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

Jesús Planella Morató
Departament de Física, Universitat de Girona
Josep Lluís Pelegrí
Departament d'Oceanografia Física i Tecnològica, Institut de Ciències del Mar (CSIC)
Marta Martín Rey
Instituto de Geociencias, UCM-CSIC
Anna Olivé Abelló
Departament d'Oceanografia Física i Tecnològica, Institut de Ciències del Mar (CSIC)
Xavier Vallès
Fight AIDS Foundation
Josep Roca
Epidemiology Unit, Hospital Universitari Germans Trias i Pujol
Carlos Rodrigo Gonzalo de Liria
Department of Pediatrics, Hospital Universitari Germans Trias i Pujol
Oriol Estrada
Directorate for Innovation and Interdisciplinary Cooperation, Northen Metropolitan Region, Institut Català de la Salut
Ignasi Vallès Casanova (ignasi.valles@mail.huji.ac.il)
Institute of Earth Sciences, The Hebrew University of Jerusalem

Article

Keywords:

Posted Date: November 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2206639/v1

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

Jesús Planella-Morató[1,2,3*], Josep L. Pelegrí[1*], Marta Martín-Rey[4,5], Anna Olivé-Abelló[1], Xavier Vallès [6,7], Josep Roca[8], Carlos Rodrigo Gonzalo de Liria[9], Oriol Estrada[10], Ignasi Vallès-Casanova[1,11*]

1. Departament d'Oceanografia Física i Tecnològica, Institut de Ciències del Mar, CSIC, Barcelona, Spain
2. Department of Physics, University of Girona, Girona, Spain
3. University School of Health and Sport (EUSES), University of Girona, Girona, Spain
4. Instituto de Geociencias, UCM-CSIC, Madrid, Spain
5. Departamento de Física de la Tierra y Atrofísica, Universidad Complutense de Madrid, Madrid, Spain
6. Fight AIDS and Infectious Diseases Foundation, Badalona, Spain
7. Fundació Institut per la Recerca Germans Trias i Pujol, Badalona, Spain
8. Epidemiology Unit, Hospital Universitari Germans Trias i Pujol, Institut Català de la Salut, Badalona, Spain
9. Department of Pediatrics, University Hospital Germans Trias i Pujol, Northern Metropolitan Region from Barcelona, Institut Català de la Salut, Barcelona, Spain
10. Directorate for Innovation and Interdisciplinary Cooperation, Northern Metropolitan Region from Barcelona, Institut Català de la Salut, Barcelona, Spain
11. Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem, Israel

*Corresponding authors:
Jesús Planella Morató: jesus.planella@udg.edu
Josep L. Pelegrí: pelegri@icm.csic.es
Ignasi Vallès Casanova: ignasi.valles@mail.huji.ac.il

Abstract

Numerous studies have explored whether and how the spread of the coronavirus disease 2019 (COVID-19) responds to environmental conditions without reaching unique or consistent answers. Sociodemographic factors such as variable population density or mobility as well as the lack of effective epidemiological monitoring difficult establishing 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 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 in the seasonal dynamics of the COVID-19 spread, which may be used to improve COVID-19 forecast models.

Introduction

A cluster of atypical pneumonias in Wuhan (China) in December 2019 disclosed the new coronavirus pathogen SARS-CoV-2 and the related clinical entity (COVID-19)[1]. This new threat was categorised by the World Health Organization (WHO) as a public health emergency of international concern on January 30 and as global pandemic on March 11, 2020[2]. As of early August 2022, COVID-19 has claimed almost 6.4 million notified deaths and about 570 million confirmed cases[3], besides the important socio-economical disruptions due the lockdowns and contention measures. The air-borne transmissivity of the new coronavirus was soon established, as well as its high infectiousness, in contrast with the previous SARS-CoV-1 and MERSV[4] .

The ongoing evolution of SARS-CoV-2 is entering a new scenario where humans are experiencing reinfections with new variants of this pathogen. Despite the availability of vaccines and the increasing rates of vaccination, the transmission of SARS-CoV-2 remains high. The limited duration of protective immunity against infection and the high genomic variability of the pathogen increase the rate of repeated infection[5,6] and extend the persistence of coronavirus over time[7]. This is especially evident with recent variants like Delta (B.1.617.2) and Omicron (B.1.1.529) and their sub-variants (BA.04 and BA.05)[8]. This new scenario may likely evolve towards an endemic disease, possibly controlled by weather conditions[9] that will cause outbreaks or seasonal peaks similar to most common respiratory infections[10]. Therefore, one of the critical questions is to determine the character and extent of this seasonality and which weather variables may have a greater incidence on the transmission dynamics of SARS-CoV-2.

In temperate regions, weather conditions such as temperature and humidity modulate the transmissibility of many common respiratory viruses such as influenza[11,12], favoring higher rates of transmission during winter[13]. Specifically, numerous studies have explored the possible relationship between climatic factors and the spread of SARS-CoV-2[14,15,16,17,18,19,20]. Several of these studies have concluded that low temperatures and low relative humidity favor the spread of SARS-CoV-2. However, models solely based on atmospheric variables have failed to predict the incidence of the disease probably due to inconsistences in the counting system of infected population[21], as well as because of incomplete consideration of relevant social variables, including the masking effect of the social restrictions imposed during the succeeding SARS-CoV-2 waves.

In the new unfolding epidemiological scenario, where policy interventions and social distancing measures are already residual, the seasonal character 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. Here we assess the relationship between the spread of COVID-19 and the atmospheric conditions in the region of Catalonia (northwestern Mediterranean) from September 2020 to December 2020. A climate-dependent COVID-19 predictive model is developed based on the cross-correlation results between a simple infection index and time-lagged atmospheric variables. The model is validated externally during the third COVID-19 outbreak, from December 2020 to February 2021 to assess its predictive performance.

Methods

Health data processing and normalization

The health data used in this study are available from the Catalan Transparency Portal database of the Catalan government and they have been processed as follows. First, only positive cases detected by PCR ($N_{PCR,+}$) were selected for the analysis. Then, detected cases were grouped into Health Regions (HRs, a total of nine with mean area about 3600 km^2) and, for each HR dataset, data were broken down into Basic Health Areas (BHAs, a total of 372 with mean area about 86 km^2 and population ranging between 5000 and 25,000 people each). Afterwards, a rectangular grid of 0.1° x 0.1° latitude (*lat*) – longitude (*lon*), covering all Catalonia, was generated. Health data were assigned to each point of the grid according to their location in the basic areas; grid points located inside one same BHA contain the same health data. Points located within the BHA that did not report data during the pandemic were excluded from the further analysis. As a result, time series for health data at each grid point were generated for the period of time analysed. Then, time series $N_{PCR,+}$ (t, lat, lon) were normalised by dividing the time series by the area and the population size of each BHA; the resulting series are positive PCR cases for 100,000 inhabitants and squared kilometre (cases per 10^5 inhab km^2).

We applied two corrections to the normalised time series $N_{PCR,+}$ (t, lat, lon). The first one considers 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 week at each BHA was computed, and a weekday factor was applied to the previous normalised dataset. A second PCR correction factor was defined as:

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

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 in account the availability of the total number of PCR tests to detect positive cases during the period of time analysed; this was particularly important during the first wave of the pandemic, as the number of available tests was significantly lower than during the other waves, which underestimated the number of infected people at the early stages of the pandemic.

The normalised time series obtained at each grid point after applying both correction factors, $N^*_{PCR,+}(t, lat, lon)$, 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 $\widehat{N}_{PCR,+}(t, lat, lon)$ at each point of the grid covering Catalonia.

Daily infection index

The final dataset $\widehat{N}_{PCR,+}(t, lat, lon)$ was used to define an infection index, to be used for monitoring the risk of contagious 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:

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

This pandemic parameter, which is obtained directly from the health dataset, provides information on the people infected daily with the virus with respect to the total population that is potentially infectious, which is estimated as the people infected during the prior 10 days[22].

After processing the health data, the BHAs were classified according to population density. In low-density populated areas, reporting of positive cases experienced some difficulties during 2020, 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^{-2}) for further analysis (Fig.1). One of the selected BHAs is located in the city of Barcelona (BCN-10A, d_{BCN} = 11873 inhab km^{-2}) and five more are included in districts located in towns of the Barcelona metropolitan area: Gava (GVA-2, d_{GVA} = 1847 inhab km^{-2}), Sant Just Desvern (SJD, d_{SJD} = 2340 inhab km^{-2}), Sant Vicenç dels Horts (SVH-2, d_{SVH} = 2409 inhab km^{-2}), Rubi (RUB-3, d_{RUB} = 644 inhab km^{-2}) and Terrassa (TRS-E, d_{TRS} = 1864 inhab km^{-2}). 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} = 1573 inhab km^{-2}), and a district of the town of Lleida, located in the interior of Catalonia (LLEI-2, d_{LLEI} = 1123 inhab km^{-2}).

