



## OPEN ACCESS

## EDITED BY

Zhimin Tao,  
Jiangsu University, China

## REVIEWED BY

Andrew W. Bartlow,  
Los Alamos National Laboratory (DOE),  
United States  
Till Frank,  
University of Connecticut, United States

## \*CORRESPONDENCE

Jesús Planella-Morató  
✉ [jesus.planella@udg.edu](mailto:jesus.planella@udg.edu)  
Josep L. Pelegrí  
✉ [pelegr@icm.csic.es](mailto:pelegr@icm.csic.es)  
Ignasi Vallès-Casanova  
✉ [ignasi.valles@ieo.csic.es](mailto:ignasi.valles@ieo.csic.es)

RECEIVED 10 May 2024

ACCEPTED 01 November 2024

PUBLISHED 05 December 2024

## CITATION

Planella-Morató J, Pelegrí JL, Martín-Rey M,  
Olivé Abelló A, Vallès X, Roca J, Rodrigo C,  
Estrada O and Vallès-Casanova I (2024)  
Environmental predictors of SARS-CoV-2  
infection incidence in Catalonia  
(northwestern Mediterranean).  
*Front. Public Health* 12:1430902.  
doi: 10.3389/fpubh.2024.1430902

## COPYRIGHT

© 2024 Planella-Morató, Pelegrí, Martín-Rey,  
Olivé Abelló, Vallès, Roca, Rodrigo, Estrada  
and Vallès-Casanova. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Environmental predictors of SARS-CoV-2 infection incidence in Catalonia (northwestern Mediterranean)

Jesús Planella-Morató<sup>1,2,3\*</sup>, Josep L. Pelegrí<sup>1\*</sup>,  
Marta Martín-Rey<sup>4</sup>, Anna Olivé Abelló<sup>1</sup>, Xavier Vallès<sup>5,6,7</sup>,  
Josep Roca<sup>8</sup>, Carlos Rodrigo<sup>9,10</sup>, Oriol Estrada<sup>11</sup> and  
Ignasi Vallès-Casanova<sup>1,12,13\*</sup>

<sup>1</sup>Departament d'Oceanografia Física i Tecnològica, Institut de Ciències del Mar, CSIC, Barcelona, Spain, <sup>2</sup>Departament de Física, Universitat de Girona, Girona, Spain, <sup>3</sup>University School of Health and Sport (EUSES), University of Girona, Girona, Spain, <sup>4</sup>Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid, Madrid, Spain, <sup>5</sup>Fundació Lluita contra les Infeccions, Badalona, Spain, <sup>6</sup>Fundació Institut per la Recerca Germans Trias i Pujol, Badalona, Spain, <sup>7</sup>Programa de Salut Internacional Institut Català de la Salut (PROSICS), Badalona, Spain, <sup>8</sup>Epidemiology Unit, Hospital Universitari Germans Trias i Pujol, Institut Català de la Salut, Badalona, Spain, <sup>9</sup>Department of Pediatrics, Institut de Recerca Germans Trias i Pujol, Badalona, Spain, <sup>10</sup>Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, <sup>11</sup>Directorate for Innovation and Interdisciplinary Cooperation, Northern Metropolitan Region from Barcelona, Institut Català de la Salut, Barcelona, Spain, <sup>12</sup>Hebrew University of Jerusalem, Jerusalem, Israel, <sup>13</sup>Centro Oceanográfico de Santander, Instituto Español de Oceanografía, IEO-CSIC, Santander, Spain

Numerous studies have explored whether and how the spread of the SARS-CoV-2, the causative agent of coronavirus disease 2019 (COVID-19), responds to environmental conditions without reaching consistent answers. Sociodemographic factors, such as variable population density and mobility, as well as the lack of effective epidemiological monitoring, make it difficult to establish robust correlations. Here we carry out a regional cross-correlation study between nine atmospheric variables and an infection index ( $I_c$ ) estimated from standardized positive polymerase chain reaction (PCR) test cases. The correlations and associated time-lags are used to build a linear multiple-regression model between weather conditions and the  $I_c$  index. Our results show that surface pressure and relative humidity can largely predict COVID-19 outbreaks during periods of relatively minor mobility and meeting restrictions. The occurrence of low-pressure systems, associated with the autumn onset, leads to weather and behavioral changes that intensify the virus transmission. These findings suggest that surface pressure and relative humidity are key environmental factors that may be used to forecast the spread of SARS-CoV-2.

## KEYWORDS

environmental sciences, epidemiology, COVID-19, atmospheric variables, weather-infection correlation, virus spread, climate change

## 1 Introduction

Soon after the onset of the COVID-19 disease in December 2019 in Wuhan (China), it became clear that the newly identified causative agent (SARS-CoV-2) had higher transmissivity than the previous coronavirus SARS-CoV-1 and MERSV (1). The declaration of a global pandemic on 11 March 2020 confirmed the worst fears of the high infectiousness capacity of SARS-CoV-2. In this situation, and according to the World Health Organization (WHO),

governments implemented a package of measures to deal with the virus. However, by late August 2023 the pandemic of COVID-19 had claimed almost 7 million notified deaths and about 770 million confirmed cases (94), besides the important socio-economic disruptions due to the lockdowns and contention measures. During that time, the scientific community stepped in to help searching for the origins of SARS-CoV-2 and determining the mechanisms that caused its spreading.

As an air-borne disease and with the purpose of developing climate-based predictive tools, the pandemic research soon focused on the climatic and environmental factors that might affect its transmissibility and virulence (2). Several studies explored the relationship between weather conditions and virus transmission, using meteorological or environmental variables as prognostic tools to reproduce the propagation of COVID-19. The different types of models used to predict the spread of the virus are listed in Table 1 along with their associations with the weather variables. Non-linear models include polynomial regressions (PR) and generalized additive models (GAM), incorporating time-lagged weather effects (DLNM). Other efforts have used more simple linear models that work reasonably well, such as single or multiple linear regression (MLR) models and generalized linear (GLM) models. The results show the existence of significant correlations for temperature and relative humidity with virus transmission, with the transmission of SARS-CoV-2 and severity of COVID-19 increasing during dry and cold weather conditions in European countries (Table 1), modulated by large scale atmospheric patterns (3–5). However, the results remain unclear for other variables, such as precipitation, wind speed, solar radiation or surface pressure (Table 1).

It is now clear that external factors such as social behavior (e.g., mobility, use of indoor spaces) and mitigation policies (e.g., use of face masks, vaccination) play the main role in the transmission of SARS-CoV-2 (6, 7). Nevertheless, it has been observed that a small but yet significant fraction of the occurrence and persistence of SARS-CoV-2 is explained by the weather conditions (1–10, 93). When considering the atmospheric variables individually, their predictive skill does not exceed 10%, a figure that increases to about 18% as several variables are considered jointly (8). The complex non-linear interrelation between the weather variables can substantially deviate the estimates obtained from the models. The connection between weather factors is not easily established from observational data, remaining as a challenging task (11). In this context, MLR models are applied to datasets to visualize trends but they have to be analyzed carefully in order to obtain adequate results.

From a mid-term preventive perspective, it is important to improve our capacity to predict seasonal patterns and future outbreaks of SARS-CoV-2, including those new emerging variants with high transmissibility patterns (i.e., Omicron and derivatives). From a long-term adaptative perspective, in the context of the current climate change scenario, a better understanding of the impact of climatic and environmental factors in respiratory diseases becomes crucial to design effective adaptation and mitigation strategies.

The ongoing evolution of SARS-CoV-2 has entered a new scenario where humans are experiencing reinfections with sub-variants of the Omicron lineage, characterized by very high infectiousness. The impact of vaccines on transmissibility of SARS-CoV-2 Omicron variants remains low, as the combination of relatively short immunity against infection and high pathogen genomic variability increase the rate of repeated infection (12, 13) and extend the persistence of coronavirus over time (14, 15). This new scenario has evolved toward

an endemic disease, possibly controlled by weather conditions that cause outbreaks or seasonal peaks similar to most common respiratory infections (16, 17).

In the new unfolding epidemiological scenario, where policy interventions and social distancing are already residual, the seasonality of COVID-19 deserves further attention. The predominance of Omicron-related variants, whose infectiousness is highly independent of the vaccinated status, emphasizes the need for a better understanding and characterization of the climatic factors that impact the susceptibility to infection. The dry and warm future scenario for the Euro-Mediterranean region (18) further stresses the need for identifying the main meteorological and environmental precursors for respiratory diseases, and establishing what are their lead mean times.

Here we assess the relationship between the spread of SARS-CoV-2 and the atmospheric conditions in Catalonia (northwestern Mediterranean) from September 2020 to December 2020. A multiple linear regression, based on cross-correlation results between a simple infection index and time-lagged atmospheric variables, is used to model and forecast the impact of weather on the COVID-19 incidence. The model is validated externally during the third COVID-19 outbreak, from December 2020 to February 2021, in order to assess its predictive performance. These results are critically compared with updated findings, on the capacity of environmental factors to predict the seasonal dynamics of the SARS-CoV-2 spread (Table 1).

## 2 Materials and methods

### 2.1 Health data

The health data used in this study are available from the database of the Catalan government (see Data Availability section) and were processed using the MATLAB software (MATLAB R21017a). Due to the limited accuracy of antigen tests, which depends on the time-window considered and results in false negatives, only positive cases detected through PCR ( $N_{PCR,+}$ ) were selected for the analysis. Detected cases were grouped into Health Regions (HRs, a total of nine with mean area about 3,600 km<sup>2</sup>) and, for each HR dataset, were broken down into Basic Health Areas (BHAs, a total of 372 with mean area about 86 km<sup>2</sup> and population ranging between 5,000 and 25,000 people each). Afterwards, a rectangular grid of 0.1° × 0.1° latitude (*lat*) – longitude (*lon*) was generated covering all Catalonia. Health data were assigned to each grid point according to their location in the BHAs, so that grid points located inside one same BHA contain the same health data. BHAs that did not report data during the pandemic were excluded from the preliminary data analysis. As a result, time series for health data at each grid point were generated for the period of time analyzed. These time series were normalized dividing by the area and the population size of each BHA; the resulting series  $N_{PCR,+}(t, lat, lon)$  are positive PCR cases for 100,000 inhabitants and square kilometer (cases per 10<sup>5</sup> inhab km<sup>2</sup>).

We applied two corrections to the normalized time series  $N_{PCR,+}(t, lat, lon)$ . The first one considered the observed weekday biases, e.g., typically as a result of increased counts after the weekend. For this purpose, a histogram of mean confirmed cases for each day of the

**TABLE 1** Associations between different weather variables and virus propagation indicated by colors: red (positive), blue (negative), yellow (no association), green (unclear), pink (depending on the value) and grey (not analyzed).

Related Work	Model approach	Spatial Resolution	Temporal Resolution	$T_{mean}$	$RH$	$Rad$	$P$	$Prec$	$w$	$T_{min}$	$T_{max}$
Sanchez-Lorenzo et al. (40)	Non-Linear (PR)	Regional	Monthly	Blue	Blue	Grey	Grey	Grey	Grey	Grey	Grey
Fang et al. (73)	Non-Linear (PR)	Regional	Daily	Blue	Grey	Grey	Grey	Grey	Grey	Grey	Grey
Gao et al. (74)	Non-Linear (PR)	Global	Daily	Blue	Red	Grey	Grey	Grey	Grey	Grey	Grey
Prata et al. (75)	Non-Linear (GAM)	Regional	Daily	Pink	Grey	Grey	Grey	Grey	Grey	Grey	Grey
Qi et al. (76)	Non-Linear (GAM)	Regional	Daily	Pink	Pink	Grey	Grey	Grey	Grey	Grey	Grey
Karim and Akter (77)	Non-Linear (GAM)	Regional	Daily	Blue	Blue	Grey	Grey	Grey	Grey	Grey	Grey
Fu et al. (3)	Non-Linear (DLNM)	Regional	Daily	Blue	Blue	Grey	Grey	Grey	Grey	Grey	Grey
Yuan et al. (78)	Non-linear (GAM)	Global	Daily	Pink	Pink	Grey	Grey	Grey	Grey	Pink	Grey
Metelmann et al. (79)	Non-linear (GAM)	Regional	Daily	Blue	Blue	Grey	Grey	Red	Grey	Red	Grey
Islam et al. (59)	Non-linear (DLNM)	Regional	Daily	Blue	Blue	Grey	Grey	Grey	Grey	Red	Blue
Fong and Smith (80)	Non-linear (DLNM)	Regional	Daily	Blue	Grey	Grey	Grey	Grey	Grey	Grey	Grey
Ai et al. (81)	Non-linear (DLNM)	Global	Daily	Blue	Green	Blue	Grey	Green	Grey	Grey	Grey
Ladha et al. (82)	Linear (MLR)	Local	Daily	Yellow	Yellow	Grey	Grey	Grey	Grey	Grey	Yellow
Hu et al. (83)	Linear (MLR)	Regional	Daily	Grey	Grey	Grey	Grey	Grey	Grey	Pink	Pink
Hoogeveen et al. (84)	Linear (MLR)	Regional	Daily	Blue	Yellow	Blue	Grey	Grey	Yellow	Grey	Grey
Culqui et al. (4)	Linear (MLR)	Global	Daily	Green	Red	Grey	Grey	Red	Blue	Grey	Grey
Matheew et al. (85)	Linear (MLR)	Regional	Daily	Blue	Blue	Grey	Grey	Grey	Red	Grey	Grey
Alnaser et al. (86)	Linear (MLR)	Global	Daily	Blue	Grey	Blue	Grey	Grey	Grey	Grey	Grey
Aidaoo et al. (87)	Linear (MLR)	Regional	Daily	Blue	Blue	Grey	Red	Grey	Red	Grey	Grey
Lin et al. (88)	Linear (MLR)	Global	Monthly	Grey	Blue	Grey	Yellow	Yellow	Yellow	Green	Green
Liu et al. (89)	Linear (GLM)	Regional	Daily	Blue	Blue	Grey	Grey	Grey	Grey	Grey	Grey
Culqui et al. (4)	Linear (GLM)	Regional	Daily	Yellow	Yellow	Grey	Grey	Grey	Grey	Grey	Grey

$T_{mean}$ : temperature;  $RH$ , relative humidity;  $Rad$ , solar radiation;  $P$ , surface pressure;  $Prec$ , precipitation;  $w$ , wind speed;  $T_{min}$ , minimum temperature;  $T_{max}$ , maximum temperature.

week at each BHA was computed, and a weekday factor was applied to the previous normalized dataset. A second PCR correction factor was defined as in Equation 1:

$$f_{PCR} = \frac{PCR_+(t, lat, lon)}{PCR_{tot}(t, lat, lon)}, \tag{1}$$

which represents the fraction between the daily positive tests ( $PCR_+$ ) and the total PCR tests ( $PCR_{tot}$ ) done during that day in each BHA region. This factor takes into account the availability of PCR tests, of particular importance during the first wave of the pandemic, as the number of available tests was significantly lower than during the other waves and the number of infected people at the early stages of the pandemic was underestimated.

