sustainability

MDPI

Article

Dynamic Complex Network Analysis of $PM_{2.5}$ Concentrations in the UK, Using Hierarchical Directed Graphs (V1.0.0)

Parya Broomandi [1,2,3], Xueyu Geng [1], Weisi Guo [1,4,5], Alessio Pagani [5], David Topping [5,6] and Jong Ryeol Kim [2,*]

1 School of Engineering, The University of Warwick, Coventry CV4 7AL, UK; parya.broomandi@warwick.ac.uk (P.B.); xueyu.geng@warwick.ac.uk (X.G.); weisi.guo@cranfield.ac.uk (W.G.)
2 Department of Civil and Environmental Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan 010000, Kazakhstan
3 Department of Chemical Engineering, Masjed-Soleiman Branch, Islamic Azad University, Masjed-Soleiman, Iran
4 School of Aerospace, Transport, and Manufacturing, Cranfield University, Bedford MK43 0AL, UK
5 The Alan Turing Institute, London NW1 2DB, UK; apagani@turing.ac.uk (A.P.); david.topping@manchester.ac.uk (D.T.)
6 School of Earth, Atmospheric and Environmental Science, University of Manchester, Manchester M13 9PL, UK
* Correspondence: jong.kim@nu.edu.kz; Tel.: +7-7172-70-9136

Abstract: The risk of a broad range of respiratory and heart diseases can be increased by widespread exposure to fine atmospheric particles on account of their capability to have a deep penetration into the blood streams and lung. Globally, studies conducted epidemiologically in Europe and elsewhere provided the evidence base indicating the major role of $PM_{2.5}$ leading to more than four million deaths annually. Conventional approaches to simulate atmospheric transportation of particles having high dimensionality from both transport and chemical reaction process make exhaustive causal inference difficult. Alternative model reduction methods were adopted, specifically a data-driven directed graph representation, to deduce causal directionality and spatial embeddedness. An undirected correlation and a directed Granger causality network were established through utilizing $PM_{2.5}$ concentrations in 14 United Kingdom cities for one year. To demonstrate both reduced-order cases, the United Kingdom was split up into two southern and northern connected city communities, with notable spatial embedding in summer and spring. It continued to reach stability to disturbances through the network trophic coherence parameter and by which winter was construed as the most considerable vulnerability. Thanks to our novel graph reduced modeling, we could represent high-dimensional knowledge in a causal inference and stability framework.

Keywords: atmospheric pollution; causality; stability; complex network; $PM_{2.5}$

Citation: Broomandi, P.; Geng, X.; Guo, W.; Pagani, A.; Topping, D.; Kim, J.R. Dynamic Complex Network Analysis of $PM_{2.5}$ Concentrations in the UK, Using Hierarchical Directed Graphs (V1.0.0). *Sustainability* **2021**, *13*, 2201. https://doi.org/10.3390/su13042201

Received: 17 November 2020
Accepted: 22 January 2021
Published: 18 February 2021

1. Introduction

1.1. Background and Rationale

Both local emissions (by both stationary and mobile sources) and regional transport processes can be the source of atmospheric particulate matter. It is demanding and challenging to achieve causal inference between primary (emitted directly by the emission sources) and secondary (originated from the transformation of gaseous pollutants in the atmosphere). For instance, not only do sources such as road traffic result in the magnitude of anthropogenic PM emissions and give rise to $PM_{2.5}$ formation [1,2], but also meteorological conditions can impact $PM_{2.5}$ concentrations via deposition and dispersion. Overcoming the high dimensionality challenge and compressing the concentration data into a two-dimensional (2D) network are necessities, owing to the high data complexity and dimensionality resulted from the contribution of atmospheric chemistry transport processes and a range of emission sources in ambient $PM_{2.5}$ concentrations. Current and future caps are set on anthropogenic emissions of primary- and secondary-precursor components of $PM_{2.5}$ at national scale and from individual sources by European legislation [3].

Moreover, it is widely acknowledged that ambient PM originates from both transport and transboundary emissions [3], so developing efficacious mitigation emissions at local scale would be challenging [3–6].

A great number of studies have made use of a variety of techniques to tackle challenges in computational performance of chemical transport models. Solving atmospheric chemical kinetics is a stiff numerical problem, with choice of solvers used reflecting the need to ensure numerical stability [7]. As a result, the main componential cost of atmospheric chemistry model (e.g., GEOS-Chem) belongs to chemical kinetics integration (50–90%) [8–10]. Dynamical reduction (adaptive solvers) in solving the chemical mechanism was previously demonstrated to increase the efficiency of the integration at the expense of a reduction in accuracy [11]. Other attempts to reduce the computationally of chemical kinetics include repro-modeling (utilizing polynomial functions to approximate the chemical kinetics) [12], approximation of quasi-steady state [13], and separation process of slow and fast species [14]. Other studies use reduced chemical mechanisms with fewer species [13,15].

Recent attempts have also used machine learning to replace the use of traditional integrators [16]. For example, using a neural network emulator for an atmospheric chemistry box model, an order-of-magnitude speed-up was found, but once applied over several time intervals, the new implementation suffered from fast error reproduction [15]. Numerical emulators are capable in directly forecasting the air pollution levels across future time intervals [17]. This approach was also applied in chemistry–climate simulations with the focus on model which can forecast the mean concentrations of special species, such as hydroxyl radical (OH), and ozone (O_3) over several time intervals [18,19]. The replacement of random forest regression as a suitably trained machine-learning-based approach was investigated by Keller and Evans (2019) for the gas-phase chemistry in atmospheric chemistry transport models such as GEOS-Chem [20]. As noted within this particular study, this approach suffers also from some limitations, including (a) being only applicable within the data range used for the training, (b) studying scenarios with significant changes in the emissions (being outside of used data for the training) can lead to inaccurate predictions by the model, and (c) algorithm of machine learning might not capture model resolution caused by the non-linearity of chemistry [20].

The chemical transport models require emission inventory data (local or regionally originated) and a meteorological core to predict the dispersion and deposition of pollutants such as $PM_{2.5}$. Besides the notable amount of required data, high-performance computing (HPC) platforms are required to deploy and evaluate model outputs, not least including experience with the pre- and post-processing software environments. As a result, recent attempts have also tried to investigate the feasibility of machine learning in studying the spatiotemporal distribution of air pollutants such as $PM_{2.5}$. In previous studies, to overcome such uncertainty associated with chemistry-transport models (CTMs) in air-quality models, different techniques, such as a three-layer feedforward neural network (FNN), multiple additive regression trees (MART), and a deep feedforward neural network (DFNN), were deployed [21–26]. According to their studies, machine learning has a great capacity to improve air-quality forecasts, such as estimations with satellite data, including data gap filling, prediction algorithms, and source estimations. However, its applicability in air-quality forecasting suffers from some limitations, including (a) the existence of a data gap for comprehensive investigations, preventing breakthroughs in disruptive predicting performance improvements; (b) diverse learning algorithms and deep networks, as well as comprehensive datasets, are required to be compared for extracting the best complex nonlinear features; and (c) machine-learning techniques for air-quality predictions are limited in dealing with pollutant observations from ground-level monitoring networks and meteorological parameters [24,27,28].

