

Article

# Reconstruction of the Temporal Correlation Network of All-Cause Mortality Fluctuation Across Italian Regions: The Importance of Temperature and Among-Nodes Flux

Guido Gigante <sup>1,\*</sup>  and Alessandro Giuliani <sup>2</sup> 

<sup>1</sup> Radiation Protection and Computational Physics, Istituto Superiore di Sanità, 00161 Rome, Italy

<sup>2</sup> Environment and Health Department, Istituto Superiore di Sanità, 00161 Rome, Italy

\* Correspondence: [guido.gigante@iss.it](mailto:guido.gigante@iss.it)

**Abstract:** All-cause mortality is a very coarse grain, albeit very reliable, index to check the health implications of lifestyle determinants, systemic threats and socio-demographic factors. In this work, we adopt a statistical-mechanics approach to the analysis of temporal fluctuations of all-cause mortality, focusing on the correlation structure of this index across different regions of Italy. The correlation network among the 20 Italian regions was reconstructed using temperature oscillations and traveller flux (as a function of distance and region's attractiveness, based on GDP), allowing for a separation between infective and non-infective death causes. The proposed approach allows monitoring of emerging systemic threats in terms of anomalies of correlation network structure.

**Keywords:** complex networks; dynamical systems; epidemiology; time series



**Citation:** Gigante, G.; Giuliani, A. Reconstruction of the Temporal Correlation Network of All-Cause Mortality Fluctuation Across Italian Regions: The Importance of Temperature and Among-Nodes Flux. *Entropy* **2023**, *25*, 21. <https://doi.org/10.3390/e25010021>

Academic Editor: Matteo Convertino

Received: 21 November 2022

Revised: 19 December 2022

Accepted: 19 December 2022

Published: 23 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The monthly-based all-cause death rate fluctuations of the 20 Italian regions are highly correlated in time. This happens even in the absence of recognisable macroscopic parameters, such as massive epidemics. In this work, we tried and built a phenomenological model of the observed between-region correlations based on the traveller flux among the network having, as nodes, the regions and, as edges, the mutual traveller fluxes estimated by a simple exponential model having the distance between regions and Gross Domestic Product (GDP) as major determinants. The above model was complemented by the well-known biphasic effect of temperature on all-cause mortality [1–5]. The problem can be interpreted as the reconstruction of a network wiring in which the between nodes (regions) edge strength corresponds to the observed temporal correlation of the relative death rate fluctuations in time (*Y*-network) by the network wiring generated by the combination of between nodes fluxes and temperature effects (*X*-network).

The strategy of analysis was as follows: the (extremely high) between-region correlation was normalised by what was expected by the observed (well-known) biphasic effect of seasonality. The crude effect of seasonality (when partially out) had, as a consequence, the effect of lowering correlations, but we still have a very high residual correlation for a more refined model.

The biphasic effect (high mortality in winter and summer) on all-cause mortality was hypothesised as derived from an infective component prevailing in winter and a non-infective component prevailing in summer. This interpretation stems from the higher diffusion of viral infections in winter and cardiovascular (often from older people's dehydration) in summer. The winter (infectious) component was modelled using the between regions travellers flux (exponentially decaying with distance) complemented by the 'attractiveness' of each region proportional to its GDP. Thus we generated a 'between-region flux network' using a SIR-like model. The summer (non-infectious) model was formalised using

a linear function of the month-specific average temperature of each region. This allows us to take into consideration the effect of local heat waves.

A model encompassing the above-sketched elements (*X*-network) was fitted to the observed death rates, producing the correlation network (*Y*-network). This minimalistic model was able to reconstruct the death rate oscillation in time and the observed between-region correlation network with high fidelity (corr = 0.993 and corr = 0.841, respectively).

In this work, we demonstrate that weighted edge correlation networks are a very powerful method in epidemiological studies, allowing the tracing of the dynamics of mortality (morbidity) patterns and potentially discovering anomalies relevant to public health. Taking into account the most general definition of a system as ‘...a set of interacting units with relationships among them’ [6], we can safely state that Italy, as for death-rate fluctuations, due to the high temporal correlation among its regions, is a proper system. This allows for the sensible use of second-order statistics (such as correlations), adding unique information content to environmental and epidemiological studies that, in the great majority of cases, rely on the exploitation of a single variable (e.g., death rate fluctuations in a given area) in terms of a set of covariates (e.g., pollution, age structure, etc.).

## 2. Materials and Methods

In order to model the possible order parameters shaping the observed (and partially unexpected) very dense among regions correlation structure of the monthly-based 2011–2019 time series of all-cause mortality, we tried to keep to a minimum both the a priori hypotheses and the number of fitted parameters. This modelling choice was dictated by both the lack of any strong theory on all-cause mortality and to avoid overfitting problems.

Thus we limit ourselves to inserting, as ‘explanatory variables’, the two-phase effect of temperature and an index derived by commuter flux among different regions modulated by the GDP of each region (considered a proxy of region attractiveness).

In the following, we will call  $n_{im}$  the number of deaths recorded in region  $i$  during the  $m$ -th month of recording; accordingly, we will denote  $T_{im}$  as the average temperature in the same region during the same month.

### 2.1. Bi-Phasic Effect of Temperature

To account for temperature effects, we assume that  $n_{im}$  is Poisson distributed with  $\langle n_{im} \rangle = \lambda_i(T_{im})$ :

$$\lambda_i^{\text{biphasic}}(T) = e^{-a_{c,i}T + b_{c,i}} + e^{a_{h,i}T + b_{h,i}}, \quad (1)$$

where  $a_{c,i} > 0$  and  $a_{h,i} > 0$  ( $c$  and  $h$  stand for ‘cold’ and ‘hot’, respectively). This is a convex function with a minimum at:

$$T_{\min,i} = \frac{b_{c,i} - b_{h,i} + \log(a_{c,i}/a_{h,i})}{a_{c,i} + a_{h,i}}. \quad (2)$$

The four parameters ( $a_{c,i}$ ,  $b_{c,i}$ ,  $a_{h,i}$  and  $b_{h,i}$ ) are fitted, for each region  $i$ , by maximising the log-likelihood (through the `scipy.optimize.minimize` function, with TNC method [7,8]):

$$\mathcal{L}\mathcal{L} = \sum_m n_{im} \log(\lambda_i^{\text{biphasic}}(T_{im})) - \lambda_i^{\text{biphasic}}(T_{im}). \quad (3)$$

The temperature  $T_{im}$  is computed by associating the main administrative centre of each region with the three closest weather stations for which we have temperature readings.  $T_{im}$  is then taken, for each month, as the weighted average of the three stations, with the weight proportional to the inverse of the distance between each station and the main administrative centre.