The normalized time series at each grid point, obtained after applying the two correction factors, were smoothed with a three-day moving average. Further, an interpolant for scattered grid points was applied to estimate daily values in BHAs with no reported cases. The result was a daily COVID-19 health time series  $\hat{N}_{PCR,+}(t, lat, lon)$  at each point of the grid covering Catalonia.

## 2.2 Daily infection index

The final dataset  $\hat{N}_{PCR,+}(t, lat, lon)$  was used to define an infection index that monitors the contagion risk for the population in a specific area and day. The daily infection index at each grid point  $I_c(t, lat, lon)$  is computed as in Equation 2:

$$I_c(t, lat, lon) = \frac{\hat{N}_{PCR,+}(t, lat, lon)}{\sum_{i=1}^{10} \hat{N}_{PCR,+}(t-i, lat, lon)}, \quad (2)$$

This pandemic parameter can be obtained directly from the health dataset, providing information on the people infected daily with respect to the total population that is potentially infectious, which is estimated as the people infected during the prior 10 days.

After processing the health data, the BHAs were classified according to population density. In low-density populated areas, during 2020 there were some difficulties in the reporting of positive cases, mostly related to the local availability of tests. In order to avoid this problem, we selected eight densely populated areas (population density  $d \geq 500$  inhab km<sup>-2</sup>) for further analysis (Figure 1). One of the selected BHAs is located in the city of Barcelona (BCN-10A,  $d_{BCN} = 11,873$  inhab km<sup>-2</sup>) and five more are included in districts located in towns of the Barcelona metropolitan area: Gava (GVA-2,  $d_{GVA} = 1,847$  inhab km<sup>-2</sup>), Sant Just Desvern (SJD,  $d_{SJD} = 2,340$  inhab km<sup>-2</sup>), Sant Vicenç dels Horts (SVH-2,  $d_{SVH} = 2,409$  inhab km<sup>-2</sup>), Rubí (RUB-3,  $d_{RUB} = 644$  inhab km<sup>-2</sup>) and Terrassa (TRS-E,  $d_{TRS} = 1,864$  inhab km<sup>-2</sup>). The last two BHAs belong to urban areas away from the city of Barcelona: a district of the town of Tarragona, located by the coast (TRG-2,  $d_{TRG} = 1,573$  inhab km<sup>-2</sup>), and a district of the town of Lleida, located in the interior of Catalonia (LLEI-2,  $d_{LLEI} = 1,123$  inhab km<sup>-2</sup>).

## 2.3 Weather data

*In situ* temperature, relative humidity, surface pressure, solar radiation and precipitation data were obtained from a network of 187 automatic weather stations spread along Catalonia and managed by the Meteorological Service of Catalonia (MSC), which is a public institution dependent on the Government of Catalonia (see Data Availability section). For each atmospheric variable, the original 30-min data available since 2009 were averaged to daily values. This allowed

estimating the maximum and minimum daily temperatures, the daily thermal amplitude and the difference in mean temperature between consecutive days. The final time series were first smoothed with a 3-day running-average filter and then used to obtain the time series at each BHA by spatial interpolation in the region. In summary, the nine atmospheric variables chosen for assessing the weather impact on the COVID-19 propagation are daily mean values of mean temperature ( $T_{mean}$ ), relative humidity ( $RH$ ), shortwave solar radiation ( $Rad$ ), and surface pressure ( $P$ ), as well as daily precipitation ( $Prec$ ), daily minimum ( $T_{min}$ ) and maximum temperature ( $T_{max}$ ), daily thermal amplitude ( $DTA$ ), and difference in mean temperature between consecutive days ( $\Delta T$ ).

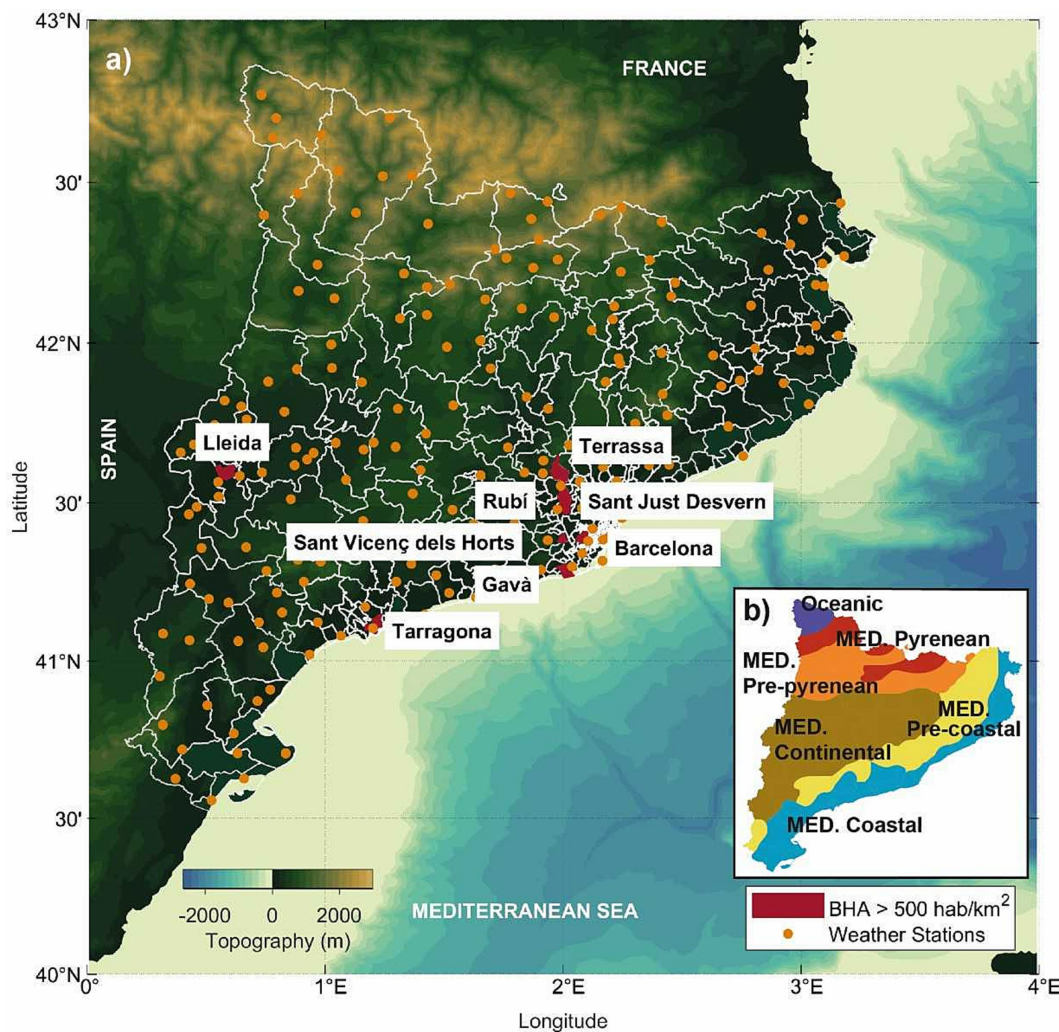
## 2.4 Cross-correlation analysis

In June 21, 2020, following a substantial decrease in the number of infections and deaths by COVID-19 and coinciding with the end of the Spanish school year, the Spanish government opened a period with no restrictions in mobility and distancing that was named 'new-normality'. This allowed a substantial fraction of the Catalan population to spend a few weeks of July–August in holiday destinations. This implied large internal mobility to locations away from their registered residence, disabling a proper normalization of infections in terms of resident population. Hence, we assess the impact of weather on the propagation of COVID-19 immediately after the holiday season, between September 2020 and February 2021. Concretely, an internal validation is done from September 1 to November 18, 2020 – the setup period – which covers a period of relative normality, when most families were back to their homes for work and the start of the academic course, and before the onset of the second COVID-19 wave. Additionally, an external validation is done from November 19, 2020, to February 2, 2021 – the forecast period – which covers a period after the end of the second COVID-19 wave and before a substantial fraction of the population was vaccinated, and includes the third COVID-19 wave.

The relationship between the local weather and health variables is explored through a time-lagged cross-correlation analysis between each of the nine atmospheric variables and the  $I_c$  index. To evaluate the similarity between the two series, the atmospheric variables are shifted backwards in time with respect to  $I_c$ . Following our initial hypothesis, that infection is driven by weather conditions, only negative lags are considered for further analysis; these negative time lags ( $\tau < 0$ ) indicate that changes in the infection index follow the atmospheric variables. According to the average maximum reported COVID-19 incubation days (91) and to the  $I_c$  definition, the maximum time lag selected was  $\tau_{max} = 10$  days. This lead–lag correlation analysis will allow us to determine the lead time of each weather parameter that affects the COVID-19 transmission. In order to quantify the impact of weather on the spread of the virus, we assume that an atmospheric variable affects the virus propagation if the sample correlation coefficient ( $CCF$ ) between this variable and the infection index is significant below a threshold of  $\alpha = 5\%$ .

## 2.5 Selection of climatic variables to build the model

Once the weather precursors and their characteristic lead time are identified, the COVID-19 propagation is modeled using a MLR model



**FIGURE 1**  
 (a) Basic health areas (BHAs, delimited in white) and automatic weather stations (orange dots) in Catalonia. Those BHAs selected for this study, with a population density  $d \geq 500$  inhab  $\text{km}^2$ , are drawn in red. The colour bar shows the topography. (b) The inset shows the bioclimates in Catalonia: Mediterranean coastal, Mediterranean pre-coastal, Mediterranean continental, Mediterranean pre-Pyrenean, following the Mediterranean Pyrenean and Oceanic, Meteorological Service of Catalonia (101).

for each BHA. This model expresses the predicted infection index  $I_{c,pred}$  in terms of  $p$  local climatic predictors [Equation 3]:

$$I_{c,pred}(t; X_1, \dots, X_p) = c_0 + \sum_{j=1}^p c_j \cdot X_j(t + \tau_j^*), \quad (3)$$

where  $t$  is time,  $X_j$  indicates any of the local predictors of the model,  $c_0$  is the constant coefficient for the model,  $c_j$  ( $j \in [1, p]$ ) is the regression coefficient for the  $X_j$  predictor, and  $\tau_j^*$  is the characteristic lag for the  $X_j$  predictor. The characteristic time lag (or lead time) is defined as the time interval for which the highest correlation coefficient between the predictors and the observed  $I_c$  index is obtained. Hence, a total of  $2p + 1$  parameters are obtained for each BHA.

## 2.6 Building the model: model descriptors and statistics

Before building the model, we explore the potential collinearity effects between the predictors as detailed next. First, the correlation

coefficients  $r_{ij}$  between two predictors, namely  $X_i$  and  $X_j$ , are computed. Second, correlation  $t$ -tests are done to evaluate whether the predictors have a significant linear relationship. The  $t$ -statistic  $t_{TS,ij}$  associated with each combination of predictors is calculated as in Equation 4:

$$t_{TS,ij} = \frac{r_{i,j} \sqrt{n-2}}{\sqrt{1-r_{i,j}^2}}, \quad (4)$$

where  $n$  is the size of the sample. The statistics follows a  $t$ -distribution with  $n - 2$  degrees of freedom  $t_{TS,ij} \cong t_{n-2}$ . Finally, if the two predictors are correlated, the degree of collinearity between them is evaluated using the variance inflation factor, defined as in Equation 5:

$$VIF_j = \frac{1}{1-r_j^2}, \quad (5)$$

where the parameter  $r_j$  indicates the coefficient of determination of the variable  $j$  regressed on the remaining predictors. If  $VIF_j$  is less

than 2.5, we consider that collinearity is not significant and both predictors can be used to build the model (19).