1.2. Importance and Impact

Atmospheric particulate matter influences human health [29,30] and climate change through radiative forcing [31]. The burden of global health from exposure to ground-level $PM_{2.5}$ is considerable. Based on the Global Burden of Disease project in 2005, being exposed to ambient $PM_{2.5}$ concentrations caused 76 million disability-adjusted life years and 3.2 million premature deaths [3,32]. In Europe, exposure to ambient $PM_{2.5}$ is still considered a burning health issue. From 2010 to 2012, the European Environment Agency reported that 10–14% of the urban people in the EU28 countries were exposed to $PM_{2.5}$ more than the EU annual-mean $PM_{2.5}$ reference value (25 μg m^{-3}), while 91–93% were exposed to concentrations more than the WHO annual-mean $PM_{2.5}$ (10 μg m^{-3}) [33,34]. In the United Kingdom, around 29,000 deaths a year can be attributed to long-term exposure to PM from anthropogenic sources [35,36]. Moreover, the daily emergency hospital admissions (for respiratory and cardiovascular conditions) and mortality can be escalated because of acute exposure to air pollution events [37]. Focusing on two air pollution events (March–April 2014) with the highest $PM_{2.5}$ concentrations, about 600 casualties were recorded due to the acute $PM_{2.5}$ exposure (3.9% of total all-cause death) during these 10 days of air pollution event.

It is complicated to meet the standards focused on $PM_{2.5}$ by the notable chemical heterogeneity. PM long-term exposure has been recognized as more critical than the short-term (daily) exposure to higher PM concentrations that was first attributed to effects on human health [38,39]. The foundation has been laid for calculation of health effects from exposure in Europe and the United Kingdom by long-term impact studies, which are crucial and significant [35]. Changes in legislation stemmed from changes in the direction of studies towards $PM_{2.5}$, related to the evidence that long-term PM levels play a major role, in company with short-term peaks, regarding health outcomes [40,41].

1.3. Modeling Challenges

Challenges related to conventional modeling of PM evolution to infer local and regional impacts involve the necessity to embed a range of emission sources, chemical complexity, and transformative processes in Eularian models. In this pioneering study, for the first time, the potential for compressing ambient $PM_{2.5}$ network data into two-dimensional (2D) network was explored, creating a straightforward graph to infer causality and stability. A well-timed study and strategic investments in local and national air-quality monitoring networks necessitate an evaluation on the usefulness, or not, of network design. Even though a sparse distributed network was focused on in this study, future applications for local networks across cities are also discussed. For instance, in a graph, each node in the graph represents a city, which indicates a temporal signal ($PM_{2.5}$) and is linked to other cities, if they indicate a strong association with regard to either correlation (undirected) or Granger causality (directed). It is also possible that we can understand how UK cities cross-pollute across regional and national distances.

2. Materials and Methods

2.1. Ground-Level $PM_{2.5}$ Data

$PM_{2.5}$ concentrations were monitored hourly, at 15 monitoring stations, in various cities in the UK, from the Automatic Urban and Rural Network (AURN) (https://uk-air.defra.gov.uk/data/openair) [42], as illustrated in Figure 1, and coordinates given in SI–List S1 (Supplementary Materials Table S1). The study period was divided into four seasons (meteorological seasons) (Supplementary Materials Table S2). The validity of the data was checked before averaging the $PM_{2.5}$ concentration. Only valid data for 20 h a day were averaged, representing the daily $PM_{2.5}$ concentration, and stations with valid data above 85% were chosen to study in this work.

Figure 1. Studied stations in the UK.

2.2. Calculation of Cross-Correlation for Spatial Distribution of Ambient $PM_{2.5}$ in the United Kingdom

In the present study, the hourly based cross-correlation (XCROSS) using PAST (PAleontological Statistics) [43–46] version 3.25, for all site pairs (106 pair of cities) in four seasonal windows (spring, summer, autumn, and winter), was calculated in order to analyze and investigate the similarity of $PM_{2.5}$ concentration time series between each pair of cities. These intervals were selected to capture and examine the impact of seasonal variations on the calculated similarity among ambient $PM_{2.5}$ concentrations. Based on a previous similar study conducted in Switzerland to characterize the spatial distribution and seasonal changes of PM_{10} and $PM_{2.5}$ concentrations, using long-term monitoring data [34], we decided to choose 70% as our threshold cross-correlation.

2.3. Calculating Granger Causality in Ambient $PM_{2.5}$ Network in the United Kingdom

The Granger causality test statistically ascertains if one time series can cause the other. Therefore, it is used to perceive whether or not previous values of a time series hold the information about the future values of another time series. This method was implemented (employing Eviews, version 11) [47] to each pair of cities in the network, throughout various seasons. Following this, statistically significant results ($p < 0.05$) were used to determine which time series contain information about the future values of another.

Both x and y time series (x and y represent $PM_{2.5}$ concentration series for various stations in our network) are presumed to be stationary, which was not the case in this study. Consequently, de-trending was firstly implemented before applying the Granger causality test [48,49].

To maintain the same degree of freedom (DF) (mathematically, DF symbolizes the number of dimensions of the domain of a random vector, or the number of components that should be perceived before the vector is completely established), with yearly data, the lag number is typically small (1 or 2 lags). The suitable lag number is 1 to 8 for quarterly data (in our case). In the case of having monthly data, 6, 12, or 24 lags are used, given there are enough data points. The number of lags is crucial, since a different number of lags results in different test results. Consequently, the optimal lag number of 7 ensures the stability of the model in this case study (based on Akaike Information Criterion (AIC)).

There is feasibility of causation in one or both directions (x Granger-causes y, and y Granger causes x). The lowest p-value was considered as a basis to choose direction. For instance, based on our analysis, we infer that "activities" in Manchester in spring are statistically impacting on concentrations measured in Preston with a p-value= 5×10^{-29}, whereas Preston is statistically impacting on Manchester with a p-value= 3×10^{-8}. Thus, the first statement aforementioned (pollution from Manchester is impacting on Preston's concentrations) is the correct one to be selected, owing to its lower p-value. The language chosen reflecting the statistical inference for the network analysis is worth being noted. In addition, the importance of mapping of inference to atmospheric behavior and known challenges around $PM_{2.5}$ source apportionment that were discussed should not be ignored.