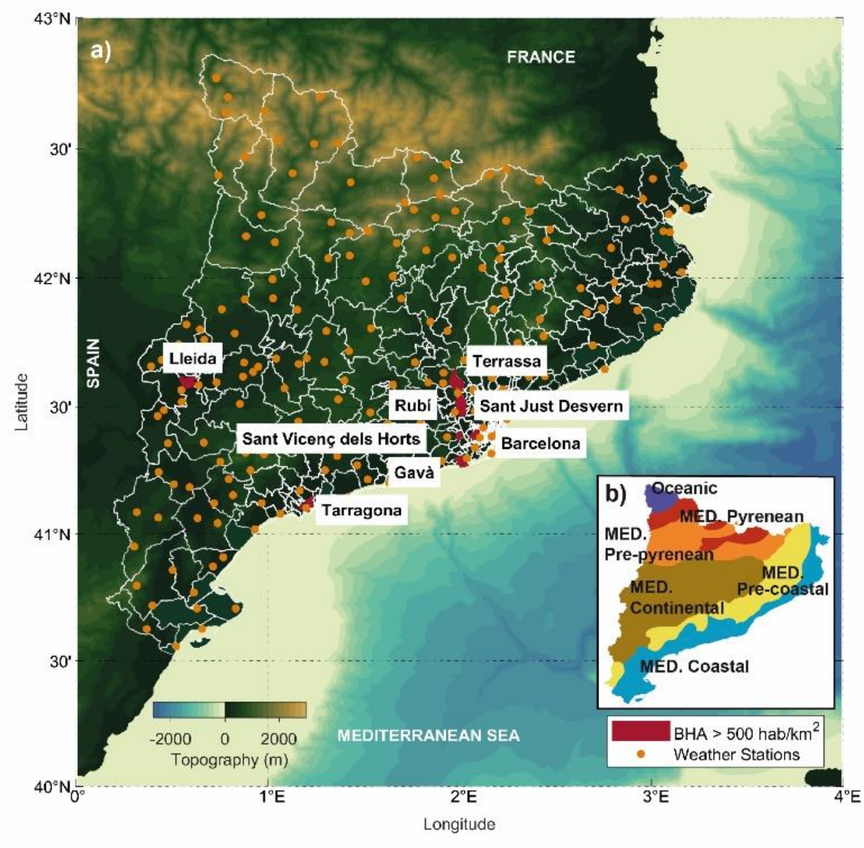

Figure 1. (a) Basic health areas (BHAs, delimited in white) and automatic weather stations (orange dots) in Catalonia. Those BHA selected for this study, with a population density $d \geq 500$ inhab km^2, are drawn in red. (b) Bioclimates in Catalonia according to the climatic conditions: Mediterranean coastal, Mediterranean pre-coastal, Mediterranean continental, Mediterranean pre-Pyrenean, Mediterranean Pyrenean and Oceanic[23] published on the Meteorological Service of Catalonia (MSC).

Health data selection

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 academic courses, the Spanish government opened a period with no mobility and distance restrictions 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 have assessed the impact of weather on the propagation of COVID-19 between September 2020 and February 2021. An internal validation is done from September 1 to November 18, 2020, which is the setup period. This 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. An external validation is done from November 19, 2020, to February 28, 2021, which is the forecast period. It covers a period after the end of the second COVID-19 wave and before a substantial fraction of the population was vaccinated.

Weather data processing

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. For each atmospheric variable, the original 30-minute data available since 2009 from the Meteorological Service of Catalonia (MSC) were averaged to daily values. This allowed estimating the maximum and minimum daily temperatures and hence the daily thermal amplitude. Mean temperature difference between consecutive days was also computed. The final time series were smoothed with a run-averaged filter of three days and 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 impact of weather on the propagation of the COVID-19 are: daily mean temperature (T_{mean}), relative humidity (*RH*) , shortwave solar radiation (*Rad*), surface pressure (*P*), daily thermal amplitude (*DTA*), daily minimum temperature (T_{min}), daily maximum temperature (T_{max}), temperature difference between consecutive days (ΔT), and daily precipitation (*Prec*).

Cross-correlation Analysis

The relationship between the local weather and health variables was explored through a time-lagged cross-correlation analysis between each of the nine atmospheric variables and the I_c index. To explore the similarity between the two series, the atmospheric variables were shifted forwards and backwards in time with respect to I_c. Negative time lags ($\tau < 0$) indicate that changes in the infection index follow the atmospheric variables. Following our initial hypothesis, that infection is driven by weather conditions, only negative lags are considered for further analysis; considering the average maximum reported COVID-19 incubation days[22] and according to the I_c definition, the maximum time lag considered was set to $\tau_{max} = 10$ days. In order to quantify the impact of weather on the spread of the virus, we assumed that an atmospheric variable affects the virus propagation if the sample correlation coefficient (*CCF*) between this variable and the infection index is significant at a 95% confidence level.

Selection of climatic variables to build the model

The propagation of the virus was modelled using a multiple linear regression model for each BHA, with the predicted infection index $I_{c,pred}$ expressed in terms of p local climatic predictors:

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

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 τ_j^* is the characteristic lag for the X_j predictor. The characteristic time lag is defined as the time interval that produces the highest correlation coefficient between the predictors and the observed I_c index. Hence, a total of $2p+1$ parameters are fitted for each BHA.

Building the model: Model descriptors and statistics

Before building the model, we explored the potential collinearity effects between the predictors as detailed next. First, the correlation coefficients $r_{i,j}$ between two predictors, namely X_i and X_j, were computed. Second, correlation t-tests were 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:

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

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 were correlated, the degree of collinearity between them was evaluated using the variance inflation factor, defined as:

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

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 considered that collinearity was not significant and both predictors could be used to build the model[24].

After inspecting the collinearity between predictors, we tested the regression coefficients separately for each BHA in order to select the predictors included in the final model. The significance for the regression coefficients was assessed using the t-test. Since these tests can be very conservative, we applied the forward stepwise estimation method to decide whether a candidate predictor must be included in the model, as follows. First, we selected the predictor with the highest correlation coefficient with the infection index. Then, this predictor and the infection index are fitted 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 highest correlation with the infection index was added to the linear model and r^2_{adj}, was recomputed. The F-test statistics was used to decide whether the addition of the remaining predictor made a significant contribution to the model. The F-statistic was calculated as:

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

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 confidence level > 90 %. If the F-value is significant at this confidence level (p-value ≤ 0.1), the model improves significantly with the addition of the new predictor, which is maintained in the model. The final model includes all predictors that passed this partial F-test.

Additionally, the final model was 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 and is calculated as:

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

where SSR, SSE and n are defined in Eq. (6) and k is the amount 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 for that statistic. If the p-value was higher than a 0.05 significance level, we concluded that there is enough statistical evidence that the final model fits the observations better than the intercept-only model.

Finally, the assumptions inherent to the linear regression model were analysed through the residuals from the fitted model; statistical tests were implemented to complement the graphical information of the residual plots. In this way, the Kolmogorov-Smirnov (K-S) test was conducted at a 95% confidence interval in order to examine if the residuals are normally distributed[25]. Thus, the test statistic was computed to evaluate whether the gap between the empirical and normal (hypothesized) distributions of the residuals is significant at the considered confidence interval. In addition, the White test for heteroscedasticity[26] at 95% confidence level was applied 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 χ^2 distribution, is smaller than the significance level of the test, then the statistic is not significant and there is no evidence of heteroscedasticity in the final model.

Model validation during the setup period

The final model was 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. Note that the time for this validation has been shifted backwards due to the characteristic lags for P and RH found in the correlation analysis, as introduced in Eq. (3). In the LOOCV procedure, the infection index I_c for the day i ($I_{c,i}$) was excluded and the model was fitted using the remaining data. During this process, each of these training subsets provided an individual model, which was expected to be slightly different from the original one, and each model was used to predict the infection index $I_{c,[i]}$ with the i-th case removed from the sample. The prediction error for each model was computed as $e_{[i]} = I_{c,i} - I_{c,[i]}$; it is a measure of how close the prediction is to the observation when this observation is omitted. The absolute percentage error in each measurement was obtained scaling each prediction error against its corresponding observed value, that is $APE_{[i]} = 100 \cdot | I_{c,i} - I_{c,[i]} | / I_{c,[i]}$ .Then, the overall performance of the model was estimated using the mean of the absolute percentages calculated previously ($MAPE_{CV}$). The prediction error $e_{[i]}$ was also used to estimate the root mean square of that errors ($MAPE_{CV}$) when the predictions were obtained by cross-validation. Finally, other statistical parameters were determined from the linear regression fit of the observed values $I_{c,i}$ to the cross-validated ones $I_{c,[i]}$. A significance test was implemented to evaluate 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. The linear fit allows to compute the R-squared of the cross-validation, q^2_{CV} . The F-test of overall significance was performed to investigate whether the predictions obtained from the predictor variables, explain a significant part of the variance observed in the responses compared to data obtained from a model with no predictors.

The statistical parameters obtained by the application of the LOOCV method were compared to the corresponding ones obtained from the data used to build the model. Table 1 summarizes the expressions of the statistical parameters used for construction and validation of the model[27].

Model building and External validation	LOOCV (Internal validation)
$MAPE = \frac{1}{n}\sum_{i=1}^{n} APE_i = \frac{100}{n} \cdot \sum_{i=1}^{n} \frac{\lvert I_{c,i} - I_{c,pred,i} \rvert}{I_{c,i}}$	$MAPE_{CV} = \frac{1}{n}\sum_{i=1}^{n} APE_{[i]} = \frac{100}{n} \cdot \sum_{i=1}^{n} \frac{\lvert I_{c,i} - I_{c,[i]} \rvert}{I_{c,i}}$
$RMSE = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}\left(I_{c,i} - I_{c,pred,i}\right)^2}{n}}$	$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{n} e_{[i]}^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}\left(I_{c,i} - I_{c,[i]}\right)^2}{n}}$
$r^2 = 1 - \frac{\sum_{i=1}^{n}\left(I_{c,i} - I_{c,pred,i}\right)^2}{\sum_{i=1}^{n}\left(I_{c,i} - \bar{I}_c\right)^2}$ $r_{adj}^2 = 1 - (1 - r^2)\frac{n-1}{n-k-1}$	$q_{CV}^2 = 1 - \frac{\sum_{i=1}^{n}\left(I_{c,i} - I_{c,[i]}\right)^2}{\sum_{i=1}^{n}\left(I_{c,i} - \bar{I}_c\right)^2}$

Table 1. Summary of the expressions used to determine the statistic parameters (*MAPE*, *RMSE* and 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 validation during the forecast period

The generalization of the final model was investigated using independent meteorological and health datasets for the period until February 2021. This forecast period, which includes the pandemic's third wave in Catalonia, allows testing the predictive ability of the model in each of the BHAs. Mobility and social restrictions in Catalonia decreased at the end of November, 2020, but the increase of positive cases led to new restrictions on mobility (including curfew at night) and restrictions in bars, cafes and restaurants. The number of positive cases was maximum around January 15, 2021, followed by a decrease in the number of infections, falling to the levels before the onset of the third wave in February. On the other hand, vaccination in Catalonia began on December 27, 2020. The percentage of people who had received their first vaccine was low (< 10 %) until February, and increased to about 30% by the end of April 2021. Despite this minor fraction of vaccinated population and the reset of some mobility limitations, we have used 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 was done using the same procedure as described for the internal validation. First, the values of the infection index $I_{c,i}$ during this time period were calculated from the health data time series. Second, the predicted values $I_{c,pred,i}$ were estimated from the linear regression model. Third, the predicted and observed values were compared and the statistical significance of the linear fit, the slope and the intercept were assessed. Finally, the prediction errors for each value of $I_{c,i}$ were 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 were 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 1).