After assessing the collinearity between predictors, we test the regression coefficients separately for each BHA in order to select the weather predictors included in the final model. The significance for the regression coefficients is assessed using the *t*-test. Since these tests can be very conservative, we apply the forward stepwise estimation method (20) to decide whether a candidate predictor must be included in the model, as follows. First, we select the predictor with the highest correlation coefficient with the infection index, and fit this predictor and the infection index to a linear regression. As the model is linear, we use the adjusted coefficient of determination,  $r^2_{adj}$ , to evaluate the goodness of the fit. In our case, this coefficient represents the percentage of the variation in the infection index that can be explained by the variation in the predictors, taking into account the size and the number of independent variables in the model. Next, the predictor that has the second highest correlation with the infection index is included in the linear model and  $r^2_{adj}$  is recomputed. The partial *F*-test statistics is used to decide whether the addition of that predictor makes a significant contribution to the model, calculated as in Equation 6:

$$F_{k_2-k_1, n-k_2-1} = \frac{\frac{(SSR_2 - SSR_1)}{(k_2 - k_1)}}{\frac{SSE_2}{(n - k_2 - 1)}}, \quad (6)$$

where the sub-indexes 1 and 2 correspond to the models with the remaining predictor removed or added, respectively. The terms *SSR* and *SSE* indicate the sum of the squares due to regression (i.e., the variability of  $I_c$  explained by the regression) and the sum of the squares due to error (i.e., the variability of  $I_c$  not explained by the regression) for the corresponding models. Finally, *n* represents the size of sample and *k* the amount of predictors used in the corresponding linear regression. The *F*-statistic is inspected at  $\alpha = 10\%$ . If the *F*-value is significant at this level, the model improves significantly with the addition of the new predictor, which is hence maintained in the model. This second step of the forward stepwise (FS) method is repeated for the remaining predictors, which are added one by one from higher to lower correlation. The final model includes all predictors that pass this partial *F*-test.

The FS method has been widely used to select the set of variables to be included in the models because it is easy to implement and reduces the model complexity removing irrelevant or redundant variables (21–23). However, this method has been also questioned to have an inherent bias (24, 25, 95, 96) because it assigns the variables with the highest correlations a higher chance to end up in the final model (26, 27). After considering the potential drawbacks of the method, our model has been assessed by means of a new model constructed using the unbiased two-step method (24). The results obtained from this linear model at each specific BHA are compared to the results of the FS model (see [Supplementary Information](#)) showing similar results.

The final FS model is assessed using the joint *F*-test. This test allows deciding whether the linear regression used in the model provides a better fit to the observations than a model with no predictors (intercept-only model). The test statistic, which is denoted by *F*, has a Fisher distribution that is calculated as in Equation 7:

$$F = \frac{SSR}{SSE} \frac{(n - k - 1)}{k} \quad (7)$$

where *SSR*, *SSE* and *n* are defined in Equation 6 and *k* is the number of predictors in the model. For each BHA, the linear fit to the data allows obtaining the *F*-statistic along with its corresponding *p* value. If the *p* value is higher than a 5% significance level ( $\alpha$ ), we conclude that the final model fits the observations better than the intercept-only model.

Finally, the assumptions in the linear regression model are analyzed through the residuals from the fitted model; statistical tests are implemented to complement the graphical information. First, the Kolmogorov–Smirnov (K-S) test is conducted at  $\alpha = 5\%$  in order to examine if the residuals are normally distributed (28). This allows assessing whether the gap between the empirical and normal (hypothesized) distributions of the residuals is significant at the considered significance level. In addition, the White test for heteroscedasticity (29), with  $\alpha = 5\%$ , is used to assess if the regression errors have a non-constant variance. If the *p* value associated with the test statistic, which follows a Chi-square  $\chi^2$  distribution, is smaller than the significance level of the test, then there is no evidence of heteroscedasticity in the final model.

## 2.7 Model validation during the setup period

The final model is internally validated for the setup period, from September 1 to November 18, 2020, through the implementation of the leave-one-out cross-validation (LOOCV) method. In the LOOCV procedure, the infection index  $I_c$  for the day *i* ( $I_{c,i}$ ) is excluded and the model fitted using the remaining data. Each of these training subsets provides an individual model, which is expected to be slightly different from the original one, and then these models are used to predict the infection index  $I_{c,[i]}$  with the *i*-th case removed from the sample. The prediction error from these models, computed as  $e_{[i]} = I_{c,i} - I_{c,[i]}$ , is a measure of how close the prediction is to the observation when this observation is omitted. The absolute percentage error in each measurement is obtained scaling each prediction error against its corresponding observed value:  $APE_{[i]} = 100|I_{c,i} - I_{c,[i]}|/I_{c,[i]}$ .

The overall performance of the model is estimated using the mean of the absolute percentages previously calculated ( $MAPE_{CV}$ ). Then, the value of that mean ( $MAPE_{CV}$ ) is compared to the mean of the absolute percentages obtained when the model was built (*MAPE*) in order to internally evaluate the consistency of the model. Finally, other statistical parameters are determined from the linear regression fit of the observed values  $I_{c,i}$  to the cross-validated ones  $I_{c,[i]}$ : a significance test assesses the deviations of the slope and the *y*-intercept to the expected values, which are  $\beta_1 = 1$  for the slope and  $\beta_0 = 0$  for the *y*-intercept, respectively; a linear fit allows computing the R-squared of the cross-validation,  $q^2_{CV}$ ; and an *F*-test of overall significance tells whether the predictions explain a significant part of the variance observed in the responses as compared to data obtained from a model with no predictors.

The statistical parameters obtained by the application of the LOOCV method are finally compared to the corresponding ones obtained from the data used to build the model. Table 2 summarizes

TABLE 2 Summary of the expressions used to determine the statistic parameters (MAPE, RMSE, standard R-squared  $r^2$  and its adjusted version  $r^2_{adj}$ ) in the internal and external model validations (left), and the internal cross-validation using the LOOCV method (right).

Model building and External validation	LOOCV (Internal validation)
$MAPE = \frac{1}{n} \sum_{i=1}^n APE_i = \frac{100}{n} \cdot \sum_{i=1}^n \frac{ I_{c,i} - I_{c,pred,i} }{I_{c,i}}$	$MAPE_{CV} = \frac{1}{n} \sum_{i=1}^n APE [i] = \frac{100}{n} \cdot \sum_{i=1}^n \frac{ I_{c,i} - I_{c,[i]} }{I_{c,i}}$
$RMSE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (I_{c,i} - I_{c,pred,i})^2}{n}}$	$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^n e_{[i]}^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (I_{c,i} - I_{c,[i]})^2}{n}}$
$r^2 = 1 - \frac{\sum_{i=1}^n (I_{c,i} - I_{c,pred,i})^2}{\sum_{i=1}^n (I_{c,i} - I_c)^2}$ $r^2_{adj} = 1 - \left(1 - r^2\right) \frac{n-1}{n-k-1}$	$q^2_{CV} = 1 - \frac{\sum_{i=1}^n (I_{c,i} - I_{c,[i]})^2}{\sum_{i=1}^n (I_{c,i} - I_c)^2}$

the expressions of the statistical parameters used for construction and validation of the model (30).

### 2.8 Model validation during the forecast period

The final model is generalized using independent meteorological and health datasets for the period from the end of the setup period until February 2021. This forecast period, which includes the pandemic’s third wave in Catalonia, allows testing the predictive skill of the model in each of the BHAs. Mobility and social restrictions in Catalonia were reduced at the end of November 2020, followed by an increase of positive cases (third wave) that led to some restrictions on mobility (including curfew at night) and restaurants. The number of positive cases peaked about January 15, 2021, followed by a decrease in the number of infections, in February falling to those levels previous to the onset of the third wave. On the other hand, vaccination in Catalonia began on December 27, 2020. The percentage of people who had received their first vaccine remained low (< 10%) until February, and increased to about 30% by the end of April 2021. Despite this minor fraction of the vaccinated population and the reset of some mobility limitations, we have considered the data from November 19, 2020, to February 2, 2021, to validate the final model for all eight BHAs.

The validation for this forecast period is done following the same procedure as described for the internal validation. First, the values of the infection index  $I_{c,i}$  during this time period are calculated from the health data time series. Second, the predicted values  $I_{c,pred,i}$  are estimated from the multiple linear regression model. Third, the predicted and observed values are compared and the statistical significance of the linear fit, the slope and the intercept are assessed. Finally, the prediction errors for each value of  $I_{c,i}$  are computed as the differences between observed and predicted values,  $e_i = I_{c,i} - I_{c,pred,i}$ , normalized by each observation  $I_{c,i}$  in order to obtain the mean absolute percentage error of the external validation (MAPE). Additionally, the same statistical parameters estimated for the internal validation are calculated to assess the overall performance of the final model with this new dataset, including  $r^2_{ext}$ ,  $r^2_{ext,adj}$  and RMSE (Table 2).

## 3 Results

### 3.1 Seasonality of COVID-19 in Catalonia

The first and second COVID-19 outbreaks in Catalonia took place, respectively, in March–April and October 2020. Both outbreaks have been assessed through the number of positive PCR cases, normalized in terms of population and area, and the infection  $I_c$  index (see Methods) (Figure 2). Values of  $I_c$  higher than 1 indicate that the number of positive cases at any day were higher than the summation of all positive cases during the previous 10 days, characterizing situations of high transmissibility for small number of infections. Note that the range of normalized PCR cases varied substantially between locations, from maximum values exceeding 3.5 per  $10^5$  inhab  $km^2$  for the most populated values (Figures 2a,d) and decreasing to peak values about 0.05 cases per  $10^5$  inhab  $km^2$  in the least populated BHAs (Figure 2h). However, the peak infection index ranged between about 1 and 6, with larger and more intermittent peaks in the least populated BHAs.

The Spanish first state of alarm lasted between March 14 and June 21, 2020, with strict social interaction and mobility restrictions. After this last date and until the end of October, these measures were similar to the pre-pandemic period, leveraged by non-pharmaceutical interventions (NPI) like minor mobility restrictions and compulsory face masks. The Spanish holiday season takes place in July and August, characterized by major mobility from high- to low-density rural and coastal areas, a fact that affects substantially the demography and hardens any normalization effort. For this reason, for our analyses we have focused on two post-holiday periods, as discussed below.

During the first pandemic wave, the BHA in Barcelona (BCN-10A) reached the highest number of confirmed cases with 3.9 cases per  $10^5$  inhab  $km^2$  followed by the BHAs of Sant Vicenç dels Horts (SVH-2) and Sant Just Desvern (SJD) with 3.7 and 1.9 cases per  $10^5$  inhab  $km^2$ , respectively (Figures 2a,c,d). These maxima occurred before the Spanish government banned all non-essential activities on March 14. Following this first lockdown and coinciding with the lowest percentage of mobility registered during 2020 (31), the number of confirmed cases sharply decreased in all BHAs. During the first half of April the number of positive cases diminished progressively for Gava (GVA-2), Rubi (RUB-3), Terrassa (TRS-E) and