2.4. Trophic Coherence

Trophic coherence is defined as a way of labeling the hierarchical levels (trophic levels, as derived from food webs and predation levels) and hierarchically restructuring a directed network. Trophic levels have been depicted to be an efficacious compressed metric to infer stability on extensive directed networks with no precise input–output definition. The bottom (basal) nodes represent where all energy originates (e.g., main source of pollution), and the coherence of the entire network is considered a proxy for stability against disturbances. The trophic level (s_i) of node i represents the mean trophic level of its in-neighbors:

$$s_i = 1 + \frac{1}{k_i^{in}} \sum_j a_{ij} s_j \quad (1)$$

where $k_i^{in} = \sum_j a_{ij}$ is the number of in-neighbors of the node i, and a_{ij} is the adjacency matrix of the graph. Basal nodes k_i^{in} have trophic level $s_i = 1$ by convention [50]. In this study, the initial stage was introducing basal nodes in order to interpret trophic coherence in a directed causal network.

Based on this definition, stations with a high trophic level are receptors, while stations with a low trophic level are $PM_{2.5}$ sources. The trophic level of a station is the average level of all the stations from which it receives $PM_{2.5}$ pollutant plus 1. The associated trophic difference of each edge equals $x_{ij} = s_i - s_j$. Generally, p(x) (the distribution of trophic differences) has a mean value of 1, and the variance of this distribution is smaller when the network is more trophically coherent. The measurement of the trophic coherence of network is the incoherence parameter q, which is the standard deviation of p(x):

$$q = \sqrt{\frac{1}{L} \sum_{ij} a_{ij} x_{ij}^2 - 1}, \quad (2)$$

where $L = \sum_{ij} a_{ij}$ is the edges (the number of connections) between the nodes (stations) in the network. Q with the values of more than 0 demonstrates less coherent networks. However, when $q = 0$, the network is perfectly coherent.

3. Results

3.1. Spatial Distribution of $PM_{2.5}$ across the United Kingdom

While analyzing the cross correlation of the hourly values between the different sites, appealing information about the spatial distribution of the $PM_{2.5}$ concentrations across the United Kingdom can be collected. The first group (northern, Group A) includes Preston (Pre), Birmingham (Bir), Newcastle (New), Liverpool (Liv), Nottingham (Not), Chesterfield (Chest), Manchester (Man), and Leeds, whilst the second one (southern, Group B) includes Southampton (South), Bristol (Bri), Plymouth (Ply), Oxford (Oxf), Norwich (Nor), and London (two stations named LonB and LonR). For spring, summer, and autumn (Fall), the value of XCROSS changes, but the combination of groups does not (Figure 2). In winter, the combination of cities in and out of clusters changes (Figure 2D). Figure 2 visualizes the connected cities seasonally, generating a directed dynamic network.

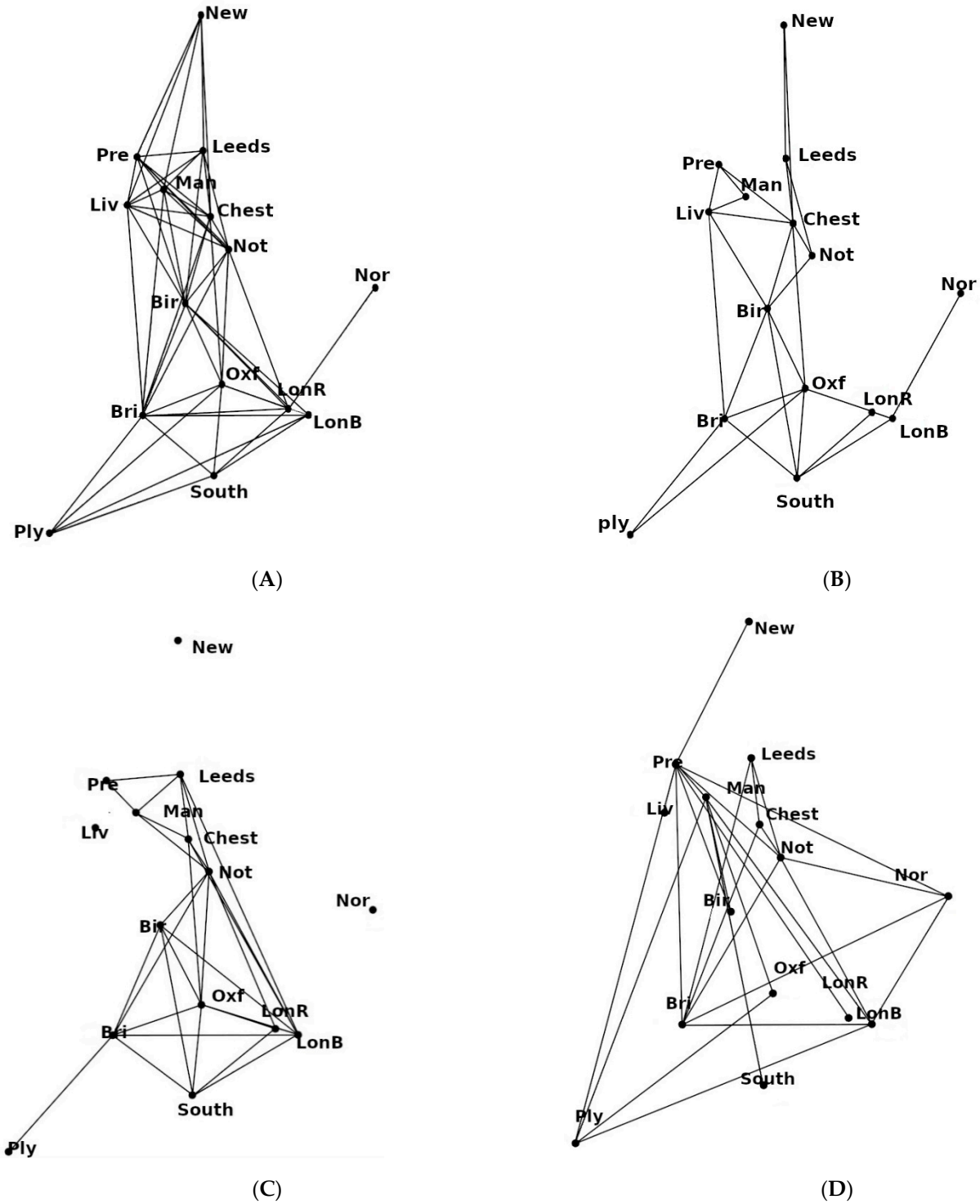

Figure 2. Cross-correlation-based dynamic network including (**A**) spring window, (**B**) summer window, (**C**) autumn window, and (**D**) winter window, in 2017–2018, in the UK.

