Check for updates

# Climatic signatures in the different COVID-19 pandemic waves across both hemispheres

Alejandro Fontal [1], Menno J. Bouma[1], Adrià San-José[1], Leonardo López[1], Mercedes Pascual[2] and Xavier Rodó [1,3] ✉

The roles of climate and true seasonal signatures in the epidemiology of emergent pathogens, and that of SARS-CoV-2 in particular, remain poorly understood. With a statistical method designed to detect transitory associations, we show, for COVID-19 cases, strong consistent negative effects of both temperature and absolute humidity at large spatial scales. At finer spatial resolutions, we substantiate these connections during the seasonal rise and fall of COVID-19. Strong disease responses are identified in the first two waves, suggesting clear ranges for temperature and absolute humidity that are similar to those formerly described for seasonal influenza. For COVID-19, in all studied regions and pandemic waves, a process-based model that incorporates a temperature-dependent transmission rate outperforms baseline formulations with no driver or a sinusoidal seasonality. Our results, so far, classify COVID-19 as a seasonal low-temperature infection and suggest an important contribution of the airborne pathway in the transmission of SARS-CoV-2, with implications for the control measures we discuss.

G iven the periodic influenza pandemics in the twentieth century—the so-called Spanish flu in 1918, Asian flu in 1957 and Hong Kong flu in 1968—the medical community has been apprehensive about the emergence of novel pathogens with an extensive reach in a more connected world[1]. Unlike influenza, with its well appreciated ability to mutate[2], the few relatively benign coronaviruses did not originally cause alarm. This was the case until the 2003 outbreak of severe acute respiratory syndrome (SARS) in China[3] and Middle East respiratory syndrome (MERS) in Saudi Arabia in 2012[4], which featured severe symptoms and high mortality. Both outbreaks were short-lived, and isolation of cases proved sufficient to overcome the restricted ability of these pathogens to transmit between humans. The emergence of the novel SARS-CoV-2, with a similar geographic and zoonotic origin as SARS[5], was declared a pandemic by the World Health Organization (WHO) in March 2020[6] and has rapidly spread over extensive regions. So far, over 220 million cases have been recorded, with over 4.7 million deaths[7] and a high mortality, not unlike influenza, in older age groups[8].

Studies of the 2009 H1N1 influenza viruses suggested that low relative humidity (RH) and high population density shaped the early spread of the virus. Those conditions also favored persistence of the virus during the weeks at the beginning of the pandemic, underscoring an increasingly recognized role of humidity in influenza dynamics. Climatic factors were deemed relevant when designing control and prevention measures[9]. Determining the role of climate drivers in the transmission dynamics of emergent pathogens remains fraught with difficulties as a result of limited data, nonlinear responses and multiple potential confounding factors at the early stages of pandemics[10]. Examples of such factors include processes sharing trends, variable intervention measures such as lockdowns, and the rising use of non-pharmacological interventions. An ample availability of susceptible individuals fueling transmission is also thought to lessen the importance of climate conditions[11]. The results of a mathematical model addressing the potential for seasonality in the population dynamics of COVID-19 suggested an inability of

climate forcing to establish seasonality given the large number of available susceptible, non-immune individuals. Nevertheless, initial observations suggested that SARS-CoV-2 could be a seasonal disease, a possibility that was not advanced for the short-lived outbreaks of SARS and MERS. Specifically, the initial propagation of COVID-19 emerged in a latitudinal band between 30° N and 50° N, with low humidity levels and temperatures between 5 °C and 11 °C (ref. [12]), and these weather sensitivities were reported from China, with absolute humidity (AH) negatively associated with daily death counts of COVID-19[13]. SARS-CoV-2 further resembled influenza with respect to its winter appearance. Recent perspectives have reviewed and evaluated the evidence for climate-driven seasonality, indicating contradictory results and confounding effects of other factors at these early stages of the pandemic[10,14].

At a more mechanistic level, some observations on the long-range transmission of SARS in 2003[15] raised the scope for aerial transmission and a role of aerosols[3], which would bypass short-range control measures. The intangible connection between airborne and seasonal transmission has been implicit in the scientific interest that followed the SARS 2003 outbreak, particularly for influenza. Both the nature of seasonality and the plausible substantial contribution of airborne transmission were explored[16–18], and attempts were made to mechanistically connect temperature and humidity to the environmental persistence of bio-aerosols and seasonality. Strikingly, the results of these influenza studies on aerosols are similar to those of initial SARS-CoV-2 investigations[19], implying a possible contribution of aerosols to long-range transmission. With the third—and in some cases the fifth—pandemic wave under way in the Northern Hemisphere, and the need to project future dynamics under vaccination and the different duration of immunity, the question of whether COVID-19 is a genuine seasonal disease becomes increasingly central, with implications for determining judicial intervention measures. More generally, the answer can inform our understanding of airborne transmission in other respiratory viruses.

Retrospective consideration of the first and second waves provides an opportunity to address climate drivers influencing the

rising and waning phases of the epidemic, even in the presence of containment measures. A statistical method for transient correlations (scale-dependent correlation analysis, or SDC) is applied to the time series of reported COVID-19 cases to identify similar temporal variation to that of temperature and humidity over localized windows of time. The consistency of the results across temporal and spatial scales is also examined. Transitory associations can arise from nonlinear responses involving thresholds, whereby a climate factor acts as a more dominant limiting factor of transmission intensity in a given critical range. Because of its ability to detect small localized changes in epidemic shape, the statistical approach is able to uncover climate effects, despite confounding public health interventions and the very limited effect of trends at these scales (Methods)[20–22]. The new multiscale SDC (MSDC) implementation provides an opportunity to further confirm the consistency of associations, assessing their stability across scales. As a comparison, preferential meteorological ranges are also derived with the same approach for seasonal influenza in Japan. Finally, a complementary analysis of COVID-19 is implemented with a process-based epidemiological model for which the transmission rate is driven by climate. Its ability to capture observed cases and deaths is compared to that of alternate formulations with either a constant or a seasonal transmission term. We replicate this analysis at two different spatial scales—regional and city levels.

## Results

**Global and country-level role of temperature in COVID-19.** To address the association of temperature ($T$) and AH in the initial phase of the invasion by SARS-CoV-2 and before changes in human behavior and public health policies were put into place, we first considered a measure of the initial growth rate of cases across countries. For this, we relied on an estimate of the reproductive number of the disease, $R_0$, in the 20 days after the first 20 cases of COVID-19 were officially notified for each of 162 countries spanning five continents (Extended Data Fig. 1, Methods and Supplementary Information). The global maps for the climate variables at those corresponding times in each country and for population density are shown in Extended Data Fig. 2. A negative relationship is observed between $R_0$ and $T$, as well as AH, at the global scale on the basis of linear pairwise regression models (Extended Data Fig. 1b,c and Extended Data Fig. 2). Although these regressions account for 10–25% of the variance in $R_0$, the two regressions are consistent and statistically significant ($P < 0.001$), irrespective of the COVID-19 measure employed (for example, $R_0$ or cumulative cases; see also sensitivity tests in Extended Data Fig. 3).

To address whether the negative relationship with both $T$ and AH is merely the result of a coincident timing between trends in epidemic spread and the dominant wintertime conditions in temperate regions, we analyzed the evolution of the disease–climate associations locally in time and at several disaggregated spatial scales for groups of countries in all five continents (Fig. 1). To this end, we relied on a statistical method, SDC, which was specifically developed to identify similar patterns of variation by means of local or transitory (linear) correlations between two time series given a moving window of time (Methods and refs. [20–22]). Application of this analysis to other diseases known to be climate-sensitive (for example, cholera, malaria and Kawasaki disease) has successfully shown that associations can be discontinuous in time, alternating temporal intervals with highly significant correlations and those with low or non-existent ones[23,24]. Such transient coupling can be expected in the response of nonlinear systems to an external driver, especially when the functions relating particular parameters to the driver are themselves nonlinear, so that effects are more evident in particular ranges of the forcing variable. SDC is essentially a pattern-recognition tool that allows one to consider the scale over which to evaluate correlations, as the window size can be systematically varied

to focus on increasingly local patterns of variation, for example (Methods)[21]. A new extension of this method, MSDC, is presented in this study to examine the stability of correlations across all scales.

Figure 1 shows the results of SDC analyses for COVID-19 cases and $T$ for time windows of two and a half months (a scale of $s = 75$ days) when grouping countries according to region and latitude in the two hemispheres. Similar patterns are found for cases and AH (Extended Data Fig. 4). Strong negative transient associations are obtained for short time lags between the disease and climate time series, with consistent patterns worldwide. Interestingly, the negative relationship is seen during both the first and second waves of the pandemic, and for both the rising and declining phases, with a break during summertime in all continents (Fig. 1 and Extended Data Fig. 4). Transient positive correlations of varying intensity are also detected during the warmer months across locations, varying in intensity and not consistently high. Whether these reflect real patterns linked to mass gatherings of young people in vacation resorts, as in Spain in the summer 2021, is discussed later. To further examine the association patterns, we next considered the reported cases at the smaller spatial scale of individual countries in Europe (Fig. 2, for France, the United Kingdom, Italy, Spain and Germany), the first most affected continent following the emergence of the virus in China. SDC results show similar transient and negative associations for $T$ and AH with COVID-19 cases ($s = 75$ days). The negative relationship occurs largely in synchrony across the different countries, for the same time intervals during the waning of the first epidemic wave as $T$ and AH rise, and also during the rise of the second wave in the fall as $T$ and AH fall, with a break in between. For this same temporal window ($s = 75$), a similar negative relationship holds also for individual regions within three of these highly affected European countries for which data at the higher spatial resolution of individual provinces are available (namely, Lombardy, Thüringen and Catalonia, Supplementary Figs. 1 and 2). These results are also shown separately for the first and second waves in Lombardy, Thüringen and Catalonia (Supplementary Fig. 2). Locally in time, these associations account for large fractions of the variability in COVID-19 cases in the three regions (over 80% at scales of $s = 21$ days; Supplementary Fig. 2). Their discontinuity explains the lower values obtained with the traditional correlation coefficient, since by definition this quantity averages over the whole length of the time series[20].

We can also inspect how these transient couplings change as a function of the temporal lag (in days) between the two time series. As the magnitude of the change in the climate variable $T$ or AH increases, the lag tends to shorten. As shown in Fig. 2b, this behavior is seen for the direction of change that is presumably relevant for influencing transmission (increasing $T$ and AH in the fall of the first wave and decreasing $T$ and AH in the rise of the second wave). The shorter lags may reflect a faster speed of community transmission, concomitant with a more intense effect of the meteorological factors. Zooming in even more locally in time with a window of about two weeks ($s = 14$), SDC reveals limited intervals of very highly significant correlations at these regional levels for Lombardy (Italy; Fig. 3) and for Catalonia, Spain (Extended Data Fig. 5). Intervals where the climate covariate accounts for over 80–90% of the variability in cases alternate with complete decoupling between the time series. Interestingly, the strongest negative associations are observed during the rise of the second wave in these European regions. Positive associations are weaker and infrequent at this scale.

We next interrogate our SDC results to address the existence of threshold values in $T$ and AH for the spread of COVID-19. To this end, we examine whether intervals of coupling become more frequent depending on the range of the climate covariates. This behavior would indicate the existence of a critical range within which responses are heightened and, conversely, outside which the forcing becomes ineffective. As an example, a meta-analysis was
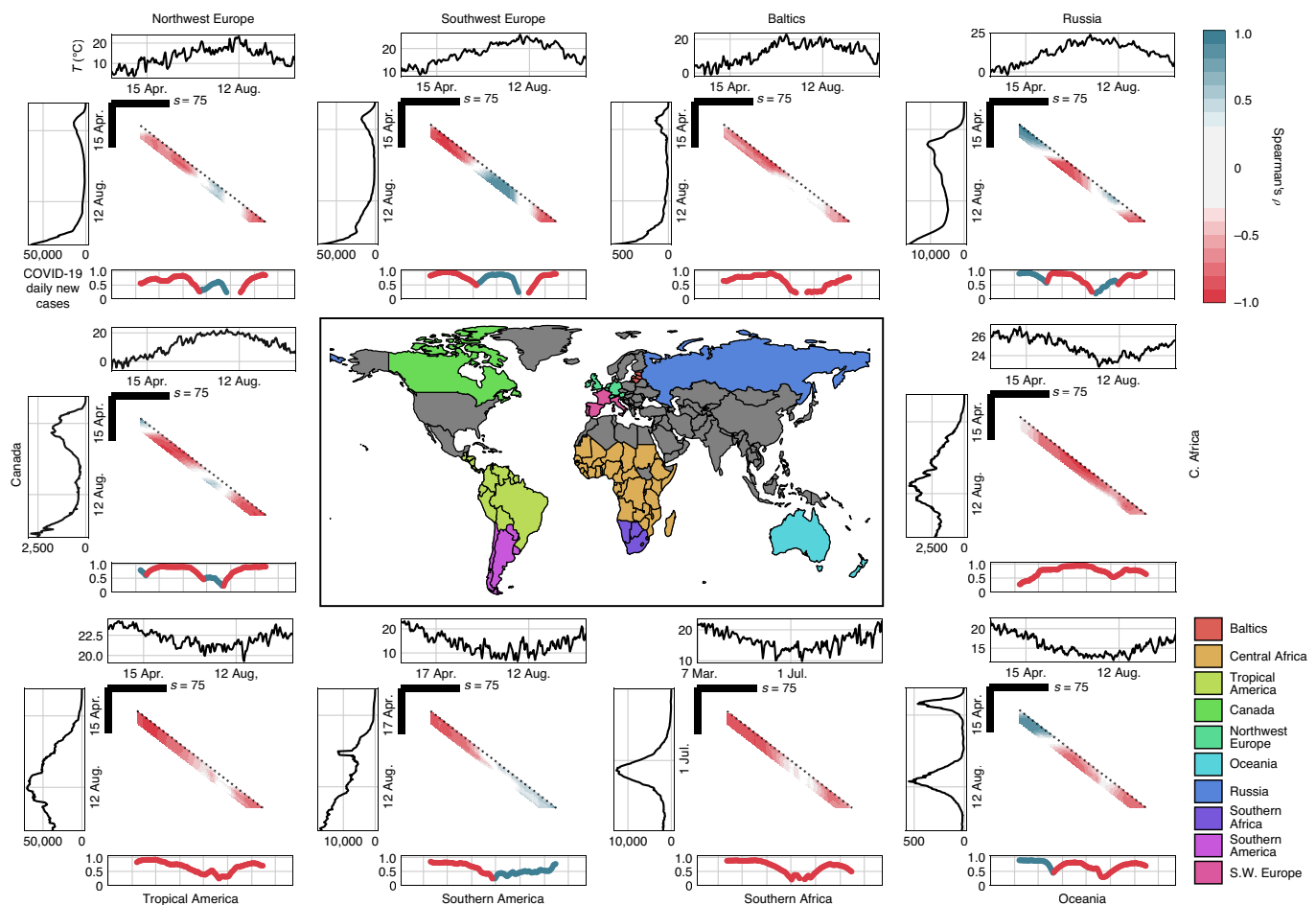
**Fig. 1 | SDC analyses for the aggregated COVID-19 cases and temperature for groups of contiguous countries.** Two-way SDC analyses for the time series of COVID-19 daily new cases versus mean temperature $T$ in groups of contiguous countries. These daily climate conditions are weighted averages based on the relative population of the relevant countries from March to October 2020. The temporal scale or window size, $s$, used for the local correlations is 75 days, and the lag between the placement of the windows in the two time series varies from 0 to 21 days. In the SDC plots, the two time series are shown respectively to the left and top of the central correlation grid, for reported cases (left plots) with time running down and for the climate factor (top plots) with time running right. Each cell of the grid is colored according to the Spearman correlation coefficient, with the row and column corresponding to the position of the two respective time windows of size $s$ along each of the time series, with the lag between these positions corresponding to the distances from the diagonal (lag 0 to +21 days). The panel below each plot shows the maximal correlation coefficient obtained at each time point (vertically, and therefore relative to the time of the climate time series). Only correlations found to be significant in a non-parametric randomized test ($\alpha = 0.05$) are shown and colored. Red (blue) points in the grid correspond to negative (positive) correlations, according to the specified scale (top right).