Results

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 contagious or infection index I_c (see Methods) (Fig. 2). Values of I_c higher than 1 indicate that the number of positive cases at any day are 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 varies substantially between locations, from maximum values in excess of 3.5 per 10^5 inhab km^2 for the most populated values, decreasing to peak values about 0.05 cases per 10^5 inhab km^2 in the least populated BHAs. However, the peak infection indexes 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 mask. In the absence of lockdown, an increase of the SARS-CoV-2 transmissivity has been found to occur in areas of high-density population[28], likely related with environmental factors[19].

During the first pandemic wave, the Barcelona (BCN-10A) reached the highest number of confirmed cases with 3.9 cases per 10^5 inhab km^2 followed by 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. These maxima were reached 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[29], the number of confirmed cases sharply decreased in all BHAs. During the first half of April the number of positive cases descended progressively for Gava (GVA-2), Rubi (RUB-3), Terrassa (TRS-E) and Lleida (LLEI-2) (Fig. 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, Fig. 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^2. The single exception was LLEI-2, which reached 1.5 cases per 10^5 inhab km^2 (Fig. 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 due to socio-economic conditions[30].

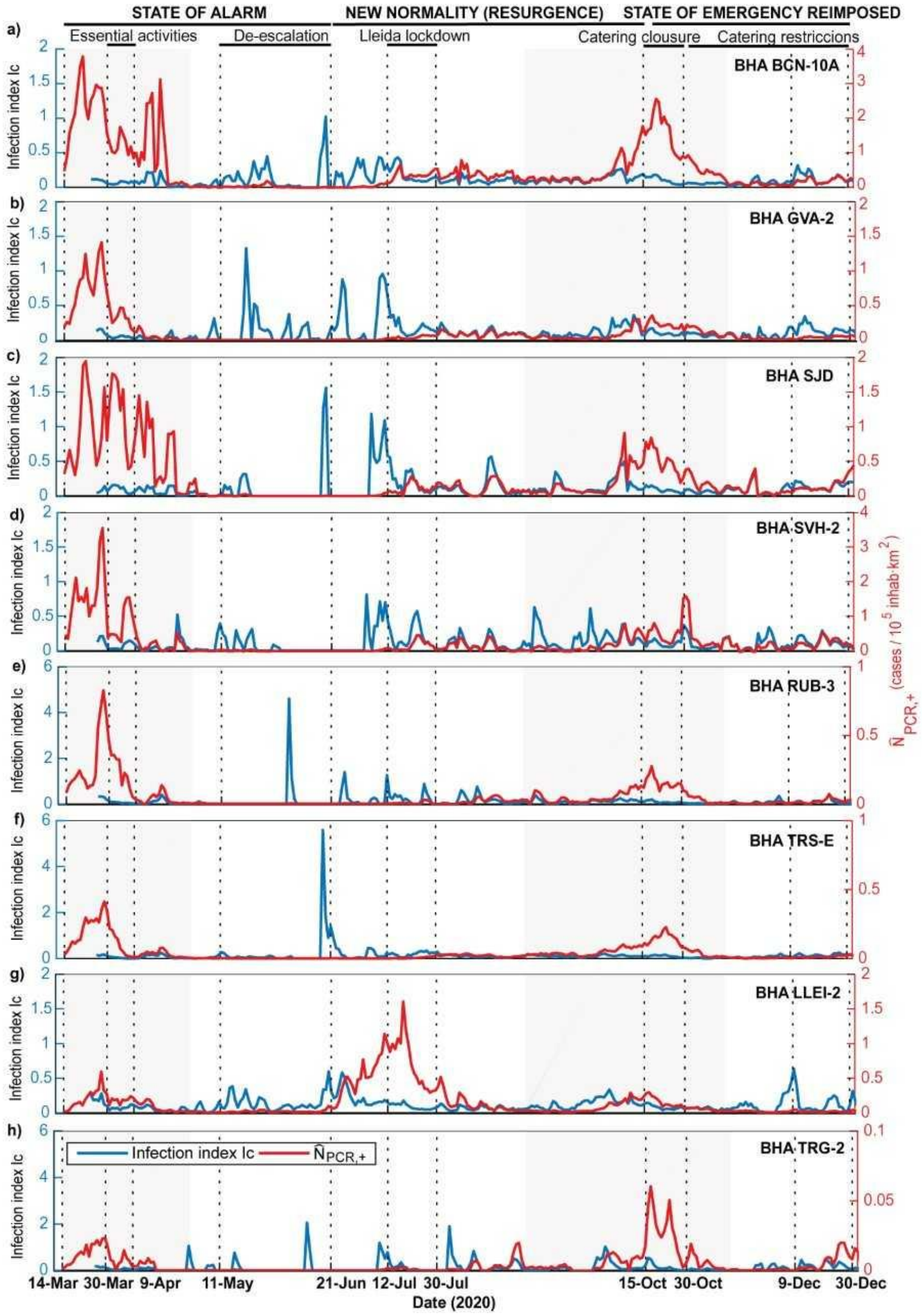

Figure 2. (a-h) Temporal evolution during 2020 of the number of COVID-19 cases in the eight selected Catalan BHAs, presented both in terms of the normalised number of cases $\widehat{N}_{PCR,+}$ as determined through the PCR positive tests (red line) and the rate of infections I_c (blue line). In each panel, the grey shaded areas indicate the first and the second pandemic wave periods and the vertical dotted lines delineate the duration of the main social and mobility restrictions imposed in Catalonia due to the COVID-19 disease.

During this first wave the infection index I_c remained low, indicating that the government measures were effective to prevent contagion. In contrast, several high I_c peaks appeared intermittently during summer (Fig. 2) without any major response on the standardised PCR, with the exception of LLEI-2 (Fig. 2g). This suggests that there were irregular infection episodes taking place although 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 favourable 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 (SVH-2 and BCN-10A) showed again the highest values, respectively with 1.7 and 2.7 cases per 10^5 inhab km^2, although these values were lower than during the first pandemic (Fig. 2a,d). In contrast, in the coastal town of Tarragona (TRG-2) values reached 0.07 per 10^5 inhab km^2, higher than during the first wave (Fig. 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 Rubi (RUB-3). The mobility restrictions remained low until the end of October, therefore the increase in COVID-19 transmission may be associated with weather conditions[31,19]; indeed, during this time period, several cold fronts circulated from west to east in a row, a typical autumn scenario (see Figs. S1 and S2, Supplementary information).

Correlation between weather variables and the infection index

In order to explore the role of weather on the second COVID-19 wave in our study region, we analyse the time-lagged correlations between local infection indicators and weather data for each of the eight BHAs. We use nine daily-averaged atmospheric variables for each BHA, during the period from September 1 and November 15, 2020 (see Methods). The initial selection of humidity and temperature was 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[32,33,34]. Additionally, we include solar radiation, precipitation, surface pressure, minimum and maximum temperature, daily thermal amplitude and mean temperature difference between consecutive days.

The evolution of the normalised number of cases and the infection index can be compared with changes in the atmospheric variables as it is shown in Fig. 3 for BCN-10A (see in Fig. S2, Supplementary information, for all BHAs). Oscillations in surface pressure, temperature, relative humidity and solar radiation are associated with the passage of cold fronts in this area (see, for example, the oscillation in September 7 to 10). The time series also reveals oscillations in the infection index, which apparently appear several days after the atmospheric changes. 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.