As the networks are very spatial (i.e., distance is considered a major impedance factor), a general measure was studied on how it is spatially embedded. The pairs of stations were split up into groups, according to the distance (Table 1). To measure the level of spatial embeddedness, a relationship between cross-correlation and distance between each pair of cities was analyzed and studied (Table 1). Less spatial embeddedness of network was observed when the distance rose to over 200 km (for all seasons), whilst a very high spatially embedded part of the network was formed below 100 km for all seasons. A

major part of the network (100 km) was formed in cluster A, as follows: 89%, 67%, 54%, and 60% during winter, spring, summer, and autumn, respectively. This value in cluster A declined (for all seasons) when the distance increased between pairs of cities, reaching the value of zero during winter and autumn. As the distance between cities in cluster A was predominantly over 100 km, the significant part of the network in cluster B was formed below 200 km (100–200 km), with the percentages as follows: 23%, 38%, 52%, and 46% during winter, spring, summer, and autumn, respectively. This value in cluster B reduced (for all seasons) when the distance increased between pairs of cities, reaching the value of zero during autumn, whilst it was 19% for the distance over 200 km during winter. When the distance between cities was over 200 km, the number of outliers (pairs of connected cities out of Groups A and B) reached its highest values of 81%, 40%, and 100% during winter, spring, and autumn, respectively. When the distance was below 100 km (the same decreasing trend was observed in both groups), the number of paired cities in the network reduced by 50% between winter and spring. The network was weakened by 50% for distances below 200 km. Intriguingly, the network was boosted by 17%, compared to spring, during winter, when the distance between cities rose to above 200 km.

Table 1. The relationship between cross-correlation (XCROSS) of the daily values of $PM_{2.5}$ and distance of the cities in UK.

Distance	Pair of Connected Cities in Network	Pair of Connected Cities in Group A	Pair of Connected Cities in Group B	Outliers (Pair of Connected Cities Out of Groups)
		Spring		
<100 km	18 (43%)	12 (67%)	6 (33%)	0
<200 km	42 (81%)	24 (57%)	16 (38%)	2 (5%)
>200 km	10 (19%)	3 (30%)	3 (3%)	4 (40%)
		Summer		
<100 km	13 (52%)	7 (54%)	6 (46%)	0
<200 km	25 (90%)	12 (48%)	13 (52%)	0
>200 km	3 (10%)	2 (67%)	1 (33%)	0
		Autumn		
<100 km	15 (54%)	9 (60%)	6 (40%)	0
<200 km	28 (93%)	9 (27%)	13 (46%)	9 (27%)
>200 km	2 (7%)	0	0	2 (100%)
		Winter		
<100 km	9 (35%)	8 (89%)	1 (11%)	0
<200 km	26 (41%)	14 (54%)	6 (23%)	6 (23%)
>200 km	37 (59%)	0	7 (19%)	30 (81%)

3.2. Granger Causality Test

The major result achieved in this study shows that cities with the strongest cross-correlation have the lowest *p*-value (below 5%) (Figure 3). As previously mentioned, in spring, results statistically indicate that activity in Manchester is causing concentrations to change in Preston with *p*-value= 5×10^{-29} (i.e., Manchester past $PM_{2.5}$ data provide information about the future $PM_{2.5}$ values of Preston), and Bristol is causing change in Oxford with a *p*-value of 9×10^{-28}. Liverpool is causing change in Preston with a *p*-value of 7×10^{-17} in summer. Chesterfield is causing change in Nottingham with a *p*-value of 1×10^{-7} in winter, whilst Manchester is causing change in Preston with *p*-value = 6×10^{-23} in autumn. The results appear very spatial, and the distance is a crucial impedance factor. The distance between all paired cities was below 50 km. According to Table 2, the order of *p*-value rises when the distance between a pair of cities rises.

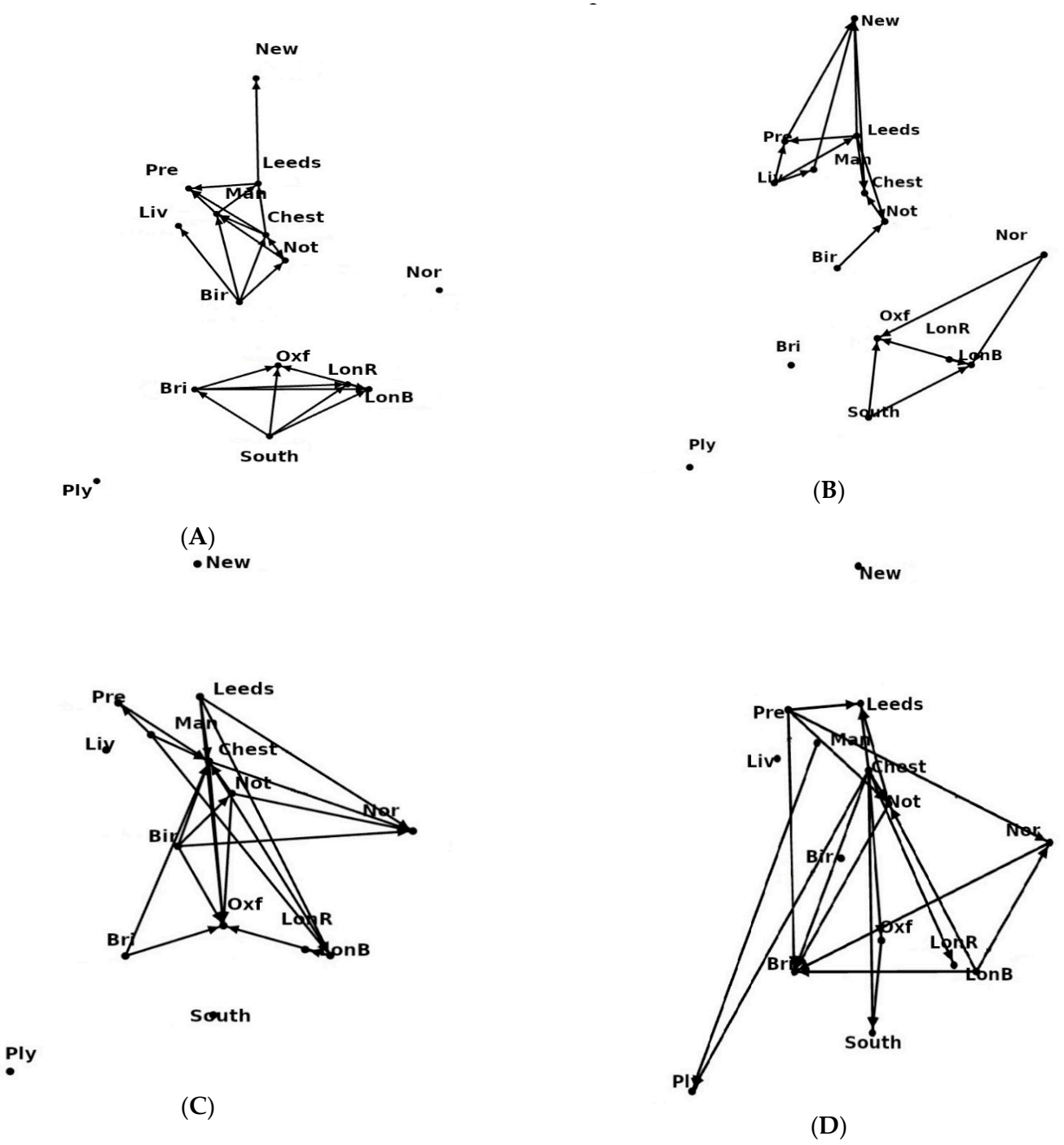

Figure 3. Granger-based dynamic network, including (**A**) spring window, (**B**) summer window, (**C**) autumn window, and (**D**) winter window, in 2017–2018, in the UK.