performed for Italy, where detailed regional case data are available. We jointly analyzed all SDC correlations between COVID-19 cases and each of the climate covariates ($T$ or AH) by pooling together the results for all the individual regions of Italy (Fig. 3, Supplementary Fig. 5 and Methods; for Spain, results are shown in Extended Data Fig. 6). The resulting distributions, shown in Supplementary Fig. 4 and Extended Data Fig. 6, show the proportion of the total possible comparisons (pairs of time intervals between the two time series) that fall into a given range of the climate covariate, with those proportions subdivided into significant positive and negative correlations (Fig. 3e,f) and including non-significant ones (Supplementary Fig. 4). Distributions are also shown for only the positive and negative correlations along the range of climate intervals (Fig. 3e,f), and for the first two waves separately (Supplementary Figs. 5 and 6 and Extended Data Fig. 6). The mode in these distributions indicates preferential intervals for $T$ and AH, respectively, where the climate effects are most evident and beyond which decoupling is likely (roughly 12–18 °C and 4–12 g m$^{-3}$, respectively; Fig. 3g,h). Although similar results are obtained for the first and second waves separately, these ranges

should only be seen as indicative, given the short records yet available (Supplementary Fig. 6 and Extended Data Fig. 6).

To further examine the consistency of the identified relationships with $T$ and AH, we applied MSDC—an extension of SDC developed to inspect the evolution of transient correlations at all scales at once in the same graphical display (Methods)—to countries in other continents that experienced a later arrival of SARS-CoV-2 than Europe, namely South Africa, Argentina and Canada, from February to December 2020 (Extended Data Fig. 7). The top and central plots in Extended Data Fig. 7 for each country present, respectively, the most negative and the most positive significant Spearman correlation found for every pair of fragments of a given window size and date. (To visualize timing with respect to the epidemic progression, the bottom plot displays the seven days moving average of daily reported new cases for each of the countries.) Negative correlations exhibit continuity as scale $s$ is increased, and also transition to larger values (indications of causal effects). By contrast, positive correlations are much more limited in the range of scales, largely restricted to the very small scales and appearing either at times of minima or maxima of COVID-19 incidence, but not during the rise or the
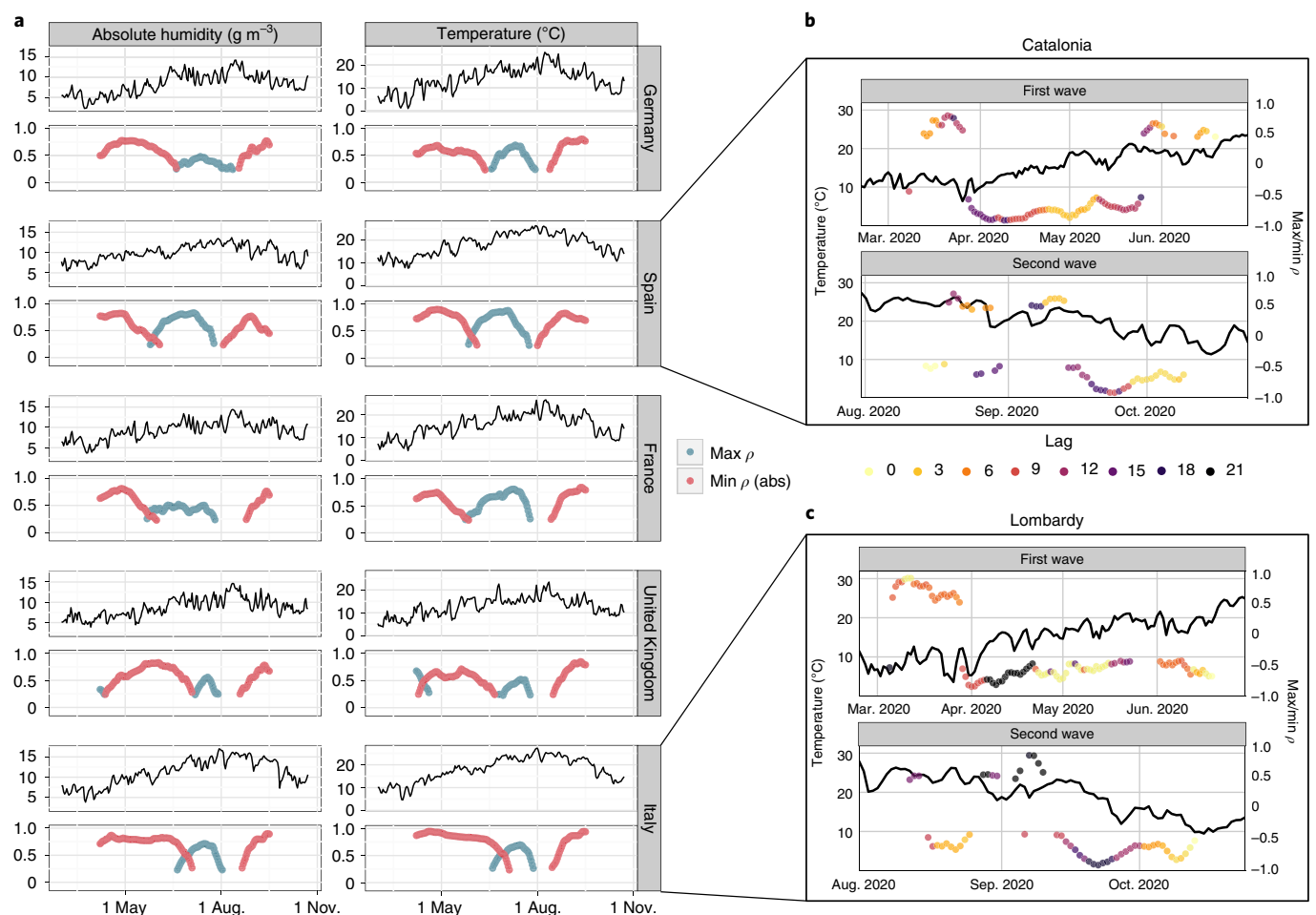
**Fig. 2 | Transitory correlations between meteorological factors and COVID-19 incidence for several European countries. a**, Panels from SDC analysis for the maximum absolute correlations between daily time series for the meteorological variable (AH, left; $T$, right) and COVID-19 new cases, as a function of time (the position of the local window of time). These panels correspond to those shown below the SDC grid of local correlations in Fig. 1, but here at the country level for five European countries (Germany, France, Italy, Spain and the United Kingdom). As before, $s = 75$, and the maximum correlation is chosen for lags from 0 to 21 days. The colors correspond to maximum negative (red) and positive (blue) correlations for each day. **b**, The resulting maximum correlations for the Spanish region of Catalonia with the corresponding time lag indicated. This dominant lag (the one for the maximum $\rho$ value among the significant comparisons between a lag at 0 and +21 days) is overlaid on the temperature time series for each time point, with shorter lags being represented with lighter colors and longer lags with darker ones. The fragment size ($s$) used in this panel is 21 days, which allows us to localize and identify multiple shorter transient couplings between COVID-19 incidence and meteorological factors. **c**, As in **b**, but for the Italian region of Lombardy. Note that, in **b** and **c**, when the meteorological factor appears to be more related to the transmission of SARS-CoV-2 and the spread of COVID-19 in the population, the lag shortens, indicating a more intense and fast association.

fall of the pandemic waves. Total attributed variability is therefore large for negative correlations, but much more limited for the positive patterns. When the positive correlations are only present at the smallest scales and do not persist as we move up in scale, they most probably reflect spurious associations arising from random fluctuations (Extended Data Fig. 7).

**A mechanistic model incorporating explicit temperature data.**
To further evaluate the role of climate factors on disease dynamics from a more mechanistic perspective, we also implemented a process-based stochastic model that incorporates an effect of temperature in the transmission rate. The model structure and corresponding parameters are shown in Fig. 4 (see Methods for additional description). The basic model, which divides the human population into seven compartments (susceptible, exposed, confined, infected, quarantined, recovered and dead individuals), already exhibited good performance in terms of predictive skill for the first waves of the COVID-19 pandemic in different countries[25].

Here, we extend its formulation to include temperature as a driver in the transmission rate. For simplicity, $T$ was chosen for this purpose rather than AH based on its stronger relationship with COVID-19 in the above SDC results, and given the similarity of the temporal patterns of the two climate variables. For the purpose of comparison, the baseline formulation omits temperature and includes a constant transmission rate. A seasonally forced version includes a sinusoidal variation of this term with a period of 12 months. The model incorporating $T$ is better able to capture the data for all the different waves and regions (Figs. 4 and 5 and Supplementary Table 1). Figure 4c,d shows model fits for the three different models for recovered cases, active cases and deaths in Catalonia and Lombardy, respectively. Comparisons of the mean squared error for these models are displayed in Fig. 5 and Extended Data Fig. 8. To address the different spatial scales of aggregation, the same mechanistic model was fitted to the epidemic data for the city of Barcelona (Fig. 5a,b and Extended Data Fig. 8). Here, too, the model incorporating $T$ shows better performance overall
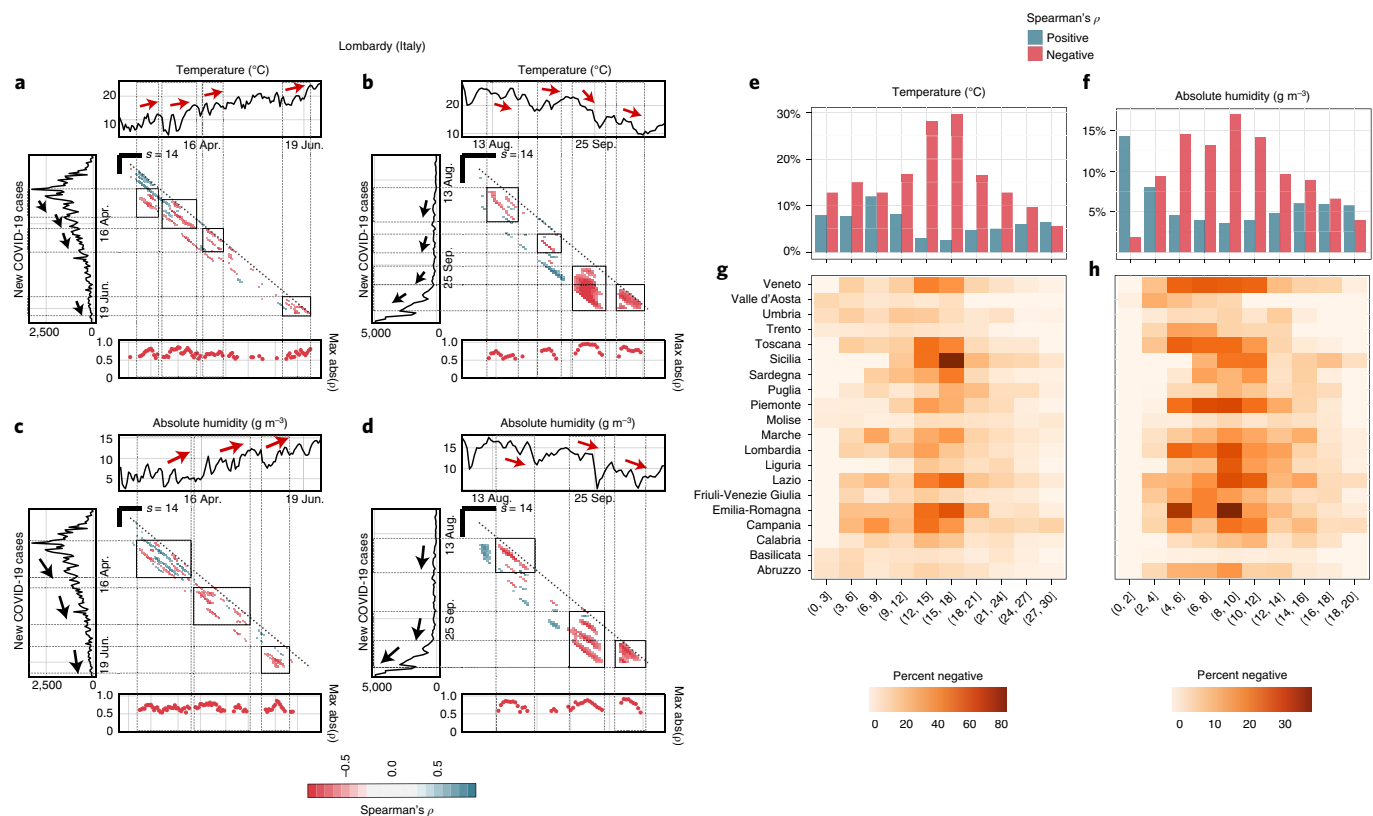
**Fig. 3 | Analysis of the effect of *T* and AH in the spread of COVID-19 in Italy. a–d,** Short-term transient correlations over time intervals of two weeks at the spatial resolution of region, here for Lombardy in Italy. As for Fig. 1, the graphical results of SDC analysis are shown for the correlations between daily COVID-19 cases and *T* (**a,b**) or AH (**c,d**) and for the different parts of the pandemic. Here, a smaller window size is used ($s = 14$ days) to examine even shorter-term transient associations. The panels in **a** and **c** showcase the results for the waning phase of the first wave (from late February to July), and the panels in **b** and **d** those for the rising phase of the second wave (from August to late October). The time intervals of significant correlations are indicated with boxes within the central grid, with the corresponding timings in the two time series highlighted with red arrows for the meteorological factor and with black arrows for the response variable, COVID-19 cases. The pairs of arrows for a given box show clear patterns of opposite temporal change (negative correlation) in the climate and epidemiological variables. For clarity, in this case the bottom panels below each plot show only the maximum negative correlation values obtained as the analysis window is moved in time (with red dots indicating negative correlations). **e,f,** Frequency distributions of negative and positive correlations as a function of range of the climate covariate (numbers in parentheses), for *T* (**e**) and AH (**f**). The frequency of windows with negative (or positive) correlations is obtained from SDC analysis of COVID-19 cases and the given climate factor in all Italian regions ($n = 20$; for $s = 21$ days). The percentages correspond to the percentages of total comparisons (within each of the ranges of the climate variables) that were statistically significant in either of the directions. The distributions are non-uniform with a mode in a particular range of the climate variables. This is indicative of a nonlinear relationship where the negative associations are most likely to be observed in a particular meteorological range. **g,h,** This same frequency distribution but across the different Italian regions. As can be seen, there is an overrepresentation of significant negative monotonic relationships in temperate conditions (AH of 4–12 g m⁻³, *T* of 12–18 °C).