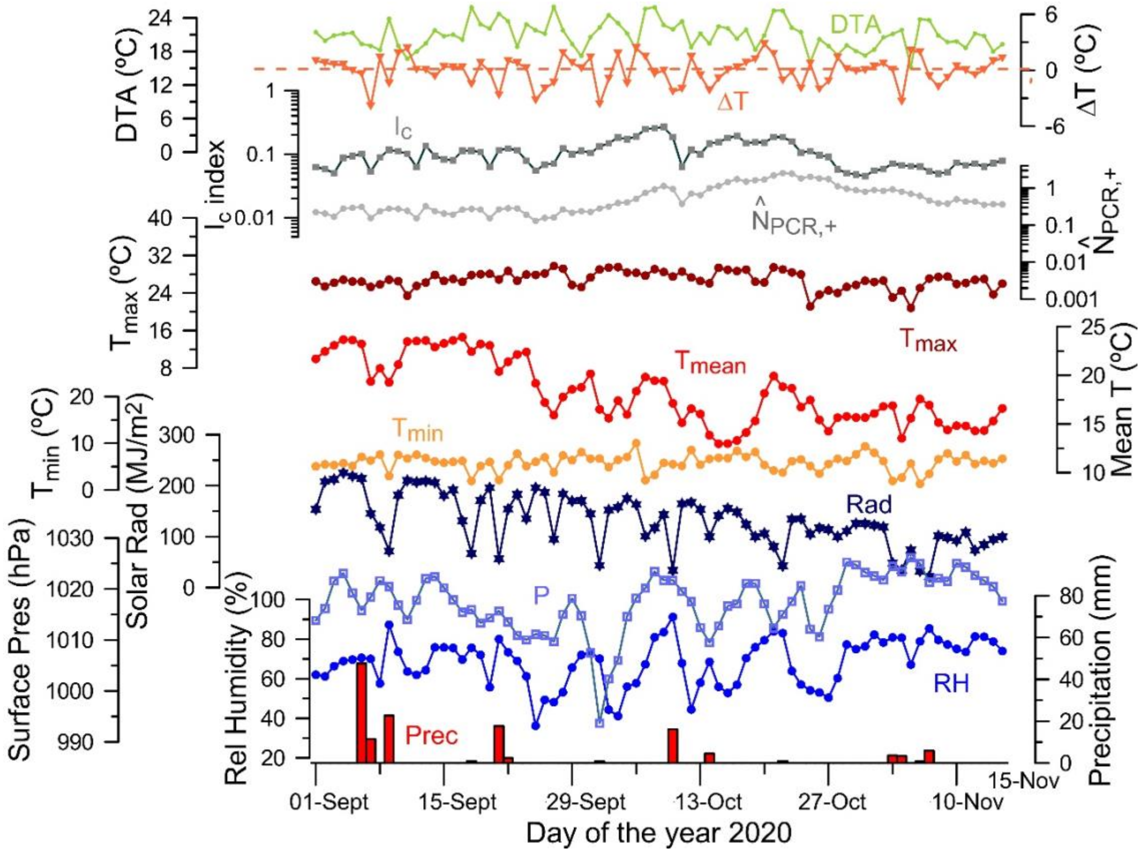

Figure 3. Time series of daily values of atmospheric variables during the second outbreak (September 1 to November 15) together with the normalized number of cases ($\widehat{N}_{PCR,+}$) and the infection index (I_c) for BCN-10A. The atmospheric daily variables are mean temperature (T_{mean}), relative 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 mean temperature difference between consecutive days (ΔT) . The units for these variables are indicated in their corresponding axes.

Our results show consistent significant negative correlation between surface pressure (*P*) and relative humidity (*RH*) and 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 is significant at $> 99\%$ confidence level in all BHAs for specific time lags(τ). The average cross-correlation coefficient for all BHAs is maximum ($CCF \cong$ -0.5) at $\tau = 7$ days, although some areas have even higher coefficients at shorter lags (Fig. S3, Supplementary Information). On the other hand, the statistical significance of $RH - I_c$ decreases below 99% in four of the BHAs (GVA-2, SJD, SVH-2 and TRS-E) but still remains significant at $> 95\%$ confidence level. In this case, the highest values for the correlation coefficients ($CCF \cong$ - 0.35) are obtained for τ = 3 to 5 days. The other meteorological variables evaluated do not show consistent correlation scores (Fig. S3, Supplementary information). In particular, the variables derived from temperature and precipitation are, in general, poorly correlated with the infection index. Only two BHAs (LLEI-2 and TRG-2) have significant correlations at the longest lags, $\tau = 10$ days. The daily thermal amplitude is significant at $> 95\%$ confidence level 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 (Fig. S3, Supplementary information).

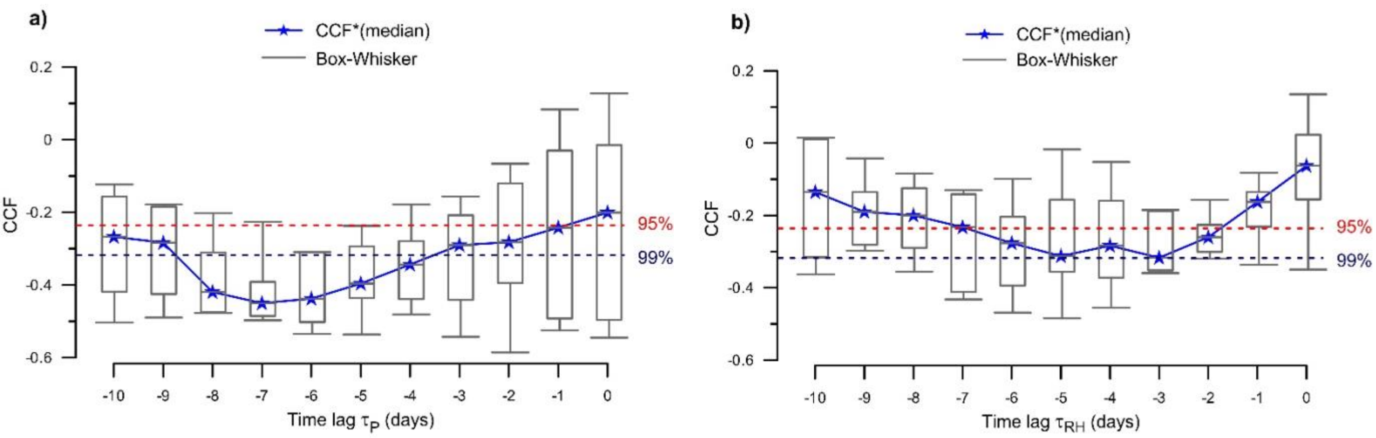

Figure 4. Composite box plots of the cross-correlation coefficients (*CCF*) of a) surface pressure and b) relative humidity with respect to I_c as a function of lag time, calculated using the eight reference BHAs. The lower and upper ends of the box represent the first and third quartiles, respectively, and the median (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 95% (red line) and 99% (blue line) confidence levels.

We may 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 (Fig. 4). For each lag, we select the median of the *CCF* box plots (hereinafter CCF^*) as the representative value of the set. For both variables, CCF^* shows a well-defined valley where the negative correlation is highest (Fig. 4). Surface pressure has the largest absolute correlation at $\tau_{P,min}$ = - 7 days and the relative humidity has minima at both $\tau_{RH,min}$ = - 3 days and - 5 days. For both variables, the τ_{min} values occur at CCF^* significant confidence levels. For surface pressure, CCF^* is actually significant at 99% confidence level for $\tau_P \in$ [-8, -4]; in particular, the smallest interquartile ranges (IRQ) for the *CCF* distributions are in this lag interval (IRQ ≅ 0.1), indicative of a minimum dispersion of the *CCF* values among the different BHA (see Fig. 4a). The smallest dispersion is found at $\tau_{P,min}$ = - 7 days, with IRQ ($\tau_{P,min}$) ≅ 0.05. 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 behaviour, but in a range of $\tau_{RH} \in$ [-6, -2], although the confidence level decreases to 95% (Fig. 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 ($\tau_{RH,min}$) ≅ 0.1. These results indicate that the infection index is negatively correlated with the surface pressure conditions 7 days before and with the relative humidity conditions 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 (τ^*) for the surface pressure and the relative humidity in the study area.

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

The cross-correlation analysis indicates that surface pressure and relative humidity are the only variables that bear statistical significant correlation with I_c (> 95 % confidence level) in all BHAs. Hence, we identify them as potential predictors for the time evolution of COVID-19 in Catalonia (Spain), with one single time lag for each predictor over the entire region. However, the infection index differs substantially between the different BHAs, likely reflecting specific demographic and geographic characteristics (Fig. 1). Hence, we propose a multiple linear regression model of the infection index in terms of these two variables, with the same time lags for all BHAs but allowing for changes in the coefficients as follows (Eq. 3 in Methods):

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

where $c_i \ \forall$ [1,3] are the model regression coefficients for each BHA, the predictor 1 is the surface pressure *P* (measured in hPa) and the predictor 2 is 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 τ^{*}_{j} is the characteristic for the j predictor. The time-lags between the weather and health variables cause that the atmospheric time series start and end later than the health time series.

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 during the second outbreak (September 1 to November 15, 2020, for the atmospheric variables and time lagged for the health variables, the setup period). This includes obtaining common time lags for all BHAs and individual regression coefficients for each BHA. The model is then used to forecast I_c ($I_{c,pred}$) 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, Supplementary information). Note that after February 2021, over 10% of the Catalan population had already received their first vaccine (Chaudhuri et al., 2022)35, likely undermining the use of more recent data for external validation. The results and statistics of the model validation are extensively explained in different sections of the Supplementary document.

Fitting the model I_c predictions to the observations for the setup period (see Model parameters and statistics, Supplementary information) shows that the two predictors, P and RH, have a significant contribution to the model (> 90% confidence level) in four of the eight BHA (BCN-10A, SJD, RUB-3 and LLEI-2). For the other four stations, it turns 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 at 95% confidence level in all BHAs and their values vary between [-10, -3] $\times 10^{-3}$ hPa^{-1} (mean(c_1) = -5.90×10^{-3} hPa^{-1}; std(c_1) = 1.90×10^{-3} hPa^{-1}) and [-4, -1] $\times$ 10^{-3} (mean(c_2) = -1.91×10^{-3} ; std(c_2) $=1.12\cdot\times10^{-3}$). 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 (Fig. 4a,b).

During the pandemic's second outbreak or setup period, as expected, the model captures well the general behaviour of I_c. This is confirmed by the significant correlation coefficients (r) obtained between the $I_{c,pred}$ and I_c. 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-val < 0.01). We conclude that our simple climate-dependent model reproduces the main changes anomalies in the infection index during the second COVID-19 outbreak in Catalonia (fall 2020).