### 2.2. Analysis of Commuter Flux

We denote  $c_{ij}$  as the number of daily commuters from region  $j$  to region  $i$ ; and  $d_{ij}$  (The inter-region distance,  $d_{ij}$ , we used is the distance between the main administrative

centres—“capoluogo”—of each region.) as the distance between the same two regions; note that  $c_{ij}$ , unlike  $d_{ij}$ , is not symmetric. We hypothesise an exponentially decaying relationship between the flux and the distance:

$$c_{ij}^{\text{dist}}(d) = \kappa \left( e^{-\frac{d}{d_0}} + b_0 \right) \tag{4}$$

where the parameters ( $\kappa$ ,  $d_0$  and  $b_0$ ) are fitted by maximising the log-likelihood (through the `scipy.optimize.minimize` function, with the TNC method; the zeros—no commuters from region  $j$  to region  $i$ —are not included in the fit):

$$\mathcal{LL} = - \sum_{i,j|c_{ij}>0} \left( \log(c_{ij}) - \log(c_{ij}^{\text{dist}}(d_{ij})) \right)^2 \tag{5}$$

Here we are assuming a log-normal distribution for  $c_{ij}$  around the expected value  $c_{ij}^{\text{dist}}$ .

Defining  $c_i \equiv \sum_j c_{ij}$ —the number of commuters to region  $i$ , and calling  $\text{GDP}_i$  the GDP of region  $i$ , we hypothesise the existence of a linear relationship:

$$c_i^{\text{gdp}}(\text{GDP}) = \kappa \text{GDP} \tag{6}$$

whose slope  $\kappa$  is fitted by maximising the log-likelihood (through the `scipy.optimize.minimize` function, with the TNC method):

$$\mathcal{LL} = - \sum_i \left( \log(c_i) - \log(c_i^{\text{gdp}}(\text{GDP}_i)) \right)^2 \tag{7}$$

We are assuming that  $c_i$  follows a log-normal distribution around the expected value  $c_i^{\text{gdp}}$ . This is in contrast with the assumption of log-normality for  $c_{ij}$ , since the sum of log-normal variables is not itself log-normal. Yet, in many cases, this is a good approximation [9].

Finally, we performed a fit that considers the two effects together: the exponential decay with distance and the linear dependence on the GDP of the region of destination; indicating with  $\text{pop}_j$  the population of region  $j$ , we have:

$$c_{ij}^{\text{fit}} = \kappa \text{pop}_j \text{GDP}_i e^{-\frac{d_{ij}}{d_0}} \tag{8}$$

As above, the fit procedure finds the best parameters  $\kappa$  and  $d_0$  by maximising the log-normal log-likelihood (through the `scipy.optimize.minimize` function, with the TNC method; the zeros—no commuters from region  $j$  to region  $i$ —are not included in the fit).

### 2.3. Total Flux

We hypothesise that the total flux of persons  $f_{ij}$  comprises, beyond the daily commuters  $c_{ij}$ , an ‘episodic’ component  $e_{ij}$  of more irregular movements:

$$f_{ij} = c_{ij} + e_{ij} \tag{9}$$

Starting from the results for  $c_{ij}$ , we make the assumption that  $e_{ij}$  is an exponentially decaying function of the distance between regions and a linear function of the GDP of the region of arrival  $i$  and of the population of the region of departure:

$$e_{ij} = \kappa^e \text{pop}_j \text{GDP}_i e^{-d_{ij}/d_0^e} \tag{10}$$

With respect to Equation (8), we expect  $d_0^e > d_0$ , since episodic travels, in contrast with frequent ones, are likely less affected by the distance to travel.

### 2.4. Sir Network Model

The full flux-temperature model includes two different effects. The first one is related to the non-infective component of mortality:

$$\lambda_{im}^{\text{flux}} = \text{pop}_i \left( e^{a_h T_{im} + b_h} + \rho_{0i} \right) + \dots, \tag{11}$$

where  $\lambda_{im}^{\text{flux}}$  is the model expectation for the number of deaths in region  $i$  at month  $m$ ; and  $\rho_{0i}$  is a baseline mortality rate for region  $i$ . Note that this effect is akin to the warm-season component of Equation (1), but here in the flux-temperature model, for the sake of parsimony, we lose the individualised behaviour of each region  $i$ , and all regions respond to high temperatures in the same way.

The second effect takes into account the infective component of mortality. We make the simplifying assumption that, in each month, a new infectious disease starts spreading; at the end of the month, a fraction  $\mu$  of the people ‘recovered’ from the disease dies; the following month, the process starts afresh. The spreading of the disease follows a SIR (Susceptible, Infected, Recovered) model [10] on the flux network. Defining the two matrices:

$$\phi_{ij} = \begin{cases} \frac{f_{ij}}{\text{pop}_j} & \text{for } i \neq j \\ 1 - \frac{\sum_k f_{kj}}{\text{pop}_j} & \text{for } i = j \end{cases} \tag{12}$$

$$\hat{\phi}_{ij} = \left( \frac{\phi_{ij}}{\sum_k \phi_{ik}} \right)^T \tag{13}$$

the dynamics of the model reads:

$$\dot{S}_i = - \sum_l \hat{\phi}_{il} \left[ \frac{\beta}{\text{pop}_l} \left( \sum_j \phi_{lj} S_j \right) \left( \sum_j \phi_{lj} I_j \right) \right] \tag{14}$$

$$\dot{I}_i = \sum_l \hat{\phi}_{il} \left[ \frac{\beta}{\text{pop}_l} \left( \sum_j \phi_{lj} S_j \right) \left( \sum_j \phi_{lj} I_j \right) \right] - \gamma I_i \tag{15}$$

$$\dot{R}_i = \gamma I_i, \tag{16}$$

where  $S_i$ ,  $I_i$  and  $R_i$  are the number, respectively, of susceptible, infected and recovered individuals in region  $i$ ;  $\beta$  measures the rate at which susceptible individuals get infected ( $S \rightarrow I$ ); and  $\gamma$  is the rate of recovery,  $I \rightarrow R$ . The model, therefore, consists of 60 coupled differential equations.