(Fig. 4c,d) and lower residual estimates than any other model (Fig. 5a,b and Supplementary Table 1). In particular, the fitting to the first wave appears clearly improved with respect to its counterparts, possibly indicating a stronger temperature control for this early epidemic stage. Although notable effects are seen also for the other waves (and are most apparent for the active cases), the improvements are modest in the second wave for Lombardy (Figs. 4c,d and 5a,d), Thüringen (Fig. 5a,b) and in the second and third waves for Catalonia (Figs. 4c and 5a,b) and Barcelona (Fig. 5a,b) (also Extended Data Fig. 8). These results are in strong agreement with those obtained with SDC analysis on the enhanced extraction of a *T* signature in this first COVID-19 pandemic wave.

Finally, the cutoff analysis conducted for COVID-19 (Supplementary Fig. 6) was also applied to records of seasonal influenza (A and B) in the city of Kawasaki in Japan from 2014 to 2021 (Fig. 6b). Seasonal influenza is a better known respiratory disease for which climate effects have been documented, in particular in relation to humidity (Extended Data Fig. 9). Transient negative

correlations are clearly evident for influenza A and influenza B with both AH and *T* (Extended Data Fig. 9). As for COVID-19, the number of positive correlations is, by comparison, minimal (Extended Data Fig. 9). The ranges or thresholds for *T* and AH derived from the SDC analysis indicate (1) a slight systematic delay of up to one or two months in the occurrence of influenza B with respect to influenza A (Extended Data Fig. 9) and (2) *T* and AH relevant ranges similar to those obtained for COVID-19 (Fig. 6b). Interestingly, these ranges are slightly different between the two viral strains, suggesting small differences in climatic niches.

## Discussion
Taken together, our findings support the existence of strong external drivers of transmission intensity, as suggested by a uniform summer recession despite a variety of intervention measures across countries[25]. They also support the view of COVID-19 as a true seasonal infection, similar in that sense to seasonal influenza and to the more benign circulating coronaviruses[26].
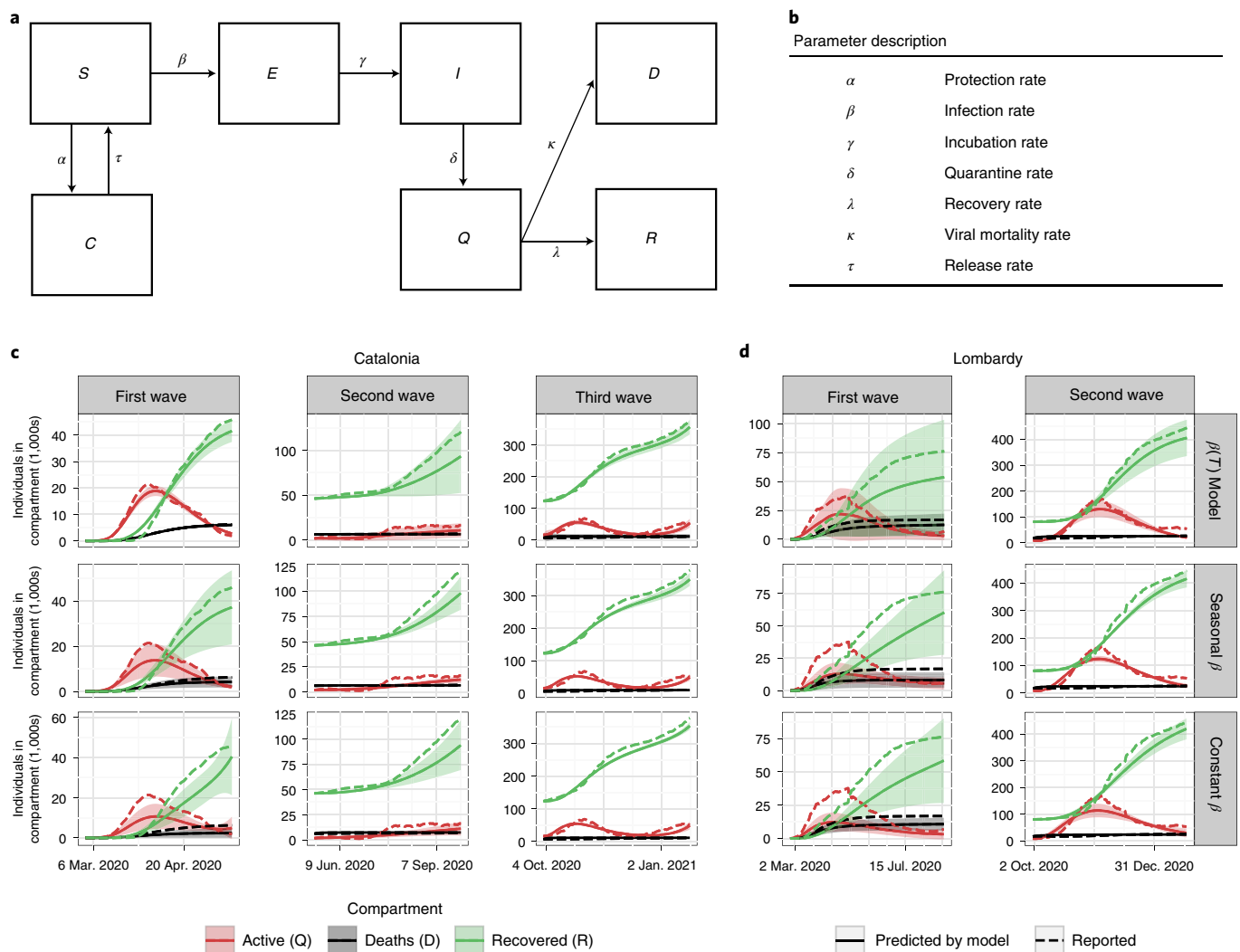
**Fig. 4 | Compartmental process-based model diagram and fits. a,b**, Diagram summarizing the structure of the compartmental process-based model (**a**) and table of the parameters for the transitions between these compartments (**b**). **c,d**, Model fitted values (solid lines, with the shaded areas representing the standard deviations (s.d.) of the predictions) compared to the officially reported values (dashed lines) for the active, recovered and deaths compartments in the regions of Catalonia (**c**) and Lombardy (**d**).

Concomitant temporal variation of ecological conditions, human behavior and disease incidence over the annual cycle can produce true but also spurious correlations with climate parameters[21]. Our results are consistent with those of ref. [27] on an effect of temperature and specific humidity on the spread of COVID-19, which is certainly encouraging given that the methods applied in this work (SDC) and their work (distributed-lag nonlinear models) are different. Our analysis is also complementary in considering data from different countries and spatial scales. The analyses in our study were formulated to circumvent the spurious correlations arising from such coincident seasonal variation, as well as effects of potential confounders (Supplementary Discussion). A similar approach could contribute to the further understanding of the complex global seasonality of influenza[28,29].

Further support for a transient but strong role of climate in the modulation of COVID-19 transmission is provided here by the results of the implemented mechanistic model. Model fits and model comparisons consistently indicate a climate effect across the different epidemic waves, for different regions and individual cities (Fig. 5 and Extended Data Fig. 8). The results for Barcelona are fully aligned with those for Catalonia and other regions. When inspecting results by waves, this observation is further reinforced, as the

mean squared error for the temperature-driven model is always the lowest, especially for the first epidemic wave. Similarly, increasing the scale resolution does not modify the results by SDC at the largest spatial scales considered (those of region and country).

The problem of co-linearity in climate factors has challenged the search for environmental drivers of influenza in the past, and very similar issues now apply to COVID-19[27]. Both $T$ and AH appear here equally capable of accounting for the observed COVID-19 variation. It is well known that air of rising temperature can contain more water, and for geophysical reasons these two parameters are usually strongly coupled over extensive spatial scales. Also, both parameters can act in the same direction in their biological effects. For example, the environmental persistence of an exhaled infectious virus particle—associated with a small (<5 μm) or a larger (100 μm) droplet[30]—it increases under laboratory conditions with lower temperatures and lower humidity for both influenza[31], SARS[32] and SARS-CoV-2[33]. Although prolonged environmental survival at lower temperatures applies to most of the viruses that have been studied[34], a less humid environment shortens the survival for most upper airway infections such as the rhino- and adenoviruses that cause the common cold[35]. Importantly, however, these viruses lack an envelope and are not seasonally defined[36]. By contrast, enveloped
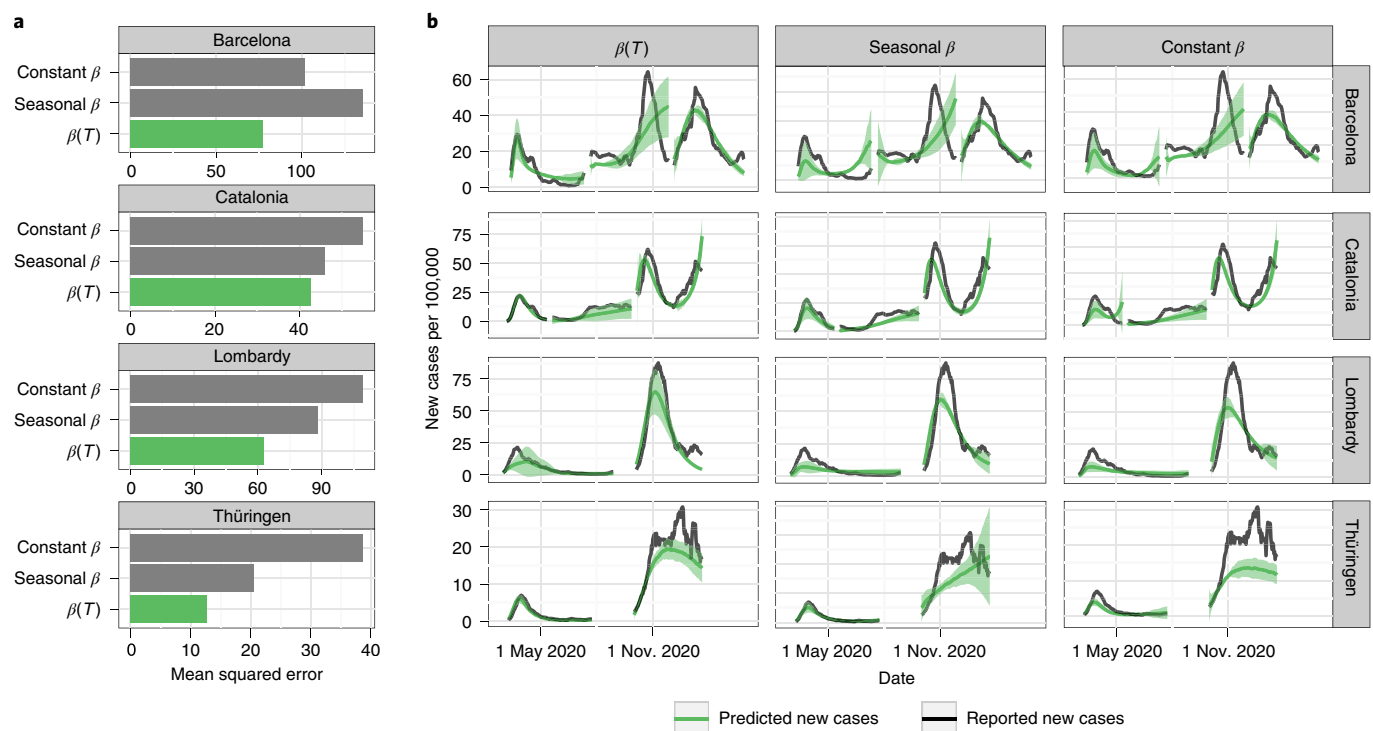
**Fig. 5 | Goodness-of-fit comparison among the different model parameters.** Comparison of the compartmental process-based model fits for the different regions, periods and parameters used. **a**, The mean squared error for the models fitted to the data across all periods and for each of the regions. The values corresponding to the models with the temperature-dependent transmission rate $\beta$ are highlighted in green. **b**, The model predicted curves (in green) corresponding to the mean of 100 stochastic simulations are compared to the seven-day rolling mean of reported daily new cases in each region (in gray).
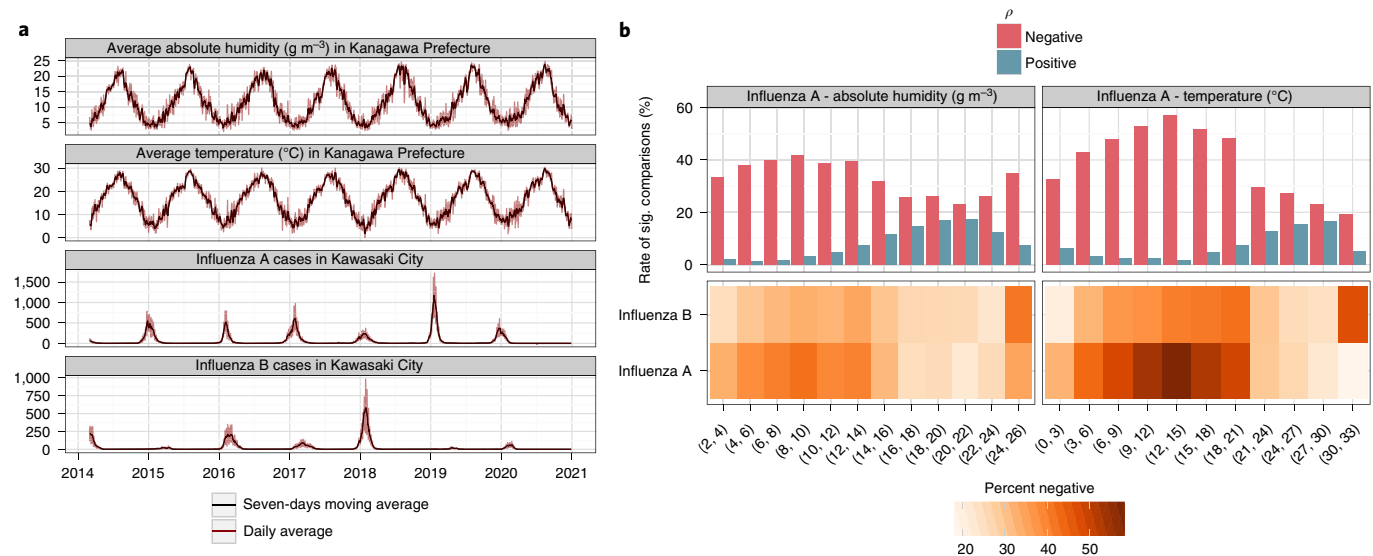