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 (Fig. 5). The mobility and social measures progressively relaxed between late October 2020 and early May 2021, when the state-of-alarm was revoked; during this period, however, social interactions increased substantially during the festivities in late December and early January. During this third wave, the infection rate followed specific features depending on the analysed BHAs. For example, in BCN-10A, GVA-2 and LLEI-2, there is a clear peak in mid-December followed by two secondary peaks in late December and early January (Figs. 5a,b,g). However, this is not the case for SJD, that had peaks of comparable amplitude (Fig.5c), or for TRS-E, where the situation was reverted and the two secondary peaks took place in December and the main peak during the New Year's Eve (Fig. 5f).

We apply our climate-dependent multiple-regression model to the forecast period (November 19, 2020, to February 2, 2021), in order to forecast the infection index for each health area ($I_{c,pred}$) during the pandemic's third outbreak (Fig. 5). Despite differences in amplitude, most BHAs illustrate three peaks in I_c during mid- and end-December 2020 and in January 2021 (Fig.5). Outstandingly, the predictions reproduce the basic features in the observed index (I_c) for most BHAs. The model reproduces quite well the increase of infection index I_c during mid-December and January in the LLEI-2 ($r = 0.33$), SJD ($r = 0.42$), GVA-2 ($r = 0.40$) and BCN-10A ($r = 0.33$). Notice that for GVA-2 and BCN-10A the model simulates the enhancement of infection rate 5-7 days earlier than in the observations (Fig.5a,b). However, the model generally fails to reproduce the third peak of I_c at the end of December. 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 (Fig. 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 behaviour between the observed and predicted values (Fig. 5d).

The relative success of the climate model is remarkable if we consider that the entire forecast period was very anomalous in terms of limited mobility and social interaction, with restrictions until early May but with temporary relaxation during the holiday season. 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-val > 0.01), as it happens in several BHAs between November 19 and the end of the year (Fig S7, Supplementary information).

Discussion

We have explored the effects of weather conditions on the propagation of the coronavirus SARS-CoV-2 in eight highly-populated basic health areas (BHAs) of Catalonia, in the north-western Mediterranean area, between March 2020 and February 2021. For this purpose, we have proposed, developed and validated a predictive model that assesses the impact of weather on the COVID-19 infection rate. The transmission of the virus has been monitored using a simple parameter, the daily infection index I_c, which is defined as the ratio between the people infected with the virus at any day and the potential infectious population at that same day (people infected with the virus during the 10 previous days). Nine weather variables (mean temperature, relative humidity, solar radiation, surface pressure, daily thermal amplitude, minimum and maximum temperature, precipitation and mean temperature difference between two consecutive days) have been explored as the potential environmental drivers of the virus expansion.

A cross-correlation analysis between each weather variable and the I_c index was done from September 1 to November 15, 2020, a period of low mobility and social restrictions. The results of this cross-correlation analysis show significant correlations for surface pressure (P) and relative humidity (RH), at confidence levels respectively above 99% and 95%, in all BHAs; in contrast, all other climatic variables are either poorly correlated with I_c or exhibit non-coherent correlations. From our knowledge, this is the first time that surface pressure is proposed as a key factor for virus transmissivity. A predictive model for I_c solely based on P and RH provides consistent and significant results. The cross-correlation analysis indicates that the highest correlations occur when I_c lags P and RH by 7 days and 3 days, respectively.

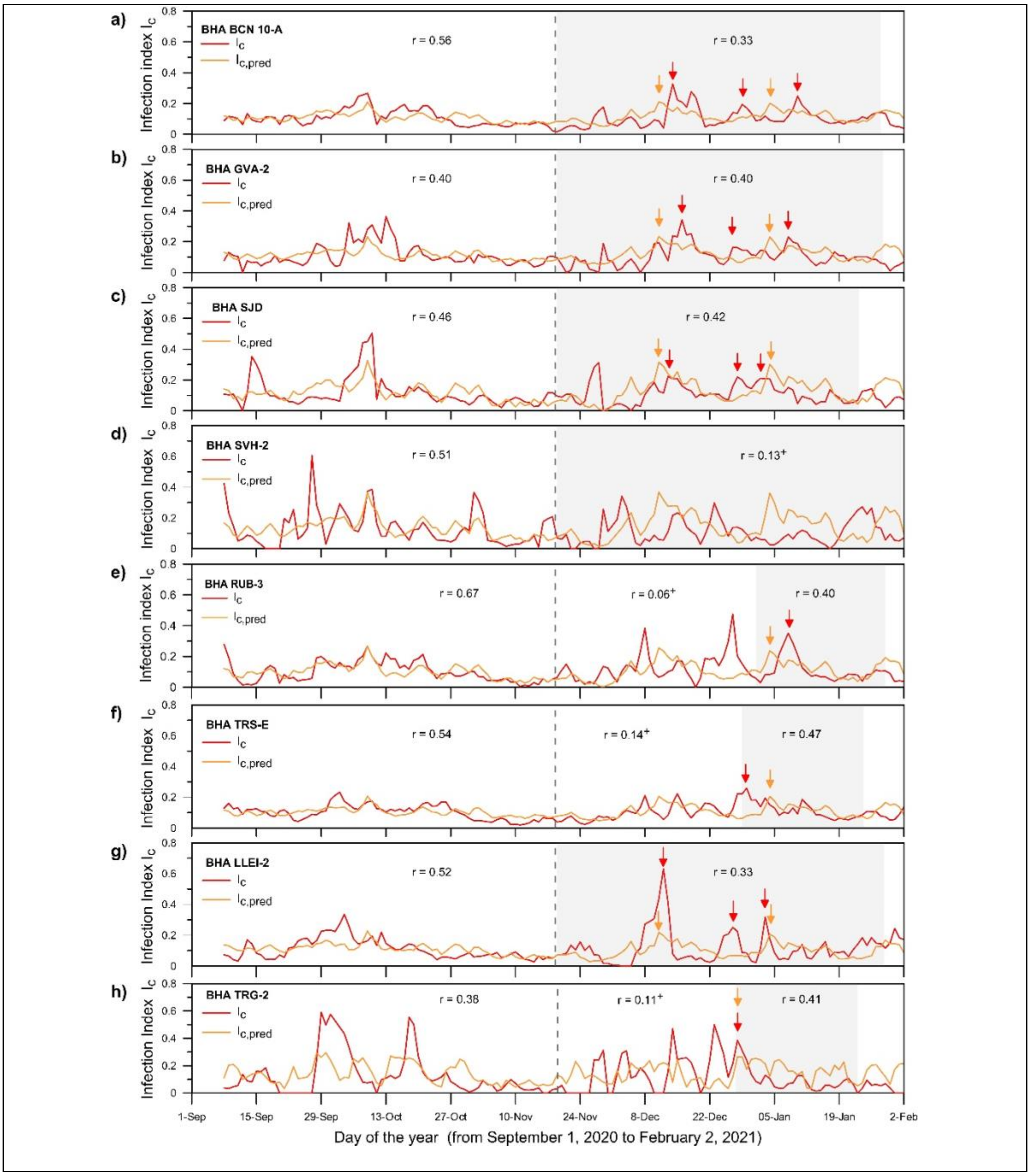

Figure 5. Results for the eight BHAs (panels a) through h)) showing the temporal evolution of the observed infection I_c index (red lines) and the corresponding predicted $I_{c,pred}$ values (light lines) as obtained from the climate model between September 1, 2020, and February 2, 2021. The vertical dashed line on November 18, 2020, delineates the setup and forecast periods. The grey 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 minimum 90% confidence level; in several cases it corresponds to time periods of low normalized number of cases (see Fig. S7, Supplementary information). The arrows indicate the main and the secondary observed (red) and predicted (light orange) I_c peaks during the forecast periods.

We find that during the model setup period, when there were no mobility and social restrictions and there was not yet a vaccine available, surface pressure *P* and relative humidity *RH* work well as COVID-19 predictors. A multiple-regression model that uses only five parameters, with time lags obtained from the composite analysis of all eight BHAs, has predictive capacity with a confidence degree above 95%. In particular, the model regression coefficients are negative for all health areas, indicating that a decrease of *P* and *RH* causes an increase of the I_c index after several days. The linearity of the model implies that the highest values of I_c are found between 3 and 7 days after *P* and *RH* reach their minima.

The climate-dependent model works reasonably well when applied to the posterior predictive period, with a predictive skill significant at a 90% level in four out of the eight BHAs (BCN-10A, GVA-2, SJD and LLEI-2), and remaining 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 14% to 45% of the variability of the infection index I_c can be explained by the weather conditions in the selected eight BHAs of Catalonia during the setup period (September 1 to November 15), and that 11% to 22% of the infection rate is attributed to the weather component during the predictive period. The reduction in this latter period is likely associated to the changing social interaction and mobility measures, much greater than during the setup period.

The proposed mechanisms for weather-mediated changes in respiratory disease include the effects of weather on virus survival in surfaces and outdoors, changes in the susceptibility to the disease, and also variations in human social behaviour. To our knowledge, surface pressure has not been previously described as a driver of SARS-CoV-2 transmission. 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 behaviour. The most evident is to seek indoor spaces and for much longer time periods under bad weather conditions, which are in turn more suitable conditions of infection spreading, through increased persistence of the virus in the air and surfaces[36], and enhanced close contacts[37]. Accordingly, a scientific report lead by WHO-China commission concluded that 78-85% of transmissions occurred within household settings during the first wave, suggesting that transmission occurs during close and prolonged contact[2]. Moreover, in enclosed spaces with inadequate ventilation, small infected droplets and particles can remain suspended for minutes to hours[38,39].