The reasoning behind the model is as follows. The term  $\sum_j \phi_{ij} S_j$  represents the number of susceptible individuals in region  $i$ , at a given instant in time, due to the flux from other regions (minus the flux out of region  $i$  itself—the diagonal elements  $\phi_{ii}$ ); for the infected, it is  $\sum_j \phi_{ij} I_j$ . In a classical SIR model, the number of newly infected individuals  $dI$  is given by  $\beta \frac{SI}{\text{pop}}$ ; in our case:

$$dI_l = \frac{\beta}{\text{pop}_l} \left( \sum_j \phi_{lj} S_j \right) \left( \sum_j \phi_{lj} I_j \right). \tag{17}$$

At the end of the day, the reverse flux  $f_{ji}$  (people moving back from region  $i$  to region  $j$ ) will redistribute the newly infected in proportion to the fraction of susceptible individuals contributed by each region  $j$ ; this is given by:

$$dI_i = \sum_l \hat{\phi}_{il} dI_l, \tag{18}$$

that, together with Equation (17), gives the infinitesimal increment of infected people entering Equation (15) (first term on the left).

The initial conditions are always in the form of one ‘patient zero’ in region  $i_0$  at time  $t = 0$ , so that  $S^i(t = 0) = \text{pop}_i$  for  $i \neq i_0$ , and  $S^{i_0}(t = 0) = \text{pop}_{i_0} - 1$ ; accordingly,

$I^i(t = 0) = 0$  for  $i \neq i_0$ , and  $I^{i_0}(t = 0) = 1$ . Since  $i_0$  is not known, we assume it to be a random variable distributed such that:

$$p(i_0 = i) \propto \text{GDP}_i; \tag{19}$$

this amounts to assuming that the external flux to region  $i$  (people coming to region  $i$  from outside Italy) is proportional to the GDP of the region itself.

For each month, we evolve Equations (14)–(16) for 30 days ( $t \in [0, 30]$ ); the equations are integrated using the Euler method, with step size  $dt = 1$  day. Finally, this infective component of mortality is incorporated into the model:

$$\lambda_{imi_0}^{\text{flux}} = \text{pop}_i \left( e^{a_h T_{im} + b_h} + \rho_{0i} \right) + \mu R_{im}^{i_0}(t = 30), \tag{20}$$

where with  $R_{im}^{i_0}(t = 30)$  we designate the total number of recovered individuals for region  $i$  at the end of month  $m$ , when the patient zero was located in region  $i_0$  (all months are assumed, for simplicity, to have 30 days).

To also incorporate seasonal effects in the infective dynamics, we make  $\beta$  a function of the temperature:

$$\beta_{im} = e^{-a_\beta T_{im} + b_\beta}, \tag{21}$$

with  $a_\beta > 0$ ; with the additional constraint that  $\beta < \frac{1}{dt}$  (the condition  $\beta = \frac{1}{dt}$  amounts to having all the population infected in a single  $dt$ ; larger values lead, in the Euler approximation, to unphysical solutions).

Considering Equations (9) and (11) (parameters  $a_h$ ,  $b_h$  and  $\rho_{0i}$ ), Equation (10) ( $\kappa^e$  and  $d_0^e$ ), Equation (20) ( $\mu$ ), Equation (21) ( $a_\beta$  and  $b_\beta$ ), alongside Equations (14)–(16) ( $\gamma$ ), the model comprises 28 parameters; of which 20 (the  $\rho_{0i}$ ) are simply used to offset the different mortality rates in different regions (due, for example, to distinct age structures). These parameters are fitted to the data by maximising the Poisson log-likelihood:

$$\mathcal{LL} = \frac{1}{N_{\text{counts}}} \sum_{i_0} p(i_0) \left[ \sum_{i,m} n_{im} \log(\lambda_{imi_0}^{\text{flux}}) - \lambda_{imi_0}^{\text{flux}} \right], \tag{22}$$

where  $N_{\text{counts}}$  is the number of terms in the sum  $\sum_{i,m}$  (if we consider  $n_{\text{batch}}$  different months, having 20 regions,  $N_{\text{counts}} = 20 \cdot n_{\text{batch}}$ ), and  $p(i_0)$  is given by Equation (19).

To this likelihood, we added two prior likelihoods to constrain the parameters of the model. The first is a soft flat prior, not-null in the range  $[0, 0.2]$  for the quantity  $\frac{\sum_k f_{ki}}{\text{pop}_i}$  (see Equation (9)); the log-prior becomes quadratic outside the allowed (flat probability) range; the factor in front of the quadratic term is chosen large enough to practically prevent leaving the allowed range. This constrains the fraction of the population leaving a region every day to less than 20%. The second log-prior is quadratic in  $d_0^e$  (Equation (10)):

$$\log(p(d_0^e)) = -9.74 \cdot 10^{-7} (d_0^e)^2 + \text{const}, \tag{23}$$

to penalise very high spatial decay constants  $d_0^e$  for the episodic component of the flux.

The maximisation has been carried out, in this case, through the Adam optimiser [11], with default parameters ( $\beta_1 = 0.9$  and  $\beta_2 = 0.99$ ) and a learning rate decreasing at each optimisation step according to:

$$lr(\text{step}) = \frac{10^{-3}}{\left(1 + \frac{\text{step}}{10^4}\right)^{0.75}}. \tag{24}$$

The training set consists of the monthly death counts for each of the 20 regions for 96 (out of 108) randomly chosen months in the period 2011–2019. We reserved 12 months ( $12 + 96 = 108$ ) as test data; these months were selected to have one exemplar of each

calendar month—January to December; since the dataset spans only 9 years, 3 randomly-selected years contributed two months (6 months apart, e.g., April–October) to the test data.

At each step,  $n_{\text{batch}}$  months (with month, we here denote one specific month in a specific year; so, in the training set, we have 96 months) are randomly selected from the training set (the same month can appear multiple times in the batch). For each month  $m$ , a patient zero-region  $i_{0m}$  is randomly extracted with a probability given by Equation (19). The computed log-likelihood is then:

$$\mathcal{L}\mathcal{L}_{\text{batch}} = \frac{1}{20 \cdot n_{\text{batch}}} \sum_{i, m, i_{0m}} n_{im} \log(\lambda_{imi_{0m}}^{\text{flux}}) - \lambda_{imi_{0m}}^{\text{flux}}, \quad (25)$$

a stochastic approximation of the total log-likelihood of Equation (22).

