



OPEN

Data analysis evidence beyond correlation of a possible causal impact of weather on the COVID-19 spread, mediated by human mobility

Denis-Răducu Nichita^{1,2}✉, Mihai Dima¹, Loredana Boboc¹ & Marian-Gabriel Hâncean³

Previous correlative and modeling approaches indicate influences of environmental factors on COVID-19 spread through atmospheric conditions' impact on virus survival and transmission or host susceptibility. However, causal connections from environmental factors to the pandemic, mediated by human mobility, received less attention. We use the technique of Convergent Cross Mapping to identify the causal connections, beyond correlation at the country level, between pairs of variables associated with weather conditions, human mobility, and the number of COVID-19 cases for 32 European states. Here, we present data-based evidence that the relatively reduced number of cases registered in Northern Europe is related to the causal impact of precipitation on people's decision to spend more time at home and that the relatively large number of cases observed in Southern Europe is linked to people's choice to spend time outdoors during warm days. We also emphasize the channels of the significant impact of the pandemic on human mobility. The weather-human mobility connections inferred here are relevant not only for COVID-19 spread but also for any other virus transmitted through human interactions. These results may help authorities and public health experts contain possible future waves of the COVID-19 pandemic or limit the threats of similar human-to-human transmitted viruses.

The COVID-19 pandemic has produced one of the worst public health crises in the recent history of humankind. It has disrupted global health and economies, with consequences yet to amass. A possible strategy toward a comprehensive understanding of the processes behind its expansion and contraction consists of isolating the pandemic's main forcing factors, quantifying their impact, and identifying the channels through which their influences are transmitted¹.

Biological factors² are linked with the characteristics of the virus (e.g., transmissivity and resilience to changing living environments) and with nasal antiviral immunity (e.g., affected by cold exposure). Activities involving human interactions (like transport, work, study, shopping, relaxation, gatherings, and social events) can leverage the impact of these types of factors on the pandemic.

The fact that human society is embedded in a natural shell represented by the atmosphere, land, and ocean suggests a potential impact of environmental conditions on the virus' spread. As the development of herd immunity in populations takes a relatively long time, the global threat of SARS-CoV-2 remains high. Consequently, the effectiveness of public health measures depends on their alignment with environmental factors that influence the virus's spread³.

Coalescing evidence indicates that temperature and humidity are potential factors of COVID-19 spread^{4–6}, which proceeds through several paths. First, temperature and humidity modulate the virus's survival and spread⁷. Evidence collected from China suggests that in winter, temperature decrease enhances the COVID-19 survival time in the atmosphere, facilitating its transmission⁸. Consistent with these findings, the available measurements indicate that respiratory epidemics are significantly higher in winter than in summer⁹. Additionally, high temperature and humidity reduce COVID-19 circulation, possibly through faster evaporation, which can prevent

¹Faculty of Physics, University of Bucharest, Bucharest, Romania. ²Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest-Măgurele, Romania. ³Faculty of Sociology and Social Work, University of Bucharest, Bucharest, Romania. ✉email: denis.nichita@theory.nipne.ro; denisnichita@gmail.com

the spread of droplets responsible for transmission¹⁰. However, during the summer of 2020, several countries in the Northern Hemisphere experienced a significant increase in daily COVID-19 cases, dubbed the second pandemic wave³. These apparently contradictory aspects underline the need for a better understanding of the mechanisms through which environmental conditions affected the SARS-CoV-2 virus' spread. Moreover, recent results posit that through the modulation of airway defense mechanisms and the host antiviral innate immunity against respiratory virus infections, environmental factors influence host susceptibility¹¹. Furthermore, the impact of relative humidity on droplet spray in short-range transmission and on aerosol in long-range transmission could have an impact on COVID-19 spread, as it stimulates viruses and droplets to combine into larger drops in humid regions^{11–13}. Lastly, human behavioral patterns (e.g., human mobility) affect exposure time to specific weather conditions and contact rates between infected and susceptible individuals¹⁴.

The COVID-19 spread through human networks depends on social activities (e.g., work, commute, relaxation), occupations¹⁵, socio-demographic features^{16,17}, and the cycle of working days and weekend^{12,18}. Among these, human mobility is pivotal in virus transmission by modulating the people-to-people contact rate¹⁹. Whereas most previous studies have addressed the environmental factors' impact on virus survival time, host susceptibility, droplet spray in short-range and aerosols in long-range transmission, to our knowledge, research has approached less their influence on the pandemic through human mobility.

This study investigates the potential influences of atmospheric conditions on the number of COVID-19 cases through human mobility based on data from 32 European countries. The analyzed variables are temperature, precipitation, human mobility related to Parks (hereafter *Parks*), Retail and Recreation (hereafter *Recreation*), Transit stations (hereafter *Transit*), and the number of new daily cases (hereafter *Cases*) normalized to 1 million inhabitants, for each of the 32 countries (see Data availability). Note that *Recreation* refers to mobility to non-home, generally indoors recreational places and *Parks* to outdoor meeting points. So, our results can be interpreted as having an outdoors mobility variable (*Parks*), an indoors mobility variable (*Recreation*) and *Transit*. However, this correspondence is not exact in the case of *Recreation* (which also includes outdoors places) and consequently we stick to the first nomenclature. These variables cover the period from April 4, 2020, to December 31, 2021, to exclude data potentially influenced by vaccination. Potential causal links between pairs of time series are explored using the Convergent Cross-Mapping (CCM) technique, a state space time-embedding method for identifying causality between time series²⁰. An extension of this method, Time Delayed CCM (TDCCM), is used when causality manifests with a lag (See Methods).

Results

Impact of weather on Cases in Italy

We present findings obtained using TD-CCM and CCM methods for time series data related to Italy, a country that was pivotal during the initial stages of the pandemic²¹, in Europe (Fig. 1). This example is illustrative for our investigation of the causal relationships from weather variables to COVID-19 spread through mobility indicators. Notice that the *Precipitation* time series has daily variations, while the *Parks* record is dominated by the annual cycle (Fig. 1a,b). The *Cases* signal also includes several fluctuations extending over a few months. The focus in this work is on day-to-day variations, so using the method of Singular Spectrum Analysis, the annual cycle and the inter-monthly fluctuations are removed from *Parks* and *Cases* time series. The two pairs of time series formed from the initial un-altered *Precipitation* and residual *Parks records*, on one side, and residual *Parks* and residual *Cases* records, on another side, are used in the CCM algorithm to identify causality (Fig. 1a,b,c).

Cross-mapping from *Parks* to *Precipitation* as a function of the time lag leads to a clearly defined causal spike at $\tau = 0$, suggesting an in-phase causal influence of *Precipitation* on *Parks* time series (the green vertical line in Fig. 1d). The fact that the peak is narrow and much higher than the noise level associated with other lags indicates a robust causal signal. A necessary condition for causality is convergence above a statistical significance level, specifically above the 95% threshold with respect to cross-maps between randomized surrogate time series (see Methods, Statistical Significance). For this pair of time series, the strength of convergence is $\rho = 0.46$, far above the 95% significance level (Fig. 1e). The robustness of the causal link is reflected by the difference between the convergence and significance levels (see Methods) and has a value of $\beta = 0.34$. Therefore, *Precipitation* in Italy impacts population mobility, as reflected in the *Parks* variable.

The TDCCM representation of the cross-map from *Cases* to *Parks* exhibits a local peak at $\tau = -1$ days (Fig. 1f). The CCM is convergent up to $\rho = 0.44$ with the robustness of $\beta = 0.023$ (see Methods for the algorithm regarding the optimal lag choice for the low robustness pandemic—mobility links). Although less robust, it is statistically significant and equally strong as the previous link. Therefore, a complete causal chain, *Precipitation* → *Parks* → *Cases*, is indicated. Following the same structure (*Temperature* and *Precipitation* cause *Cases* through three human mobility indicators), compound causal links for other countries are further investigated. Details on the choice of embedding dimension and other CCM parameters are given in the Supplementary Information Figure S2, Supplementary Table S1, and Methods.

Impact of weather on Cases in Europe

The analysis for the case of Italy is extended to all 32 European countries. The statistically significant causal strength and robustness of the identified links are synthesized in Fig. 2. We also take into account spatial frequency (the number of countries in which a given link is statistically significant). In this section, the qualitative aspects of the results, which are expressions of the quantitative estimations, are presented, which are synthesized in Supplementary Table S2 and Methods.

