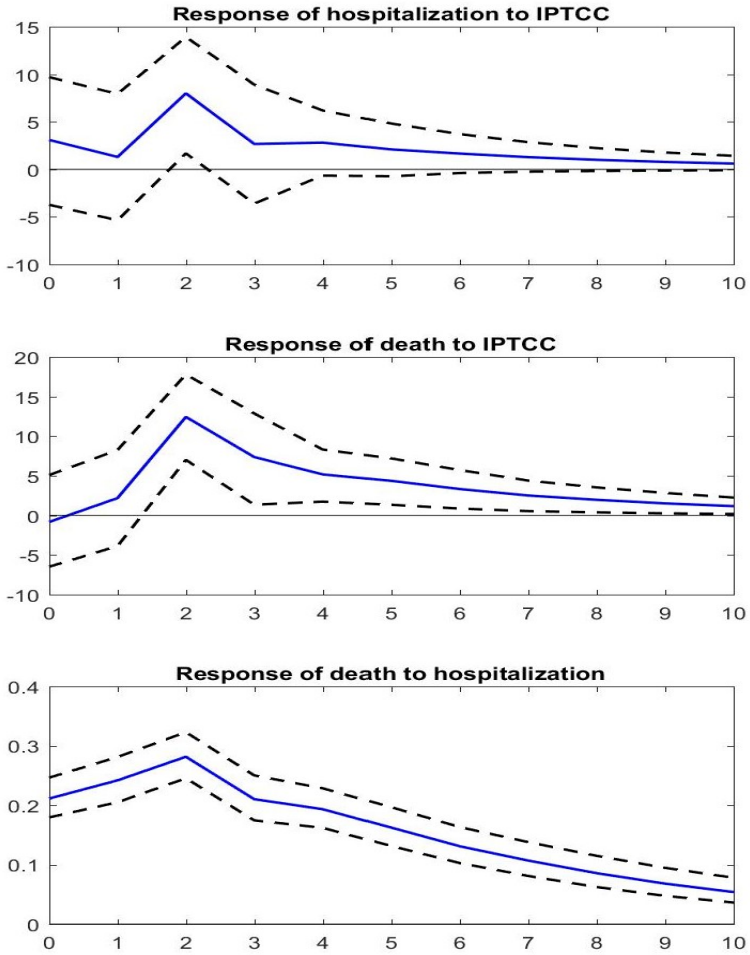
Table A2: Correlation coefficients between hospitalizations, deaths and the IPTCC

Region	corr(hospitlzation,death)	corr(hospitalization,IPTCC)	corr(death,IPTCC)
Allier	0.890	0.808	0.806
Alpes-Maritimes	0.800	0.588	0.663
Aude	0.838	0.573	0.640
Bas-Rhin	0.879	0.775	0.642
Bouches-du-Rhône	0.830	0.528	0.544
Calvados	0.817	0.675	0.713
Cantal	0.665	0.765	0.414
Charente	0.722	0.733	0.701
Charente-Maritime	0.771	0.445	0.459
Cher	0.897	0.686	0.701
Corrèze	0.812	0.598	0.412
Corse-du-Sud	0.750	0.382	0.327
Côte-d'Or	0.934	0.752	0.775
Doubs	0.894	0.802	0.732
Eure	0.902	0.651	0.686
Finistère	0.732	0.562	0.616
Gard	0.836	0.531	0.574
Gironde	0.818	0.577	0.665
Haut-Rhin	0.931	0.674	0.526
Haute-Corse	0.571	0.222	-0.001
Haute-Garonne	0.888	0.575	0.562
Haute-Marne	0.882	0.715	0.741
Haute-Saône	0.940	0.802	0.726
Haute-Vienne	0.895	0.707	0.736
Hautes-Alpes	0.756	0.745	0.551
Hautes-Pyrénées	0.822	0.755	0.697
Hérault	0.829	0.521	0.644
Ille-et-Vilaine	0.842	0.614	0.671
Indre-et-Loire	0.778	0.599	0.597
Isère	0.895	0.775	0.756
Landes	0.806	0.687	0.634
Loire	0.896	0.741	0.783
Loire-Atlantique	0.844	0.547	0.583
Loiret	0.806	0.610	0.671
Lot-et-Garonne	0.738	0.600	0.683
Marne	0.883	0.769	0.768
Mayenne	0.830	0.576	0.713
Meurthe-et-Moselle	0.910	0.703	0.670
Nord	0.885	0.699	0.784
Orne	0.822	0.676	0.643
Paris	0.904	0.493	0.550
Pas-de-Calais	0.890	0.710	0.805
Puy-de-Dôme	0.873	0.841	0.775
Pyrénées-Atlantiques	0.900	0.685	0.790
Pyrénées-Orientales	0.842	0.568	0.475
Rhône	0.895	0.682	0.732
Sarthe	0.795	0.645	0.674
Saône-et-Loire	0.959	0.740	0.754
Seine-Maritime	0.880	0.652	0.759
Val-d'Oise	0.859	0.580	0.639