For the first  $10^4$  optimisation steps,  $n_{\text{batch}} = 10$ . From that step onwards,  $n_{\text{batch}} = 100$ ; and, to the log-likelihood, we added a ‘regularisation’ term:

$$\mathcal{L}\mathcal{L}_{\text{corr}} = -\frac{v_{\text{corr}}}{190} \sum_{i>j} (\text{corr}_{ij} - \text{corr}_{ij}^0)^2 \quad (26)$$

where  $\text{corr}_{ij}^0$  is the actual correlation between the monthly deaths of region  $i$  and region  $j$ ; whereas  $\text{corr}_{ij}$  is the corresponding correlations produced by the model (on the specific batch); factor  $\frac{1}{190}$  normalises the sum  $\sum_{i>j}$ , which comprises 190 terms. We set  $v_{\text{corr}} = 8.76 \times 10^2$ .

We monitored the log-likelihood (Equation (22)) on the test data during the training; since it never substantially decreased (that would suggest some level of over-fitting), we interrupted the optimisation after  $10^6$  steps, when improvement on the training set appeared extremely slow.

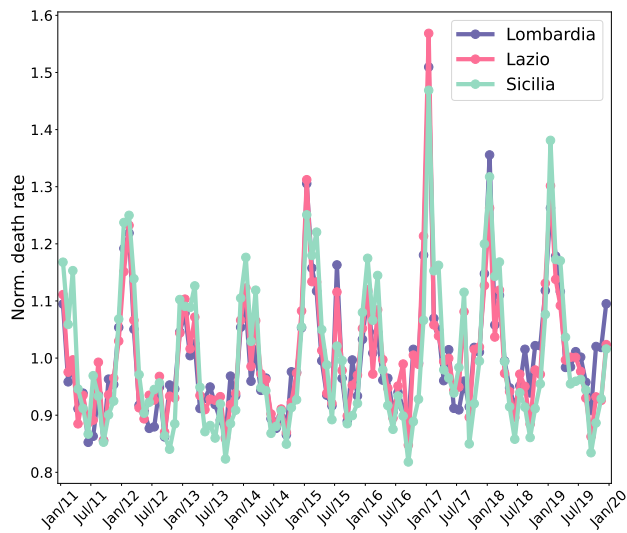
All the computations were performed with custom code written in Python; core functions were just-in-time compiled, and their gradient was computed, where necessary, through the Jax package (<https://github.com/google/jax>, accessed on 9 December 2022). The complete code, reproducing all the reported results, can be found at <https://github.com/GuidoGigante/All-cause-mortality-fluctuation-across-Italian-regions>, accessed on 9 December 2022.

### 3. Results

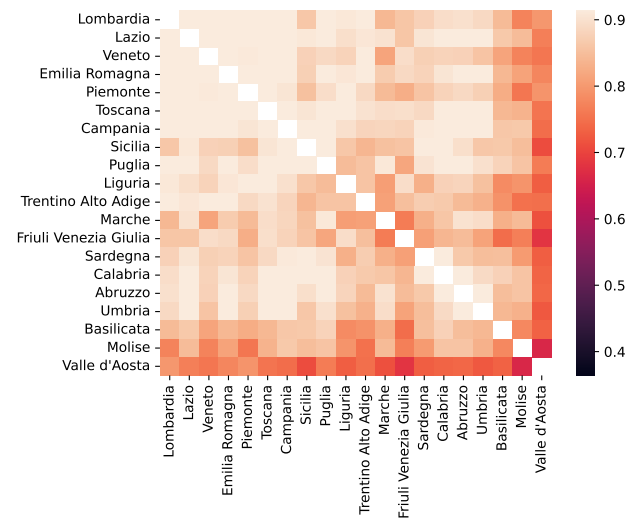
The course of monthly death rates, normalised to the mean over the entire period, is strikingly similar for different regions. This can be appreciated in Figure 1a, where we show three regions (chosen to be representative of the north—Lombardia, centre—Lazio, and south—Sicilia, of Italy).

Such observations are made more quantitative in Figure 1b, which shows the between-region correlation matrix of monthly death rate fluctuations (correlations computed on 108 data points) relative to the different regions. As evident from the figure, the between-region correlations are extremely high ( $0.865 \pm 0.063$ ), with smaller and less densely populated regions (i.e., Valle d’Aosta and Molise) endowed (as expected) by a lower (albeit very significant) average correlation strength (0.739 and 0.805, respectively).

To check if the bi-phasic effect of temperature was sufficient to get rid of the observed correlations (in the presence of a substantially similar age structure across the different Italian regions), a quantitative model taking into account the temperature effect was fitted to the different regions’ mortality data (see Section 2).



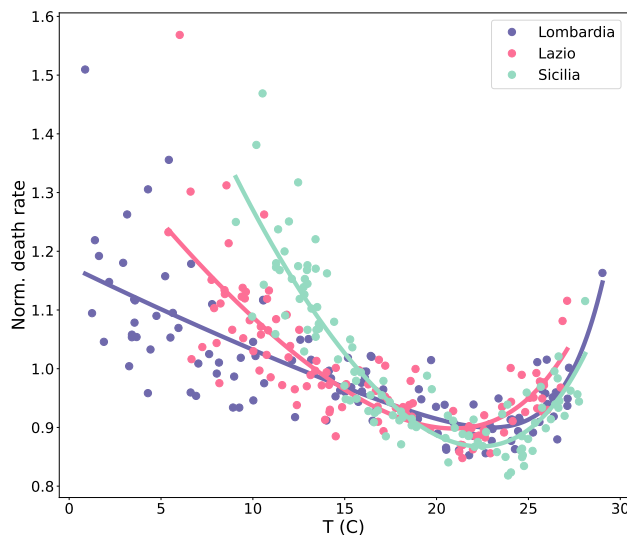
(a)