Parks is influenced by *Temperature* and *Precipitation* across all of Europe (See Fig. 2Aa,d,Ba,d). The *Temperature* → *Parks* average strength of causality is lower than the *Precipitation* → *Parks* one (Supplementary

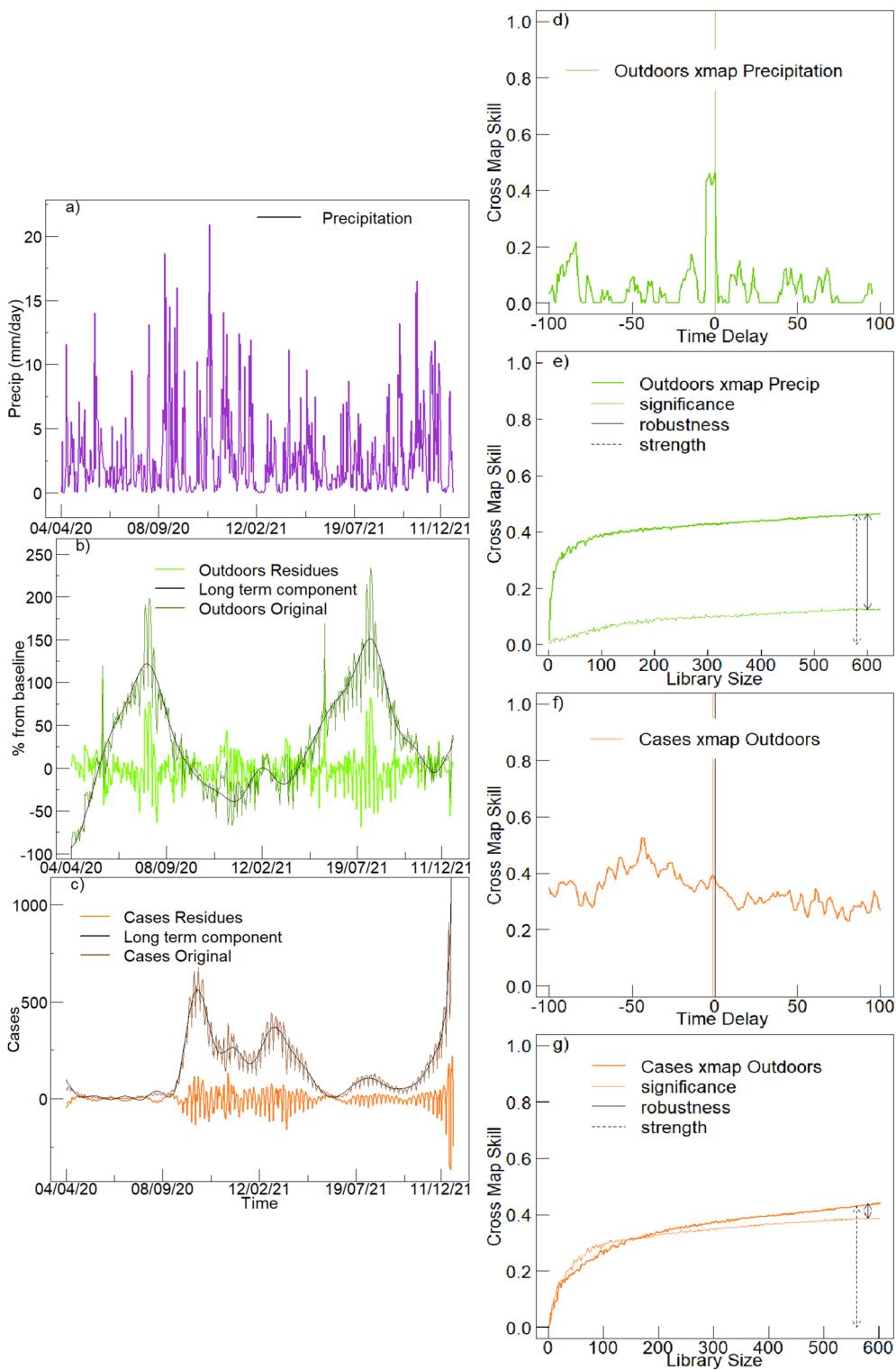


Figure 1. Convergent Cross-Mapping (CCM) analyses of one complete causal chain for Italy. Time series of variables used for causality analysis: (a) *Precipitation* (mm/day) (b) *Parks* mobility (% from baseline), and (c) the number of daily *Cases* scaled to 1 million inhabitants. The data spans the period April 4, 2020–December 31, 2021. In panels b) and c) dark green and dark orange lines represent the original time-series, black lines represent the long-term component identified through SSA to be removed from the original time-series; the bright green (bright orange) is the residual time series, which includes only day-to-day and week-to-week variations submitted to CCM. *Precipitation* is the only time series used unaltered in the CCM analysis. Panels (d) and (e) depict the complete CCM analysis associated with the *Precipitation* → *Parks* causal link. d) Time delay CCM between *Precipitation* and *Parks* with 0-lag for the influence of *Precipitation* on *Parks*. e) convergent, statistically significant, and very robust CCM for the influence of *Precipitation* on *Parks*. Panels (f) and (g) depict the complete CCM analysis associated with the *Parks* → *Cases* causal link: f) TDCM between *Parks* and *Cases* with negative lag for the influence of *Parks* on *Cases*; g) convergent and statistically significant CCM for the influence of *Parks* on *Cases*.

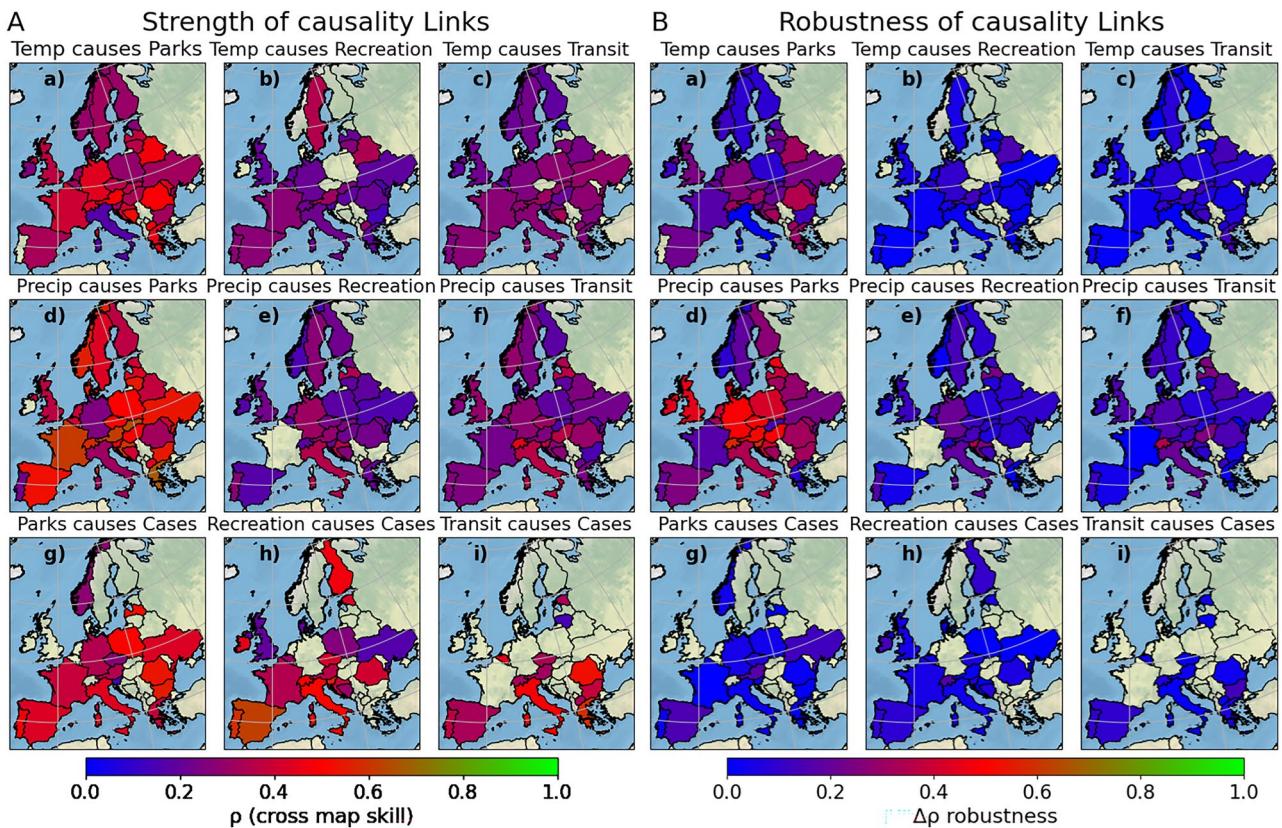


Figure 2. Maps of weather→mobility→pandemic causality links. In panel (A), each image reports the asymptotic cross-map skill of the CCM between the indicated variables following the ρ color bar. If a country does not have a color, it does not have a statistically significant causal link. The first two rows report the influence of weather factors on the social indicators: *Temperature influences Parks* (a), *Recreation* (b), and *Transit* (c), and similarly, *Precipitation influences Parks* (d), *Recreation* (e), and *Transit* (f). The last row represents the influence of the aforementioned social indicators on the number of daily cases of COVID-19 (*Cases*). From top to bottom, each column shows a potentially complete causal chain starting from *Precipitation/Temperature* to *Cases* through a sociological variable like *Parks* in (a)-(d)-(g), *Recreation* in (b)-(e)-(h) and *Transit* (c)-(f)-(i). In panel B, images follow the same structure as in panel A but report the Robustness instead of the Strength. The map imagery was obtained using Python libraries geopandas 0.14.3 <https://pypi.org/project/geopandas/> and Cartopy 0.22.0 <https://pypi.org/project/Cartopy/> with earth data taken from <https://www.naturalearthdata.com/>.