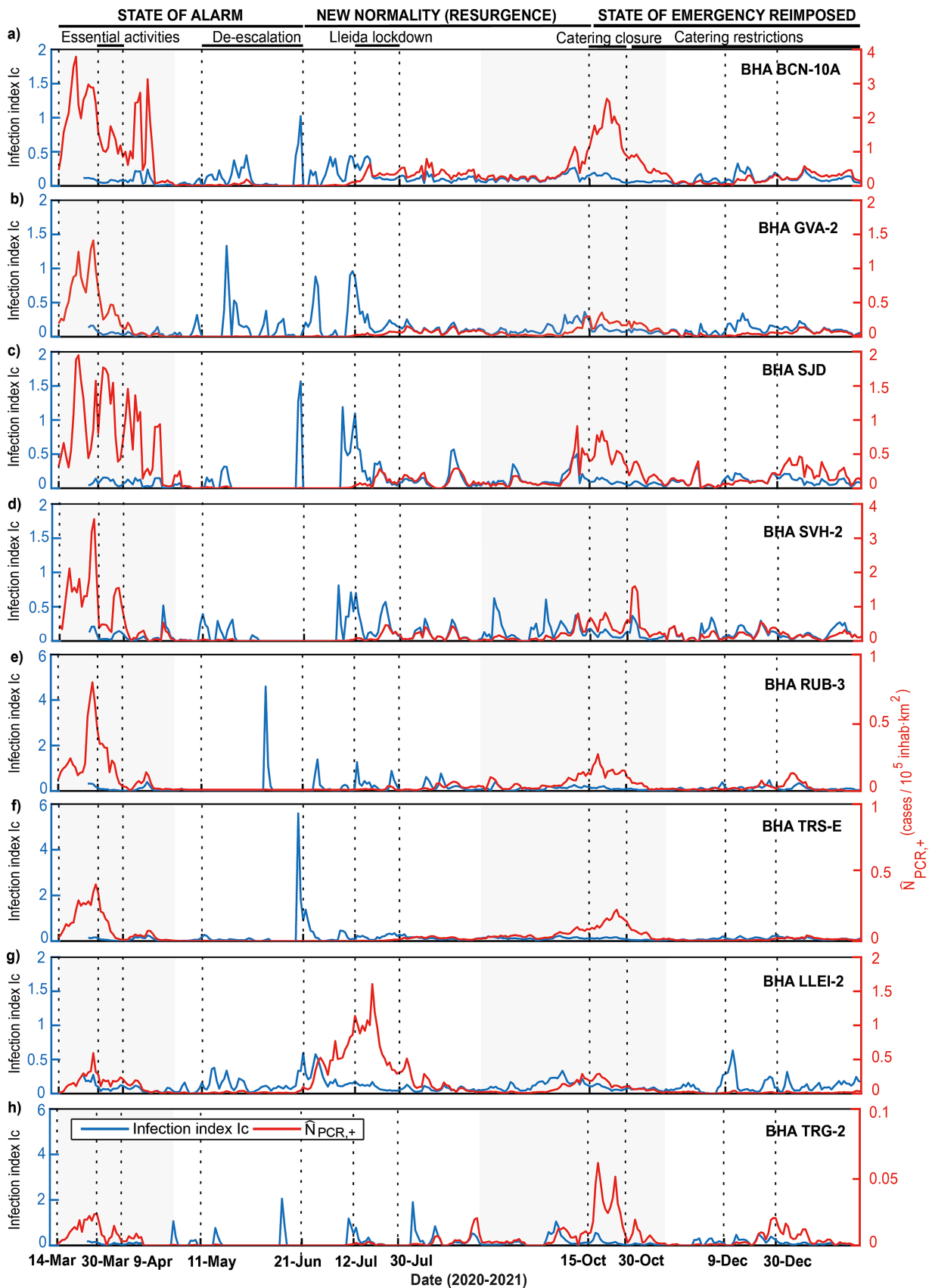


FIGURE 2 (a-h) Temporal evolution during 2020 and beginning of 2021 of the number of COVID-19 cases in the eight selected Catalan BHAs, presented both in terms of the normalized number of cases  $\hat{N}_{PCR,+}$  as determined through the PCR positive tests (red line) and the rate of infections  $I_c$  (blue line).

(Continued)



FIGURE 2 (Continued)

In each panel, the gray shaded areas indicate periods that include the first and the second pandemic waves and the vertical dotted lines delineate the duration of the main social and mobility restrictions imposed in Catalonia.

Lleida (LLEI-2) (Figures 2b,e,f,g), whereas for the remaining BHAs (with the highest density population) there were still several intermittent important peaks (e.g., BCN-10-A and SJD, Figures 2a,c). In the second half of April the number of cases reduced drastically, flattening the curve for all BHAs and leading to a gradual leverage of social restrictions and an increase in mobility. In June 2020, when most of the mobility restrictions had stopped, the number of positive PCR remained low, not exceeding 0.5 cases per  $10^5$  inhab km<sup>2</sup>. The single exception was LLEI-2, which reached 1.5 cases per  $10^5$  inhab km<sup>2</sup> (Figure 2g). In this particular case, the enhancement in virus transmission was associated with seasonal agricultural workers living in overcrowded conditions, which acted as reservoirs and further spreaders of the infection (32).

During this first wave the  $I_c$  index remained low, indicating that the government measures were effective to prevent contagion. In contrast, several high  $I_c$  peaks appeared intermittently during summer (Figure 2) without any major response on the standardized PCR, with the exception of LLEI-2 (Figure 2g). This suggests that, despite the existence of several irregular infection episodes, the initial low numbers of infected people and the intermittency of these events did not allow the number of infected people to grow. It could be argued that the weather conditions were neither favorable to spread the infection but, because of the extremely high mobility during this period, this is very difficult to assess.

During the second pandemic wave, the highly-populated Barcelona metropolitan area (BCN-10A, SJD and SVH-2) showed again the highest values, respectively with 1.7 and 2.7 cases per  $10^5$  inhab km<sup>2</sup>, although these values were lower than during the first wave of the pandemic (Figures 2a,d). In contrast, in the coastal town of Tarragona (TRG-2) values reached 0.07 per  $10^5$  inhab km<sup>2</sup>, higher than during the first wave (Figure 2h). The normalized number of positive PCR cases behaved similarly in all BHAs, rising in the second half of September and peaking in late October. Throughout summer, the  $I_c$  index remained intermittent and relatively high in all BHAs except Terrassa (TRS-E) and Rubí (RUB-3). The mobility restrictions remained low until the end of October, suggesting that the increase in COVID-19 transmission could have been influenced by weather conditions (33, 34). Indeed, during this time period, several cold fronts circulated from west to east in a row, a typical autumn scenario (Supplementary Figures S1, S2).

### 3.2 Correlation between weather variables and the infection index

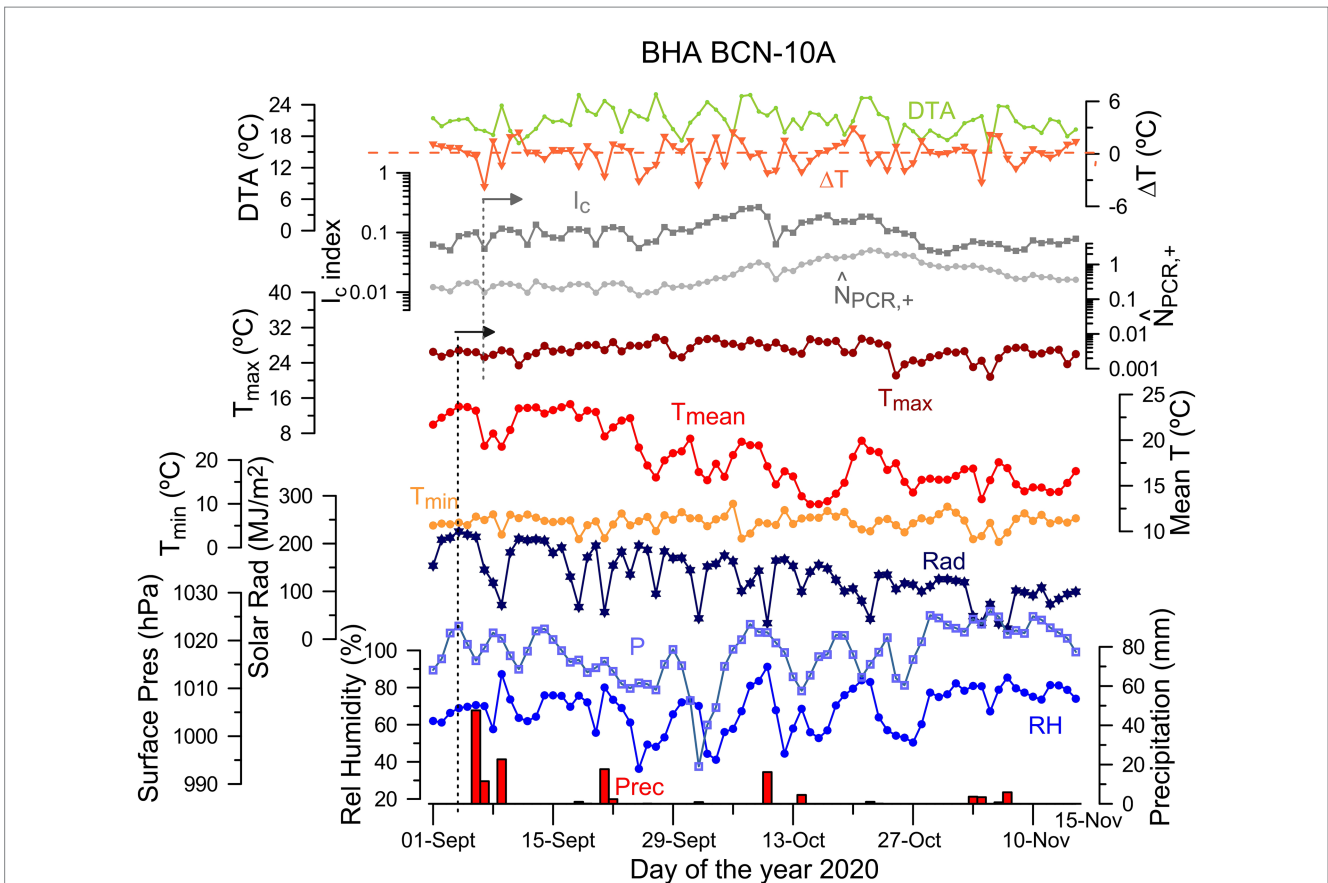
In order to explore the role of weather conditions on the second COVID-19 wave in our study region, we analyze the time-lagged correlations between local infection indicators and weather data for each of the eight BHAs. A total of nine daily-averaged atmospheric variables for each BHA are used during the period from September 1 to November 18, 2020 (see Methods). The initial selection of humidity and temperature is based on previous research on SARS-CoV-2 and

other respiratory viruses such as influenza, which explored the impact of seasonal variations of these variables on virus survival in the environment or on host susceptibility (35, 36, 97). Additionally, we include daily mean values of solar radiation, and surface pressure, as well as daily precipitation, daily minimum and maximum temperature, daily thermal amplitude and the difference in mean temperature between consecutive days.

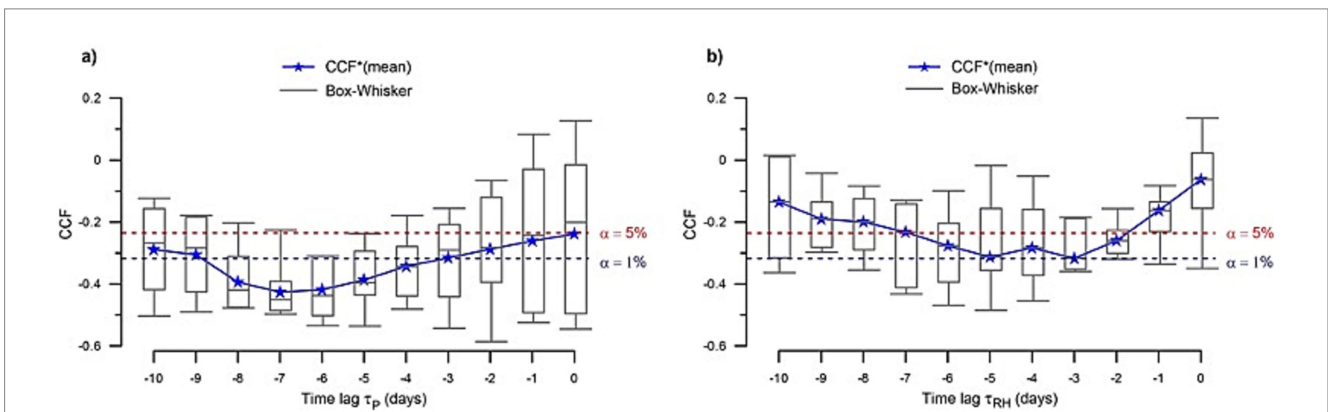
The evolution of the normalized number of cases and the infection index is compared with changes in the atmospheric variables, as shown in Figure 3 for BCN-10A (see Supplementary Figure S2, for all BHAs). Oscillations in surface pressure, temperature, relative humidity and solar radiation are associated with the passage of cold fronts in this area (e.g., the September 7–10 variation in Figure 3). The time series also reveals changes in the infection index, which seem to show up several days after the atmospheric changes (gray dashed line in Figure 3). We hence explore the correlation of  $I_c$  with the entire selected set of atmospheric variables in order to determine their possible influence on the spread of the virus.

Our results show consistent significant negative correlation of surface pressure ( $P$ ) and relative humidity ( $RH$ ) with the lagged infection index  $I_c$  for all BHAs. A negative correlation indicates that a decrease in  $P$  and/or  $RH$  enhances the spread of the virus several days later. For  $P - I_c$ , the correlation coefficients are statistically significant ( $p < 0.01$ ) in all BHAs for specific time lags ( $\tau_p$ ). The cross-correlation  $P - I_c$  coefficients are high for all BHAs ( $CCF \cong -0.5$ ), with a time lag fairly constant at about  $-7$  days, although some areas have even higher coefficients at shorter lags (Supplementary Figure S3). For  $RH - I_c$ , the cross-correlation coefficients are statistically significant with  $p < 0.01$  in four of the BHAs (GVA-2, SJD, SVH-2 and TRS-E) and at  $p < 0.05$  in the remaining four BHAs. In this case, the highest values for the correlation coefficients ( $CCF \cong -0.5$ ) correspond to  $\tau_{RH}$  between  $-5$  to  $-3$  days (Supplementary Figure S3). It means that an increase of the infection index in Catalonia is preceded by low surface pressure (7 days before) and dry conditions (3–5 days before). The other meteorological variables evaluated do not show consistent correlation scores (Supplementary Figure S3). In particular, the variables derived from temperature and precipitation are poorly correlated with the infection index, except for LLEI-2 and TRG-2 that have significant correlations at  $\tau = -10$  days. The daily thermal amplitude is significant ( $p < 0.05$ ) in five out of the eight BHAs, but the cross-correlation function for these BHAs has variable time lags. A similar situation appears in the case of the shortwave solar radiation (Supplementary Figure S3). These results suggest that daily thermal amplitude and shortwave radiation can contribute additionally in the COVID-19 transmission in specific locations. Hereafter, we focus on the common meteorological patterns that affect and can be used as predictors for our entire area of study.