In an ordered pair, $G = (N, E)$, or a directed graph [51], N is a set of nodes (i.e., stations) and E is a set of ordered pairs of nodes, called edges (i.e., the probability values for F statistics). The hierarchical structure of a directed graph can be shown by its trophic coherence property. The general idea is that hierarchical systems enjoy fewer feedback loops and less cascade impacts. To measure the coherence of the seasonal causal network to demonstrate how trophic distance is firmly associated with edges concentrated around its mean value (which is always 1), the incoherence parameter (q) was employed [52]. An incoherent network in our seasonal datasets can be seen in Table 3.

Table 2. Comparison among Granger causality results (*p*-values) in different seasons.

Source	Target	Distance (km)	*p*-Value
Spring			
Manchester	Preston	43.66	5×10^{-29}
Bristol	Oxford	91.78	9×10^{-28}
Summer			
Liverpool	Preston	42.62	7×10^{-17}
Leeds	Newcastle	131	5×10^{-11}
Autumn			
Manchester	Preston	43.66	6×10^{-23}
Chesterfield	Oxford	165.11	3×10^{-20}
Winter			
Chesterfield	Nottingham	36.17	1×10^{-7}
Chesterfield	Bristol	213.74	7×10^{-6}

Table 3. Incoherence factor of seasonal directed networks in current study.

Directed Network	Incoherence Factor (q)
Spring	0.69
Summer	0.37
Autumn	0.49
Winter	0.35

It worth mentioning that, if we had perfect coherence (q = 0), then there would be a source of pollution that is affecting others. If we had perfect incoherence (q = 1), then all the cities would be polluting each other equally (Table 3). This gives us an idea of both the nature and geography of the transport ecosystem for different seasons, as well as its stability. The bottom (basal) nodes represent where all energy originates (e.g., main source of pollution in terms of $PM_{2.5}$ data), and the coherence of the entire network is considered to be a proxy for stability against disturbances (incoherence parameter (q)) [52]. Based on this definition, stations with a high trophic level are $PM_{2.5}$ receptors, while stations with a low trophic level are $PM_{2.5}$ sources (Figure 4).

Spring was known as the highly incoherent season, with q = 0.69, while winter was a less incoherent network (q = 0.35). In Figure 4, based on the parameter definition, the basal nodes with the low trophic level stand for the major pollution source nodes, whilst stations that act as receptors in the causal network are the ones with high trophic levels. During spring, the network was well formed, owing to well mixing of the lower atmospheric layer. In Group A, Birmingham was known as a pollution source with low trophic level, whilst in Group B, Southampton represented a pollution source with low trophic level.

Table 3 shows a similar incoherence factor for winter and summer. With a q value of 0.3–0.4, this suggests having similar stability but different source of pollution [50]. The summer and winter periods have similar values but different sources. Figure 4B (summer) suggests that the sources of the network are Liverpool, London Road, Southampton, Norwich, and Birmingham. Meanwhile, in winter (Figure 4D), the sources of the network are Chesterfield, Manchester, Preston, and London Bexely. Winter is also inferred to be represented as a national network, while summer is more local.

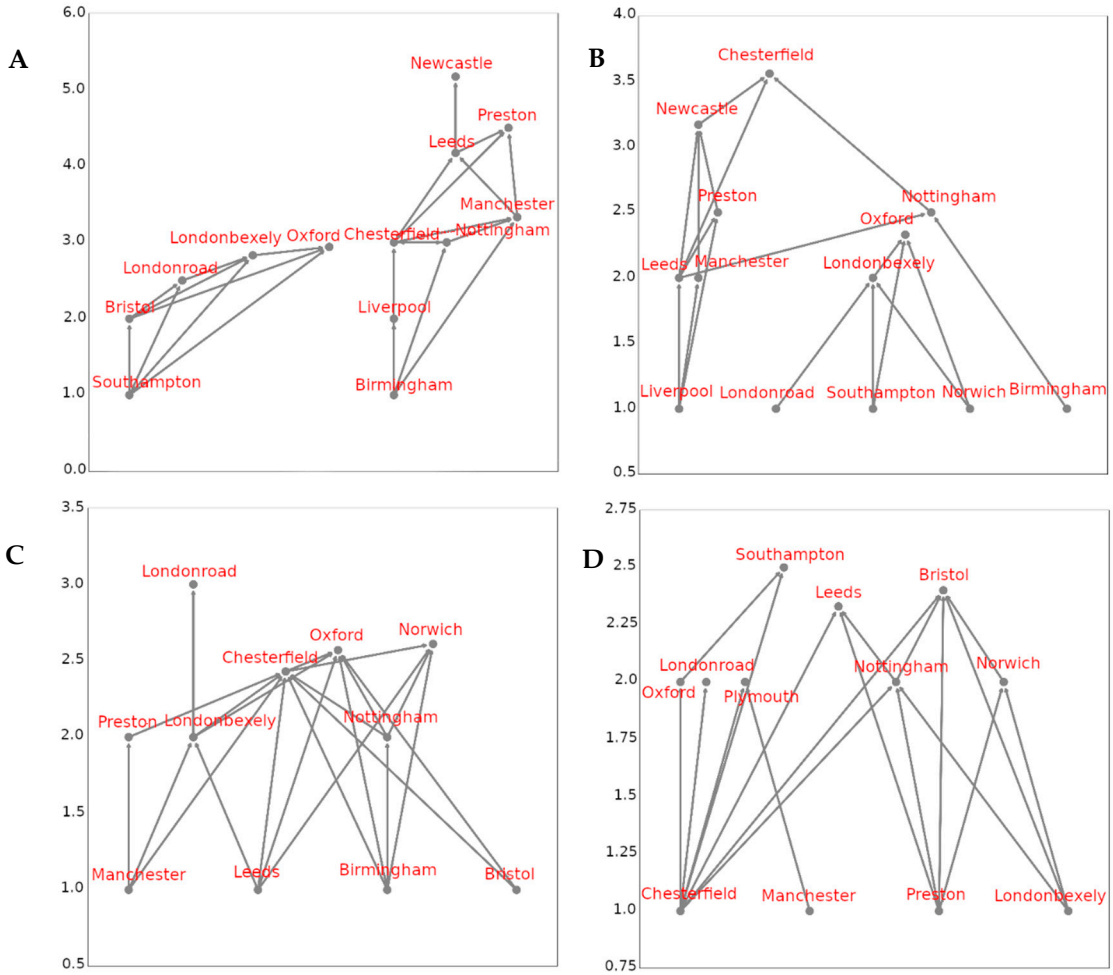

Figure 4. The hierarchical structure and basal nodes of causal network, including (**A**) spring window, (**B**) summer window, (**C**) autumn window, and (**D**) winter window, in 2017–2018, in the UK.