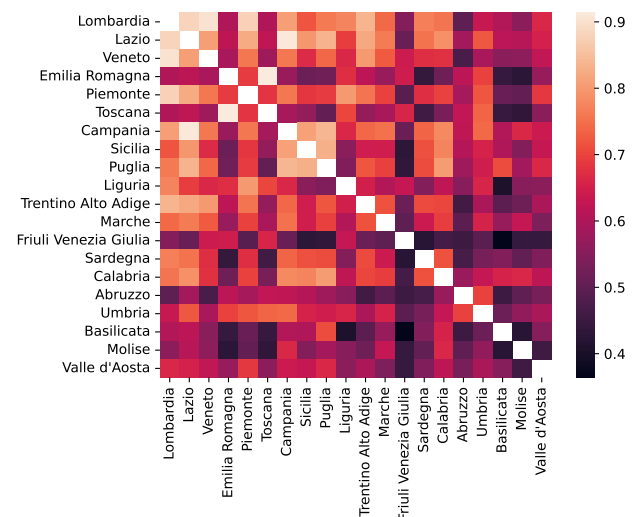
(b)

**Figure 1.** Death rates in different regions are extremely correlated. (a) Actual time series for three regions, normalised to have an average value equal to one. The three lines present a strikingly similar course. (b) Pairwise between-region correlations (regions are ordered—left to right and top to bottom—according to decreasing GDP). A trend with GDP is appreciable, with smaller and less densely populated regions (i.e., Valle d’Aosta and Molise) endowed with lower (albeit still high) correlations.

All regions showed very similar relations between death rate fluctuations and temperature with the expected bi-phasic relation with two winter and summer peaks and a minimum at intermediate temperature values (spring and autumn) (see Figure 2a reporting three representative regions’ data; continuous lines: see Equation (1)).



(a)

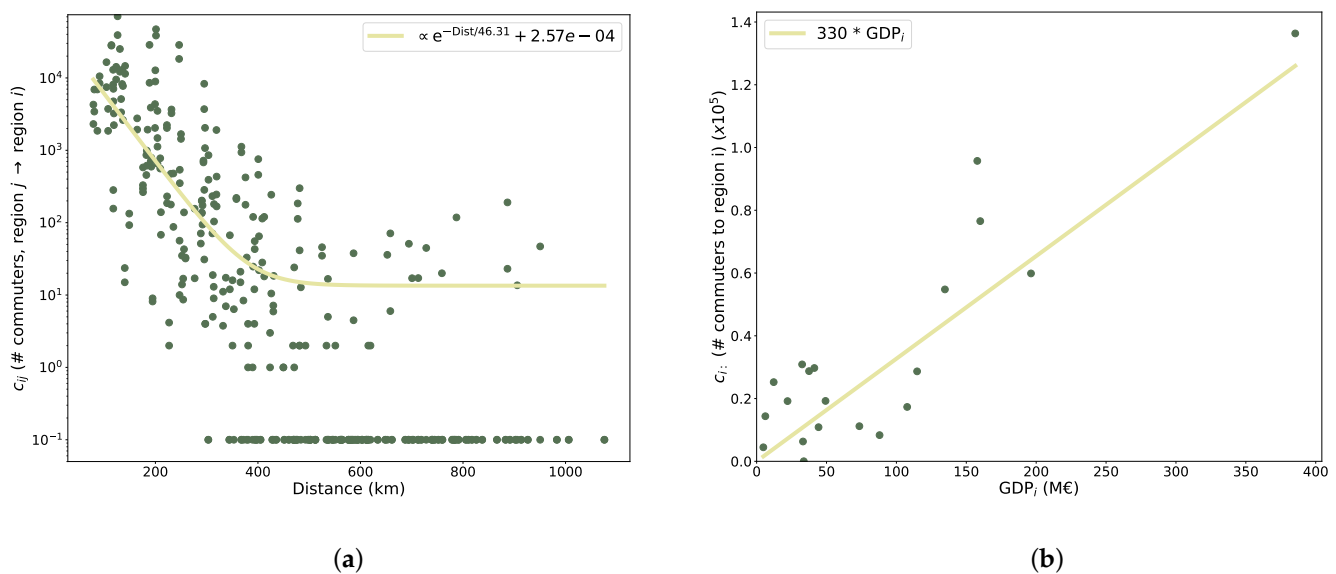


(b)

**Figure 2.** (a) Normalised death rates for three regions as a function of the temperature. The continuous lines are the results of a fit (see Equations (1) and (3)). (b) Pairwise between-region correlations for the time series of the deaths when the fitted effect of the temperature is subtracted from the raw numbers. Correlations drastically decrease (colour scale as in Figure 1b) but remain large.

By normalising the time series of death rate fluctuations by temperature effect, the between-region correlation drastically decreases ( $0.63 \pm 0.12$ ), therefore, confirming the expected effect of temperature on mortality (Figure 2b; colour scale as in Figure 1b). Notwithstanding that the residual entity correlation is still high, asking for some other relevant factor to be taken into consideration.

We hypothesise that strong correlations among regions also arise for an infective component, continually spreading from region to region at small time scales (less than a month), driven by the movement of people from one region to another. First, we examined the data about the flux of daily commuters between regions; such data show a clear dependence both on spatial distance (exponential decay; see Figure 3a; the continuous line is the result of a fit, see Equations (4) and (5)) and the GDP of the region of arrival, having the role of an ‘attractiveness’ factor (linear dependence; see Figure 3b, where the continuous line is the result of a fit; see Equations (6) and (7)).



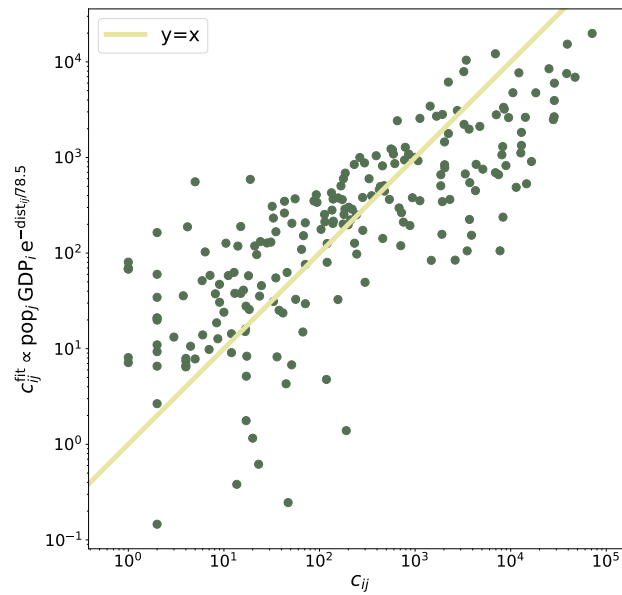
**Figure 3.** Determinants of the commuters flux. (a) The flux between two regions decays exponentially with the distance to travel (continuous line, exponential fit; see Equation (4)). The points at the bottom of the graph are zeros (not allowed in logarithmic scale and not considered in the fit). (b) The flux increases linearly with the GDP of the region of destination (continuous line, linear fit; see Equations (6) and (7)).