We use all eight BHAs to produce a composite box-plot of  $CCF$  as a function of time lag, for either  $P$  and  $RH$  with respect to  $I_c$  (Figure 4). For each lag, we select the mean of the  $CCF$  values for each box plot (hereinafter  $CCF^*$ ) as the representative value of the set. For both variables,  $CCF^*$  shows a well-defined valley where the negative



**FIGURE 3**  
Time series of daily values of atmospheric variables during the period that includes the second outbreak (September 1 to November 18, 2020) together with the normalized number of cases ( $\hat{N}_{PCR,+}$ ) and the infection index ( $I_c$ ) for BCN-10A. The atmospheric daily variables are mean temperature ( $T_{mean}$ ), relative humidity ( $RH$ ), solar radiation ( $Rad$ ), precipitation ( $Prec$ ), surface pressure ( $P$ ), minimum and maximum temperature ( $T_{min}$ ,  $T_{max}$ ), daily thermal amplitude ( $DTA$ ) as well as the difference in mean temperature between consecutive days ( $\Delta T$ ). The units for these variables are indicated in their corresponding axes. The vertical black and gray dashed lines indicate the arrival of a cold front on September 7 and the changes observed in the health series a few days later.



**FIGURE 4**  
Composite box plots of the cross-correlation coefficients ( $CCF$ ) of (a) surface pressure and (b) relative humidity with respect to  $I_c$  as a function of lag time, calculated using the eight reference BHAs. The lower and upper ends of the box represent the first and third quartiles, respectively, and the mean ( $CCF^*$ ) is indicated by a blue star. The whiskers extend to the most extreme value within 1.5 IQR (interquartile range) from the box ends. Horizontal dashed lines indicate the statistical significance of the coefficients at  $\alpha = 5\%$  (red line) and  $\alpha = 1\%$  (blue line).

correlation is highest (Figure 4). Surface pressure has the largest absolute correlation  $CCF^* = -0.42$  at  $\tau_{p,min} = -7$  days (with standard deviation  $std(CCF) = 0.08$ ) and the relative humidity has minima at

both  $\tau_{RH,min} = -3$  ( $CCF^* = -0.28$ ,  $std(CCF) = 0.08$ ) days and  $-5$  days ( $CCF^* = -0.27$ ,  $std(CCF) = 0.14$ ). For both variables, the  $\tau_{min}$  values occur at  $CCF^*$  significant levels.

For surface pressure,  $CCF^*$  is actually significant at  $\alpha = 10\%$  for  $\tau_P \in [-8, -4]$ ; in particular, the smallest interquartile ranges (IRQ) for the  $CCF$  distributions are in this lag interval ( $IRQ \cong 0.1$ ), indicative of a minimum dispersion of the  $CCF$  values among the different BHA (Figure 4a). The smallest dispersion is found at  $\tau_{P,min} = -7$  days, with  $IRQ(\tau_{P,min}) \cong 0.05$  and  $std(CCF) = 0.08$ . We conclude that most of the inspected BHAs show the highest correlations between  $I_c$  and  $P$  at  $\tau_{P,min} = -7$  days.

In the case of the relative humidity, we observe a similar behavior, with  $CCF^*$  values statistically significant in a range of  $\tau_{RH} \in [-6, -2]$  days (Figure 4b). In this interval, the interquartile range (IQR) for the  $CCF$  coefficients is about 0.2, and the smallest value takes place at  $\tau_{RH} \cong -3$  days, with  $IRQ(\tau_{RH,min}) \cong 0.1$  ( $std(CCF) = 0.08$ ).

These results indicate that the infection index was negatively correlated with the surface pressure conditions 7 days before and with the relative humidity conditions about 3 days before. We conclude that a decrease (increase) in  $P$  or  $RH$  leads to an increase (decrease) in  $I_c$  and vice versa some 7 and 3 days later, respectively. These values for the time lags are chosen as the characteristic lags ( $\tau^*$ ) for the surface pressure and the relative humidity in the study area.

### 3.3 Surface pressure and relative humidity as predictors of COVID-19 variability

The cross-correlation analysis reveals that surface pressure and relative humidity are the only meteorological variables that bear statistical significant correlation with  $I_c$  ( $p < 0.05$ ) in all BHAs. Therefore, the model is built using only two candidate predictors ( $P$  and  $RH$ ), hence requiring five parameters for each BHA (Equation 3). However, the results of the composite plots of the  $CCF$  (see section 3.2) evidence similar time lags for each predictor over the entire region. The substantial variations in the coefficients suggest that the intensity of the response is region-dependent, likely reflecting specific demographic and geographic characteristics (Figure 1), but the similarity of the time lags advocates for the existence of analogous responses to environmental forcing. Hence, despite the infection indexes differ substantially between different BHAs, we propose a multiple linear regression model of the infection index (Equation 3) in terms of these two variables, with changing coefficients but one same time lag for all BHAs.

This approach reduces the total number of free parameters for each BHA to only three, so the model is expressed as in Equation 8:

$$I_{c,pred}(t; P, RH) = c_0 + c_1 \cdot P(t + \tau_P^*) + c_2 \cdot RH(t + \tau_{RH}^*) \quad (8)$$

where  $c_i$  ( $\forall i \in [0,2]$ ) are the model intercept and regression coefficients for each BHA, the predictor 1 corresponds to the surface pressure  $P$  (measured in hPa) and the predictor 2 corresponds to relative humidity  $RH$  (measured as a percentage of absolute humidity relative to the maximum saturation value for that temperature). The temporal variable ( $t \geq 0$ ) is the day counter for the selected time period and  $\tau_j^*$  is the characteristic time lag for the  $j$  predictor. These lags between the weather and health variables indicate the leading times of the atmospheric parameters in the COVID-19 transmission. In practice, it means that the atmospheric time series cover the period

from September 1 to November 15, 2020, while the health time series refer to the period September 8 to November 18, 2020.

The setup and application of the climate-dependent COVID-19 model is explained in detail in Methods. Briefly, the model is first developed using the  $P$ ,  $RH$  and  $I_c$  time series for a period that includes the second outbreak (September 1 to November 18, 2020, the setup period), obtaining the time lags and regression coefficients for each BHA. Since the FS method may possibly introduce some bias in the regression coefficients (see section 2.6 in Methods), the results of our model have been compared with those from a simpler unbiased method, where all weather variables are simultaneously assessed. The comparison between the two models shows that the bias introduced in our FS model is negligible.

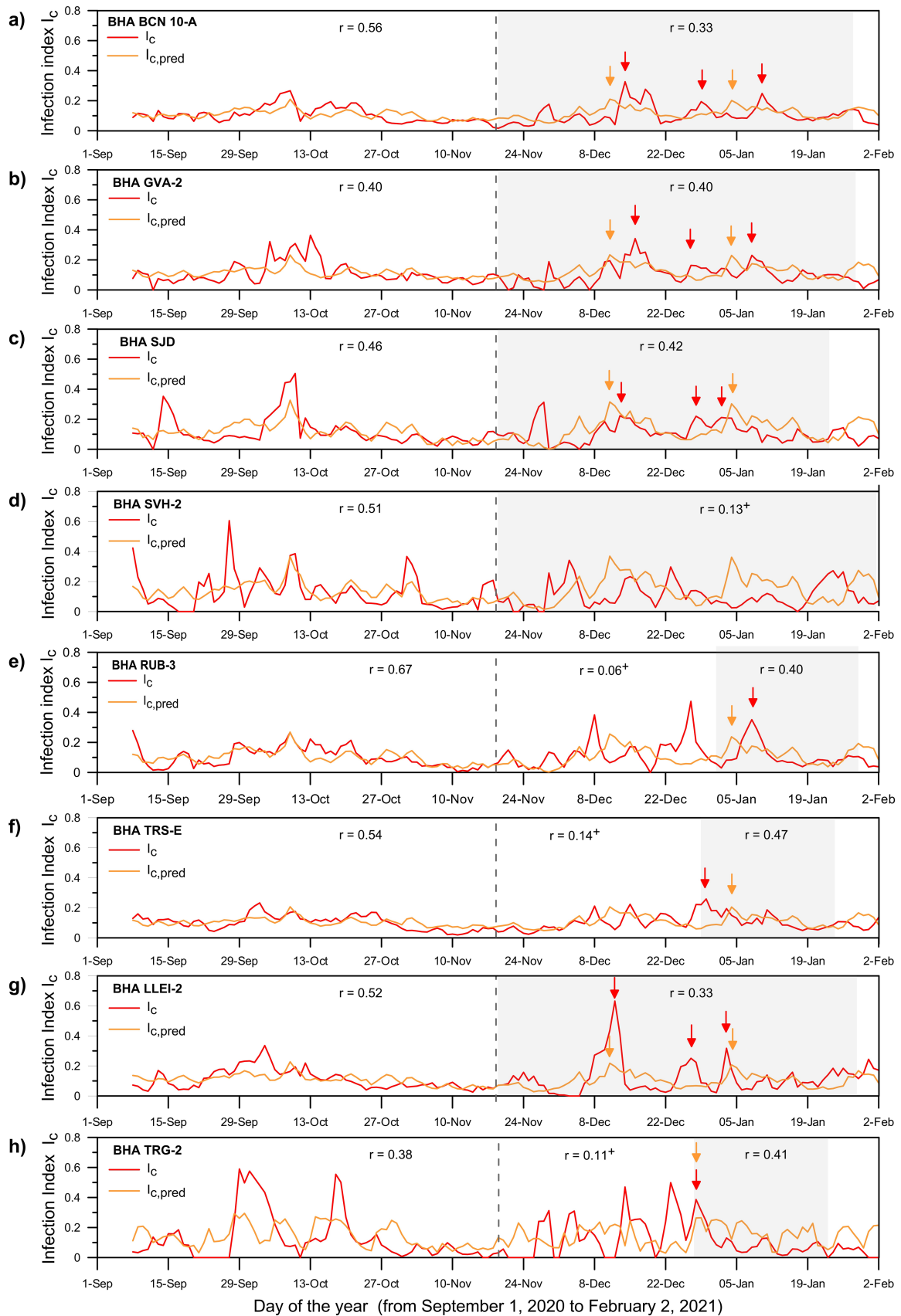
After this validation, the FS model is used to forecast the infection index  $I_{c,pred}$  for an independent dataset (November 19, 2020, to February 2, 2021, the forecast period). The validations for both the setup and forecast periods are conducted through the leave-one-out cross-validation (LOOCV) method (see External validation of the model in the Supplementary Information). Note that after February 2021, over 10% of the Catalan population had already received their first vaccine (37), undermining the use of more recent data for external validation. The results and statistics of the model validation are extensively explained in the Supplementary Information.

Fitting the model infection index  $I_{c,pred}$  predictions to the observations for the setup period (see Model parameters and statistics in the Supplementary Information) shows that the two predictors,  $P$  and  $RH$ , have a significant contribution to the model ( $p < 0.1$ ) in four of the eight BHAs (BCN-10A, SJD, RUB-3 and LLEI-2). For the other four stations, it turns out that one single predictor,  $P$  or  $RH$ , is enough to characterize the evolution of  $I_c$ . Specifically, the model does not significantly improve by adding  $P$  in TRG-2 or by adding  $RH$  in GVA-2, SVH-2 or TRS-E. The regression coefficients,  $c_1$  and  $c_2$ , are significant ( $p < 0.05$ ) in all BHAs and their values vary between  $[-10, -3] \times 10^{-3}$  hPa<sup>-1</sup> ( $mean(c_1) = -5.90 \times 10^{-3}$  hPa<sup>-1</sup>;  $std(c_1) = 1.90 \times 10^{-3}$  hPa<sup>-1</sup>) and  $[-4, -1] \times 10^{-3}$  ( $mean(c_2) = -1.91 \times 10^{-3}$ ;  $std(c_2) = 1.12 \times 10^{-3}$ ). The negative regression coefficients indicate that a decrease (increase) in the infection index occurs when  $P$  and  $RH$  increases (decreases) several days before, confirming the results of the correlation analysis (Figures 4a,b).

During the pandemic second outbreak or setup period, as expected, the model captures moderately well the general behavior of  $I_c$ . This is confirmed by the significant correlation coefficients ( $r$ ), derived from the R-squared ( $r^2$ ) of the corresponding linear regressions between the  $I_{c,pred}$  and  $I_c$  (Supplementary Figure S5 and Supplementary Table S19). In particular, RUB-3, TRS-E or BCN-10A exhibit the highest correlations, with  $r = 0.67$ , 0.54 and 0.56, respectively; SVH-2 and LLEI-2 show correlations higher than 0.5, and the remaining three BHAs show lower yet statistically significant correlations ( $0.3 < r < 0.5$ ;  $p < 0.01$ ). We conclude that our simple climate-dependent model reproduces several main changes in the infection index during the second COVID-19 outbreak in Catalonia (fall 2020).