4. Discussion

The Impact of Meteorological Parameters on Network Structure

According to the prior analysis, this connection (network) demonstrates that meteorological conditions and diurnal emissions from a broad range of common sources (i.e., traffic), rather than locally specific sources and events, dominate the relative variations of the concentrations of fine particulate matter for lengthy periods [34]. The meteorology, during wintertime, is marked by forming an efficient hindrance for the distribution, frequent inversions, and homogenization of PM. Therefore, only firm spatially embedded parts of network (below 100 km with the highest percentage of restored network) could "withstand" meteorological impacts and further parts (over 100 km) commenced to collapse from a network perspective. In winter, the compelling reason of connecting the cities out of the initial network (81% of connected cities were out of the initial network with distance over 200 km) might be the higher average seasonal wind speeds (based on all studied stations), presumably owing to the balance among further dilution and shorter transport times at higher wind speeds, which takes less time for PM dispersion and deposition over longer distances [38].

In fact, it is a well-known fact that $PM_{2.5}$ concentrations and formation mechanisms can be substantially influenced by changes in weather condition (e.g., wind speed, wind direction, rainfall, and temperature) [1,3]. Besides primary sources, secondary sources are contingent on the abundance of precursors and meteorological conditions. Secondary aerosols have a major role in $PM_{2.5}$ concentrations in the United Kingdom, where a signifi-

cant proportion of transboundary secondary $PM_{2.5}$ is made of nitrate particles in the form of ammonium nitrate transferred from Europe [1,3]. Common transboundary sources can be one compelling reason for connection within a network.

The association among wind direction and $PM_{2.5}$ can provide a better picture of the origins of the measured $PM_{2.5}$ concentrations. Taking this into an account, there is a remarkable coherence throughout the patterns across the Group A and Group B in the United Kingdom. As a result, there is a slight change between cities in the south (Group B) and those in the north or in the vicinity of northern part (Group A) of the United Kingdom [38]. High $PM_{2.5}$ concentrations in Group B (southern sites) are more attributable to winds from the east through to southeast, which are frequently attributable to a blocking high pressure over the Nordic countries, leading to a south-easterly or easterly air flow that results in the transportation of emissions from the east of Europe, north of Germany, and Belgium and the Netherlands to the southern cities in the United Kingdom [38,53].

Besides this, high $PM_{2.5}$ concentrations in Group A (northern cities or in the vicinity of northern part) are more paramount attributable to the winds blowing from the northeast through to east, drawing air flow (probably to commence blowing when a low pressure goes up the English Channel) northward across European emission sources (to chiefly be emission sources of precursors of secondary PM), out into the North Sea, then stretching toward northern parts of the United Kingdom from a north-easterly direction [53].

The general framing of our approach is at the national level, trying to demonstrate (via a data-driven correlation and causal network) the statistical relationship between data from multiple cities. This data-driven low-dimensional network enables us to examine seasonal trends and infer root causal mechanisms. We believe this approach requires evaluations across multiple scales. Nonetheless, we believe this approach will offer an additional approach to traditional models where inference of causality remains challenging. Of course, what our model lacks is the relationship back to the physical flow models, and our future work will incorporate this. Machine-learning models are used to predict, but we are here to infer causality and demonstrate topological patterns via the network.

5. Conclusions

In this study, $PM_{2.5}$ concentrations were applied in 14 cities in the United Kingdom, for one year, to deduce an undirected correlation and a directed Granger causality network. Both network cases (Groups A and B) were shown, with two robust spatial communities split up the United Kingdom into the northern and southern city clusters, with more spatial embedding in summer and spring.

According to the Granger causality test, it is inferred that $PM_{2.5}$ data of cities with the most significant cross-correlation (having the lowest p-value) provide information about the future $PM_{2.5}$ values in the network. On the other hand, there are certainly several caveats with this statement, some of which are mentioned in the discussions around known impacts from meteorological and source variability. The directed network was used to infer stability to disturbances through the trophic coherence parameter, whereby it was found that winter had the highest vulnerability.

As previously mentioned, this connection (network) indicates that relative variations of the urban background $PM_{2.5}$ concentrations [34] applying this sparse network data were dominated by meteorological conditions and emissions from regional origins in lieu of specific local origins and events. It is known that PM with emission sources from continental Europe, presumably as secondary PM, can take a major role in impacting $PM_{2.5}$ levels in different parts of the United Kingdom [38]. However, this study suffers from some limitations, such as a short period of time over which the network was analyzed. In addition, to gain a better understanding of network, assessing a predictive network-based $PM_{2.5}$ model utilizing meteorological parameters, and contributions from established clusters in the United Kingdom, would be beneficial. This work has the role of indicator for information that can be derived from an undirected correlation and a directed Granger causality network. Further work is required to be done, in parallel with additional data that

likely support the extracted relationships, i.e., source apportionment data and transport activity. The approach may be well suited to more local networks, as well, i.e., monitoring stations throughout a city.

Code availability: The code required to compute the trophic level of each node in the network, the trophic difference, and, lastly, the trophic coherence (q) of the network, with all required scripts to replicate the results in the current study, is available at https://github.com/kohyar88/PM2.5--Trophic--Coherence-/tree/v1.0.0, with the DOI number 10.5281/zenedo.3661483.

Supplementary Materials: The following are available online at https://www.mdpi.com/2071-1050/13/4/2201/s1, Table S1: Fifteen monitoring stations in different cities (from UK Air Defra dataset website) shown in Figure 1 and coordinates. Table S2: Time-split of the meteorological seasons during the studied period.