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 2-14 days[40]. The 3-days lag for relative humidity is more difficult to justify, even if it still lays between the estimated incubation period bounds. The linear model predicts that dry conditions will favour the propagation of the virus (an increase in I_c) about 3 days later and wet conditions will tend to inhibit it. Low humidity has been considered as a key weather factor associated with the transmission and stability of respiratory viruses such as influenza. This effect has been related to susceptibility and severity of influenza infection through disruption of local mucosal immunity of the respiratory tract[41,32,12,42], and recent studies support an inverse relation between humidity and the spreading of SARS-CoV-2, consistent with our findings[43,44]. However, a recent study[45] found that coronavirus have a different pattern of weather susceptibility as compared with the influenza virus, with an increase of transmission (above 80%) in England and Wales (UK) during periods of high relative humidity, becoming a better predictor than specific or absolute humidity. The positive effect of relative humidity towards SARS-CoV-2 infectiousness has also been found in other studies[46,47,15,16,17,48].

The disparity of findings related to humidity suggests that there may be geographically-dependent factors[41] that possibly constrain the humidity conditions and the virus response. Our results show that surface pressure is the most influential weather variable, 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 crossing 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 driver while relative humidity has a more secondary role. Something similar could be happening with other possibly secondary variables such as sunlight radiation, which has been related to susceptibility to SARS-CoV-2 and other airborne viruses[49].

Our results do show that there is a significant role of weather in epidemic outbreaks, although the lack of immunity (high susceptibility) of a population remains as the main driver[50]. The inability of current vaccines and previous infections to fully prevent the infection of new

SARS-CoV-2 variant and the definitive leverage of social restrictions, increase the interest on weather conditions. Under this post-pandemic induction phase context, weather conditions may substantially influence the onset and extend of new waves and outbreaks (as observed in other respiratory viruses), and even lead to the establishment of seasonal patterns of SARS-CoV-2 infection in the near future. Our results have identified relative humidity and, particularly, surface pressure as useful predictors to be included in more complex epidemiological models of the spread of COVID-19 and other respiratory viruses.

Data availability

The data analyzed during the current study is available 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-meteorol-giques-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.

Authors contribution

J.P.M. wrote the main manuscript text and performed the statistical analysis and the model experiments. I.V.C. analysed the health and meteorological data and wrote the main manuscript text. J.L.P. contributed to the conception or design of the work and revised the main manuscript. M.M.R. made substantial contributions to the conception of the work. A.O.A. analysed the meteorological data and critically revised the manuscript. X. V. made substantial contributions to the conception of the work and critically revised the manuscript. J.R. provided the health data. C.R.G.L. and O.E. critically revised the manuscript.

Competing interests

The authors declare no competing interests or other interests that might be perceived to influence the results and/or discussion reported in this paper.

References

1. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med*. 382, 727–733 (2020)

2. Gomes, C. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). *Braz. J. Implantol Health Sci*. 2,12 (2020)

3. World Health Organitzation World Health Organization (WHO). COVID-19 weekly epidemiological update https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---3-august-2022 (2020)

4. Kirtipal N., Bharadwaj S. & Kang S. G. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infect. Genet. Evol.* https://doi.org/10.1016/j.meegid.2020.104502 (2020)

5. Tegally, H. et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021)

6. Hadj Hassine, I. Covid-19 vaccines and variants of concern: A review. *Reviews in medical virology*, *32*(4), e2313 (2022)

7. Sciscent, B. Y., Eisele, C. D., Ho, L., King, S. D., Jain, R. & Golamari, R. R. COVID-19 reinfection: the role of natural immunity, vaccines, and variants. *J Community Hosp Intern Med Perspect*, *11*(6), 733-739 (2021)

8. López-Muñoz, I. 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. 28(10):1999-2008 (2022)

9. Callaway, E. Beyond Omicron: what's next for COVID's viral evolution. Nature, 600(7888), 204-207 (2021).

10. Fontanet, A., Autran, B., Lina, B., Kieny, M. P., Karim, S. S. A., & Sridhar, D. SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet*, *397*(10278), 952-954 (2021)

11. Lowen, A. C., Mubareka, S., Steel, J., & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS pathogens*, *3*(10), e151 (2007)

12. Marr L. C., Tang J. W., Van Mullekom J. & Lakdawala S. S. Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *J. R. Soc. Interface* 16:20180298 (2019)

13. Lowen, A. C., & Steel, J. Roles of humidity and temperature in shaping influenza seasonality. *Journal of virology*, 88(14), 7692-7695. (2014)

14. Sil, A. & Kumar, V. N. Does weather affect the growth rate of COVID-19, a study to comprehend transmission dynamics on human health. *J. Saf. Sci. Resil.* **1**, 3–11 (2020).

15. Sarkodie, S. A. & Owusu, P. A. Impact of meteorological factors on COVID-19 pandemic: Evidence from top 20 countries with confirmed cases. *Environ. Res.* **191**, 110101 (2020)

16. Ahlawat, A., Wiedensohler, A. & Mishra, S. K. An overview on the role of relative humidity in airborne transmission of SARS-CoV-2 in indoor environments. *Aerosol Air Qual. Res.* **20**, 1856–1861 (2020).

17. Islam, A. R. T. *et al.* Effect of meteorological factors on COVID-19 cases in Bangladesh. *Environ. Dev. Sustain.*, 1–24 (2020).

18. Rostami, A. et al. SARS-CoV-2 seroprevalence worldwide: a systematic review and meta-analysis. *Clinical Microbiology and Infection*, *27*(3), 331-340 (2021).

19. Smith, T. P. et al. Temperature and population density influence SARS-CoV-2 transmission in the absence of nonpharmaceutical interventions. *PNAS*, *118*(25), e2019284118 (2021)

20. Briz-Redón, Á., & Serrano-Aroca, Á. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Science of the total environment*, *728*, 138811; https://doi.org/10.1016/j.scitotenv.2020.138811 (2020)

21. Gutiérrez-Hernández, O., & García, L. V. (2020). ¿Influyen tiempo y clima en la distribución del nuevo coronavirus (SARS CoV-2)? Una revisión desde una perspectiva biogeográfica. *Investigaciones Geográficas*, (73), 31-55 (2020)

22. European Centre for Disease Prevention and Control. Guidance for discharge and ending isolation of people with COVID-19, https://www.ecdc.europa.eu/en/publications-data/covid-19-guidance-discharge-and-ending-isolation (2020)

23. Martín-Vide, J. El Clima. Geografia General dels Països Catalans (ed. *Enciclopèdia Catalana,* Barcelona, 1992)

24. 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.* 52(4), 1957-1976 (2018)

25. Massey F. J. Jr. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc*. 46, 68–78 (1951)

26. White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*. 48 (4): 817–838 (1980)

27. Ji, Z. G. Hydrodynamics and water quality: modeling rivers, lakes, and estuaries. John 1038 Wiley & Sons (2017)

28. Martins-Filho, P. R. Relationship between population density and COVID-19 incidence and mortality estimates: A county-level analysis. *Journal of infection and public health*, 14(8), 1087 (2021)

29. Henriquez, J., Gonzalo-Almorox, E., Garcia-Goñi, M., and Paolucci, F. The first months of the COVID-19 pandemic in Spain. *Health Policy and Technology*, *9*(4), 560-574 (2020)

30. Güell, B. & Garcés-Mascareñas, B. Agricultural seasonal workers in times of Covid-19 in Spain. *Advancing Alternative Migration Governance* (2020)

31. Liu, X., Huang, J., Li, C., Zhao, Y., Wang, D., Huang, Z., and Yang, K. The role of seasonality in the spread of COVID-19 pandemic. *Environmental research*, *195*, 110874 (2021)

32. Kudo, E., Song, E., Yockey, L. J., Rakib, T., Wong, P. W., Homer, R. J., & Iwasaki, A. Low ambient humidity impairs barrier function and innate resistance against influenza infection. *PNAS*, *116*(22), 10905-10910 (2019)

33. Mu, Y. et al. Transmission of SARS-CoV-2 virus and ambient temperature: a critical review. *Environ Sci Pollut Res* **28,** 37051–37059 (2021)

34. Wu, 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. 729*, 139051 (2020)

35. Chaudhuri, S., Giménez-Adsuar, G., Saez, M., & Barceló, M. A. PandemonCAT: Monitoring the COVID-19 Pandemic in Catalonia, Spain. *International Journal of Environmental Research and Public Health*, *19*(8), 4783 (2022)

36. Dinoi A. 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.* https://doi.org/10.1016/j.scitotenv.2021.151137 (2022)

37. Akaishi T. et al. COVID-19 transmission in group living environments and households. *Sci Rep.* 11(1), 11616 (2021)

38. CDC Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission. https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html (2020)

39. Chen, B., Jia, P. & Han, J. Role of indoor aerosols for COVID-19 viral transmission: a review. *Environ Chem Lett* **19,** 1953–1970 (2021)

40. Salzberger B. et al. Epidemiology of SARS-CoV-2. *Infection.* 49(2), 233-239 (2021)

41. Moriyama, M., Hugentobler, W. J., Iwasaki, A. Seasonality of Respiratory Viral Infections. Annual Review of Virology, 7(1), annurev-virology-012420-022445 (2020)

42. Yang, W., Elankumaran, S. & Marr L. C. Relationship between Humidity and Influenza A Viability in Droplets and Implications for Influenza's Seasonality. *PLoS ONE* 7(10): e46789 (2012)

43. Xie, J. & Zhu, Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Sci. Total Environ.* **724**, 138201 (2020).

44. Pan, J. et al. Warmer weather unlikely to reduce the COVID-19 transmission: an ecological study in 202 locations in 8 countries. *Sci. Total Environ.* **753**, 142272 (2020).