### 3.4 Forecasting the infection index during the third wave

The pandemic third outbreak started with an increase of the infection rate in early December 2020 but declined to the pre-outbreak levels in January 2021 (red lines, Figure 5). The mobility



**FIGURE 5**  
 Results for the eight BHAs (panels a through h) showing the temporal evolution of the observed  $I_c$  (red lines) and predicted  $I_{c,pred}$  (orange lines) infection index for the period between September 1, 2020, and February 2, 2021. The vertical dashed line on November 18, 2020, separates the setup and

(Continued)

FIGURE 5 (Continued)

forecast periods. The gray shaded areas indicate selected portions of the forecast interval with the highest correlation coefficient between the two series. The correlation coefficients for each time interval are displayed in the upper part of each panel. The symbol “\*” indicates time intervals when the two series are not statistically significant at a maximum  $\alpha$  of 1%; in several cases it corresponds to time periods of low normalized number of cases (Supplementary Figure S6). The arrows indicate the observed (red) and predicted (orange)  $I_c$  peaks during the forecast periods.

and social measures progressively relaxed after November 2020 and the state-of-alarm was revoked in early May 2021. Social interactions increased substantially during Spanish public holidays in early December and Christmas time, leading to a third wave that had specific features for each BHA. For example, in BCN-10A, GVA-2 and LLEI-2, the infection rate shows a marked peak in mid-December followed by two secondary peaks in late December and early January (Figures 5a,b,g). However, this is not the case for SJD, which had peaks of comparable amplitude (Figure 5c), or for TRS-E, where the situation was reversed and the two secondary peaks took place in December while the main peak occurred during the New Year's Eve (Figure 5f).

We apply our climate-dependent multiple-regression model to the forecast period (November 19, 2020, to February 2, 2021), in order to predict the infection index for each health area ( $I_{c, \text{pred}}$ ) during the pandemic's third outbreak (Figure 5). Despite differences in amplitude, most BHAs illustrate three peaks in  $I_c$  during mid- and end-December 2020 and in January 2021 (red lines, Figure 5). Outstandingly, the predictions reproduce the basic features in the observed index ( $I_c$ ) for most BHAs (light orange lines, Figure 5). In particular, the model reproduces the increase of  $I_c$  during mid-December and January in LLEI-2 ( $r=0.33$ ), SJD ( $r=0.42$ ), GVA-2 ( $r=0.40$ ) and BCN-10A ( $r=0.33$ ) (Figures 5a–c,g). Notice that for GVA-2 and BCN-10A the climate-based model simulates the enhancement of the infection rate 5–7 days earlier than in the observations (Figures 5a,b). However, the model generally fails to reproduce the  $I_c$  peak at the end of December. This suggests that this peak may have been largely driven by the relaxation of social restrictions during the Christmas festivities. For the remaining BHAs (TRG-2, TRS-E and RUB-3), the correlation remains significant (0.40 to 0.47) when only the second half of the period (December 25 to January 20) is considered (Figures 5e,f,h). However, the model fails completely to forecast  $I_c$  in SVH-2 for the entire period, showing even an out-of-phase behavior between the observed and predicted values (Figure 5d). The same behavior is also recognizable when the observations in SVH-2 are compared to the observed values for the other areas (e.g., BCN-10A, GVA-2 or SJD).

The relative success of the climate model is remarkable if we consider that the forecast included periods of limited mobility and social interactions, with temporary relaxation during the holiday seasons. Finally, it is important to note that when the normalized number of cases is low then the observed and predicted infection indexes show no significant correlation ( $p \geq 0.01$ ), as it also happens in several BHAs between November 19 and the end of the year (Supplementary Figure S6).

## 4 Discussion

We have explored the spread of SARS-CoV-2 infection in Catalonia (northwestern Mediterranean) for periods when there were limited mobility and social restrictions and there was not yet a vaccine available.

In this way we have minimized the effects associated with decreased SARS-CoV-2 transmissivity related to acquired immunization and lockdown-mobility restrictions in areas with high-density population (38). Our results show significant correlations of both surface pressure ( $P$ ) and relative humidity ( $RH$ ) with the daily infection index ( $I_c$ ), at  $\alpha=1\%$  and  $\alpha=5\%$ , respectively, in all basic health areas (BHAs); solar radiation ( $Rad$ ) and daily thermal amplitude ( $DTA$ ) are also significantly correlated with  $I_c$  ( $p < 0.05$ ) but only in some BHAs while consistent correlation does not exist between the other meteorological variables evaluated and  $I_c$ . This allows building a simple multiple linear regression model for the infection index with only two predictors,  $P$  and  $RH$ ; in this model, the regression coefficients for  $P$  and  $RH$  are negative for all BHAs studied, indicating that a decrease of  $P$  and  $RH$  causes an increase of the  $I_c$  index after several days. The regression model has predictive capacity with a significance level below 5% for the setup period. Remarkably, the cross-correlation of  $I_c$  and either  $P$  or  $RH$  provides consistent and significant results, strongly suggesting that  $P$  and  $RH$  lead the COVID-19 outbreaks by 7 days and 3 days lag, respectively.

Outstandingly, our climate-dependent model shows a moderate skill to forecast the SARS-CoV-2 transmission during a later period despite real limitations, such as the accuracy of the daily counts of COVID-19, and potential restraints, such as the interrelation between weather variables. The predictive skill of our model is significant at a 90% level in four out of the eight BHAs (BCN-10A, GVA-2, SJD and LLEI-2), and remains significant during shorter periods for three of the other areas (RUB-3, TRS-E and TRG-2; correlation scores between 0.33 and 0.47). This means that weather conditions are able to explain between 14 and 45% of the variability of the infection index  $I_c$  in the selected eight BHAs of Catalonia during the setup period (September 1 to November 15), and between 11 and 22% during the predictive period. The reduction in this latter period is likely related to the changing social interaction and mobility measures, much greater than during the setup period.

Our findings are also consistent with the results of other studies, in particular those conducted in the Iberian Peninsula. Fernández-Ahúja and Martínez (39) found that surface pressure and infection cases are negatively correlated at a regional scale. Sanchez-Lorenzo et al. (40) suggested that dry conditions favor the spread of the virus during the initial stage of the pandemic. No association between infection cases and mean daily temperature was found in another study in Catalonia (90). Our results corroborate these findings on a smaller spatial scale during other pandemic waves and using a different approach, i.e., a MLR analysis with time lags.

The mechanisms for weather-mediated changes in respiratory disease include the effects of weather on virus survival in the air and surfaces of outdoors/indoors spaces, changes at the individual-level susceptibility toward the disease, and also variations in social human behavior. The 7-days lag between a decrease in surface pressure and the onset of a peak of infection agrees with the incubation period of the SARS-CoV-2 variants circulating during the study period [mean of 5.7 days and a range between 2 and 14 days; (41)]. The predictor capacity of surface level pressure on the virus expansion may possibly

arise from both direct and indirect causes. A pressure change is the main indicator for the passage of low- and high-pressure frontal systems that bring substantial changes in weather, such as temperature, precipitation or wind velocity. Further, rapid changes in weather conditions may affect the susceptibility to airborne virus infection with disruption of local mucosal immunity. An indirect effect may be the weather-related changes in human behavior, with the most evident response under bad weather conditions being to remain indoors, where the virus can persist longer. In enclosed spaces with inadequate ventilation, small infected droplets and particles can remain suspended from minutes to hours (42, 43). Furthermore, the chances of close contacts increase hence leading to an enhanced virus spreading capacity (44–46). Indeed, it is widely accepted that most of the infections occur in indoor spaces, frequently as super-spreading events (47–49). A scientific report led by the WHO-China commission concluded that 78–85% of transmissions occurred within household settings during the first wave, indicating that transmission is likely to occur during close and prolonged contact scenarios (92). Further, the outdoor chains of transmission, which are much more influenced by direct weather conditions, also define the chances that infected individuals go indoors and provoke super-spreading events (50).

The 3-days lag for relative humidity is more difficult to justify, even if it still lies between the estimated incubation bounds. The linear model predicts that dry conditions will favor the propagation of the virus (an increase in  $I_t$ ) about 3 days later and wet conditions will tend to inhibit it. Low humidity, indicative of dry weather, has been identified as a key factor associated with the transmission and stability of respiratory viruses such as influenza (51). Dry weather conditions enhance the susceptibility and severity of influenza infection through disruption of local mucosal immunity of the respiratory tract (35, 51–53); additionally, they increase the stability of the SARS-CoV-2 in the environment (54, 55). Accordingly, some studies support an inverse relationship between humidity and the spreading of SARS-CoV-2, consistent with our findings (56, 57). However, a positive effect of relative humidity toward SARS-CoV-2 infectiousness has also been found in other studies (58–63); in particular, a recent study in England and Wales has found that coronavirus have a different pattern of weather susceptibility as compared with the influenza virus, with an increase of transmission (above 80%) during periods of high relative humidity, which behaves as a better predictor than specific or absolute humidity (64).

The disparity of findings related to humidity could be explained by the convex relationship between virus stability and humidity, which is highest under both high and low  $RH$  (54). It also suggests that there may be geographically-dependent factors that modulate local humidity conditions and the virus response (51). Several studies suggest that dry conditions may have favored the fast spread of the virus in Spain at the onset of the pandemic (40, 65), although there is no full consensus (66). A recent review underlines the significant association between geographic factors (e.g., location, demography) and SARS-CoV-2 spreading at regional or local levels, with the potential interplay of meteorological factors (67). Future studies should consider the succession of weather episodes that can foster or inhibit virus expansion and their interaction between outdoors and indoors conditions (51), e.g., episodes of dry weather that favor environmental outdoors stability of the virus followed by rainy episodes that increase the chances of indoors transmission.

Our results show that surface pressure is the major weather driver of the SARS-CoV-2 spread, and indeed this is probably the main atmospheric indicator for the arrival of frontal systems. Depending on latitude and location – e.g., west versus east coasts of continents – these systems will typically arrive from different directions and cross either land (dry) or sea (wet) regions, hence driving a decrease or an increase in humidity. This idea fits with our finding that surface pressure is the main weather parameter influencing the spreading of the virus while relative humidity plays a more secondary role. Something similar could happen with other possibly secondary variables such as sunlight radiation, which we observe to have significant correlations with the infection index in some BHAs and has been related to SARS-CoV-2 and other airborne viruses (68).

According to a recent comprehensive review (69), temperature is the weather variable that bears major relation with the different infection variables indicative of SARS-CoV-2 spread (number of confirmed COVID-19 cases, number of COVID-19 death cases, cumulative incidence, clinically compatible or PCR based diagnostic cases, among others), with warm and wet climates generally (but not always) decreasing COVID-19 incidence. This same review also shows that high humidity is generally associated with higher incidence/prevalence of the infection but with substantial variability. This review also identifies that surface pressure in some cases is correlated to the occurrence of COVID-19, but this correlation appears much less frequent than temperature or humidity (69).

Models based solely on atmospheric variables have failed to predict the incidence of the disease probably due to the presence of other factors beyond the interrelations between different weather variables. These include: an incomplete consideration of relevant social variables, such as the mobility and distancing restrictions imposed during the succeeding SARS-CoV-2 waves; the diverse infection outputs and inconsistencies in the counting system of infected population (70); and the broad range of indexes used to measure the epidemic activity. In our case, the absence of a stagewise analysis could be considered a limitation (71), as different events, such as the start of the school year in mid-September and the Christmas holidays, might have influenced infection dynamics. However, this limitation is only relevant during the Christmas period due to the substantial increase in social interactions. Schools, on the other hand, did not significantly contribute to transmission among children, as the circulating strains at the time had minimal impact on this group (98–100). Moreover, the infection index  $I_t$  used in our study provides a robust estimate of SARS-CoV-2 activity, as it is based on direct, systematic laboratory-based screening at the community level.

It is important to emphasize that the primary aim of our study has been to show that weather conditions, as reflected by a selected number of variables, do have an effect on the spread of COVID-19. The objective was not to obtain good fits but rather to identify atmospheric factors that do affect COVID-19 infections. From this perspective, our results show that surface pressure and relative humidity do have a significant influence on epidemic outbreaks. Nevertheless, the relatively low correlation between these weather variables and the infection index confirms that the lack of population immunity remains as the primary driver, with high susceptibility associated with the absence of naturally acquired immunity, the low effectivity of vaccination against infection, and the limited vaccine coverage (72). The inability of current vaccines and past infections to produce a long-standing immunity, and the definitive leverage of most of the social restrictions by the end of 2022, underscore

the importance of understanding the interplay between weather conditions and the epidemiological dynamics of COVID-19. In this post-pandemic induction phase, weather conditions may substantially affect the onset and extent of new waves and outbreaks as observed in other respiratory viruses. This study illustrates the potential of linear multiple-regression models as a useful tool to identify the right combination of weather variables that preconditions the development of respiratory pathogens and pandemics.