**Fig. 6 | Comparison of weather patterns and influenza A and B incidences in Kawasaki City. a**, Representation of daily averages of AH and $T$ in the Japanese prefecture of Kanagawa, followed by the daily reported cases of influenza A and B in Kawasaki City from 2014 to 2020. **b**, Representation of the critical ranges of $T$ and AH when performing an SDC analysis with $s = 30$ days, similar to the one shown in Fig. 3e–h.

viruses, including measles, variola and varicella, are infections reaching the lower airways, are prevalent during the cooler seasons and are predominantly or substantially airborne. It seems plausible that some viruses have adapted to this phase in the seasonal cycle and their concomitant features of temperature and humidity, ruling out the identification of a single or dominant driver of transmission. Our results point to SARS-CoV-2 belonging to this cluster of

viruses. Regardless, plausible modes of action of temperature and humidity (Supplementary Discussion) suggest the importance of airborne transmission. Low humidity levels can reduce the size of bio-aerosols by evaporation when these cool-weather viruses are exhaled[37]. 'Droplet nuclei' smaller than 5 μm in size can bypass the nasal zone defenses and enter the deeper parts of the lungs. SARS and SARS-CoV-2 specifically target the angiotensin-converting

enzyme 2 protein for cell entry[38], where this protein is most abundant in alveolar tissue[39]. During the colder winter months in higher latitudes, indoor heating generates a microclimate with low levels of RH, a parameter that better reflects the drying power of air than AH. However, both indoor AH and RH in heated accommodation correspond well with the routinely measured outdoor AH in colder climates[18], and therefore also the gridded climatology used here for studying this parameter.

The permissive role of low temperatures for the transmission of SARS-CoV-2 appears to contradict the warm-weather-related transmission and severity of the second wave in the United States for July–August 2020 (not shown) and the 2021 summer wave, as well as the second or third waves in 2021 in tropical countries such as India and in Latin America. In rich countries of the Northern Hemisphere, the cooled indoor microclimate during these months could accommodate transmission, as in the many outbreaks observed in the meat-processing plants of the United States[40] and Europe[41] where cooled air is mechanically reventilated. In countries such as Argentina, Brazil, Colombia and Peru, the limited vaccine supply with its slow roll-out, weak health systems and fragile economies that make stay-at-home orders difficult to impose or maintain can all combine to explain the recent surges in spring of 2021. The implications, and the possibility that new variants emerging at these latitudes might be better adapted to stable tropical conditions, remain to be thoroughly examined.

The positive correlations we found for the summer months of 2020 (for example, Fig. 1) during the seasonal low of cases may also reflect community transmission in public temperature-controlled buildings such as care homes[41,42]. They may also result from massive social clustering and increased contact during warm-weather leisure activities outdoors[43]. Indeed, air conditioning in the subtropics was also proposed to explain SARS transmission[32], a situation bearing similarities with influenza[44]. Similarly, the Northern Hemisphere experienced an important uptick in cases in summer of 2021 in countries such as Spain and Portugal, the target of massive seaside tourism in Europe. Regardless of explanation, these positive associations were less frequent overall, and less robust and weaker when zooming into smaller temporal window sizes.

Our study has several limitations, the most important one stemming from the limited length of the epidemiological records so far available. Also limiting are the different social and environmental forcing factors influencing the epidemiology of COVID-19, which can be country-dependent. When SDC is employed at a single scale of analysis, it can potentially identify spurious correlations, an aspect largely circumvented by the new multiscale analysis. Interestingly, SDC sensitivity can help track other locally relevant epidemiological dynamics potentially linked to human behavior, such as summer gatherings in 2020 inside building premises or under the effect of air conditioning in hot climates, producing positive associations. Such associations do not arise from direct meteorological preferences of the virus but from effects of human behavior on virus transmission. They would not be detected by fixed regression methods. Data-reporting differences across locations and different sensitivities of virus strains might alter the current results, as data for these new variants (delta, delta-plus, lambda) have not been analyzed here. Variation in reporting errors add uncertainty to the analysis, although SDC appears robust to identifying associations even in the presence of a strong weekly cycle. This robustness opens the door to the development, in the near future, of more curated and tailored climate services and early-warning systems for COVID-19.

The identified climate forcing should persist for novel virus variants with increased transmissibility and an enhanced ability to partially evade protection from previous exposure. It should also persist at lower incidence levels were the disease to become endemic, therefore defining the annual timing for vaccination. Strong current policies to curtail transmission, including

lockdowns, where effective, should limit the role of climate drivers. This effect can already be seen when considering the most recent reported cases through the winter in regions of Italy (Extended Data Fig. 6). The frequency of local significant correlations with our method decreases markedly and the critical range of negative correlations is no longer evident (Extended Data Fig. 6b). By contrast, for regions of Spain, where control efforts have been more variable in implementation, this range remains apparent and consistent with the earlier waves (Extended Data Fig. 6b), even though climate conditions are shared by the two countries (Extended Data Fig. 6c). Public health interventions to curb transmission of COVID-19 have focused in this initial phase on reducing the contact rate between people through social distancing, and on hand washing, decontamination of infected surfaces and face covering. With the exception of face covering, these measures have emphasized the importance of short-range transmission, the default assumption on respiratory diseases caused by exhaled droplets[45] with a restricted spatial range[46]. The role of smaller exhaled droplets that become airborne as bio-aerosols or 'droplet nuclei' for extended periods, and may cover long distances in a viable state, has been proposed for COVID-19 but has remained unresolved[47–50]. Seasonality, and the role of low temperatures and the associated low humidity, can be mechanistically linked to viable SARS-CoV-2 aerosols, supporting the relevance of the airborne transmission route implicated in other studies[51]. This link warrants an emphasis on 'air hygiene' through improved indoor ventilation[52] to more effectively intervene in the unfolding pandemic. It also underscores the need to include appropriate meteorological parameters in the evaluation and planning of pharmacological and behavioral control measures.

## Methods

**Global statistical analysis.** Our first attempt to identify plausible effects of meteorological covariates on COVID-19 spread applied a comparative regression analysis. To this end, we focused on the exponential onset of the disease, as it is the epidemic phase that allows for a better comparison between countries or regions, without the confounding effect of intervention policies. We first determined, for each of the spatial units (either countries or NUTS (nomenclature of territorial units for statistics) 2 regions), the day in which 20 or more cumulative cases were officially reported. We then fitted the first-order polynomial function $f(t) = x_0 + rt$ for the next 20 days of log-transformed data, where $t$ represents time (in days) and $x_0$ is the value at initial condition $t = 0$. The $r$ parameter can be understood as the exponential growth rate, and is then used to estimate the basic reproduction number ($R_0$) using the estimated serial interval $T$ for COVID-19 of 4.7 days[53], such that $R_0 = 1 + rT$ (ref. [54]). (We note that we are interested here in the relationship between the reproductive number and not in the actual inference of $R_0$.) Once $R_0$ was obtained for all our spatial units, we filtered our meteorological data to match the same fitting period (with a 10-day negative delay to account for an incubation and reporting lapse) for every spatial unit. To compute a single average of the meteorological variables per regional unit, we computed a weighted average on the basis of the population contribution of each grid cell to the total population of the region. We did so to have an aggregated value that would better represent the impact of these factors on the population transmission of COVID-19, as the same variation in weather in a high-density urban area is more likely to contribute to a change in population-level transmission than that of an unpopulated rural area. We then averaged the daily values of temperature and AH for each country and computed univariate linear models for each of these variables as predictors of $R_0$. Given the somewhat arbitrary criteria to select the dates to estimate the $R_0$ in each country, a sensitivity analysis was run to test the robustness of the regressions to changes in the related parameters. We tested 70 different combinations of two parameters: the total number of days used for the fit (18–27) and the threshold of cumulative COVID-19 cases used to select the initial day of the fit (15–45). We also calculated the weather averages by shifting the selected dates accordingly. Then, a linear model for each of the estimates was fitted for both $T$ and AH. A summary of the distribution of parameter estimates (the regression slope coefficients and the $R^2$ of the models) is shown in Extended Data Fig. 3.

**Bivariate time-series analysis with scale-dependent correlations.** To examine associations between cases and climate factors in more detail, SDC was performed on the daily time series of both COVID-19 incidence and a given meteorological variable. SDC is an optimal method for identifying dynamical couplings in short and noisy time series[20,21]. In general, Spearman correlations between incidence and a meteorological time series assess whether there is a monotonic relation between the variables. SDC analysis was specifically developed to study transitory

associations that are local in time at a specified temporal scale corresponding to the size of the time intervals considered ($s$). The two-way implementation (TW-SDC) is a bivariate method that computes non-parametric Spearman rank correlations between two time series, for different pairs of time intervals along these series. Different window sizes ($s$) can be used to examine increasingly finer temporal resolution. The results are sensitive to the value of this window size, $s$, with expected significant and highest correlation values at the scale of the transient coupling between variables. Correlation values decrease in magnitude as window size increases, and averages are computed over too long a time interval. Values can also decrease and become non-significant for small windows when correlations are spurious. Here, the method was applied for windows of different length (from $s = 75$ to 14 days) and, despite a weekly cycle showing up in some cases for small $s$, results removing this cycle were robust. We therefore did not remove this cycle.

The results are typically displayed in a figure with the following subplots: (1) the two time series, to the left and top of the matrix of correlation values, respectively; (2) the matrix or grid of correlation values itself in the center, with significant correlations colored in blue when positive and in red when negative, with rows and columns corresponding to the temporal localization of the moving window along the time series on the left and top, respectively; (3) a time series at the bottom, below this grid, with the highest significant correlations for a given time (vertically, and therefore for the variable that acts as the driver, here the meteorological time series). To read the results, one starts at the diagonal and moves vertically down from it to identify a given lag for which significant correlations are found (the closest to the main diagonal). In some of the SDC figures, the time intervals with high local correlations are highlighted with boxes. These intervals alternate with other ones (left blank) for which no significant correlation is found. All colored areas correspond to significance levels of at least $P < 0.05$. A new presentation of results is also used for the influenza analysis, in which the two time series are superimposed in the same plot with the significant correlations shown in a panel below as a function of time and lag (Extended Data Fig. 9).

Significance is assessed with a non-parametric randomization test (see ref. [20] for further details and for examples illustrating the method). For the baseline test, SDC calculates Spearman correlations (at $\alpha = 0.05$) between two white-noise time series at each fragment size $s$ for a non-parametric permutation test (Supplementary Fig. 7). The indices of the series are randomly re-ordered, breaking their temporal shape. This permutation test enables a first estimate of the probability of finding significant spurious correlations, and it can thus be used as a non-parametric significance test for pairs of any length for the time series of interest. As seen in Supplementary Fig. 7, the threshold decays rapidly as the fragment size grows, with high values becoming rarer the longer the time series are compared. As a second test, SDC evaluates the unwanted effects of the internal autocorrelation in a time series. This effect might artificially inflate the correlation obtained between two time series, and should therefore be taken into consideration properly in a significance comparison. To estimate this effect, we generate pairs of autocorrelated (red) noise series of 365 time steps, with a mean $\mu = 0$ and standard deviation $\sigma = 1$ and varying degrees of autocorrelation (autocorrelation parameter $\varphi$ in Supplementary Fig. 8) from 0 (equivalent to white noise) to 0.95, in steps of 0.05. We repeated this procedure 20 times for every unique combination of parameters to achieve a robust estimate. The method then searches for significant couplings in either direction. This is carried out for each of these synthetic time-series pairs with an SDC analysis (for example, with $s = 30$), yielding, for each of them, the false discovery rate, which accounts for a type I error and provides the rate of significant couplings at $\alpha = 0.05$. In Supplementary Fig. 8b, we show the average false discovery rate of the tests for each pair of autocorrelation values. As shown, the chance of finding a spurious coupling increases monotonically as a function of the autocorrelation evaluated.

Our approach focuses on analyzing temporal associations in one location at a time, and comparing across these locations the patterns of association themselves, including their timing (for example, in the waning or rising phase of epidemic waves). This allows comparison of results among distinct regions, despite differences in control measures and disease epidemic state. Time-series analyses applied to each location do not present the typical problems of comparative spatial regression studies, which might be biased by uncontrollable confounding effects across spatial locations.