45. Nichols GL, Gillingham EL, Macintyre HL, Vardoulakis S, Hajat S, Sarran CE, Amankwaah D, Phalkey R. Coronavirus seasonality, respiratory infections and weather. *BMC Infect Dis*. 21(1), 1101; https://doi.org/10.1186/s12879-021-06785-2 (2021)

46. Zheng, H. L., Guo, Z. L., Wang, M. L., Yang, C., An, S.Y. & Wu, W. Effects of climate variables on the transmission of COVID-19: a systematic review of 62 ecological studies. *Environ Sci Pollut Res Int*. 28(39), 54299-54316 (2021)

47. Sil, A. & Kumar, V. N. Does weather affect the growth rate of COVID-19, a study to comprehend transmission dynamics on human health. *J. Saf. Sci. Resil.* **1**, 3–11 (2020).

48. Ward, M. P., Xiao, S., & Zhang, Z. The role of climate during the COVID-19 epidemic in New South Wales, Australia. *Transboundary and Emerging Diseases*, 67(6), 2313-2317 (2020)

49. Tang, L., Liu, M., Ren, B., Wu, Z., Yu, X., Peng, C. & Tian, J. 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*. 751:141816 (2021)

50. Baker, R. E., Yang, W., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B.T. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science*. 369(6501), 315-319 (2020)

Figures

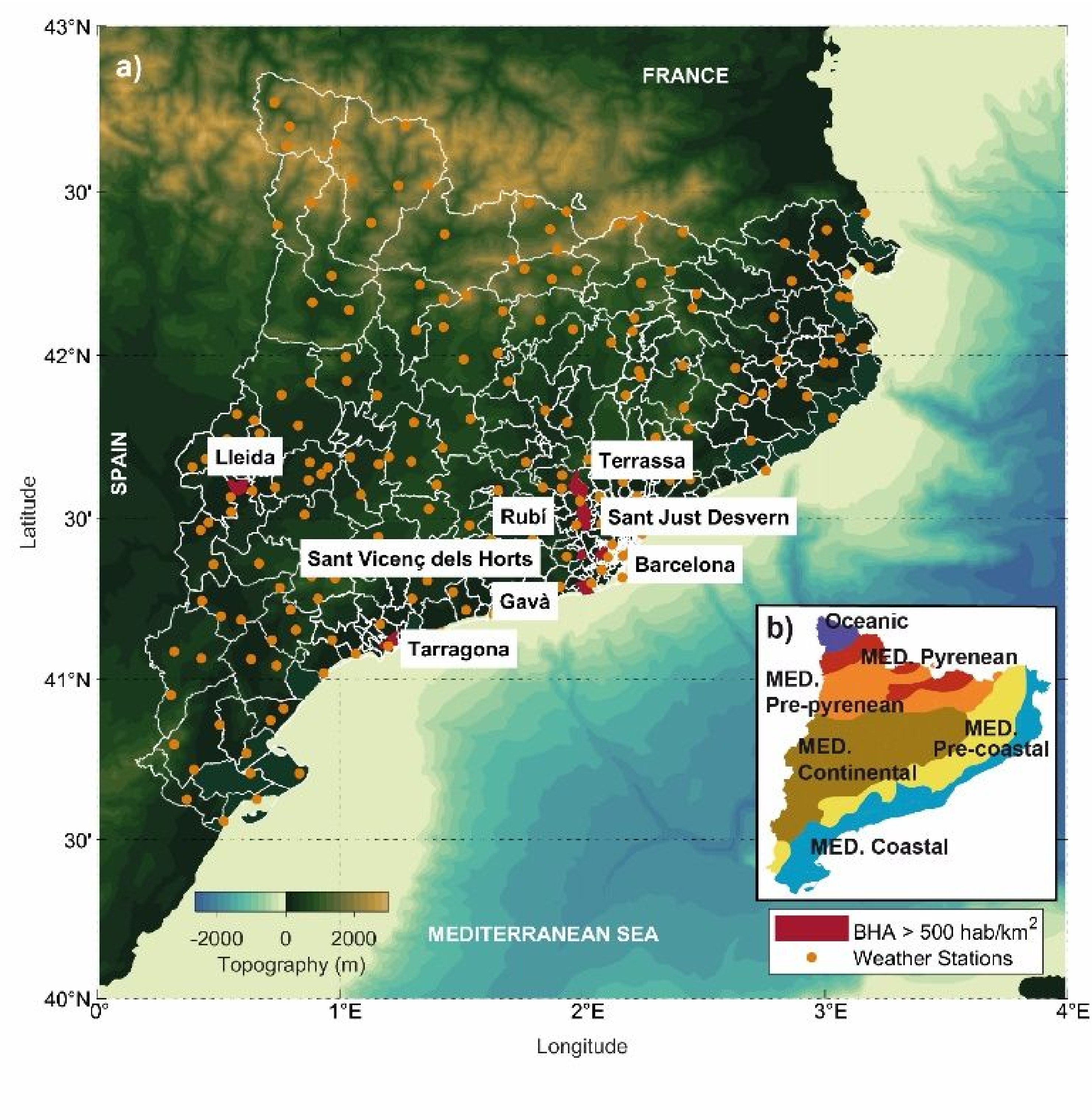

Figure 1

(a) Basic health areas (BHAs, delimited in white) and automatic weather stations (orange dots) in Catalonia. Those BHA selected for this study, with a population density d ³ 500 inhab km^2, are drawn in red. (b) Bioclimates in Catalonia according to the climatic conditions: Mediterranean coastal, Mediterranean pre-coastal, Mediterranean continental, Mediterranean pre-Pyrenean, Mediterranean Pyrenean and Oceanic[23] published on the Meteorological Service of Catalonia (MSC).

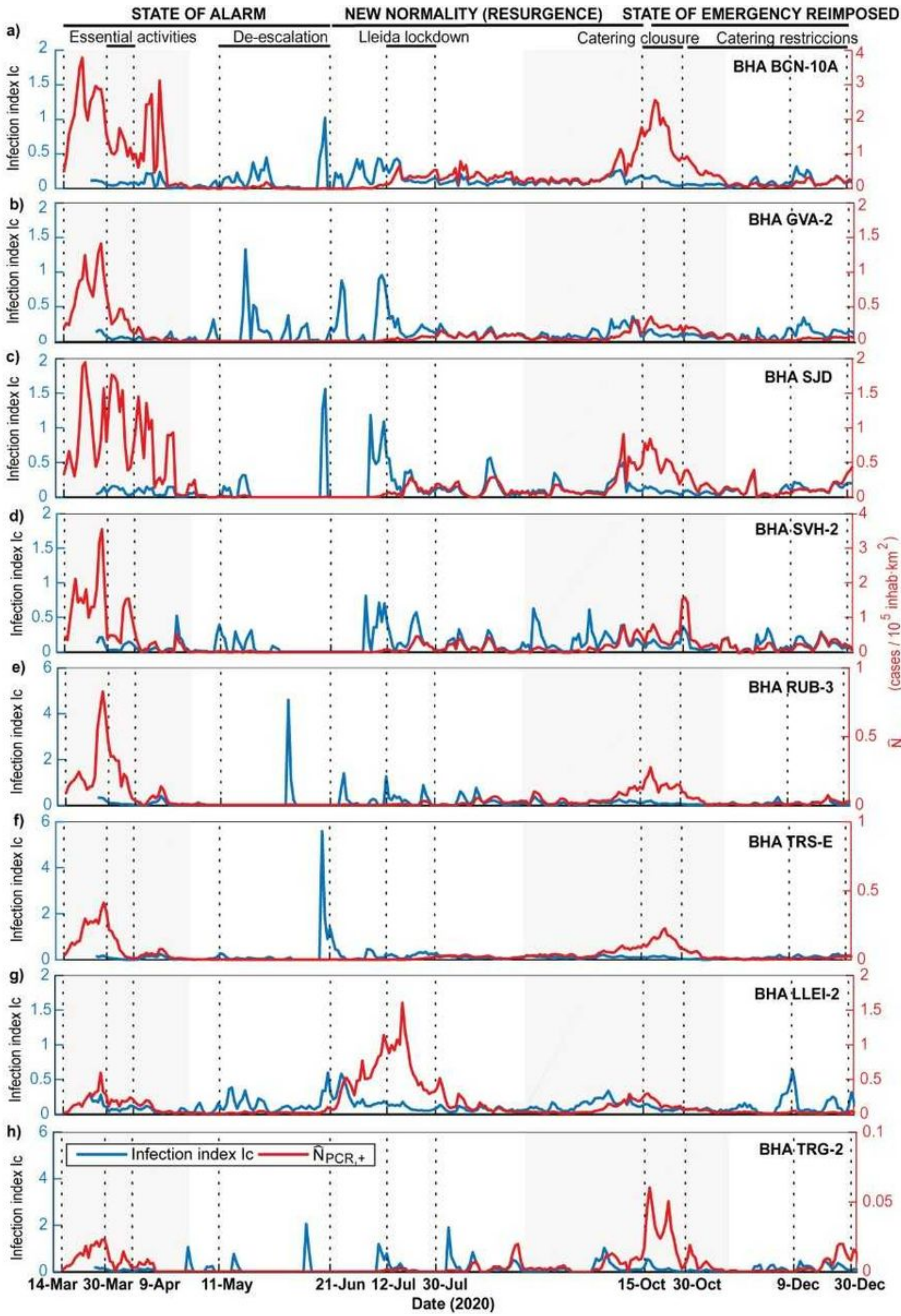

Figure 2

(a-h) Temporal evolution during 2020 of the number of COVID-19 cases in the eight selected Catalan BHAs, presented both in terms of the normalised number of cases as determined through the PCR positive tests (red line) and the rate of infections I_c (blue line). In each panel, the grey shaded areas indicate the first and the second pandemic wave periods and the vertical dotted lines delineate the duration of the main social and mobility restrictions imposed in Catalonia due to the COVID-19 disease.