Table S2). Further down the chain, for the *Parks-Cases* link, although the causal signals are less spatially frequent (Fig. 2Ag,Bg), they are equally strong (Supplementary Table S2).

The links from *Temperature* and *Precipitation* to *Recreation* have a similar spatial fingerprint, with a slightly weaker signal (Fig. 2Ab,e). As we generally find, *Precipitation's* causal footprint is more robust (Fig. 2Be, Supplementary Table S2) compared to that of *Temperature* (Fig. 2Bb, Supplementary Table S2). The *Recreation*→*Cases* connections are less geographically frequent, but they are the strongest among the three mobility indicators (Fig. 2Ah, Supplementary Table S2).

Temperature→*Transit* (Fig. 2Ac,Bc) and *Precipitation*→*Transit* links are similarly distributed as the weather→*Parks* connections (Fig. 2Af,Bf, Supplementary Table S2), but the *Transit*→*Cases* links are the least spatially frequent out of all mobility to pandemic connections (Fig. 2Ai,2Bi, Supplementary Table S2).

The causal impact of *Cases* on mobility

As inferred from the CCM analyses, the influence of COVID-19 spread on mobility indicators is equally strong with the other links but less common among countries (Fig. 3). *Cases*→*Parks* (Fig. 3Aa,Ba) and *Cases*→*Recreation* (Fig. 3Ab,Bb) links encode similar statistics in spatial frequency (Supplementary Table S2), while the *Recreation* and *Transit* channels have the strongest causal fingerprints on average (Supplementary Table S2). While, on average, the strength of the pandemic→mobility is similar to that of the mobility→pandemic, the maximum ρ links are higher for the former direction than the latter (Supplementary Table S2).

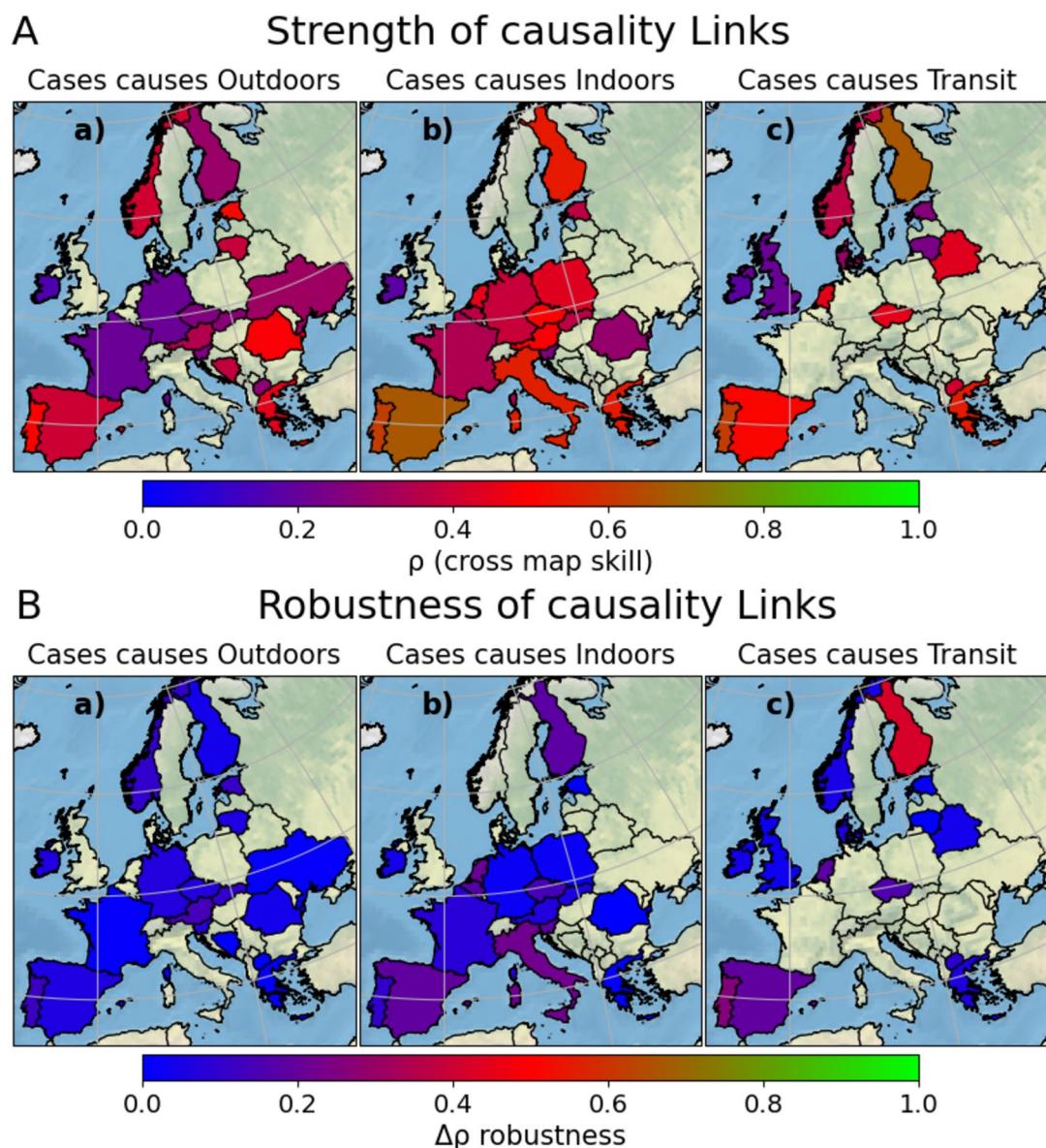


Figure 3. Maps of pandemic → mobility causality links. **(A)** Maps of causality links from the number of daily cases to mobility indicators **(a)** Cases causes Parks, **(b)** Cases causes Recreation, and **(c)** Cases causes Recreation. The associated CCM's maximum cross-map skill value determines each country's color, as indicated in the color bar. If a country does not have a color, it does not have a statistically significant causal link. **(B)** Robustness value for the strength values in panel **(A)**. Panel **(B)** has the same structure as panel **(A)**. The map imagery was obtained using Python libraries geopandas 0.14.3 <https://pypi.org/project/geopandas/> and Cartopy 0.22.0 <https://pypi.org/project/Cartopy/> with earth data taken from <https://www.naturalearthdata.com/>.

Complete causal chains

In order to investigate the spatial distribution of the causal channels through the three mediums, the numbers of complete weather → mobility → pandemic chains for European countries are determined and synthesized in Fig. 4. An example of a complete causal chain is *Temperature* → Parks → Cases, or *Precipitation* → Recreation → Cases. Italy is associated with the largest number of complete causal chains, six (three from *Temperature* and three from *Precipitation*), together with Austria. A total of 24 out of 32 countries show at least one complete causal chain, with an average of 3 chains per country.

Discussion

The first stage of our analysis comprised of causality assimilation between individual links (Fig. 2) belonging to pairs of two variables from the three systems taken into consideration here: weather (with *Temperature* and *Precipitation*), mobility (with *Parks*, *Recreation*, *Transit*) and the pandemic (with *Cases*). All the individual links are compared using three criteria: CCM strength, CCM robustness, and spatial frequency.

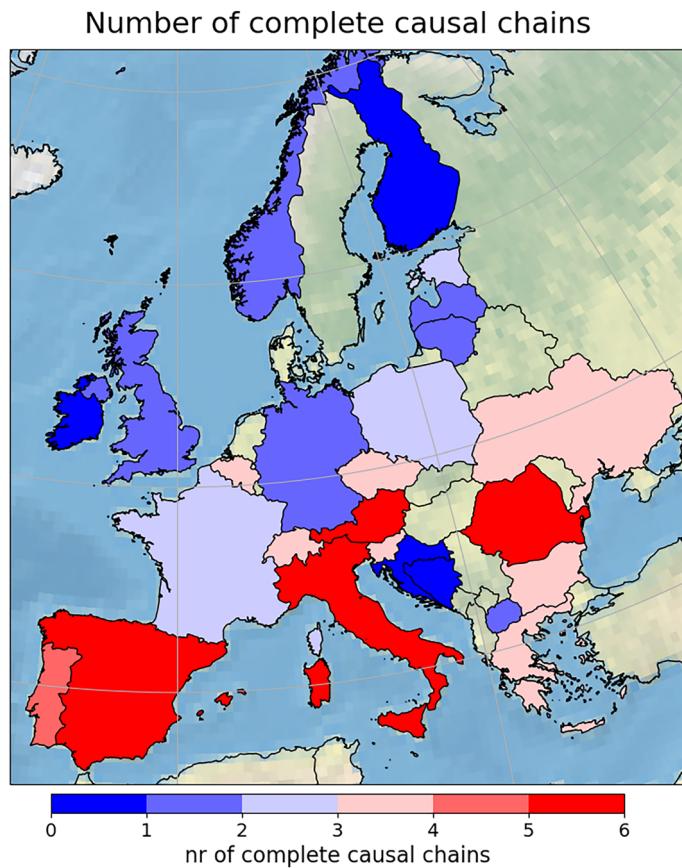


Figure 4. Map of complete causal chains. The number of complete causal chains for each country is counted, having the structure of a weather factor (*Temperature/Precipitation*) → mobility indicator (*Parks/Recreation/Transit*) → the number of daily cases (*Cases*). The maximum number of complete causal chains is six. If a country is left uncolored, it may contain particular causal links but no complete causal chains. The map imagery was obtained using Python libraries geopandas 0.14.3 <https://pypi.org/project/geopandas/> and Cartopy 0.22.0 <https://pypi.org/project/Cartopy/> with earth data taken from <https://www.naturalearthdata.com/>.