One reason why couplings between two variables in ecological or epidemiological systems may be transient couplings is the existence of thresholds above or below which responses to forcing are weak or absent. To interrogate our analyses for the existence of critical thresholds/ranges for optimal transmission of the virus (inferred from COVID-19 disease outcomes), we pooled all negative and positive significant correlations performed at a size $s = 21$ days between each meteorological variable and COVID-19 cases. We then computed the proportion of those negative correlations among all possible correlations for a given fragment size obtained for each bin of $T$ or AH values and plotted their distribution (Fig. 3 and Supplementary Figs. 4–6) for all individual regions in Italy.

Singular spectrum analysis (SSA) involves the spectral decomposition (eigenvalues and corresponding eigenvectors) of a covariance matrix obtained by lagging the time-series data for a prescribed number of lags $M$ called the embedding dimension. There are two crucial steps in this analysis for which there are no formal results, but useful rules of thumb: (1) the choice of $M$ and (2) the grouping of the eigenvectors to define the specific major components and

reconstruct them. Typically, the grouping of the eigencomponents is based on the similarity and magnitudes of the eigenvalues, their power (variance of the data they account for) and the peak frequency of the resulting reconstructed components. For selection of the embedding dimension, one general strategy is to choose it so that at least one period of the lowest-frequency component of interest can be identified, that is, $M > f_s/f_r$, where $f_s$ is the sampling rate and $f_r$ the minimum frequency. Another strategy is that $M$ be large enough that the $M$-lagged vector incorporates the temporal scale of the time series that is of interest. The larger the $M$, the more detailed the resulting decomposition of the signal. In particular, the most detailed decomposition is achieved when the embedding dimension is approximately equal to half of the total signal length. A compromise must be reached, however, as a large $M$ implies increased computation, and too large a value may produce mixing of components. SSA is especially well suited for separating components corresponding to different frequencies in nonlinear systems. Here, we applied it to remove the weekly cycle.

**MSDC analysis.** MSDC provides a scan of the SDC analyses over a range of different scales (here, $S$ from 5 to 100 days at 5-day intervals), by selecting the maximum correlation values (positive or negative) closer to the diagonal. The goal is to consider the evolution of transient correlations at all scales pooled together in a single analysis. The MSDC plot displays time on the $x$ axis and scale ($S$) on the $y$ axis, and positive and negative correlations either jointly or separately. The rationale behind MSDC is that correlations at very small scales can occur by chance because of coincident similar patterns, but that as one moves up to larger scales (by increasing $S$), the correlation patterns that are spurious tend to vanish, whereas those reflecting mechanistic links increase in strength. This increase in correlation values should occur up to the real scale of interaction, decreasing afterwards. By 'real', we mean here the temporal scale covering the extent of the interaction between the driver and the response process (in this case, the response of disease transmission to a given climate factor). Thus, continuity of the same sign correlations together with transitions to larger values are indicative of causal effects, whereas the rapid vanishing of small-scale significant correlations signals spurious ones.

**Process-based model.** *Description.* The dynamical model is a discrete stochastic model that incorporates seven different compartments: S, E, I, C, Q, R and D. The model structure is illustrated in Fig. 4. The transition probabilities of the stochastic model are based on the corresponding rates of the transitions between classes in the deterministic (mean-field) model (specified in Fig. 4b). These probabilities are defined as follows. $P(e) = (1.0 - \exp(-\beta \, dt))$ is the probability of infection exposure of the susceptible class, where $\beta = (1/N)(\beta_I I + \beta_Q Q)$ is the infection rate (of the deterministic model). $P(i) = (1.0 - \exp(-\gamma \, dt))$ is the probability that an new exposed individual becomes infectious, where $\gamma$ denotes the incubation rate. $P(r) = (1.0 - \exp(-\Lambda \, dt))$ is the recovery probability, where $\lambda_0(1 - \exp(\lambda_1 t))$ is the (deterministic) recovery rate. $P(p) = (1.0 - \exp(-\alpha \, t))$ is the protection probability, where $\alpha = \alpha_0 \exp(\alpha_1 t)$. $P(d) = (1.0 - \exp(-K \, dt))$ is the mortality probability, with $K = k_0 \exp(k_1 t)$. $P(re) = (1.0 - \exp(-\tau \, dt))$ is the release probability from confinement, where $\tau = \tau_0 \exp(\tau_1 t)$. Finally, $P(q) = (1.0 - \exp(-\delta \, dt))$ is the detection probability, where $\delta$ is the quarantine rate (for example, at which infected individuals are isolated from the rest of the population).

In the model, both infected non-detected and infected detected individuals can infect susceptible ones. In the model incorporating temperature in the transmission rate, the respective values of $\beta_I$ and $\beta_Q$ are calculated as follows:

$$\beta_I(t) = \beta_I \, T_{\text{inv}}(t); \quad \beta_Q(t) = \beta_Q \, T_{\text{inv}}(t)$$

where $T_{\text{inv}} = f\left(\frac{1 - T(t)}{\bar{T}}\right)$, with $\bar{T}$ corresponding to the overall mean of the temperature time series and $f(\cdot)$ to a Savitzky–Golay filter, used to smooth the temperature series with a window size of 50 data points and a polynomial order of 3. When the infection rate is constant, we simply omit the temperature term. For further comparison, in a third model, $\beta$ is specified with a sinusoidal function of period equal to 12 months and an estimated phase.

The number of individuals transitioning from compartment $i$ to $j$ at time $t$ are determined by means of binomial distributions $P(Xi, P(y))$, where $Xi$ corresponds to one of the compartments S, E, I, Q, R, D, C, and $P(y)$ to the respective transition probability defined above. Thus,

- $e(t) = P(S(t), P(e))$, new exposed individuals at time $t$
- $p(t) = P(S(t), P(p))$, protected individuals at time $t$
- $i(t) = P(E(t), P(i))$, new infected not detected individuals at time $t$
- $q(t) = P(I(t), P(q))$, new infected and detected individuals at time $t$
- $r(t) = P(Q(t), P(r))$, total recovered individuals at time $t$
- $d(t) = P(Q(t), P(d))$, total dead individuals at time $t$
- $re(t) = P(C(t), P(re))$, individuals released from confinement at time $t$

Then, the final dynamics are given by the following equations:

$$S(t) = S(t - dt) - e(t) - p(t) + re(t)$$

$$E(t) = E(t - dt) + e(t) - i(t)$$

$$I(t) = I(t - dt) + i(t) - q(t)$$

$$Q(t) = Q(t - \mathrm{d}t) + q(t) - r(t) - d(t)$$

$$R(t) = R(t - \mathrm{d}t) + r(t)$$

$$D(t) = D(t - \mathrm{d}t) + d(t)$$

$$C(t) = C(t - \mathrm{d}t) + p(t) - re(t)$$

*Calibration.* The model was implemented using Python and calibrated by means of the least squares algorithm of the scipy library. The error function minimized with this algorithm was obtained from the normalized residuals on the basis of total cases (Q + R + D) and deaths (D).

To search parameter space, we ran 100 calibrations starting from different initial choices of parameter combinations. The tolerance for termination in the change of the cost function was set to $1 \times 10^{-10}$. Tolerance for termination by the norm of the gradient was also set to $1 \times 10^{-10}$, and the tolerance for termination by the change of the independent variables was set to $1 \times 10^{-10}$. The solver was the lsmr method (which is suitable for problems with sparse and large Jacobian matrices) with a differential step of $1 \times 10^{-5}$. With this configuration, each fitting run usually converged after ~500 iterations.

*Validation.* To compare the model including an effect of $T$ in the transmission rate to those without it, we calculated the chi-square, Akaike information criterion (AIC) and Bayesian information criterion (BIC) indices for the residuals obtained from the optimization process. The resulting values are shown in Supplementary Table 1.

Our choice of $T$ to modulate the infection rate ($\beta$) instead of AH underlies the fact that the temporal dynamics of both factors roughly follow the same shape, with the advantage that $T$ shows less oscillatory behavior than AH. This fact adds stability to the model when the inverse relationship is used in the calculation of $\beta$ (Supplementary Information). This selection is further reinforced by the results from the SDC analyses, which yielded larger correlations for temperature, even when penalizing for the larger autocorrelation structure.

Our choice to modulate $\beta$ using $T$ instead of AH follows from the fact that the temporal dynamics of both climate variables present roughly the same shape, with the advantage that $T$ exhibits weaker oscillations. This less fluctuating pattern provides stability to the model fitting when the inverse relationship is used in the calculation of $\beta$ (Supplementary Information). Additionally, the transient correlations obtained with SDC yielded higher values for $T$ than for AH (even when accounting for concurrent levels of autoregression in the two variables).

## Data availability
Data on COVID-19 daily incidence counts were retrieved from the ECDC dataset[55] for the nationally aggregated data and from the COVID-19 Data Hub[56] for the NUTS 2-disaggregated incidence data. National and regional boundaries were obtained from shapefiles provided by GISCO services by Eurostat[57]. Meteorological data were retrieved from ERA5 reanalysis through API requests to the Copernicus Climate Data Store (CDS)[58]. We obtained temperature (at 2 m), sea level pressure, total precipitation and dew point temperature for a 0.5° × 0.5° grid for the country-wise analysis and for a 0.25° × 0.25° grid for the rest. From the 2 m temperature and the dew point temperature, we derived relative humidity, and used the atmos Python package[59] to calculate AH. The Gridded Population of the World (GPWv4), developed by the Center for International Earth Science Information Network (CIESIN) at Columbia University, was obtained from ref. [60] at a resolution of 2.5 min. We did not include data from the United States in our analyses given the large geographical extent of this country, which includes different climatic zones, and the largely asynchronous implementation of intervention policies across its different states. Data for influenza A and B incidence in the city of Kawasaki were retrieved from the Real-time Surveillance website (https://kidss.city.kawasaki.jp/en/realsurveillance/opendata) of the Japanese National Epidemiological Surveillance of Infectious Diseases[61] as total daily reported cases from 1 March 2014 to 31 December 2020. Daily temperature averages were obtained by averaging the hourly values for the same period obtained via the ERA5 reanalysis for grid cells spanning the municipality of Kawasaki (coordinates (139.5, 35.5) and (139.75, 35.5) in a 0.25° × 0.25° grid). Source data for Figs. 1 to 5 and Extended Data Figs. 1 to 6 are available in a Code Ocean repository[62]. Source data for Figure 6 are provided with this paper.

## Code availability
Code for the SDC analyses has been implemented as a Python package available in PyPi (sdcpy[63]). Code to ensure the reproducibility of the analysis is available in Code Ocean[62] and GitHub (https://github.com/AlFontal/covid_climate_signatures).

## References
1. Fan, V. Y., Jamison, D. T. & Summers, L. H. Pandemic risk: how large are the expected losses? *Bull. World Health Org.* **96**, 129–134 (2018).
2. Shao, W., Li, X., Goraya, M. U., Wang, S. & Chen, J.-L. Evolution of influenza a virus by mutation and re-assortment. *Int. J. Mol. Sci.* **18**, 1650 (2017).
3. Roy, C. & Milton, D. Airborne transmission of communicable infection—the elusive pathway. *N. Engl. J. Med.* **350**, 1710–1712 (2004).
4. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
5. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
6. WHO Director-General's opening remarks at the media briefing on COVID-19. *WHO* (11 March 2020); https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
7. Worldometer Coronavirus Update (Live): 62,125,065 Cases and 1,451,937 Deaths from COVID-19 Virus Pandemic (accessed September 18 2021) https://www.worldometers.info/coronavirus/#countries
8. Faust, J. S. & del Rio, C. Assessment of deaths from COVID-19 and from seasonal influenza. *JAMA Intern. Med.* **180**, 1045–1046 (2020).
9. Vittecoq, M. et al. Does the weather play a role in the spread of pandemic influenza? A study of H1N1pdm09 infections in France during 2009–2010. *Epidemiol. Infect.* **143**, 3384–3393 (2015).
10. Carlson, C. J., Gomez, A. C. R., Bansal, S. & Ryan, S. J. Misconceptions about weather and seasonality must not misguide COVID-19 response. *Nat. Commun.* **11**, 4312 (2020).
11. 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**, 315–319 (2020).
12. Sajadi, M.M. et al. Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19). *JAMA network open* **3**, e2011834–e2011834 (2020).
13. Ma, Y. et al. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Total Environ.* **724**, 138226 (2020).
14. Bashir, M. F. et al. Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Sci. Total Environ.* **728**, 138835 (2020).
15. Yu, I. T. et al. Evidence of airborne transmission of the severe acute respiratory syndrome virus. *N. Engl. J. Med.* **350**, 1731–1739 (2004).
16. Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission and seasonality. *Proc. Natl Acad. Sci. USA* **106**, 3243–3248 (2009).
17. Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316 (2010).
18. 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).
19. Liu, Y. et al. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature* **582**, 557–560 (2020).
20. Rodo, X. & Rodriguez Arias, M. A. A new method to detect transitory signatures and local time/space variability structures in the climate system: the scale-dependent correlation analysis. *Clim. Dyn.* **27**, 441–458 (2006).
21. Rodríguez-Arias, M. N. & Rodó, X. A primer on the study of transitory dynamics in ecological series using the scale-dependent correlation analysis. *Oecologia* **138**, 485–504 (2004).
22. Rodó, X. Reversal of three global atmospheric fields linking changes in SST anomalies in the Pacific, Atlantic and Indian oceans at tropical latitudes and midlatitudes. *Clim. Dyn.* **18**, 203–217 (2001).
23. Rodó, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. ENSO and cholera: a nonstationary link related to climate change? *Proc. Natl Acad. Sci. USA* **99**, 12901–12906 (2002).
24. Rodó, X. et al. Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan. *Proc. Natl Acad. Sci. USA* **111**, 7952–7957 (2014).
25. López, L. & Rodó, X. The end of social confinement and COVID-19 re-emergence risk. *Nat. Hum. Behav.* **4**, 746–755 (2020).
26. Audi, A. et al. Seasonality of respiratory viral infections: will COVID-19 follow suit?. *Front. Public Health* **8**, 567184 (2020).
27. 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.* **12**, 3602 (2021).
28. Lipsitch, M. & Viboud, C. Influenza seasonality: lifting the fog. *Proc. Natl Acad. Sci. USA* **106**, 3645–3646 (2009).
29. Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S. & Sugihara, G. Global environmental drivers of influenza. *Proc. Natl Acad. Sci. USA* **113**, 13081–13086 (2016).
30. *Airborne Transmission of SARS CoV 2—A Virtual Workshop* (The National Academies of Science, Engineering and Medicine, 2020); https://www.

nationalacademies.org/event/08-26-2020/
airborne-transmission-of-sars-cov-2-a-virtual-workshop

31. Irwin, C. K. et al. Using the systematic review methodology to evaluate factors that influence the persistence of influenza virus in environmental matrices. *Appl. Environ. Microbiol.* **77**, 1049–1060 (2011).