Var	0.918	0.567	0.581
Vaucluse	0.903	0.660	0.683
Vienne	0.827	0.634	0.652
Yonne	0.815	0.737	0.695
National average	0.843	0.647	0.644

4. Dynamic responses following an increase in the 'false' IPTCC index

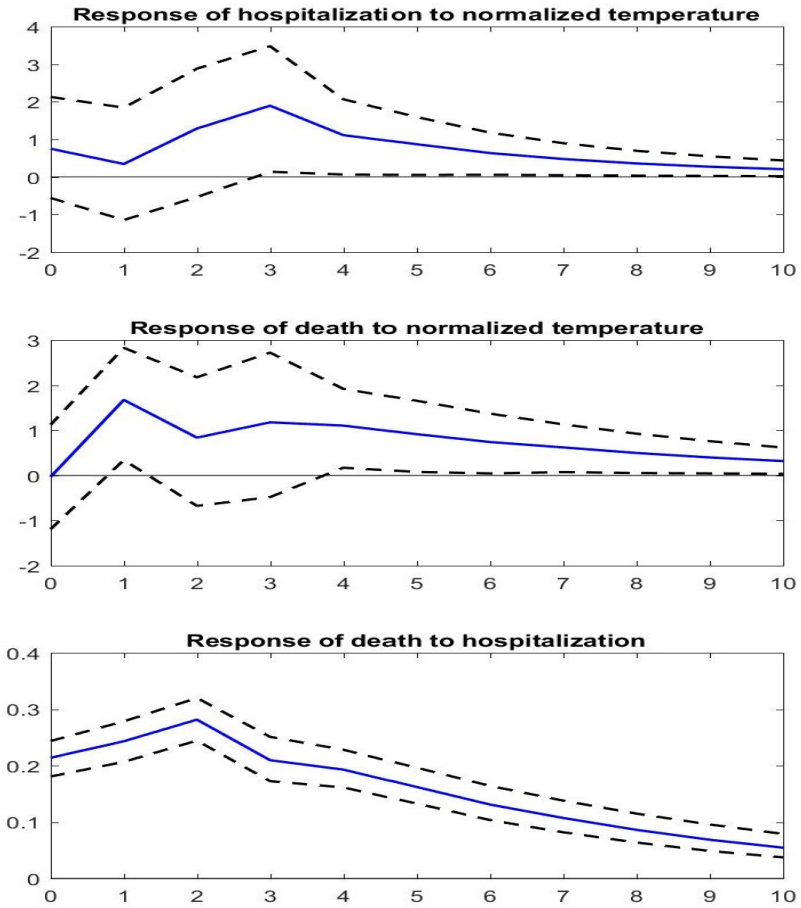
Figure A4: Responses of hospitalizations and deaths following an increase in the 'false' IPTCC index



Notes: The solid line gives the estimated impulse responses. Dashed lines give the 90% confidence intervals generated by Monte Carlo with 5,000 repetitions. The size of the increase in the 'false' IPTCC is set to 10-point increase. The size of the increase in hospitalizations is set to one percent increase. The responses are the percent change in the number of hospitalizations and deaths.