Out of all the 2-variable causal connections identified here, the influence from *Precipitation* on *Parks* mobility is the strongest, most robust, and spatially frequent one. (Fig. 2Ad,Bd). While *Parks* has the most frequent, strong, and robust channel for the weather → mobility link, the *Recreation* variable has the strongest channel for the mobility → pandemic segment (Fig. 2Ah). The *Parks* → *Cases* link follows closely, topping the *Recreation* → *Cases* link in frequency but not in strength (Fig. 2Ag, Supplementary Table S2). Overall, when considering strength, robustness, and frequency, the influence of weather factors on COVID-19 spread appears to manifest mainly through the *Parks* and *Recreation* channels.

The causal links between the weather factors and the mobility indicators are more robust and more spatially frequent, but the connections from mobility to the pandemic, albeit fewer of them, are stronger. As expected, this also shows that weather is not the primary factor determining the pandemic; a lot of weather → mobility links do not find a continuous causal link in the mobility → pandemic part. The fact that *Recreation* → *Cases* is the strongest causal link among the mobility → pandemic ones is not surprising, being consistent with the existing literature showing that the virus mainly transmits in indoors gathering locations^{22,23}. However, the fact that the 3-variable causal channel from *Precipitation* through *Parks* to *Cases* is so strong, robust, and spatially frequent may be somewhat counterintuitive at first sight. This influence is so pronounced in our analysis likely because it is a result of two causal mechanisms that are at play: (1) changes in contact rate modulated by weather factors, in combination with (2) the environmental conditions' impact on the virus survival probability, host susceptibility, or aerosol long-range transmission. Unlike this, *Recreation* mobility implies time spent in closed spaces, which are not under direct impact of weather conditions, and therefore, the impact of this factor on *Cases* manifests only through one causal mechanism: changes in the contact rate. Possibly because of this reason also, the spatial extension of the *Recreation* → *Cases* links is smaller than that of *Outdoor* → *Cases* connection (Fig. 2Ag,h). Furthermore, *Transit* likely implies less time spent in close spaces than *Recreation* and also a lower contact rate. Therefore, the spatial extension of the *Transit* → *Cases* links, together with its average strength and robustness, is reduced with respect to the other two channels (Fig. 2Ah,i). Notice that the value of the cross-map in the CCM analyses does not exceed 0.5 in any of the links found. This is to be expected since weather has only a marginal, modulating effect on the complex dynamics of the pandemic.

These results presented so far emphasize the influences of weather conditions on human decisions about involvement in different types of social activities, which in turn affects the people-to-people contact rates and the number of cases. While CCM provides information about the dynamical connection of variables, it is silent on the sign of that dependence. This could be approximatively inferred from correlation under the weak assumption that the dynamics are linear (see Methods, Linear correlations, and Supplementary Figure S5). *Precipitation* is negatively correlated with *Parks*, *Recreation*, and *Transit*, while all three mobility indices are positively correlated with *Cases* in most countries (See Methods, Linear Correlations for further details). This likely reflects people's tendency to stay at home on days with significant precipitations, and consequently, the virus is less likely to spread. Therefore, *Precipitation* acts as a natural lockdown mechanism, reducing the transmission of the virus. Note that this channel is the strongest, most robust, and most frequent out of all three, implying that *Precipitation* plays a key role in containing the spread of the virus. On the other hand, *Temperature* is positively correlated with all three mobility indices, having the inverse effect: people tend to favor social gatherings of all sorts when the temperature on a certain day is relatively high. This increases the rate of transmission of the virus. These conclusions could explain the peaks of cases recorded during the summer of 2020 in countries of the Northern Hemisphere³.

There is also a causal impact of pandemic dynamics on mobility indicators (Fig. 3), which could be generated by social distancing measures imposed by governments and by the fear of contagion which determines people to spend more time at home. One notes the contrast between the positive-signed influence of mobility on the pandemic (Figs. 2, S5) and the seemingly negatively-signed influence of the latter on the former. This is a mark of non-linear dynamics involving negative feedback. For instance, if mobility increases the number of cases, the state will take stronger measures to limit it, which in turn decreases the number of cases. When cases are low, the measures are relaxed, increasing the cases right back up (if population immunity has not been reached). One can also think of two processes: one where an increase in mobility increases contact rate and, therefore, the number of cases and another one where the increase in cases decreases mobility through governmental isolation measures and fear of contagion. We find a predominantly positive statistical correlation between mobility and pandemic in more than 50% of the countries analyzed (See Methods, Linear correlations). However, the percentage is much lower in the mobility ↔ pandemic than in the case of weather → mobility, where causality was unidirectional (mobility does not influence weather) and there was only one process at play. Indeed, when performing partial linear correlations to control for other covariates, while the weather → mobility signs remain unchanged, there appears to be a predominantly negative sign between mobility and pandemic variables, emphasizing the causal mechanism of Fig. 3. Consequently, because of the inherent non-linearity of the systems, we interpret the positive signs as capturing the mobility → pandemic influence, while the negative sign pertains to the reverse causal direction.

Having analyzed the individual links, in the last stage of our work, we track the transitive influence from the weather to the pandemic by counting the total number of causal chains in each country. We recognize that on the one hand, some weather → mobility links may not find continuation further down the causal chain, while on the other hand, some mobility → pandemic links may be caused internally by socio-demographic conditions with no influence from the weather. However, by tracing the complete causal chains, we can reveal the influence of *Temperature* and *Precipitation* on the viral spread (Fig. 4).

The southern Mediterranean countries, like Italy (6), Portugal (5), Greece (4), and Spain (6) have a relatively high number of complete causal chains, while northern countries like Finland (1), Norway (2), Denmark (0), and Sweden (0) show lower numbers. In middle Europe, there is a mixture of these two situations. This is consistent with the high number of days/year with precipitation for northern European countries²⁴, which contribute to fewer gatherings and a reduced contact rate, with limited impact on *Cases* and a reduced number of complete causal chains. On the other hand, in Mediterranean countries, where the number of days/year with precipitation is relatively low²⁴, *Temperature* plays a dominant role, with warm days favoring human interactions, increased contact rate, and number of daily cases. This qualitative interpretation is confirmed by a quantitative estimation, according to which the mentioned northern countries added a total of 2.78 million cases as of December 2021, while the southern countries registered 15.02 million cases. Of course, the distribution of the total number of cases is not solely and primarily caused by weather but also has internal policy components, socioeconomic and demographic conditions (like in Sweden), etc. These results just underline the fact that weather is a contributing factor in containing or expanding the pandemic. Relaxed policies, for example, are more effective in the North of Europe, where precipitation already acts as a natural lockdown mechanism.

Most previous studies have focused on the influence of atmospheric conditions on pandemics through their impact on virus survival, host susceptibility, and small particle transmission through air^{2,7,9–13}. Here, we focus on the influence of temperature and precipitation on the number of COVID-19 cases mediated by population mobility. In contrast with previous ecological studies based on correlations^{25–27}, we use the CCM technique to identify causal connections based on pairs of time series and cautiously use correlation only as an indicator of the sign of causality. However, even though CCM reveals dynamic information and not mere correlation, ecological study design (based on time series) still represents a more challenging condition to identify causal mechanisms than Randomized Control Trials (RCTs). This is because, on the one hand, our results could be affected by ecological bias (difference between individual and aggregate level associations), but also because of a possible bias due to measured and unmeasured covariates and confounders.

Another limitation of our study is represented by the coarse-grained details of population networks in the mobility time series. As shown in previous studies^{15,16,19}, the structure of the human network plays a key role in the distribution of the disease. We leave for future work the investigation of weather's influence on a more intricate network structure, which in turn affects the propagation of the virus. Related to this, another way of aggregating the weather data could be using a population-weighted measure. Mobility indices and Cases are clearly affected by the distribution of population and the understanding of the causal patterns present here could

benefit from a more elaborate geographical approach. However, not taking the distribution of population into account only comes to our disadvantage, since we're using weather data from places where no one lives. The fact that we find causality despite using aggregate data without population weighting means that the causal signal is strong enough to be identifiable by CCM.