The two determinants (distance and GDP) are considered together in Figure 4, where the actual number of commuters (from one region to another; the flux is not symmetric) is compared to the result of the fitted model (see Equation (8)); the good agreement of the reconstructed flux with the real one (the continuous line is the identity line) supports the assumptions of the model.

Starting from these results, we make the hypothesis that the total flux of people between regions is made of the commuters flux plus an ‘episodic’ flux, unknown but with the same functional form (exponential decay with distance; linear dependence on GDP). Then, we built, on the total-flux matrix, a SIR-like model [10] that takes into account the exchange of infected people between regions. Temperature impacts the model in two ways. The first increases with the temperature and is akin to the high-temperature (rightmost) arm of the model of Figure 2a. The other modulates the contagiousness of the disease (higher for lower temperatures).

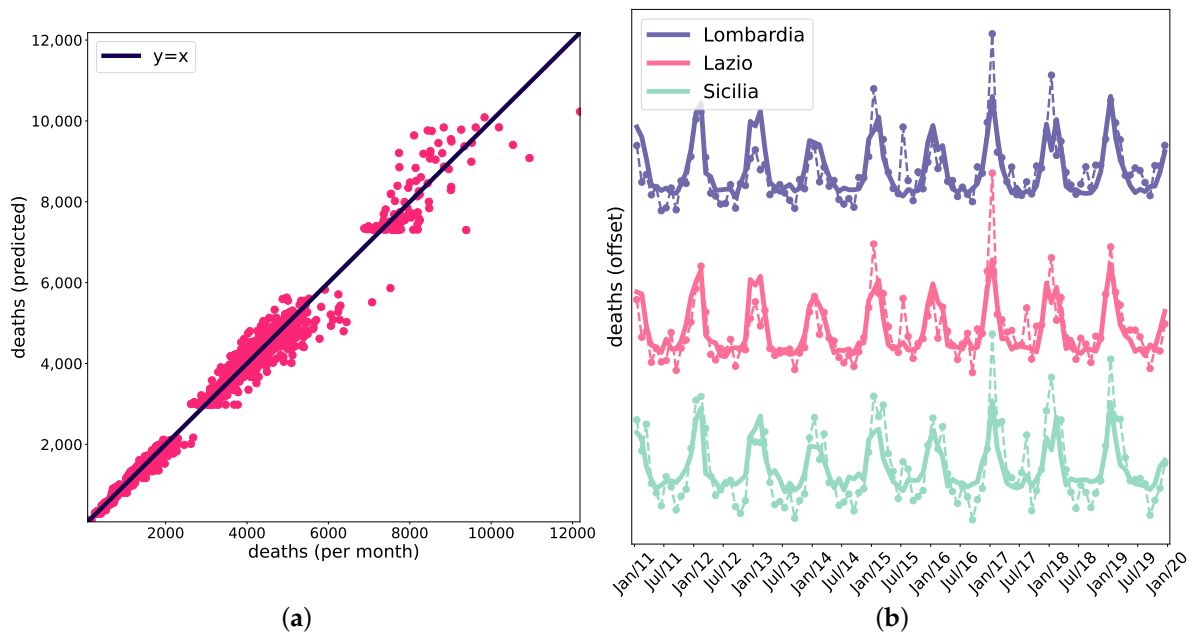
In the model, each month a new disease starts spreading from a given region (chosen according to a probability distribution); at the end of the month, a fraction of the ‘recovered’ people dies. Added to this effect are the high-temperature mortality and, finally, a generic, temperature-independent, region-specific mortality.





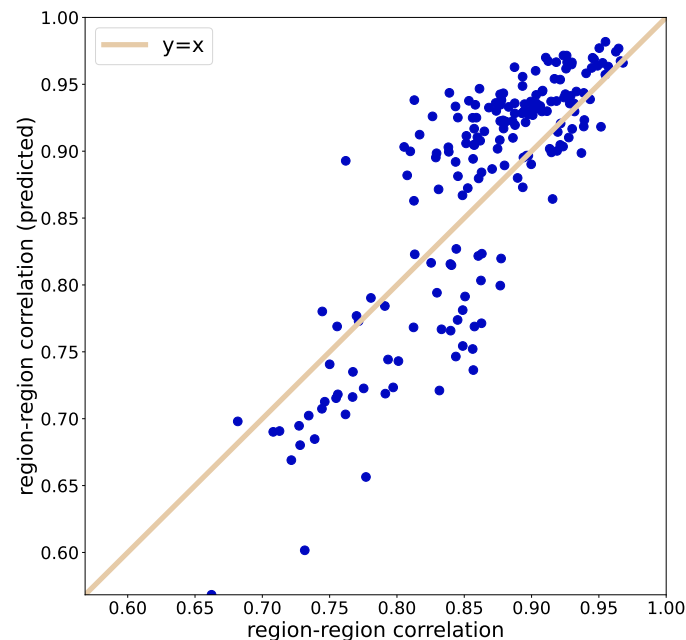
**Figure 4.** Actual commuters flux *vs.* the flux reconstructed by a fitted model that decays exponentially with the distance and grows linearly with the GDP of the region of destination (see Equation (8)). The continuous line is the identity line. Only non-zero entries of the commuters’ matrix are displayed and considered in the fitting procedure.

We fitted the model’s parameters (see Material and Methods) to the data. Figure 5a shows the death counts for all the regions and all the considered months against the death counts generated by the model. The model can reproduce a large part of the observed variability (corr = 0.993; the continuous line is the identity line). This can be appreciated, as the deaths evolve over time, in Figure 5b, for three different regions (dashed lines: data; continuous line: model). Note that the three time series are offset vertically to make the comparison data *vs.* model clearer.



**Figure 5.** (a) Actual deaths *vs.* the deaths expected by the model (corr = 0.993; the continuous line is the identity). (b) Time-series of the deaths for three regions; dashed lines: data; continuous lines: model.

Finally, we compare, in Figure 6, the observed between-region correlations and the correlations between the time series produced by the model for each region. To a large extent, the model can capture the variability of the correlation among the regions ( $\text{corr} = 0.841$ ; the continuous line is the identity line).