Author Contributions: Conceptualization, methodology, software, data curation, and writing—original draft preparation, P.B.; conceptualization, methodology, and writing—reviewing and editing, W.G.; writing—reviewing and editing, D.T.; software, A.P.; conceptualization and project administration, X.G.; conceptualization, supervision and project administration, J.R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This project obtained financial support from European Union's Horizon 2020 research and innovation programme Marie Skłodowska-Curie Actions Research and Innovation Staff Exchange (RISE), under grant agreement No. 778360 and NU project (Nazarbayev Research Fund SOE2017003).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

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

References

1. AQEG: Mitigation of United Kingdom PM2.5 Concentrations, Air Quality Expert Group, UK Department for Environment, Food and Rural Affairs, London, UK. PB13837. Available online: http://ukair.defra.gov.uk/assets/documents/reports/cat11/1508060903_DEF-PB14161_Mitigation_of_UK_PM2.5.pdf (accessed on 1 November 2015).
2. Munir, S. Analysing temporal trends in the ratios of PM2.5/PM10 in the UK. *Aerosol Air Qual. Res.* **2016**, *17*, 34–48. [CrossRef]
3. Vieno, M.; Heal, M.; Williams, M.; Carnell, E.; Nemitz, E.; Stedman, J.; Reis, S. The sensitivities of emissions reductions for the mitigation of UK PM2.5. *Atmos. Chem. Phys.* **2016**, *16*, 265–276. [CrossRef]
4. Donkelaar, A.; Martin, R.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P. Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [CrossRef] [PubMed]
5. Meerschman, E.; Cockx, L.; Islam, M.M.; Meeuws, F.; Van Meirvenne, M. Geostatistical Assessment of the Impact of World War I on the Spatial Occurrence of Soil Heavy Metals. *AMBIO* **2011**, *40*, 417–424. [CrossRef] [PubMed]
6. Zhang, Y.H.; Hu, M.; Zhong, L.J.; Wiedensohler, A.; Liu, S.C.; Andreae, M.; Wang, W.; Fan, S. Regional integrated experiments on air quality over Pearl River Delta 2004 (PRIDE-PRD2004): Overview. *Atmos. Environ.* **2008**, *42*, 6157–6173. [CrossRef]
7. Sandu, A.; Sander, R. Technical Note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1. *Atmos. Chem. Phys.* **2006**, *5*. [CrossRef]
8. Eastham, S.; Long, M.; Keller, C.; Lundgren, E.; Yantosca, R.; Zhuang, J.; Li, C.; Lee, C.; Yannetti, M.; Auer, B.; et al. GEOS-Chem High Performance (GCHP v11-02c): A next-generation implementation of the GEOS-Chem chemical transport model for massively parallel applications. *Geosci. Model. Dev.* **2018**, *11*, 2941–2953. [CrossRef]
9. Hu, L.; Keller, C.; Long, M.; Sherwen, T.; Auer, B.; Da Silva, A.; Nielsen, J.; Pawson, S.; Thompson, M.; Trayanov, A. Global simulation of tropospheric chemistry at 12.5 km resolution: Performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM). *Geosci. Model. Dev.* **2018**, *11*, 4603–4620. [CrossRef]
10. Nielsen, J.E.; Pawson, S.; Molod, A.; Auer, B.; da Silva, A.M.; Douglass, A.R.; Duncan, B.; Liang, Q.; Manyin, M.; Oman, L.D. Chemical Mechanisms and Their Applications in the Goddard Earth Observing System (GEOS) Earth System Model. *J. Adv. Model. Earth Syst.* **2017**, *9*, 3019–3044. [CrossRef]
11. Cariolle, D.; Moinat, P.; Teyssèdre, H.; Giraud, L.; Josse, B.; Lefèvre, F. ASIS v1.0: An adaptive solver for the simulation of atmospheric chemistry. *Geosci. Model. Dev.* **2017**, *10*, 1467–1485. [CrossRef]
12. Turányi, T. Parameterization of reaction mechanisms using orthonormal polynomials. *Comput. Chem.* **1994**, *18*, 45–54. [CrossRef]

13. Whitehouse, L.E.; Tomlin, A.S.; Pilling, M.J. Systematic reduction of complex tropospheric chemical mechanisms, Part II: Lumping using a time-scale based approach. *Atmos. Chem. Phys.* **2004**, *4*, 2057–2081. [CrossRef]
14. Young, T.R.; Boris, J.P. A numerical technique for solving stiff ordinary differential equations associated with the chemical kinetics of reactive-flow problems. *J. Phys. Chem.* **1997**, *81*, 2424–2427. [CrossRef]
15. Kelp, M.; Tessum, C.; Marshall, J. Orders-of-magnitude speedup in atmospheric chemistry modeling through neural network-based emulation. *arXiv* **2018**, arXiv:1808.03874.
16. Porumbel, I.; Petcu, A.C.; Florean, F.G.; Hritcu, C.E. Artificial neural networks for modeling of chemical source terms in CFD simulations of turbulent reactive flows. In *Modeling and Optimization of the Aerospace, Robotics, Mechatronics, Machines-Tools, Mechanical Engineering and Human Motricity Fields*; Applied Mechanics and Materials; Trans Tech Publications: Stafa-Zurich, Switzerland, 2014; Volume 555, pp. 395–400.
17. Mallet, V.; Stoltz, G.; Mauricette, B. Ozone ensemble forecast with machine learning algorithms. *J. Geophys. Res.* **2009**, *114*. [CrossRef]
18. Nicely, J.; Salawitch, R.; Canty, T.; Anderson, D.; Arnold, S.; Chipperfield, M.; Emmons, L.; Flemming, J.; Huijnen, V.; Kinnison, D.; et al. Quantifying the Causes of Differences in Tropospheric OH within Global Models. *J. Geophys. Res. Atmos.* **2017**, *122*. [CrossRef]
19. Nowack, P.; Braesicke, P.; Haigh, J.; Abraham, N.; Pyle, J.; Voulgarakis, A. Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. *Environ. Res. Lett.* **2018**, *13*. [CrossRef]
20. Keller, C.; Evans, M. Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. *Geosci. Model. Dev.* **2019**, *12*, 1209–1225. [CrossRef]
21. Kamali, N.; Zare Shahne, M.; Arhami, M. Implementing Spectral Decomposition of Time Series Data in Artificial Neural Networks to Predict Air Pollutant Concentrations. *Environ. Eng. Sci.* **2015**, *32*, 379–388. [CrossRef]
22. Memarianfard, M.; Hatami, A.M. Artificial neural network forecast application for fine particulate matter concentration using meteorological data. *Glob. J. Environ. Sci. Manag.* **2017**, *3*, 333–340. [CrossRef]
23. Shamsoddini, A.; Aboodi, M.; Karami, J. Tehran air pollutants prediction based on Random Forest feature selection method. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, 483–488. [CrossRef]
24. Zuo, X.; Guo, H.; Shi, S.; Zhang, X. Comparison of Six Machine Learning Methods for Estimating $PM_{2.5}$ Concentration Using the Himawari-8 Aerosol Optical Depth. *J. Indian Soc. Remote Sens.* **2018**, *48*, 1277–1287. [CrossRef]
25. Karimian, H.; Li, Q.; Wu, C.; Qi, Y.; Mo, Y.; Chen, G.; Sachdeva, S.; Zhang, X. Evaluation of Different Machine Learning Approaches in Forecasting $PM_{2.5}$ Mass Concentrations. *Aerosol Air Qual. Res.* **2019**, *19*, 1400–1410. [CrossRef]
26. Shafiee, M.; Feghhi, S.A.H.; Rahighi, J. Experimental performance evaluation of ILSF BPM data acquisition system. *Measurement* **2017**, *100*, 205–212. [CrossRef]
27. Chang, S.; Chang, C.L.; Li, L.T.; Liao, S. Reinforcement Learning for Improving the Accuracy of $PM_{2.5}$ Pollution Forecast Under the Neural Network Framework. *IEEE Access* **2019**, *1*, 9864–9874. [CrossRef]
28. Zhao, R.; Gu, X.; Xue, B.; Zhang, J.; Ren, W. Short period $PM_{2.5}$ prediction based on multivariate linear regression model. *PLoS ONE* **2018**, *13*, e0201011. [CrossRef]
29. Festy, B. *Review of Evidence on Health Aspects of Air Pollution—REVIHAAP Project*; Technical Report; World Health Organization Regional Office for Europe: Copenhagen, Denmark, 2013.
30. WHO. Air quality guidelines, global update 2005. In *Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide*; World Health Organization Regional Office for Europe: Copenhagen, Denmark, 2006; Available online: http://www.euro.who.int/_data/assets/pdf_file/0005/78638/E90038.pdf (accessed on 1 January 2015).
31. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA; 1535p.
32. Lim, S.; Vos, T.; Flaxman, A.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; Al-Mazroa, M.; Amann, M.; Anderson, H.; Andrews, K.; et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2224–2260. [CrossRef]
33. EEA. *Air Quality in Europe—2014 Report*; EEA Report No. 5/2014; European Environment Agency, Publications Office of the European Union: Copenhagen, Denmark, 2014; Available online: http://www.eea.europa.eu/publications/air-quality-in-europe-2014 (accessed on 1 January 2014).
34. Gehrig, R.; Buchmann, B. Characterising seasonal variations and spatial distribution of ambient PM10 and PM2.5 concentrations based on long-term Swiss monitoring data. *Atmos. Environ.* **2003**, *37*, 2571–2580. [CrossRef]
35. COMEAP. *The Mortality Effects of Long-Term Exposure to Particulate Air Pollution in the United Kingdom*; Department of Health Committee on the Medical Effects of Air Pollution: Chilton, UK, 2010; ISBN 978-0-85951-685-3. Available online: http://comeap.org.uk/documents/reports.html (accessed on 1 December 2014).
36. Gowers, A.M.; Miller, B.G.; Stedman, J.R. *Estimating Local Mortality Burdens Associated with Particulate Air Pollution*; Public Health England: Oxfordshire, UK, 2014.
37. Macintyre, H.L.; Heaviside, C.; Neal, L.S.; Agnew, P.; Thornes, J.; Vardoulakis, S. Mortality and emergency hospitalizations associated with atmospheric particulate matter episodes across the UK in spring 2014. *Environ. Int.* **2016**, *97*, 108–116. [CrossRef]
38. Harrison, R.M.; Laxen, D.; Moorcroft, S.; Laxen, K. Processes affecting concentrations of fine particulate matter ($PM_{2.5}$) in the UK atmosphere. *Atmos. Environ.* **2012**, *46*, 115–124. [CrossRef]