The *Temperature* and *Precipitation* variables used in this work, while representative of the weather system, are by no means exhaustive. Other studies used variables such as humidity⁴, wind^{28,29}, and atmospheric circulation³⁰, which could influence the virus's transmission directly or through changes in population measures, such as mobility. Also, other mobility variables, such as Residential, Workplaces, Grocery&pharmacy³¹ may mediate the influence of weather factors on COVID-19. Lastly, different measures for the time evolution of the pandemic could have been chosen, such as the mortality rate or cumulative cases. All these choices of variables could be employed in the ecological study design presented here, which overcomes the limitations of correlation using Convergent Cross Mapping.

Convergent Cross Mapping is not the only analytic method available for the study of causality non-linear systems. Information Flow, for example could be equally suitable in analyzing the causal patterns observed here^{32–34}, but also other methods such as Structural Equation Modelling³⁵.

One possible concern regarding the analysis in our paper could be that weather has potential intertemporal substitution effects. For example, suppose that precipitation today just leads individuals to delay outings until tomorrow (or whenever it stops raining), not canceling those outings forever. In that case, precipitation-via-mobility will just affect the temporal pattern, not the cumulative amount of cases. However, Time Delay Convergent Cross Mapping can distinguish such a situation. TDCCM is a similar method to CCM, with the difference that instead of plotting cross-map skill against library length, one plots cross-map skill against time delay. Consequently, it explores potential causal relationships, considering various delays between the two variables. Our analyses show that causality has the strongest signal for synchronous variations (no time delay between variables; see Supplementary Figs. S2a–d and 1d). The width of the peak at half-maximum is an artifact of the optimal embedding dimension (see Fig. S6), and the maximum lag lies within the embedding vector, $-(E-1) \leq \tau \leq 0$, consistent with a contemporaneous interaction³⁶. Moreover, the cumulative effect is smaller than the lag 0 effect for all countries for the *Precip* causes *Parks* links, as measured by partial correlations (see Methods). Therefore, in this case, we do not detect any intertemporal substitution effects as the causal relationships between weather and mobility appear to manifest without any delay.

In this study, we present empirical evidence of consistent causal influences of *Precipitation* and *Temperature* on the pandemic spread through population mobility channels, represented by *Parks*, *Recreation*, and *Transit*. *Parks* mobility mediates the strongest and most robust influence of *Precipitation* on *Cases*, with this weather variable acting as a natural lockdown mechanism and therefore limiting the virus' spread without human intervention. Negative correlations are also found between precipitation and the other two mobility indices. This type of causal connection can explain why northern Europe shows a lower number of cases than its southern counterpart.

Unlike this, the impact of *Temperature* through *Parks*, *Recreation* and *Transit* on *Cases* is positively correlated, causality being the strongest through *Recreation* mobility. This reveals people's propensity to participate in social gatherings during warm weather conditions, which increases the person-to-person contact rate, especially in closed spaces. Southern Mediterranean countries may be the most vulnerable to this causal mechanism, and our results may explain the increased number of cases in these countries during summer.

Moreover, these results could help in designing measures targeting an optimal impact of non-pharmaceutical interventions on pandemics and a minimum negative impact on the economy. For example, even though we do not have control over environmental conditions, Northern European countries might indulge in more relaxed measures since *Precipitation* already acts as a natural lockdown mechanism. However, in Southern countries, more drastic measures are required since *Temperature* is the dominating weather factor. More flexible isolation measures could also be imposed, taking into consideration weather prognosis, since on rainy days, people will tend not to gather together, therefore naturally containing the spread of the virus.

The causal links between mobility indicators and the number of cases could also be relevant for spreads related to viruses other than COVID-19. Complementary, the influence of the pandemic on population mobility and its impact on the economy is also general and could hold true for other viruses.

Methods

Singular spectrum analysis (SSA)

Some of the time series show relatively long-term variations. For example, *Parks* has a pronounced annual cycle signal (Fig. 1b), whereas *Cases* displays fluctuations extending over a few months (Fig. 1c). Such oscillatory signals include only 1-to-3 complete cycles during the 21-month time window of interest. Consequently, one cannot infer any reliable statistical significance based on them. Therefore, these components are separated and removed from the time series through a preliminary Singular Spectrum Analysis (SSA). In effect, the residual signals contain only inter-daily and inter-weekly variations, on which our study is focused (Fig. 1a–c).

SSA is a powerful technique to identify and separate trends, oscillatory signals, and noise in time series^{37,38}. It extracts information from short and noisy time series by providing data-adaptive filters by using an eigenvector and eigenvalues analysis applied in the time domain³⁹. The shapes of the trends are inferred from the data without requiring a preliminary assumption about their form. Unlike classical Fourier analysis, the oscillations identified with SSA can be modulated in amplitude and phase. The identification and removal of signals that are not of interest or the reconstruction of the initial time series based only on a reduced number of components can be very effective in increasing the signal-to-noise ratio in data.

Convergent cross mapping

Convergent Cross Mapping (CCM) is an empirical method based on the theory of dynamical systems that uses observational time series to identify causal footprints beyond correlation, comprising necessary and sufficient criteria for causality²⁰.

In simple terms, CCM performs the following operations. Let us assume that X causes Y. In dynamical systems, if X causes Y, Y contains information about X, not vice versa. So, instead of using X to predict Y to check for causality, CCM employs Y to predict X. The prediction is done using a time-embedded reconstruction. Take the time series of Y, $Y(t)$. Shift the time series by one unit, obtaining $Y(t-1)$ and then again obtaining $Y(t-2)$. Build a 3-dimensional system composed of $Y(t)$, $Y(t-1)$, and $Y(t-2)$. This is called an embedding. Using the vector $[Y(t), Y(t-1), Y(t-2)]$, CCM aims to estimate the value of $X(t)$ for all values of t . This is the technique called cross-mapping (see Supplementary Fig. S1 for a causality circuit, which visually illustrates cross-mapping). After all values of $X(t)$ have been estimated, Pearson's correlation coefficient, ρ , is employed between X_{observed} and $X_{\text{predicted}}$ to check how good the prediction was. The stronger the causal influence of X and Y is, the higher the cross-map skill, ρ will be. The prediction of X is made in consecutive steps of larger and larger library lengths, meaning that at first, not all the values of the time series of Y are used, but rather only a certain proportion of them. Adding more and more values of Y to the prediction of X should increase the estimation ability. This increase in cross-map skill is called convergence and is a necessary criterion for causality.

In more technical terms, the identification of causal footprints is made by leveraging Takens' Theorem of state space reconstruction, according to which one can reconstruct the manifold of the trajectories of a multi-dimensional system using only lags of the time series of only one variable. Suppose two variables belong to the same dynamical system. In that case, Takens' Theorem states that both of their time series can be used to build an embedding manifold to reproduce the original attractor manifold since both manifolds are diffeomorphic with it as well as with each other. Thus, a practical criterion for causality is born. Taking the time series of the effect, constructing a time-embedded manifold, and estimating the contemporaneous states of the cause (a technique called cross-mapping), one could obtain an indication that the effect contains information about the cause and that the latter has influenced the former. It is important to note that CCM works even when diffeomorphism is not satisfied between the cause and the effect's manifolds. This is because CCM relies on the fact that causes contain information about the effects and not on the strict geometrical relationship between the manifolds. The dimension E of the time embedding has to be predetermined before the CCM analysis (see Embedding dimension and embedding lag below).

However, in order to infer causality, one also needs convergence, which is represented by the increase in cross-map skill with the number of points (or library size) used for prediction. Ideally, if small increments in library size are used, one obtains an asymptotic increase to a saturation level. This reflects that the available causal information has been exhaustively harnessed, and the level of convergence is a measure of the strength of the influence. Convergence gives confidence to the cross-map, ensures noise reduction, diminishes sample sensitivity, and implies that the causal relationship identified is consistent across different segments of the time series, implying generalizability of the causal relationship across different time scales and conditions. Cross-mapping and convergence represent necessary and sufficient criteria for causality.

Time delayed convergent cross-mapping

Causality can manifest with a lag. Time Delayed Convergent Cross-Mapping (TDCCM) is a representation in which the cross-map skill at the maximum library length is computed as a function of the prediction time (or time delay). The causal signal and lag are identified by a peak in the cross-map at the time delay when the effect manifests relative to the cause. The time of prediction is negative since the prediction runs backward to the past (from effect to cause). Suppose the maximum cross-prediction is situated at a positive lag. In that case, that direction of causation is spurious and should be disqualified because it violates the causal dogma of causes preceding effects. Therefore, a strong unidirectional forcing manifested in TDCCM at a negative symmetric lag is the first indicator of a causal signal.