32. Chan, K. H. et al. The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.* **2011**, 734690 (2011).

33. Morris, D. H. et al. Mechanistic theory predicts the effects of temperature and humidity on inactivation of SARS-CoV-2 and other enveloped viruses. *Elife* **10**, e65902 (2021).

34. Woese, C. Thermal inactivation of animal viruses. *Ann. N. Y. Acad. Sci.* **83**, 741–751 (1960).

35. Boone, S. A. & Gerba, C. P. Significance of fomites in the spread of respiratory and enteric viral disease. *Appl. Environ. Microbiol.* **73**, 1687–1696 (2007).

36. Moriyama, M., Hugentobler, W. J. & Iwasaki, A. Seasonality of respiratory viral infections. *Annu. Rev. Virol.* **7**, 83–101 (2020).

37. Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **3**, 1470–1476 (2007).

38. Zhao, Y. et al. Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *Am. J. Resp. Crit. Care Med.* **202**, 756–759 (2020).

39. Hamming, I. et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* **203**, 631–637 (2004).

40. Dyal, J. W. COVID-19 among workers in meat and poultry processing facilities-19 states, April 2020. *Morb. Mortal. Wkly Rep* **69**, 557–561 (2020).

41. ECDC Heating, Ventilation and Air-Conditioning Systems in the Context of COVID-19: First Update. *ECDC* (11 November 2020); https://www.ecdc. europa.eu/en/publications-data/ heating-ventilation-air-conditioning-systems-covid-19

42. Lynch, R. M. & Goring, R. Practical steps to improve air flow in long-term care resident rooms to reduce COVID-19 infection risk. *J. Am. Med. Directors Assoc.* **21**, 893–894 (2020).

43. Azuma, K., Kagi, N., Kim, H. & Hayashi, M. Impact of climate and ambient air pollution on the epidemic growth during COVID-19 outbreak in Japan. *Environ. Res.* **190**, 110042 (2020).

44. Moriyama, M. & Ichinohe, T. High ambient temperature dampens adaptive immune responses to Influenza A virus infection. *Proc. Natl Acad. Sci. USA* **116**, 3118–3125 (2019).

45. Chapin, C. V. *The Sources and Modes of Infection* (Wiley, 1912).

46. Keene, C. H. Airborne contagion and air hygiene. William Firth Wells. *J. School Health* **25**, 249–249 (1955).

47. Lednicky, J. A. et al. Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients. *Int. J. Infect. Dis.* **100**, 476–482 (2020).

48. Somsen, G. A., Rijn, C. V., Kooij, S., Bem, R. A. & Bonn, D. Small droplet aerosols in poorly ventilated spaces and SARS-CoV-2 transmission. *Lancet Resp. Med.* **8**, 658–659 (2020).

49. Morawska, L. & Milton, D. K. It is time to address airborne transmission of coronavirus disease 2019 (COVID-19). *Clin. Infect. Dis.* **71**, 2311–2313 (2020).

50. Tanne, J. H. Covid-19: CDC publishes then withdraws information on aerosol transmission. *BMJ* (24 September 2020) https://doi.org/10.1136/bmj.m3739

51. Zhang, R., Li, Y., Zhang, A. L., Wang, Y. & Molina, M. J. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 14857–14863 (2020).

52. Jayaweera, M., Perera, H., Gunawardana, B. & Manatunge, J. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. *Environ. Res.* **188**, 109819 (2020).

53. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286 (2020).

54. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B Biol. Sci.* **274**, 599–604 (2007).

55. ECDC Download historical data (to 14 December 2020) on the daily number of new reported COVID-19 cases and deaths worldwide. *ECDC* (14 December 2020); https://www.ecdc.europa.eu/en/publications-data/ download-todays-data-geographic-distribution-covid-19-cases-worldwide

56. Guidotti, E. & Ardia, D. COVID-19 data hub. *J. Open Source Softw.* **5**, 2376 (2020).

57. *NUTS - GISCO* (Eurostat) https://ec.europa.eu/eurostat/web/gisco/geodata/ reference-data/administrative-units-statistical-units/nuts#nuts21/

58. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

59. atmos-python: atmospheric sciences utility library. *GitHub* https://github. com/atmos-python/atmos (2021).

60. SEDAC Gridded population of the World, Version 4 (GPWv4): Population Count, Revision 11. *SEDAC* (Columbia University, 2018); https://sedac.ciesin. columbia.edu/data/set/gpw-v4-population-count-rev11

61. KIDSS Kawasaki City Infectious Disease Surveillance System. *KIDSS* (October 2021); https://kidss.city.kawasaki.jp/en/realsurveillance/opendata

62. Fontal, A., Bouma, M.J., San Jose, A., Pascual, M. and Rodo, X. Climatic signatures in the different COVID-19 pandemic waves across both hemispheres https://doi.org/10.24433/CO.5600300.v1 (2021).

63. Fontal, A. *AlFontal/sdcpy* (Zenodo, 2021); https://doi.org/10.5281/zenodo. 4949813

## Author contributions
A.F. participated in methodology, software, formal analysis, data curation, visualization and writing, reviewing and editing the manuscript. M.J.B. contributed to writing the original draft, and reviewing and editing the manuscript. A.S.J. contributed to software, formal analysis and data curation. L.L. contributed to software, formal analysis and data curation. M.P. contributed to writing the original draft, and reviewing and editing the manuscript. X.R. contributed to conceptualization, methodology, supervision, software, writing the original draft, and reviewing and editing the manuscript. All authors participated in the discussion and interpretation of results.

## Competing interests
The authors declare no competing interests.

## Additional information
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s43588-021-00136-6.

**Correspondence and requests for materials** should be addressed to Xavier Rodó.

**Peer review information** *Nature Computational Science* thanks Muhammad Farhan Bashir and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Handling editor: Fernando Chirigati, in collaboration with the *Nature Computational Science* team.