39. Pope, C.A.; Dockery, D.W. Health Effects of Fine Particulate Air Pollution: Lines that Connect. *J. Air Waste Manag. Assoc.* **2006**, *56*, 709–742. [CrossRef] [PubMed]
40. Defra. *The Air Quality Strategy for England, Scotland, Wales and Northern Ireland*; Defra: London, UK, 2007. Available online: www.defra.gov.uk/environment/quality/air/airquality/strategy/index.htm (accessed on 1 July 2007).
41. Official Journal. Directive 2008/50/EC of the European Parliament and the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal of the European Union L152, 1e44. Available online: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF.11.6.2008 (accessed on 1 May 2008).
42. DEFRA. *Air Pollution in the UK 2011*; Department for Environment Food and Rural: London, UK, 2012. Available online: http://uk-air.defra.gov.uk/assets/documents/annualreport/air_pollution_uk_2011_issue_2.pdf (accessed on 1 December 2014).
43. Hammer, O.; Harper, D.; Ryan, P. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **2011**, *4*, 1–9.
44. Shafiee, M.; Feghhi, S.A.H.; Rahighi, J. Analysis of de-noising methods to improve the precision of the ILSF BPM electronic readout system. *J. Instrum.* **2016**, *11*, 11. [CrossRef]
45. Shafiee, M.; Feghhi, S.A.H.; Rahighi, J. Numerical Analysis of the Beam Position Monitor Pickup for the Iranian Light Source Facility. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* **2016**, *847*. [CrossRef]
46. Shafiee, M.; Behnamian, H.; Feghhi, S.A. A study of wake potentials for the pick-ups in storage ring. *J. Instrum.* **2017**, *12*, 12006. [CrossRef]
47. Software, Q.M. *Eviews, Version 11*; IHS Global Inc.: Irvine, CA, USA, 2019.
48. Papagiannopoulou, C.; Decubber, S.; Miralles, D.G.; Demuzere, M.; Verhoest, N.E.C.; Waegeman, W. *Analyzing Granger Causality in Climate Data with Time Series Classification Methods BT—Machine Learning and Knowledge Discovery in Databases*; Altun, Y., Das, K., Mielikäinen, T., Malerba, D., Stefanowski, J., Read, J., Žitnik, M., Eds.; Springer: Cham, Switzerland, 2017; Volume 10536, pp. 15–26.
49. Papagiannopoulou, C.; Miralles, D.; Verhoest, N.; Dorigo, W.; Waegeman, W. A non-linear Granger causality framework to investigate climate–vegetation dynamics. *Geosci. Model. Dev. Discuss.* **2016**, 1–24. [CrossRef]
50. Pagani, A.; Mosquera, G.; Alturki, A.; Johnson, S.; Jarvis, S.; Wilson, A.; Guo, W.; Varga, L. Resilience or robustness: Identifying topological vulnerabilities in rail networks. *R. Soc. Open Sci.* **2019**, *6*, 181301. [CrossRef]
51. Bang-Jensen, J.; Gutin, G.Z. *Digraphs: Theory, Algorithms and Applications*, 2nd ed.; Springer: London, UK, 2008.
52. Johnson, S.; Domínguez-García, V.; Donetti, L.; Muñoz, M. Trophic coherence determines food-web stability. *Proc. Natl. Acad. Sci. USA* **2014**, *111*. [CrossRef]
53. Barry, R.G.; Chorley, R.J. *Atmosphere, Weather and Climate*, 9th ed.; Routledge: Abingdon, UK, 2010; pp. 272–274.