TDCCM is more than a mere tool to identify the lag. It actually shows how robust the causal signal is. For example, if the cause influences the effect with a lag of five days, ideally, a peak of the cross-mapping signal at $\tau = -5$ appears. The peak should be over and above the random cross-mapping between the time series at different lags, and its width at half of the height should be of the order of several time units. If a peak is narrow and well above the noise in the TDCCM representation, then it corresponds to a robust causal link. At its essence, robustness measures how well the causal connection is encrypted in the time series. In practice, the encryption of the causal signal is hindered by process noise, measurement errors, numerical truncations, etc. Also, multiple causal signals from other variables in a dynamical system can induce several peaks in the TDCCM, further complicating the clear identification of a maximum.

The TDCCM representation of the cross-map from Cases to all three variables of mobility is usually less robust than the case of mobility to weather, for example. Because of that, we don't see a clear, narrow signal in TDCCM; instead, we see a few fluctuations that hinder the choice of a lag that reflects the true causal connection. In order to find the optimal lag for these links, we have to take into consideration the fact that the estimated reporting time from infection can be at most 2 weeks and that a maximum at a lag within the embedding vector, $-(E-1) \leq \tau \leq 0$ is considered instantaneous³⁶. As a result, we choose a local maximum at a lag between -30 years and 30 years, together with an embedding dimension of $E=1-20$. If the difference between $|\tau|$ and E is larger than 14 days, we disregard that link as non-causal since it's a result of a random fluctuation and not a causal signal manifested at a lag that makes physical sense. If the optimal lag τ is positive, we also disregard that lag as being non-sensical, violating the dogma of causes preceding effects.

Equipped with a corresponding lag and direction of causality, a check for convergence and statistical significance is required. If the asymptotic level of convergence is above the statistical significance level, then a causal link between the time series is detected. Whereas the strength of the causal link is reflected by the level of convergence (ρ), the robustness of the causal signal (β) can be measured by the difference between the convergence and the 95th significance level: $\beta = \rho - \rho_{\text{sign}}$ (Fig. 1 and Supplementary Fig. S2). To show that the accuracy of a TDCCM representation is reflected in the difference between the strength of CCM and the statistical significance level, we have plotted four clear TDCCM signals (Supplementary Fig S2 a–d) and their corresponding CCM representations (Supplementary Fig. S2e–h). One can see that accurate signals correspond to high robustness. Contrastingly, less accurate TDCCM causal signals (Supplementary Fig. S2i–l) show lower robustness (Supplementary Fig. S2m–p).

The robustness of the link measured this way is always less or equal to the strength of causality but not necessarily correlated with it. There could be a relatively weak (e.g., $\rho \approx 0.3$) but very robust causal signal ($\beta \approx 0.25$) and also a very strong ($\rho \approx 0.8$) but not so robust one ($\beta \approx 0.01$). If the strength of causality is too low ($\rho < 0.1$), it is not an extremely robust causal signal. Note that robustness has meaning only in conjunction with the strength of the causal link.

Embedding dimension and embedding lag

Various methods exist to find the embedding dimension necessary to compute cross-mapping, such as Simplex Projection⁴⁰. Here, the cross-mapping of one variable is computed from another for different embedding dimensions (E), and the dimension for which cross-mapping is maximum is the optimal one. Further increasing E after this point is redundant and may result in saturation or a decrease in cross-map skill. We label this representation Embedding Dimension CCM (EDCCM)⁴¹. The EDCCM for the two CCMs in Fig. 1 of the main article are plotted in Supplementary Fig S3. The cross-prediction maxima occur at E=6 for *Parks* xmap *Precip* and E=9 for *Cases* xmap *Parks*.

For all analyses in this work, an embedding lag of l=1 was used.

The parameters of the CCM analyses used in Fig. 1 in the main material are given in Supplementary Table S1 and Supplementary Fig. S3, and the statistics presented in the main text for Fig. 2 are shown in Supplementary Table S2.

Strong unidirectional forcing

Sometimes a very strong and robust causal signal may induce a false positive signal in the TDCCM representation in the opposite direction. In Supplementary Fig. S4 we illustrate this phenomenon by comparing the TDCCMs of a weakly coupled model with a strongly coupled one. Variable X is the independent variable, and it causes variable Y. The model equations are included below. A weak causality model has $\beta = 0.1$, while a strongly coupled one has $\beta = 0.9$. Initial conditions are $X(1) = 0.4$ and $Y(1) = 0.2$. A plot of the time series of variables X and Y and the bidirectional TDCCM is shown in Supplementary Fig. S4. One can observe that the coupled model shows two maxima when, in fact, there's only one causal signal. This can also happen with real-time series, and in that case, further independent arguments must be brought forward to support reversed causality (physical mechanism, other statistical methods like Information Flow³²). The virtual causal peak is nevertheless for a positive lag, which indicates its falseness. In real-time series, noise can influence the location of the peak so as to shift it to a negative lag. Also, in an instantaneous causal interaction, a peak in the opposite direction of causality cannot be avoided (since the lag is 0, or $\tau < 2^*E-1$)²⁰.

$$X(t+1) = X(t)(3.8 - 3.8X(t))$$

$$Y(t+1) = Y(t)(3.5 - 3.5Y(t) - \beta X(t))$$

Statistical significance

To quantify the statistical significance of each convergent cross map, we build 100 surrogates for the cause's time series using two models: SWAP and Ebisuzaki. We perform CCM between the effect and the 100 surrogate causes and consider that ρ of the 95th cross-map order represents the significance level. This operation is performed for both models, and the least favorable and highest one is plotted in the CCM figures.

The SWAP model consists of choosing a random point in the time series and swapping the two segments separated by the point. This changes local interdependencies in the time series. The Ebisuzaki model consists of randomly changing the phases of the time series while preserving the power spectrum⁴². These randomization procedures ensure that the cross maps obtained in the study are not due to mere chance.

Linear correlations

CCM is a powerful method that accommodates non-linear dynamics. Correlation does not imply causation, but under the weak assumption of linearity, correlation may indicate information about the sign of causality. As such, linear fits on scatter plots are computed to investigate the sign of causality. However, the weather-mobility-pandemic connections analyzed are complex and non-linear, which may hinder a clear causative sign. Moreover, bidirectional causality may further complicate the linear fit. For example, treating *Parks* as the cause, its increase should boost the number of cases because of a higher contact rate, while treating *Cases* as a cause should decrease the *Parks* mobility due to lockdowns and fear of contagion. Therefore, the mobility ↔ pandemic links have two competing processes with opposite signs of causality. When computing a sign for the causal connection using correlation, we don't know which variable is the cause and which one is the effect and thus cannot know if the

given sign belongs to the CCM results of Fig. 2Ag–i or Fig. 3Aa–c. However, the weather → mobility links are causally unidirectional, and a more consistent sign is expected. Thus, we found a negative correlation between *Precipitation* and the three mobility indicators and a positive correlation between *Temperature* and mobility. When it comes to the slope between *Cases* and mobility, using linear correlation, we predominantly find a positive correlation in more than 50% of the countries. However, not all of them show this sign, and in some countries, it is negative, reflecting two differently-signed competing processes.

The data used for scatter plots are the original data from which the long-term component was removed with SSA and normalized with standard deviation. We make nine scatter plots and nine linear fits for each country and assess their slope and p-value (see Supplementary Fig. S5 for Italy). We deem a fit statistically significant if its p-value is less than $p < 0.05$.

The *Precip* vs. mobility fits are negative in all countries (100%), with a very high rate of statistical significance. The *Temperature* vs. mobility links are all above 80% positively correlated with a rate of statistical significance as high as 100%. Lastly, *Cases* vs. mobility are generally positively correlated for all links in at least 50% of the countries within each link (statistical significance rates ranging from 29.4% for *Cases* vs. *Parks* to 72.4% for *Cases* vs. *Recreation*). The lower percentage, in this case, is due to the two competing processes mentioned in the previous paragraph (increased contact rate vs isolation and fear of contagion). The highest percentage of positive signs is for *Cases* vs *Recreation* (90.6%), which is not surprising since *Recreation* is the strongest causal channel among the mobility → pandemic ones. In the *Parks* and *Transit* links, the pandemic → mobility influence is a little more prominent but still not dominating. The exact statistics are given below:

- *Precip* vs. *Parks*: 32/32 (100.0%) negative, of which 32/32 (100.0%) were statistically significant
- *Precip* vs. *Recreation*: 32/32 (100.0%) negative, of which 28/32 (87.5%) were statistically significant
- *Precip* vs. *Transit*: 32/32 (100.0%) negative, of which 32/32 (100.0%) were statistically significant
- *Temp* vs. *Parks*: 32/32 (100.0%) positive, of which 32/32 (100.0%) were statistically significant
- *Temp* vs. *Recreation*: 27/32 (84.4%) positive, of which 13/27 (48.1%) were statistically significant
- *Temp* vs. *Transit*: 31/32 (96.9%) positive, of which 24/31 (77.4%) were statistically significant
- *Cases* vs. *Parks*: 17/32 (53.1%) were positive, of which 5/17 (29.4%) were statistically significant
- *Cases* vs. *Recreation*: 29/32 (90.6%) were positive, of which 21/29 (72.4%) were statistically significant
- *Cases* vs. *Transit*: 21/32 (65.6%) positive, of which 14/21 (66.7%) were statistically significant