**Reprints and permissions information** is available at www.nature.com/reprints.
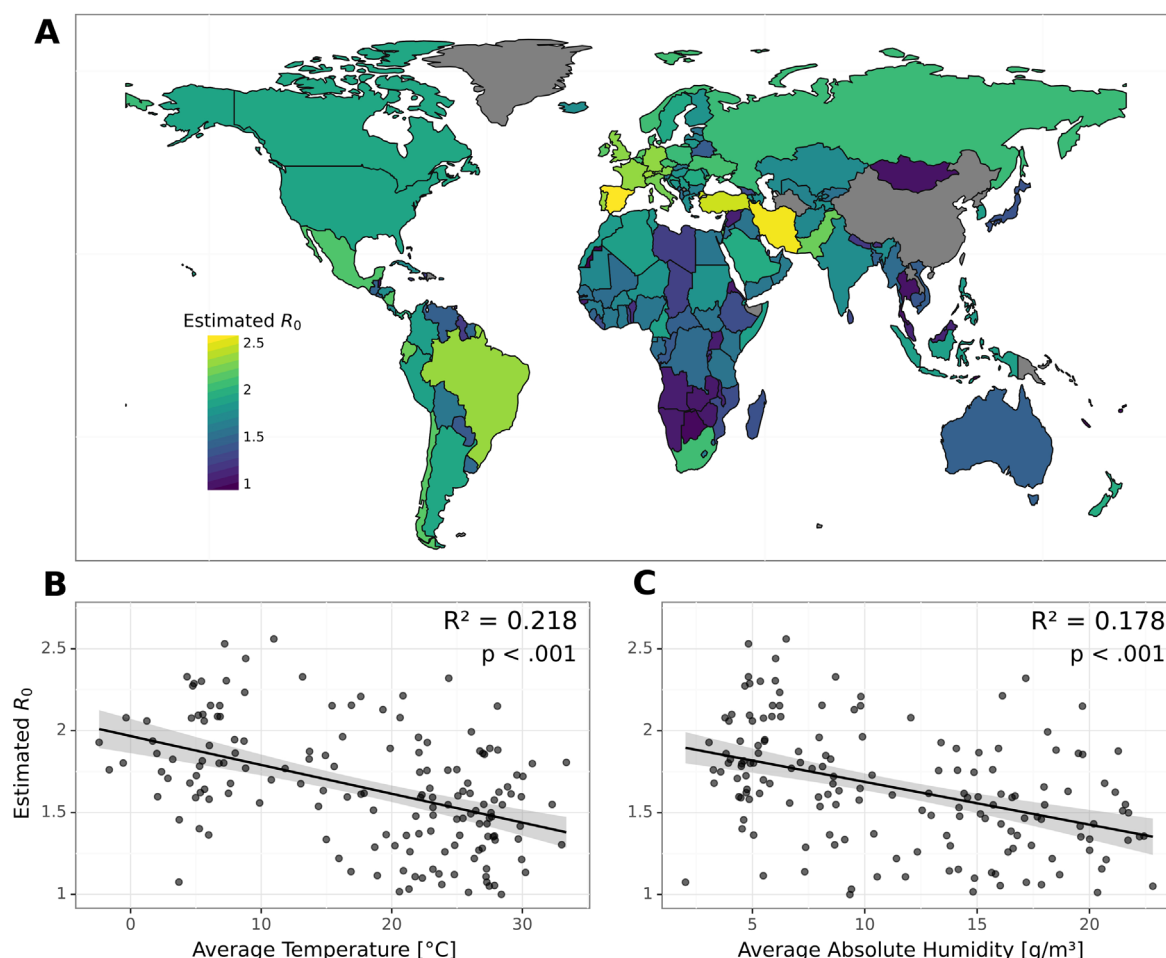
**Extended Data Fig. 1 | Comparison of the country-wise estimated R0 during the initial phases of the pandemic in contrast with their meteorological conditions.** Estimated basic reproduction number (R0) during the initial phases of the pandemic for all countries included in the study (n=162). (A) To focus on the initial rise of the first wave, these estimates are based on the cumulative COVID-19 cases during the first 20 days following the 20th confirmed case in each country (see Methods). Scatterplots showing the variation of the estimated R0 for the different countries as a function of their average temperatures [°C] (B) and absolute humidities [g/m³] (C) during the initial phase of the pandemic. Black lines and $R^2$ values correspond to the estimates of an univariate linear model in both cases, fitted by Ordinary Least Squares (with $p < 0.001$ in both cases in a two-tailed T-test). Of the 162 countries included globally, the numbers corresponding to each continent are the following: Africa (n=50), Asia & Oceania (n=45), Europe (n=38), North America (n=16), and South America (n=12). For comparison, the R0 estimates obtained after the first 100 reported cases (rather than 20) are shown in Extended Data Fig. 2(A). Similar spatial patterns are observed. The spatial distribution of average temperatures and absolute humidity, as well as the national weights based on relative population density, used to compute these averages, are shown in Extended Data Fig. 2(B-D).
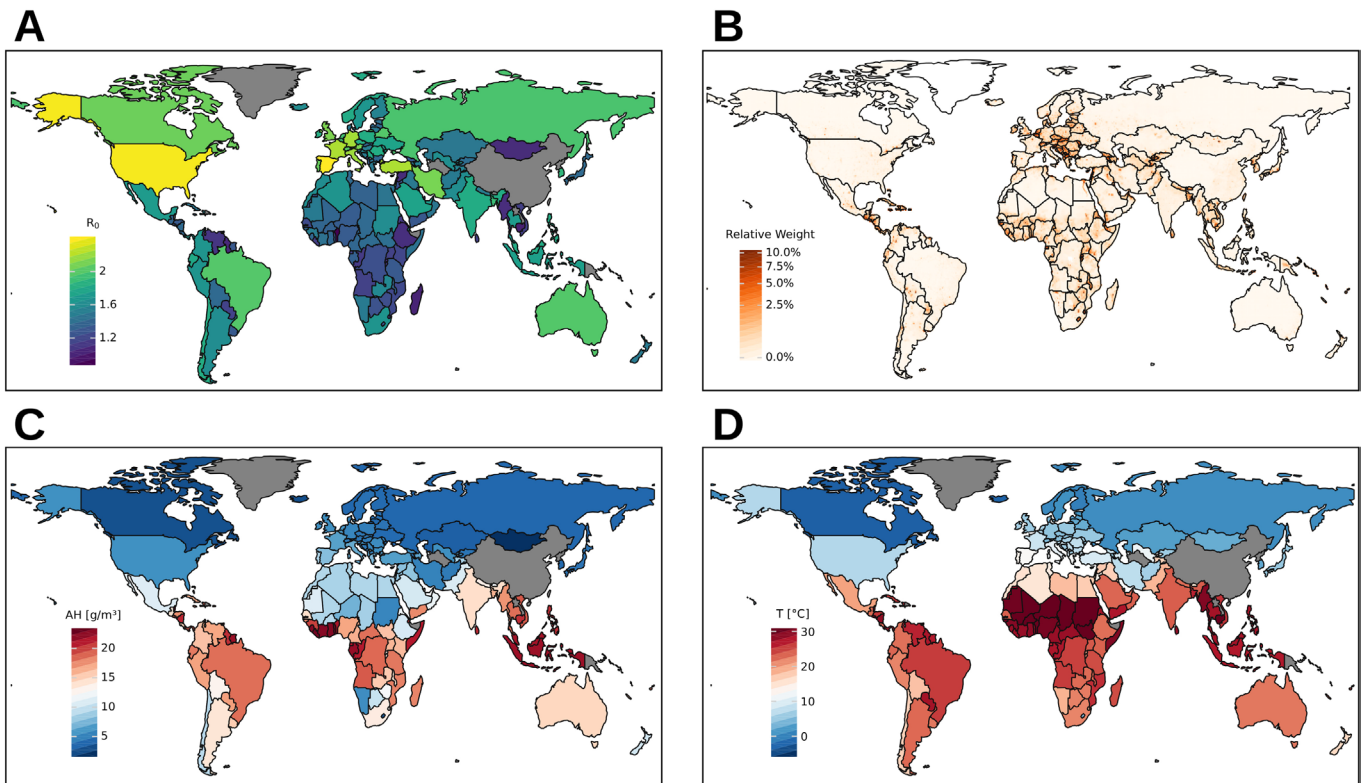
**Extended Data Fig. 2 | Worldwide maps for estimated initial R0, weather variables, and population density.** (A) Geographical distribution for R0 when it is estimated using the period of 20 days after the 100th COVID19 case (for comparison with Extended Data Fig. 1). (B) Geographical distribution of relative weights given to each cell in a 1×1° grid. Weights represent the fraction of the population living in each cell for every country. Maps in C and D correspond to the averages of AH and T +/-10 days before/after the notification day of the 20th COVID19 case in each country, respectively.
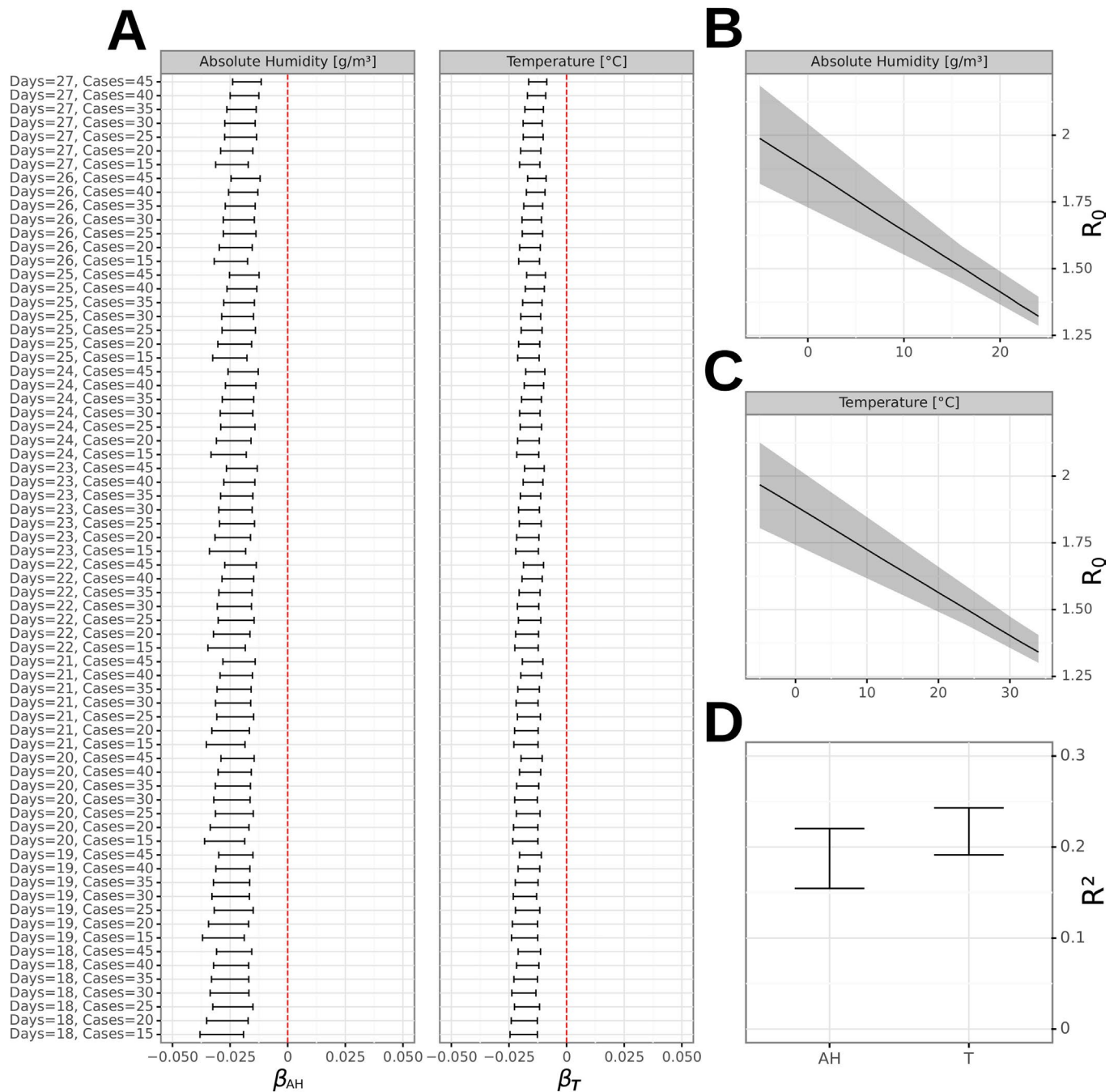
**Extended Data Fig. 3 | Regression sensitivity analysis.** Sensitivity tests for the OLS regression models shown in Extended Data Fig. 1. (A) 95% CI of the estimated slopes of the OLS regressions ($\beta_{AH}$ and $\beta_T$) for different values of the number of days (18 to 27) and initial number of cases (15 to 45) used for estimating the R0 (see Methods). B and C show the distribution of all the regression lines obtained when considering the variation in days and cases shown in A, for AH and T, respectively. Solid line is the median regression line and the shaded interval corresponds to the area occupied by the most extreme regression lines among the 70 sets of parameters tested. (D) Full range of $R^2$ estimates obtained as a result of the sensitivity analysis. Results support the robustness and stability of results in Extended Data Fig. 1.
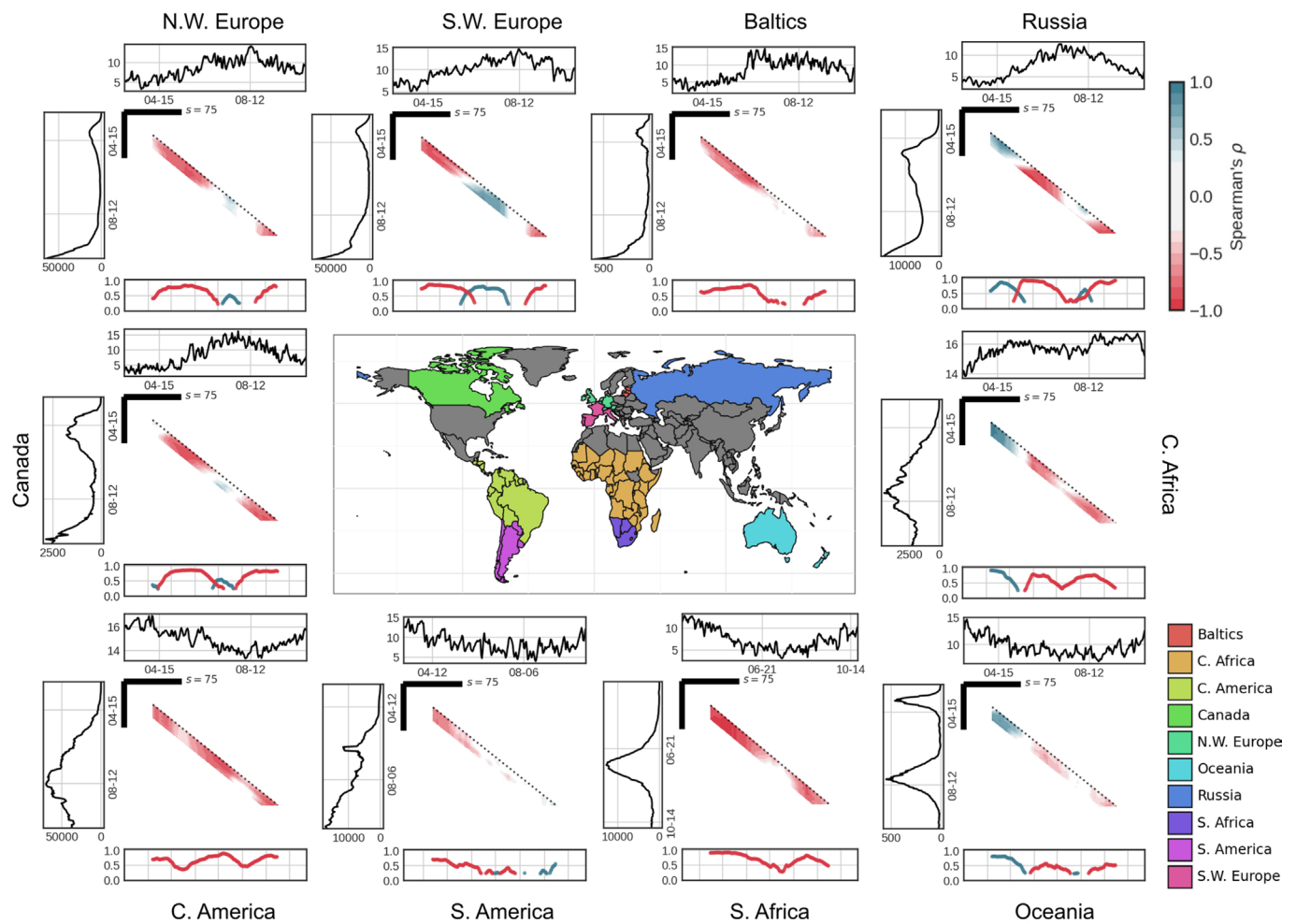
**Extended Data Fig. 4 | SDC analyses for the aggregated COVID-19 cases and absolute humidity for groups of contiguous countries.** Idem as Figure 1 but for Absolute Humidity (g/m³).