**Figure 6.** Between-region correlation: data vs. reconstructed from the model ( $\text{corr} = 0.841$ ; the continuous line is the identity line).

#### 4. Discussion

We aptly reconstructed the strong correlation among temporal series of all-cause monthly death rates relative to the 20 Italian regions by a model encompassing non-infectious (mainly summer) and infectious (winter) components. This last component was modelled in terms of a set of SIR equations taking into account both the across-regions commuter exchange (daily flow) and the more irregular traveller flux (longer time flow). The different mathematical treatments of summer and winter components allowed for a neat increase in the reconstruction of both global death rates and among-regions correlation strength concerning crude seasonality.

The reconstruction of the observed correlation network by our model (Figure 6), excluding the hypothesis of contemporary arising ‘epidemic sources’ across all the regions (that have near-zero probability), confirms the reliability of the proposed model. Overall, we can consider Italy as a proper ‘integrated system’ that, thanks to both a rich exchange flux among regions and the sharing of ‘heat waves’, reaches a general coherence in death rate fluctuations. This coherent behaviour acts as a largely invariant ‘mean field’ governing all-cause monthly (and thus unaffected by longer time fluctuations in age class distribution) death rate fluctuations.

The existence of a very stable correlation network among Italian regions can be profitably used as a tool for the epidemiological surveillance of the territory: the arising of an anomalous value of the correlation degree of a region can be intended as the presence of an emerging source of risk (of both infectious and/or environmental origin). Thanks to the intrinsic redundancy of the correlation matrix, any (even transient) change reverberate on the entire network, allowing for a more sensible detection of instabilities: one clear example is the case of Recurrence Quantification Analysis (RQA, [12]). RQA relies upon the construction of a distance (correlation in the case of angular metrics) matrix between subsequent epochs of a time (or space [13])-dependent signal. When in the presence of regime changes driven by a slowly varying control parameter, RQA metrics (at odds with

usual statistical indexes) exactly determine the entity and time (spatial) location of regime change [14,15]. Analogous considerations hold for correlation matrices and any other network system [16] induced by external (e.g., epidemics) or internal (e.g., deterioration of living conditions) driving forces [17].

The application of statistical-mechanics-inspired tools in public health is still in its infancy [18,19] and/or confined to very specific issues [20,21]. In this work, relying upon a very general consideration by Alexander Gorban and colleagues, ‘It is useful to analyse correlation graphs’ [22], we demonstrate how a raw (albeit very reliable) indicator, all-cause mortality, is amenable to a statistical mechanics approach opening new avenues to epidemiological and environmental research.

## 5. Conclusions

All-cause mortality is considered a general indicator of the general health status of a population in the context of a particular age structure. This is why many epidemiological studies investigate both temporal and spatial fluctuation of all-cause mortality, looking for a correlation with socioeconomic [23], environmental [24–26] and physiological/pathological conditions [27].

Moreover, the reliability of all-cause mortality statistics, due to its coarse-grain character, makes the analysis of its fluctuations a very important viewpoint to estimate the impact of epidemic threats [28].

Our approach stems from the above considerations on all-cause mortality to explore a still neglected dimension of this index: the relative weight of purely stochastic and deterministic (coherent) components of spatio-temporal fluctuations of all-cause mortality. The presence of external drivers increasing the correlation of such fluctuations was already observed [24] in terms of departure from ‘optimal temperature’, here we highlight another correlation source linked to the population fluxes among different areas modelled in terms of both daily commuters and ‘relative attractiveness’ of different regions in terms of GDP.

This choice allowed us to get rid of a very dense correlation network among the 20 Italian regions for monthly-based fluctuation rates in the 2011–2019 period. The presence of very high temporal correlations among different regions was partly unexpected and constituted a very important result for monitoring the onset of emerging public health threats in terms of alterations of such correlation structures.

**Author Contributions:** Conceptualisation, methodology and writing, A.G. and G.G.; software, formal analysis and visualisation, G.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The daily deaths for each Italian district (‘comune’), starting from 2011, can be downloaded from the site of the Italian Institute of Statistics (ISTAT), following this link: [https://www.istat.it/storage/dati\\_mortalita/Dataset-decessi-comunali-giornalieri\\_regioni%28excel%29\\_5-21-ottobre-2021.zip](https://www.istat.it/storage/dati_mortalita/Dataset-decessi-comunali-giornalieri_regioni%28excel%29_5-21-ottobre-2021.zip) (accessed on 24 June 2022). The geographical coordinates for each *comune* can be found in the following GitHub project: <https://github.com/MatteoHenryChinaski/Comuni-Italiani-2018-Sql-Json-excel>, notably the file *italy\_geo.xlsx* (accessed on 24 June 2022). The data concerning the temperature have been obtained, on 23 June 2022, from the National Centers for Environmental Information (<https://www.noaa.gov/>), with the following order specifications: Begin date: 2012-01-01 00:00; End date: 2021-12-31 23:59; Data types: PRCP, SNWD, TAVG, TMAX, TMIN; Custom Flags: Station Name, Geographic Location, Include Data Flags. The number of daily commuters between regions has been obtained from the data made available by ISTAT at the link: [https://www.istat.it/storage/cartografia/matrici\\_pendolarismo/matrici\\_pendolarismo\\_2011.zip](https://www.istat.it/storage/cartografia/matrici_pendolarismo/matrici_pendolarismo_2011.zip) (data description can be found at <https://www.istat.it/it/archivio/157423>; accessed on 28 June 2022). The number of commuters we used is the sum of commuters from any *comune* belonging to one region towards any *comune* of another region. The GDP for each region has been retrieved from