5. Dynamic responses following an increase in the normalized temperature index

Figure A5: Responses of hospitalizations and deaths following an increase in the normalized temperature index.



Notes: The solid line gives the estimated impulse responses. Dashed lines give the 90% confidence intervals generated by Monte Carlo with 5,000 repetitions. The size of the increase in the normalized temperature index is set to 10-point increase. The size of the increase in hospitalizations is set to one percent increase. The responses are the percent change in the number of hospitalizations and deaths.

6. Granger causality in a model extended for air pollution indicators

We have redone our computations of the Granger causality by including the following indicators: the concentration (in µg/m^3) of atmospheric particulate matter, either for particulates with a diameter of 10 µm (PM_{10}) and for those with a diameter of 2.5 µm ($PM_{2.5}$), of nitrogen dioxide (NO_2) and of ozone (O_3). We take the weekly average of the daily concentrations registered in all the Météo-France stations that monitor the pollution. Since the coverage of pollution is less important than that of climate, the number of regions used in the estimation is lower. It is equal to 21 for PM_{10}, 18 for $PM_{2.5}$ and NO_2, and 23 for O_3.

Equation (1) was replaced by:

$$y_{it} = \sum_{s=1}^{p} \alpha_s \, y_{it-s} + \sum_{s=1}^{p} \beta_s \, cl_{it-s} + \sum_{s=1}^{p} \gamma_s \, ap_{it-s} + \mu_i + \delta_i . t + \varepsilon_{it},$$

$$i = 1, \dots N \; and \; t = 1, \dots, T \qquad (A1)$$

where ap_{it-s} is the considered air pollution indicator. Remark that, given the small number of regions with data on air pollution a common national time effect is not included in this latter equation.

As previously, the null hypothesis of no Granger causality from climatic conditions to hospitalizations or deaths is H_0: $\beta_1 = \beta_2 = \cdots = \beta_p$. The corresponding Wald test statistic is $W = \hat{\theta}' R' [\hat{\sigma}^2 R (X'X)^{-1} R']^{-1} R \hat{\theta}$ where $\hat{\theta}$ is the estimator of $\theta = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_p)'$, R is a known $(p \times 2p)$ matrix with $R = [0 : I_p]$, X is a $(N(T - p) \times 3p)$ matrix of $\tilde{y}_{it-1}, \dots, \tilde{y}_{it-p}, \widetilde{cl}_{it-1}, \dots, \widetilde{cl}_{it-p}, \widetilde{ap}_{it-1}, \dots, \widetilde{ap}_{it-p}$ that are the transformations of corresponding variables after removing region specific effects and region specific trends, and $\hat{\sigma}^2$ is the estimated variance of the residual in Equation (A1). Under the

null hypothesis, W follows a chi-squared distribution of a degree of freedom equal to p (which correspond to the number of constraints to be tested that corresponds to the lag length). In the same vein, based on Equations (A1), we also apply the Granger causality from air pollution to hospitalizations or deaths (controlling for climate conditions, IPTCC), with the corresponding hypothesis H_0: $\gamma_1 = \gamma_2 = \cdots = \gamma_p$.

Table A3 reports the results of Granger non-causality from IPTCC to hospitalizations and deaths controlling for and air pollution indicator and from air pollution to hospitalizations and deaths controlling for the IPTCC. When accounting for the potential impact of air pollution on hospitalizations and deaths, at the 5% level of significance (actually, even at the 1% level), we cannot accept the null hypothesis of no Granger causality from IPTCC to either hospitalizations or deaths induced by SARS-CoV-2. However, we cannot reject the null of hypothesis of no Granger causality from air pollution to either hospitalizations or deaths. This does not imply that air pollution does not contemporaneously affect hospitalizations and deaths, since Granger causality assesses how lag values of air pollution are useful to predict hospitalizations and deaths. The contemporaneous impact can be computed with the VAR by the response to an innovation or a structural shock of air pollution (see below).