# Catalonia (Spain)
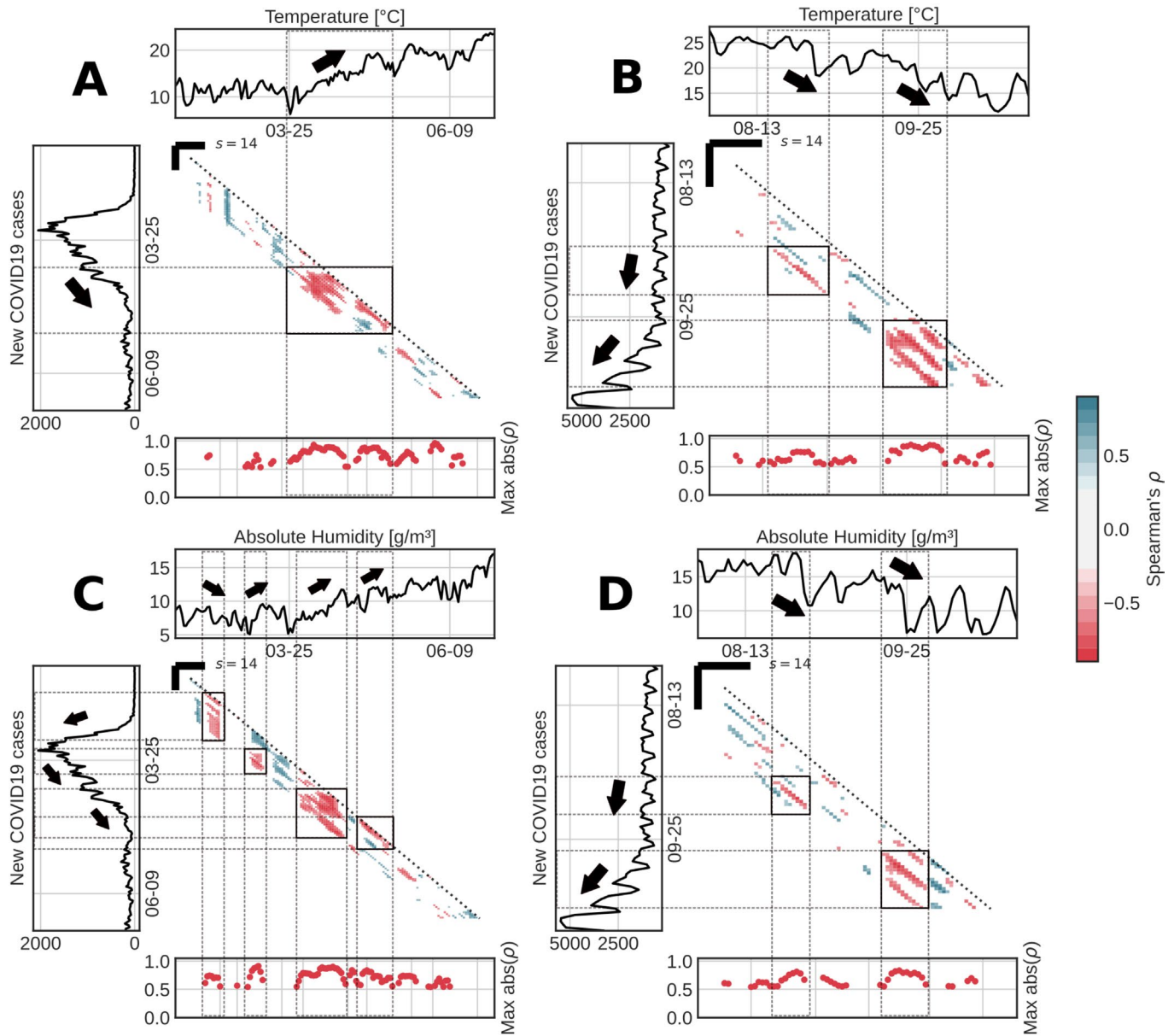


**Extended Data Fig. 5 | Scale Dependent Correlation analyses of T/AH with COVID-19 cases during the initial wave and summer in Catalonia, Spain.** Idem as Fig. 3 (A-D) but for Catalonia (Spain).
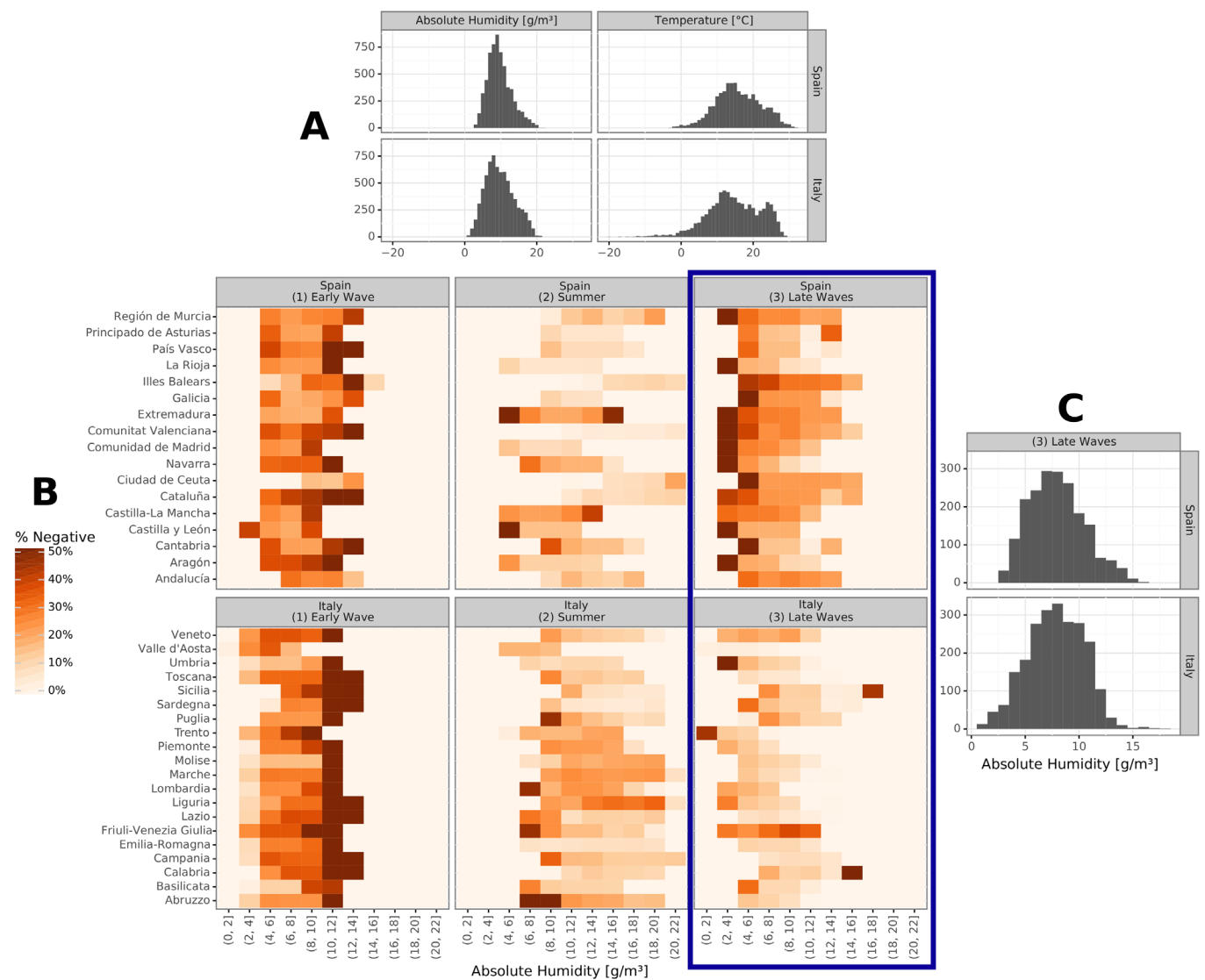
**Extended Data Fig. 6 | Distribution of weather conditions and its effect on COVID-19 spread across waves in Spain and Italy.** (A) Distribution of daily absolute humidity and temperature averages for Italy and Spain during the whole considered period (15 February 2020 to 31 January 2021). (B) Comparison of the frequency of significant negative local correlations for different ranges of absolute humidity found in SDC analyses (s=21) for Italian and Spanish regions across the three defined periods of the COVID19 pandemic (see methods). (C) Distribution of daily absolute humidity averages in the same regions during the late waves of the pandemic (October 2020 to January 2021).
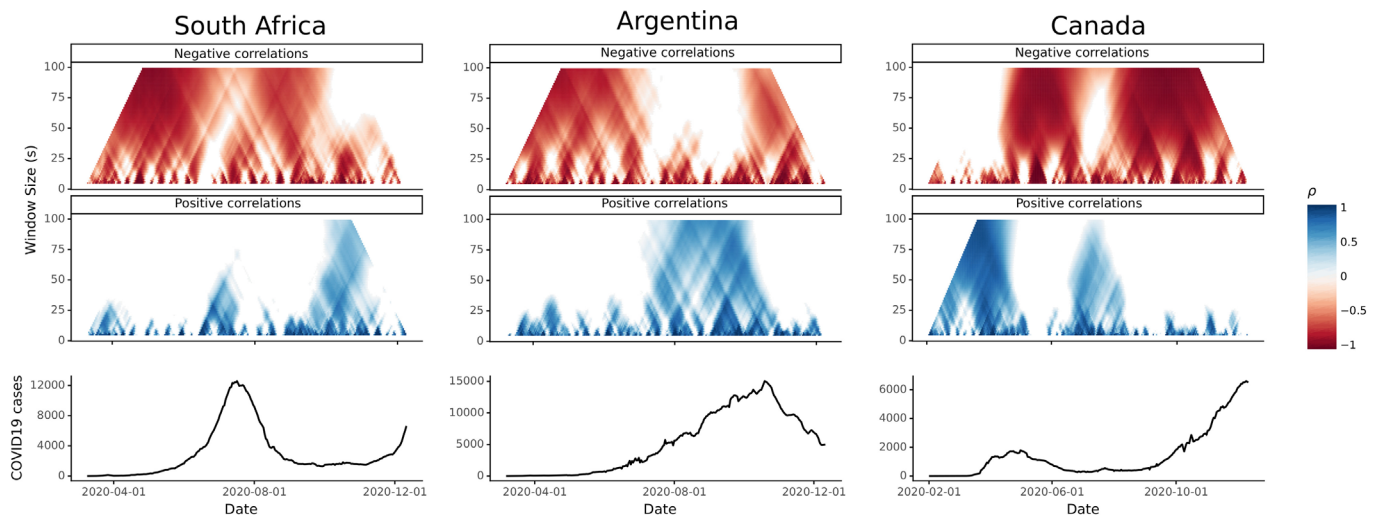
**Extended Data Fig. 7 | Multiscale SDC analysis of COVID-19 and temperature in South Africa, Argentina and Canada.** Representation of the results of running multiple two-way SDCs at scales ranging from 5 to 100 days for the average daily temperatures against the number of reported daily new cases of COVID19 in South Africa, Argentina in Canada, spanning from February to December 2020. The top panels (red) represent the minimal significant Spearman correlation found for every pair of window size and date, while the central panels (blue) represent the maximal significant correlation for the same pair of variables. The bottom panels display the 7-days moving average of daily reported new cases for each of the regions.
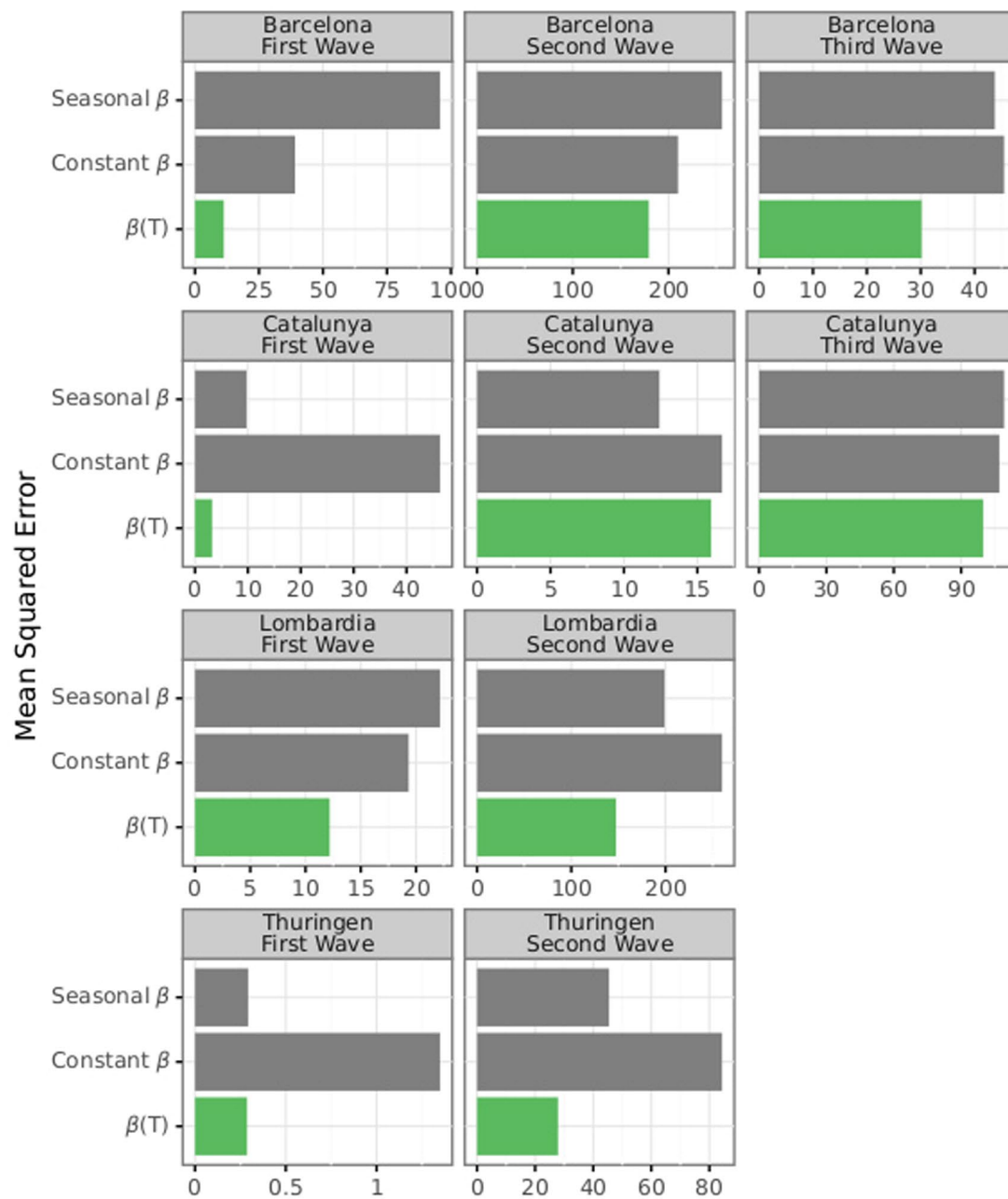
**Extended Data Fig. 8 | Mathematical model's goodness of fit.** Mean squared error of each of the models, regions and periods in the study fitted with the process-based model.
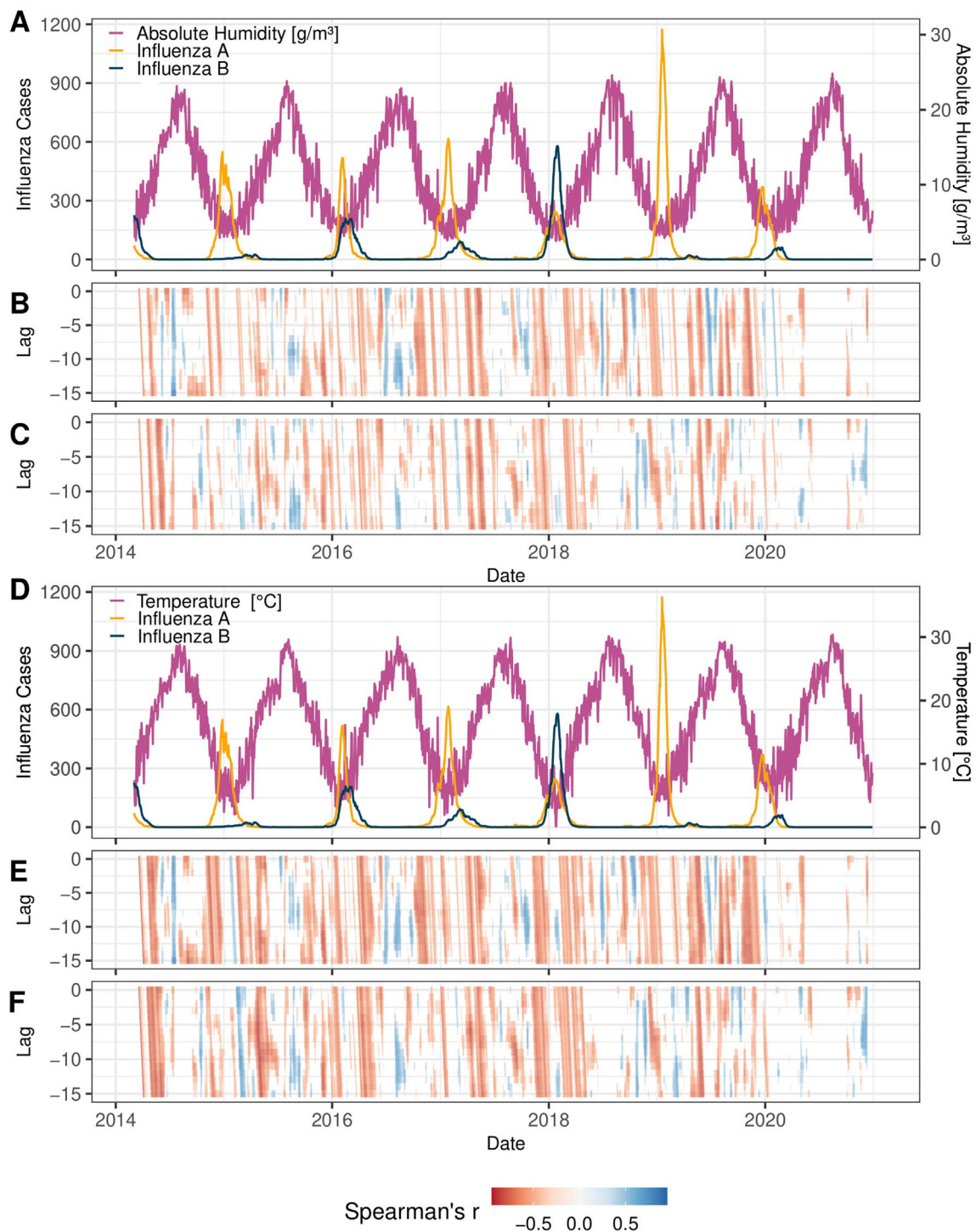
**Extended Data Fig. 9 | Comparison of historical Influenza incidence and weather conditions in Kawasaki City.** A and D show the daily averages for absolute humidity and temperature in the Kanagawa prefecture against the daily reported cases of Influenza A and B in Kawasaki City. Below, the results of an SDC analysis performed at s=30 for absolute humidity against Influenza A (B) and Influenza B (C), and temperature against Influenza A (E) and Influenza B (F).