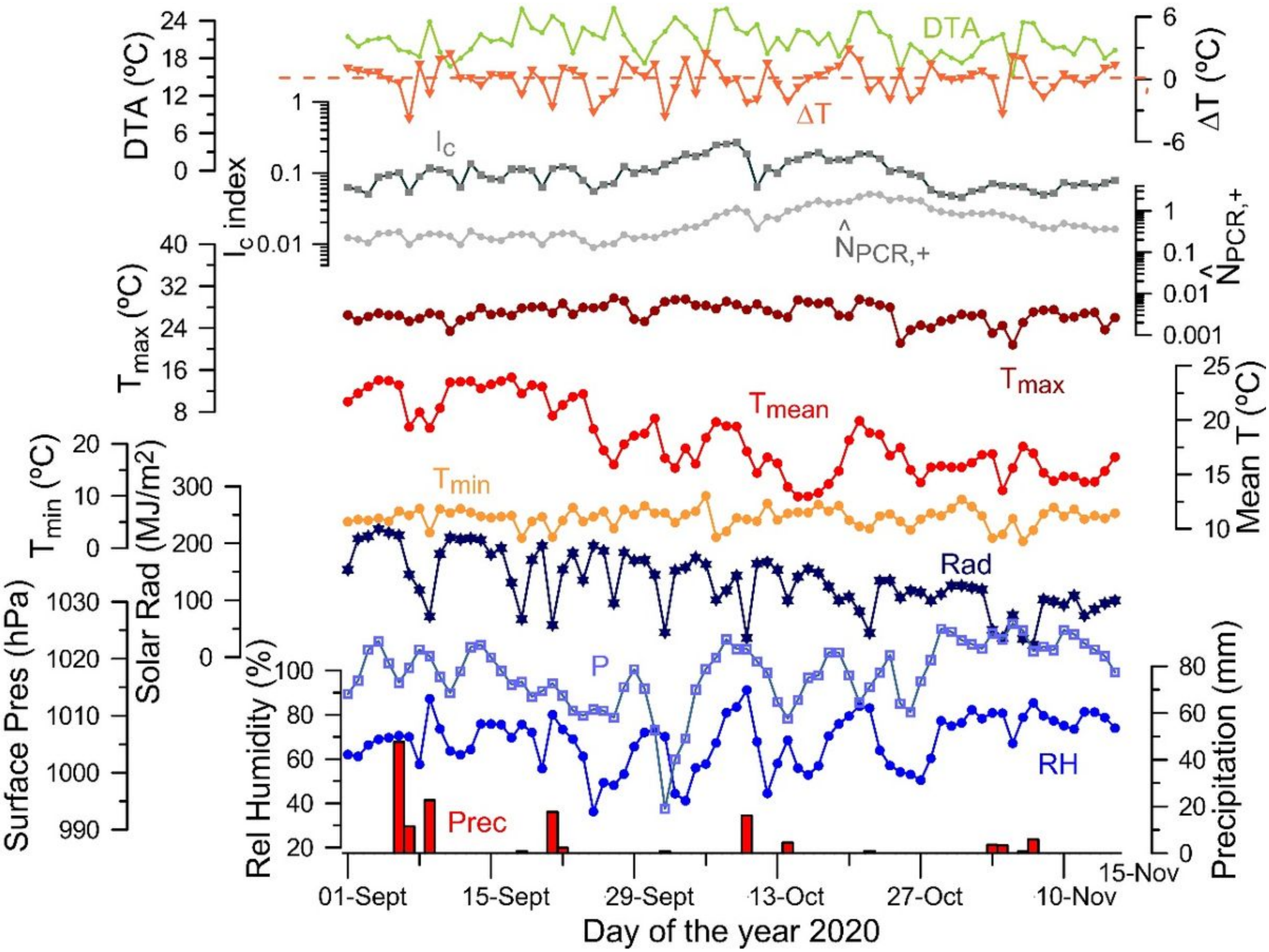

Figure 3

Time series of daily values of atmospheric variables during the second outbreak (September 1 to November 15) together with the normalized number of cases () 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 mean temperature difference between consecutive days (D *T*) . The units for these variables are indicated in their corresponding axes.

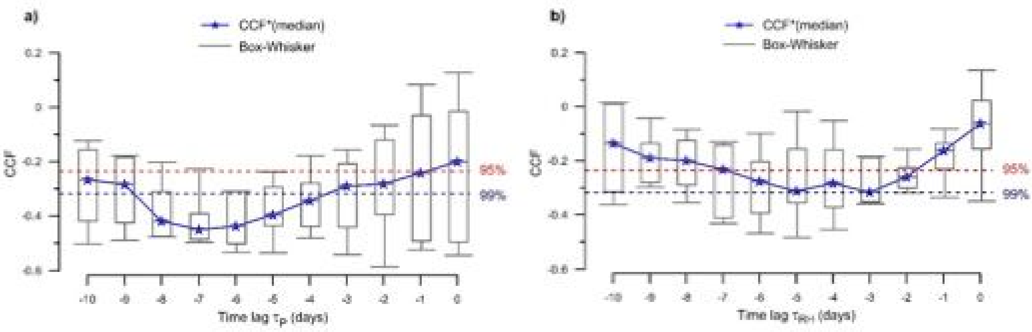

Figure 4

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 median (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 95% (red line) and 99% (blue line) confidence levels.

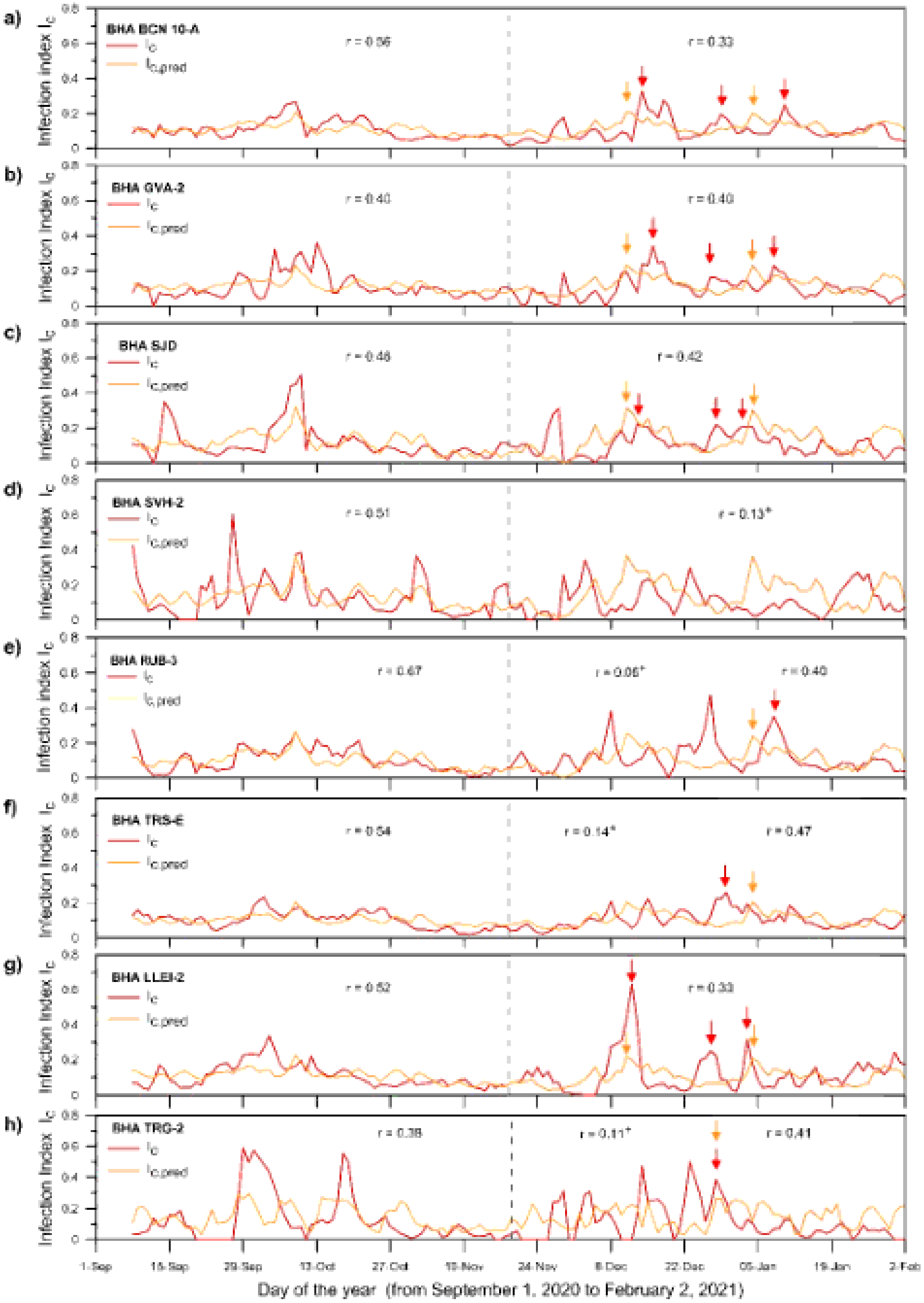

Figure 5

Results for the eight BHAs (panels a) through h)) showing the temporal evolution of the observed infection I_c index (red lines) and the corresponding predicted $I_{c,pred}$ values (light lines) as obtained from the climate model between September 1, 2020, and February 2, 2021. The vertical dashed line on November 18, 2020, delineates the setup and forecast periods. The grey 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 minimum 90% confidence level; in several cases it corresponds to time periods of low normalized number of cases (see Fig. S7, Supplementary information). The arrows indicate the main and the secondary observed (red) and predicted (light orange) I_c peaks during the forecast periods.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PlanellaMoratoetalsuppmaterials1.docx