## Data availability statement

The datasets analyzed for this study can be found in the Catalan Transparency Portal databases for each data set: automatic weather stations (XEMA) from the Meteorological Service of Catalonia, <https://analisi.transparenciacatalunya.cat/Medi-Ambient/Dades-meteorologiques-de-la-XEMA/nzvn-apee>, and open database of COVID19 record from the Health Department of Catalonia, <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-a-Catalunya-per-rea-/xuwf-dxjd>. The datasets are also available from the corresponding authors on reasonable request.

## Ethics statement

The studies involving humans were approved by Institut de Recerca Hospital Universitari Germans Trias i Pujol. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because the human samples used in this study were acquired from data distributed publicly by the Catalan Health Agency.

## Author contributions

JP-M: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing, Formal analysis, Visualization, Supervision. JLP: Conceptualization, Investigation, Methodology, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing. MM-R: Conceptualization, Methodology, Writing – review & editing. AOA: Data curation, Writing – review & editing, Visualization. XV: Conceptualization, Investigation, Writing – review & editing, Supervision. JR: Writing – review & editing. CR: Writing – review & editing, Funding acquisition.

## References

- Kirtipal N, Bharadwaj S, Kang SG. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infect Genet Evol.* (2020) 85:104502. doi: 10.1016/j.meegid.2020.104502
- Ayres JG, Forsberg B, Annesi-Maesano I, Dey R, Ebi KL, Helms PJ, et al. Climate change and respiratory disease: European Respiratory Society position statement. *Eur Respir J.* (2009) 34:295–302. doi: 10.1183/09031936.00003409
- Fu S, Wang B, Zhou J, Xu X, Liu J, Ma Y, et al. Meteorological factors, governmental responses and COVID-19: evidence from four European countries. *Environ Res.* (2021) 194:110596. doi: 10.1016/j.envres.2020.110596
- Culqui DR, Díaz J, Blanco A, Lopez JA, Navas MA, Sanchez-Martinez G, et al. Short-term influence of environmental factors and social variables COVID-19 disease in Spain during first wave (Feb–may 2020). *Environ Sci Pollut Res.* (2022) 29:50392–406. doi: 10.1007/s11356-022-19232-9
- Zoran MA, Savastru RS, Savastru DM, Tautan MN, Baschir LA, Tenciu DV. Exploring the linkage between seasonality of environmental factors and COVID-19 waves in Madrid, Spain. *Process Saf Environ Protect.* (2021) 152:583–600. doi: 10.1016/j.psep.2021.06.043
- Ayala-Aldana N, Monleon-Getino A, Canela-Soler J, Radeva P, Rodenas J. A bootstrap method based on linear regression to estimate COVID-19 ecological risk in Catalonia. *World J Adv Res Rev.* (2023) 17:324–32. doi: 10.30574/wjarr.2023.17.1.0047

OE: Writing – review & editing. IV-C: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing, Supervision.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work has been financially supported by the project CLIMATE –COVID19 (Ref. 202030E222) funded by the Spanish National Research Council (CSIC). The work of MM-R has been further supported by Ramón y Cajal (RYC2022-038454-I, funded by MCIN/AEI/10.13039/501100011033 and co-funded by the FSE+, European Union), the DISTROPIA (PID2021-125806NB-I00) and OFF (TED2021-130106B-I00) projects funded by the Spanish Ministry of Science and Innovation. The work of AOA has been supported by the FPU program, funded by the Spanish government (Ministerio de Educacion, Cultura y Deporte, reference FPU17/03796). The authors also recognize the institutional support of the Spanish Government through the Severo Ochoa Center of Excellence accreditation (CEX2019-000928-S).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2024.1430902/full#supplementary-material>

7. Sera F, Armstrong B, Abbott S, Meakin S, O'Reilly K, von Borries R, et al. A cross-sectional analysis of meteorological factors and SARS-CoV-2 transmission in 409 cities across 26 countries. *Nat Commun.* (2021) 12:5968. doi: 10.1038/s41467-021-25914-8
8. Ma Y, Pei S, Shaman J, Dubrow R, Chen K. Role of meteorological factors in the transmission of SARS-CoV-2 in the United States. *Nat Commun.* (2021) 12:3602. doi: 10.1038/s41467-021-23866-7
9. Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. *Proc Natl Acad Sci.* (2020) 117:27456–64. doi: 10.1073/pnas.2008590117
10. Pica N, Bouvier NM. Environmental factors affecting the transmission of respiratory viruses. *Curr Opin Virol.* (2012) 2:90–5. doi: 10.1016/j.coviro.2011.12.003
11. Hao Z, Hu S, Huang J, Hu J, Zhang Z, Li H, et al. Confounding amplifies the effect of environmental factors on COVID-19. *Infect Dis Model.* (2024) 9:1163–74. doi: 10.1016/j.idm.2024.06.005
12. Hassine IH. Covid-19 vaccines and variants of concern: a review. *Rev Med Virol.* (2022) 32:e2313. doi: 10.1002/rmv.2313
13. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature.* (2021) 592:438–43. doi: 10.1038/s41586-021-03402-9
14. López-Muñoz I, Torrella A, Pérez-Quílez O, Castillo-Zuza A, Martró E, et al. Secondary attack rate of SARS-CoV-2 in vaccinated and unvaccinated household contacts during viral replacement from Delta to omicron variant in Spain. 2022. *Emerg Infect Dis.* (2022) 28:1999–2008. doi: 10.3201/eid2810.220494
15. Sciscient BY, Eisele CD, Ho L, King SD, Jain R, Golamari RR. COVID-19 reinfection: the role of natural immunity, vaccines, and variants. *J Community Hosp Intern Med Perspect.* (2021) 11:733–9. doi: 10.1080/20009666.2021.1974665
16. Callaway E. Beyond omicron: what's next for COVID's viral evolution. *Nature.* (2021) 600:204–7. doi: 10.1038/d41586-021-03619-8
17. Fontanet A, Autran B, Lina B, Kieny MP, Karim SSA, Sridhar D. SARS-CoV-2 variants and ending the COVID-19 pandemic. *Lancet.* (2021) 397:952–4. doi: 10.1016/S0140-6736(21)00370-6
18. IPCC. Climate change 2023: synthesis report. A report of the intergovernmental panel on climate change. Contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change. Geneva: IPCC (2023).
19. Johnston R, Jones K, Manley D. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual Quant.* (2018) 52:1957–76. doi: 10.1007/s11135-017-0584-6
20. Nisbet R, Elder J, Miner GD. Handbook of statistical analysis and data mining applications. San Diego, CA: Academic press (2009).
21. Ayvat P, Omeroglu SK. Mortality estimation using APACHE and CT scores with stepwise linear regression method in COVID-19 intensive care unit: a retrospective study. *Clin Imaging.* (2022) 88:4–8. doi: 10.1016/j.clinimag.2022.04.017
22. Cheshmehzangi A, Li Y, Li H, Zhang S, Huang X, Chen X, et al. A hierarchical study for urban statistical indicators on the prevalence of COVID-19 in Chinese city clusters based on multiple linear regression (MLR) and polynomial best subset regression (PBSR) analysis. *Sci Rep.* (2022) 12:1964. doi: 10.1038/s41598-022-05859-8
23. Zampoglou P, Loukidis D. Investigation of the importance of climatic factors in COVID-19 worldwide intensity. *Int J Environ Res Public Health.* (2020) 17:1–25.
24. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. Hoboken, New Jersey: John Wiley and Sons (2013).
25. Steyerberg EW, Eijkemans MJ, Habbema JDF. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol.* (1999) 52:935–42. doi: 10.1016/S0895-4356(99)00103-1
26. Greenwood CJ, Youssef GJ, Letcher P, Macdonald JA, Hagg LJ, Sanson , et al. A comparison of penalised regression methods for informing the selection of predictive markers. *PLoS One.* (2020) 15:e0242730. doi: 10.1371/journal.pone.0242730
27. Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP. Why do we still use stepwise modelling in ecology and behaviour? *J Anim Ecol.* (2006) 75:1182–9. doi: 10.1111/j.1365-2656.2006.01141.x
28. Massey FJ Jr. The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc.* (1951) 46:68–78. doi: 10.1080/01621459.1951.10500769
29. White HA. Heteroskedasticity-consistent covariance matrix estimator and a direct test for Heteroskedasticity. *Econometrica.* (1980) 48:817–38. doi: 10.2307/1912934
30. Ji ZG. Hydrodynamics and water quality: Modeling rivers, lakes, and estuaries. Hoboken (New Jersey): John Wiley and Sons (2017).
31. Henriquez J, Gonzalo-Almorox E, Garcia-Goñi M, Paolucci F. The first months of the COVID-19 pandemic in Spain. *Health Policy Technol.* (2020) 9:560–74. doi: 10.1016/j.hlpt.2020.08.013
32. Güell B, Garcés-Mascareñas B. (2020). Agricultural seasonal workers in times of Covid-19 in Spain. Available at: [https://www.cidob.org/publicaciones/serie\\_de\\_publicacion/proyecto\\_papers/admigov/agricultural\\_seasonal\\_workers\\_in\\_times\\_of\\_covid\\_19\\_in\\_spain](https://www.cidob.org/publicaciones/serie_de_publicacion/proyecto_papers/admigov/agricultural_seasonal_workers_in_times_of_covid_19_in_spain) (Accessed: March 13, 2024)
33. Liu X, Huang J, Li C, Zhao Y, Wang D, Huang Z, et al. The role of seasonality in the spread of COVID-19 pandemic. *Environ Res.* (2021) 195:110874. doi: 10.1016/j.envres.2021.110874
34. Smith TP, Flaxman S, Gallinat AS, Kinoshita SP, Stemkovski M, Unwin HJT, et al. Temperature and population density influence SARS-CoV-2 transmission in the absence of nonpharmaceutical interventions. *PNAS.* (2021) 118:e2019284118. doi: 10.1073/pnas.2019284118
35. Kudo E, Song E, Yockey LJ, Rakib T, Wong PW, Homer RJ, et al. Low ambient humidity impairs barrier function and innate resistance against influenza infection. *PNAS.* (2019) 116:10905–10. doi: 10.1073/pnas.1902840116
36. Mu Y, Shao M, Zhong B, Zhao Y, Leung KMY, Giesy JP, et al. Transmission of SARS-CoV-2 virus and ambient temperature: a critical review. *Environ Sci Pollut Res Int.* (2021) 28:37051–9. doi: 10.1007/s11356-021-14625-8
37. Chaudhuri S, Giménez-Adsuar G, Saez M, Barceló MA. PandemonCAT: monitoring the COVID-19 pandemic in Catalonia, Spain. *Int J Environ Res Public Health Switz.* (2022) 19:4783.
38. Martins-Filho PR. Relationship between population density and COVID-19 incidence and mortality estimates: a county-level analysis. *J Infect Public Health.* (2021) 14:1087–8.
39. Fernández-Ahúja JML, Martínez JLF. Effects of climate variables on the COVID-19 outbreak in Spain. *Int J Hyg Environ Health.* (2021) 234:113723. doi: 10.1016/j.ijheh.2021.113723
40. Sanchez-Lorenzo A, Vaquero-Martínez J, Calbó J, Wild M, Santurtún A, Lopez-Bustins JA, et al. Did anomalous atmospheric circulation favor the spread of COVID-19 in Europe? *Environ Res.* (2021) 194:110626. doi: 10.1016/j.envres.2020.110626
41. Salzberger B, Buder F, Lampl B, Ehrenstein B, Hitzentbichler F, Holzmann T, et al. Epidemiology of SARS-CoV-2. *Infection.* (2021) 49:233–9. doi: 10.1007/s15010-020-01531-3
42. Centers for Disease Control and Prevention (CDC). (2020). Scientific brief. SARS-CoV-2 and potential airborne transmission. <https://stacks.cdc.gov/view/cdc/105949> [Accessed: March 12, 2024]
43. Chen B, Jia P, Han J. Role of indoor aerosols for COVID-19 viral transmission: a review. *Environ Chem Lett.* (2021) 19:1953–70. doi: 10.1007/s10311-020-01174-8
44. Akaishi T, Kushimoto S, Katori Y, Kure S, Igarashi K, Takayama S, et al. COVID-19 transmission in group living environments and households. *Sci Rep.* (2021) 11:11616. doi: 10.1038/s41598-021-91220-4
45. Dinoi A, Feltracco M, Chirizzi D, Trabucco S, Conte M, Gregoris E, et al. A review on measurements of SARS-CoV-2 genetic material in air in outdoor and indoor environments: implication for airborne transmission. *Sci Total Environ.* (2022) 809:151137. doi: 10.1016/j.scitotenv.2021.151137
46. Zhao L, Qi Y, Luzzatto-Fegiz P, Cui Y, Zhu Y. COVID-19: effects of environmental conditions on the propagation of respiratory droplets. *Nano Lett.* (2020) 20:7744–50. doi: 10.1021/acs.nanolett.0c03331
47. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, et al. Superspreading events in the transmission dynamics of SARS-CoV-2: opportunities for interventions and control. *PLoS Biol.* (2020) 18:e3000897. doi: 10.1371/journal.pbio.3000897
48. Kyriakopoulos AM, Papaefthymiou A, Georgilas N, Doulberis M, Kountouras J. The potential role of super spread events in SARS-COV-2 pandemic; a narrative review. *Arch Acad Emerg Med.* (2020) 28:e74
49. Rambo APS, Gonçalves LF, Gonzáles AI, Rech CR, Paiva KM, Haas P. Impact of super-spreaders on COVID-19: systematic review. *Sao Paulo Med J.* (2021) 139:163–9. doi: 10.1590/1516-3180.2020.0618.R1.10122020
50. de Crane D'Heysselaer S, Parisi G, Lissou M, Bruyère O, Donneau AF, Fontaine S, et al. Systematic review of the key factors influencing the indoor airborne spread of SARS-CoV-2. *Pathogens.* (2023) 12:382. doi: 10.3390/pathogens12030382
51. Moriyama M, Hugentobler WJ, Iwasaki A. Seasonality of respiratory viral infections. *Annu Rev Virol.* (2020) 7:83–101. doi: 10.1146/annurev-virology-012420-022445
52. Marr LC, Tang JW, Van Mullekom J, Lakdawala SS. Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *J R Soc Interface.* (2019) 16:20180298. doi: 10.1098/rsif.2018.0298
53. Yang W, Elankumaran S, Marr LC. Relationship between humidity and influenza a viability in droplets and implications for Influenza's seasonality. *PLoS One.* (2012) 7:e46789. doi: 10.1371/journal.pone.0046789
54. Chin A.W.H., Chu J.T.S., Perera M.R.A., Hui K.P.Y., Yen, H.-L., Chan M.C.W., et al. (2020). Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe* 1:e10. doi: 10.1016/S2666-5247(20)30003-3
55. Riddell S, Goldie S, Hill A, Eagles D, Drew TW. The effect of temperature on persistence of SARS-CoV-2 on common surfaces. *Virol J.* (2020) 17:1–7. doi: 10.1186/s12985-020-01418-7
56. Pan J, Yao Y, Liu Z, Meng X, Ji JS, Qiu Y, et al. Warmer weather unlikely to reduce the COVID-19 transmission: an ecological study in 202 locations in 8 countries. *Sci Total Environ.* (2020) 753:142272