Partial correlations

Even though we only use correlation as an approximative indication of the sign of the causal relation identified with CCM, we also calculated partial correlations for the above pairs in order to control and remove the possible confounding effect of the other measured causal factors, as shown below. Of course, the confounding effect is just regarding the sign, not causality, since this is a mere correlation. The results are consistent with the simple linear correlations when it comes to the unidirectional weather → mobility links and also for the *Cases* vs *Recreation* link (since *Recreation* is the strongest channel for virus transmission). The exception situations are for *Cases* vs *Parks* and *Cases* vs *Transit* where we find a negative sign in more than 50% of the countries. This occurs because of the elimination of the *Recreation* variable and a proliferation of the negatively signed causal signal coming from *Cases*. These results show the influence of *Cases* through the limitations imposed by governments to explicitly not go outside, together with the travel restrictions. The statistics are shown below:

- *Precip* vs. *Parks* controlling for (*Temp*): 32/32 (100.0%) negative, of which 32/32 (100.0%) were statistically significant
- *Precip* vs. *Recreation* controlling for (*Temp*): 32/32 (100.0%) negative, of which 28/32 (87.5%) were statistically significant
- *Precip* vs. *Transit* controlling for (*Temp*): 32/32 (100.0%) negative, of which 32/32 (100.0%) were statistically significant
- *Temp* vs. *Parks* controlling for (*Precip*): 32/32 (100.0%) positive, of which 32/32 (100.0%) were statistically significant
- *Temp* vs. *Recreation* controlling for (*Precip*): 27/32 (84.4%) positive, of which 12/27 (44.4%) were statistically significant
- *Temp* vs. *Transit* controlling for (*Precip*): 30/32 (93.8%) positive, of which 23/30 (76.7%) were statistically significant
- *Cases* vs. *Parks* controlling for (*Recreation*, *Transit*): 22/32 (68.8%) negative, of which 7/22 (31.8%) statistically significant
- *Cases* vs. *Recreation* controlling for (*Parks*, *Transit*): 26/32 (81.2%) positive, of which 18/26 (69.2%) were statistically significant
- *Cases* vs. *Transit* controlling for (*Outdoors*, *Recreation*): 20/32 (62.5%) negative, of which 13/20 (65.0%) were statistically significant

We calculate the slope coefficient between *Precip* and *Parks* at lag 0 and the sum of the coefficients for lags of *Precipitation* between – 10 and – 1. As shown already above in the Linear Correlations section all contemporaneous correlations are negative at lag 0. The statistically significant sums of coefficients for ten lags of *Precip* are positive for 12 countries and negative for 5. The rest are not statistically significant. Here are the results:

Positive: 12

Austria: 1.9080

Belarus: 0.8255

Bosnia: 0.2072
 Bulgaria: 0.7389
 Czechia: 0.5166
 Estonia: 1.0157
 Ireland: 0.5978
 North Macedonia: 0.3849
 Norway: 3.4622
 Romania: 1.8031
 Slovenia: 0.6583
 Spain: 0.6113
Negative: 5
 Italy: - 0.6766.
 Moldova: - 0.0966
 Portugal: - 1.4202
 Serbia: - 0.6079
 United Kingdom: - 0.5638

However, any sum coefficient for any country is smaller than the corresponding coefficient for lag = 0. We believe this suggests that even though there may be some intertemporal substitution effects, these effects are smaller on our data than the contemporaneous ones. Moreover, TDCCM doesn't show any evidence of intertemporal substitution but rather only common embedding dimension effects (Fig. S6). As correlations don't imply causation, we believe that intertemporal substitution effects, if present, are only marginal.

Data availability

For the CCM analysis in Fig. 2, 3, and 4 in the main material, we have chosen only countries with a population larger than one million inhabitants, comprising the following states: Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Moldova, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, and the United Kingdom. We have excluded from the list Albania and Serbia, for which there is no complete available data, as well as Andorra, Liechtenstein, Luxembourg, Malta, Monaco, Montenegro, San Marino, Iceland, and Vatican City, which all have a population below 1 million. As the pandemic unfolded, Google devised *Community Mobility Reports* to provide insights about people's responses to the mobility policies³¹ imposed worldwide to contain COVID-19 spread. These reports have tracked changes in human movement from logs of the Google Maps app across geographical regions in many countries (e.g., the number of visits to grocery stores in Italy in a specific time window). We use three daily human mobility indicators. The first one comprises the movements in parks, public beaches, marinas, plazas, and public gardens (hereafter *Parks*). The second, retail and recreation mobility (hereafter, *Recreation*), refers to mobility in restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theatres. The third mobility indicator is related to public transport hubs such as subway, bus, and train stations (hereafter, *Transit*). These variables show how human mobility changes relative to a baseline score computed by Google between January 3 and February 6, 2020. Further, we use the number of daily new cases (hereafter *Cases*) normalized to 1 million inhabitants in each of the 32 analyzed countries. At the country level, variations in these variables are quantified by time series of daily values extending between April 4, 2020, and December 31, 2021. We use this time period to analyze data that was not affected by vaccination. We collected the data set used for this study from Google: COVID-19 Community Mobility Reports (<https://www.google.com/covid19/mobility/>). Covid-19 data has been downloaded from John Hopkins University (<https://coronavirus.jhu.edu/about/how-to-use-our-data>). For replication purposes, sample data for Italy can be downloaded from the following electronic repository: <https://doi.org/10.6084/m9.figshare.23739681>. The weather variables used in the analyses are the daily temperature at 2 m height (T2m), expressed in K, and total precipitation (Tp), expressed in mm/day, from the ERA5 dataset^{43,44}. Tp reflects the total liquid quantity on Earth's surface, in both liquid and frozen form (i.e., rain, snow). These ERA5 variables are distributed on $0.5^\circ \times 0.5^\circ$ horizontal grids and on 37 atmospheric levels ranging from the ground to 80 km altitude. Time series of daily area averages of these two variables were constructed for each country for the April 12th 2020–December 31st 2021 period.

Code availability

The code used in this analysis is available through <https://doi.org/10.6084/m9.figshare.23739681>.

Received: 10 August 2023; Accepted: 17 July 2024

Published online: 01 August 2024

References

- 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**(1), 4312. <https://doi.org/10.1038/s41467-020-18150-z> (2020).
- Huang, D. *et al.* Cold exposure impairs extracellular vesicle swarm-mediated nasal antiviral immunity. *J. Allergy Clin. Immunol.* **151**(2), 509-525.e8. <https://doi.org/10.1016/j.jaci.2022.09.037> (2023).
- Paraskevis, D. *et al.* A review of the impact of weather and climate variables to COVID-19: In the absence of public health measures high temperatures cannot probably mitigate outbreaks. *Sci. Total Environ.* **768**, 144578. <https://doi.org/10.1016/j.scitotenv.2020.144578> (2021).
- Mecenas, P., Bastos, R. T. R. M., da Vallinoto, A. C. R. & Normando, D. Effects of temperature and humidity on the spread of COVID-19: A systematic review. *PLoS ONE* **15**(9), 0238339. <https://doi.org/10.1371/journal.pone.0238339> (2020).