the Wikipedia page [https://it.wikipedia.org/wiki/Regioni\\_d%27Italia](https://it.wikipedia.org/wiki/Regioni_d%27Italia) (GDP is ‘Prodotto interno lordo’ or PIL, in Italian; accessed on 10 October 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- English, L.K.; Ard, J.D.; Bailey, R.L.; Bates, M.; Bazzano, L.A.; Boushey, C.J.; Brown, C.; Butera, G.; Callahan, E.H.; De Jesus, J.; et al. Evaluation of dietary patterns and all-cause mortality: A systematic review. *JAMA Netw. Open* **2021**, *4*, e2122277. [[CrossRef](#)] [[PubMed](#)]
- Bilinski, A.; Emanuel, E.J. COVID-19 and excess all-cause mortality in the US and 18 comparison countries. *Jama* **2020**, *324*, 2100–2102. [[CrossRef](#)] [[PubMed](#)]
- Foster, H.M.; Celis-Morales, C.A.; Nicholl, B.I.; Petermann-Rocha, F.; Pell, J.P.; Gill, J.M.; O'Donnell, C.A.; Mair, F.S. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: A prospective analysis of the UK Biobank cohort. *Lancet Public Health* **2018**, *3*, e576–e585. [[CrossRef](#)] [[PubMed](#)]
- Martinez, G.S.; Diaz, J.; Hooyberghs, H.; Lauwaet, D.; De Ridder, K.; Linares, C.; Carmona, R.; Ortiz, C.; Kendrovski, V.; Adamonyte, D. Cold-related mortality vs heat-related mortality in a changing climate: A case study in Vilnius (Lithuania). *Environ. Res.* **2018**, *166*, 384–393. [[CrossRef](#)] [[PubMed](#)]
- Nielsen, J.; Mazick, A.; Glismann, S.; Mølbak, K. Excess mortality related to seasonal influenza and extreme temperatures in Denmark, 1994–2010. *BMC Infect. Dis.* **2011**, *11*, 1–13. [[CrossRef](#)] [[PubMed](#)]
- Miller, J.G. The nature of living systems. *Behav. Sci.* **1976**, *21*, 295–319. [[CrossRef](#)]
- Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
- Nash, S.G. Newton-type minimization via the Lanczos method. *SIAM J. Numer. Anal.* **1984**, *21*, 770–788. [[CrossRef](#)]
- Mitchell, R.L. Permanence of the log-normal distribution. *JOSA* **1968**, *58*, 1267–1272. [[CrossRef](#)]
- Hethcote, H.W. The mathematics of infectious diseases. *SIAM Rev.* **2000**, *42*, 599–653. [[CrossRef](#)]
- Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
- Marwan, N.; Romano, M.C.; Thiel, M.; Kurths, J. Recurrence plots for the analysis of complex systems. *Phys. Rep.* **2007**, *438*, 237–329. [[CrossRef](#)]
- Colafranceschi, M.; Colosimo, A.; Zbilut, J.P.; Uversky, V.N.; Giuliani, A. Structure-related statistical singularities along protein sequences: A correlation study. *J. Chem. Inf. Model.* **2005**, *45*, 183–189. [[CrossRef](#)] [[PubMed](#)]
- Trulla, L.; Giuliani, A.; Zbilut, J.; Webber, C., Jr. Recurrence quantification analysis of the logistic equation with transients. *Phys. Lett. A* **1996**, *223*, 255–260. [[CrossRef](#)]
- Casdagli, M. Recurrence plots revisited. *Phys. D: Nonlinear Phenom.* **1997**, *108*, 12–44. [[CrossRef](#)]
- Liu, X.; Li, D.; Ma, M.; Szymanski, B.K.; Stanley, H.E.; Gao, J. Network resilience. *Phys. Rep.* **2022**, *971*, 1–108. [[CrossRef](#)]
- Gorban, A.N.; Smirnova, E.V.; Tyukina, T.A. Correlations, risk and crisis: From physiology to finance. *Phys. A Stat. Mech. Its Appl.* **2010**, *389*, 3193–3217. [[CrossRef](#)]
- Park, J.; Choi, J.; Choi, J.Y. Network analysis in systems epidemiology. *J. Prev. Med. Public Health* **2021**, *54*, 259. [[CrossRef](#)]
- Huo, J.; Zhao, H. Dynamical analysis of a fractional SIR model with birth and death on heterogeneous complex networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *448*, 41–56. [[CrossRef](#)]
- Sangeet, S.; Sarkar, R.; Mohanty, S.K.; Roy, S. Quantifying Mutational Response to Track the Evolution of SARS-CoV-2 Spike Variants: Introducing a Statistical-Mechanics-Guided Machine Learning Method. *J. Phys. Chem. B* **2022**, *126*, 7895–7905. [[CrossRef](#)]
- Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of influential spreaders in complex networks. *Nat. Phys.* **2010**, *6*, 888–893. [[CrossRef](#)]
- Gorban, A.; Tyukina, T.; Pokidysheva, L.; Smirnova, E. It is useful to analyze correlation graphs: Reply to comments on “Dynamic and thermodynamic models of adaptation”. *Phys. Life Rev.* **2022**, *40*, 15–23. [[CrossRef](#)] [[PubMed](#)]
- Dadgar, I.; Norström, T. Is there a link between all-cause mortality and economic fluctuations? *Scand. J. Public Health* **2022**, *50*, 6–15. [[CrossRef](#)] [[PubMed](#)]
- Burkart, K.G.; Brauer, M.; Aravkin, A.Y.; Godwin, W.W.; Hay, S.I.; He, J.; Iannucci, V.C.; Larson, S.L.; Lim, S.S.; Liu, J.; et al. Estimating the cause-specific relative risks of non-optimal temperature on daily mortality: A two-part modelling approach applied to the Global Burden of Disease Study. *Lancet* **2021**, *398*, 685–697. [[CrossRef](#)] [[PubMed](#)]
- Yu, W.; Mengersen, K.; Hu, W.; Guo, Y.; Pan, X.; Tong, S. Assessing the relationship between global warming and mortality: Lag effects of temperature fluctuations by age and mortality categories. *Environ. Pollut.* **2011**, *159*, 1789–1793. [[CrossRef](#)]
- Li, T.; Zhang, Y.; Wang, J.; Xu, D.; Yin, Z.; Chen, H.; Lv, Y.; Luo, J.; Zeng, Y.; Liu, Y.; et al. All-cause mortality risk associated with long-term exposure to ambient PM<sub>2.5</sub> in China: A cohort study. *Lancet Public Health* **2018**, *3*, e470–e477. [[CrossRef](#)]

27. Blair, S.N.; Kohl, H.W.; Barlow, C.E.; Paffenbarger, R.S.; Gibbons, L.W.; Macera, C.A. Changes in physical fitness and all-cause mortality. *Jama* **1995**, *273*, 1093–1098. [[CrossRef](#)]
28. Jdanov, D.A.; Galarza, A.A.; Shkolnikov, V.M.; Jasilionis, D.; Németh, L.; Leon, D.A.; Boe, C.; Barbieri, M. The short-term mortality fluctuation data series, monitoring mortality shocks across time and space. *Sci. Data* **2021**, *8*, 1–8. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.