57. Xie J, Zhu Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Sci Total Environ.* (2020) 724:138201. doi: 10.1016/j.scitotenv.2020.138201
58. Ahlawat A, Wiedensohler A, Mishra SK. An overview on the role of relative humidity in airborne transmission of SARS-CoV-2 in indoor environments. *Aerosol Air Qual Res.* (2020) 20:1856–61. doi: 10.4209/aaqr.2020.06.0302
59. Islam ARMT, Hasanuzzaman M, Azad MAK, Salam R, Toshi FZ, Khan MSI, et al. Effect of meteorological factors on COVID-19 cases in Bangladesh. *Environ Dev Sustain.* (2021) 23:9139–62. doi: 10.1007/s10668-020-01016-1
60. Sarkodie SA, Owusu PA. Impact of meteorological factors on COVID-19 pandemic: evidence from top 20 countries with confirmed cases. *Environ Res.* (2020) 191:110101. doi: 10.1016/j.envres.2020.110101
61. Sil A, Kumar VN. Does weather affect the growth rate of COVID-19, a study to comprehend transmission dynamics on human health. *J Saf Sci Resil.* (2020) 1:3–11. doi: 10.1016/j.jnlssr.2020.06.004
62. Ward MP, Xiao S, Zhang Z. The role of climate during the COVID-19 epidemic in New South Wales. *Australia Transbound Emerg Dis.* (2020) 67:2313–7. doi: 10.1111/tbed.13631
63. Zheng HL, Guo ZL, Wang ML, Yang C, An SY, Wu W. Effects of climate variables on the transmission of COVID-19: a systematic review of 62 ecological studies. *Environ Sci Pollut Res Int.* (2021) 28:54299–316. doi: 10.1007/s11356-021-15929-5
64. Nichols GL, Gillingham EL, Macintyre HL, Vardoulakis S, Hajat S, Sarran CE, et al. Coronavirus seasonality, respiratory infections and weather. *BMC Infect Dis.* (2021) 21:1101. doi: 10.1186/s12879-021-06785-2
65. Valero C, Barba R, Marcos DP, Puente N, Riancho JA, Santurtún A. Influence of weather factors on the incidence of COVID-19 in Spain. *Med Clin (Barc).* (2022) 159:255–61. doi: 10.1016/j.medcli.2021.10.010
66. Abdollahi A, Rahbaralam M. Effect of temperature on the transmission of COVID-19: a machine learning case study in Spain. *MedRxiv.* (2020). [Preprint]. Available at: <https://www.medrxiv.org/content/10.1101/2020.05.01.20087759v1> (Accessed November 12, 2024).
67. Vandelli V, Palandri L, Coratza P, Rizzi C, Ghinoi A, Righi E, et al. Conditioning factors in the spreading of Covid-19 - does geography matter? *Heliyon.* (2024) 10:e25810. doi: 10.1016/j.heliyon.2024.e25810
68. Tang L, Liu M, Ren B, Wu Z, Yu X, Peng C, et al. Sunlight ultraviolet radiation dose is negatively correlated with the percent positive of SARS-CoV-2 and four other common human coronaviruses in the U.S. *Sci Total Environ.* (2021) 751:141816. doi: 10.1016/j.scitotenv.2020.141816
69. Moazeni M, Rahimi M, Ebrahimi A. What are the effects of climate variables on COVID-19 pandemic? A systematic review and current update. *Adv Biomed Res.* (2023) 12:33. doi: 10.4103/abr.abr\_145\_21
70. Gutiérrez-Hernández O, García LV. ¿Influyen tiempo y clima en la distribución del nuevo coronavirus (SARS CoV-2)? Una revisión desde una perspectiva biogeográfica. *Investigaciones Geográficas.* (2020) 73:31–55. doi: 10.14198/INGEO2020.GHVG
71. Frank TD. COVID-19 epidemiology and virus dynamics. Cham, Switzerland: Springer (2022).
72. Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science.* (2020) 369:315–9. doi: 10.1126/science.abc2535
73. Fang ZG, Yang SQ, Lv CX, An SY, Guan P, Huang DS, et al. The correlation between temperature and the incidence of COVID-19 in four first-tier cities of China: a time series study. *Environ Sci Pollut Res Int.* (2022) 29:41534–43.
74. Gao M, Zhou Q, Yang X, Li Q, Zhang S, Yung KKL, et al. Nonlinear modulation of COVID-19 transmission by climate conditions. *Meteorol Appl.* (2021) 28:e1985. doi: 10.1002/met.1985
75. Prata DN, Rodrigues W, Bermejo PH. Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil. *Sci Total Environ.* (2020) 729:138862. doi: 10.1016/j.scitotenv.2020.138862
76. Qi H, Xiao S, Shi R, Ward MP, Chen Y, Tu W, et al. COVID-19 transmission in mainland China is associated with temperature and humidity: a time-series analysis. *Sci Total Environ.* (2020) 728:138778. doi: 10.1016/j.scitotenv.2020.138778
77. Karim R, Akter N. Effects of climate variables on the COVID-19 mortality in Bangladesh. *Theor Appl Climatol.* (2022) 150:1463–75. doi: 10.1007/s00704-022-04211-4
78. Yuan J, Wu Y, Jing W, Liu J, Du M, Wang Y, et al. Association 1079 between meteorological factors and daily new cases of COVID-19 in 188 countries: a time 1080 series analysis. *Sci The Total Environ.* (2021) 780:146538. doi: 10.1016/j.scitotenv.2021.146538
79. Metelmann S, Pattini K, Brierley L, Cavalerie L, Caminade C, Blagrove MSC, et al. Impact of climatic, demographic and disease control factors on the 981 transmission dynamics of COVID-19 in large cities worldwide. *One Health.* (2021) 12:100221. doi: 10.1016/j.onehlt.2021.100221
80. Fong FC, Smith DR. Exposure-lag response of air temperature on COVID-19 incidence in twelve Italian cities: a meta-analysis. *Environ Res.* (2022) 212:113099. doi: 10.1016/j.envres.2022.113099
81. Ai H, Nie R, Wang X. Evaluation of the effects of meteorological factors on COVID-19 prevalence by the distributed lag nonlinear model. *J Transl Med.* (2022) 20:1–9. doi: 10.1186/s12967-022-03371-1
82. Ladha N, Bhardwaj P, Charan J, Mitra P, Goyal JP, Sharma P, et al. Association of environmental parameters with COVID-19 in Delhi. *India Indian J Clin Biochem.* (2020) 35:497–501. doi: 10.1007/s12291-020-00921-6
83. Hu CY, Xiao LS, Zhu HB, Zhu H, Liu L. Correlation between local air temperature and the COVID-19 pandemic in Hubei. *China Front Public Health.* (2021) 8:604870. doi: 10.3389/fpubh.2020.604870
84. Hoogveen MJ, Kroes AC, Hoogveen EK. Environmental factors and mobility predict COVID-19 seasonality in the Netherlands. *Environ Res.* (2022) 211:113030. doi: 10.1016/j.envres.2022.113030
85. Matthew OJ, Eludoyin AO, Oluwadiya KS. Spatio-temporal variations in COVID-19 in relation to the global climate distribution and fluctuations. *Spa Spatiotemporal Epidemiol.* (2021) 37:100417. doi: 10.1016/j.sste.2021.100417
86. Alnaser WE, Alnaser NW, Almarabbeh AJ, Al-Ansari AM, Al-Aali HH, Ghanem AZ. The relation between SARS-CoV-2 cases in Bahrain and meteorological parameters: 27 months of data. *Front Built Environ.* (2022) 8:997380. doi: 10.3389/fbuil.2022.997380
87. Aidoo EN, Adebajji AO, Awashie GE, Appiah SK. The effects of weather on the spread of COVID-19: evidence from Ghana. *Bull Natl Res Cent.* (2021) 45:20–0. doi: 10.1186/s42269-021-00484-3
88. Lin S, Fu Y, Jia X, Ding S, Wu Y, Huang Z. Discovering correlations between the COVID-19 epidemic spread and climate. *Int J Environ Res and Public Health.* (2020) 17:7958. doi: 10.3390/ijerph17217958
89. Liu J, Zhou J, Yao J, Zhang X, Li L, Xu X, et al. Impact of meteorological factors on the COVID-19 transmission: a multi-city study in China. *Sci Total Environ.* (2020) 726:138513. doi: 10.1016/j.scitotenv.2020.138513
90. Briz-Redón Á, Serrano-Aroca Á. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Sci Total Environ.* (2020) 728:138811. doi: 10.1016/j.scitotenv.2020.138811
91. European Centre for Disease Prevention and Control (ECDC). (2020). Guidance for discharge and ending isolation of people with COVID-19, 2020. Available at: <https://www.ecdc.europa.eu/en/publications-data/covid-19-guidance-discharge-and-ending-isolation> (Accessed: March 13, 2024)
92. Gomes C. Report of the WHO-China joint Mission on coronavirus disease 2019 (COVID-19). *Braz J Implantol Health Sci.* (2020) 2:1–4.
93. He Z, Chin Y, Huang J, He Y, Akinwunmi BO, Yu S, et al. Meteorological factors and domestic new cases of coronavirus disease (COVID-19) in nine Asian cities: a time series analysis. *medRxiv.* (2020). [Preprint]. Available at: <https://www.medrxiv.org/content/medrxiv/early/2020/04/18/2020.04.15.20066613.full.pdf> (Accessed November 12, 2024).
94. World Health Organization (WHO). (2023). Weekly epidemiological update on COVID-19. Available at: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---1-september-2023> (Accessed September 1, 2023)
95. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat. Soc. Series B Stat. Methodol.* (1996) 58:267–288.
96. Zhou DJ, Chahal R, Gotlib IH, Liu S. Comparison of lasso and stepwise regression in psychological data. *Methodology.* (2024) 20:121–143. doi: 10.5964/meth.11523
97. Wu Y, Jing W, Liu J, Ma Q, Yuan J, Wang Y, et al. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Sci. Total Environ.* (2020) 729:139051. doi: 10.1016/j.scitotenv.2020.139051
98. Barcellini L, Forlanini F, Sangjorgio A, Gambacorta G, Alberti L, Meta A, et al. Does school reopening affect SARS-CoV-2 seroprevalence among school-age children in Milan? *PLoS One.* (2021) 16:e0257046. doi: 10.1371/journal.pone.0257046
99. Boey L, Roelants M, Merckx J, Hens N, Desombere I, Duysburgh E, et al. Age-dependent seroprevalence of SARS-CoV-2 antibodies in school-aged children from areas with low and high community transmission. *Eur. J. Pediatr.* (2022) 181:571–578. doi: 10.1007/s00431-021-04222-9
100. Poletti M, Raballo A. Coronavirus disease 2019 and effects of school closure for children and their families. *JAMA Pediatr.* (2021) 175:210. doi: 10.1001/jamapediatrics.2020.3586
101. Martín Vide J. *El Clima.* Geografía General Dels Països Catalans, Enciclopèdia Catalana. (1992).