5. Fontal, A. *et al.* Climatic signatures in the different COVID-19 pandemic waves across both hemispheres. *Nat. Comput. Sci.* **1**(10), 655–665. <https://doi.org/10.1038/s43588-021-00136-6> (2021).
6. 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**(1), 3602. <https://doi.org/10.1038/s41467-021-23866-7> (2021).
7. Sajadi, M. M. *et al.* Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19). *JAMA Netw. Open* **3**(6), e2011834. <https://doi.org/10.1001/jamanetworkopen.2020.11834> (2020).
8. Guo, X.-J., Zhang, H. & Zeng, Y.-P. Transmissibility of COVID-19 in 11 major cities in China and its association with temperature and humidity in Beijing, Shanghai, Guangzhou, and Chengdu. *Infect. Dis. Poverty* **9**(1), 87. <https://doi.org/10.1186/s40249-020-00708-0> (2020).
9. Kumar, M. *et al.* A chronicle of SARS-CoV-2: Seasonality, environmental fate, transport, inactivation, and antiviral drug resistance. *J. Hazard. Mater.* **405**, 124043. <https://doi.org/10.1016/j.jhazmat.2020.124043> (2021).
10. Demongeot, J., Flet-Berliac, Y. & Seligmann, H. Temperature decreases spread parameters of the new Covid-19 case dynamics. *Biology* **9**(5), 94. <https://doi.org/10.3390/biology9050094> (2020).
11. Moriyama, M., Hugentobler, W. J. & Iwasaki, A. Seasonality of respiratory viral infections. *Annu. Rev. Virol.* **7**(1), 83–101. <https://doi.org/10.1146/annurev-virology-012420-022445> (2020).
12. Bhardwaj, R. & Agrawal, A. Likelihood of survival of coronavirus in a respiratory droplet deposited on a solid surface. *Phys. Fluids* **32**(6), 061704. <https://doi.org/10.1063/5.0012009> (2020).
13. Yang, X.-D., Li, H.-L. & Cao, Y.-E. Influence of meteorological factors on the COVID-19 transmission with season and geographic location. *Int. J. Environ. Res. Public. Health* **18**(2), 484. <https://doi.org/10.3390/ijerph18020484> (2021).
14. Wu, Y., Mooring, T. A. & Linz, M. Policy and weather influences on mobility during the early US COVID-19 pandemic. *Proc. Natl. Acad. Sci.* **118**(22), e2018185118. <https://doi.org/10.1073/pnas.2018185118> (2021).
15. Hâncean, M.-G. *et al.* Occupations and their impact on the spreading of COVID-19 in urban communities. *Sci. Rep.* **12**(1), 14115. <https://doi.org/10.1038/s41598-022-18392-5> (2022).
16. Hâncean, M.-G. *et al.* The role of age in the spreading of COVID-19 across a social network in Bucharest. *J. Complex Netw.* **9**, cnab026–cnab026 (2021).
17. Hâncean, M.-G. *et al.* Disaggregated data on age and sex for the first 250 days of the COVID-19 pandemic in Bucharest, Romania. *Sci. Data* **9**(1), 253. <https://doi.org/10.1038/s41597-022-01374-7> (2022).
18. Willem, L., Kerckhove, K. V., Chao, D. L., Hens, N. & Beutels, P. A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PLoS ONE* **7**(11), e48695. <https://doi.org/10.1371/journal.pone.0048695> (2012).
19. Hâncean, M.-G., Slavinec, M. & Perc, M. The impact of human mobility networks on the global spread of COVID-19. *J. Complex Netw.* **8**(6), cnaa041. <https://doi.org/10.1093/comnet/cnaa041> (2021).
20. Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Science* **338**(6106), 496–500. <https://doi.org/10.1126/science.1227079> (2012).
21. Remuzzi, A. & Remuzzi, G. COVID-19 and Italy: What next?. *The Lancet* **395**(10231), 1225–1228. [https://doi.org/10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9) (2020).
22. Qian, H. *et al.* Indoor transmission of SARS-CoV-2. *Indoor Air* **31**(3), 639–645. <https://doi.org/10.1111/ina.12766> (2021).
23. Li, Y., Nazaroff, W. W., Bahnfleth, W., Wargocki, P. & Zhang, Y. The COVID-19 pandemic is a global indoor air crisis that should lead to change: A message commemorating 30 years of indoor air. *Indoor Air* **31**(6), 1683–1686. <https://doi.org/10.1111/ina.12928> (2021).
24. Marchi, M. *et al.* ClimateEU, scale-free climate normals, historical time series, and future projections for Europe. *Sci. Data* **7**, 428. <https://doi.org/10.1038/s41597-020-00763-0> (2020).
25. Majumder, P. & Ray, P. P. A systematic review and meta-analysis on correlation of weather with COVID-19. *Sci. Rep.* **11**(1), 10746. <https://doi.org/10.1038/s41598-021-90300-9> (2021).
26. McClymont, H. & Hu, W. Weather variability and COVID-19 transmission: A review of recent research. *Int. J. Environ. Res. Public. Health* **18**(2), 396. <https://doi.org/10.3390/ijerph18020396> (2021).
27. Fadli, A. *et al.* Simple correlation between weather and COVID-19 pandemic using data mining algorithms. *IOP Conf. Ser. Mater. Sci. Eng.* **982**(1), 012015. <https://doi.org/10.1088/1757-899X/982/1/012015> (2020).
28. Feng, Y., Marchal, T., Sperry, T. & Yi, H. Influence of wind and relative humidity on the social distancing effectiveness to prevent COVID-19 airborne transmission: A numerical study. *J. Aerosol Sci.* **147**, 105585. <https://doi.org/10.1016/j.jaerosci.2020.105585> (2020).
29. Rendana, M. Impact of the wind conditions on COVID-19 pandemic: A new insight for direction of the spread of the virus. *Urban Clim.* **34**, 100680. <https://doi.org/10.1016/j.uclim.2020.100680> (2020).
30. Sanchez-Lorenzo, A. *et al.* Did anomalous atmospheric circulation favor the spread of COVID-19 in Europe?. *Environ. Res.* **194**, 110626. <https://doi.org/10.1016/j.envres.2020.110626> (2021).
31. COVID-19 Community Mobility Reports. <https://www.google.com/covid19/mobility/> (accessed 2024-02-21).
32. Liang, X. S. Information flow within stochastic dynamical systems. *Phys. Rev. E* **78**(3), 031113. <https://doi.org/10.1103/PhysRevE.78.031113> (2008).
33. Liang, X. S. Unraveling the cause-effect relation between time series. *Phys. Rev. E* **90**(5), 052150. <https://doi.org/10.1103/PhysRevE.90.052150> (2014).
34. Liang, X. S. Normalizing the causality between time series. *Phys. Rev. E* **92**(2), 022126. <https://doi.org/10.1103/PhysRevE.92.022126> (2015).
35. Kaplan, D. W. *Structural Equation Modeling: Foundations and Extensions* 2nd edn. (SAGE Publications Inc, 2008).
36. Ye, H., Deyle, E. R., Gilarranz, L. J. & Sugihara, G. Distinguishing time-delayed causal interactions using convergent cross mapping. *Sci. Rep.* **5**(1), 14750. <https://doi.org/10.1038/srep14750> (2015).
37. Vautard, R., Yiou, P. & Ghil, M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Phys. Nonlinear Phenom.* **58**(1), 95–126. [https://doi.org/10.1016/0167-2789\(92\)90103-T](https://doi.org/10.1016/0167-2789(92)90103-T) (1992).
38. Ghil, M. *et al.* Advanced spectral methods for climate time series. *Rev Geophys.* **2002**, 1003–1043 (2002).
39. Allen, M. R. & Smith, L. A. Optimal filtering in singular spectrum analysis. *Phys. Lett. A* **234**(6), 419–428. [https://doi.org/10.1016/S0375-9601\(97\)00559-8](https://doi.org/10.1016/S0375-9601(97)00559-8) (1997).
40. Sugihara, G. & May, R. M. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**(6268), 734–741. <https://doi.org/10.1038/344734a0> (1990).
41. Dima, M., Nichita, D. R., Lohmann, G., Ionita, M. & Voiculescu, M. Early-onset of atlantic meridional overturning circulation weakening in response to atmospheric CO₂ concentration. *Npj Clim. Atmos. Sci.* **4**(1), 1–8. <https://doi.org/10.1038/s41612-021-00182-x> (2021).
42. Ebisuzaki, W. A method to estimate the statistical significance of a correlation when the data are serially correlated. *J. Clim.* **10**(9), 2147–2153. [https://doi.org/10.1175/1520-0442\(1997\)010%3c2147:AMTET%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010%3c2147:AMTET%3e2.0.CO;2) (1997).
43. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**(730), 1999–2049. <https://doi.org/10.1002/qj.3803> (2020).
44. Muñoz-Sabater, J. *et al.* ERA5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **13**(9), 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021> (2021).

Acknowledgements

This work was supported by the Executive Unit for Financing Higher Education, Research, Development, and Innovation (UEFISCDI) (PN-III-P4-ID-PCE-2020-2828).

Author contributions

D.R.N. has contributed to the main idea of the article, to the preparation of the COVID-19 data, performing the analyses for Figs. 1, 2, 3 and 4 as well as the analyses in Supplementary Fig. S1–6 and Supplementary Tables 1–2, the interpretation of the results and writing of the manuscript. M.D. has contributed to the main idea of the article, the interpretation of the results, and the writing of the manuscript. L.B. has gathered and prepared the weather data, as well as performed analyses for Fig. 2. M.-G.H. has contributed to gathering the mobility data and writing the manuscript. D.R.N. has contributed to the main idea of the article, to the preparation of the COVID-19 data, performing the analyses for Fig. 1, 2, 3 and 4 as well as the analyses in Supplementary Fig S1–6 and Supplementary Tables 1–2, the interpretation of the results and writing of the manuscript. M.D. has contributed to the main idea of the article, the interpretation of the results, and the writing of the manuscript. L.B. has gathered and prepared the weather data, as well as performed analyses for Fig. 2. M.-G.H. has contributed to gathering the mobility data and writing the manuscript.

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/s41598-024-67918-6>.

Correspondence and requests for materials should be addressed to D.-R.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024