Table A3: Granger causality from the IPTCC to hospitalizations and deaths, controlling for air pollution

Model with measures of PM_{10}		
Hypothesis	Test statistics	P-value
IPTCC does not Granger-cause hospitalizations	17.389	0.001
IPTCC does not Granger-cause deaths	28.478	0.000
Air pollution does not Granger-cause hospitalizations	2.775	0.428
Air pollution does not Granger-cause deaths	1.467	0.690
Model with measures of $PM_{2.5}$		
Hypothesis	Test statistics	P-value
IPTCC does not Granger-cause hospitalizations	12.830	0.005
IPTCC does not Granger-cause deaths	23.401	0.000
Air pollution does not Granger-cause hospitalizations	5.456	0.141
Air pollution does not Granger-cause deaths	4.869	0.182
Model with measures of NO_2		
Hypothesis	Test statistics	P-value
IPTCC does not Granger-cause hospitalizations	20.351	0.000
IPTCC does not Granger-cause deaths	28.828	0.000
Air pollution does not Granger-cause hospitalizations	1.513	0.679
Air pollution does not Granger-cause deaths	3.178	0.365
Model with measures of O_3		
Hypothesis	Test statistics	P-value
IPTCC does not Granger-cause hospitalizations	22.939	0.000
IPTCC does not Granger-cause deaths	33.400	0.000
Air pollution does not Granger-cause hospitalizations	2.093	0.553
Air pollution does not Granger-cause deaths	4.830	0.186

Notes: The test statistic is a Wald statistic which follows, under the null hypothesis, a chi-squared distribution of 3 (the number of constraints that corresponds to the lag length).

7. Estimations of the epidemiological responses to a change in the IPTCC in models that include an air pollution indicator

As for Granger causality, the estimations of the dynamic responses of the epidemiological variables to a change in the IPTCC can be done in a model augmented with an air pollution indicator (denoted as in A1, ap_{it} that can be the concentration of PM_{10}, $PM_{2.5}$, NO_2 or O_3). Equation (2) was replaced by:

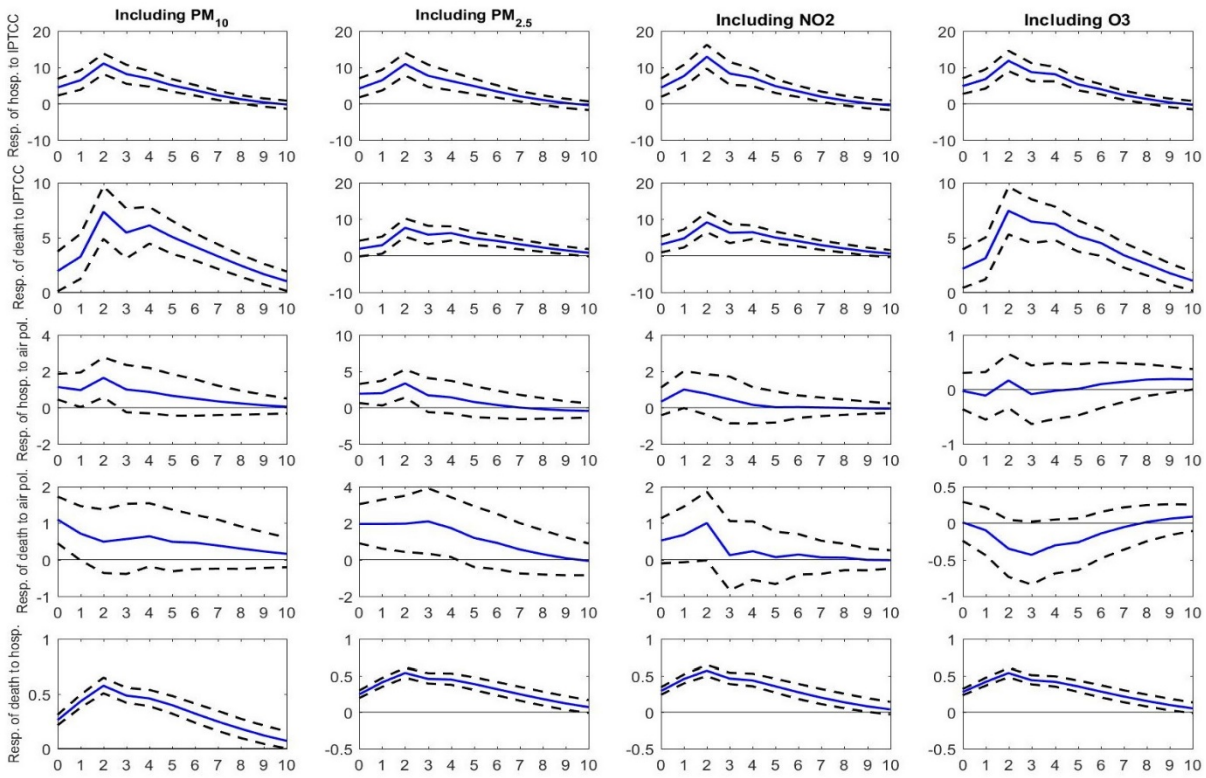
$$Z_{it} = \sum_{s=1}^{p} A_s Z_{it-s} + \sum_{s=0}^{p} b_s cl_{it-s} + u_i + d_i.t + v_{it}, \;\; i = 1, \dots N \; and \; t = 1, \dots, T \qquad (A2)$$

where A_s are 3×3 matrices of coefficients associated with $Z_{it} = (ap_{it-s}, hosp_{it-s}, death_{it-s})'$, b_s are 3×1 vector of coefficient associated with cl_{it-s}, $u_i = (u_i^1, u_i^2, u_i^3)'$ is a vector of region fixed-effects; $d_i.t = (d_i^1, d_i^2, d_i^3)'.t$ represent region specific-time (linear) trend; $v_{it} = (v_{it}^1, v_{it}^2, v_{it}^3)'$ is a 3-dimensional vector of errors satisfying $E(v_{it}) = 0$ and $E(v_{it}v_{is}') = \Omega.\mathbb{1}\{t = s\}$ for all t and s. It is worth noting that air pollution is included in the endogenous vector Z_{it}, since it can be affected by the pandemic disease. As in Equation (A1), given the small number of regions with data on air variable a common national time effect is not included in Equation (A2).

As in Equation (2), to identify the response between endogenous variables (air pollution, deaths and hospitalization, we need to identify the structural shocks η_{it} of these endogenous variables as follow: $\eta_{it} = A_0 v_{it}$ where A_0 is (3×3) matrix such that $E(\eta_{it}\eta_{it}') = I_3$ or $A_0 A_0' = \Omega$. We identified A_0 based on Cholesky decomposition by setting A_0 as the unique lower-triangular Cholesky factor of Ω. This identification relies on the reasonable assumption that air pollution may contemporaneously influence hospitalizations and deaths, hospitalizations may contemporaneously influence deaths, while hospitalizations can

potentially influence air pollution only with, and deaths can potentially influence air pollution and hospitalizations only with lags. Results are reported in Figure A6. We first see that the control for air pollution does not qualitatively alter the dynamic response of hospitalizations and deaths to the IPTCC. Second, we see that the two measures of concentration of atmospheric particulate matter (namely PM_{10} and $PM_{2.5}$) have a significant impact on hospitalizations and deaths induced by SARS-CoV-2, while the concentrations of nitrogen dioxide (NO_2) and of ozone (O_3) have no impact. We remark that even if PM_{10} and $PM_{2.5}$ do not Granger cause hospitalizations and deaths, they have significant contemporaneous impact inducing a significant dynamic impact.

Figure A6. Dynamic responses of hospitalizations and deaths to IPTCC and air pollution



Note: The solid line gives the estimated impulse responses. The dashed lines give the 90% confidence intervals generated by Monte Carlo with 5,000 repetitions. The size of the increase in the IPTCC is set to a 10-point increase. The size of the increase in hospitalizations is set to a one-percent increase. The responses are the percentage change in the number of hospitalizations and deaths.